OTORITAS JASA KEUANGAN

TATA KELOLA KECERDASAN ARTIFISIAL PERBANKAN INDONESIA

ARTIFICIAL INTELLIGENCE GOVERNANCE FOR INDONESIAN BANKING

Daftar Isi
Table of Contents

Daftar Tabel
List of Tables

Daftar Gambar
List of Figures

Daftar Grafik
List of Charts

Kata Pengantar
Foreword

Dian Ediana Rae
Kepala Eksekutif Pengawas Perbankan
merangkap Anggota Dewan Komisioner Otoritas Jasa Keuangan

Sambutan Kepala Eksekutif Pengawas Perbankan Merangkap Anggota Dewan Komisioner OJK

Assalaamu'alaikum warahmatullahi wabarakatuh, Salam Sejahtera bagi kita semua, Syalom, Om Swastyastu, Namo Buddhaya, Salam Kebajikan.

Puji dan syukur kami panjatkan kehadirat Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya, Otoritas Jasa Keuangan dapat menyelesaikan penyusunan Tata Kelola Kecerdasan Artifisial Perbankan Indonesia (*Artificial Intelligence Governance for Indonesia Banking*) sebagai panduan bagi perbankan di Indonesia untuk memastikan teknologi kecerdasan artifisial (AI) (termasuk *advanced AI systems*) dikembangkan dan diterapkan secara bertanggung jawab.

Pengembangan dan penerapan sistem kecerdasan artifisial di sektor perbankan disadari berpotensi mentransformasi industri perbankan dengan mendorong inovasi, memberdayakan pengambilan keputusan yang lebih cerdas serta menciptakan pengalaman yang lebih

Foreword from the Chief Executive of Banking Supervision Also Serving as a Member of the OJK Board of Commissioners

Assalamu'alaikum Wr. Wb. Greetings to all of us, Om Swastyastu, Namo Buddhaya, Greetings of Virtue

We praise and express our gratitude to God Almighty, because by His grace and blessings, the Financial Services Authority (Otoritas Jasa Keuangan) has been able to complete the preparation of the Artificial Intelligence Governance for Indonesia Banking as a guideline for banks in Indonesia to ensure that artificial intelligence technology (including advanced AI systems) is developed and implemented responsibly.

The development and implementation of artificial intelligence systems in the banking sector are recognized to have the potential to transform the banking industry by driving innovation, empowering smarter decision-making, and creating more personalized and

personal dan menarik bagi nasabah. Namun demikian, pengembangan dan penerapan sistem kecerdasan artifisial di sektor perbankan dalam berbagai *use cases* harus dilakukan secara bertanggung jawab, agar penerapan kecerdasan artifisial mampu memberikan manfaat yang diharapkan sesuai dengan potensi yang dimilikinya serta dengan pengelolaan risiko yang terkendali, sehingga mampu melindungi nasabah termasuk menjaga stabilitas sistem perbankan serta stabilitas sistem keuangan secara luas. Karenanya, Tata Kelola Kecerdasan Artifisial Perbankan Indonesia ini disusun dengan memperhatikan berbagai regulasi, standar maupun panduan, baik dalam dan luar negeri, serta disesuaikan dengan konteks kebutuhan dan pendekatan bagi sektor perbankan Indonesia dengan tetap mengedepankan aspek kehati-hatian.

Adopsi teknologi pada sektor perbankan termasuk *emerging technology* seperti kecerdasan artifisial harus dilakukan bank melalui serangkaian tahapan dengan memperhatikan berbagai aspek, agar proses adopsi teknologi dapat berjalan lancar dan seluruh potensi risiko yang timbul telah dipertimbangkan dengan baik. Disamping itu, pengembangan dan penerapan kecerdasan artifisial di lingkungan perbankan juga membutuhkan kolaborasi dari seluruh pihak yang terlibat baik dari dalam maupun luar bank (*AI actors* sebagai pihak yang terlibat dalam pengembangan, penerapan, dan pengelolaan sistem di sepanjang siklus hidup kecerdasan artifisial (*AI life cycle*)). Karenanya, dalam dokumen Tata Kelola Kecerdasan Artifisial Perbankan Indonesia ini, pengembangan dan penerapan sistem kecerdasan artifisial di sektor perbankan secara bertanggung jawab dilakukan di sepanjang siklus hidup kecerdasan artifisial, sehingga setiap tahapan yang akan dilakukan bank didasarkan atas pertimbangan kebutuhan dan tujuan yang diharapkan, pengelolaan risiko yang terkendali, mengedepankan aspek kehati-hatian, serta melibatkan kesiapan seluruh sumber daya bank, sehingga sistem kecerdasan artifisial bank mampu beroperasi dalam tingkat kinerja yang dapat dipercaya dan dapat diandalkan.

Dengan mengadopsi kerangka strategis, membangun budaya inovatif, dan mengelola aspek etika, bank dapat memaksimalkan potensi kecerdasan artifisial, mengedepankan pendekatan yang adaptif, serta mengelola sumber daya secara optimal dalam merespons dinamika perubahan. Selain itu, pengalokasian sumber daya yang tepat menjadi kunci agar pengembangan dan

engaging experiences for customers. However, the development and implementation of artificial intelligence systems in various use cases within the banking sector must be carried out responsibly so that AI applications can deliver the expected benefits according to their potential while managing risks effectively, thus ensuring protection for customers, maintains the stability of the banking system, and supports broader financial system stability. Therefore, this Artificial Intelligence Governance for Indonesia Banking is prepared by considering various regulations, standards, and guidelines both domestic and international, tailored to the needs and approaches relevant to Indonesia's banking sector while prioritizing prudence.

Adoption of technology in the banking sector, including emerging technologies such as artificial intelligence, must be carried out by banks through a series of stages that take into account various aspects to ensure the technology adoption process runs smoothly and all potential risks are well considered. In addition, the development and implementation of artificial intelligence within the banking environment also require collaboration from all parties involved, both inside and outside the bank (AI actors as parties involved in the development, implementation, and management of systems throughout the AI life cycle). Therefore, in this Artificial Intelligence Governance for Indonesia Banking document, the responsible development and implementation of artificial intelligence systems in the banking sector are carried out throughout the AI lifecycle, this ensures that every stage undertaken by the bank is based on considerations of needs and expected objectives, controlled risk management, prioritization of prudence, and involvement of all bank resources' readiness so that the AI system can operate at a reliable and trustworthy performance level.

By adopting a strategic framework, fostering an innovative culture, and managing ethical aspects, banks can maximize the potential of artificial intelligence, prioritize an adaptive approach, and optimally manage resources in responding to dynamic changes. Additionally, proper resource allocation is key to ensuring that AI development and implementation align

penerapan AI berjalan sesuai dengan tujuan strategis bank dan tetap selaras dengan ketentuan regulasi yang berlaku.

Tata Kelola Kecerdasan Artifisial Perbankan Indonesia ini juga disusun untuk melengkapi berbagai rangkaian kebijakan akselerasi transformasi digital perbankan yang telah diterbitkan oleh OJK, antara lain Cetak Biru Transformasi Digital Perbankan, POJK No. 11/POJK.03/2022 tentang Penyelenggaraan Teknologi Informasi oleh Bank Umum, SEOJK No. 29/SEOJK.03/2022 tentang Ketahanan dan Keamanan Siber bagi Bank Umum, SEOJK No. 24/SEOJK.03/2023 tentang Penilaian Tingkat Maturitas Digital Bank Umum, dan Panduan Resiliensi Digital (*Digital Resilience*).

Akhir kata, kami menyampaikan apresiasi kepada seluruh pihak yang telah memberikan masukan, komentar, dan saran yang konstruktif dalam penyusunan Tata Kelola Kecerdasan Artifisial Perbankan Indonesia. Harapan kami, dokumen Tata Kelola Kecerdasan Artifisial Perbankan Indonesia ini agar dapat menjadi acuan minimal bagi sektor perbankan dalam mengembangkan dan menerapkan sistem kecerdasan artifisial, mengingat teknologi kecerdasan artifisial akan terus mengalami perkembangan dan tantangan yang dinamis, sehingga perlu untuk saling melengkapi dalam merespon dinamika perubahan dengan berbagai kerangka regulasi, standar, panduan maupun kebijakan lain yang relevan, dengan tetap mengedepankan pengelolaan risiko dan aspek kehati-hatian.

Semoga Tuhan Yang Maha Kuasa senantiasa meridhoi semua niat baik yang ingin kita lakukan dalam upaya untuk mewujudkan perbankan nasional di era digital yang penuh daya saing dan berintegritas, antisipatif terhadap berbagai risiko yang melekat, serta lebih resilien, adaptif dan terus dapat berkontribusi optimal dalam mendukung pertumbuhan perekonomian nasional untuk sebesar-besarnya kesejahteraan masyarakat Indonesia.

Wassalaamu'alaykum warahmatullahi wabarakatuh, Om Shanti Shanti Shanti Om, Namo Buddhaya, Salam kebajikan.

Dian Ediana Rae

Kepala Eksekutif Pengawas Perbankan
Merangkap Anggota Dewan Komisioner
Otoritas Jasa Keuangan

with the bank's strategic objectives and remain compliant with applicable regulations.

This Artificial Intelligence Governance for Indonesia Banking is also developed to complement various policies accelerating digital transformation in banking issued by OJK, including the Digital Transformation Blueprint for Banking, POJK No. 11/POJK.03/2022 concerning the Implementation of Information Technology by Commercial Banks, SEOJK No. 29/SEOJK.03/2022 on Cyber Resilience and Security for Commercial Banks, SEOJK No. 24/SEOJK.03/2023 on the Assessment of Digital Maturity Levels of Commercial Banks, and the Digital Resilience Guidelines.

In closing, we express our appreciation to all parties who have provided input, comments, and constructive suggestions in the preparation of the Artificial Intelligence Governance for Indonesia Banking. It is our hope that this document can serve as a minimum reference for the banking sector in developing and implementing artificial intelligence systems, considering that AI technology will continue to evolve and face dynamic challenges, therefore, it is necessary to complement each other in responding to dynamic changes through various regulatory frameworks, standards, guidelines, and other relevant policies, while still prioritizing risk management and prudence.

May God Almighty continue to bless all the good intentions we strive to realize in building a national banking sector in the digital era that is competitive and with integrity, anticipative of inherent risks, as well as more resilient, adaptive, and continuously able to contribute optimally in supporting national economic growth for the greatest welfare of the Indonesian people.

Wassalaamu'alaykum warahmatullahi wabarakatuh, Om Shanti Shanti Shanti Om, Namo Buddhaya, Greetings of Virtue.

Dian Ediana Rae

Chief Executive of Banking Supervision
Also Serving as a Member of The OJK
Board of Commissioners

HALAMAN INI SENGAJA DIKOSONGKAN

THIS PAGE IS INTENTIONALLY LEFT BLANK

Bab 1
Latar Belakang

Chapter 1
Background

A. Definisi Kecerdasan Artifisial

1. Jenis Kecerdasan Artifisial

Kecerdasan artifisial atau *artificial intelligence* (AI) telah menjadi kekuatan transformatif dalam teknologi modern, mencakup berbagai kemampuan yang meniru kecerdasan manusia dalam mesin dan perangkat lunak. Dalam cakupan yang lebih luas, teknologi AI bertransformasi menjadi beberapa cabang utama, masing-masing menawarkan fungsi dan aplikasi unik yang merevolusi cara menyelesaikan masalah dan membuat keputusan. *Machine Learning* (ML) adalah bagian dari AI yang berfokus pada pengembangan model dan algoritma yang memungkinkan komputer untuk belajar dan menemukan solusi secara mandiri melalui analisis *dataset*. Dengan memanfaatkan jaringan saraf tiruan (*artificial neural networks*), ML memungkinkan sistem untuk terus meningkatkan performanya dalam menyelesaikan masalah kompleks.

Lebih jauh lagi, *Deep Learning* (DL) memanfaatkan jaringan saraf tiruan *multilayer* untuk menangani tugas pembelajaran yang rumit. DL unggul dalam memproses data dengan jumlah besar, seperti teks atau gambar, sehingga sangat berguna untuk aplikasi yang memerlukan pengenalan pola (misalnya *face recognition*). Salah satu kemajuan dalam DL adalah *Generative AI* (GenAI), yang berfokus pada pembuatan konten baru dari input tidak terstruktur seperti teks, gambar, atau audio. Dengan memanfaatkan DL dan *dataset* yang sangat besar, GenAI menghadirkan kemampuan revolusioner seperti percakapan layaknya manusia (contohnya ChatGPT untuk teks).

Berbeda dengan GenAI, *Predictive AI* berfokus pada memproyeksikan hasil, tren, atau kejadian di masa depan melalui analisis data historis. Dengan mengidentifikasi pola yang menggunakan teknik AI, termasuk ML dan DL, *Predictive AI* membantu pengambilan keputusan berbasis data, seperti mendeteksi penipuan melalui analisis prediktif.

Gelombang terbaru dalam perkembangan AI adalah *Agentic AI,* yang mewakili gelombang ketiga AI setelah *Predictive AI* dan *Generative AI*. *Agentic AI* memperkenalkan sistem yang mampu membuat keputusan secara otonom, berkolaborasi, dan belajar secara mandiri. Teknologi ini memungkinkan interaksi dengan agen AI lain dan mengotomatisasi tugas-tugas kompleks, seperti robotika untuk operasi otonom, kendaraan tanpa pengemudi untuk navigasi (*autonomous vehicles*), dan asisten cerdas yang secara mandiri mengelola tugas.

A. Definition of Artificial Intelligence

1. Types of Artificial Intelligence

Artificial Intelligence (AI) has become a transformative force in modern technology, encompassing various capabilities that mimic human intelligence in machines and software. In a broader scope, AI technology is evolving into several main branches, each offering unique functions and applications that revolutionize problem-solving and decision-making. Machine Learning (ML) is a subset of AI that focuses on developing models and algorithms that enable computers to learn and find solutions independently through dataset analysis. By leveraging artificial neural networks, ML allows systems to continuously improve their performance in solving complex problems.

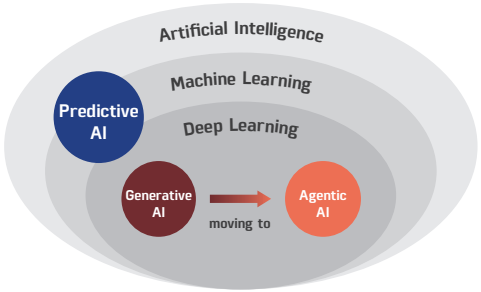
Furthermore, Deep Learning (DL) utilizes multilayer artificial neural networks to handle complex learning tasks. DL excels at processing large amounts of data, such as text or images, making it highly useful for applications that require pattern recognition (e.g., face recognition). One of the advancements in DL is Generative AI (GenAI), which focuses on creating new content from unstructured inputs like text, images, or audio. By leveraging DL and vast datasets, GenAI offers revolutionary capabilities such as human-like conversation (for example, ChatGPT for text generation).

Unlike GenAI, Predictive AI focuses on projecting outcomes, trends, or future events through the analysis of historical data. By identifying patterns using AI techniques, including ML and DL, Predictive AI supports data-driven decision-making, such as detecting fraud through predictive analysis.

The latest wave in AI development is Agentic AI, which represents the third wave of AI after Predictive AI and Generative AI. Agentic AI introduces systems capable of autonomous decision-making, collaboration, and independent learning. This technology enables interaction with other AI agents and automates complex tasks, such as robotics for autonomous operations, autonomous vehicles for navigation, and intelligent assistants that independently manage tasks.

Gambar 1. Jenis Teknologi Kecerdasan Artifisial

Figure 1. Types of Artificial Intelligence Technology



Source: BCG (2023); WEC (2024); Bernard Marr (Forbes) (2024).

2. Perbedaan Otomatisasi dan Kecerdasan Artifisial

Pendekatan teknologi terus berkembang, mulai dari mengotomasi tugas rutin hingga menciptakan sistem yang mampu bertindak secara mandiri. Hal ini mencerminkan perjalanan inovasi yang membuka peluang baru dalam berbagai industri dan kehidupan manusia. Dalam konteks ini, terdapat empat pendekatan utama yang membedakan otomatisasi dan evolusi AI, serta masing-masing memiliki keunikan dan tujuan berbeda.

Robotic Process Automation (RPA) berfokus pada automasi tugas-tugas repetitif dan alur kerja tanpa menghasilkan *output* yang baru. Kemampuan utamanya adalah meniru interaksi manusia dengan sistem, tanpa menciptakan metode interaksi baru. Pendekatan ini mengandalkan aturan imitasi, bukan pembelajaran mandiri, dan umum digunakan untuk entri data dan automasi proses.

Traditional AI membawa kemajuan dalam pengenalan pola, analisis data, regresi, prediksi, dan klasifikasi. Berbeda dengan RPA, teknologi AI ini menggunakan algoritma *machine learning* yang lebih terstruktur untuk mengolah data yang ada. Namun, pembelajaran dalam *Traditional AI* jarang bersifat *real-time*, dan difokuskan pada dukungan dalam pengambilan keputusan, manajemen risiko, segmentasi pelanggan, prediksi.

Generative AI merupakan lompatan selanjutnya dari AI dengan kemampuannya menghasilkan data dan *output* baru seperti teks, gambar, audio, atau kode. Teknologi ini menggunakan pembelajaran mandiri tanpa pengawasan dengan representasi ruang laten (tersembunyi), menjadikannya alat yang kuat untuk augmentasi kemampuan manusia, seperti pembuatan konten kreatif.

Agentic AI menghadirkan paradigma baru dalam dunia AI. Teknologi ini mampu mengambil keputusan, belajar, dan bertindak secara otonom. Dengan pendekatan *reinforcement learning* dan pembelajaran tanpa pengawasan,

2. The Difference between Automation and Artificial Intelligence (AI)

Technological approaches continue to evolve, ranging from the automation of routine tasks to the creation of systems capable of independent action. This reflects a journey of innovation that unlocks new opportunities across industries and everyday life. In this context, there are four main approaches that distinguish automation from the evolution of AI, each with its own uniqueness and distinct objectives.

Robotic Process Automation (RPA) focuses on automating repetitive tasks and workflows without generating new outputs. Its main capability is to mimic human interactions with systems, without creating new interaction methods. This approach relies on predefined rules rather than self-learning and is commonly used for data entry and process automation.

Traditional AI brings advancements in pattern recognition, data analysis, regression, prediction, and classification. Unlike RPA, this AI technology utilizes structured machine learning algorithms to process existing data. However, learning in Traditional AI is rarely real-time and is focused on supporting decision-making, risk management, customer segmentation, and predictions.

Generative AI represents the next leap in AI with its ability to generate new data and outputs such as text, images, audio, or code. This technology utilizes unsupervised self-learning with latent space representation, making it a powerful tool for augmenting human capabilities, particularly in creative content creation.

Agentic AI introduces a new paradigm in the world of AI. This technology is capable of making decisions, learning, and acting autonomously. With reinforcement learning and unsupervised learning approaches,

Gambar 2. Perbedaan Otomatisasi dan Kecerdasan Artifisial

Figure 2. Difference between Automation and Artificial Intelligence

	Robotic Process Automation (RPA)	Traditional AI	Generative AI	Agentic AI
Functions	• Automation of work flow and repetitive tasks. • Does not generate "new" output.	Pattern recognition, regression, analysis/prediction, classification.	Content generation (e.g., text, images, codes, etc.).	Autonomous decision making and action execution.
Core Competencies	• Emulating human interaction with the system • Does not create new interaction methods.	• Analysis, application, and prediction based on existing data/models. • Almost no real-time learning.	• Generating new data and outputs. • Real time learning, independent correction..	Interacts with other systems, learns and acts in real time.
Learning	• Based on imitating rules (copying)s • No learning.	• Single algorithm for machine learning. • More structured and controlled compared to Generative AI.	Independent learning, no supervision, latent (hidden) space representation.	Reinforcement learning, unsupervised learning.
Use Case	Task automation, data entry, process automation.	Supporting human, risk management, customer segmentation, prediction.	Human augmentation; creation of text, images, audio, codes.	AI Assistant and autonomous teams

Source: Compiled from various souces

Agentic AI dapat berinteraksi dengan sistem AI lain dan membentuk tim otonom. Penggunaannya meliputi sistem otonom seperti kendaraan tanpa pengemudi dan asisten AI yang dapat mengelola tugas secara mandiri.

Agentic AI can interact with other AI systems and form autonomous teams. Its applications include autonomous systems such as self-driving vehicles and AI assistants that can manage tasks independently.

3. Lanskap Kecerdasan Artifisial

3. Artificial Intelligence Landscape

Teknologi telah berkembang pesat, mulai dari sistem deterministik yang sederhana hingga teknologi dengan kecerdasan tinggi dan prediktif. Perjalanan ini dimulai dengan algoritma klasik, yaitu sistem berbasis aturan yang dirancang untuk tugas-tugas tertentu. Contohnya, *chatbot* sederhana yang menjawab pertanyaan dasar atau *spam filter* pada *email*. Teknologi ini masih membutuhkan pemrograman eksplisit dan modifikasi manual jika ingin digunakan untuk domain lain. Dalam algoritma klasik digunakan algoritma berbasis aturan atau yang telah ditentukan sebelumnya yang dirancang untuk melakukan tugas spesifik atau menyelesaikan masalah tertentu. Jika terdapat tugas atau *domain* yang berbeda, maka dibutuhkan pemrograman eksplisit dan modifikasi untuk mengakomodasi tugas tersebut.

Technology has rapidly evolved, from simple deterministic systems to high-intelligence and predictive technologies. This journey began with classical algorithms, which are rule-based systems designed for specific tasks. For example, simple chatbots that answer basic questions or spam filters in emails. These technologies still requires explicit programming and manual modifications to be applied to other domains. In classical algorithms, rule-based or predefined algorithms are used, designed to perform specific tasks or solve particular problems. If there are different tasks or domains, explicit programming and modifications are needed to accommodate those tasks.

Dalam perkembangan selanjutnya, teknologi AI muncul dengan kemampuan rendah atau sering disebut *Narrow AI*. Sistem ini adalah AI yang lebih pintar dari algoritma klasik, karena bisa melakukan tugas-tugas tertentu dengan baik, seperti pengenalan suara (contohnya *Siri*, *Alexa*, dan *Google Assistant*), wajah (seperti *Apple Face ID* dan *Google Photos*), sistem rekomendasi produk (seperti *Amazon*), dan personalisasi konten untuk *streaming platform* (seperti *Netflix* dan *Spotify*). Namun, *Narrow AI* dirancang hanya untuk menjalankan tugas tertentu dalam batasan dan definisi terstruktur yang telah ditentukan.

Transformasi AI selanjutnya adalah *Generative AI*, yang menjadi *game-changer* dalam perkembangan teknologi AI. *Generative AI* mampu menghasilkan *output* baru, seperti teks, gambar, musik, atau bahkan kode program. Contohnya adalah *ChatGPT* untuk teks, *DALL-E* untuk gambar, *Runway ML* dan *Synthesia* untuk mengedit video, *Jukebox* untuk menciptakan musik berdasarkan genre tertentu, *Copilot* dan *Codex* untuk membantu pembuatan kode dalam aplikasi, dan Jasper.ai yang menyediakan solusi *marketing* dan *sales content*.

In subsequent developments, AI technology emerged with limited capabilities, often referred to as Narrow AI. This system is more advanced than classical algorithms because it can perform specific tasks well, such as voice recognition (e.g., Siri, Alexa, and Google Assistant), facial recognition (e.g. Apple Face ID and Google Photos), product recommendation systems (e.g. as Amazon), and content personalization for streaming platforms (e.g. Netflix and Spotify). However, Narrow AI is designed solely to carry out specific tasks within predefined structured boundaries and definitions.

The next transformation in AI is Generative AI, which has become a game-changer in the development of AI technology. Generative AI can produce new outputs, such as text, images, music, or even program code. Notable examples include ChatGPT for text, DALL-E for images, Runway ML and Synthesia for video editing, Jukebox for creating music in specific genres, Copilot and Codex for assisting in code creation within applications, and Jasper.ai, which provides marketing and sales content solutions.

Perkembangan terkini adalah konsep *General AI* atau AI Umum, yang masih bersifat hipotetis saat ini. *General AI* dalam artian lain disebut sebagai "*ultimate level*" dari AI, karena dirancang untuk memahami, belajar, dan menyelesaikan berbagai tugas di berbagai *domain*, mirip dengan kemampuan kognitif manusia. Namun, secara praktik teknologi ini belum sepenuhnya hadir di dunia nyata.

The latest development is the concept of General AI, or Artificial General Intelligence (AGI), which remains hypothetical at this time. General AI, in other words, is referred to as the "ultimate level" of AI, as it is designed to understand, learn, and accomplish various tasks across multiple domains, similar to human cognitive abilities. However, in practice, this technology has not yet fully materialized in the real world.

B. Perkembangan Kecerdasan Artifisial

1. Perkembangan Kecerdasan Artifisial secara Global

Dalam laporan Fortune Business Insights (2023), *top three industries* yang paling banyak mengadopsi teknologi AI secara global adalah industri teknologi informasi (TI) dan telekomunikasi, sektor jasa keuangan, dan otomotif. Lebih lanjut, laporan ini menyoroti bahwa perusahaan TI dan telekomunikasi adalah hal yang umum untuk adopsi teknologi AI, namun sektor perbankan adalah yang paling diluar prediksi karena secara historis, industri perbankan dikenal sebagai industri yang konservatif, lebih mengutamakan keamanan, regulasi ketat, dan stabilitas.

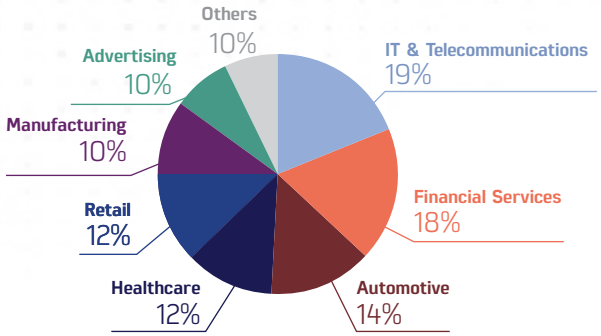
B. The development of Artificial Intelligence (AI).

1. Global Development of Artificial Intelligence (AI)

According to the Fortune Business Insights report (2023), the top three industries that have adopted AI technology globally are the information technology (IT) and telecommunications industry, the financial services sector, and the automotive industry. While AI adoption is expected in IT and telecommunications, the report highlights the banking sector as a surprising frontrunner. Traditionally known for its conservative approach–prioritizing security, strict regulations, and stability. However, the data reveals that the banking industry

Grafik 1. Industri dengan Adopsi Kecerdasan Artifisial Terbesar

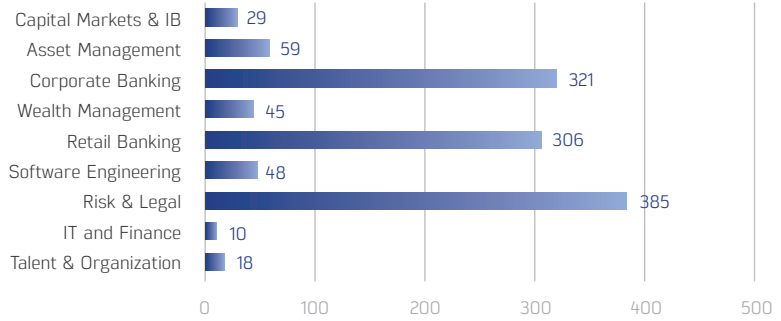
Chart 1. Industries with The Biggest Artificial Intelligence Adoption



Source: Fortune Business Insights (2023)

Grafik 2. Potensi Ekonomi yang Diciptakan AI berdasarkan Segmen dan Fungsi

Chart 2. Economic Potential Generated by AI based on Segment and Function



Source: McKinsey (2023)

Namun data tersebut mengungkapkan bahwa industri perbankan menjadi salah satu industri yang cepat mengadopsi teknologi AI dalam era transformasi digital.

Sejalan dengan hal tersebut, dalam laporan Business Insider (2023), sekitar 80% bank telah menyadari potensi keunggulan teknologi AI, dalam hal ini adalah *machine learning*. Dengan teknologi tersebut, bank diproyeksikan dapat menghemat biaya sebesar US$447 miliar. Sementara itu, *Generative* AI diproyeksikan dapat memberikan manfaat tambahan bagi bank, dengan nilai manfaat mencapai US$340 miliar.

Selanjutnya, berdasarkan laporan McKinsey (2023) jika melihat potensi ekonomi yang diciptakan AI berdasarkan segmen dan fungsi pada bank, area pada risiko dan hukum berpotensi menghasilkan nilai sebesar US$385 miliar, diikuti oleh perbankan korporasi (US$321 miliar) dan perbankan ritel (US$306 miliar).

In line with this, the Business Insider report (2023) states that approximately 80% of banks have recognized the potential advantages of AI technology, particularly in the area of machine learning. With this technology, banks are projected to save costs amounting to US$447 billion. Meanwhile, Generative AI is expected to provide additional benefits for banks, with the value of these benefits reaching US$340 billion.

Moreover, according to the McKinsey report (2023), when examining the economic potential created by AI based on segments and functions within banks, the area of risk and compliance has the potential to generate a value of US$385 billion, followed by corporate banking (US$321 billion) and retail banking (US$306 billion).

2. Penggunaan Kecerdasan Artifisial di Sektor Perbankan

Sektor keuangan merupakan sektor yang sangat potensial dalam perkembangan implementasi teknologi AI. Secara umum terdapat beberapa fungsi/tugas di industri jasa keuangan yang dapat dilakukan dengan memanfaatkan teknologi AI, seperti pemrosesan dokumen, manajemen risiko, penjualan dan jasa, manajemen dan deteksi kecurangan (*fraud management and detection*), deteksi serangan siber dan penguatan keamanan siber, layanan pelanggan (*customer service*), dll.

Di sektor perbankan, teknologi AI khususnya *Predictive AI* dan *Generative AI* dapat dimanfaatkan untuk melakukan fungsi-fungsi antara lain, pemasaran dan penjualan (*marketing and sales*), *prospecting and onboarding* konsumen, pengembangan produk, operasional, nasihat keuangan (*financial advice*), dukungan kepada pelanggan (*customer support*). risiko dan kepatuhan, hingga dukungan terhadap *corporate function* seperti pengelolaan sumber daya manusia, manajemen rapat, dan penyusunan laporan keuangan.

2. Use Cases of Artificial Intelligence (AI) in the Banking Sector

The financial sector presents significant potential for the implementation of AI technology. Several functions within the financial services industry can be enhanced through AI, including document processing, risk management, sales and customer service, fraud detection and management, cyberattack detection, and cybersecurity enhancement.

In the banking sector, AI technologies–particularly Predictive AI and Generative AI–can be applied across various functions, including marketing and sales, customer prospecting and onboarding, product development, operations, financial advisory services, and customer support. Additionally, AI can support risk management, regulatory compliance, and corporate functions such as human resource management, meeting coordination, and financial report preparation.

Tabel 1. Penggunaan Kecerdasan Artifisial di Sektor Keuangan

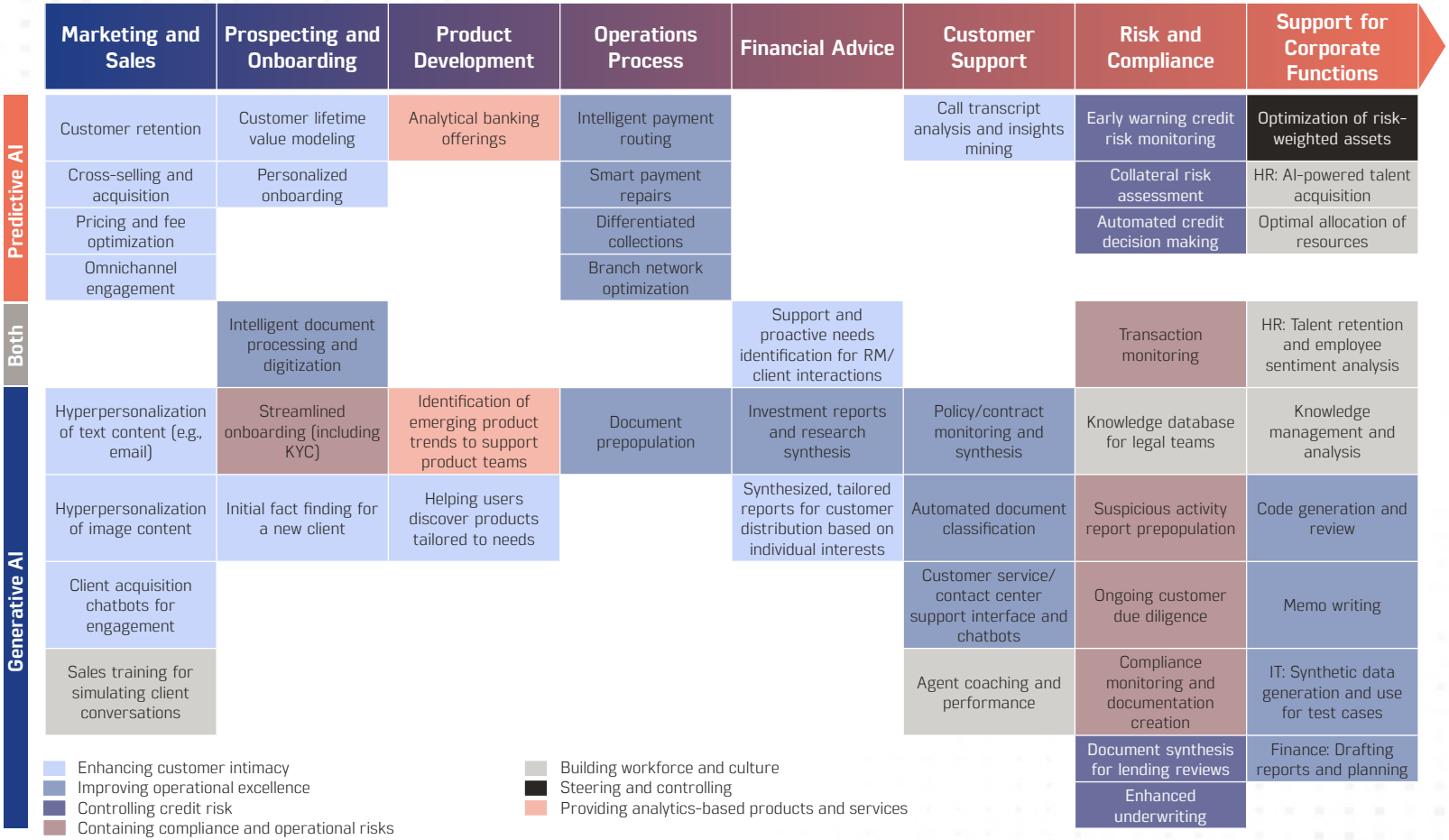
Table 1. Use Cases of Artificial Intelligence in the Financial Sector

Industry	Use cases	Value Delivered
Banking	• Customer Service • Fraud Detection • Personalized offering and product • Fraud Detection • Know Your Customer (KYC) process • Process automation and document processing	• Digitization and enhanced customer experienced • Operational efficiency • Improve accuracy
Corporate finance	• Portfolio optimization • Capital Budgeting • Compliance • Investment management • Financial statement analysis	• Optimal capital allocation • Manage risks • Improve efficiency and accuracy
Capital markets	• Stock movement projections • Algorithmics trading • Investment planning • Trade executions • Portfolio rebalancing	Helps individual and institutional investor to formulate and execute capital market strategies
Payment system	• Cyber Fraud detection • Customer Service • Streamlined Processing and Settlement • Data-driven insights • Personalized offer and recommendations	• Automate payment processing • Improve efficiency • Faster transaction • Optimal payment acceptance
Insurance	• Credit scoring • Fraud detection • Claims management • Claims Adjudication • Insurance distribution	Improve operational efficiency and reduce risks
Fintech	• Investment management • Fraud identification • Personalized sales • Secure transactions	Improve decision making and customer interaction

Source: Compiled from various sources

Gambar 3. Peluang Implementasi *Predictive AI* dan *Generative AI* pada Bank

Figure 3. Opportunities to Implement Predictive AI and Generative AI in Banks



Source: BCG (2023)

Selain *Predictive AI* dan *Generative AI*, industri perbankan juga dapat memanfaatkan *Agentic AI* yang dianggap dapat memberikan dampak yang lebih besar terhadap produktivitas dibandingkan *Predictive AI* dan *Generative AI*. Beberapa area/fungsi yang dapat dilakukan oleh *Agentic AI* antara lain interaksi dengan pelanggan (*customer interaction*), produk dan kalkulasi harga (*product and pricing*), kepatuhan dan pencegahan kecurangan (*compliance and fraud detection*), dan intelegensi pasar (*market intelligence*).

In addition to Predictive AI and Generative AI, the banking industry can also leverage Agentic AI, which is considered to have a greater impact on productivity compared to Predictive AI and Generative AI. Some areas where Agentic AI can be applied include customer interaction, product and pricing optimization, compliance and fraud detection, and market intelligence.

3. Dampak Penggunaan Kecerdasan Artifisial pada Pendapatan dan Biaya Bank

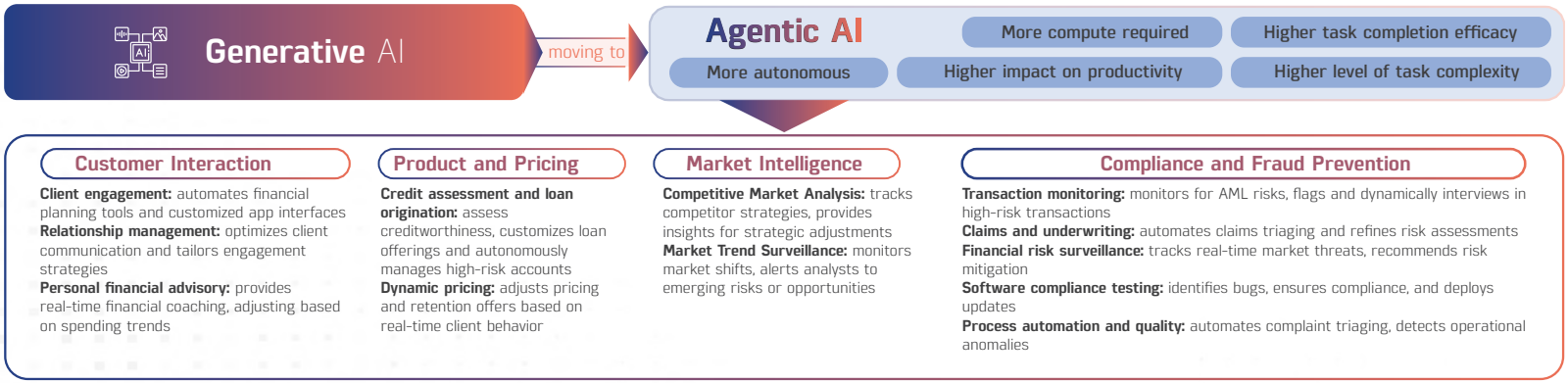
Teknologi AI berpotensi menjadi *tools* yang dapat memberikan dampak signifikan terhadap bisnis dan memperbesar kinerja keuangan pada industri perbankan. Laporan Deloitte (2024) menyebutkan bahwa banyak bank telah berinvestasi pada inovasi strategis seperti teknologi *cloud* dan digitalisasi. Namun, belum semua bank mencapai peningkatan keuntungan secara signifikan dari investasi ini. Sementara teknologi AI memiliki potensi untuk meningkatkan efisiensi biaya di bank, seperti:

3. Impact of Artificial Intelligence Usage on Bank Revenue and Costs

AI technology has the potential to be a tool that can significantly impact business and enhance financial performance in the banking industry. A Deloitte report (2024) states that many banks have invested in strategic innovations such as cloud technology and digitalization. However, not all banks have realized substantial profit gains from these investments. In contrast, AI offers promising opportunities to improve cost efficiency across various banking functions, such as:

Gambar 4. Peluang Implementasi *Agentic AI* pada Bank

Figure 4. Opportunities to Implement Agentic AI in Banks



Source: World Economic Forum, How Agentic AI will transform financial services with autonomy, efficiency and inclusion (2024)

a. Efisiensi Total Biaya Staf (0-15%)

Teknologi AI mampu mengurangi biaya tenaga kerja dengan mempercepat produktivitas. Dengan teknologi AI, proses yang memerlukan banyak tenaga kerja manual dapat diotomatisasi, sehingga memungkinkan pencapaian *output* yang lebih tinggi dengan lebih sedikit sumber daya dan waktu. Hal ini mencakup proses seperti *data entry*, analisis data, dan pelaporan, yang dapat diotomatisasi untuk efisiensi lebih tinggi.

b. Penghematan Biaya Staf TI (10-20%)

Teknologi AI dapat memangkas waktu dan biaya dalam pembuatan, implementasi, dan pemeliharaan sistem, perangkat lunak, serta infrastruktur teknologi. Teknologi AI juga dapat digunakan untuk memantau sistem secara otomatis, mendeteksi masalah, dan memperbaikinya tanpa perlu intervensi manual, yang dapat mengurangi beban kerja tim TI.

c. Penghematan dari Penurunan Nilai Aset (10-15%)

Kemampuan teknologi AI dalam melakukan penilaian risiko kredit yang lebih baik dapat membantu mengurangi kerugian yang disebabkan oleh piutang tak tertagih. Model risiko kredit berbasis AI dapat mengevaluasi risiko debitur dengan lebih akurat, yang pada akhirnya mengurangi tingkat kerugian atau cadangan penurunan nilai yang perlu disiapkan.

d. Peningkatan Deteksi *Financial Crime/Fraud*

Penerapan AI dalam deteksi *financial crime* dan *fraud* dapat mengurangi biaya litigasi, ganti rugi, dan kerugian yang disebabkan oleh penipuan. Algoritma AI dapat mendeteksi pola perilaku mencurigakan dan anomali dalam transaksi secara *real-time*, memungkinkan pencegahan yang lebih proaktif dan respons yang cepat dalam menangani ancaman penipuan.

Dari sisi pertumbuhan pendapatan terdapat potensi peningkatan seperti:

a. Peningkatan 5-7% pada Pendapatan Perdagangan

Dengan algoritma analisis pasar berbasis AI, bank dapat menganalisis data pasar secara *real-time* untuk mengidentifikasi tren serta memperkirakan pergerakan pasar di masa depan. Teknologi ini memungkinkan keputusan *trading* yang lebih cepat dan tepat, yang dapat meningkatkan peluang keuntungan dalam aktivitas *trading*.

a. Total Staff Cost Efficiency (0-15%)

AI technology can significantly reduce labor costs by enhancing productivity. By automating processes that traditionally require extensive manual labor, AI enables higher output with fewer resources and less time. This includes tasks such as data entry, data analysis, and reporting, which can be streamlined for greater efficiency.

b. IT Staff Cost Savings (10-20%)

AI technology can reduce time and costs in the creation, implementation, and maintenance of systems, software, and technology infrastructure. It can also be used to automatically monitor system performance, detect issues, and resolve them without manual intervention—thereby reducing the workload on IT teams and improving operational efficiency.

c. Savings from Asset Depreciation (10-15%)

AI technology's ability to enhance credit risk assessment can help reduce losses resulting from bad debts. AI-based credit risk models can assess debtor risk more accurately, ultimately reducing the level of losses or impairment reserves that need to be set aside.

d. Improvement in Financial Crime/ Fraud Detection

The application of AI in detecting financial crime and fraud can reduce litigation costs, compensation expenses, and losses resulting from fraudulent activities. AI algorithms can detect suspicious behavior patterns and anomalies in transactions in real-time, allowing for more proactive prevention and quick responses to addressing fraud threats.

From the revenue growth perspective, there is potential for increases such as:

a. Increase of 5-7% in Trading Revenue

With AI-based market analysis algorithms, banks can analyze market data in real-time to identify trends and forecast future market movements. This technology enables faster and more accurate trading decisions, thereby enhancing profit opportunities in trading activities.

b. Peningkatan 1-2% pada Biaya/Komisi

AI membantu mempertahankan pelanggan melalui analisis prediktif yang memungkinkan perusahaan memahami kebutuhan dan preferensi pelanggan dengan lebih baik. Dengan layanan yang lebih relevan dan pengalaman pelanggan yang lebih baik, bank dapat meningkatkan retensi pelanggan, yang pada akhirnya meningkatkan pendapatan dari biaya dan komisi.

c. Peningkatan 5-10% pada Pendapatan/Biaya Bunga

Dengan pemasaran yang dipersonalisasi, AI dapat menargetkan calon pelanggan dengan lebih efektif, menghasilkan peningkatan dalam akuisisi pelanggan. Analisis data pelanggan memungkinkan kampanye pemasaran yang disesuaikan secara individual, yang meningkatkan konversi dan memperkuat hubungan pelanggan.

d. Peningkatan 2-3% pada Pendapatan Bunga Bersih

Teknologi AI memungkinkan penetapan suku bunga pinjaman yang lebih akurat berdasarkan penilaian risiko kredit secara individual. Dengan memahami profil risiko kredit setiap nasabah, bank dapat menentukan suku bunga pinjaman yang sesuai. Ini memungkinkan penetapan suku bunga yang kompetitif dan lebih tepat sasaran, yang dapat meningkatkan pendapatan bunga bank dan meminimalkan risiko *default*.

b. Increase of 1-2% in Fees/ Commissions

AI supports customer retention through predictive analytics that enable companies to better understand customer needs and preferences. With more relevant services and an improved customer experience, banks can strengthen customer loyalty and ultimately increase revenue from fees and commissions

c. Increase of 5-10% in Interest Income/Costs

With personalized marketing, AI can target potential customers more effectively, resulting in an increase in customer acquisition. By analyzing customer data, banks can deliver individually tailored marketing campaigns that improve conversion rates and strengthen customer relationships.

d. Increas of 2-3% in Net Interest Income

AI technology enables more accurate loan interest rate setting based on individual credit risk assessments. By understanding the credit risk profile of each customer, banks can determine appropriate loan interest rates. This allows for competitive and more targeted interest rate setting, which can enhance the bank's interest income and minimize default risk.

C. Kesiapan Adopsi Kecerdasan Artifisial di Indonesia

BCG Center of Public Economics melakukan survei kesiapan adopsi AI yang diukur menggunakan ASPIRE yaitu *Ambition* (Ambisi), *Skills* (Kemampuan), *Policy and Regulation* (Regulasi dan Kebijakan), *Investment* (Investasi), *Research and Innovation* (Riset dan Inovasi), dan *Ecosystem* (Ekosistem). Berdasarkan data ASPIRE tersebut, Indonesia berada pada top 50% sebagai salah satu negara yang telah menerapkan AI yang didukung pada penilaian aspek Ambisi, Kebijakan dan Regulasi, serta Ekosistem. Namun, negara-negara yang berada di top 50% tersebut masih terdapat kekurangan pada aspek Investasi dan Riset/Inovasi.

Secara keseluruhan, Indonesia berada pada posisi yang lebih baik daripada rata-rata global. Laporan Oxford Insights (2023) mengungkapkan bahwa Indonesia berada pada peringkat ke-42 dari 193 negara dan Indonesia mencatatkan tingkat kesiapan sekitar 58,14% dibandingkan rata-rata dunia sebesar 47,42%. Hal ini menunjukkan bahwa fondasi dasar untuk transformasi

C. Artificial Intelligence Adoption Readiness in Indonesia

The BCG Center for Public Economics conducted a survey on AI adoption readiness using ASPIRE, which stands for Ambition, Skills, Policy and Regulation, Investment, Research and Innovation, and Ecosystem. According to ASPIRE data, Indonesia ranks within the top 50% of countries adopting AI, driven by strong performance in Ambition, Policy and Regulation, and Ecosystem. However, countries in this group, including Indonesia, still face challenges in the areas of Investment and Research and Innovation.

Overall, Indonesia is in a better position than the global average. The Oxford Insights report (2023) reveals that Indonesia ranks 42nd out of 193 countries, with a readiness score of approximately 58.14% compared to the global average of 47.42%. This indicates that the foundational groundwork for digital transformation is beginning

digital sudah mulai terbentuk, meskipun infrastruktur digital dan sumber daya manusia masih perlu ditingkatkan.

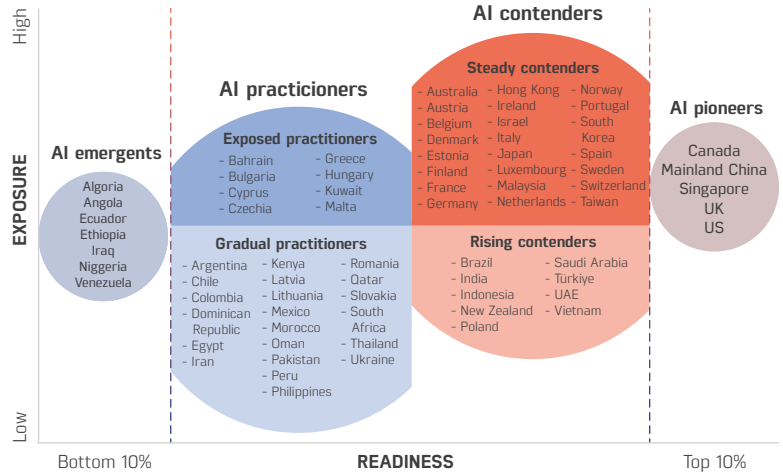
Selain survei, BCG mengeluarkan laporan Matriks Tingkat Kematangan AI terhadap 73 ekonomi global yang memberikan pandangan luas tentang adopsi AI secara global. Sebagian besar ekonomi telah mengadopsi AI secara bertahap, tetapi terdapat beberapa negara yang dianggap sebagai pelopor AI karena tingkat adopsi AI yang tinggi.

to take shape, although digital infrastructure and human resources still need improvement.

In addition to the survey, BCG released a report on the AI Maturity Matrix for 73 global economies, providing a broad view of AI adoption worldwide. Most economies have adopted AI gradually, but there are several countries considered pioneers in AI due to their advanced levels of AI adoption.

Gambar 5. Matriks Tingkat Kematangan Kecerdasan Artifisial

Figure 5. AI Maturity Matrix



Sources: BCG Center for Public Economics; BCG Analysis
Note: Within each archetype, economic appear in alphebetical order

Dengan berfokus pada dua aspek penting, laporan BCG ini menawarkan pendekatan unik untuk melihat dinamika global dalam adopsi AI. Pertama, paparan setiap ekonomi terhadap disrupsi yang didorong oleh AI. Paparan didefiniskan sebagai potensi AI untuk memengaruhi suatu sektor dalam ekonomi secara negatif atau positif. Kemudian setiap ekonomi dinilai kesiapannya untuk memanfaatkan potensi pertumbuhan AI dan untuk mengurangi potensi risiko yang ditimbulkan oleh AI. Matriks yang dihasilkan menyatukan beberapa faktor tersebut sehingga menghasilkan 6 (enam) level pengembangan dan potensi ekonomi AI.

Dari 73 negara yang dinilai, Kanada, Tiongkok, Singapura, Inggris, dan AS dikategorikan sebagai pelopor AI. Kelima negara tersebut telah mencapai tingkat kesiapan yang tinggi dengan inovasi, pengembangan bakat, serta regulasi dan etika AI. Adapun terdapat beberapa negara dengan kesiapan AI yang tinggi berada tepat di belakang negara-negara pelopor AI tersebut. Kelompok ini mencakup negara-negara yang sudah mapan, negara-negara berkembang seperti India, Arab Saudi, dan Uni Emirat Arab yang memiliki kebijakan dan melakukan investasi untuk adopsi AI pada tingkat yang lebih tinggi. Seiring

By focusing on two important aspects, the BCG report offers a unique approach to understanding the global dynamics of AI adoption. First, it examines each economy's exposure to AI-driven disruptions. Exposure is defined as the potential for AI to impact a sector within the economy either negatively or positively. It then assesses each country's readiness to harness AI's growth potential while mitigating associated risks.. The resulting matrix integrates multiple factors, producing six levels of AI economic development and potential.

Among the 73 countries assessed, Canada, China, Singapore, the United Kingdom, and the United States are categorized as AI pioneers. These five countries have achieved a high level of readiness through innovation, talent development, and AI regulation and ethics. In addition, several countries with high AI readiness are closely following the AI pioneers. This group includes advanced nations and developing countries such as India, Saudi Arabia, and the United Arab Emirates, which have policies and make investments for higher levels of AI adoption. As their innovation capabilities strengthen,

dengan semakin kuatnya kemampuan inovasi, negara-negara ini akan semakin kompetitif dan berpengaruh di bidang AI. Indonesia sendiri merupakan salah satu negara yang dikategorikan sebagai *Rising Contenders* yaitu negara-negara dengan eksposur AI yang masih rendah tetapi memiliki komitmen untuk meningkatkan adopsi AI.

these countries will become increasingly competitive and influential in the field of AI. Indonesia itself is categorized as a Rising Contender, which refers to countries with low AI exposure but a commitment to enhancing AI adoption.

Sejalan dengan hal tersebut, dukungan dan komitmen Indonesia terhadap perkembangan AI ditandai dengan diterbitkannya Strategi Nasional Kecerdasan Artificial Indonesia 2020 – 2045, Surat Edaran Menteri Komunikasi dan Informatika (SE Menkominfo) tentang Etika Kecerdasan Artifisial, serta Panduan Kode Etik AI dan Terpercaya di Industri Teknologi Finansial yang dikeluarkan oleh Otoritas Jasa Keuangan (OJK).

In line with this, Indonesia's support and commitment to the development of AI are marked by the issuance of the National Artificial Intelligence Strategy 2020 – 2045, the Minister of Communication and Information's Circular Letter (SE Menkominfo) on AI Ethics, and the AI and Trustworthy Code of Ethics Guidelines in the Financial Technology Industry issued by the Indonesian Financial Services Authority (OJK).

Strategi Nasional Kecerdasan Artifisial Indonesia 2020 – 2045 memiliki 5 (lima) hal yang menjadi fokus utama, yaitu: 1) berorientasi pada kemaslahatan umat manusia; 2) bernafaskan nilai-nilai pancasila; 3) andal, aman dan terbuka, dapat dipertanggungjawabkan; 4) sinergitas antara pemangku kepentingan; dan 5) penerapan asas-asas UU No. 11 Tahun 2019 tentang Sistem Nasional Ilmu Pengetahuan dan Teknologi.

The National Artificial Intelligence Strategy of Indonesia 2020 – 2045 focuses on five main aspects, namely: 1) oriented towards the welfare of humanity; 2) infused with the values of Pancasila; 3) reliable, safe, and open, accountable; 4) synergy among stakeholders; and 5) implementation of the principles of Law No. 11/2019 on the national system of science and technology.

SE Menkominfo tentang Etika Kecerdasan Artifisial mengatur bahwa AI harus memenuhi nilai-nilai berikut: 1) inklusivitas; 2) kemanusiaan; 3) keamanan; 4) aksesibilitas; 5) transparansi; 6) kredibilitas dan akuntabilitas; 7) pelindungan data pribadi; 8) pembangunan dan lingkungan berkelanjutan; dan 9) kekayaan intelektual.

Selanjutnya pada Panduan Kode Etik AI dan Terpercaya di Industri Fintech terdapat prinsip dasar pedoman perilaku kecerdasan buatan, yaitu: 1) berasaskan pancasila; 2) bermanfaat; 3) wajar dan akuntabel; 4) transparan dan dapat dijelaskan; dan 5) ketangguhan dan keamanan.

Di sektor perbankan, OJK bersama World Bank melakukan pengumpulan informasi melalui kegiatan survei terhadap beberapa bank yang diidentifikasi telah menerapkan teknologi AI dalam kegiatan bisnisnya. Survei tersebut bertujuan untuk mengetahui tingkat adopsi dan implementasi AI, serta tantangan dan permasalahan yang dihadapi bank dalam implementasi AI. Hasil survei yang diperoleh adalah sebagai berikut:

The Minister of Communication and Information's Circular Letter on AI Ethics stipulates that AI must adhere to the following values: 1) inclusivity; 2) humanity; 3) security; 4) accessibility; 5) transparency; 6) credibility and accountability; 7) personal data protection; 8) sustainable development and environment; and 9) intellectual property.

Furthermore, the AI and Trustworthy Code of Ethics Guidelines in the Fintech Industry outline the fundamental principles of AI behavior, namely: 1) based on Pancasila; 2) beneficial; 3) fair and accountable; 4) transparent and explainable; and 5) resilience and security.

In the banking sector, OJK, in collaboration with the World Bank, conducted a survey targeting banks identified as having implemented AI technology in their business operations. The survey aimed to assess the level of AI adoption and implementation, as well as the challenges and issues encountered by banks in AI implementation. The survey findings are summarized as follows:

Tabel 2. Rangkuman Hasil Survei atas Kesiapan Implementasi Kecerdasan Artifisial di beberapa Bank

Table 2. Summary of Survey on Artificial Intelligence Implementation Readiness in Several Banks

AI Adoption Level	The level of AI technology adoption varies among the banks surveyed, reflecting their focus on specific business needs and risk profiles. Machine learning (ML) models are widely integrated, with most banks surveyed already using them and others piloting these technologies. These models are commonly applied for risk assessment, predictive analysis and anomaly detection, which is in line with global banking trends.
Use Cases	The AI models currently deployed by the surveyed banks mostly consist of ML models that focus on risk assessment. Five of the eight banks have deployed ML models for fraud detection and credit scoring, and all eight banks plan to expand the use of such models in the coming years. These deployments are consistent with global banking practices, where fraud detection and credit scoring are among the most common AI use cases.
Plan for 3-5 years	The surveyed banks confirmed that they intend to accelerate AI adoption in the next 3 to 5 years. GenAI is expected to be adopted across the surveyed banks.
Resources Allocation	The allocation of AI-related resources, both financial and human resources, is aligned with the scale and approach of AI adoption in each bank. Banks with smaller financial budgets compensate by dedicating more human resources to develop and manage AI models internally. Conversely, banks with larger financial budgets often utilize external cloud resources, reducing the need for large internal HR allocations.
AI Governance Implementation	Banks with lower levels of AI adoption often lack specific organizational structures, specific policies, and risk management procedures tailored for AI. In contrast, most banks surveyed indicated that they have established or are in the process of developing a customized AI strategy and governance framework.
Benefits of AI	Banks that primarily utilize AI for risk assessment report several benefits, such as reducing process execution time and operational costs. Frequently cited applications include loan processing and customer onboarding activities that comply with KYC procedures. AI-driven fraud detection models improve accuracy and the ability to minimize false positives, thus significantly improving operational efficiency in fraud prevention.

Source: Survey of Banks in 2024 (reprocessed)

Bab 2

Risiko dan Tantangan

Chapter 2

Risks and Challenges

A. Deepfakes

Deepfakes adalah media buatan yang dihasilkan oleh GenAI, dalam hal ini adalah *Generative Adversarial Networks* (GAN), yang dapat berupa konten baru visual atau audio yang sangat realistis, sehingga dapat menciptakan manipulasi digital yang tampak seperti video atau suara asli seseorang (fabrikasi). Kasus terkait *deepfakes* ini meningkat seiring kemampuan dan aksesibilitas GenAI yang semakin meningkat, serta semakin majunya terobosan dalam penggunaan teknologi *deep learning*. Hal ini tercermin dari peningkatan signifikan atas ketersediaan alat dan teknologi *deepfakes*, yang dapat dilihat melalui lonjakan jumlah perdagangan *deepfake-related tools* pada forum *dark web* sebesar 223% pada rentang Q1 2023 dan Q1 2024. Lebih lanjut, konten media baik berupa gambar, video, atau audio dapat dengan sangat mudah diproduksi untuk meniru seseorang atau pihak terkait untuk menginisiasi transaksi penipuan atau tindakan ilegal lainnya. *Deepfakes* termasuk *emerging threats* bagi lembaga jasa keuangan yang dapat memicu penyebaran disinformasi, *social engineering*, sehingga dapat mengurangi kepercayaan masyarakat pada sistem keuangan.

A. Deepfakes

Deepfakes are artificial media generated by GenAI, specifically through Generative Adversarial Networks (GAN), which can produce highly realistic visual or audio content, thereby creating digital manipulations that appear as authentic videos or voices of individuals (fabrication). Cases related to deepfakes have increased alongside the growing capabilities and accessibility of GenAI, as well as advancements in deep learning technology. This is reflected in the significant rise in the availability of deepfake tools, evidenced by a 223% surge in the trading of deepfake-related tools on dark web forums between Q1 2023 and Q1 2024. Furthermore, media content, whether in the form of images, videos, or audio, can be easily generated to impersonate individuals or related parties to initiate fraudulent transactions or other illegal activities. Deepfakes represent emerging threats to financial service institutions, potentially fueling the spread of misinformation and social engineering, which can undermine public trust in the financial system.

Tabel 3. Skenario *Deepfakes*

Table 3. Deepfake Scenarios

Target for Impersonation by Fraudsters	Affected Parties	Scenario	Goal	Example
C-suite impersonation	Related Financial Institutions (i.e., banks, asset/wealth management, etc.)	A deepfake audio allows for the impersonation of an executive to initiate transactions, fund transfers, or access restricted information.	Fraud, obtaining sensitive information (such as trade secrets, Personally Identifiable Information (PII), and other non-public data).	A type of business email compromise (BEC) fraud targeted a finance employee at a multinational company in Hong Kong, who received a message from a C-level executive (Head of Finance) to carry out an urgent confidential transaction, accompanied by a falsified video conference (deepfake). (February 2024)

Target for Impersonation by Fraudsters	Affected Parties	Scenario	Goal	Example
Bank Customers	Bank Customers	Banks that implement voice authentication, without additional requirements, allow fraudsters (in this cae, those who have voice samples or Personally Identifiable Information - PII) to initiate fund transfers and withdrawals, as well as other illicit transactions.	Consumer Fraud	
Third-party relationship	Financial Institutions	Fraudsters use deepfakes to impersonate employees of financial institutions or external entities (third-parties) to gain access or withdraw funds from financial entities.	Fraud	A deepfake related to a BEC involving a company director resulted in the transfer of USD 35 million by a bank manager in Hong Kong. (2020)
Individual/public	Financial Services Institutions	Fraudsters use GenAI to create fake identity documents to register with companies for the purposes of espionage, evading punishment, and other malicious activities.	Espionage, taking advantage, evading sanctions, or simply gaining access.	A technology company unknowingly hired an IT employee from North Korea who applied using a falsified identity (deepfake). (July 2024)
Public figur	C-Level Management	Impersonation of public figures (deepfake) to deceive bank executives or C-suite management to embarrass and discredit their security procedures and checks.	Discrediting, influencing reputation.	The President of the European Central Bank (ECB), Christine Lagarde, and the Chair of the U.S. Federal Reserve, Jerome Powell, were embarrassed by a deepfake of Ukrainian President Zelensky during a phone call. Fraudsters released clips of their personal statements. (Reported April 2023)

Source:: FS-ISAC (2024)

B. Kotak Hitam (*Black Box*)

Black box AI, pada dasarnya, merujuk pada sistem AI yang dikarenakan tingkat kompleksitas tinggi, cara kerja internalnya tidak dapat sepenuhnya dipahami atau dijelaskan, bahkan oleh pembuatnya. Istilah "*black box*" secara metaforis menggambarkan ketidakjelasan sistem ini, karena *input* dan *output* dapat diamati, tetapi proses yang terjadi di dalamnya sebagian besar tidak diketahui dan tidak dapat diuraikan.

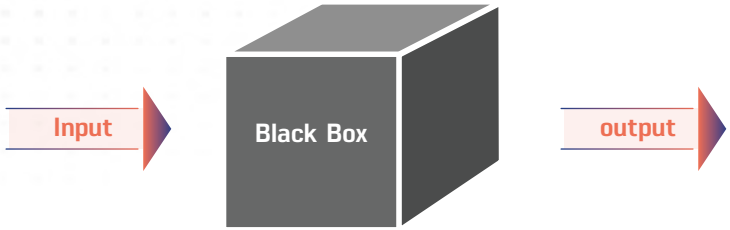
B. Black Box

Black box AI essentially refers to AI systems whose internal workings are not fully understandable or explainable due to their high level of complexity, even by their creators. The term "black box" metaphorically describes the opaqueness of these systems, as the inputs and outputs can be observed, but the processes occurring within are largely unknown and cannot be deciphered.

Karakteristik AI ini, khususnya dalam konteks *deep learning*, dapat menghasilkan prediksi atau keputusan yang sangat akurat, tetapi penalarannya tetap misterius. Kurangnya transparansi ini sering kali menimbulkan masalah kepercayaan, akuntabilitas, dan etika dalam penerapan AI. Umumnya, pendekatan *black box* AI digunakan dalam level *"deep neural networks"*, dimana model AI yang digunakan telah terlatih mengolah data dalam jumah besar dan bobot internal serta parameter dari algoritmanya telah

The characteristics of this AI, particularly in the context of deep learning, can produce highly accurate predictions or decisions, yet the reasoning behind them remains mysterious. This lack of transparency often raises issues of trust, accountability, and ethics in the application of AI. Generally, the black box AI approach is used in deep neural networks, where the AI model is trained to process large amounts of data, and the internal weights and parameters of its algorithms are adjusted during the learning process. In the banking

Gambar 6. Ilustrasi *Black Box* (Kotak Hitam)
Figure 6. Black Box Illustration



Source: Eastgate Software (2024)

disesuaikan. Pada industri perbankan, model ini seringkali digunakan dalam pengambilan keputusan terkait kredit/pinjaman.

Terdapat beberapa *potential issues* yang dapat diakibatkan dari karakteristik *black box* AI ini, yaitu:

a. Bias AI, terjadi bias informasi AI (akibat algoritma AI yang tidak dapat dijelaskan), menyebabkan hasil yang menyimpang/salah, berpotensi menyinggung, tidak adil, bahkan membahayakan kelompok tertentu.

industry, this model is frequently employed in decision-making related to credit/loans.

There are several potential issues that may arise from the characteristics of black box AI, namely:

a. AI Bias, occurs when there is bias in AI information (due to the unexplainable algorithms), leading to skewed/incorrect results that may be offensive, unfair, or even harmful to certain groups.

b. Transparansi dan akuntabilitas, karena tingkat kompleksitas yang tinggi, pengembang cenderung sulit memahami dan mengawasi bagaimana model AI bekerja, sehingga kurangnya pemahaman menyebabkan berkurangnya transparansi dan meminimalkan akuntabilitas.

c. Kurangnya fleksibilitas, apabila model ingin digunakan untuk peruntukan lain (*use case* serupa tapi berbeda), maka aturan atau parameter yang digunakan oleh model AI harus diperbarui/disesuaikan dari awal, dimana prosesnya akan sangat rumit dan membutuhkan waktu.

d. Validitas, *output* atau keputusan yang dihasilkan oleh *black box* AI seringkali sulit untuk divalidasi dan direplikasi.

e. Celah keamanan, yakni input data yang dapat dimanipulasi sehingga memengaruhi kualitas *output* yang dihasilkan (dalam hal ini, kurang tepat atau bahkan berbahaya) dan akibat modelnya yang sulit dipahami, maka hampir tidak mungkin untuk melakukan upaya perbaikan *decision-making process* yang berjalan. Selain itu, rendahnya tingkat transparansi

b. Transparency and accountability, due to the high level of complexity, developers tend to find it challenging to understand and monitor how the AI model operates, resulting in a lack of understanding that diminishes transparency and minimizes accountability.

c. Lack of flexibility, if the model is to be used for other purposes (similar but different use cases), the rules or parameters used by the AI model must be updated/reconfigurated from scratch, a process that will be very complicated and time-consuming.

d. Validity, the outputs or decisions generated by black box AI are often difficult to validate and replicate.

e. Security vulnerabilities, which refer to input data that can be manipulated, thereby affecting the quality of the outputs produced (e.g., being inaccurate or even harmful). Due to the model's complexity, it becomes nearly impossible to make improvements to the ongoing decision-making process. Additionally, the low level

model, menyebabkan sulit untuk menentukan celah keamanan lainnya pada model.

of transparency in the model makes it difficult to identify other potential security vulnerabilities within the model.

C. Bias pada AI

Sebagaimana yang telah dibahas pada bagian sebelumnya, bias merupakan salah satu risiko yang paling rentan terjadi dalam proses pembuatan keputusan yang menggunakan AI. Bias AI, juga disebut sebagai *machine learning bias* atau *algorithm bias*, mengacu pada sistem AI yang menghasilkan output bias yang pada umumnya diakibatkan oleh data yang digunakan tidak *representative*, algoritma yang tidak netral, atau faktor *user* yang mengoperasikan sistem. Bias dapat ditemukan pada beberapa komponen dalam sistem AI, seperti pada data yang digunakan untuk melatih AI (*data-training*), algoritma yang digunakan untuk memproses data, hingga *output* yang dihasilkan oleh algoritma tersebut. Hal ini menyebabkan keputusan yang dihasilkan menjadi tidak adil dan cenderung diskriminatif, sehingga dapat merugikan kelompok tertentu atau memperkuat stereotip yang telah ada.

Berikut adalah beberapa jenis bias dalam AI:

a. *Sampling bias*, terjadi ketika data pelatihan tidak mewakili populasi yang menjadi target, yang dapat mengakibatkan penurunan kinerja dan prediksi yang bias terhadap kelompok tertentu. Contoh: Algoritma pengenalan wajah yang sebagian besar dilatih pada individu kulit putih tidak berfungsi dengan baik pada individu atau kelompok dari ras lain.

b. *Algorithmic bias*, terjadi akibat desain dan implementasi algoritma memberikan bobot lebih besar pada atribut tertentu dan menghasilkan keputusan yang tidak adil. Contoh: Algoritma yang memprioritaskan usia atau jenis kelamin sehingga menghasilkan bias dalam proses perekrutan.

c. *Representation bias*, terjadi ketika kumpulan data tidak secara akurat mewakili populasi yang akan dimodelkan, sehingga menyebabkan prediksi menjadi tidak akurat. Contoh: Kumpulan data medis yang kurang mewakili perempuan, menyebabkan diagnosis kurang akurat untuk pasien perempuan. Bias ini termasuk pula yang dapat terjadi pada proses pembuatan data, gambar, atau teks sintetis.

d. *Confirmation bias*, terjadi ketika AI terlalu bergantung pada keyakinan yang sudah ada (*existing beliefs*) atau tren pada data yang sudah ada sebelumnya. Lebih lanjut, bias

C. AI Bias

As mentioned in the previous section, bias is one of the most significant risks in the decision-making process that utilizes AI. AI bias, also known as machine learning bias or algorithm bias, refers to AI systems that produce biased outputs, generally caused by non-representative data, non-neutral algorithms, or user factors operating the system. Bias can arise in several components of the AI system, such as in the data used to train the AI (training data), the algorithms used to process the data, and the outputs generated by those algorithms. This results in decisions that are unfair and tend to be discriminatory, potentially harming certain groups or reinforcing existing stereotypes.

The following are several types of AI bias:

a. Sampling bias occurs when the training data does not represent the target population, which can lead to decreased performance and biased predictions against certain groups. For example, a facial recognition algorithm trained primarily on images of white individuals may perform poorly when identifying individuals from other racial or ethnic groups, leading to inaccurate or biased outcomes.

b. Algorithmic bias, occurs when the design and implementation of an algorithm give greater weight to certain attributes, resulting in unfair decisions. For example, an algorithm that prioritizes age or gender may produce bias in the recruitment process.

c. Representation bias, occurs when the dataset does not accurately represent the population that is intended to be modeled, leading to inaccurate predictions. For example, a medical dataset that underrepresents women may result in less accurate diagnoses for female patients. This bias can also occur in the processes of generating synthetic data, images, or text.

d. Confirmation bias, occurs when AI relies too heavily on existing beliefs or trends in previously available data. Furthermore, this bias is closely related

ini juga erat kaitannya dengan risiko bias yang dapat muncul akibat peran/keterlibatan manusia dalam seluruh *lifecycle* pengembangan dan evaluasi model AI. Misalnya, jika sebuah algoritma AI yang digunakan dalam perekrutan mempelajari bahwa mayoritas kandidat yang berhasil di masa lalu adalah laki-laki, maka algoritma tersebut mungkin akan mengutamakan pelamar laki-laki di masa mendatang.

to the risk of bias that can arise from human involvement in the entire lifecycle of AI model development and evaluation. For example, if an AI algorithm used in recruitment learns that the majority of successful candidates in the past were male, the algorithm may prioritize male applicants in the future.

e. *Measurement bias*, merupakan salah satu jenis bias data, dapat terjadi ketika keakuratan data bervariasi di antara kelompok. Contoh: Model AI menggunakan biaya perawatan kesehatan sebagai *proxy* dalam mengidentifikasi pasien berisiko tinggi mengalami kondisi serius, namun hubungan antara biaya dan risiko bervariasi berdasarkan ras. Misalnya, karena pasien kulit hitam cenderung memiliki biaya medis yang lebih rendah akibat hambatan akses dan kepercayaan terhadap sistem perawatan kesehatan, model AI tersebut secara keliru mengidentifikasi mereka sebagai berisiko rendah, meskipun kondisi kesehatannya mungkin sama dengan pasien non-kulit hitam.

e. Measurement bias, is a type of data bias that can occur when the accuracy of data differs among groups. For example, an AI model using healthcare costs as a proxy to identify high-risk patients for serious conditions may find that the relationship between cost and risk varies by race. Because Black patients tend to have lower medical costs due to access barriers and distrust of the healthcare system, the AI model may incorrectly identify them as low-risk, even though their health conditions may be similar to those of non-Black patients.

f. *Interaction bias*, terjadi ketika sistem AI berinteraksi dengan manusia secara bias, sehingga mengakibatkan perlakuan tidak adil. Contoh: *Chatbot* yang merespons pria dan wanita secara berbeda, dapat mengakibatkan komunikasi yang bias.

f. Interaction bias, occurs when AI systems interact with humans in a biased manner, resulting in unfair treatment. For example, a chatbot that responds more positively to men than to women may lead to biased communication.

D. Keamanan Siber

Selain risiko yang dapat ditimbulkan akibat penggunaan AI, terdapat pula isu terkait keamanan dari AI itu sendiri. *Cyber security risk* meliputi bagaimana peretas melakukan serangan dengan memanipulasi input dari model AI sehingga model menghasilkan *output* yang tidak sesuai/merugikan. Serangan dimaksud secara garis besar terbagi menjadi beberapa kategori, yaitu:

a. *Data privacy attacks*. Dalam serangan privasi data, peretas menyimpulkan informasi sensitif dari kumpulan data pelatihan dengan menganalisis parameter atau mengajukan permintaan (*query*) pada model untuk mendapatkan data/informasi sensitif dimaksud.

b. *Adversarial inputs*. Peretas berpotensi menggunakan input atau muatan yang secara eksplisit dirancang untuk melewati pengklasifikasi sistem AI, hal ini dapat menghasilkan *output* AI yang tidak sesuai.

c. *Model extraction*. Dalam serangan ini, peretas mencoba mereplikasi model itu sendiri atau memahami cara kerja sebuah model AI tanpa izin. Hal ini dapat mengancam integritas dan kerahasiaan model.

D. Cyber Security

In addition to the risks posed by AI, there are also issues related to the security of AI itself. Cybersecurity risks include how hackers conduct attacks by manipulating the inputs of AI models, causing the models to produce outputs that are incorrect or harmful. These attacks can be broadly categorized into several types, namely:

a. Data privacy attacks. In data privacy attacks, hackers infer sensitive information from training datasets by analyzing parameters or submitting queries to the model to obtain the sensitive data/information in question.

b. Adversarial inputs. Hackers may potentially use inputs or payloads that are explicitly designed to bypass the AI system's classifiers, which can result in incorrect AI outputs.

c. Model extraction. In this type of attack, hackers attempt to replicate the model itself or understand how an AI model works without permission. This can threaten the integrity and confidentiality of the model.

d. *Training data poisoning*. *Data poisoning* adalah kontaminasi (penyuntikan data yang rusak atau berbahaya) ke dalam *dataset* yang digunakan untuk melatih sistem AI yang dapat berdampak negatif terhadap proses pembelajaran atau *output* yang dihasilkan.

E. *Human Issues* & Tantangan Lain Dalam Penggunaan AI

1. Kepemimpinan

Kepemimpinan memegang peran penting dalam keberhasilan implementasi dan adopsi AI dalam organisasi. Namun, banyak pemimpin menghadapi tantangan seperti kurangnya keterampilan digital,

d. Training data poisoning. Data poisoning is the contamination (injection of corrupted or malicious data) into the dataset used to train AI systems, which can negatively impact the learning process or the outputs produced.

E. Human Issues & Other Challenges in the Use of AI

1. Leadership

Leadership plays a crucial role in the successful implementation and adoption of AI within organizations. However, many leaders face challenges such as a lack of digital skills, limited understanding of AI technology, and minimal experience

pemahaman yang terbatas mengenai teknologi AI, dan minimnya pengalaman dalam mengelola sistem AI sehingga memperlambat proses adopsi AI secara efektif.

2. Pengetahuan Terkait AI

a. Kekurangan Pendidikan: Dibutuhkan program pendidikan dan pelatihan AI yang komprehensif untuk menjembatani kesenjangan dalam pemahaman dan keterampilan pada penggunaan AI.

b. Kurikulum yang Ketinggalan: Integrasi AI yang lambat dalam sistem pendidikan menghasilkan pelatihan yang tidak memadai, sehingga karyawan, bahkan di perusahaan teknologi canggih, sering kali tidak siap untuk bekerja dengan sistem AI.

c. Kekurangan Talenta: Kelangkaan tenaga ahli AI yang terampil menghambat adopsi, menghalangi inovasi, dan meningkatkan ketergantungan pada sekelompok kecil ahli dengan biaya yang tinggi.

3. Ketergantungan terhadap AI

Meskipun AI memberikan manfaat besar dalam hal produktivitas, ketergantungan berlebih pada AI dapat menciptakan beberapa risiko, seperti:

a. Tantangan Kolaborasi: Efisiensi dan kecepatan yang dihasilkan oleh sistem AI dapat mengurangi nilai kerja kolaboratif manusia dari waktu ke waktu.

b. Kekhawatiran Identitas Profesional: Penggunaan AI dapat memengaruhi identitas profesional seseorang, yang pada akhirnya dapat menghambat implementasi AI di tempat kerja.

4. Kehilangan Pekerjaan

Automasi yang didorong oleh AI menimbulkan ancaman terhadap pekerja dengan keterampilan rendah. Seiring perkembangan teknologi AI yang terus meningkatkan efisiensi, tenaga kerja harus beradaptasi dengan menguasai keterampilan baru agar tetap relevan dalam lanskap yang terus berubah.

5. Transparansi

Dalam hal transparansi, AI yang dianggap ideal adalah yang memiliki kemampuan untuk menjelaskan proses pengambilan keputusan. Hal ini penting agar manajemen memiliki kepercayaan terhadap model AI yang digunakan. Namun, tantangan muncul ketika sistem AI, khususnya yang berbasis *deep*

in managing AI systems, which can slow down the process of effectively adopting AI.

2. AI Knowledge

a. Lack of Education: Comprehensive AI education and training programs are needed to bridge the gap in understanding and skills related to the use of AI.

b. Outdated Curriculum: The slow integration of AI into the education system results in inadequate training, causing employees, even in advanced technology companies, to often be unprepared to work with AI systems.

c. Talent Shortage: The scarcity of skilled AI professionals hinders adoption, obstructs innovation, and increases reliance on a small group of experts, driving up costs.

3. Overreliance on AI

Although AI offers significant benefits in terms of productivity, excessive reliance on AI can create several risks, such as:

a. Collaboration Challenges: The efficiency and speed generated by AI systems can diminish the value of human collaborative work over time.

b. Concerns About Professional Identity: The use of AI can impact an individual's professional identity, which may ultimately hinder the implementation of AI in the workplace.

4. Job Displacement

AI-driven automation poses a threat to low-skilled workers. As AI technology continues to evolve and enhance efficiency, the workforce must adapt by acquiring new skills to remain relevant in the ever-changing landscape.

5. Transparency

In terms of transparency, the ideal AI is one with the the ability to explain its decision-making processes. This is important for management to have confidence in the AI models being used. However, challenges arise when AI systems, particularly those based on deep learning, become increasingly

learning, menjadi semakin kompleks. Kompleksitas ini membuat proses pengambilan keputusan oleh sistem AI sulit untuk dipahami, sehingga dapat menimbulkan keraguan dan resistensi dalam mengadopsi teknologi tersebut.

6. Kebijakan terkait AI

a. Inovasi AI yang berkembang pesat sering kali tidak sejalan dengan lambatnya perkembangan struktur hukum, sehingga menciptakan kekosongan regulasi. Akibatnya, perusahaan sering kali terjebak dalam kerangka hukum yang bersifat tambal sulam tanpa panduan yang preskriptif.

b. Tidak adanya *fiduciary duties* dalam penggunaan AI dapat menyebabkan perusahaan lebih memprioritaskan efisiensi dan profitabilitas dibandingkan dengan pertimbangan etis atau kepentingan publik.

c. Tantangan lainnya adalah kesulitan dalam menyelaraskan kebijakan AI dengan berbagai peraturan di tingkat global. Perbedaan regulasi, seperti perlindungan data pribadi, memaksa perusahaan untuk menerapkan model AI yang berbeda-beda di setiap pasar. Fragmentasi ini tidak hanya menambah kerumitan operasional tetapi juga menurunkan efisiensi perusahaan.

complex. This complexity makes the decision-making processes of AI systems difficult to understand, potentially leading to doubt and resistance in adopting the technology.

6. AI Policies

a. Rapidly evolving AI innovations often outpace the development of legal frameworks, creating regulatory gaps. As a result, companies often find themselves trapped in a patchwork legal framework without prescriptive guidance.

b. The absence of fiduciary duties in the use of AI can lead companies to prioritize efficiency and profitability over ethical considerations or public interest.

c. Another challenge is the difficulty in aligning AI policies with various regulations at the global level. Regulatory differences, such as data protection laws, require companies to implement different AI models in each market. This fragmentation not only adds operational complexity but also reduces the efficiency of the company.

F. Potensi Risiko oleh Gen-AI dan *Trustworthy AI Characteristics*

Dalam konteks GenAI, terdapat beberapa risiko atas penggunaan GenAI yang dapat "memperburuk" *existing risk* dari AI dan menciptakan risiko yang unik atau spesifik. Amerika Serikat melalui National Institute of Standards and Technology (NIST) telah menerbitkan *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* untuk mengatasi tantangan dan risiko unik yang terkait dengan sistem GenAI. Sistem ini, seperti *large language model* (LLM) dan generator gambar, menghadirkan kompleksitas yang sangat berbeda dari sistem AI tradisional.

Dokumen ini mendefinisikan risiko-risiko baru atas penggunaan GenAI, serta memberikan serangkaian tindakan yang disarankan (*suggested actions*) untuk membantu organisasi mengatur, memetakan, mengukur, dan mengelola risiko-risiko tersebut. Untuk masing-masing risiko, dokumen ini juga mengidentifikasi karakteristik *Trustworthy* AI seperti apa yang dibutuhkan agar risiko tersebut dapat dikelola dengan baik.

F. Potential Risks of Gen-AI and Trustworthy AI Characteristics

In the context of GenAI, there are several risks associated with the use of GenAI that can "exacerbate" existing risks of AI and create unique or specific risks. The United States, through the National Institute of Standards and Technology (NIST), has published the Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile to address the unique challenges and risks associated with GenAI systems. These systems, such as large language models (LLMs) and image generators, present complexities that are significantly different from traditional AI systems.

The document defines the new risks associated with the use of GenAI and provides a set of suggested actions to help organizations regulate, map, measure, and manage these risks. For each risk, the document also identifies the characteristics of Trustworthy AI that are necessary to effectively manage those risks.

Tabel 4. Risiko GenAI dan Karakteristik *Trustworthy* AI yang Relevan

Table 4. GenAI Risks and Relevant Characteristics of a Trustworthy AI

GenAI Risks	Details	Relevant Characterstics of a Trustworthy AI
CBRN Information or Capabilities	Enabling criminals to more easily access chemical, biological, radiological, or nuclear (CBRN) weapons and/or related knowledge, information, materials, tools, or technologies that can be misused to assist in the design, development, production, or use of CBRN weapons or other hazardous materials. Large language models (LLMs) can facilitate the analysis or synthesis of this information, particularly by individuals who lack formal scientific training or expertise.	Safe, Explainable and Interpretable
Confabulation	The production of content that appears to be true but is incorrect or false (known as "hallucination") can mislead or deceive users.	Fair with Harmful Bias Managed, Safe, Valid and Reliable, Explainable and Interpretable
Dangerous, Violent, or Hateful Content	Facilitating the production and access to content related to violence, incitement, radicalization, or threats, as well as recommendations for self-harm or engaging in criminal/illegal activities.	Safe, Secure and Resilient
Data Privacy	During a cyber attack, large language models (LLMs) can reveal sensitive information that has been included in the training data of the system. This issue is referred to as "data memorization," and it can pose significant privacy risks even for small-scale training sample data.	Accountable and Transparent, Privacy Enhanced, Safe, Secure and Resilient
Environmental Impacts	The high utilization of computational resources in the training and operation of GenAI models can have negative environmental impacts due to carbon emissions.	Accountable and Transparent, Safe
Harmful Bias or Homogenization	GenAI systems can amplify the occurrence of bias, potentially having adverse effects on individuals, groups, communities, organizations, and society. For example, when asked to create images of a CEO, doctor, lawyer, and judge, current text-to-image models often underrepresented women and/or racial minorities, as well as individuals with disabilities.	Fair with Harmful Bias Managed, Valid and Reliable
Human-AI Configuration	The regulation or interaction between humans and AI systems that can encourage humans to improperly anthropomorphize or form emotional attachments to GenAI systems	Accountable and Transparent, Explainable and Interpretable, Fair with Harmful Bias Managed, Privacy Enhanced, Safe, Valid and Reliable
Information Integrity	Facilitating the generation of false or inaccurate content or information that can lead to the widespread dissemination of disinformation or misinformation. This can erode public trust in accurate and valid information.	Accountable and Transparent, Safe, Valid and Reliable, Interpretable and Explainable
Information Security	Previously, cyber attacks often required high-level technical expertise. However, with the assistance of GenAI, non-technical users could potentially design and launch cyber attacks with simple instructions. Additionally, GenAI can be used to automatically analyze software code and identify vulnerabilities that can be exploited. This could lead to an increase in the number of cyber attacks.	Privacy Enhanced, Safe, Secure and Resilient, Valid and Reliable

GenAI Risks	Details	Relevant Characterstics of a Trustworthy AI
Intellectual Property	Facilitating the production or replication of copyrighted content, trademarks, or licensed materials without permission, as well as enabling plagiarism or illegal replication.	Accountable and Transparent, Fair with Harmful Bias Managed, Privacy Enhanced
Obscene, Degrading, and/ or Abusive Content	GenAI can facilitate the production and access to non-consensual intimate imagery (NCII) and child sexual abuse material (CSAM). The obscene, crude, or dehumanizing content generated by GenAI can cause harm to the privacy, psychological, and emotional well-being of specific individuals or groups.	Fair with Harmful Bias Managed, Safe, Privacy Enhanced
Value Chain and Component Integration.	The GenAI value chain involves various third-party components such as datasets, models, and software that may be obtained or used without proper procedures, thereby reducing transparency and accountability. This complexity of integration also makes it difficult to identify the sources of issues within the system.	Accountable and Transparent, Explainable and Interpretable, Fair with Harmful Bias Managed, Privacy Enhanced, Safe, Secure and Resilient, Valid and Reliable

Source: NIST - AI RMF

G. AI dan Stabilitas Sistem Keuangan

Pada 14 November 2024, Financial Stability Board (FSB) menerbitkan laporan berjudul "The Financial Stability Implications of Artificial Intelligence" yang membahas penerapan AI di sektor keuangan serta dampaknya terhadap stabilitas keuangan. Laporan tersebut menguraikan bahwa AI menawarkan berbagai manfaat, termasuk peningkatan efisiensi operasional, kepatuhan peraturan, produk keuangan yang dipersonalisasi, dan analisis data tingkat lanjut. Namun, AI juga dapat menimbulkan kerentanan tertentu yang berpotensi meningkatkan risiko sistemik dan mengancam stabilitas keuangan. Beberapa kerentanan utama yang diidentifikasi oleh FSB meliputi:

G. AI and The Stability of Financial Systems

On November 14, 2024, the Financial Stability Board (FSB) published a report titled "The Financial Stability Implications of Artificial Intelligence," which discusses the application of AI in the financial sector and its impact on financial stability. The report outlines that AI offers various benefits, including enhanced operational efficiency, regulatory compliance, personalized financial products, and advanced data analysis. However, AI can also introduce certain vulnerabilities that may increase systemic risks and threaten financial stability. Some of the key vulnerabilities identified by the FSB include:

1. Ketergantungan dan Konsentrasi Pihak Ketiga

Penggunaan AI di sektor keuangan sering kali melibatkan layanan dari pihak ketiga, seperti penyedia komputasi *cloud*, model AI, atau perangkat lunak yang dikembangkan oleh pihak eksternal. Ketergantungan besar pada sejumlah kecil penyedia layanan tersebut dapat menciptakan risiko konsentrasi. Jika terjadi gangguan, kegagalan, atau insiden keamanan pada penyedia tersebut, dampaknya dapat meluas ke seluruh sektor keuangan. Selain itu, kurangnya diversifikasi penyedia layanan membatasi kemampuan institusi keuangan untuk memitigasi risiko dan mengurangi ketahanan operasional.

1. Dependency and Concentration of Third Parties

The use of AI in the financial sector often involves services from third parties, such as cloud computing providers, AI models, or software developed by external entities. A significant reliance on a small number of these service providers can create concentration risks. In the event of disruptions, failures, or security incidents at these providers, the impact can extend throughout the financial sector. Additionally, the lack of diversification among service providers limits the ability of financial institutions to mitigate risks and undermines operational resilience.

2. Korelasi Pasar

AI digunakan secara luas untuk pemodelan risiko, analisis pasar, dan pengambilan keputusan otomatis di berbagai lembaga keuangan. Penggunaan algoritma serupa oleh berbagai institusi keuangan dapat menyebabkan perilaku homogen (*herding behavior*) di mana banyak institusi bereaksi serupa terhadap kejadian tertentu, sehingga memperburuk volatilitas pasar, mengakibatkan ketidakseimbangan pasar atau penurunan likuiditas secara tiba-tiba.

3. Risiko Siber

Seperti yang telah diuraikan di atas, penerapan AI di sektor keuangan dapat memperluas potensi serangan oleh peretas. Infrastruktur AI yang terintegrasi dengan sistem keuangan yang bersifat kritikal juga dapat menjadi target serangan siber yang berdampak signifikan terhadap keamanan dan stabilitas keuangan.

4. Risiko Model, Kualitas Data, dan Tata Kelola

AI yang digunakan dalam sektor keuangan sangat bergantung pada kualitas data, desain model, dan tata kelola yang efektif. Data yang tidak lengkap, bias, atau tidak akurat dapat mengakibatkan hasil analisis yang keliru dan keputusan yang salah. Selain itu, model AI yang kompleks dan tidak transparan (*black box models*) sulit diaudit dan divalidasi, sehingga mengurangi akuntabilitas dan mengakibatkan hasil yang tidak dapat diandalkan.

2. Market Correlation

AI is widely used for risk modeling, market analysis, and automated decision-making across various financial institutions. The use of similar algorithms by different financial institutions can lead to homogeneous behavior (herding behavior), where many institutions respond similarly to certain events, thereby exacerbating market volatility and resulting in market imbalances or sudden declines in liquidity.

3. Cyber Risks

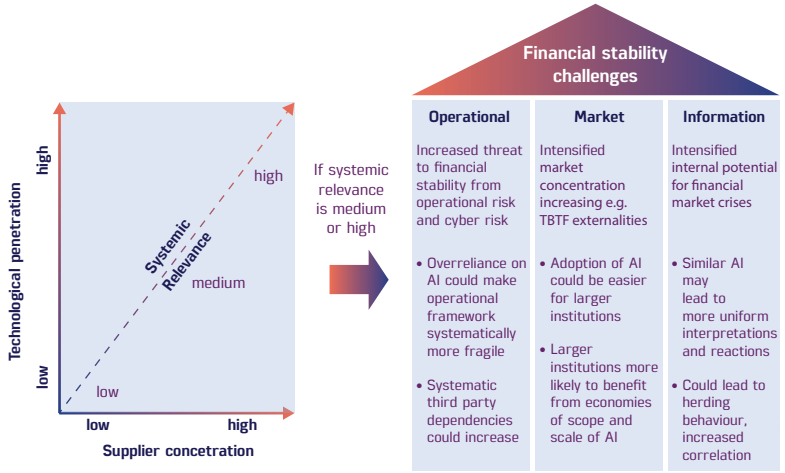
As outlined above, the application of AI in the financial sector can increase the potential for attacks by hackers. AI infrastructure integrated with critical financial systems can also become targets of cyber attacks, significantly impacting security and financial stability.

4. Model Risk, Data Quality, and Governance

AI used in the financial sector heavily relies on data quality, model design, and effective governance. Incomplete, biased, or inaccurate data can lead to erroneous analysis results and poor decision-making. Additionally, complex and opaque AI models (black box models) are difficult to audit and validate, thereby reducing accountability and resulting in unreliable outcomes.

Gambar 7. Amplifikasi Sistemik oleh AI dan Tantangan terhadap Stabilitas Keuangan

Figure 7. Systemic Amplification by AI and Challenges to Financial Stability



Source: European Central Bank (2024)

Sejalan dengan FSB, European Central Bank mempublikasi kajian "The rise of artificial intelligence: benefits and risks for financial stability" sebagai

In line with the FSB, the European Central Bank published a study titled "The Rise of Artificial Intelligence: Benefits and Risks for Financial Stability" as part of

bagian dari Financial Stability Review, Mei 2024. Kajian ini menjelaskan bahwa tingkat risiko sistemik akan meningkat seiring dengan meningkatnya penetrasi teknologi AI di lembaga keuangan dan konsentrasi penyedia AI. Jika kedua faktor ini tinggi, relevansi sistemik AI menjadi medium hingga tinggi, yang dapat mengancam stabilitas keuangan melalui berbagai tantangan berikut:

1. Risiko Operasional

a. Ketergantungan yang berlebihan pada AI dapat membuat kerangka operasional menjadi lebih rapuh secara sistemik.

b. Ketergantungan pada pihak ketiga yang bersifat sistemik dapat meningkat, yang mengarah pada potensi *single-point-of-failure* jika piha ketiga tersebut mengalami kegagalan dalam operasional.

2. Risiko Pasar

a. Adopsi AI lebih mudah dilakukan oleh lembaga keuangan yang besar dengan sumber daya lebih baik.

b. Lembaga keuangan yang besar lebih mungkin mendapatkan keuntungan dari skala ekonomi dalam penerapan AI, yang dapat memperlebar kesenjangan antara pemain besar dan kecil. Hal-hal ini dapat meningkatkan konsentrasi pasar yang semakin intensif, termasuk dampak dari eksternalitas *too big to fail.*

3. Risiko Informasi

a. Penggunaan AI serupa secara luas dapat menghasilkan interpretasi dan reaksi yang lebih seragam di pasar.

b. Pola perilaku yang seragam (*herding behaviour*) dan peningkatan korelasi dapat memperburuk volatilitas dan mengurangi ketahanan pasar.

the Financial Stability Review in May 2024. This study explains that the level of systemic risk will increase as the penetration of AI technology in financial institutions and the concentration of AI providers rise. when both of these factors are high, the systemic relevance of AI rises medium to high, which can threaten financial stability through various challenges, including:

1. Operational Risks

a. Excessive dependence on AI can increase systemic fragility within the operational framework.

b. Dependence on systemic third parties may increase, leading to potential single points of failure if those third parties experience operational failures.

2. Market Risk

a. The adoption of AI is easier for larger financial institutions with better resources.

b. Larger financial institutions are more likely to benefit from economies of scale in AI implementation, which can widen the gap between large and small players. These factors may contribute to increased market concentration, including the effect of "too big to fail" externalities.

3. Information Risks

a. Widespread use of similar AI can lead to more uniform interpretations and reactions in the market.

b. Uniform behavior patterns (herding behavior) and increased correlation can exacerbate volatility and reduce market resilience.

HALAMAN INI SENGAJA DIKOSONGKAN

THIS PAGE IS INTENTIONALLY LEFT BLANK

Bab 3
Benchmark Regulasi di Berbagai Negara

Chapter 3
Regulation Benchmark Across Countries

Mempertimbangkan besarnya dampak potensial dari penerapan berbagai sistem AI, diperlukan suatu tata kelola AI (AI *governance*) yang merupakan kerangka hukum, aturan, praktik, dan proses yang digunakan untuk memastikan teknologi AI dikembangkan dan digunakan secara bertanggung jawab.

Sehubungan dengan hal tersebut, beberapa negara dan lembaga telah menerbitkan berbagai regulasi, standar maupun panduan untuk mendukung berbagai pihak dalam adopsi tata kelola AI, dengan menggunakan pendekatan yang beragam. Pendekatan yang dilakukan oleh para pembuat kebijakan dalam penerapan sistem AI antara lain:

a. Kerangka peraturan yang mengatur penggunaan AI, seperti *European Union's Artificial Intelligence Act* (EU AI Act).

b. Regulasi yang secara eksplisit difokuskan pada *generative AI*, seperti *China's Interim Administrative Measures for Generative Artificial Intelligence Services*.

c. Regulator di sektor tertentu, seperti sektor keuangan, juga telah menerbitkan pedoman khusus. Misalnya, *Monetary Authority of Singapore* (MAS) menerbitkan dokumen *Principles to Promote Fairness, Ethics, Accountability and Transparency* (FEAT) *in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector*;

d. Penerbitan standar terkait AI dengan kolaborasi lintas yurisdiksi. Sebagai contoh adalah inisiatif yang didorong oleh *the Organisation for Economic Co-operation and Development* (OECD), *the US National Institute of Standards and Technology* (NIST), *the United Nations Educational, Scientific and Cultural Organization* (UNESCO), *the International Organization for Standardization* (ISO), dan *the Group of Seven* (G7).

e. Strategi regulasi yang beragam, seperti kerangka Uni Eropa yang secara eksplisit mengadopsi metodologi berbasis risiko, regulasi mandiri seperti di Amerika Serikat, regulasi bersama seperti di Eropa, serta kerangka Tiongkok yang terutama mengikuti pendekatan *principle-based* dan cenderung mengadopsi pendekatan regulasi secara *top-down* (*top-down regulatory approach*).

Considering the significant potential impact of implementing various AI systems, an AI governance framework is essential. This framework consists of legal structures, rules, practices, and processes used to ensure that AI technology is developed and used responsibly.

In this regard, several countries and institutions have issued various regulations, standards, and guidelines to support different parties in adopting AI governance, using diverse approaches. The approaches taken by policymakers in the implementation of AI systems include:

a. Regulatory frameworks governing the use of AI, such as the European Union's Artificial Intelligence Act (EU AI Act).

b. Regulations explicitly focused on generative AI, such as China's Interim Administrative Measures for Generative Artificial Intelligence Services.

c. Regulators in specific sectors, such as the financial sector, have also issued specific guidelines. For example, the Monetary Authority of Singapore (MAS) published the document "Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector".

d. The issuance of AI-related standards through cross-jurisdictional collaboration. An example is the initiative driven by the Organisation for Economic Co-operation and Development (OECD), the US National Institute of Standards and Technology (NIST), the United Nations Educational, Scientific and Cultural Organization (UNESCO), the International Organization for Standardization (ISO), and the Group of Seven (G7).

e. Diverse regulatory strategies, such as the European Union framework that explicitly adopts a risk-based methodology, self-regulation as seen in the United States, joint regulation as practiced in Europe, and the Chinese framework that primarily follows a principle-based approach and tends to adopt a top-down regulatory approach.

f. Perangkat instrumen tata kelola AI yang mencakup 4 (empat) pendekatan utama dalam regulasi:

1. Tata kelola industri secara mandiri (*industry self-governance*)

   Merupakan "*private ethical codes and councils*" yang merujuk pada kode etik dan dewan etika yang dibentuk oleh institusi swasta untuk memastikan bahwa aktivitas dan keputusan mereka dalam penerapan AI memenuhi dan mematuhi standar etika tertentu. Contoh, *Microsoft Aether Committee and Responsible AI Standard Playbook, Google AI Principles, Bosch Ethical Guidelines for AI, IBM's AI Ethics Board.*

2. *Soft law* (termasuk standar teknis/technical standards), yang mencakup:

   a. Non-binding international agreements (contoh, OECD/G20 AI Principles, UNESCO Recommendation on the Ethics of AI, G7 Principles).

   b. *National AI principles/ethics frameworks* (contoh, *US White House AI Bill of Rights, Singapore Model AI Governance Framework for Generative AI*).

   c. *Technical standards* (contoh, *IEEE P70xx series, ISO/IEC 23894:2023, NIST AI Risk Management Framework, UK AI Standards Hub*).

f. Governance instruments for AI that encompass 4 (four) main approaches in regulation:

1. Industry self-governance

   This refers to private ethical codes and councils, which are ethical codes and ethics boards established by private institutions to ensure that their activities and decisions in the application of AI meet and comply with certain ethical standards. Examples include the Microsoft Aether Committee and Responsible AI Standard Playbook, Google AI Principles, Bosch Ethical Guidelines for AI, and IBM's AI Ethics Board.

2. Soft law (including technical standards) which covers:

   a. Non-binding international agreements (e.g., OECD/G20 AI Principles, UNESCO Recommendation on the Ethics of AI, G7 Principles).

   b. National AI principles/ethics frameworks (e.g., US White House AI Bill of Rights, Singapore Model AI Governance Framework for Generative AI).

   c. Technical standards (e.g., IEEE P70xx series, ISO/IEC 23894:2023, NIST AI Risk Management Framework, UK AI Standards Hub).

3. *Regulatory sandboxes*

   Regulasi ini untuk mendukung ruang uji coba/pengembangan inovasi sebagai sarana, mekanisme dan lingkungan yang terkendali untuk memfasilitasi uji coba serta pengembangan inovasi sistem dan ekosistem sistem AI. Contoh, Brazil regulatory sandbox pilot for AI and data protection, Singapore AI Verify toolkit.

4. *Hard law,* yang mencakup:

   a. *New horizontal AI law*, yakni undang-undang atau regulasi yang mengatur AI secara luas dan berlaku untuk semua sektor industri, tanpa membedakan jenis penggunaan atau bidang tertentu. Contoh, *EU AI Act, Brazil AI Bill, Chile AI Bill.*

   b. Memperbarui atau menerapkan undang-undang atau regulasi yang ada. Contoh, undang-undang mengenai pelindungan data pribadi, hak asasi manusia, anti diskriminasi, kejahatan siber, kekayaan intelektual, persaingan/antimonopoli, pengadaan.

   c. Undang-undang atau peraturan secara sektoral atau *targeted.* Contoh, *New York City Local Law 144 of 2021 on Automated*

3. Regulatory sandboxes

   This regulation supports innovation testing/development spaces as a means, mechanism, and controlled environment to facilitate the testing and development of AI systems and ecosystems. Examples include Brazil's regulatory sandbox pilot for AI and data protection, and Singapore's AI Verify toolkit.

4. Hard law which covers:

   a. New horizontal AI law, which refers to laws or regulations that govern AI broadly and apply to all industry sectors, without distinguishing between specific types of use or fields. Examples include the EU AI Act, Brazil AI Bill, and Chile AI Bill.

   b. Updating or implementing existing laws or regulations. Examples include laws regarding personal data protection, human rights, anti-discrimination, cybercrime, intellectual property, competition/antitrust, and procurement.

   c. Sectoral or targeted laws or regulations. Examples include New York City Local Law 144 of 2021 on Automated

Employment Decision Tools, US semiconductor export controls, Chinese regulations on recommendation algorithms, 'deep synthesis' technologies, and generative AI.

5. Penerapan AI tunduk pada *existing* regulasi (regulasi yang sudah ada atau sudah berlaku)

   Regulasi dimaksud berfokus lebih dari sekadar teknologi AI itu sendiri, seperti undang-undang yang berfokus pada privasi, antidiskriminasi, keamanan produk. Sebagai contoh, perusahaan yang mengembangkan AI dan menawarkan model dan sistem *generative AI* di Uni Eropa (UE) antara lain juga harus memenuhi ketentuan:

   a. Pelindungan data (*EU Regulation* No. 2016/679 mengenai *General Data Protection Regulation*/GDPR).

   b. Kekayaan intelektual (larangan pelanggaran hak kekayaan intelektual) (EU Directive No.2001/29/EC mengenai *harmonisation of certain aspects of copyright and related rights in the information society*).

Employment Decision Tools, US semiconductor export controls, Chinese regulations on recommendation algorithms, 'deep synthesis' technologies, and generative AI.

5. The application of AI is subject to existing regulations (regulations that are already in place) where the regulations focus on more than just the AI technology itself, such as laws that focus on privacy, anti-discrimination, and product safety. For example, companies developing AI and offering generative AI models and systems in the European Union (EU) must also comply with the provisions of these regulations:

   a. Data protection (EU Regulation No. 2016/679 regarding the General Data Protection Regulation/GDPR).

   b. Intellectual property (prohibition of intellectual property rights infringement) (EU Directive No. 2001/29/EC regarding the harmonization of certain aspects of copyright and related rights in the information society).

   c. Layanan digital yang mencakup larangan penyebaran konten/ujaran yang merugikan (EU Directive No. 2022/2065 mengenai *Digital Services Act*).

Pada 16 Agustus 2024, UNESCO menerbitkan *Consultation Paper on AI Regulation - Emerging Approach Across the World*, yang memperkenalkan konsep-konsep utama regulasi AI, menyajikan lanskap global regulasi AI, dan menguraikan 9 (sembilan) pendekatan regulasi terkait AI dengan memberikan beberapa kasus spesifik dari berbagai negara untuk menggambarkan setiap pendekatan. 9 (sembilan) pendekatan regulasi AI ini tidak eksklusif dan dapat digabung dalam satu atau lebih regulasi, sebagai berikut:

   c. Digital services that include the prohibition of the dissemination of harmful content/speech (EU Directive No. 2022/2065 regarding the Digital Services Act).

On August 16, 2024, UNESCO published the "Consultation Paper on AI Regulation - Emerging Approach Across the World," which introduces key concepts of AI regulation, presents the global landscape of AI regulation, and outlines 9 (nine) regulatory approaches related to AI, providing specific case studies from various countries to illustrate each approach. These 9 (nine) AI regulatory approaches are not exclusive and can be combined in one or more regulations, as follows:

Tabel 5. Pendekatan Regulasi AI

Table 5. AI Regulatory Approaches

1	**Principles-Based Approach**	Providing guidance to stakeholders through a set of fundamental principles aimed at developing and using AI systems ethically, responsibly, human-centric, and in respect of Human Rights (HR). Examples include UNESCO's "Recommendations on the Ethics of AI" and OECD's "Recommendation of the Council on Artificial Intelligence"
2	**Standards-Based Approach**	Delegating (partially or entirely) regulatory authority by the state to organizations that produce technical standards to guide the interpretation and application of mandatory rules. Examples include AI-related standards published by ISO (International Organization for Standardization), IEC (International Electrotechnical Commission), and BSI (British Standards Institution).
3	**Agile and Experimentalist Approach**	Implementing flexible regulatory schemes, such as regulatory sandboxes and other testing environments, that allow organizations to test business models, methods, infrastructure, and tools under more flexible regulatory conditions with oversight and guidance from public authorities. Examples include various regulatory sandboxes designed to support innovation (related to AI).
4	**Facilitating and Enabling Approach**	Facilitating and enabling an environment that encourages all stakeholders involved in the AI lifecycle to develop and use AI systems responsibly, ethically, and in compliance with Human Rights.
5	**Adapting Existing Laws Approach**	Amending sector-specific regulations (e.g., health, finance, education, justice) and cross-sector regulations (e.g., criminal law, procurement, data protection laws, labor laws) to make gradual improvements to the existing regulatory framework
6	**Access to Information and Transparency Mandates Approach**	Requiring the implementation of transparency instruments that allow the public to access basic information about AI systems.
7	**Risk-Based Approach**	Establishing obligations and requirements in accordance with the risk assessment associated with the implementation and use of specific AI tools in particular contexts.
8	**Rights-Based Approach**	Establishing obligations or requirements to protect the rights and freedoms of individuals.
9	**Liability Approach**	stablishing responsibilities and sanctions for the use of problematic AI systems. For example, the EU AI Act in the European Union imposes administrative fines of up to EUR 35 million for violations of the law.

Pertumbuhan jumlah dan variasi regulasi, standar, dan panduan AI diperkirakan akan terus berlanjut di masa mendatang. Hal ini didasarkan atas pertimbangan bahwa para pembuat kebijakan akan terus memberikan perhatian penuh terhadap pengelolaan risiko AI (termasuk di sektor perbankan) seiring dengan perkembangan teknologi dan inovasi AI itu sendiri, serta untuk mendorong inovasi yang bertanggung jawab sekaligus melindungi pengguna AI.

Saat ini, belum ada regulasi, standar maupun panduan yang dapat di klaim sebagai "*best practices*" atau menjadi pendekatan tunggal dalam tata kelola AI, karena sistem AI yang masih terus berkembang. Berbagai regulasi, standar maupun panduan yang ada, seharusnya saling memperkuat dan sering kali bersinggungan dengan berbagai perangkat regulasi dan kebijakan lainnya.

The growth in the number and variety of AI regulations, standards, and guidelines is expected to continue in the future. This is based on the consideration that policymakers will maintain a strong focus on managing AI risks (including in the banking sector) as technology and AI innovations evolve, while also promoting responsible innovation and protecting AI users.

Currently, there are no regulations, standards, or guidelines that can be claimed as "best practices" or serve as a singular approach to AI governance, due to the continuously evolving nature of AI systems. The various existing regulations, standards, and guidelines should reinforce each other and often intersect with other regulatory instruments and policies.

Selain itu, berbagai perangkat regulasi dan kebijakan tersebut, terlepas dari kelebihan dan kelemahannya, dapat menjadi opsi dan referensi yang dapat disesuaikan, dimodifikasi, dan diadopsi oleh pembuat kebijakan sesuai dengan konteks kebutuhan. Proses ini perlu dilakukan dengan mengedepankan prinsip kehati-hatian dan melibatkan berbagai pemangku kepentingan (*meaningful participation*).

A. Indonesia

Indonesia sendiri saat ini telah memiliki beberapa publikasi atau pedoman terkait implementasi AI yang tertuang dalam:

1. Strategi Nasional Kecerdasan Artifisial Indonesia 2020-2045
   a. Strategi nasional kecerdasan artifisial difokuskan pada 4 (empat) area fokus, yaitu:
      1. Etika dan Kebijakan.
      2. Pengembangan Talenta.
      3. Infrastruktur dan Data.
      4. Riset dan Inovasi Industri.
   b. Selain area-area fokus, pemerintah Indonesia menetapkan 5 (lima) bidang prioritas untuk mensukseskan misi-misi strategi nasional Indonesia untuk kecerdasan artifisial, yaitu:
      1. Layanan Kesehatan.
      2. Reformasi Birokrasi.
      3. Pendidikan dan Riset.
      4. Ketahanan Pangan.
      5. Mobilitas dan Kota Pintar.
   c. Nilai-nilai etika kecerdasan artifisial:
      1. Berorientasi pada kemaslahatan umat manusia (manusia sebagai pengawas, kekokohan dan keamanan teknis, tata kelola data dan privasi, transparansi; kesejahteraan sosial dan lingkungan, keanekaragaman, non-diskriminasi, dan keadilan).
      2. Bernafaskan nilai-nilai Pancasila.
      3. Andal, aman, dan terbuka, dapat dipertanggungjawabkan.
      4. Sinergitas antara pemangku kepentingan.
      5. Penerapan asas-asas UU No.11 Tahun 2019 (keimanan dan ketakwaan kepada Tuhan Yang Maha Esa, kemanusiaan, keadilan, kemaslahatan, keamanan dan keselamatan, kebenaran ilmiah, transparansi, aksesibilitas, penghormatan terhadap pengetahuan tradisional dan kearifan lokal, kedaulatan negara).
2. Surat Edaran Menteri Komunikasi dan Informatika No.9 Tahun 2023 tentang Etika Kecerdasan Artifisial.
   a. Maksud dari Surat Edaran Menteri ini sebagai pedoman etika dalam:

In addition, these various regulatory instruments and policies, despite their strengths and weaknesses, can serve as options and references that can be adapted, modified, and adopted by policymakers according to contextual needs. This process should prioritize the cautionary principle and involve meaningful participation by various stakeholders.

A. Indonesia

Indonesia itself currently has several publications or guidelines related to the implementation of AI, which are outlined in:

1. Indonesian National Strategy on Artificial Intelligence 2020-2045
   a. The national artificial intelligence strategy is focused on 4 (four) key areas:
      1. Ethics and Policy.
      2. Talent Development.
      3. Infrastructure and Data.
      4. Research and Industrial Innovation.
   b. In addition to the focus areas, the Indonesian government has established 5 (five) priority sectors to successfully achieve the missions of the national strategy for artificial intelligence, namely:
      1. Healthcare Services.
      2. Bureaucratic Reform.
      3. Education and Research.
      4. Food Security.
      5. Mobility and Smart Cities.
   c. Ethical Values of Artificial Intelligence:
      1. Human-centered for the common good (humans as overseers, technical robustness and safety, data governance and privacy, transparency; social and environmental welfare, diversity, non-discrimination, and justice).
      2. Incorporates the values of Pancasila.
      3. Reliable, safe, open, and accountable.
      4. Synergy among stakeholders. Application of the principles of Law No. 11 of 2019 (faith and devotion to God Almighty, humanity, justice, the common good, security and safety, scientific truth, transparency, accessibility, respect for traditional knowledge and local wisdom, and national sovereignty).
2. Circular Letter of the Minister of Communication and Information No. 9 of 2023 on Artificial Intelligence Ethics.
   a. The purpose of this Ministerial Circular Letter is to serve as an ethical guideline in:

1. Membuat dan merumuskan kebijakan internal perusahaan, penyelenggara sistem elektronik lingkup publik dan penyelenggara sistem elektronik lingkup privat mengenai data dan etika internal kecerdasan artificial.
2. Pelaksanaan konsultasi, analisis, dan pemrograman yang berbasis kecerdasan artifisial sesuai dengan ketentuan peraturan perundang-perundangan.

b. Tujuan dari Surat Edaran Menteri ini untuk memberikan acuan nilai dan prinsip etika bagi pelaku usaha, penyelenggara sistem elektronik lingkup publik, dan penyelenggara sistem elektronik lingkup privat yang memiliki aktivitas pemrograman berbasis kecerdasan artifisial.

c. Surat Edaran ini memuat antara lain:

1. Penyelenggaraan kemampuan Kecerdasan Artifisial mencakup kegiatan konsultasi, analisis, dan pemrograman. Penggunaan teknologi Kecerdasan Artifisial termasuk ke dalam subset dari *machine learning, natural language processing, expert system, deep learning, robotics, neural networks*, dan *subset* lainnya.
2. Penyelenggaraan teknologi Kecerdasan Artifisial memperhatikan nilai Etika Kecerdasan Artifisial meliputi, 1) Inklusivitas 2) Kemanusiaan 3) Keamanan 4) Aksesibilitas 5) Transparansi 6) Kredibilitas dan Akuntabilitas 7) Pelindungan Data Pribadi 8) Pembangunan dan Lingkungan Berkelanjutan 9) Kekayaan Intelektual.
3. Pelaksanaan dan tanggung jawab dalam penyelenggaraan Kecerdasan Artifisial.

1. Formulating and establishing internal policies for companies, public electronic system organizers, and private electronic system organizers regarding data and internal ethics of artificial intelligence.
2. The implementation of consultations, analyses, and programming based on artificial intelligence in accordance with the provisions of laws and regulations.

b. The purpose of this Ministerial Circular Letter is to provide a reference for values and ethical principles for business actors, public electronic system organizers, and private electronic system organizers that engage in artificial intelligence-based programming activities.

c. The Ministerial Circular Letter contains, among others:

1. The implementation of artificial intelligence capabilities includes activities such as consultations, analyses, and programming. The use of artificial intelligence technology falls within subsets of machine learning, natural language processing, expert systems, deep learning, robotics, neural networks, and other subsets.
2. The implementation of artificial intelligence technology considers the values of artificial intelligence ethics, which include: 1) Inclusivity 2) Humanity 3) Security 4) Accessibility 5) Transparency 6) Credibility and Accountability 7) Protection of Personal Data 8) Sustainable Development and Environment 9) Intellectual Property.
3. Implementation and responsibilities in the administration of artificial intelligence.

3. OJK bersama dengan beberapa asosiasi *fintech* (*financial technology*) juga telah menerbitkan panduan terkait AI bagi industri teknologi keuangan, yaitu Panduan Kode Etik Kecerdasan Artifisial (*Artificial Intelligence*/ AI) yang Bertanggung Jawab dan Terpercaya di Industri Teknologi Finansial. Panduan ini disusun dalam rangka mitigasi risiko dan mengoptimalkan AI di industri *fintech* yang memerlukan kerangka perilaku (*code of conduct*) yang dapat menjadi panduan bagi Penyelenggara *Fintech* dan pihak terkait untuk memastikan aplikasi berbasis AI yang digunakan telah memenuhi prinsip-prinsip berasaskan Pancasila, bermanfaat (*beneficial*), wajar dan akuntabel (*fair and accountable*), transparan dan dapat dijelaskan (*transparent and explicable*), ketangguhan dan keamanan (*robustness and security*).

3. The Financial Services Authority (OJK), together with several fintech associations, has also published guidelines related to AI for the financial technology industry, namely the Code of Ethics for Responsible and Trustworthy Artificial Intelligence (AI) in the Financial Technology Industry. This guideline is developed to mitigate risks and optimize AI in the fintech industry, which requires a code of conduct that can serve as a reference for fintech organizers and related parties to ensure that AI-based applications used adhere to the principles based on Pancasila, are beneficial, fair and accountable, transparent and explicable, have robustness and security.

B. European Union's Artificial Intelligence Act (EU AI Act)

Undang-undang AI yang berlaku bagi Uni Eropa (*EU Regulation* No. 2024/1689) ditandatangani anggota parlemen Uni Eropa (UE) pada 13 Juni 2024 dan mulai berlaku sejak 1 Agustus 2024, dan merupakan undang-undang pertama di dunia yang mengatur tentang AI serta disusun secara komprehensif dan mengikat, yang menetapkan kerangka kerja untuk penggunaan dan penyediaan sistem AI di UE.

Dalam undang-undang ini, sistem AI didefinisikan sebagai sistem berbasis mesin yang dirancang untuk beroperasi dengan berbagai tingkat otonomi dan dapat menunjukkan kemampuan adaptasi setelah diterapkan. Sistem ini, baik dengan tujuan yang eksplisit maupun implisit, mengolah *input* yang diterima untuk menghasilkan *output* berupa prediksi, konten, rekomendasi, atau keputusan yang dapat memengaruhi lingkungan fisik maupun virtual.

Sistem AI dirancang untuk beroperasi dengan berbagai tingkat otonomi yang bervariasi, yang berarti sistem AI memiliki tingkat kemandirian tertentu dalam bertindak tanpa keterlibatan manusia serta kemampuan untuk beroperasi tanpa intervensi manusia. Adaptasi yang dapat ditunjukkan oleh sistem AI setelah diterapkan mengacu pada kemampuannya belajar mandiri serta memungkinkan sistem untuk berubah selama penggunaannya.

Sistem AI dapat digunakan secara mandiri atau sebagai bagian dari suatu produk, baik yang terintegrasi secara fisik ke dalam produk tersebut (*embedded*). Contohnya adalah sistem AI yang tertanam dalam kendaraan otonom (*self-driving* atau *autonomous vehicles*) untuk navigasi, pengenalan objek, dan pengambilan keputusan. Selain itu, terdapat pula sistem AI yang berfungsi untuk produk tanpa perlu terintegrasi secara fisik (*non-embedded*). Misalnya, AI pada bank digital yang mendukung pelaksanaan deteksi *fraud* (*fraud detection*), di mana AI tidak tertanam dalam perangkat fisik bank, melainkan beroperasi melalui sistem *cloud* untuk mendeteksi transaksi yang mencurigakan

B. European Union's Artificial Intelligence Act (EU AI Act)

The AI law applicable to the European Union (EU Regulation No. 2024/1689) was signed by members of the European Parliament (EP) on June 13, 2024, and will come into effect on August 1, 2024. It is the first law in the world that regulates AI comprehensively and bindingly, establishing a framework for the use and provision of AI systems in the EU.

In this law, AI systems are defined as machine-based systems designed to operate with varying levels of autonomy and can demonstrate adaptive capabilities after being deployed. These systems, whether with explicit or implicit purposes, process the inputs they receive to generate outputs in the form of predictions, content, recommendations, or decisions that can affect both physical and virtual environments.

AI systems are designed to operate with varying levels of autonomy, which means that AI systems possess a certain degree of independence in acting without human involvement, as well as the ability to function without human intervention. The adaptation that AI systems can demonstrate after deployment refers to their ability to learn independently, allowing the systems to change during their usage.

AI systems can be used independently or as part of a product, either physically integrated into the product (embedded) or functioning separately (non-embedded). An example of an embedded AI system is the AI integrated into autonomous vehicles for navigation, object recognition, and decision-making. Additionally, there are non-embedded AI systems that support products without needing physical integration. For instance, AI in digital banking that facilitates fraud detection operates through cloud systems to identify suspicious transactions, rather than being embedded in the bank's physical devices.

Undang-undang ini mengatur dan mempromosikan penggunaan sistem AI di pasar Eropa secara aman dengan menghormati hak asasi manusia, melindungi kesehatan, keselamatan, dan lingkungan. Selain itu, undang-undang ini juga menetapkan aturan mengenai cara AI dapat dipasarkan, digunakan, dan dipantau di Uni Eropa, melarang praktik-praktik AI tertentu, serta mendukung inovasi.

Lebih detil, beberapa aspek pengaturan lain dalam EU AI Act meliputi:

1. Definisi/terminologi:
   a. Perbedaan model AI (*AI model*) dan sistem AI adalah model AI terutama mengacu pada *general-purpose AI model*/GPAI model yang digunakan untuk dapat diadaptasi dalam membangun berbagai sistem AI. Sebagai contoh, *large language models* (LLM) GPT-4 adalah model AI. *Chatbot* ChatGPT yang dibangun di atas GPT-4 adalah sistem AI.
   b. Penyedia (*provider*) adalah pihak (individu/lembaga) yang membuat, mengembangkan, menyediakan sistem AI atau model AI (GPAI model). Biasanya perusahaan *software* atau *hardware* di bidang AI.
   c. Pengguna (*deployer*) adalah pihak (individu/lembaga) yang menggunakan sistem AI. Sebagai contoh, perusahaan yang menyediakan sistem AI kepada karyawannya untuk optimalisasi operasional perusahaan, atau lembaga keuangan yang membeli dan menggunakan *Chatbot* AI untuk mendukung layanan pelanggan, disebut sebagai *deployer*.
2. Penyedia dan pengguna sistem AI harus memastikan tingkat literasi AI yang memadai terkait penggunaan sistem AI, termasuk pihak yang akan terkena dampaknya.
3. Kategori risiko dari sistem AI:

This law regulates and promotes the safe use of AI systems in the European market while respecting human rights and protecting health, safety, and the environment. Furthermore, this law establishes rules regarding how AI can be marketed, used, and monitored in the European Union, prohibits certain AI practices, and supports innovation.

More specifically, several other regulatory aspects within the EU AI Act include:

1. Definitions/Terminology:
   a. The distinction between AI models and AI systems is that AI models primarily refer to general-purpose AI models (GPAI) that can be adapted to build various AI systems. For example, large language models (LLM) like GPT-4 are considered AI models. In contrast, the ChatGPT chatbot built on GPT-4 is classified as an AI system.
   b. Providers are parties (individuals or institutions) that create, develop, and supply AI systems or AI models (GPAI models). Typically, these are software or hardware companies operating in the field of AI.
   c. Users (deployers) are parties (individuals or institutions) that utilize AI systems. For example, a company that provides AI systems to its employees for operational optimization, or a financial institution that purchases and uses an AI chatbot to support customer service, is referred to as a deployer.
2. Providers and users of AI systems must ensure an adequate level of AI literacy related to the use of AI systems, including the parties that will be affected by their implementation.
3. Risk Categories of AI Systems:

Gambar 8. Klasifikasi Tingkat Risiko Sistem AI

Figure 8. Classification of AI System Risk Levels

Source: EU AI Act

a. Risiko tidak dapat diterima (*unacceptable risk*)

untuk melindungi hak-hak *fundamental* dan keselamatan masyarakat, UE melarang sistem AI yang memiliki risiko tinggi dan melanggar hak-hak fundamental, seperti:

1. Teknik manipulatif atau subliminal, dimana menggunakan teknik di luar kesadaran individu untuk memanipulasi keputusan mereka dengan cara yang merugikan.
2. Eksploitasi kerentanan, dengan memanfaatkan kerentanan individu (usia, disabilitas, atau kondisi sosial-ekonomi) untuk memanipulasi perilaku mereka sehingga menyebabkan kerugian signifikan.
3. *Social scoring* (penilaian sosial), dengan menilai individu berdasarkan perilaku sosial atau karakteristik pribadi yang menyebabkan perlakuan tidak adil atau tidak proporsional.
4. Prediksi pelanggaran kriminal, dengan memprediksi risiko seseorang melakukan kejahatan berdasarkan profil atau karakteristik pribadi, kecuali mendukung penilaian manusia dengan fakta objektif.

a. Unacceptable Risk

To protect fundamental rights and public safety, the EU prohibits high-risk AI systems that violate fundamental rights, such as:

1. Manipulative or subliminal techniques, which use methods outside of an individual's awareness to manipulate their decisions in a harmful way.
2. Exploitation of vulnerabilities, by taking advantage of individuals' vulnerabilities (age, disabilities, or socio-economic conditions) to manipulate their behavior, resulting in significant harm.
3. Social scoring, by assessing individuals based on social behavior or personal characteristics that lead to unfair or disproportionate treatment.
4. Criminal offense prediction, by predicting an individual's risk of committing a crime based on their profile or personal characteristics, unless it supports human judgment with objective facts.
5. Pengumpulan data wajah tanpa target, menyusun *database* pengenalan wajah dari gambar internet atau rekaman CCTV secara acak.
6. Inferensi emosi, yakni menyimpulkan emosi di tempat kerja atau lembaga pendidikan, kecuali untuk tujuan medis atau keselamatan.
7. Kategorisasi biometrik, dengan menggunakan data biometrik yang mengungkap ras, pandangan politik, agama, atau orientasi seksual, kecuali untuk penegakan hukum.
8. Identifikasi biometrik *real-time* di ruang publik dilarang untuk penegakan hukum, kecuali untuk kasus tertentu (mencari korban, mencegah ancaman serius, atau menangkap tersangka pelanggaran serius).

b. Risiko tinggi (*high risk*)

Sistem AI dikategorikan berisiko tinggi jika penggunaannya berpotensi menimbulkan bahaya terhadap kesehatan, keselamatan, atau hak asasi manusia. Sistem ini harus memenuhi persyaratan ketat, dan terhadap penyedia (*provider*) serta pengguna (*deployer*) diwajibkan untuk mematuhi aturan yang ketat.

5. Collection of facial data without targeting, compiling a facial recognition database from internet images or random CCTV footage.
6. Emotion inference, which involves inferring emotions in the workplace or educational institutions, except for medical or safety purposes.
7. Biometric categorization, by using biometric data that reveals race, political views, religion, or sexual orientation, except for law enforcement purposes.
8. Real-time biometric identification in public spaces is prohibited for law enforcement, except in specific cases (searching for victims, preventing serious threats, or apprehending suspects of serious offenses).

b. High risk

AI systems are categorized as high-risk if their use has the potential to cause harm to health, safety, or human rights. These systems must meet strict requirements, and both providers and users are required to comply with stringent regulations.

1. Klasifikasi sistem AI sebagai sistem berisiko tinggi:

a. Sistem AI sebagai komponen keselamatan:

Sistem AI yang menjadi bagian dari produk yang tunduk pada legislasi harmonisasi UE (misalnya, mainan, aplikasi medis seperti operasi robotik, kendaraan bermotor) dan memerlukan penilaian kesesuaian pihak ketiga. Yang dimaksud dengan penilaian kesesuaian pihak ketiga adalah proses evaluasi yang dilakukan oleh pihak independen (pihak ketiga) untuk memastikan bahwa suatu produk, sistem, atau layanan memenuhi standar, persyaratan hukum, atau regulasi tertentu.

b. Sistem AI untuk tujuan khusus. Hal ini mencakup:

i. Biometrik (identifikasi jarak jauh).

ii. Infrastruktur penting (lalu lintas jalan, suplai air, gas, listrik).

iii. Pendidikan (penilaian hasil belajar, akses ke institusi pendidikan).

iv. Ketenagakerjaan (alat perekrutan atau pengambilan keputusan terkait hubungan kerja).

v. Akses layanan penting (kesehatan, evaluasi kredit, asuransi).

1. Classification of AI systems as high-risk systems:

a. AI systems as safety components:

AI systems that are part of products subject to EU harmonization legislation (e.g., toys, medical applications such as robotic surgery, motor vehicles) and require third-party conformity assessment. The term "third-party conformity assessment" refers to the evaluation process conducted by an independent party (third party) to ensure that a product, system, or service meets specific standards, legal requirements, or regulations.

b. AI systems for specific purposes. This includes:

i. Biometric (long range identification).

ii. Key infrastructure (traffic management, water supply, gas, electricity).

iii. Education (learning outcome assessment, access to educational institutions).

iv. Employment (recruitment tools or decision-making related to employment relationships).

v. Access to essential services (healthcare, credit evaluation, insurance).

vi. Penegakan hukum (penilaian risiko kriminal atau keandalan bukti).

vii. Migrasi, suaka dan kontrol perbatasan.

viii. Administrasi peradilan (penerapan hukum).

ix. Proses demokrasi (mempengaruhi hasil pemilu).

x. Sistem AI melakukan pembuatan profil orang perseorangan.

c. Karena kategori risiko tinggi sangat luas, EU AI Act mengatur pengecualian untuk sistem AI yang tidak menimbulkan risiko bahaya yang signifikan terhadap kesehatan, keselamatan, atau hak-hak asasi manusia, termasuk tidak memengaruhi hasil pengambilan keputusan secara material, dan sistem AI memenuhi kondisi:

i. Melakukan tugas prosedural yang sempit.

ii. Meningkatkan hasil aktivitas manusia yang telah diselesaikan sebelumnya.

iii. Mendeteksi pola pengambilan keputusan atau penyimpangan dari pola pengambilan keputusan sebelumnya, sepanjang sistem AI tidak dimaksudkan untuk menggantikan atau memengaruhi penilaian manusia yang telah diselesaikan sebelumnya tanpa adanya review oleh manusia secara tepat.

vi. Law enforcement (criminal risk assessment or reliability of evidence).

vii. Migration, asylum, and border control.

viii. Administration of justice (application of the law).

ix. Democratic processes (influencing election outcomes).

x. AI systems that create profiles of individuals.

c. Due to the breadth of high-risk category, the EU AI Act regulates exceptions for AI systems that do not pose significant risks of harm to health, safety, or human rights, including not materially affecting decision-making outcomes, and the AI systems meet the conditions:

i. Performing narrow procedural tasks.

ii. Enhancing the outcomes of previously completed human activities.

iii. Decision-making or deviation from previous decision-making patterns, as long as the AI system is not intended to replace or influence previously completed human judgments without appropriate human review.

iv. Melakukan tugas persiapan untuk penilaian yang relevan terhadap tujuan penggunaan khusus sebagaimana di atas.

2. Persyaratan untuk sistem AI berisiko tinggi

Sistem AI berisiko tinggi harus memenuhi persyaratan persyaratan ketat, yang mencakup beberapa area:

a. Manajemen risiko.

b. Kualitas dan tata kelola data.

c. Dokumentasi teknis yang komprehensif.

d. Pencatatan yang konsisten (pembuatan, penyimpanan, dan pemeliharaan dokumentasi).

e. Transparansi dan penyediaan informasi kepada pengguna (*deployer*).

f. Pengawasan oleh manusia selama periode penggunaan sistem AI.

g. Memastikan keakuratan, ketahanan, dan keamanan siber.

Beberapa aspek yang perlu menjadi perhatian:

a. Pelatihan, validasi, dan pengujian dari *datasets* (kumpulan data) harus relevan, cukup representatif, lengkap, dan bebas dari kesalahan.

b. Sistem AI harus dirancang untuk memungkinkan pengawasan manusia saat digunakan.

iv. Performing preparatory tasks for relevant assessments related to the specific aforementioned purposes.

2. Requirements for high-risk AI systems

High-risk AI systems must meet strict requirements, which encompass several areas, including:

a. Risk Management.

b. Data quality and management.

c. Comprehensive technical documentation.

d. Consistent record-keeping (Creation, storage, and maintenance of documentation).

e. Transparency and provision of information to users (deployers).

f. Human oversight during the use of the AI system.

g. Ensuring accuracy, robustness, and cybersecurity.

Several aspects that require attention:

a. Training, validation, and testing of datasets must be relevant, sufficiently representative, complete, and free from errors.

b. The AI system must be designed to allow for human oversight during its use.

c. Sistem AI berisiko tinggi harus dirancang dan dikembangkan sedemikian rupa sehingga mencapai tingkat akurasi, ketahanan, dan keamanan siber yang sesuai, yang bekerja secara konsisten di sepanjang siklus hidupnya.

3. Kewajiban yang berlaku bagi penyedia (*provider*) sistem AI berisiko tinggi:

a. Identifikasi penyedia sistem AI berisiko tinggi

Penyedia atau pihak yang memenuhi kriteria penyedia adalah:

i. Pihak yang mengembangkan sistem AI atau model AI (GPAI model).

ii. Pihak yang memasarkan atau menyediakan sistem AI dengan nama atau merek dagang sendiri.

iii. Setiap distributor, importir, pengguna (*deployer*), dan pihak ketiga lainnya yang memenuhi persyaratan:

- Mencantumkan nama atau merek dagangnya pada sistem AI berisiko tinggi yang telah dipasarkan atau digunakan.
- Melakukan modifikasi substansial pada sistem AI berisiko tinggi.
- Memodifikasi GPAI model menjadi sistem AI berisiko tinggi.

c. High-risk AI systems must be designed and developed in such a way that they achieve appropriate levels of accuracy, robustness, and cybersecurity, consistently working throughout their lifecycle.

3. Requirements applicable to providers of high-risk AI systems:

a. Identification of providers of high-risk AI systems

Providers or those meeting the criteria of a provider are:

i. The party that develops the AI system or AI model (GPAI model).

ii. The party that markets or provides the AI system under its own name or trademark.

iii. Any distributor, importer, user (deployer), and other third parties that meet the requirements:

- Listing its name or trademark on the high-risk AI system that has been marketed or used.
- Making substantial modifications to the high-risk AI system.
- Modifying the GPAI model into a high-risk AI system.

Aspek yang perlu menjadi perhatian:

i. Jika terjadi perubahan yang dapat memengaruhi kepatuhan sistem AI berisiko tinggi terhadap ketentuan yang berlaku (misalnya, perubahan sistem operasi atau arsitektur perangkat lunak), atau ketika tujuan sistem berubah, sistem AI tersebut harus dianggap sebagai sistem AI baru yang harus menjalani penilaian kesesuaian baru.

ii. Perubahan terjadi pada algoritma dan kinerja sistem AI setelah dipasarkan atau digunakan, yang secara terus "belajar" dan secara otomatis mengadaptasi cara fungsi dijalankan, tidak dianggap sebagai modifikasi substansial, asalkan perubahan tersebut telah ditentukan sebelumnya oleh penyedia (*provider*) dan dinilai pada saat *conformity assessment* (proses sistematis untuk menentukan apakah suatu produk, layanan, proses, atau sistem memenuhi kesesuaian persyaratan, standar, atau regulasi tertentu).

b. Kewajiban penyedia (*provider*) sistem AI berisiko tinggi:

i. Conformity assessments

- Penilaian kesesuaian (*conformity assessments*) dilakukan oleh penyedia atau pihak ketiga

Aspects requiring attention:

i. If there are changes that may affect the compliance of the high-risk AI system with applicable regulations (e.g., changes in operating system or software architecture), or when the purpose of the system changes, the AI system must be considered a new AI system that must undergo a new conformity assessment.

ii. Changes occur in the algorithms and performance of the AI system after it has been marketed or used, which continuously "learns" and automatically adapts how functions are performed, are not considered substantial modifications, provided that these changes have been predetermined by the provider and assessed during the conformity assessment (a systematic process to determine whether a product, service, process, or system meets specific conformity requirements, standards, or regulations).

b. Requirements for providers of high-risk AI systems:

i. Conformity assessments

- Conformity assessments are conducted by the provider or an independent third party to

independen untuk memastikan sistem AI memenuhi persyaratan yang berlaku bagi sistem AI berisiko tinggi, sebelum sistem AI digunakan atau diluncurkan ke pasar UE.

- Setelah melakukan penilaian kesesuaian, penyedia (*provider*) harus memberi label sistem AI mereka dengan tanda "CE", yang menunjukkan kesesuaian dengan standar UE dan kepatuhan terhadap Undang-Undang AI.
- Selanjutnya, penyedia (*provider*) harus mendaftarkan sistem mereka pada basis data untuk sistem AI berisiko tinggi yang dapat diakses publik.

ii. Sistem manajemen mutu:

- Setelah sistem AI dipasarkan, penyedia (*provider*) harus membuat dan mendokumentasikan sistem manajemen mutu yang memastikan kepatuhan terhadap EU AI Act.
- Sistem manajemen mutu harus didokumentasikan secara sistematis dan teratur dalam bentuk kebijakan, prosedur, dan instruksi tertulis.
- penyedia juga menerapkan sistem untuk melaporkan insiden serius sebagai bagian dari tanggung jawab pemantauan paska pemasaran mereka.

ensure that the AI system meets the applicable requirements for high-risk AI systems, before the AI system is used or launched in the EU market.

- After conducting the conformity assessment, the provider must label their AI system with the "CE" mark, indicating compliance with EU standards and adherence to the AI Act.
- Subsequently, the provider must register their system in a publicly accessible database for high-risk AI systems.

ii. Quality Management System:

- After the AI system is marketed, the provider must create and document a quality management system that ensures compliance with the EU AI Act.
- The quality management system must be systematically and regularly documented in the form of policies, procedures, and written instructions.
- The provider also implements a system for reporting serious incidents as part of their post-marketing monitoring responsibilities.

- Dalam hal *deployer* mengidentifikasi adanya insiden serius terkait sistem AI, harus menginformasikan kepada penyedia.
- penyedia harus menyimpan catatan terperinci dan *log* yang dibuat secara otomatis oleh sistem AI berisiko tinggi mereka.
- Untuk membantu manajemen risiko, penyedia model AI harus merilis dokumentasi terperinci tentang karakteristik umum, kapabilitas dan keterbatasan, logika umum sistem AI dan algoritma, arsitektur sistem, metodologi dan teknik pelatihan, *dataset* pelatihan yang digunakan, dan prosedur validasi dan pengujian yang

- In the event that the deployer identifies a serious incident related to the AI system, they must inform the provider.
- The provider must maintain detailed records and logs automatically generated by their high-risk AI system.
- To assist in risk management, the AI model provider must release detailed documentation on general characteristics, capabilities and limitations, the general logic of the AI system and algorithms, system architecture, training methodologies and techniques, the training datasets used, and the validation and testing

digunakan, serta dokumentasi tentang sistem manajemen risiko yang relevan.

- Mengambil tindakan korektif/perbaikan yang diperlukan untuk menyesuaikan sistem AI berisiko tinggi yang dideteksi atau dicurigai terdapat ketidakpatuhan suatu sistem atau tidak sesuai dengan UU, dengan menonaktifkan atau menarik kembali jika sistem AI yang telah dipasarkan/digunakan.

c. Pengguna (*deployer*) (dalam hal ini, badan hukum publik, badan layanan publik, dan pengguna yang memenuhi kriteria) harus melakukan penilaian dampak terhadap hak asasi manusia yang dapat ditimbulkan oleh penggunaan sistem AI, sebelum menggunakan sistem tersebut.

procedures employed, as well as documentation on relevant risk management systems.

- Taking necessary corrective actions to adjust the high-risk AI system that is detected or suspected of non-compliance with the law, including deactivating or recalling the AI system that has been marketed/used.

c. Users (deployers) (in this case, public legal entities, public service bodies, and users that meet the criteria) must conduct an impact assessment on human rights that may arise from the use of the AI system, before using the system.

c. Risiko terbatas yang memerlukan transparansi (*limited risk/transparency risk*)

Untuk beberapa sistem AI, transparansi adalah persyaratan penting, terutama dalam hal terdapat risiko bahwa sistem tersebut dapat atau akan digunakan untuk menipu atau memanipulasi orang tanpa sepengetahuan atau persetujuan mereka. Sistem AI dimaksud dikategorikan sebagai

c. Limited risk requiring transparency

For some AI systems, transparency is an important requirement, especially in cases where there is a risk that the system can or may be used to deceive or manipulate people without their knowledge or consent. Such AI systems are categorized as limited risk (or specific transparency risk) and are subject to transparency

risiko terbatas (atau risiko transparansi khusus) dan dikenakan kewajiban transparansi, meskipun tidak menutup kemungkinan akan berdampak risiko tinggi sehingga tetap berlaku kewajiban bagi sistem AI berisiko tinggi.

4 (empat) kategori *tools* atau aplikasi dari sistem AI dimaksud meliputi:

a. Chatbot

Penyedia (*provider*) *chatbot* harus memastikan bahwa pengguna (*deployer*) menyadari bahwa mereka sedang berbicara dengan mesin. Penyedia bertanggung jawab untuk mengungkapkan bahwa pengguna berinteraksi dengan AI (bukan dengan manusia).

b. Aplikasi *Deepfake*

Penyedia (*provider*) harus mengungkapkan bahwa konten telah dibuat atau dimanipulasi secara artifisial, dengan pengecualian jika terkait aspek hukum atau jika konten tersebut merupakan bagian dari karya atau program yang bersifat artistik, kreatif, satir, atau fiktif.

c. Sistem *Generative AI*

Berfokus pada sistem AI yang menghasilkan konten audio, gambar, video, atau teks secara sintetis (secara buatan atau artifisial, bukan berasal dari sumber alami atau manusia secara langsung), termasuk sistem AI yang menghasilkan atau memanipulasi konten yang ada, yang berkaitan dengan *deepfake*.

Untuk sistem AI generatif, penyedia harus memastikan bahwa *output* sistem AI ditandai dalam format yang dapat dibaca mesin dan dapat dideteksi sebagai hasil buatan atau manipulasi, misalnya dengan menggunakan tanda air (*watermarking*), kecuali jika konten tersebut telah direviu oleh pihak yang memiliki tanggung jawab editorial atas konten tersebut.

d. Sistem Pengenalan Emosi atau Kategorisasi Biometrik

Pihak yang menerapkan sistem pengenalan emosi atau sistem kategorisasi biometrik harus memberi tahu pihak yang terdampak dari penggunaan sistem AI tersebut, misalnya *Facial Emotion Recognition* (FER), yaitu teknologi yang menganalisis ekspresi wajah dari gambar statis maupun video untuk mengungkap informasi tentang keadaan emosional seseorang.

obligations, although it does not rule out the possibility of having high-risk impacts, thus the requirements for high-risk AI systems still apply.

4 (four) categories of tools or applications of the aforementioned AI systems include:

a. Chatbot

The provider of the chatbot must ensure that users (deployers) are aware that they are interacting with a machine. The provider is responsible for disclosing that users are interacting with AI (not with a human).

b. Deepfake Applications

The provider must disclose that the content has been artificially created or manipulated, except in cases related to legal aspects or if the content is part of a work or program that is artistic, creative, satirical, or fictional.

c. Generative AI Systems

Focusing on AI systems that generate audio, images, video, or text synthetically (artificially or not derived from natural or direct human sources), including AI systems that generate or manipulate existing content, which relates to deepfakes.

For generative AI systems, the provider must ensure that the output of the AI system is marked in a machine-readable format and can be detected as a result of artificial creation or manipulation, for example, by using watermarking, unless the content has been reviewed by a party with editorial responsibility for that content.

d. Emotion Recognition Systems or Biometric Categorization

The parties implementing emotion recognition systems or biometric categorization systems must inform the individuals affected by the use of such AI systems, for example, Facial Emotion Recognition (FER), which is a technology that analyzes facial expressions from static images or videos to reveal information about a person's emotional state.

c. Risiko minimal/rendah (*minimal risk*)

Sistem AI lain yang tidak menjadi target EU AI Act dapat dikembangkan dan digunakan sesuai dengan kerangka hukum yang berlaku dan tidak menimbulkan kewajiban tambahan apa pun. Contoh sistem AI risiko minimal antara lain sistem AI untuk *spam filter*. Penyedia (*provider*) sistem ini secara sukarela dapat mematuhi standar AI yang dapat dipercaya (*trustworthy AI*) dan kode etik.

4. *General-Purpose AI* (GPAI) *Models*.

Dewan UE (The EU Council) mendefinisikan GPAI sebagai teknologi yang "dimaksudkan oleh penyedia (*provider*) untuk menjalankan fungsi yang berlaku secara umum" dan yang dapat digunakan dalam "banyak konteks dan diintegrasikan ke dalam banyak sistem AI lainnya". Ketika GPAI dapat digunakan sebagai sistem AI berisiko tinggi, GPAI harus mematuhi persyaratan peraturan sebagaimana persyaratan yang berlaku untuk sistem AI berisiko tinggi.

EU AI Act membedakan antara *general-purpose AI models* (GPAI *models*) dan *general-purpose AI systems* (GPAI *systems*):

a. GPAI *models* tidak membentuk sistem AI sendiri. Model tersebut merupakan komponen penting dari sistem AI dan memerlukan komponen lebih lanjut untuk menjadi sistem AI.

b. GPAI *systems* berarti sistem AI yang didasarkan pada GPAI *models* yang memiliki kemampuan untuk melayani berbagai tujuan, baik untuk penggunaan langsung maupun untuk integrasi dalam sistem AI lainnya.

Dengan demikian, meskipun model AI merupakan komponen penting dari sistem AI, model tersebut tidak membentuk sistem AI sendiri. Model AI memerlukan penambahan komponen lebih lanjut (seperti misalnya antarmuka pengguna) untuk menjadi sistem AI. Model AI biasanya terintegrasi ke dalam dan menjadi bagian dari sistem AI. Ketentuan dalam undang-undang ini berlaku khusus untuk GPAI *models* dan bukan untuk GPAI *systems*, yang menunjukkan bahwa dalam konteks khusus *general-purpose AI* (GPAI), ketentuan UU mengatur teknologi itu sendiri, bukan aplikasinya. EU AI Act mengharuskan penyedia (*provider*) model GPAI untuk memenuhi persyaratan khusus:

a. Menyusun dokumentasi teknis, mencakup informasi tentang cara model dirancang dan dikembangkan serta karakteristik

c. Low/minimal Risk

AI systems that are not targeted by the EU AI Act can be developed and used in accordance with applicable legal frameworks without imposing any additional obligations. Examples of minimal-risk AI systems include AI systems for spam filtering. Providers of these systems may voluntarily comply with trustworthy AI standards and ethical codes.

4. General-Purpose AI (GPAI) models.

The EU Council defines General Purpose AI (GPAI) as technology that is "intended by the provider to perform functions that are generally applicable" and that can be used in "many contexts and integrated into many other AI systems." When GPAI can be used as a high-risk AI system, it must comply with regulatory requirements as applicable to high-risk AI systems.

The EU AI Act distinguishes between general-purpose AI models (GPAI models) and general-purpose AI systems (GPAI systems):

a. GPAI models do not constitute AI systems on their own. These models are essential components of AI systems and require additional components to become an AI system.

b. GPAI systems refer to AI systems that are based on GPAI models and have the capability to serve various purposes, both for direct use and for integration into other AI systems.

Thus, although AI models are essential components of AI systems, these models do not constitute AI systems on their own. AI models require additional components (such as user interfaces) to become AI systems. AI models are typically integrated into and become part of AI systems. The provisions in this law specifically apply to GPAI models and not to GPAI systems, indicating that in the specific context of general-purpose AI (GPAI), the provisions of the law regulate the technology itself, not its applications. The EU AI Act requires providers of GPAI models to meet these specific requirements:

a. Preparing technical documentation, including information on how the model is designed and developed, as

utama model, termasuk proses pelatihan, pengujian dan hasil evaluasi.

b. Menyusun informasi dan dokumentasi untuk diberikan kepada penyedia (*provider*) hilir yang bermaksud mengintegrasikan GPAI *models* ke dalam sistem AI mereka sendiri, agar mereka dapat memahami kemampuan dan keterbatasan model dan menggunakan model atau bahkan menyempurnakan model dengan tepat.

c. Mematuhi aturan terkait hak cipta;

d. Menyediakan ringkasan yang cukup terperinci tentang konten yang digunakan untuk melatih GPAI model, termasuk bagi GPAI model yang bersifat *open-source*.

e. Penyedia GPAI models yang bersifat *open-source* dan gratis, dikecualikan dari kewajiban untuk merilis dokumentasi teknis dan memberikan informasi terperinci kepada penyedia (*provider*) hilir, namun tetap mematuhi semua persyaratan lainnya seperti kebijakan hak cipta dan ringkasan data pelatihan.

well as the key characteristics of the model, including the training process, testing, and evaluation results.

b. Preparing information and documentation to be provided to downstream providers intending to integrate GPAI models into their own AI systems, so that they can understand the capabilities and limitations of the models and use or even refine the models appropriately.

c. Complying with copyright regulations.

d. Providing a sufficiently detailed summary of the content used to train the GPAI model, including for open-source GPAI models.

e. Providers of open-source and free GPAI models are exempt from the obligation to release technical documentation and provide detailed information to downstream providers, but must still comply with all other requirements such as copyright policies and training data summaries.

5. GPAI *models* dengan risiko sistemik
   a. GPAI models dikatakan memiliki risiko sistemik jika:
      1. Memiliki kapabilitas berdampak tinggi, yakni jumlah kumulatif komputasi yang digunakan untuk pelatihannya yang diukur berdasarkan *floating point operations* (FLOP) lebih besar dari 10^{25}.
      2. Model tersebut ditetapkan memiliki kapabilitas berdampak tinggi berdasarkan keputusan Komisi berdasarkan kriteria:
   a. Kompleksitas model (jumlah parameter, kualitas dataset, dan komputasi yang digunakan).
   b. Modalitas *input-output* (teks ke teks, teks ke gambar, multimodalitas, dan sebagainya).

5. GPAI models with systemic risk
   a. GPAI models are considered to have systemic risk if:
      1. They have high-impact capabilities, namely the cumulative amount of computation used for their training, measured in floating point operations (FLOP), exceeding 10^{25}.
      2. The model is determined to have high-impact capabilities based on a decision by the Commission according to the criteria:
   a. Model complexity (number of parameters, quality of the dataset, and computation used).
   b. Input-output modalities (text to text, text to image, multimodality, etc.).

c. Kapabilitas model (jumlah tugas tanpa pelatihan tambahan, adaptabilitas, otonomi, dan akses ke alat).

d. Dampak model pada pasar internal Uni Eropa (tersedia bagi setidaknya 10.000 pengguna bisnis di UE).

e. Jumlah pengguna akhir yang terdaftar.

b. Kewajiban penyedia (*provider*) GPAI *models* dengan risiko sistemik:

1. Melakukan evaluasi model dan mengidentifikasi serta mengurangi risiko sistemik.
2. Menilai dan mengurangi kemungkinan risiko sistemik, termasuk sumbernya.

c. Model capabilities (number of tasks without additional training, adaptability, autonomy, and access to tools).

d. The impact of the model on the internal market of the European Union (available to at least 10,000 business users in the EU).

e. The number of registered end users.

b. Requirements for providers of GPAI models with systemic risk:

1. Conducting model evaluations and identifying and mitigating systemic risks.
2. Assessing and mitigating the likelihood of systemic risks, including their sources.

3. Melacak, mendokumentasikan, dan melaporkan insiden serius dan kemungkinan tindakan perbaikan kepada otoritas.
4. Memastikan tingkat perlindungan keamanan siber yang memadai.
5. Menyertakan informasi tambahan dalam dokumentasi teknis mereka:

a. Strategi evaluasi, termasuk hasil, kriteria, metrik, dan identifikasi keterbatasan, secara detail.

b. Langkah-langkah untuk pengujian *adversarial* internal/eksternal (misalnya, *red-teaming*) dan adaptasi model (penyelarasan dan *fine-tuning*).

c. Arsitektur sistem secara detail (menjelaskan bagaimana komponen perangkat lunak saling terhubung dan berintegrasi dalam keseluruhan proses).

6. Ketentuan lainnya dari undang-undang ini yang perlu menjadi perhatian:

a. Sandboxes

1. Otoritas yang berwenang di negara Uni Eropa membentuk setidaknya satu AI *regulatory sandboxes* (mekanisme uji coba terbatas) di tingkat nasional, yang menyediakan lingkungan terkendali yang mendorong inovasi dan memfasilitasi pengembangan, pelatihan, pengujian, dan validasi sistem AI yang inovatif untuk waktu

3. Tracking, documenting, and reporting serious incidents and potential remedial actions to the authorities.
4. Ensuring an adequate level of cybersecurity protection.
5. Including additional information in their technical documentation:

a. Evaluation strategy, including results, criteria, metrics, and identification of limitations, in detail.

b. Steps for internal/external adversarial testing (e.g., red-teaming) and model adaptation (alignment and fine-tuning).

c. System architecture in detail (explaining how software components are interconnected and integrated within the overall process).

6. Other provisions of this law that require attention:

a. Sandboxes

1. The competent authorities in EU countries establish at least one AI regulatory sandbox (limited trial mechanism) at the national level, providing a controlled environment that encourages innovation and facilitates the development, training, testing, and validation of innovative AI systems for a limited time before

terbatas sebelum dipasarkan/digunakan, yang disetujui antara penyedia (*providers*) dan otoritas berwenang.

2. Otoritas yang berwenang harus memberikan, panduan, pengawasan, dan dukungan dalam *AI regulatory sandboxes* yang bertujuan untuk mengidentifikasi risiko, langkah mitigasi, dan efektivitas terkait dengan kewajiban dan pemenuhan persyaratan.
3. Pembentukan *AI regulatory sandboxes* bertujuan untuk:

a. Meningkatkan kepastian hukum untuk mencapai kepatuhan regulasi dengan ketentuan yang berlaku.

b. Mendukung berbagi praktik terbaik melalui kerja sama dengan otoritas yang terlibat dalam *AI regulatory sandboxes*.

c. Mendorong inovasi dan daya saing serta memfasilitasi pengembangan ekosistem AI.

d. Berkontribusi pada *evidence-based regulatory learning* (menghasilkan ouput pembelajaran).

e. Memfasilitasi dan mempercepat akses sistem AI ke pasar Uni Eropa, khususnya ketika disediakan oleh UMKM/perusahaan *start-up*.

they are marketed/used, as approved between providers and the competent authorities.

2. The competent authorities must provide guidance, oversight, and support in AI regulatory sandboxes aimed at identifying risks, mitigation steps, and effectiveness related to obligations and compliance with requirements.
3. The establishment of AI regulatory sandboxes aims to:

a. Enhance legal certainty to achieve regulatory compliance with applicable provisions.

b. Support the sharing of best practices through collaboration with authorities involved in AI regulatory sandboxes.

c. Encourage innovation and competitiveness, as well as facilitate the development of the AI ecosystem.

d. Contribute to evidence-based regulatory learning (producing learning outputs).

e. Facilitate and accelerate access for AI systems to the European market, particularly when provided by SMEs/start-up companies.

4. Prinsip umum *AI regulatory sandboxes*:

a. Kriteria kelayakan untuk berpartisipasi dalam *AI regulatory sandboxes*.

b. Prosedur untuk penerapan, partisipasi, pemantauan, pengakhiran *AI regulatory sandboxes*.

c. Syarat dan ketentuan yang berlaku bagi peserta.

b. Real-world testing

Pengujian sistem AI berisiko tinggi dalam "kondisi dunia nyata" di luar *AI regulatory sandboxes* dapat dilakukan oleh penyedia (*providers*)/calon penyedia jika memenuhi kondisi berikut:

1. Telah menyusun rencana "pengujian dunia nyata" dan menyerahkannya kepada otoritas untuk mendapatkan persetujuan.
2. Durasi terbatas, yaitu periode maksimal 6 (enam) bulan dengan perpajangan 6 (enam) bulan.
3. Adanya perlindungan dalam pengujian, yakni subjek pengujian harus memberikan persetujuan, pengujian tidak boleh berdampak buruk, prediksi, rekomendasi, atau keputusan sistem AI harus dapat dibatalkan atau diabaikan, dan data pengguna harus dihapus setelah pengujian.

4. General principles of AI regulatory sandboxes:

a. Eligibility criteria for participation in AI regulatory sandboxes.

b. Procedures for application, participation, monitoring, and termination of AI regulatory sandboxe.

c. Terms and conditions applicable to participants.

b. Real-world testing

Testing of high-risk AI systems in "real-world conditions" outside of AI regulatory sandboxes can be conducted by providers/potential providers if the following conditions are met:

1. A "real-world testing" plan has been developed and submitted to the authorities for approval.
2. A limited duration, specifically a maximum period of 6 (six) months with a possible extension of 6 (six) months.
3. Safeguards in testing, meaning that test subjects must provide consent, the testing must not have adverse effects, predictions, recommendations, or decisions made by the AI system must be reversible or disregarded, and user data must be deleted after testing.

4. Perlindungan khusus untuk kelompok rentan, seperti individu dengan kerentanan usia atau disabilitas.
5. Pengujian dapat dikenakan inspeksi tak terjadwal oleh otoritas untuk memastikan kepatuhan dan keselamatan.

c. Hak individu untuk mendapatkan penjelasan terkait keputusan dari sistem AI (*right to explanation of individual decision-making*).

Setiap individu atau badan hukum berhak mendapatkan penjelasan ketika keputusan dibuat berdasarkan *output* dari sistem AI berisiko tinggi yang berdampak hukum atau secara signifikan memengaruhi mereka (merugikan kesehatan, keselamatan, atau hak *fundamental*). Penjelasan harus jelas dan bermakna dari penyedia (*providers*) mengenai peran sistem AI dalam proses pengambilan keputusan serta elemen utama dari keputusan yang diambil.

d. Ketentuan eksplisit untuk lembaga keuangan

1. Penyedia (*providers*) yang merupakan lembaga keuangan yang tunduk pada persyaratan mengenai tata kelola, pengaturan, atau proses internal berdasarkan UU Jasa Keuangan yang relevan:

a. Kewajiban untuk menerapkan sistem manajemen mutu (terkait pengembangan AI) dianggap terpenuhi dengan mematuhi aturan tentang pengaturan atau proses tata kelola internal.
b. Harus menyimpan dokumentasi teknis sebagai bagian dari dokumentasi yang disimpan.
c. Harus menyimpan catatan (*log*) yang secara otomatis dibuat oleh sistem AI berisiko tinggi mereka sebagai bagian dari dokumentasi yang disimpan.

2. Pengguna (*deployers*) yang merupakan lembaga keuangan yang tunduk pada persyaratan mengenai tata kelola, pengaturan, atau proses internal berdasarkan UU Jasa Keuangan yang relevan:

a. Kewajiban pemantauan dianggap telah dipenuhi dengan mematuhi peraturan mengenai pengaturan, proses, dan mekanisme tata kelola internal.
b. Harus menyimpan catatan (*log*) sebagai bagian dari dokumentasi.
c. Untuk sistem AI berisiko tinggi yang dipasarkan atau dioperasikan oleh lembaga

4. Special protections for vulnerable groups, such as individuals with age-related vulnerabilities or disabilities.
5. Testing may be subject to unscheduled inspections by authorities to ensure compliance and safety.

c. The right of individuals to receive an explanation regarding decisions made by the AI system (right to explanation of individual decision-making).

Every individual or legal entity has the right to receive an explanation when decisions are made based on the output of high-risk AI systems that have legal effects or significantly affect them (harmful to health, safety, or fundamental rights). The explanation must be clear, meaningful, and provided by the providers regarding the role of the AI system in the decision-making process, as well as the key elements of the decision made.

d. Explicit provisions for financial institutions

1. Providers that are financial institutions subject to requirements regarding governance, regulation, or internal processes based on relevant Financial Services Laws:

a. The obligation to implement a quality management system (related to AI development) is considered fulfilled by complying with regulations regarding governance or internal processes
b. Must maintain technical documentation as part of the retained documentation.
c. Must retain logs that are automatically generated by their high-risk AI systems as part of the documentation that is kept.

2. Users (deployers) who are financial institutions are subject to requirements regarding governance, regulation, or internal processes based on relevant Financial Services Laws:

a. The monitoring obligation is considered fulfilled by complying with regulations regarding governance, processes, and internal governance mechanisms
b. Must retain logs as part of the documentation.
c. For high-risk AI systems marketed or operated by financial institutions, subject to

keuangan, tunduk pada persyaratan berdasarkan UU Jasa Keuangan mengenai tata kelola, pengaturan, atau proses internal mereka.

C. European Commission's Ethics Guidelines for Trustworthy AI

Komisi Eropa pada April 2019 menerbitkan panduan yang bertujuan untuk mendorong AI yang dapat dipercaya (*trustworthy* AI). Panduan ini terdiri dari 3 (tiga) bab yang memuat:

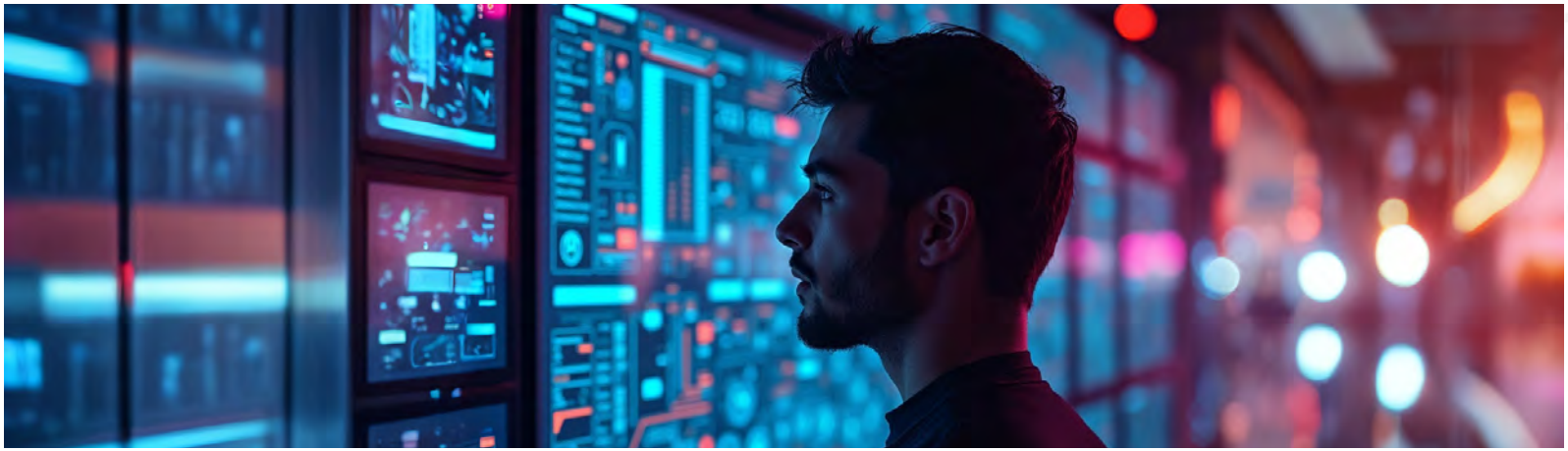
1. Pondasi dari AI yang dapat dipercaya, yang terdiri dari 4 (empat) prinsip etika yang berakar pada hak-hak fundamental dan harus dihormati untuk memastikan bahwa sistem AI dikembangkan, diterapkan, dan digunakan dengan cara yang dapat dipercaya, yaitu:
   a. Penghormatan terhadap otonomi manusia (*respect for human autonomy*):
      1. Sistem AI harus memungkinkan manusia mempertahankan kendali atas keputusan mereka sendiri.
      2. Sistem AI tidak boleh memanipulasi, menipu, atau mengendalikan manusia secara tidak sah.
      3. Sistem AI harus dirancang untuk melengkapi dan memberdayakan kemampuan kognitif, sosial, dan budaya manusia.

requirements based on Financial Services Laws regarding their governance, regulation, or internal processes.

C. European Commission's Ethics Guidelines for Trustworthy AI

The European Commission published guidelines in April 2019 aimed at promoting trustworthy AI. These guidelines contain 3 (three) chapters that include:

1. The foundation of trustworthy AI consists of 4 (four) ethical principles rooted in fundamental rights that must be respected to ensure that AI systems are developed, deployed, and used in a trustworthy manner, namely:
   a. Respect for human autonomy:
      1. AI systems must allow humans to maintain control over their own decisions.
      2. AI systems must not manipulate, deceive, or unduly control humans
      3. AI systems should be designed to complement and enhance human cognitive, social, and cultural abilities.



4. Memastikan pengawasan manusia dalam proses kerja AI, terutama di lingkungan kerja.

b. Pencegahan terhadap bahaya (*prevention of harm*):

1. Sistem AI tidak boleh menimbulkan atau berdampak buruk terhadap manusia.
2. Sistem AI harus melindungi martabat manusia, integritas mental dan fisik.
3. Sistem AI harus aman, tangguh secara teknis, dan tidak rentan terhadap penyalahgunaan.
4. Kelompok rentan harus diperhatikan dalam pengembangan dan penggunaan AI.

4. Ensure human oversight in the AI workflow, especially in the workplace.

b. Prevention of harm:

1. AI systems must not cause or have adverse effects on humans
2. AI systems must protect human dignity, mental and physical integrity.
3. AI systems must be safe, technically robust, and not vulnerable to misuse.
4. Vulnerable groups must be considered in the development and use of AI.

5. Perlu mempertimbangkan dampak AI terhadap lingkungan dan makhluk hidup lainnya.

c. Keadilan (*fairness*):

1. Sistem AI harus digunakan secara adil dan tidak mendiskriminasi individu atau kelompok tertentu.
2. Harus menjamin distribusi manfaat dan biaya yang setara serta bebas dari bias yang tidak adil.
3. Sistem AI tidak boleh menipu atau merugikan dengan menghambat kebebasan pilihan seseorang.
4. Harus ada keseimbangan antara kepentingan dan tujuan yang saling bersaing.
5. Harus ada mekanisme bagi individu untuk menantang dan mendapatkan ganti rugi terhadap

5. The impact of AI on the environment and other living beings must be taken into consideration.

c. Fairness:

1. AI systems must be used fairly and not discriminate against individuals or specific groups.
2. There must be a guarantee of equal distribution of benefits and costs, free from unfair bias.
3. AI systems must not deceive or harm by hindering an individual's freedom of choice.
4. There must be a balance between competing interests and objectives.
5. There must be mechanisms for individuals to challenge and seek redress for decisions made by

keputusan yang dibuat oleh sistem AI dan oleh manusia yang mengoperasikannya. Oleh karena itu, pihak yang bertanggung jawab atas keputusan tersebut harus dapat diidentifikasi dan proses pengambilan keputusan harus dapat dijelaskan.

d. Dapat dijelaskan (*explicability*):

1. Sistem AI harus transparan, dapat dipahami, dan dapat dijelaskan kepada pengguna.
2. *Output* dari sistem AI harus dapat dipertanggungjawabkan, terutama jika memiliki dampak signifikan terhadap individu.
3. Dalam kasus algoritma *black box*, harus ada langkah-langkah seperti auditabilitas, ketertelusuran, dan komunikasi yang transparan mengenai kemampuan sistem.
4. Tingkat keterjelasan yang dibutuhkan tergantung pada konteks dan dampak keputusan yang dihasilkan oleh sistem AI.

2. Mewujudkan AI yang dapat dipercaya.

Memberikan panduan untuk menerapkan dan mewujudkan AI yang dapat dipercaya (*trustworthy AI*) melalui 7 (tujuh) persyaratan utama yang harus dipenuhi yang didasarkan atas 4 (empat) prinsip etika, termasuk metode teknis dan non-teknis yang dapat diimplementasikan di sepanjang siklus hidup sistem AI.

AI systems and by the humans operating them. Therefore, the parties responsible for those decisions must be identifiable, and the decision-making process must be explainable.

d. Explicability:

1. AI systems must be transparent, understandable, and explainable to users.
2. The output of AI systems must be accountable, especially if it has significant impacts on individuals.
3. In the case of black box algorithms, there must be measures such as auditability, traceability, and transparent communication regarding the system's capabilities.
4. The level of clarity required depends on the context and the impact of the decisions made by the AI system.

2. Realizing Trustworthy AI.

Providing guidance for implementing and realizing trustworthy AI through 7 (seven) key requirements that must be met, based on 4 (four) ethical principles, including technical and non-technical methods that can be implemented throughout the AI system lifecycle.

4 (empat) prinsip etika di atas harus diterjemahkan ke dalam persyaratan konkret untuk mencapai AI yang dapat dipercaya. Persyaratan ini berlaku bagi berbagai pemangku kepentingan dalam siklus hidup AI, termasuk:

a. Pengembang (*developers*), mereka yang meneliti, merancang, dan mengembangkan sistem AI.
b. Pengguna AI (*deployers*), organisasi publik atau swasta yang menggunakan AI dalam proses bisnisnya untuk menyediakan produk dan layanan kepada pihak lain.
c. Pengguna akhir (*end-users*), mereka yang terlibat/berinteraksi dengan sistem AI, baik langsung atau tidak langsung.
d. Masyarakat luas (*broader society*), Semua pihak yang terdampak oleh AI, baik secara langsung maupun tidak langsung.

Beberapa aspek yang perlu diperhatikan untuk mewujudkan AI yang dapat dipercaya antara lain:

a. 7 (tujuh) persyaratan utama AI yang Tepercaya (*non-exhaustive*):

1. Kendali dan pengawasan oleh manusia (*human agency and oversight*).

Memastikan hak asasi manusia, kendali manusia, dan pengawasan manusia dalam sistem AI.

The 4 (four) ethical principles mentioned above must be translated into concrete requirements to achieve trustworthy AI. These requirements apply to various stakeholders throughout the AI lifecycle, including:

a. Developers, those who research, design, and develop AI systems.
b. AI users (deployers), public or private organizations that use AI in their business processes to provide products and services to others.
c. End-users: those who are involved/interact with the AI system, either directly or indirectly.
d. Broader society: All parties affected by AI, either directly or indirectly.

Several aspects that need to be considered to realize trustworthy AI include:

a. 7 (seven) key requirements for Trustworthy AI (non-exhaustive):

1. Human agency and oversight.

Ensuring human rights, human control, and human oversight in AI systems.

2. Ketahanan teknis dan keamanan (*technical robustness and safety*).

   Termasuk ketahanan terhadap serangan, keamanan sistem, rencana/langkah cadangan (*fallback plan*) dan keselamatan umum, akurasi, keandalan dan *reproducibility* (model AI harus mampu memberikan hasil yang konsisten dan dapat diuji ulang).

3. Privasi dan tata kelola data (*privacy and data governance*).

   Menghormati privasi, menjaga kualitas dan integritas data, serta memastikan akses data yang aman.

4. Transparansi (*transparency*).

   Mencakup keterlacakan, keterjelasan (*explainability*), dan komunikasi yang transparan.

5. Keberagaman, non-diskriminasi, dan keadilan (*diversity, non-discrimination and fairness*).

   Menghindari bias yang tidak adil, memastikan aksesibilitas dan desain universal, serta melibatkan pemangku kepentingan.

6. Kesejahteraan sosial dan lingkungan (*societal and environmental wellbeing*).

   Menjamin keberlanjutan lingkungan, mempertimbangkan dampak sosial, serta mendukung demokrasi dan kesejahteraan masyarakat.

2. Technical robustness and safety.

   Including resilience against attacks, system security, fallback plans, and overall safety, as well as accuracy, reliability, and reproducibility (AI models must be able to provide consistent and testable results).

3. Privacy and data governance.

   Respecting privacy, maintaining data quality and integrity, and ensuring secure data access.

4. Transparency.

   Including traceability, explainability, and transparent communication.

5. Diversity, non-discrimination, and fairness.

   Avoiding unfair bias, ensuring accessibility and universal design, and involving stakeholders.

6. Societal and environmental wellbeing.

   Ensuring environmental sustainability, considering social impact, and supporting democracy and societal well-being.

7. Akuntabilitas (*accountability*).

   Meliputi auditabilitas, minimalisasi dan pelaporan dampak negatif, keseimbangan kepentingan (*trade-offs*), serta mekanisme pengaduan dan ganti rugi.

Setiap pemangku kepentingan memiliki peran dalam memastikan pemenuhan persyaratan dimaksud:

1. Pengembang (*developers*) harus menerapkan persyaratan dalam desain dan pengembangan AI.
2. Pengguna AI (*deployers*) harus memastikan sistem, produk, dan layanan yang mereka gunakan memenuhi persyaratan.
3. Pengguna akhir dan masyarakat (*end-users and the broader society*) harus mendapatkan informasi tentang persyaratan ini dan dapat meminta agar persyaratan tersebut dipatuhi.

b. Metode teknis dan non-teknis untuk mewujudkan AI yang dapat dipercaya

Untuk menerapkan persyaratan AI yang dapat dipercaya, dapat digunakan metode teknis dan non-teknis yang mencakup seluruh tahap siklus hidup sistem AI dan dilaksanakan secara berkesinambungan sehubungan sistem AI terus berkembang dan beroperasi dalam lingkungan yang dinamis.

7. Accountability.

   Including auditability, minimization and reporting of negative impacts, balancing interests (trade-offs), as well as complaint and redress mechanisms.

Each stakeholder has a role in ensuring the fulfillment of the aforementioned requirements:

1. Developers must implement the requirements in the design and development of AI.
2. AI users (deployers) must ensure that the systems, products, and services they use meet the requirements.
3. End-users and the broader society must be informed about these requirements and have the ability to request compliance with them.

b. Technical and non-technical methods to realize trustworthy AI

To implement the requirements for trustworthy AI, both technical and non-technical methods can be used, covering the entire lifecycle of AI systems and carried out continuously as AI systems continue to evolve and operate in dynamic environments.

1. Metode Teknis, yang dapat diterapkan dalam tahap desain, pengembangan, dan penggunaan sistem AI. Metode yang tercantum memiliki tingkat kematangan yang berbeda-beda, yang meliputi:
   a. Arsitektur untuk AI yang Tepercaya (*Architectures for Trustworthy AI*).
   b. Etika dan kepatuhan hukum dalam desain (*Ethics and rule of law by design* atau *X-by-design*).
   c. Metode penjelasan (*Explanation methods*).
   d. Pengujian dan validasi (*Testing and validating*).
   e. Indikator Kualitas Layanan (*Quality of service indicators*).
2. Metode Non-Teknis, merupakan berbagai metode non-teknis yang dapat memainkan peran penting dalam mengamankan dan memelihara AI yang Dapat Dipercaya serta dilakukan evaluasi secara berkala, yang meliputi:
   a. Regulasi.
   b. Kode etik dan pedoman perilaku (*codes of conduct*).
   c. Standarisasi.
   d. Sertifikasi.
   e. Akuntabilitas melalui kerangka tata kelola (*governance frameworks*).
   f. Pendidikan dan kesadaran untuk menumbuhkan pola pikir etis.
   g. Partisipasi pemangku kepentingan dan dialog sosial.
   h. Tim desain yang beragam dan inklusif.
3. Penilaian

   Menetapkan daftar penilaian terkait operasionalisasi 7 (tujuh) persyaratan utama dimaksud untuk mengoperasionalkan dan memastikan *Trustworthy AI*, melalui penggunaan daftar penilaian *Trustworthy AI* yang dilakukan saat mengembangkan, menerapkan, atau menggunakan sistem AI, dan disesuaikan dengan kasus penggunaan spesifik di mana sistem tersebut diterapkan.

1. Technical methods, which can be applied in the design, development, and use stages of AI systems. The methods listed have varying levels of maturity, which include:
   a. Architectures for Trustworthy AI.
   b. Ethics and rule of law by design or X-by-design.
   c. Explanation methods.
   d. Testing and validating.
   e. Quality of service indicators.
2. Non-technical methods, which are various non-technical approaches that can play a crucial role in securing and maintaining Trustworthy AI and are subject to regular evaluation, which include:
   a. Regulation.
   b. Codes of conduct.
   c. Standardization.
   d. Certification.
   e. Accountability through governance frameworks.
   f. Education and awareness to foster an ethical mindset.
   g. Stakeholder participation and social dialogue.
   h. Diverse and inclusive design team.
3. Assesment

   Establishing an assesment list related to the operationalization of the seven key requirements to operationalize and ensure Trustworthy AI, through the use of a Trustworthy AI assesment list that is applied during the development, deployment, or use of AI systems, and tailored to the specific use cases in which the system is applied.

Daftar penilaian seperti ini tidak akan pernah sepenuhnya mencakup semua aspek. Oleh karena itu, perlu dilakukan secara berkelanjutan untuk mengidentifikasi kebutuhan, mengevaluasi solusi, dan memastikan hasil yang lebih baik sepanjang siklus hidup sistem AI, serta melibatkan pemangku kepentingan dalam proses tersebut.

Such an assesment list will never fully encompass all aspects. Therefore, it needs to be conducted continuously to identify needs, evaluate solutions, and ensure better outcomes throughout the AI system lifecycle, as well as involve stakeholders in the process.

D. United Nations Educational, Scientific and Cultural Organization (UNESCO) Recommendation on the Ethics of AI

Peningkatan pesat dalam AI telah menciptakan banyak peluang secara global serta tantangan yang menyertai yang terkait dengan etika dan risiko. Menyadari betapa pentingnya tantangan ini, UNESCO menerbitkan standar global pertama mengenai etika AI (*Recommendation on the Ethics of Artificial Intelligence*) dimana kerangka kerja (standar) ini diadopsi oleh seluruh 193 negara anggota pada November 2021. Inti dari rekomendasi tersebut adalah:

1. 4 (empat) nilai inti (*core values)* yang menjadi dasar bagi sistem AI yang berfungsi dan berkontribusi pada kebaikan manusia, individu, masyarakat, dan lingkungan:
   a. Menghormati, melindungi, dan memajukan hak asasi manusia, kebebasan fundamental, serta martabat manusia.
   b. Hidup dalam masyarakat yang damai, adil, dan saling terhubung.
   c. Menjamin keberagaman dan inklusivitas.
   d. Kelestarian lingkungan dan ekosistem.
2. 10 (sepuluh) prinsip inti (*core principles*) dalam penggunaan AI didasarkan pendekatan etika AI yang berpusat pada hak asasi manusia:
   a. Proporsionalitas dan Tidak Menyebabkan Kerugian

      Penggunaan AI harus sesuai dengan tujuan yang sah dan tidak melebihi batas yang diperlukan. Penilaian risiko harus dilakukan untuk mencegah potensi bahaya.

D. United Nations Educational, Scientific and Cultural Organization (UNESCO) Recommendation on the Ethics of AI

The rapid advancement in AI has created numerous opportunities globally, as well as accompanying challenges related to ethics and risks. Recognizing the importance of these challenges, UNESCO published the first global standards on AI ethics (Recommendation on the Ethics of Artificial Intelligence), which was adopted by all 193 member countries in November 2021. The core of the recommendation is:

1. 4 (four) core values that serve as the foundation for AI systems that function and contribute to the well-being of humanity, individuals, society, and the environment:
   a. Respecting, protecting, and promoting human rights, fundamental freedoms, and human dignity.
   b. Living in a peaceful, just, and interconnected society.
   c. Ensuring diversity and inclusivity.
   d. Environmental sustainability and ecosystem preservation.
2. 10 (ten) core principles in the use of AI are based on a human rights-centered ethical approach to AI:
   a. Proportionality and Do No Harm

      The use of AI must align with legitimate purposes and not exceed necessary limits. Risk assessments should be conducted to prevent potential harm.

b. Keselamatan dan Keamanan

Kerugian yang tidak diinginkan (risiko keselamatan) serta kerentanan terhadap serangan (risiko keamanan) harus dihindari dan ditangani oleh pelaku AI.

c. Hak atas Privasi dan Perlindungan Data

Privasi harus dilindungi dan ditingkatkan sepanjang siklus hidup AI, dengan kerangka perlindungan data yang memadai.

d. Tata Kelola Adaptif dan Kolaboratif dari Banyak Pihak

Penggunaan data harus menghormati hukum internasional dan kedaulatan negara, serta melibatkan berbagai pemangku kepentingan untuk memastikan tata kelola AI yang inklusif.

e. Tanggung Jawab dan Akuntabilitas

Sistem AI harus dapat diaudit dan ditelusuri, dengan mekanisme pengawasan, penilaian dampak, dan uji kelayakan untuk mencegah konflik dengan hak asasi manusia dan ancaman terhadap lingkungan.

f. Transparansi dan Keterjelasan

Sistem AI harus transparan dan dapat dijelaskan. Masyarakat harus diberi tahu jika sebuah keputusan dibuat berdasarkan AI. Namun, transparansi harus disesuaikan dengan konteks, mengingat potensi konflik dengan prinsip lain seperti privasi, keselamatan dan keamanan.

b. Safety and Security

Unintended harm (safety risks) and vulnerabilities to attacks (security risks) must be avoided and addressed by AI actors.

c. Right to Privacy and Data Protection

Privacy must be protected and enhanced throughout the AI lifecycle, with adequate data protection frameworks.

d. Adaptive and Collaborative Multi-Stakeholder Governance

The use of data must respect international law and national sovereignty, as well as involve various stakeholders to ensure inclusive AI governance.

e. Responsibility and Accountability

AI systems must be auditable and traceable, with mechanisms for oversight, impact assessment, and feasibility testing to prevent conflicts with human rights and threats to the environment.

f. Transparency and Clarity

AI systems must be transparent and explainable. The public should be informed if a decision is made based on AI. However, transparency should be contextualized, considering potential conflicts with other principles such as privacy, safety, and security.

g. Pengawasan dan Kontrol Manusia

AI tidak boleh menggantikan tanggung jawab dan akuntabilitas manusia sebagai pihak yang utama.

h. Keberlanjutan

AI harus dievaluasi berdasarkan dampaknya terhadap keberlanjutan, termasuk pencapaian Tujuan Pembangunan Berkelanjutan (*Sustainable Development Goals*) yang ditetapkan oleh Perserikatan Bangsa Bangsa.

i. Kesadaran dan Literasi

Pemahaman publik tentang AI dan data harus ditingkatkan melalui pendidikan yang terbuka dan mudah diakses, keterlibatan masyarakat, keterampilan digital dan pelatihan etika AI, serta literasi media dan informasi.

j. Keadilan dan Non-Diskriminasi

Para pelaku AI harus mempromosikan keadilan sosial, kesetaraan, dan non-diskriminasi, dengan pendekatan inklusif agar manfaatnya dapat dirasakan oleh semua pihak.

3. Ditetapkan 11 (sebelas) area kebijakan dalam mengoperasionalkan nilai-nilai dan prinsip-prinsip dimaksud, dan selanjutnya agar negara anggota menerapkan langkah-langkah yang efektif, termasuk, misalnya, kerangka kerja atau mekanisme kebijakan, serta memastikan kepatuhan

g. Human Supervision and Control

AI must not replace the responsibility and accountability of humans as the primary party.

h. Sustainability

AI must be evaluated based on its impact on sustainability, including the achievement of the Sustainable Development Goals (SDGs) set by the United Nations.

i. Awareness and Literacy

Public understanding of AI and data must be enhanced through open and accessible education, community engagement, digital skills, AI ethics training, and media and information literacy.

j. Fairness and Non-Discrimination

AI actors must promote social justice, equality, and non-discrimination, with an inclusive approach to ensure that the benefits are experienced by all parties.

3. 11 (eleven) policy areas have been established to operationalize the aforementioned values and principles, and member states are encouraged to implement effective measures, including frameworks or policy mechanisms, while ensuring compliance from other stakeholders

dari pemangku kepentingan lain (perusahaan sektor swasta, lembaga akademis, dan masyarakat), yaitu: 1) Penilaian dampak etika, 2) Tata kelola etika dan pengawasan, 3) Kebijakan data, 4) Pengembangan dan kerja sama internasional, 5) Lingkungan dan ekosistem, 6) Gender, 7) Budaya, 8) Pendidikan dan penelitian, 9) Komunikasi dan informasi, 10) Ekonomi dan tenaga kerja, dan 11) Kesehatan dan kesejahteraan sosial.

E. G7 Hiroshima Principles (Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System)

Para pemimpin G7 (Perancis, Amerika Serikat, Inggris, Jerman, Jepang, Italia, dan Kanada) menerbitkan prinsip tata kelola AI agar:

1. Mendukung AI yang aman, terjamin, dan tepercaya dan memberikan panduan bagi organisasi yang mengembangkan dan menggunakan sistem AI termasuk sistem AI dengan model yang paling terdepan (*advanced AI systems*) untuk memaksimalkan manfaat teknologi sekaligus memitigasi risikonya.
2. Dalam pemanfaatan peluang inovasi, organisasi tidak boleh mengembangkan atau menerapkan sistem AI dengan cara yang melemahkan nilai-nilai demokrasi, merugikan individu atau masyarakat, berdampak hukum dan HAM (memfasilitasi terorisme, kriminal, atau menimbulkan risiko besar terhadap keselamatan, keamanan), dan menjadikan umat manusia sebagai prioritas utama.

Semua pihak yang terlibat dalam siklus hidup AI didorong untuk mengikuti 12 (dua belas) prinsip tata kelola AI bagi organisasi yang mengembangkan sistem AI (*advanced AI systems*) yang berpedoman pada *Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems* (30 Oktober 2023), dengan mempertimbangkan kapasitas mereka dan peran mereka dalam siklus hidup AI:

1. Mengambil langkah-langkah yang sesuai selama pengembangan sistem AI, termasuk sebelum dan selama penerapan serta peredarannya di pasar, untuk mengidentifikasi, mengevaluasi, dan mengurangi risiko di seluruh siklus hidup AI.
2. Mengidentifikasi dan mengurangi kerentanan serta, jika diperlukan, insiden dan pola penyalahgunaan setelah sistem diterapkan, termasuk saat sudah beredar di pasar.

(private sector companies, academic institutions, and society), namely: 1) Ethical Impact Assessment, 2) Ethical Governance and Oversight, 3) Data Policy, 4) International Development and Cooperation, 5) Environment and Ecosystem, 6) Gender, 7) Culture, 8) Education and Research, 9) Communication and Information, 10) Economy and Labor, and 11) Health and Social Well-being.

E. G7 Hiroshima Principles (Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System)

The G7 leaders (France, the United States, the United Kingdom, Germany, Japan, Italy, and Canada) published AI governance principles to ensure that:

1. To support safe, secure, and trustworthy AI and provide guidance for organizations developing and using AI systems, including advanced AI systems, to maximize the benefits of technology while mitigating its risks.
2. In leveraging innovation opportunities, organizations must not develop or implement AI systems in ways that undermine democratic values, harm individuals or society, have legal and human rights impacts (facilitating terrorism, crime, or posing significant risks to safety and security), and must prioritize humanity above all.

All parties involved in the AI lifecycle are encouraged to adhere to the 12 (twelve) principles of AI governance for organizations developing AI systems (advanced AI systems) based on the Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems (October 30, 2023), taking into account their capacities and roles in the AI lifecycle:

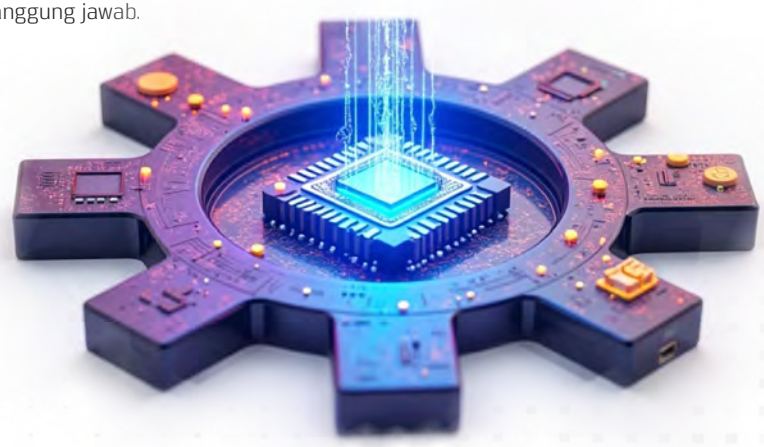
1. Take appropriate steps during the development of AI systems, including before and during their implementation and market distribution, to identify, evaluate, and mitigate risks throughout the AI lifecycle.
2. Identify and mitigate vulnerabilities and, if necessary, incidents and patterns of misuse after the system is implemented, including when it is already in the market.

3. Melaporkan secara publik tentang kemampuan, keterbatasan, serta bidang penggunaan yang tepat dan tidak tepat dari sistem AI untuk memastikan transparansi yang memadai, sehingga meningkatkan akuntabilitas.
4. Mendorong pembagian informasi yang bertanggung jawab dan pelaporan insiden di antara organisasi yang mengembangkan sistem AI tingkat lanjut, termasuk dengan industri, pemerintah, masyarakat sipil, dan akademisi.
5. Mengembangkan, menerapkan, dan mengungkapkan kebijakan tata kelola AI serta manajemen risiko berbasis pendekatan risiko, termasuk kebijakan privasi dan langkah-langkah mitigasi.
6. Berinvestasi dalam dan menerapkan langkah-langkah keamanan yang kuat, termasuk keamanan fisik, keamanan siber, dan perlindungan dari ancaman internal di seluruh siklus hidup AI.
7. Mengembangkan dan menerapkan mekanisme autentikasi serta pelacakan asal konten yang andal, jika memungkinkan secara teknis, seperti tanda air (*watermark*) atau teknik lain yang memungkinkan pengguna mengidentifikasi konten yang dihasilkan oleh AI.

3. Publicly report on the capabilities, limitations, and appropriate and inappropriate areas of use of AI systems to ensure adequate transparency, thereby enhancing accountability.
4. Encourage responsible information sharing and incident reporting among organizations developing advanced AI systems, including with industry, government, civil society, and academia.
5. Develop, implement, and disclose AI governance policies and risk management based on a risk-based approach, including privacy policies and mitigation measures.
6. Invest in and implement robust security measures, including physical security, cybersecurity, and protection against internal threats throughout the AI lifecycle.
7. Develop and implement reliable authentication mechanisms and content provenance tracking, where technically feasible, such as watermarks or other techniques that allow users to identify AI-generated content.

8. Memprioritaskan penelitian untuk memitigasi risiko sosial, keamanan, dan keselamatan serta berinvestasi dalam langkah-langkah mitigasi yang efektif.
9. Memprioritaskan pengembangan sistem AI untuk mengatasi tantangan global terbesar, terutama tetapi tidak terbatas pada krisis iklim, kesehatan global, dan pendidikan.
10. Mendorong pengembangan serta penerapan standar teknis internasional.
11. Menerapkan langkah-langkah input data yang sesuai serta perlindungan terhadap data pribadi dan kekayaan intelektual.
12. Mendorong dan berkontribusi pada penggunaan sistem AI tingkat lanjut yang dapat dipercaya dan bertanggung jawab.

8. Prioritize research to mitigate social, security, and safety risks, and invest in effective mitigation measures.
9. Prioritize the development of AI systems to address the world's greatest global challenges, particularly but not limited to climate crisis, global health, and education.
10. Encourage the development and implementation of international technical standards.
11. Implement appropriate data input measures and protections for personal data and intellectual property.
12. Encourage and contribute to the use of trustworthy and responsible advanced AI systems.



F. Amerika Serikat: *The White House Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*

Pada Oktober 2022, *White House Office of Science and Technology Policy* bekerjasama dengan akademisi, organisasi hak asasi manusia, perusahaan-perusahaan besar, dan masyarakat umum merilis *Blueprint for an AI Bill of Rights* (cetak biru untuk melindungai hak asasi di era AI). *Blueprint* ini bersifat tidak mengikat (*non-binding*), dan bukan merupakan kebijakan pemerintah AS. Perusahaan dapat memilih ikut serta atau tidak ikut serta.

Prinsip-prinsip dalam *Blueprint for an AI Bill of Rights* bertujuan untuk memastikan perlindungan terhadap publik dari bahaya dan mencegah penggunaan teknologi, data, dan sistem otomatis yang mengancam hak-hak manusia. Terdapat 5 (lima) prinsip dan praktik terkait yang mengarahkan sektor swasta dalam merancang, menggunakan, dan menerapkan AI untuk melindungi hak-hak masyarakat Amerika di era AI. Kelima prinsip tersebut meliputi:

F. The United States: The White House Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People

In October 2022, the White House Office of Science and Technology Policy, in collaboration with academics, human rights organizations, large companies, and the general public, released the Blueprint for an AI Bill of Rights. This blueprint is non-binding and does not constitute U.S. government policy. Companies can choose whether or not to participate.

The principles in the Blueprint for an AI Bill of Rights aim to ensure protection for the public from harm and to prevent the use of technology, data, and automated systems that threaten human rights. There are 5 (five) principles and related practices that guide the private sector in designing, using, and implementing AI to protect the rights of the American public in the era of AI. The five principles include:

1. *Safe and effective systems* (sistem yang aman dan efektif)

   Automated systems (AI) harus dikembangkan melalui konsultasi dengan berbagai pihak dan pemangku kepentingan untuk mengidentifikasi masalah, risiko, dan dampak potensial dari sistem. Masyarakat harus dilindungi dari sistem AI yang tidak aman atau tidak efektif. Sistem harus menjalani pengujian sebelum diterapkan, identifikasi dan mitigasi risiko, serta pemantauan berkelanjutan untuk meningkatkan keamanan dan efektivitasnya. Selain itu, hasil evaluasi independen dan laporan yang memastikan bahwa sistem aman dan efektif harus diumumkan ke publik jika memungkinkan. Aspek yang diharapkan:

   a. Melindungi masyarakat dari bahaya dengan cara yang proaktif dan berkelanjutan, melalui konsultasi, pengujian, identifikasi dan mitigasi risiko, pengawasan berdasarkan organisasi (prosedur dan tata kelola) yang jelas, pemantauan berkelanjutan.
   b. Menghindari penggunaan data yang tidak tepat, berkualitas rendah, atau tidak relevan serta bahaya dari penggunaan ulang data, melalui data yang relevan dan berkualitas tinggi, melacak dan mereviu sumber data turunan dengan cermat, batasan penggunaan kembali data di domain sensitif.
   c. Keamanan dan efektivitas sistem, melalui evaluasi independen dan pelaporan secara berkala.

1. Safe and effective systems

   Automated systems (AI) must be developed through consultation with various parties and stakeholders to identify issues, risks, and potential impacts of the systems. The public must be protected from unsafe or ineffective AI systems. Systems should undergo testing before deployment, risk identification and mitigation, as well as continuous monitoring to enhance their safety and effectiveness. Additionally, the results of independent evaluations and reports ensuring that the systems are safe and effective should be made public whenever possible. The expected aspects include:

   a. Protecting the public from harm in a proactive and sustainable manner, through consultation, testing, risk identification and mitigation, oversight based on clear organizational procedures and governance, and continuous monitoring.
   b. Avoiding the use of inappropriate, low-quality, or irrelevant data and the dangers of data reuse, through the use of relevant and high-quality data, carefully tracking and reviewing the sources of derived data, and imposing restrictions on data reuse in sensitive domains.
   c. Ensuring the security and effectiveness of systems through independent evaluations and regular reporting.

2. *Algorithmic discrimination protections* (perlindungan terhadap diskriminasi algoritmik)

   Designers, developers, dan *deployers* AI harus mengambil tindakan proaktif dan berkelanjutan untuk melindungi individu dan masyarakat dari diskriminasi oleh algoritma serta menggunakan dan merancang sistem secara adil. Masyarakat tidak boleh mengalami diskriminasi oleh algoritma dan berdampak tidak adil (negatif) pada individu berdasarkan karakteristik seperti ras, orientasi seksual, status disabilitas, dan karakteristik lain yang dilindungi oleh hukum. Aspek yang diharapkan:

   a. Melindungi masyarakat dari diskriminasi algoritmik secara proaktif dan berkelanjutan, melalui penilaian proaktif atas kesetaraan dalam desain; penggunaan data yang representatif dan *robust*; berhati-hati dalam penggunaan *proxy* berupa informasi demografis yang dapat menyebabkan diskriminasi; memastikan aksesibilitas selama desain,

2. Algorithmic discrimination protections

   Designers, developers, and deployers of AI must take proactive and ongoing actions to protect individuals and society from discrimination by algorithms and to use and design systems fairly. Members of society should not experience algorithmic discrimination and should not be negatively impacted unfairly based on characteristics such as race, sexual orientation, disability status, and other legally protected characteristics. The expected aspects include:

   a. Protecting the people from algorithmic discrimination in a proactive and sustainable manner, through proactive assessments of equity in design; the use of representative and robust data; caution in the use of proxies in the form of demographic information that may lead to discrimination; ensuring accessibility during design, development, and deployment;

pengembangan dan penerapan; penilaian kesenjangan; mitigasi kesenjangan; serta pemantauan dan mitigasi berkelanjutan.

b. Menunjukkan bahwa sistem melindungi dari diskriminasi algoritmik melalui evaluasi independen dan pelaporan terkait penilaian dampak algoritmik.

3. *Data privacy* (privasi data)

Masyarakat harus memiliki kendali atas bagaimana data tentang mereka digunakan. Perlindungan terhadap pengguna dan privasi mereka dengan meminta izin dan menghormati keputusan individu/masyarakat terkait pengumpulan, penggunaan, akses, transfer, dan penghapusan data mereka dengan cara yang tepat, dalam upaya melindungi privasi dan kebebasan sipil warga negara. Aspek yang diharapkan:

a. Melindungi privasi berdasarkan desain dan default, melalui perancangan dan pembangunan AI dengan perlindungan privasi secara *default*, pengumpulan data dan batasan cakupan *use-case*, identifikasi dan mitigasi risiko, keamanan yang dirancang untuk menjaga privasi.

b. Lindungi publik dari surveilans (pemantauan) yang tidak terkendali melalui peningkatan pengawasan terhadap surveilans, surveilans yang terbatas dan proporsional, penetapan batasan ruang lingkup pengawasan untuk melindungi hak asasi dan nilai-nilai demokrasi.

c. Menyediakan publik mekanisme terkait perizinan, akses, dan kontrol yang tepat dan berarti atas data mereka melalui penggunaan perizinan yang tidak menimbulkan praktik pengawasan yang tidak adil, permintaan persetujuan yang singkat dan langsung, tersedianya akses dan koreksi data, adanya opsi untuk penarikan dan penghapusan data, serta dukungan sistem otomatis.

d. Tunjukkan bahwa privasi data dan kontrol pengguna dilindungi, melalui evaluasi independen dan pelaporan.

4. *Notice and explanation* (pemberitahuan dan penjelasan)

Masyarakat harus mengetahui bahwa sistem otomatis sedang digunakan dan memahami bagaimana serta mengapa sistem tersebut berkontribusi pada hasil yang memengaruhi mereka. Para *designers*, *developers*, dan *deployers* sistem otomatis harus menggunakan bahasa yang sederhana dan mudah diakses untuk menjelaskan:

a. keseluruhan fungsi sistem dan peran yang dimainkan oleh otomatisasi

gap assessments; gap mitigation; as well as ongoing monitoring and mitigation.

b. Demonstrating that systems protect against algorithmic discrimination through independent evaluations and reporting related to algorithmic impact assessments.

3. Data privacy

The public must have control over how data about them is used. Protection for users and their privacy should involve obtaining consent and respecting individual/community decisions regarding the collection, use, access, transfer, and deletion of their data in appropriate ways, in an effort to protect the privacy and civil liberties of citizens. The expected aspects include:

a. Protecting privacy by design and by default, through the design and development of AI with privacy protections as the default setting, data collection and use-case scope limitations, risk identification and mitigation, and security measures designed to safeguard privacy.

b. Protecting the public from unchecked surveillance through enhanced oversight of surveillance practices, limited and proportional surveillance, and the establishment of boundaries on the scope of surveillance to protect human rights and democratic values.

c. Providing the public with appropriate and meaningful mechanisms for permission, access, and control over their data through the use of permissions that do not lead to unfair surveillance practices, concise and straightforward consent requests, availability of access and data correction, options for withdrawal and deletion of data, as well as support from automated systems.

d. Demonstrating that data privacy and user control are protected through independent evaluations and reporting.

4. Notice and explanation

The public must be aware that automated systems are being used and understand how and why these systems contribute to outcomes that affect them. Designers, developers, and deployers of automated systems should use simple and accessible language to explain:

a. The overall functions of the system and the role played by automation.

b. Pemberitahuan bahwa sistem sedang digunakan.

c. Bagaimana serta mengapa sistem tersebut memengaruhi hasil yang berdampak pada individu.

d. Individu atau organisasi yang bertanggung jawab atas sistem.

Aspek yang diharapkan:

a. Memberikan pemberitahuan penggunaan dan penjelasan secara jelas, tepat waktu, mudah dipahami, dan dapat diakses, melalui dokumentasi dalam bahasa yang mudah dipahami terkait keseluruhan sistem, akuntabel, tepat waktu dan terkini, singkat dan jelas.

b. Memberikan penjelasan tentang bagaimana dan mengapa suatu keputusan dibuat atau suatu tindakan diambil oleh AI, yang disesuaikan dengan tujuan, disesuaikan dengan target penjelasan, disesuaikan dengan tingkat risiko dan valid (akurat).

c. Menunjukkan perlindungan untuk pemberitahuan dan penjelasan, melalui pelaporan/dokumentasi yang tepat.

5. *Human alternatives, consideration and fallback* (alternatif manusia, pertimbangan, dan rencana/langkah cadangan)

Memilih AI daripada alternatif manusia didasarkan atas harapan yang wajar dalam konteks tertentu dan dengan fokus untuk memastikan kemampuan akses yang luas dan perlindungan terhadap potensi bahaya. Pengguna harus dapat memilih untuk keluar dari sistem AI kapan saja dan memiliki akses ke seseorang yang dapat dengan cepat menyelesaikan masalah yang dihadapi (dalam hal ini dengan menyediakan pertimbangan manusia dan solusi melalui proses eskalasi, serta langkah cadangan yang dapat diakses dan tepat waktu, terutama ketika sistem otomatis gagal, menghasilkan kesalahan, atau ketika seseorang ingin mengajukan banding atas dampaknya). Aspek yang diharapkan:

b. Notification that the system is in use.

c. How and why the system affects outcomes that impact individuals.

d. The individuals or organizations responsible for the system.

The expected aspects:

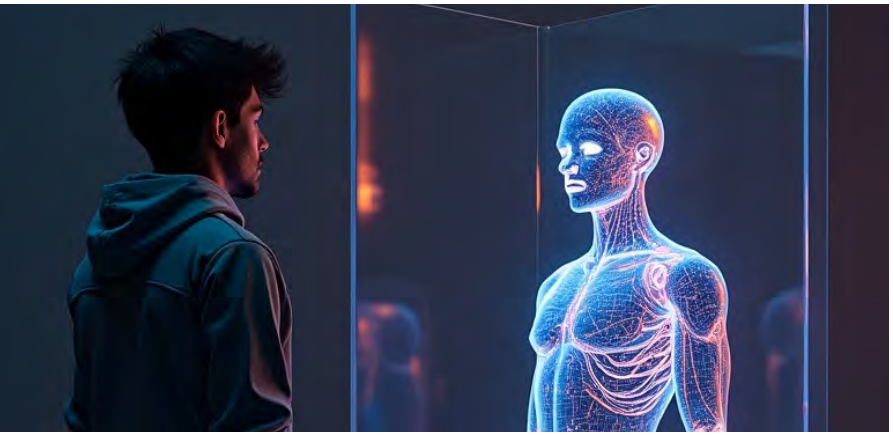
a. Providing notifications of use and explanations that are clear, timely, easily understood, and accessible, through documentation in plain language related to the overall system, ensuring accountability, timeliness, and currency, while being concise and clear.

b. Providing explanations of how and why a decision is made or an action is taken by AI, tailored to the objectives, aligned with the target audience for the explanation, adjusted according to the level of risk, and ensuring that the information is valid (accurate).

c. Demonstrating protections for notifications and explanations through appropriate reporting/documentation.

5. Human alternatives, considerations and fallback

Choosing AI over human alternatives is based on reasonable expectations in specific contexts, with a focus on ensuring broad access and protection against potential harm. Users should be able to opt out of AI systems at any time and have access to someone who can quickly resolve issues they encounter (in this case providing human consideration and solutions through an escalation process, as well as accessible and timely backup measures, especially when automated systems fail, produce errors, or when someone wishes to appeal the impacts). The expected aspects include:



a. Menyediakan mekanisme untuk memilih keluar dari AI untuk digantikan dengan alternatif manusia, yakni alternatif manusia tersedia bila diperlukan, alternatif manusia yang tepat waktu dan tidak memberatkan, pemberitahuan dan instruksi yang singkat, jelas, dan mudah diakses.

b. Memberikan pertimbangan manusia yang tepat waktu dan perbaikan melalui sistem *fallback* (rencana/langkah cadangan yang dapat digunakan dalam keadaan darurat) dan eskalasi jika AI gagal, secara proporsional, aksesibel, nyaman, adil, tepat waktu, efektif, terpelihara.

c. Menetapkan pelatihan, penilaian, dan pengawasan untuk melawan bias otomasi dan memastikan semua komponen sistem berbasis manusia berjalan efektif, melalui pelatihan dan penilaian yang tepat dan pengawasan berdasarkan struktur tata kelola yang tepat.

d. Menerapkan tambahan pengawasan dan perlindungan manusia terhadap AI yang terkait dengan domain sensitif, melalui data dan kesimpulan dengan cakupan yang sempit, disesuaikan dengan situasi, pertimbangan manusia sebelum mengambil keputusan berisiko tinggi, akses yang berarti untuk memeriksa sistem.

e. Pelaporan tentang aksesibilitas, ketepatan waktu, dan efektivitas pertimbangan manusia dan *fallback* yang dipublikasikan secara berkala selama sistem tersebut digunakan.

a. Providing mechanisms for opting out of AI to be replaced with human alternatives, ensuring that human alternatives are available when needed, timely, and not burdensome, along with notifications and instructions that are concise, clear, and easily accessible.

b. Providing timely human consideration and remediation through fallback systems (contingency plans that can be used in emergencies) and escalation processes if AI fails, ensuring that these measures are proportional, accessible, convenient, fair, timely, effective, and well-maintained.

c. Establishing training, assessment, and oversight to combat automation bias and ensure that all human-based system components operate effectively, through appropriate training and assessment, as well as oversight based on a proper governance structure.

d. Implementing additional oversight and human protections for AI related to sensitive domains, through data and conclusions with narrow scope, tailored to the situation, human consideration before making high-risk decisions, and meaningful access to review the systems.

e. Reporting on the accessibility, timeliness, and effectiveness of human consideration and fallback measures, published regularly while the system is in use.

G. India

Sebagai salah satu ekonomi digital terkemuka di Asia dan dengan pertumbuhan yang pesat dalam adopsi layanan digital dalam beberapa tahun terakhir, India ke depan juga memprioritaskan pengembangan dan adopsi AI dalam inisiatif kebijakannya. Pada Maret 2024, India mengumumkan alokasi lebih dari USD1,25 miliar untuk misi AI India (*India AI Mission*), yang mencakup berbagai aspek AI, termasuk kapasitas infrastruktur komputasi, pelatihan, inovasi, dataset, serta AI yang aman dan terpercaya (*safe and trusted AI*).

G. India

As one of the leading digital economies in Asia and with rapid growth in the adoption of digital services in recent years, India is also prioritizing the development and adoption of AI in its policy initiatives. In March 2024, India announced an allocation of over USD 1.25 billion for the India AI Mission, which includes various aspects of AI, including computing infrastructure capacity, training, innovation, datasets, as well as safe and trusted AI.

Sampai saat ini, India belum memiliki kerangka regulasi yang mengikat untuk AI (*horizontal AI law* atau *hard law*), namun demikian regulasi AI di India akan terus berkembang ke depannya dan pemerintah terus memprioritaskan kebijakan terkait AI, antara lain:

1. National Strategy for Artificial Intelligence #AIforAll

   Pada Juni 2018, National Institution for Transforming India (NITI Aayog) meluncurkan Strategi Nasional untuk AI pertama dengan *tagline* #AIforAll, yang berfungsi sebagai pendekatan inklusif terhadap AI dan berupaya untuk mengatasi beberapa tantangan global dari perspektif AI, baik itu aplikasi, penelitian, pengembangan, teknologi, dan AI yang bertanggung jawab. Strategi tersebut mengidentifikasi area penting untuk prioritas nasional dalam inovasi dan penerapan AI, yang berfokus pada 5 (lima) sektor yang diharapkan memperoleh manfaat paling besar dari AI dalam memecahkan kebutuhan masyarakat, yakni layanan kesehatan, pertanian, pendidikan, *smart city* dan dan infrastruktur, serta *smart mobility* transportasi, dengan semangat #AIforAll yang berarti kepemimpinan teknologi dalam AI untuk mencapai kebaikan yang lebih besar. Untuk benar-benar mendapatkan manfaat dari penerapan AI dalam skala besar, hambatan yang perlu diatasi untuk mencapai tujuan #AIforAll yaitu:

   a. Kurangnya keahlian yang luas dalam penelitian dan penerapan AI.
   b. Tidak adanya ekosistem data yang mendukung (akses ke *intelligent data*).
   c. Biaya sumber daya yang tinggi dan kesadaran yang rendah untuk adopsi AI.
   d. Privasi dan keamanan, termasuk kurangnya peraturan formal seputar anonimisasi data.
   e. Tidak adanya pendekatan kolaboratif untuk adopsi dan penerapan AI.

2. Principles for Responsible AI

   NITI Aayog pada Februari 2021 menerbitkan dokumen yang memuat prinsip-prinsip AI bertanggung jawab sebagai kelanjutan dari Strategi Nasional untuk AI. Dokumen tersebut menguraikan 7 (tujuh) prinsip utama untuk tata kelola sistem AI yang bertanggung jawab di India yakni:

   a. Prinsip Keamanan dan Keandalan

      AI harus diterapkan dengan andal sesuai tujuan dan dilengkapi perlindungan yang memadai untuk memastikan keselamatan para pemangku kepentingan. Risiko bagi semua pemangku kepentingan harus diminimalkan dan struktur penanganan keluhan, perawatan, dan kompensasi yang tepat harus tersedia, jika terjadi kerugian

Until now, India does not have a binding regulatory framework for AI (horizontal AI law or hard law); however, AI regulation in India will continue to evolve in the future, and the government continues to prioritize AI-related policies, including:

1. National Strategy for Artificial Intelligence #AIforAll

   In June 2018, the National Institution for Transforming India (NITI Aayog) launched the first National Strategy for AI with the tagline #AIforAll, which serves as an inclusive approach to AI and aims to address several global challenges from an AI perspective, including applications, research, development, technology, and responsible AI. The strategy identifies key areas for national priority in AI innovation and implementation, focusing on 5 (five) sectors expected to benefit the most from AI in addressing societal needs, namely healthcare, agriculture, education, smart cities and infrastructure, as well as smart mobility transportation, with the spirit of #AIforAll meaning technological leadership in AI to achieve greater good. To truly reap the benefits of large-scale AI implementation, the barriers that need to be addressed to achieve the goals of #AIforAll are:

   a. A lack of extensive expertise in AI research and implementation.
   b. The absence of a supportive data ecosystem (access to intelligent data).
   c. High resource costs and low awareness for AI adoption.
   d. Privacy and security, including the lack of formal regulations surrounding data anonymization.
   e. The absence of a collaborative approach to AI adoption and implementation.

2. Principles for Responsible AI

   NITI Aayog in February 2021 published a document containing principles of responsible AI as a continuation of the National Strategy for AI. The document outlines 7 (seven) key principles for the governance of responsible AI systems in India, namely:

   a. The Principle of Safety and Reliability

      AI should be deployed reliably as intended and sufficient safeguards must be placed to ensure the safety of relevant stakeholders. Risks to all stakeholders should be minimized and appropriate grievance redressal, care and compensation structures should be in place, in case of any unintended or unexpected harm. The AI system needs to be monitored

yang tidak diinginkan atau tidak terduga. Sistem AI perlu dipantau sepanjang siklus hidupnya sehingga berfungsi dengan cara yang dapat diterima, andal, sesuai dengan tujuan yang diinginkan.

b. Prinsip Kesetaraan

Sistem AI harus memperlakukan individu dalam keadaan yang sama dan setara terkait pengambilan keputusan.

c. Prinsip Inklusivitas dan Non-diskriminasi

Sistem AI tidak boleh menolak menghalangi kesempatan individu yang memenuhi syarat berdasarkan identitas mereka. Sistem AI juga harus memastikan bahwa tidak ada pengecualian dan ketidakadilan terhadap layanan atau manfaat. Jika terjadi keputusan yang merugikan, mekanisme pengaduan harus dirancang agar terjangkau dan dapat diakses oleh semua orang, terlepas dari latar belakang mereka.

d. Prinsip Privasi dan Keamanan

AI harus menjaga privasi dan keamanan data individu atau entitas yang digunakan untuk melatih sistem. Akses hanya boleh diberikan kepada pihak yang berwenang dengan pengamanan yang memadai.

e. Prinsip Transparansi

Desain dan cara kerja sistem AI harus dicatat dan tersedia untuk pengawasan dan audit eksternal guna memastikan penerapan yang adil, jujur, tidak memihak, dan menjamin akuntabilitas.

f. Prinsip Akuntabilitas

Semua pemangku kepentingan yang terlibat dalam desain, pengembangan, dan penerapan sistem AI harus bertanggung jawab atas tindakan mereka. Pemangku kepentingan harus melakukan penilaian risiko dan dampak serta evaluasi langsung dan tidak langsung dari sistem AI terhadap pengguna akhir, menyiapkan proses audit (baik internal maupun eksternal jika diperlukan) untuk memastikan kepatuhan terhadap prinsip, serta menciptakan mekanisme pengaduan jika terjadi dampak negatif.

g. Prinsip Perlindungan dan Penguatan Nilai-Nilai Kemanusiaan yang Positif.

AI harus mempromosikan nilai-nilai kemanusiaan yang positif dan tidak mengganggu keharmonisan sosial dalam bermasyarakat.

3. Operationalizing Principles for Responsible AI

NITI Aayog pada Agustus 2021 menerbitkan dokumen yang merupakan bagian kedua dari prinsip-prinsip AI yang bertanggung jawab. Dokumen tersebut mengidentifikasi serangkaian tindakan yang harus diadopsi ekosistem untuk mendorong AI yang bertanggung jawab. Tindakan ini dibagi diantara

through its lifecycle so it performs in an acceptable manner, reliably, according to the desired goals.

b. Principle of Equality

AI systems must treat individuals under same circumstances relevant to the decision equally.

c. Principle of Inclusivity and Non-discrimination

AI systems should not deny opportunity to a qualified person on the basis of their identity. The AI system should also strive to ensure that unfair exclusion of services or benefits does not happen. In case of an adverse decision, appropriate grievance redressal mechanism should be designed in a manner affordable and accessible to everyone irrespective of their background.

d. Principle of Privacy and Security

AI should maintain privacy and security of data of individuals or entities that is used for training the system. Access should be provided only to those authorized with sufficient safeguards.

e. Principle of Transparency

The design and functioning of the AI system should be recorded and made available for external scrutiny and audit to the extent possible to ensure the deployment is fair, honest, impartial and guarantees accountability.

f. Principle of Accountability

All stakeholders involved in the design, development and deployment of the AI system must be responsible for their actions. Stakeholders should conduct risk and impact assessments to evaluate direct and indirect potential impact of AI systems on endusers, set up an auditing process (internal and if required external) to oversee adherence to principles and create mechanisms for grievance redressal in case of any adverse impact.

g. Principle of Protection and Reinforcement of Positive Human Values.

AI should promote positive human values and not disturb in any way social harmony in community relationships.

3. Operationalizing Principles for Responsible AI

NITI Aayog in August 2021 published a document that is the second part of the principles of responsible AI. The document identifies a series of actions that the ecosystem must adopt to promote responsible AI. These actions are divided among three stakeholders: the

tiga pemangku kepentingan, yakni pemerintah, sektor swasta, dan lembaga penelitian. Di antara para pemangku kepentingan ini, tindakan tersebut lebih lanjut dikategorikan ke dalam beberapa area, dengan setiap area mengidentifikasi serangkaian langkah terkait untuk menerapkan prinsip-prinsip AI, sebagai berikut:

a. Pemerintah, merancang intervensi regulasi dan kebijakan yang ideal, meningkatkan kesadaran, aksesibilitas, dan pembangunan kapasitas, serta memfasilitasi strategi pengadaan yang tepat.

b. Sektor swasta dan lembaga penelitian, mendorong *ethics by design* (pendekatan di mana prinsip AI yang bertanggung jawab diintegrasikan ke dalam seluruh proses pengembangan teknologi AI)., menciptakan kerangka kerja untuk kepatuhan terhadap standar dan pedoman AI yang relevan, serta mendorong praktik AI yang bertanggung jawab dalam penelitian.

4. Regulasi AI yang diterbitkan oleh regulator sektoral India antara lain:

a. Di sektor keuangan, Badan Pengawas Pasar Modal India mengeluarkan surat edaran pada Januari 2019 kepada entitas di sektor pasar modal tentang persyaratan terkait aplikasi dan sistem AI dan *machine larning* yang digunakan.

government, the private sector, and research institutions. Among these stakeholders, the actions are further categorized into several areas, with each area identifying a series of related steps to implement the principles of AI, as follows:

a. The Government, designing ideal regulatory and policy interventions, creating awareness, accessibility and capacity building, and facilitating precise procurement strategiest.

b. The private sector and research institutions, incentivising ethics by design (the approach where the responsible AI principle is integrated into all processes of AI technology development)., creating frameworks for compliance with relevant AI standards and guidelines, and the promotion of Responsible AI practices in research.

4. AI regulations published by Indian sectoral regulators include:

a. In the financial sector, the Securities and Exchange Board of India issued a circular in January 2019 to entities in the capital markets regarding requirements related to the applications and systems of AI and machine learning being used.

b. Di sektor kesehatan, strategi Misi Kesehatan Digital Nasional mengidentifikasi perlunya pembuatan panduan dan standar untuk memastikan keandalan sistem AI dalam kesehatan.

5. Mempertimbangkan kemungkinan dampak AI terhadap ekonomi dan masyarakat serta untuk menghasilkan kerangka kebijakan tentang AI, Kementerian Elektronika & Teknologi Informasi India (MeitY) membentuk 4 (empat) komite tentang AI dan telah menerbitkan beberapa laporan tentang masalah keamanan, keselamatan, hukum, dan etika yang berkaitan dengan AI:

b. In the health sector, the National Digital Health Mission strategy identifies the need to create guidelines and standards to ensure the reliability of AI systems in healthcare.

5. Considering the potential impact of AI on the economy and society, as well as to generate a policy framework on AI, the Ministry of Electronics & Information Technology of India (MeitY) has formed 4 (four) committees on AI and has published several reports on issues related to security, safety, law, and ethics concerning AI:

a. *Committee on platforms and data on AI*, berfokus untuk pengembangan platform dan pengelolaan data yang dibutuhkan untuk implementasi AI.

b. *Committee on leveraging AI for identifying national missions in key sectors*, berfokus untuk mengidentifikasi misi nasional di sektor kunci yang dapat dimanfaatkan melalui AI.

c. *Committee on mapping technological capabilities, key policy enablers required across sectors, skilling, reskill, R&D* berfokus untuk memetakan kemampuan teknologi, termasuk teknologi AI yang telah ada, serta riset dan pengembangan.

a. Committee on platforms and data on AI, focusing on the development of platforms and the management of data needed for AI implementation.

b. Committee on leveraging AI for identifying national missions in key sectors, focusing on identifying national missions in key sectors that can be leveraged through AI.

c. Committee on mapping technological capabilities, key policy enablers required across sectors, skilling, reskill, R&D, focusing on mapping technological capabilities, including existing AI technologies, as well as research and development.

d. *Committee on cyber security, safety, legal and ethical issues*, berfokus untuk membahas isu-isu keamanan siber, keselamatan, masalah hukum, dan etika mengenai teknologi, termasuk AI.

6. *Digital Personal Data Protection Act* (No. 22 of 2023)

India menerbitkan Undang-Undang Pelindungan Data Pribadi Digital pada 11 Agustus 2023. Undang-undang ini dipandang mengadopsi metodologi yang mirip dengan *EU's General Data Protection Regulation* (GDPR) dalam mendefinisikan "data pribadi" dan memperluas cakupannya ke semua

d. Committee on cyber security, safety, legal and ethical issues, focusing on discussing issues of cybersecurity, safety, legal matters, and ethics regarding technology, including AI.

6. Digital Personal Data Protection Act (No. 22 of 2023)

India enacted the Digital Personal Data Protection Act on August 11, 2023. This law is seen as adopting a methodology similar to the EU's General Data Protection Regulation (GDPR) in defining "personal data" and extending its scope to all entities that

entitas yang memproses data pribadi, tanpa memandang ukuran atau status kepemilikannya. Terkait AI, beberapa ketentuan dalam undang-undang ini dipandang bertujuan untuk melindungi kemampuan melatih model AI dengan data pribadi. Misalnya, ketentuan pelindungan data pribadi digital tidak berlaku untuk data pribadi yang tersedia untuk umum, asalkan data tersebut dipublikasikan oleh individu yang bersangkutan, dan juga tidak berlaku untuk pemrosesan data pribadi yang diperlukan untuk penelitian, pengarsipan, atau tujuan statistik, yang memenuhi syarat dan kondisi tertentu. Namun, jika pemrosesan ini digunakan untuk mengambil keputusan terkait individu yang datanya diproses, ketentuan pelindungan data pribadi digital tetap berlaku.

process personal data, regardless of size or ownership status. Regarding AI, several provisions in this law are viewed as aimed at protecting the ability to train AI models with personal data. For example, the provisions for digital personal data protection do not apply to personal data that is publicly available, as long as the data is published by the individual concerned, and also do not apply to the processing of personal data necessary for research, archiving, or statistical purposes, which meet certain qualifications and conditions. However, if this processing is used to make decisions related to individuals whose data is processed, the provisions for digital personal data protection still apply.

H. Tiongkok

Tiongkok merupakan salah satu negara yang terdepan dan memiliki undang-undang AI yang paling luas, khususnya yang menyangkut *generative* AI. Persaingan strategis di arena internasional mendorong upaya pemerintah Tiongkok untuk mengendalikan pengembangan teknologi AI dan penggunaannya, sambil mengurangi ketergantungan pada rantai pasokan asing dan teknologi utama.

Strategi AI menyeluruh Tiongkok diartikulasikan dalam sebuah rencana yang dirumuskan pada tahun 2017. Baru-baru ini, negara tersebut berupaya untuk menegaskan pengaruhnya di panggung global dengan mengadopsi sikap yang jelas tentang tata kelola AI, dengan Inisiatif Tata Kelola AI Global Tiongkok.

1. The New Generation AI Development Plan (2017)

   Pada bulan Juli 2017, pemerintah pusat Tiongkok (China's State Council) mengeluarkan pemberitahuan tentang Rencana Pengembangan AI Generasi Baru (Rencana Pengembangan), yang menandai dokumen kebijakan sistematis pertama Tiongkok dalam kecerdasan buatan. Rencana pengembangan tersebut bertujuan untuk menjadikan Tiongkok sebagai pemimpin dunia dalam bidang AI pada tahun 2030. Strategi tersebut mencakup investasi signifikan dalam penelitian dan pengembangan AI, meningkatkan integrasi AI di berbagai sektor, dan mengembangkan regulasi yang komprehensif serta pedoman etika untuk teknologi AI.

   Dilaksanakan dalam tiga tahap, rencana pengembangan tersebut menguraikan tujuan strategis, tugas-tugas penting, alokasi sumber daya, dan langkah-langkah untuk

H. China

China is one of the leading countries with the most comprehensive AI laws, particularly concerning generative AI. Strategic competition in the international arena drives the Chinese government's efforts to control the development and use of AI technology while reducing dependence on foreign supply chains and key technologies.

China's comprehensive AI strategy is articulated in a plan formulated in 2017. Recently, the country has sought to assert its influence on the global stage by adopting a clear stance on AI governance, with the China Global AI Governance Initiative.

1. The New Generation AI Development Plan (2017)

   In July 2017, the China's State Council issued a notice regarding the New Generation AI Development Plan, which marked the first systematic policy document in artificial intelligence in China. The development plan aims to make China a world leader in AI by 2030. The strategy includes significant investments in AI research and development, enhancing the integration of AI across various sectors, and developing comprehensive regulations and ethical guidelines for AI technology.

   Implemented in three phases, the development plan outlines strategic goals, key tasks, resource allocation, and measures for AI development that focus on safety, security,

pengembangan AI yang berfokus pada keselamatan, keamanan, dan pertimbangan etika AI sambil mendorong inovasi dan kemajuan teknologi:

a. Pada tahun 2020, Tiongkok menargetkan teknologi dan aplikasi AI yang mencapai standar global yang canggih, yang mengubah sektor AI menjadi pendorong utama pertumbuhan ekonomi dan diharapkan dapat memberikan peluang baru untuk meningkatkan kesejahteraan publik.
b. Pada tahun 2025, tujuan Tiongkok adalah mencapai terobosan substansial dalam teori dasar AI, memposisikan teknologi dan aplikasi tertentu di garis depan inovasi global, dan AI diharapkan dapat mendorong perubahan transformatif dalam industrialisasi dan ekonomi Tiongkok secara keseluruhan dan memajukan pembentukan *intelligent society*.
c. Pada tahun 2030, visi Tiongkok adalah agar teori, teknologi, dan aplikasi AI yang secara kolektif mencapai posisi terdepan di panggung global dan menjadikan Tiongkok sebagai pusat utama inovasi AI.

The New Generation AI Development Plan ini tetap menjadi landasan strategi AI Tiongkok meskipun peraturan dan inisiatif diluncurkan di tahun-tahun berikutnya yang tetap sejalan dengan tujuan strategis yang diuraikan dalam rencana tahun 2017.

2. The Global AI Governance Initiative of China (2023)

   The Global AI Governance Initiative of China yang diperkenalkan oleh Presiden Xi Jinping pada bulan Oktober 2023 pada forum the *third Belt and Road Forum for International Cooperation* bertujuan untuk membentuk pengembangan dan tata kelola AI dalam skala global, yang menjadi langkah strategis Tiongkok untuk menegaskan pengaruhnya atas tata kelola kecerdasan artifisial global.

   Inisiatif ini terkenal karena penekanannya pada kolaborasi internasional dan tata kelola teknologi AI yang adil, dengan pendekatan seimbang yang mempertimbangkan peluang dan risiko. Inisiatif ini menentang monopoli teknologi dan mendorong kerja sama global untuk mencegah penyalahgunaan teknologi AI. Inisiatif ini juga menyoroti perlunya negara-negara berkembang untuk memiliki suara yang signifikan dalam tata kelola AI global.

   Fitur utama inisiatif ini mencakup fokus pada inklusivitas dan kesetaraan (semua negara dapat

and ethical considerations of AI while promoting innovation and technological advancement:

a. By 2020, China aimed for AI technologies and applications to reach advanced global standards, transforming the AI sector into a major driver of economic growth and expected to provide new opportunities for enhancing public welfare.
b. By 2025, China's goal is to achieve substantial breakthroughs in the fundamental theories of AI, positioning certain technologies and applications at the forefront of global innovation. AI is expected to drive transformative changes in the industrialization and overall economy of China, advancing the establishment of an intelligent society.
c. By 2030, China's vision is for its theories, technologies, and applications of AI to collectively achieve a leading position on the global stage, establishing China as a major hub for AI innovation.

The New Generation AI Development Plan remains the foundation of China's AI strategy, even as regulations and initiatives have been launched in subsequent years that align with the strategic goals outlined in the 2017 plan.

2. The Global AI Governance Initiative of China (2023)

   The Global AI Governance Initiative of China, introduced by President Xi Jinping in October 2023 at the third Belt and Road Forum for International Cooperation, aims to shape the development and governance of AI on a global scale. This initiative represents a strategic move by China to assert its influence over global artificial intelligence governance.

   This initiative is notable for its emphasis on international collaboration and fair governance of AI technology, adopting a balanced approach that considers both opportunities and risks. It opposes technological monopolies and encourages global cooperation to prevent the misuse of AI technology. Additionally, the initiative highlights the need for developing countries to have a significant voice in global AI governance.

   The key features of this initiative include a focus on inclusivity and equality (ensuring that all countries

berpartisipasi dalam pengembangan dan tata kelola AI), komitmen untuk mengembangkan norma etika, perlindungan privasi, dan struktur hukum, yang sejalan dengan anjuran global untuk mendorong regulasi yang kuat, pembangunan berkelanjutan dan memprioritaskan kesejahteraan dan keamanan manusia.

3. Perlindungan Data dan Hak Kekayaan Intelektual

   Pengembang model *generative AI* di Tiongkok harus mematuhi undang-undang yang dirancang untuk melindungi data pribadi dan hak kekayaan intelektual. Pelindungan data pribadi mengacu pada *Personal Information Protection Law* tahun 2021, dimana undang-undang ini merupakan undang-undang privasi data yang komprehensif dan sangat mirip dengan EU's GDPR, yang bertujuan untuk melindungi data pribadi dan sensitif warga negara Tiongkok dengan mengatur akses, pemrosesan, dan pembagian informasi tersebut. Sedangkan menyangkut hak cipta terhadap karya yang dihasilkan AI dan potensi karya tersebut untuk melanggar materi berhak cipta yang ada, telah banyak kasus atau tuntutan hukum yang telah diajukan ke pengadilan selama beberapa tahun terakhir terkait penerapan hukum hak cipta bagi pengembang *generative AI.*

can participate in the development and governance of AI), a commitment to developing ethical norms, privacy protection, and legal frameworks. These elements align with global recommendations to promote robust regulation, sustainable development, and prioritize human welfare and security.

3. Data Protection and Intellectual Property Rights

   Developers of generative AI models in China must comply with laws designed to protect personal data and intellectual property rights. Personal data protection refers to the Personal Information Protection Law of 2021, which is a comprehensive data privacy law that is very similar to the EU's GDPR, aimed at protecting the personal and sensitive data of Chinese citizens by regulating access, processing, and sharing of that information. Regarding copyright for works generated by AI and the potential for those works to infringe upon existing copyrighted material, many cases or lawsuits have been filed in courts over the past few years related to the application of copyright law for developers of generative AI.

4. Administrative Provisions on Algorithm Recommendation for Internet Information Services (2021)

   Menindaklanjuti dengan *Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms* (*Guiding Opinions*) yang diterbitkan oleh *Cyberspace Administration of China* (CAC), CAC bersama tiga kementerian lainnya (*Ministry of Industry and Information Technology, Ministry of Public Security, dan State Administration for Market Regulation*) menerbitkan *Administrative Provisions on Algorithm Recommendation for Internet Information Services* (*Algorithm Recommendation Provisions*) pada bulan Desember 2021 dan mulai berlaku pada tanggal

4. Administrative Provisions on Algorithm Recommendation for Internet Information Services (2021)

   Following the Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms (Guiding Opinions) issued by the Cyberspace Administration of China (CAC), the CAC, along with three other ministries (Ministry of Industry and Information Technology, Ministry of Public Security, and State Administration for Market Regulation), published the Administrative Provisions on Algorithm Recommendation for Internet Information Services (Algorithm Recommendation Provisions) in December 2021, which came into effect on March 1, 2022. This

1 Maret 2022 yang memperkenalkan kerangka peraturan untuk rekomendasi algoritmik yang digunakan dalam layanan daring di Tiongkok.

Regulasi ini dirancang untuk mengawasi penggunaan algoritma di layanan daring, seperti rekomendasi personal, pemeringkatan dan pemilihan, filter pencarian, pengambilan keputusan otomatis, dan lainnya. Aturan ini melibatkan koordinasi antara berbagai departemen pemerintah, termasuk keamanan siber, telekomunikasi, dan pengaturan pasar. Poin penting regulasi:

introduced a regulatory framework for algorithmic recommendations used in online services in China.

This regulation is designed to oversee the use of algorithms in online services, such as personalized recommendations, ranking and selection, search filtering, automated decision-making, and more. The rules involve coordination among various government departments, including cybersecurity, telecommunications, and market regulation. Key points of the regulation include:

a. Pendaftaran algoritma:
   1. Penyedia layanan daring diwajibkan untuk memberikan informasi ke dalam registri algoritma nasional.
   2. Platform media sosial yang memiliki pengaruh pada opini publik atau mobilisasi sosial harus melaporkan detail algoritma mereka kepada pemerintah, termasuk jenis, aplikasi, dan laporan evaluasi algoritma.

b. Standar untuk manajemen informasi:
   1. Penyedia layanan daring harus menerapkan sistem manajemen informasi yang mencakup registrasi pengguna, pemeriksaan pra-publikasi, keamanan data, perlindungan informasi pribadi, dan penanganan insiden keamanan.
   2. Penyedia layanan daring harus secara berkala melakukan audit, evaluasi, dan memvalidasi mekanisme tata kelola algoritma, model, data, dan hasil aplikasi serta melakukan penilaian keamanan dan memelihara catatan jaringan.

c. Larangan aktivitas ilegal, dimana penyedia layanan dilarang menggunakan algoritma untuk:
   1. Memicu perilaku ilegal atau tidak etis.
   2. Menggunakan informasi palsu atau berbahaya untuk memengaruhi pengguna.

a. Algorithm registration:
   1. Online service providers are required to submit information into the national algorithm registry.
   2. Social media platforms that have an influence on public opinion or social mobilization must report the details of their algorithms to the government, including the type, application, and evaluation reports of the algorithms.

b. Standards for information management:
   1. Online service providers must implement an information management system that includes user registration, pre-publication checks, data security, personal information protection, and security incident handling.
   2. Online service providers must periodically conduct audits, evaluations, and validate the governance mechanisms of algorithms, models, data, and application outcomes, as well as perform security assessments and maintain network records.

c. Prohibition of illegal activities, where service providers are prohibited from using algorithms to:
   1. Incite illegal or unethical behavior.
   2. Use false or harmful information to influence users.

3. Membuat akun palsu untuk manipulasi opini publik.
4. Melakukan diskriminasi harga atau praktik monopoli yang tidak wajar.
5. Menyebarkan informasi yang dapat membahayakan kesehatan fisik atau mental anak-anak, atau menyebabkan kecanduan internet.

d. Perlindungan pengguna:
1. Penyedia layanan daring harus memberi informasi kepada pengguna tentang tujuan, prinsip, dan cara kerja algoritma.
2. Pengguna memiliki hak untuk menolak algoritma yang berbasis karakteristik pribadi dan mematikan layanan rekomendasi algoritma sepenuhnya.
3. Perlindungan khusus diberikan kepada kelompok rentan seperti anak-anak, lansia, pekerja, dan konsumen.

5. Internet Information Service Deep Synthesis Management Provisions (Deep Synthesis Regulation) (2022)

Regulasi yang dikenal sebagai "*Deep Synthesis Regulation*" diterbitkan oleh CAC bersama *Ministry of Industry and Information Technology* dan *Ministry of Public Security* pada November 2022 dan berlaku sejak 10 Januari 2023. Aturan ini menjadi regulasi pertama di Tiongkok yang secara khusus mengatur layanan berbasis teknologi *deep synthesis*, seperti *deep learning*, realitas virtual, dan algoritma sintetis lainnya yang digunakan untuk menghasilkan konten sintetik (teks, gambar, audio, video, dll).

Fokus utama regulasi ini adalah mengendalikan teknologi *generative AI* (penyalahgunaan teknologi *deep synthesis*) termasuk *deepfake*, yang dianggap dapat menimbulkan risiko signifikan terhadap masyarakat, melindungi kepentingan publik, keamanan nasional, dan privasi individu, memastikan transparansi dalam penggunaan teknologi *generative AI*, menjaga kepercayaan publik terhadap informasi yang tersedia di internet. Poin utama regulasi sebagai berikut:

a. Kewajiban bagi penyedia layanan dan pendukung teknis
1. Langkah keamanan, dimana penyedia layanan wajib:
a. Menerapkan mekanisme keamanan teknis yang terkendali.
b. Melakukan evaluasi keamanan untuk model atau template yang melibatkan informasi biometrik atau kepentingan nasional.

3. Create fake accounts to manipulate public opinion.
4. Engage in price discrimination or unfair monopolistic practices.
5. Disseminate information that may harm the physical or mental health of children, or cause internet addiction.

d. Consumer protection:
1. Online service providers must inform users about the purpose, principles, and functioning of the algorithms.
2. Users have the right to refuse algorithms based on personal characteristics and to completely disable algorithmic recommendation services.
3. Special protection is provided to vulnerable groups such as children, the elderly, workers, and consumers.

5. Internet Information Service Deep Synthesis Management Provisions (Deep Synthesis Regulation) (2022)

The regulation known as the "Deep Synthesis Regulation" was issued by the CAC in conjunction with the Ministry of Industry and Information Technology and the Ministry of Public Security in November 2022 and has been in effect since January 10, 2023.

This regulation is the first in China to specifically govern services based on deep synthesis technology, such as deep learning, virtual reality, and other synthetic algorithms used to generate synthetic content (text, images, audio, video, etc.).

The main focus of this regulation is to control generative AI technology (the misuse of deep synthesis technology), including deepfakes, which are considered to pose significant risks to society. It aims to protect public interests, national security, and individual privacy, ensure transparency in the use of generative AI technology, and maintain public trust in the information available on the internet. The key points of the regulation are as follows:

a. Requirements for service providers and technical supporters
1. Security measures, where service providers are required to:
a. Implement controlled technical security mechanisms.
b. Conduct security evaluations for models or templates involving biometric information or national interests.

2. Manajemen data pelatihan, dimana penyedia layanan wajib:
   a. Mengelola data pelatihan dengan aman, khususnya jika mengandung informasi pribadi maka harus mematuhi aturan pelindungan data pribadi.
   b. Untuk data biometrik (misalnya wajah dan suara), penyedia layanan harus memastikan pengguna memberi notifikasi dan memperoleh persetujuan dari individu terkait.
3. Kebijakan pengguna dan sistem manajemen:
   a. Penyedia layanan harus memverifikasi identitas pengguna secara nyata (misalnya, menggunakan nomor identitas atau sistem otentikasi daring).
   b. Layanan tidak boleh diberikan kepada pengguna yang tidak terverifikasi.
4. Manajemen konten:
   a. Dilarang menggunakan teknologi untuk membuat atau menyebarkan berita palsu.
   b. Penyedia harus segera menangani informasi ilegal atau berbahaya dan melaporkannya ke otoritas terkait.

2. Training data management, where service providers are required to:
   a. Manage training data securely, especially if it contains personal information, in compliance with personal data protection regulations.
   b. For biometric data (such as facial and voice data), service providers must ensure that users are notified and obtain consent from the individuals concerned.
3. User policies and management systems:
   a. Service providers must verify the real identity of users (for example, by using identification numbers or online authentication systems).
   b. Services must not be provided to unverified users.
4. Content manajement:
   a. It is prohibited to use technology to create or disseminate false news.
   b. Providers must promptly address illegal or harmful information and report it to the relevant authorities.

5. Pelabelan konten yang dihasilkan AI:
   Konten yang dihasilkan oleh AI harus diberi tanda (*watermark*) yang jelas dan mencolok untuk mengindikasikan bahwa konten tersebut dibuat secara sintetis, terutama jika berpotensi membingungkan publik.

b. Kewajiban khusus bagi penyedia layanan yang mempengaruhi opini publik atau mobilisasi sosial
   1. Pendaftaran Algoritma:
      Penyedia layanan AI generatif yang mampu memengaruhi opini publik atau memobilisasi masyarakat diwajibkan untuk mendaftarkan registri algoritma di regulator negara.
   2. Penilaian Keamanan:
      Sebelum meluncurkan produk, aplikasi, atau fitur baru yang dapat memengaruhi opini publik, penyedia layanan harus melakukan penilaian keamanan.
   3. Kriteria layanan yang dicakup dalam regulasi ini:
      Yakni layanan AI generatif yang memiliki kemampuan memengaruhi opini publik atau mobilisasi sosial. Kriteria tersebut mencakup platform seperti:

5. Labeling of AI-generated content:
   AI-generated content must be clearly and prominently labeled (watermarked) to indicate that the content is synthetically created, especially if it has the potential to confuse the public.

b. Special requirements for service providers that influence public opinion or social mobilization
   1. Algorithm registration:
      Service providers of generative AI that can influence public opinion or mobilize society are required to register their algorithms with the national regulator.
   2. Security Assessment:
      Before launching new products, applications, or features that may influence public opinion, service providers must conduct a security assessment.
   3. Criteria for services covered by this regulation:
      Which is generative AI services that have the capability to influence public opinion or social mobilization. The criteria include platforms such as:

a. Forum terbuka, blog, mikroblog, grup obrolan, akun publik, video pendek, *webcast*, berbagi informasi, *embedded programs*, dan layanan informasi lainnya.

b. Layanan yang memungkinkan berbagi opini publik atau memobilisasi masyarakat untuk melakukan kegiatan tertentu.

c. Kewajiban platform distribusi aplikasi

1. Menerapkan mekanisme keamanan seperti peninjauan awal sebelum layanan ditawarkan, manajemen rutin, penanganan darurat.

2. Memeriksa penilaian keamanan dan dokumen pendaftaran dari layanan *deep synthesis*. Jika terjadi pelanggaran, mereka harus segera mengambil tindakan sesuai dengan ketentuan negara.

d. Larangan bagi pengguna

1. Penyedia layanan, pendukung teknis, dan pengguna dilarang menggunakan layanan *deep synthesis* untuk membuat, menyebarkan, atau menerbitkan informasi ilegal dan melakukan aktivitas yang mengancam keamanan nasional, merugikan kepentingan publik, atau mengganggu ketertiban ekonomi dan sosial.

a. Open forums, blogs, microblogs, chat groups, public accounts, short videos, webcasts, information sharing, embedded programs, and other information services.

b. Services that enable the sharing of public opinions or mobilize society to engage in specific activities.

c. Requirements for application distribution platforms

1. Implement security mechanisms such as initial reviews before services are offered, routine management, and emergency handling.

2. Review the security assessments and registration documents of deep synthesis services. In the event of a violation, they must promptly take action in accordance with national regulations.

d. Prohibitions for users

1. Service providers, technical supporters, and users are prohibited from using deep synthesis services to create, disseminate, or publish illegal information and engage in activities that threaten national security, harm public interests, or disrupt economic and social order.

2. Dilarang manipulasi *watermark*, dengan menggunakan metode teknis untuk menghapus, memalsukan, atau menyembunyikan *watermark* pada konten yang dihasilkan oleh teknologi AI.

6. Interim Administrative Measures for Generative AI Services (2023)

a. Ruang lingkup aturan

1. Berlaku untuk penyedia layanan *generative AI* yang menghasilkan konten berupa teks, gambar, audio, dan video untuk publik di Tiongkok.

2. Pengecualian terhadap institusi pendidikan, penelitian, bisnis, atau badan budaya, yang tidak menyediakan layanan AI generatif untuk publik.

2. Manipulation of watermarks is prohibited, using technical methods to remove, forge, or conceal watermarks on content generated by AI technology.

6. Interim Administrative Measures for Generative AI Services (2023)

a. Scope of Regulation

1. Applicable to service providers of generative AI that produce content in the form of text, images, audio, and video for the public in China.

2. Exceptions for educational institutions, research organizations, businesses, or cultural bodies that do not provide generative AI services to the public.

3. Tidak berlaku untuk penyebaran dan penggunaan AI generatif oleh lembaga pendidikan dan penelitian, bisnis, badan budaya publik, dan lembaga profesional yang tidak menyediakan layanan *generative AI* untuk publik di Tiongkok.

b. Kewajiban penyedia layanan *generative AI*

1. Kewajiban umum

a. Meningkatkan transparansi, akurasi, dan keandalan konten yang dihasilkan AI.

b. Melindungi hak privasi, data pribadi, serta mencegah diskriminasi berdasarkan ras, agama, gender, usia, atau profesi.

c. Melarang monopoli atau persaingan tidak sehat yang berbasis algoritma atau data.

3. Not applicable to the dissemination and use of generative AI by educational and research institutions, businesses, public cultural bodies, and professional organizations that do not provide generative AI services to the public in China.

b. Requirements for generative AI service providers

1. General obligation

a. Enhance the transparency, accuracy, and reliability of AI-generated content.

b. Protect privacy rights, personal data, and prevent discrimination based on race, religion, gender, age, or profession.

c. Prohibit monopolies or unfair competition based on algorithms or data.

2. Pelatihan model AI

a. Manajemen data pelatihan, dengan gunakan sumber data yang sah, menghormati hak kekayaan intelektual dan privasi pengguna.

b. Meningkatkan kualitas, akurasi, objektivitas, dan keberagaman data pelatihan.

c. Menerapkan pedoman pelabelan yang jelas dan akurat selama pengembangan model.

3. Konten dan aktivitas ilegal

a. Dilarang menghasilkan konten yang melanggar hukum, seperti yang mengancam keamanan nasional, mempromosikan kekerasan, atau menyebarkan informasi palsu.

b. Mencegah aktivitas ilegal pengguna dengan peringatan, pembatasan fitur, atau pemutusan akses.

c. Melaporkan konten ilegal kepada otoritas terkait dan memperbaiki model agar tidak mengulangi pelanggaran serupa.

4. Perlindungan pengguna

a. Menjaga kerahasiaan data pribadi pengguna dan patuhi permintaan pengguna untuk mengakses atau menghapus data mereka.

b. Menyediakan mekanisme pengaduan dan langkah untuk mencegah kecanduan AI pada anak-anak.

2. AI model training

a. Data training management, using legitimate data sources, respecting intellectual property rights and user privacy.

b. Improve the quality, accuracy, objectivity, and diversity of training data.

c. Implement clear and accurate labeling guidelines during model development.

3. Illegal content and activity

a. Prohibited from generating content that is illegal, such as content that threatens national security, promotes violence, or disseminates false informationl.

b. Prevent illegal user activities through warnings, feature restrictions, or access termination.

c. Report illegal content to the relevant authorities and improve the model to prevent similar violations from occurring again.

4. User protection

a. Maintain the confidentiality of users' personal data and comply with user requests to access or delete their data.

b. Provide a complaint mechanism and measures to prevent AI addiction in children.

5. Pelabelan konten AI, konten yang dihasilkan AI harus diberi *watermark* eksplisit (terlihat) atau implisit (tersembunyi) yang dapat dikenali dengan metode teknis.
6. Kewajiban yang lebih ketat bagi penyedia layanan *generative AI* yang mampu memengaruhi opini publik atau memobilisasi publik. Layanan AI yang berpotensi memengaruhi opini publik atau mobilisasi sosial harus menjalani penilaian keamanan dan registrasi algoritma.

c. Penegakan aturan

1. Otoritas berwenang mengawasi dan memeriksa penyedia *generative AI* secara berkala dan memberikan sanksi, termasuk peringatan, koreksi, atau penghentian layanan jika terjadi pelanggaran.
2. Jika layanan *generative AI* asing melanggar aturan, otoritas Tiongkok dapat mengambil tindakan teknis.
3. Pengguna dapat melaporkan penyedia layanan yang melanggar aturan ini.

d. Kasus penting (2024)

Pengadilan Internet Guangzhou memutuskan bahwa penyedia layanan *generative AI*, meskipun hanya menggunakan model pihak ketiga, tetap bertanggung jawab memenuhi kewajiban hukum, seperti menyediakan mekanisme pengaduan, melabeli konten AI, dan melindungi hak kekayaan intelektual. Keputusan ini menegaskan pentingnya keseimbangan antara inovasi teknologi dan regulasi hukum di Tiongkok.

5. AI content labeling: AI-generated content must be marked with an explicit (visible) or implicit (hidden) watermark that can be recognized through technical methods.
6. Stricter obligations for generative AI service providers that can influence public opinion or mobilize the public. AI services that have the potential to affect public opinion or social mobilization must undergo security assessments and algorithm registration.

c. Regulation enforcement

1. The relevant authorities oversee and inspect generative AI providers periodically and impose sanctions, including warnings, corrections, or service termination in the event of violations.
2. If foreign generative AI services violate the regulations, Chinese authorities may take technical actions.
3. Users can report service providers that violate these regulations.

d. Important case (2024)

The Guangzhou Internet Court ruled that generative AI service providers, even when using third-party models, remain responsible for fulfilling legal obligations, such as providing a complaint mechanism, labeling AI content, and protecting intellectual property rights. This decision emphasizes the importance of balancing technological innovation with legal regulation in China.

I. Singapura

Singapura saat ini belum memiliki undang-undang khusus yang mengikat terkait AI (*hard law*). Namun demikian, Singapura telah mengadopsi pendekatan *soft law* dengan instrumen yang tidak mengikat untuk memberikan pedoman bagi penggunaan AI yang etis dan bertanggung jawab. Kebijakan AI di Singapura sebagian besar dipimpin oleh Infocomm Media Development Authority (IMDA) yang merupakan lembaga pemerintah di bawah Kementerian Komunikasi dan Informasi, untuk mengawasi perkembangan teknologi digital di negara tersebut. Selain itu, strategi AI nasional dikelola oleh Smart Nation and Digital Government Office, yang menetapkan visi jangka panjang Singapura dalam AI.

1. The National AI Strategy (2019)

   Pada 2019, Singapura meluncurkan Strategi AI Nasional pertamanya untuk mengintegrasikan AI ke dalam berbagai sektor ekonomi, termasuk logistik, kesehatan, dan layanan publik. Pada 2023, strategi

I. Singapore

Currently, Singapore does not have binding hard law legislation specifically related to AI. However, Singapore has adopted a soft law approach with non-binding instruments to provide guidelines for the ethical and responsible use of AI. AI policy in Singapore is primarily led by the Infocomm Media Development Authority (IMDA), a government agency under the Ministry of Communications and Information, which oversees the development of digital technology in the country. Additionally, the national AI strategy is managed by the Smart Nation and Digital Government Office, which sets the long-term vision for Singapore in AI.

1. The National AI Strategy (2019)

   In 2019, Singapore launched its first National AI Strategy to integrate AI into various economic sectors, including logistics, healthcare, and public services. In 2023, this

ini diperbarui menjadi NAIS 2.0, yang menetapkan 15 langkah aksi dalam tiga sistem utama untuk mendorong inovasi AI dalam 3-5 tahun ke depan.

2. The Model AI Governance Framework (2020)

   Pada tahun 2020, IMDA merilis edisi kedua dari *Model AI Governance Framework* yang merupakan pengkinian dari edisi pertama yang dirilis dari tahun 2019. Panduan ini memuat etika dan tata kelola saat mengimplementasikan solusi AI sektor swasta, yang sebagian besar berfokus pada struktur tata kelola internal, keterlibatan manusia dalam pengambilan keputusan AI, manajemen operasi, dan interaksi pemangku kepentingan.

   The Model AI Governance Framework ini menekankan dua prinsip utama yakni Keputusan AI harus dapat dijelaskan, transparan, dan adil, serta AI harus memprioritaskan kepentingan manusia, meningkatkan kemampuan manusia, dan melindungi hak asasi manusia. Berdasarkan prinsip tersebut, *framework* ini memberikan panduan dalam empat aspek utama:

   a. Tata kelola internal.
   b. Keterlibatan manusia dalam pengambilan keputusan oleh AI.
   c. Manajemen operasional.
   d. Interaksi dan komunikasi dengan pemangku kepentingan.

   Selain itu, kerangka ini mengintegrasikan manajemen risiko AI dalam struktur manajemen risiko perusahaan, termasuk evaluasi *datasets* untuk memastikan akurasi dan menghindari bias, serta pemantauan dan pelaporan risiko yang komprehensif. Meskipun tidak bersifat mengikat, penerapan kerangka ini membantu organisasi mematuhi aturan nasional seperti undang-undang privasi data Singapura. IMDA juga menekankan bahwa prinsip utama dari kerangka ini adalah kepraktisan. Sebagai pendukung, IMDA menerbitkan dokumen pendamping untuk *Model Framework*, yakni:

   a. *Compendium of Use Cases* (terdiri dari 2 (dua) volume), yang membahas bagaimana berbagai organisasi di berbagai sektor dan ukuran (lokal dan internasional) mampu menerapkan dan menyelaraskan praktik tata kelola AI mereka dengan *Model Framework*., serta secara efektif menerapkan praktik tata kelola AI yang akuntabel dan memperoleh manfaat dari penggunaan AI dalam lini bisnis mereka.
   b. *Implementation and Self-Assessment Guide for Organizations*, sebagai *living document* yang merupakan panduan pendamping untuk *Model Framework* dan

strategy was updated to NAIS 2.0, which outlines 15 action steps across three main systems to promote AI innovation over the next 3-5 years.

2. The Model AI Governance Framework (2020)

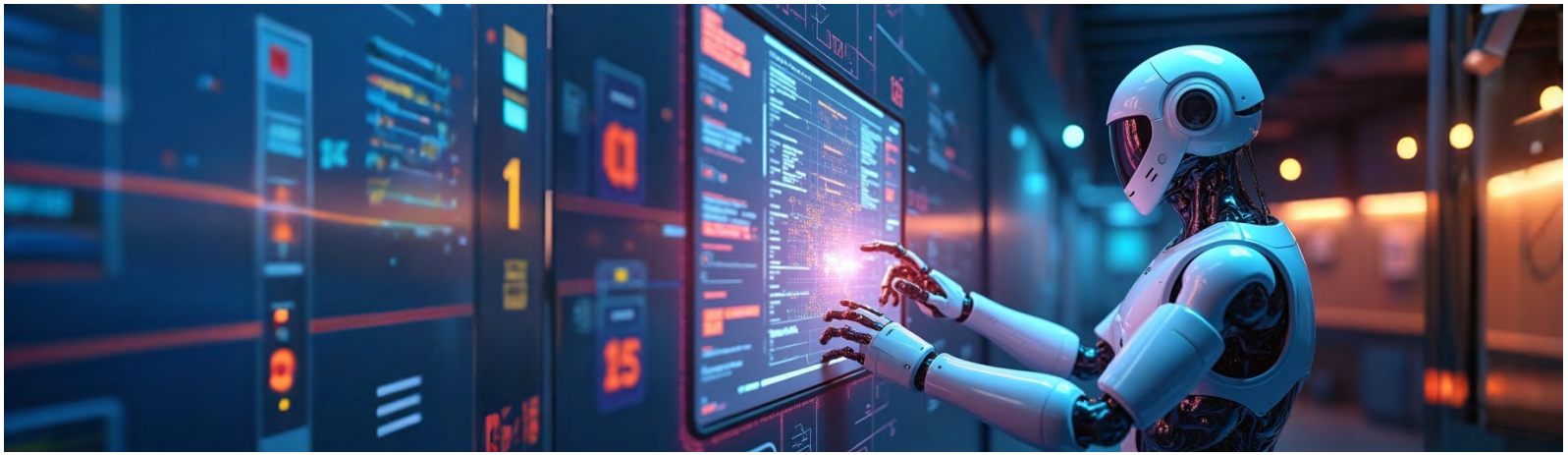
   In 2020, the IMDA released the second edition of the Model AI Governance Framework, which is an update of the first edition released in 2019. This guide includes ethics and governance considerations when implementing AI solutions in the private sector, primarily focusing on internal governance structures, human involvement in AI decision-making, operational management, and stakeholder interactions.

   The Model AI Governance Framework emphasizes two main principles: AI decisions must be explainable, transparent, and fair, and AI must prioritize human interests, enhance human capabilities, and protect human rights. Based on these principles, the framework provides guidance in four key aspects:

   a. Internal governance.
   b. Human involvement in AI decision-making.
   c. Operational management.
   d. Interaction and communication with stakeholder.

   In addition, this framework integrates AI risk management into the company's risk management structure, including the evaluation of datasets to ensure accuracy and avoid bias, as well as comprehensive risk monitoring and reporting. Although it is not binding, the implementation of this framework helps organizations comply with national regulations such as Singapore's data privacy laws. The IMDA also emphasizes that the main principle of this framework is practicality. As a supporter, the IMDA publishes a companion document for the Model Framework, namely:

   a. Compendium of Use Cases (consisting of 2 (two) volumes), which discusses how various organizations across different sectors and sizes (local and international) are able to implement and align their AI governance practices with the Model Framework, as well as effectively apply accountable AI governance practices and benefit from the use of AI in their business lines.
   b. Implementation and Self-Assessment Guide for Organizations, as a living document that serves as a companion guide to the Model Framework and aims to help



bertujuan untuk membantu organisasi menilai keselarasan praktik tata kelola AI mereka dengan *Model Framework*, mengidentifikasi potensi kesenjangan dalam proses yang ada, dan mengatasinya sebagaimana mestinya.

3. The Model AI Governance Framework for Generative AI (2024)

   Dirilis pada tanggal 30 Mei 2024 oleh IMDA dan AI Verify Foundation, dokumen ini merupakan pedoman berikutnya, setelah sebelumnya menerbitkan *The Model AI Governance Framework* yang telah dirilis tahun 2019 dan diperbarui pada tahun 2020. Penerbitan dokumen ini mempertimbangkan munculnya *generative AI* yang telah memperkuat beberapa risiko AI yang sama (misalnya, bias, penyalahgunaan, kurangnya penjelasan), dan memperkenalkan risiko baru (misalnya, halusinasi, pelanggaran hak cipta, penyelarasan nilai). Oleh karena itu, kerangka tata kelola yang ada perlu ditinjau ulang untuk mendorong ekosistem tepercaya yang lebih luas dan untuk mencapai kesimbangan antara melindungi pengguna dan mendorong inovasi.

   Model AI Governance Framework for Generative AI ini berupaya untuk menetapkan pendekatan yang sistematis dan seimbang untuk mengatasi masalah *generative AI*

organizations assess the alignment of their AI governance practices with the Model Framework, identify potential gaps in existing processes, and address them appropriately.

3. The Model AI Governance Framework for Generative AI (2024*)*

   Released on May 30, 2024, by the IMDA and AI Verify Foundation, this document serves as the next guideline, following the publication of The Model AI Governance Framework, which was released in 2019 and updated in 2020. The publication of this document takes into account the emergence of generative AI, which has reinforced several of the same AI risks (e.g., bias, misuse, lack of explainability) and introduced new risks (e.g., hallucination, copyright infringement, value alignment). Therefore, the existing governance framework needs to be reviewed to promote a broader trusted ecosystem and to achieve a balance between protecting users and encouraging innovation.

   The Model AI Governance Framework for Generative AI aims to establish a systematic and balanced approach to addressing generative AI issues

sambil terus memfasilitasi inovasi, yang mencakup 9 (sembilan) area utama sebagai berikut:

a. Akuntabilitas

Harus ada alokasi tanggung jawab yang jelas sejak awal proses pengembangan dan setelah sistem AI diterapkan, serta adanya struktur insentif yang tepat bagi berbagai pelaku dalam siklus hidup pengembangan sistem AI agar bertanggung jawab kepada pengguna akhir.

b. Data

Data berkualitas sangat penting (seperti melalui penggunaan sumber data tepercaya) untuk memastikan hasil model yang baik karena data merupakan inti dari pengembangan model. Penggunaan data yang sensitif, seperti data pribadi dan hak cipta, harus diperlakukan secara adil dan transparan.

c. Pengembangan dan Penerapan yang Terpercaya

Meningkatkan transparansi termasuk penerapan standar keamanan dan praktik terbaik industri dalam pengembangan, evaluasi, dan pengungkapan.

d. Pelaporan Insiden

Menerapkan sistem manajemen insiden untuk pemberitahuan tepat waktu, perbaikan, mitigasi risiko dan mendukung peningkatan berkelanjutan sistem AI, karena tidak ada sistem AI yang sempurna.

e. Pengujian dan Jaminan Keandalan

Pengujian AI oleh pihak ketiga dapat meningkatkan kepercayaan pengguna. Penting juga untuk mengembangkan standar umum seputar pengujian AI untuk memastikan kualitas dan konsistensi.

f. Keamanan

Generative AI membawa risiko keamanan baru yang berbeda dari perangkat lunak biasa. Oleh karena itu, kerangka kerja yang ada untuk keamanan informasi perlu diadaptasi dan alat pengujian baru perlu dikembangkan untuk mengatasi risiko ini.

while continuing to facilitate innovation, which includes 9 (nine) key areas as follows:

a. Accountability

There must be a clear allocation of responsibilities from the beginning of the development process and after the AI system is deployed, as well as the presence of appropriate incentive structures for various actors in the AI system development lifecycle to ensure accountability to end users.

b. Data

High-quality data (such as using trusted data sources) is essential to ensure good model outcomes, as data is at the core of model development. The use of sensitive data, such as personal data and copyrighted material, must be handled fairly and transparently.

c. Reliable Development and Implementation

Enhancing transparency includes the implementation of security standards and industry best practices in development, evaluation, and disclosure.

d. Incident Reporting

Implementing an incident management system for timely notification, remediation, risk mitigation, and supporting the continuous improvement of AI systems, as no AI system is perfect.

e. Testing and Reliability Assurance

Third-party testing of AI can enhance user trust. It is also important to develop common standards around AI testing to ensure quality and consistency.

f. Security

Generative AI introduces new security risks that differ from those of conventional software. Therefore, existing frameworks for information security need to be adapted, and new testing tools must be developed to address these risks.

g. Asal Konten

AI dapat memperburuk penyebaran informasi palsu (misinformasi). Oleh karena itu, transparansi terkait asal-usul konten sangat penting. Solusi teknis seperti digital *watermarking* dan kriptografi perlu digunakan dalam konteks yang tepat dalam membantu melacak sumber konten AI.

h. Riset dan Pengembangan Keamanan AI

Mempercepat penelitian dan pengembangan melalui kerjasama global antar Institut Keamanan AI untuk meningkatkan keselarasan model dengan tujuan dan nilai-nilai kemanusiaan, sertauntuk mengoptimalkan sumber daya yang terbatas dalam mencapai dampak maksimum.

i. AI untuk Kepentingan Publik

AI yang bertanggung jawab mencakup penggunaan AI untuk memberi manfaat bagi masyarakat dan bisnis agar dapat berkembang di masa depan yang didukung AI, dan tidak hanya terbatas pada mitigasi risiko, melalui peningkatan akses dan adopsi sektor publik, meningkatkan keterampilan pekerja, dan mengembangkan sistem AI secara berkelanjutan.

g. Content Source

AI can exacerbate the spread of false information (misinformation). Therefore, transparency regarding the origin of content is crucial. Technical solutions such as digital watermarking and cryptography need to be employed in the appropriate context to help trace the sources of AI-generated content.

h. AI Security Research and Development

Accelerating research and development through global collaboration among AI Security Institutes to enhance the alignment of models with human goals and values, as well as to optimize limited resources in achieving maximum impact.

i. AI for the Public Good

Responsible AI encompasses the use of AI to benefit society and businesses to thrive in a future supported by AI, and is not limited to risk mitigation. This includes enhancing access and adoption in the public sector, improving worker skills, and developing AI systems sustainably.

4. *Principles to Promote Fairness, Ethics, Accountability, and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector*

Mempertimbangkan bahwa di era di mana kecerdasan artifisial dan analisis data (*artificial intelligence and data analytics*/AIDA) semakin banyak digunakan dalam pengambilan keputusan terkait produk dan layanan keuangan, Singapura telah memperkenalkan serangkaian prinsip untuk memastikan teknologi AI digunakan secara bertanggung jawab, efektif dan etis.

Prinsip-prinsip ini bertujuan untuk mengurangi risiko yang terkait dengan AIDA, yang dapat menyebabkan penyalahgunaan sistematis dan dampak yang meluas jika tidak dikelola dengan baik. Sehubungan dengan hal tersebut, Monetary Authority of Singapore bersama dengan para pemangku kepentingan mengembangkan prinsip-prinsip ini melalui *Fairness, Ethics, Accountability and Transparency* (FEAT) untuk menyediakan kerangka kerja yang kuat bagi sektor keuangan. Prinsip FEAT bertujuan sebagai:

4. Principles to Promote Fairness, Ethics, Accountability, and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector

Considering that in this current era artificial intelligence and data analytics (AIDA) are increasingly used in decision-making related to financial products and services, Singapore has introduced a set of principles to ensure that AI technology is used responsibly, effectively, and ethically.

These principles aim to mitigate the risks associated with AIDA, which could lead to systematic abuse and widespread impacts if not managed properly. In this regard, the Monetary Authority of Singapore, together with stakeholders, developed these principles through Fairness, Ethics, Accountability, and Transparency (FEAT) to provide a robust framework for the financial sector. The FEAT principles aim to serve as:

a. Panduan dengan serangkaian prinsip dasar yang perlu dipertimbangkan bagi perusahaan yang menyediakan produk dan layanan keuangan ketika menggunakan AIDA dalam pengambilan keputusan.
b. Tata kelola operasional, untuk membantu perusahaan mengintegrasikan dan menerapkan standar tata kelola untuk AIDA dalam model bisnis mereka.
c. Meningkatkan kepercayaan publik terhadap penggunaan AIDA dengan memastikan penerapannya yang bertanggung jawab.

Prinsip FEAT memuat:

a. Fairness (Keadilan)

Justifiability (Dapat Terjustifikasi):

1. Keputusan berbasis AIDA tidak boleh secara sistematis merugikan individu atau kelompok kecuali jika keputusan tersebut dapat dibenarkan.
2. Penggunaan atribut pribadi sebagai faktor dalam keputusan AIDA harus memiliki justifikasi yang jelas.

Accuracy and Bias (Akurasi dan Bias)

1. Data dan model yang digunakan dalam keputusan AIDA harus diperiksa dan divalidasi secara berkala untuk memastikan akurasi, relevansi, serta meminimalkan bias yang tidak disengaja.
2. Keputusan oleh AIDA harus ditinjau secara berkala untuk memastikan model berfungsi sesuai dengan desain dan tujuan awalnya.

b. Ethics (Etika)

1. Penggunaan AIDA harus selaras dengan standar etika, nilai, dan kode etik perusahaan.
2. Keputusan berbasis AIDA harus mematuhi standar etika yang setidaknya sama dengan keputusan yang dibuat oleh manusia.

c. Accountability (Akuntabilitas)

Akuntabilitas Internal

1. Penggunaan AIDA dalam pengambilan keputusan harus mendapatkan persetujuan dari otoritas internal yang berwenang.
2. Perusahaan yang menggunakan AIDA bertanggung jawab atas model AIDA yang dikembangkan secara internal maupun diperoleh dari pihak eksternal.
3. Perusahaan harus secara proaktif meningkatkan kesadaran manajemen dan dewan pengurus tentang penggunaan AIDA.

Akuntabilitas Eksternal

1. Subjek data (individu yang datanya digunakan) harus diberikan akses untuk mengajukan pertanyaan, banding,

a. A guide with a set of fundamental principles that need to be considered by companies providing financial products and services when using AIDA in decision-making.
b. Operational governance, to assist companies in integrating and implementing governance standards for AIDA within their business models.
c. Enhancing public trust in the use of AIDA by ensuring its responsible application.

The FEAT principle includes:

a. Fairness

Justifiability:

1. AIDA-based decisions must not systematically disadvantage individuals or groups unless such decisions can be justified.
2. The use of personal attributes as factors in AIDA decisions must have clear justification.

Accuracy and Bias

1. The data and models used in AIDA decisions must be regularly examined and validated to ensure accuracy, relevance, and to minimize unintended bias.
2. Decisions made by AIDA must be reviewed periodically to ensure that the models function according to their original design and objectives.

b. Ethics

1. The use of AIDA must align with ethical standards, values, and the company's code of conduct.
2. AIDA-based decisions must adhere to ethical standards that are at least equivalent to those of decisions made by humans.

c. Accountability

Internal Accountability

1. The use of AIDA in decision-making must receive approval from the appropriate internal authorities.
2. Companies using AIDA are responsible for both internally developed and externally sourced AIDA models.
3. Companies must proactively raise management and board awareness regarding the use of AIDA.

External Accountability

1. Data subjects (individuals whose data is used) must be granted access to raise questions, appeals, and requests for review

dan permintaan peninjauan terkait keputusan AIDA yang berdampak pada mereka.

2. Data tambahan yang telah diverifikasi dan relevan yang diberikan oleh subjek data harus diperhitungkan saat melakukan peninjauan atas keputusan oleh AIDA.

d. Transparency (Transparansi)

1. Untuk meningkatkan kepercayaan publik, penggunaan AIDA harus diinformasikan secara proaktif kepada subjek sebagai bagian dari komunikasi umum.
2. Berdasarkan permintaan, subjek data harus diberikan penjelasan yang jelas tentang data yang digunakan dalam keputusan AIDA dan bagaimana data tersebut mempengaruhi keputusan.
3. Berdasarkan permintaan, subjek data harus diberikan penjelasan yang jelas mengenai dampak dari keputusan berbasis AIDA terhadap mereka.

regarding AIDA decisions that impact them.

2. Additional verified and relevant data provided by data subjects must be taken into account when reviewing decisions made by AIDA.

d. Transparency

1. To enhance public trust, the use of AIDA must be proactively communicated to subjects as part of general communication efforts.
2. Upon request, data subjects must be provided with a clear explanation of the data used in AIDA decisions and how that data influences the decisions.
3. Upon request, data subjects must be provided with a clear explanation regarding the impact of AIDA-based decisions on them.

J. Jepang

Teknologi yang terkait dengan AI khususnya *generative* AI terus berkembang, serta peluang dalam penggunaan AI dengan berbagai kemungkinannya menjadi terus meningkat dimana AI digunakan untuk menciptakan inovasi industri dan pemecahan tantangan sosial. Disamping itu, AI diharapkan akan berkontribusi pada terwujudnya "Masyarakat 5.0. (Society 5.0)".

Pada 19 April 2024 diterbitkan pedoman/panduan AI yakni "*AI Guidelines for Business Ver1.0*" yang bertujuan untuk mengatasi perubahan teknologi yang cepat termasuk perkembangan *generative* AI, dimana Ministry of Economy, Trade and Industry (METI) and the Ministry of Internal Affairs and Communications (MIC) mengintegrasikan dan memperbarui 3 (tiga) pedoman yang ada di Jepang terkait AI (AI R&D Guidelines (2017; MIC), AI Utilization Guidelines (2019; MIC), and Governance Guidelines for Implementation of AI Principles Ver. 1.1 (2022; METI)) serta melalui berbagai diskusi dengan para ahli.

J. Japan

The technology related to AI, especially generative AI, continues to evolve, and the opportunities in the use of AI with its various possibilities are continuously increasing, where AI is used to create industrial innovations and solve social challenges. In addition, AI is expected to contribute to the realization of "Society 5.0".

On April 19, 2024, the "AI Guidelines for Business Ver1.0" were published, aimed at addressing rapid technological changes, including the development of generative AI. The Ministry of Economy, Trade and Industry (METI) and the Ministry of Internal Affairs and Communications (MIC) integrated and updated 3 (three) existing guidelines related to AI in Japan (AI R&D Guidelines (2017; MIC), AI Utilization Guidelines (2019; MIC), and Governance Guidelines for Implementation of AI Principles Ver. 1.1 (2022; METI)) through various discussions with experts.

Common Guiding Principles (Prinsip Panduan Umum) menguraikan apa yang dikerjakan masing-masing aktor dalam kolaborasi untuk mencapai masyarakat yang dituju melalui AI, dan diformulasikan berdasarkan *Social Principles of Human-Centric AI* yang memadukan 3 (tiga) pedoman di Jepang serta mempertimbangkan tren di negara lain dan munculnya teknologi baru. Hasilnya, pedoman ini disusun berdasarkan apa yang dikerjakan masing-masing aktor dan apa yang diharapkan untuk dikerjakan dalam kolaborasi dengan masyarakat.

Common Guiding Principles outline what each actor does in collaboration to achieve the desired society through AI, and are formulated based on the Social Principles of Human-Centric AI, which integrate the 3 (three) guidelines in Japan while considering trends in other countries and the emergence of new technologies. As a result, these guidelines are structured based on what each actor is currently doing and what is expected to be done in collaboration with society.

Dengan mempertimbangkan peran spesifik mereka dalam siklus hidup AI, pihak-pihak yang menjadi sasaran Pedoman AI ini secara garis besar dikelompokkan ke dalam 3 (tiga) kategori (aktor) yakni AI *developers* (pengembang AI), AI *providers* (penyedia AI), dan AI *business users* (pengguna bisnis AI). *Data providers* (penyedia data) dan *non-business users* (pengguna nonbisnis) tidak termasuk.

Considering their specific roles in the AI lifecycle, the parties targeted by these AI Guidelines are broadly categorized into 3 (three) groups (actors): AI developers, AI providers, and AI business users. Data providers and non-business users are not included.

1. Prinsip panduan umum untuk semua pelaku

   Setiap pelaku bisnis AI harus mengembangkan, menyediakan, atau menggunakan sistem dan layanan AI dengan menghormati aturan hukum, hak asasi manusia, demokrasi, keberagaman, dan masyarakat yang adil dan jujur sesuai dengan prinsip pertama yakni berpusat pada manusia (*human-centric*).

   Undang-undang yang relevan, termasuk Konstitusi Jepang, Undang-Undang Dasar Kekayaan Intelektual, Undang-Undang tentang Pelindungan Informasi Pribadi, serta undang-undang dan peraturan lain yang relevan dan berlaku di masing-masing bidang yang berkaitan dengan AI harus dipatuhi.

   Prinisp umum:

   a. Human Centric (Berpusat pada Manusia)

      1. AI harus digunakan untuk memperluas kemampuan manusia dan memungkinkan berbagai individu mencapai kesejahteraan yang beragam.

1. Common guiding principles for all actors

   Every AI business actor must develop, provide, or use AI systems and services while respecting legal rules, human rights, democracy, diversity, and a fair and honest society in accordance with the first principle, which is human-centric.

   Relevant laws, including the Constitution of Japan, the Basic Intellectual Property Act, the Act on the Protection of Personal Information, as well as other applicable laws and regulations relevant to each field related to AI, must be complied with.

   General principles:

   a. Human-centric

      1. AI must be used to enhance human capabilities and enable various individuals to achieve diverse well-being.

2. Sadar akan risiko meningkatnya disinformasi, misinformasi, dan bias informasi yang dapat mengganggu stabilitas masyarakat, serta mengambil langkah-langkah pencegahan yang diperlukan.
3. Memastikan bahwa AI dapat diakses oleh kelompok rentan agar lebih banyak orang dapat merasakan manfaatnya.

b. Safety (Keamanan)

1. Melakukan analisis risiko yang tepat untuk mengantisipasi dan menangani berbagai risiko yang mungkin muncul.
2. Mencegah penyalahgunaan AI yang menyimpang dari tujuan awalnya, terutama dalam ruang lingkup yang tidak dapat dikendalikan oleh penyedia layanan AI.
3. Menjaga akurasi dan relevansi data yang digunakan dalam pelatihan AI sesuai dengan tujuan sistem, serta memastikan transparansi data, kepatuhan terhadap hukum, dan pembaruan model AI secara berkala.

c. Fairness (Keadilan)

1. Berupaya menghilangkan bias dan diskriminasi yang merugikan individu atau kelompok berdasarkan ras, gender, asal negara, usia, pandangan politik, agama, dan faktor lainnya.
2. Untuk mencegah AI menghasilkan keputusan yang tidak adil, penting untuk menerapkan intervensi manusia yang tepat waktu, bukan hanya mengandalkan keputusan AI sepenuhnya.

d. Privacy Protection (Perlindungan Privasi)

1. Mematuhi hukum terkait perlindungan data pribadi serta menetapkan dan menginformasikan kebijakan privasi yang jelas bagi pengguna AI.
2. Menghormati privasi pemangku kepentingan dengan mempertimbangkan konteks sosial dan harapan dari masyarakat.

e. Ensuring Security (Memastikan Keamanan)

1. Menjaga kerahasiaan, integritas, dan ketersediaan sistem serta layanan AI untuk memastikan penggunaannya yang aman dan terlindungi.
2. Meningkatkan langkah-langkah keamanan karena metode serangan terhadap AI terus berkembang setiap hari.

f. Transparency (Transparansi)

Memberikan informasi yang jelas kepada pemangku kepentingan tentang penggunaan AI, termasuk cakupan penggunaannya, metode pengumpulan data, kemampuan serta keterbatasan AI, dan cara penggunaannya yang tepat atau tidak tepat.

2. Be aware of the risks of increasing disinformation, misinformation, and information bias that can disrupt societal stability, and take the necessary preventive measures.
3. Ensure that AI is accessible to vulnerable groups so that more people can experience its benefits.

b. Safety

1. Conduct appropriate risk analysis to anticipate and address various risks that may arise.
2. Prevent the misuse of AI that deviates from its original purpose, especially in areas that cannot be controlled by AI service providers.
3. Maintain the accuracy and relevance of the data used in AI training in accordance with the system's objectives, as well as ensure data transparency, compliance with laws, and regular updates of AI models.

c. Fairness

1. Strive to eliminate bias and discrimination that harm individuals or groups based on race, gender, nationality, age, political views, religion, and other factors.
2. To prevent AI from making unfair decisions, it is important to implement timely human interventions, rather than solely relying on AI decisions.

d. Privacy Protection

1. Comply with laws related to personal data protection and establish and communicate clear privacy policies for AI users.
2. Respect the privacy of stakeholders by considering the social context and societal expectations.

e. Ensuring Security

1. Maintain the confidentiality, integrity, and availability of AI systems and services to ensure their safe and protected use.
2. Enhance security measures as attack methods against AI continue to evolve every day.

f. Transparency

Provide clear information to stakeholders about the use of AI, including the scope of its use, data collection methods, the capabilities and limitations of AI, and the appropriate or inappropriate ways to use it.

g. *Accountability* (Akuntabilitas)

1. Menyediakan informasi yang memungkinkan pelacakan (*traceability*) serta memastikan kepatuhan terhadap prinsip-prinsip panduan umum.
2. Menetapkan dan melaporkan kebijakan terkait tata kelola AI dan perlindungan privasi kepada publik.
3. Mendokumentasikan dan menyimpan informasi yang relevan dalam jangka waktu yang ditentukan agar dapat dirujuk kapan pun diperlukan.

h. *Education/Literacy* (Edukasi/Literasi)

1. Memastikan bahwa individu yang terlibat dalam pengembangan AI memiliki literasi AI yang cukup sesuai perannya.
2. Memberikan edukasi kepada pemangku kepentingan mengenai sifat kompleks AI, risiko misinformasi, dan kemungkinan penyalahgunaan AI secara sengaja.

i. *Ensuring Fair Competition* (Memastikan Persaingan yang Adil)

Menjaga lingkungan persaingan yang sehat agar bisnis dan layanan baru berbasis AI dapat berkembang, mendorong pertumbuhan ekonomi yang berkelanjutan, serta menyediakan solusi untuk tantangan sosial.

g. Accountability

1. Provide information that enables traceability and ensure compliance with the common guiding principles.
2. Establish and publicly report policies related to AI governance and privacy protection.
3. Document and retain relevant information for a specified period so that it can be referenced whenever needed.

h. Education/Literacy

1. Ensure that individuals involved in AI development have adequate AI literacy relevant to their roles.
2. Educate stakeholders about the complex nature of AI, the risks of misinformation, and the potential for intentional misuse of AI.

i. Ensuring Fair Competition

Maintain a healthy competitive environment so that new AI-based businesses and services can thrive, promote sustainable economic growth, and provide solutions to social challenges.

j. *Innovation* (Inovasi)

1. Mempromosikan internasionalisasi, diversifikasi, serta kolaborasi antara industri, akademisi, dan pemerintah untuk mendorong inovasi terbuka.
2. Memastikan keterhubungan dan interoperabilitas antara sistem/layanan AI dengan sistem/layanan AI lainnya.
3. Mematuhi spesifikasi standar yang berlaku dalam industri AI.

2. Hal yang terkait AI *developers* (pengembang AI)

Para pengembang AI perlu mempelajari terlebih dahulu dampak yang mungkin ditimbulkan oleh AI yang mereka kembangkan dan mengambil langkah-langkah yang diperlukan, karena mereka memiliki kendali langsung untuk merancang dan memodifikasi model AI.

a. Selama Proses Prapelatihan/Pelatihan Data

1. Pelatihan data yang tepat

a. Mengumpulkan data pelatihan dengan metode yang sesuai, seperti *privacy-by-design*, serta memastikan bahwa data pihak ketiga, termasuk data pribadi dan materi yang memiliki hak kekayaan intelektual, dikelola sesuai dengan peraturan hukum yang berlaku.

j. Innovation

1. Promote internationalization, diversification, and collaboration among industry, academia, and government to encourage open innovation.
2. Ensure connectivity and interoperability between AI systems/services and other AI systems/services.
3. Comply with applicable standard specifications in the AI industry.

2. Matters related to AI developers

AI developers need to first study the potential impacts that the AI they develop may cause and take the necessary steps, as they have direct control to design and modify AI models.

a. During the Data Pretraining/Training process

1. Proper data training

a. Collect training data using appropriate methods, such as privacy-by-design, and ensure that third-party data, including personal data and materials with intellectual property rights, is managed in accordance with applicable legal regulations.

b. Menerapkan langkah-langkah perlindungan data, misalnya dengan menerapkan fungsi manajemen dan pembatasan akses data.

2. Memperhatikan bias dalam data

a. Mengendalikan kualitas data dengan mempertimbangkan kemungkinan adanya bias dalam proses pelatihan data dan model AI.

b. Karena bias tidak dapat sepenuhnya dihilangkan, pastikan model AI dilatih dengan *dataset* yang representatif dan sistem AI diuji untuk meminimalkan bias.

b. Implement data protection measures, such as applying data management and access restriction functions.

2. Pay attention to bias in the data

a. Control data quality by considering the possibility of bias in the data training process and AI models.

b. Since bias cannot be completely eliminated, ensure that the AI model is trained with a representative dataset and that the AI system is tested to minimize bias.

b. Saat Mengembangkan AI

1. Pengembangan AI dengan mempertimbangkan manusia dan lingkungan

   Memperhitungkan performa AI di lingkungan yang tidak terduga serta mengembangkan metode untuk meminimalkan risiko terhadap manusia dan lingkungan.

2. Pengembangan yang berkontribusi pada penggunaan AI yang tepat

   Menetapkan kebijakan dan pedoman yang jelas untuk penggunaan AI yang aman, serta memilih model AI yang sesuai untuk pelatihan lanjutan.

3. Mengurangi bias dalam algoritma AI

a. Memperhitungkan kemungkinan bias yang timbul dari setiap elemen teknis yang membentuk model AI.

b. Memastikan model AI dilatih dengan *dataset* yang representatif guna mengurangi bias.

4. Penerapan mekanisme keamanan

   Menerapkan langkah-langkah keamanan yang sesuai berdasarkan karakteristik teknologi AI yang digunakan (*secure by design*).

b. During AI Development

1. AI development with consideration for humans and the environment

   Taking into consideration the performance of AI in unexpected environments and develop methods to minimize risks to humans and the environment.

2. Development that contributes to the appropriate use of AI

   Establish clear policies and guidelines for the safe use of AI, as well as select appropriate AI models for further training.

3. Reduce bias in AI algorithms

a. Consider the potential bias arising from each technical element that makes up the AI model.

b. Ensure that the AI model is trained with a representative dataset to reduce bias.

4. Implementation of Safety Mechanism

   Implement appropriate security measures based on the characteristics of the AI technology used (secure by design).

5. Memastikan verifikasi AI

a. Memahami bahwa kualitas *output* AI dapat berubah atau tidak mencapai tingkat akurasi yang diharapkan setelah mulai digunakan.

b. Mencatat seluruh proses pengembangan AI untuk verifikasi di masa depan serta mengambil langkah-langkah untuk mempertahankan dan meningkatkan kualitas AI.

c. Setelah Mengembangkan AI

1. Memperhatikan tren terbaru

Mengidentifikasi dan mengantisipasi metode serangan baru terhadap sistem AI yang terus berkembang setiap hari.

5. Establish AI verification

a. Understand that the quality of AI output may change or not reach the expected level of accuracy after it starts being used.

b. Document the entire AI development process for future verification and take steps to maintain and improve AI quality.

c. Post AI Development

1. Observe the latest trends

Identify and anticipate new attack methods against AI systems that are continuously evolving.

2. Menyediakan informasi kepada pemangku kepentingan

Memberikan informasi terkait keamanan AI, termasuk karakteristik teknis, mekanisme perlindungan, risiko yang dapat diprediksi, langkah-langkah mitigasi, serta penyebab dan penanganan kegagalan sistem.

3. Menjelaskan kepatuhan terhadap prinsip panduan umum kepada penyedia AI

Menjelaskan kepada penyedia AI (AI *providers*) bahwa kualitas *output* AI dapat berubah secara signifikan, yang dapat menimbulkan risiko baru.

2. Provide information to stakeholders

Provide information related to AI security, including technical characteristics, protection mechanisms, predictable risks, mitigation measures, as well as causes and handling of system failures.

3. Explain compliance with general guideline principles to AI providers

Explain to AI providers that the quality of AI output may change significantly, which can pose new risks.

4. Dokumentasi informasi pengembangan

   Menyusun dokumentasi tentang proses pengembangan AI, pengumpulan dan pelabelan data, algoritma yang digunakan, serta informasi terkait lainnya.

5. Berperan dalam inovasi AI

a. Melakukan penelitian dan pengembangan terkait kualitas, keandalan, dan metodologi pengembangan AI.

b. Berkontribusi pada pertumbuhan ekonomi yang berkelanjutan serta solusi untuk tantangan sosial.

c. Mempromosikan internasionalisasi, diversifikasi, dan kolaborasi antara industri, akademisi, dan pemerintah, termasuk mengikuti perkembangan diskusi global seperti DFFT *(Data Free Flow with Trust*) serta bergabung dengan komunitas pengembang dan organisasi akademik AI.

d. Menyediakan informasi tentang AI kepada seluruh masyarakat.

3. Hal yang terkait AI *providers* (penyedia AI)

   Penyedia AI harus menyediakan sistem dan layanan AI dengan memastikan bahwa AI dioperasikan dan digunakan dengan benar.

a. Saat Menerapkan Sistem AI

1. Tindakan untuk mengurangi risiko terhadap manusia dan lingkungan

   Mengambil langkah-langkah untuk memastikan kinerja AI tetap optimal dalam berbagai kondisi penggunaan, serta meminimalkan risiko yang dapat timbul.

2. Menjamin penggunaan AI yang tepat

a. Menggunakan AI dalam cakupan penggunaan yang telah ditetapkan oleh pengembang AI.

b. Memastikan akurasi AI, serta mengevaluasi perbedaan antara lingkungan penggunaan AI yang diharapkan oleh pengembang dan kondisi aktual.

3. Memperhatikan bias dalam konfigurasi atau data AI

a. Memastikan keadilan data, serta meninjau bias yang mungkin ada dalam informasi referensi atau layanan eksternal yang digunakan AI.

b. Secara berkala mengevaluasi *input/output* AI dan alasan di balik keputusan yang dibuat oleh model AI untuk memantau kemungkinan bias yang dihasilkan.

c. Meneliti kemungkinan adanya bias yang dapat membatasi keputusan pengguna secara tidak adil berdasarkan hasil output AI.

4. Documentation of development information

   Compile documentation on the AI development process, data collection and labeling, algorithms used, and other related information.

5. Participate in AI innovation

a. Conduct research and development related to the quality, reliability, and methodologies of AI development.

b. Contribute to sustainable economic growth and solutions for social challenges .

c. Promote internationalization, diversification, and collaboration among industry, academia, and government, including keeping up with global discussions such as DFFT (Data Free Flow with Trust) and joining AI developer communities and academic organizations.

d. Provide information about AI to the general public.

3. Matters related to AI providers

   AI providers must deliver AI systems and services by ensuring that AI is operated and used correctly.

a. During AI System Implementation

1. Actions to reduce risks to humans and the environment

   Take steps to ensure that AI performance remains optimal under various usage conditions, while minimizing potential risks that may arise.

2. Ensure the appropriate AI use

a. Use AI within the scope of usage established by the AI developers.

b. Ensure the accuracy of AI, and evaluate the differences between the expected usage environment defined by the developers and the actual conditions.

3. Pay attention to bias in AI configuration or data

a. Ensure data fairness, and review any potential biases in the reference information or external services used by AI.

b. Periodically evaluate the AI's input/output and the reasoning behind the decisions made by the AI model to monitor for potential biases that may arise; and

c. Investigate the potential for biases that may unfairly limit user decisions based on the AI output results.

4. Menerapkan mekanisme perlindungan privasi

   Menerapkan langkah-langkah perlindungan privasi, seperti mekanisme yang mengelola dan membatasi akses ke data pribadi berdasarkan teknologi yang digunakan (*privacy by design*).

5. Menerapkan mekanisme keamanan

   Mengambil langkah-langkah keamanan berdasarkan karakteristik teknologi yang digunakan (*security by design*).

6. Dokumentasi arsitektur sistem

   Menyusun dokumen yang menjelaskan arsitektur sistem dan pemrosesan data yang memengaruhi pengambilan keputusan AI.

b. Setelah Sistem atau Layanan AI Mulai Digunakan

1. Menjamin penggunaan AI yang tepat

   Secara berkala memverifikasi apakah sistem atau layanan AI digunakan sesuai tujuan yang diharapkan.

2. Tindakan terhadap pelanggaran privasi

a. Mengumpulkan informasi tentang pelanggaran privasi dalam sistem atau layanan AI.

4. Implement privacy protection mechanisms

   Implement privacy protection measures, such as mechanisms that manage and restrict access to personal data based on the technology used (privacy by design).

5. Implement security mechanisms

   Take security measures based on the characteristics of the technology used (security by design).

6. Document System Architecture

   Compile a document that explains the system architecture and data processing that influence AI decision-making.

b. Post Deployment of AI System or Services

1. Ensuring proper AI use

   Periodically verify whether the AI system or service is utillized according to its intended purpose.

2. Actions regarding privacy violations

a. Collect information about privacy violations in AI systems or services.

b. Menangani pelanggaran privasi secara tepat serta mengambil langkah-langkah untuk mencegah terulangnya pelanggaran serupa.

3. Menangani kerentanan keamanan

   Mengidentifikasi tren terkini mengenai risiko dan titik rawan dalam setiap tahap penggunaan AI, serta berupaya menghilangkan kerentanan yang ada.

4. Memberikan informasi kepada pemangku kepentingan

a. Memberikan informasi mengenai keamanan AI, termasuk karakteristik teknis sistem AI, mekanisme perlindungan, risiko yang dapat diprediksi dan cara menanganinya, kemungkinan perubahan dalam *output* atau program, penyebab kegagalan dan langkah-langkah perbaikannya, kebijakan terkait pengumpulan dan pembelajaran data oleh model AI.

b. Menjelaskan fakta bahwa AI digunakan, metode penggunaan yang tepat/tidak tepat, serta alasan dan detail pembaruan AI berdasarkan sifat dan tujuan penggunaannya.

5. Menjelaskan kepatuhan terhadap prinsip panduan umum kepada pengguna bisnis AI

b. Address privacy violations appropriately and take steps to prevent similar violations from recurring.

3. Address security vulnerabilities

   Identify current trends regarding risks and vulnerabilities at each stage of AI usage, and strive to eliminate existing vulnerabilities.

4. Provide information to stakeholders

a. Provide information regarding AI security, including the technical characteristics of AI systems, protection mechanisms, predictable risks and how to address them, potential changes in output or programs, causes of failures and corrective measures, and policies related to data collection and learning by AI models.

b. Explain the fact that AI is used, the appropriate/inappropriate methods of use, as well as the reasons and details of AI updates based on the nature and purpose of its use.

5. Explain compliance with general guiding principles to AI business users

Mendorong pengguna bisnis AI untuk menggunakan AI dengan benar, serta memperhatikan akurasi dan keterbaruan data yang digunakan untuk pembelajaran AI, risiko pembelajaran model AI yang tidak sesuai saat menggunakan *in-context learning*, peringatan terkait input data pribadi yang tidak sesuai ke dalam sistem AI.

6. Dokumentasi perjanjian layanan

   Menyusun dan menyediakan perjanjian layanan untuk pengguna AI serta menyajikan kebijakan privasi terkait penggunaan AI.

4. Hal-Hal terkait AI *business users* (pengguna bisnis AI)

   Penting bagi pengguna bisnis AI untuk selalu menggunakan AI dalam cakupan penggunaan yang ditetapkan oleh penyedia AI. Selain itu, mereka diharapkan mengoperasikan sistem AI sesuai kebutuhan serta mempelajari aspek yang diperlukan untuk menggunakan AI secara lebih efektif. Aspek yang perlu menjadi perhatian saat menggunakan sistem dan layanan AI:

Encourage AI business users to use AI correctly, while paying attention to the accuracy and freshness of the data used for AI learning, the risks of inappropriate AI model learning when using in-context learning, and warnings regarding the inappropriate input of personal data into AI systems.

6. Documentation of service agreements

   Compile and provide service agreements for AI users and present privacy policies related to the use of AI.

4. Matters related to AI business users

   It is important for AI business users to always use AI within the scope of use established by the AI provider. Additionally, they are expected to operate the AI system according to their needs and to learn the necessary aspects to use AI more effectively. Aspects that need to be considered when using AI systems and services include:

a. Penggunaan AI yang tepat dengan memperhatikan keamanan
   1. Mematuhi petunjuk penggunaan yang ditetapkan oleh penyedia AI serta menggunakan AI dalam cakupan penggunaan yang telah ditentukan selama proses desain.
   2. Memahami tingkat akurasi dan risiko *output* AI, serta memastikan berbagai faktor risiko sebelum menggunakan output AI.

b. Memperhatikan bias dalam data *input* atau *prompt*

   Memasukkan data yang telah terjamin keadilannya, memperhatikan bias dalam *prompt*, serta bertanggung jawab dalam menentukan apakah hasil output AI akan digunakan dalam bisnis.

c. Tindakan terhadap *input* data pribadi yang tidak tepat dan pelanggaran privasi
   1. Menghindari memasukkan data pribadi secara tidak tepat ke dalam sistem dan layanan AI.
   2. Mengumpulkan informasi tentang pelanggaran privasi dalam sistem dan layanan AI, serta mengambil langkah-langkah yang diperlukan untuk mencegah terjadinya pelanggaran.

a. Proper use of AI with attention to security
   1. Compliance with the usage instructions set by the AI provider and use AI within the scope of use defined during the design process.
   2. Understand the accuracy level and risks of AI output, and ensure various risk factors are considered before using AI output.

b. Attention to biases in input data or prompts

   Inputing data that has been ensured for fairness, pay attention to biases in prompts, and be responsible in determining whether the AI output will be used in business.

c. Actions regarding inappropriate input of personal data and privacy violations
   1. Avoid entering personal data inappropriately into AI systems and services.
   2. Collect information about privacy violations in AI systems and services, and take the necessary steps to prevent such violations from occurring.

d. Pelaksanaan langkah-langkah keamanan

Mematuhi instruksi keamanan yang ditetapkan oleh penyedia AI.

e. Menyediakan informasi kepada pemangku kepentingan

1. Menggunakan data input yang telah terjamin keadilannya dan memperhatikan bias dalam *prompt* saat memperoleh *output* AI.
2. Jika hasil *output* digunakan untuk pengambilan keputusan bisnis, informasi terkait harus disampaikan kepada pemangku kepentingan.

f. Memberikan penjelasan kepada pemangku kepentingan

1. Menyediakan informasi yang jelas dan mudah diakses bagi pemangku kepentingan terkait, termasuk cara menyediakan data dan formatnya berdasarkan karakteristik dan tujuan penggunaan AI, kebijakan privasi, cara menghubungi pihak terkait.

d. Implementation of security measures

Compliance with the security instructions set by the AI provider.

e. Provide information to stakeholders

1. Use input data that has been ensured for fairness and pay attention to biases in prompts when obtaining AI output.
2. If the output results are used for business decision-making, relevant information must be communicated to stakeholders.

f. Provide explanation to stakeholders

1. Provide clear and easily accessible information for relevant stakholders, including how to provide data and its format based on the characteristics and purpose of AI usage, privacy policies, and how to contact relevant parties.

2. Jika hasil output AI akan digunakan sebagai referensi dalam evaluasi individu atau kelompok tertentu, maka keputusan akhir harus dibuat oleh manusia secara rasional.
3. Menyediakan *help desk* untuk menangani pertanyaan dari pemangku kepentingan, memberikan penjelasan, dan menerima permintaan dengan bekerja sama dengan penyedia AI.

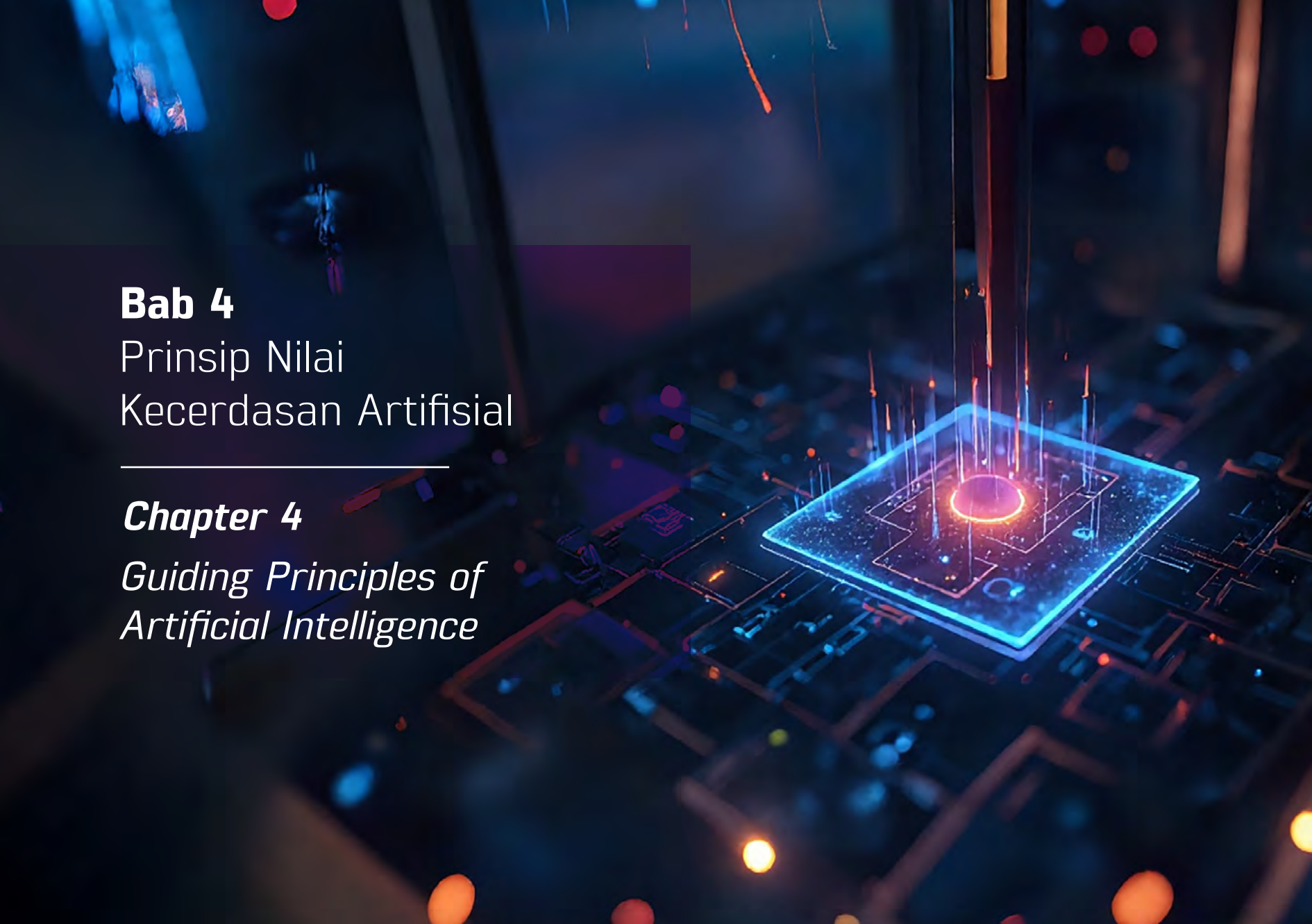
g. Pemanfaatan dokumen dan kepatuhan terhadap perjanjian

1. Menyimpan dan menggunakan dokumen terkait sistem dan layanan AI yang disediakan oleh penyedia AI dengan benar.
2. Mematuhi perjanjian layanan yang ditetapkan oleh penyedia AI.

2. If the AI output will be used as a reference in the evaluation of specific individuals or groups, the final decision must be made rationally by a human.
3. Provide a help desk to handle inquiries from stakeholders, offer explanations, and accept requests in collaboration with the AI provider.

g. Utilization of documents and compliance with agreements

1. Proper storage and utilization of documentation related to the AI system and services provided by the AI provider.
2. Compliance with the service agreements established by the AI provider.



Bab 4
Prinsip Nilai Kecerdasan Artifisial

Chapter 4
Guiding Principles of Artificial Intelligence

Nilai dasar untuk mencapai AI yang bertanggung jawab dan dapat dipercaya memuat nilai-nilai utama yang menjadi acuan bagi seluruh pemangku kepentingan dalam ekosistem AI. Nilai-nilai ini harus selaras dengan nilai yang berlaku secara nasional di negara atau masyarakat dan prinsip umum yang berlaku secara internasional. Nilai-nilai ini bersifat setara satu sama lain dan diperlukan agar bank dapat mengimplementasikan AI secara memadai. Implementasi AI yang bertanggung jawab dan dapat dipercaya tidak hanya tercermin dari adopsi dan kepatuhan terhadap beberapa nilai, tetapi harus dapat menginternalisasi semua nilai sebagai kesatuan utuh yang komprehensif.

The fundamental principles for promoting responsible and trustworthy AI encompass core values that provide a guiding framework for all stakeholders within the ecosystem. These values must align with nationally applicable values and comply with internationally accepted general principles. These values are equal to one another and are necessary for the banks to adequately implement AI. The implementation of responsible and trustworthy AI is not only reflected in the adoption and compliance with certain values but must also internalize all values as a comprehensive whole.

Gambar 9. Prinsip Dasar Kecerdasan Artifisial yang Bertanggung Jawab dan Dapat Dipercaya

Figure 9. Basic Principles of Responsible and Trustworthy Artificial Intelligence

A. Nilai-Nilai Utama untuk Mewujudkan Kecerdasan Artifisial yang Bertanggung-jawab dan Dapat Dipercaya

1. *Reliability* (Keandalan)

Keandalan sangat penting untuk memastikan bahwa keputusan yang dihasilkan oleh AI dapat diandalkan sesuai strategi yang dilakukan bank untuk mencapai tujuan. Model AI yang andal adalah yang mampu menghasilkan output melalui sistem yang dapat dijelaskan (*explainable*) dan mudah dipahami, aman (*secure*), serta dihasilkan oleh sistem yang andal dan tangguh (*resilient*).

a. *Explainable* (Dapat Dijelaskan), *Output* yang dihasilkan oleh sistem AI dapat diverifikasi dan dijustifikasi. Pemilik data, dalam hal ini adalah nasabah, berhak memahami bagaimana data mereka digunakan dan bagaimana sistem AI

A. Core Values to Realize Responsible and Trustworthy Artificial Intelligence

1. Reliability

Reliability is crucial to ensure that decisions made by AI are dependable according to the strategy implemented by Banks to achieve their objective. A reliable AI model is one that can produce output through an explainable and easily understandable system, is secure, and is generated by a robust and resilient system.

a. Explainable, the output generated by the AI systems can be verified and justified. Data owners, in this case customers, have the right to understand how their data is used and how the AI system makes decisions.

mengambil keputusan. Hal ini mencakup penyediaan informasi yang jelas dan mudah dipahami, sehingga pemilik data dapat menguji hasil tersebut. Namun, dalam beberapa kasus, persyaratan eksplainabilitas dapat berdampak negatif pada akurasi dan kinerja sistem AI. Penyederhanaan variabel (agar lebih mudah dipahami) untuk mengatasi masalah yang kompleks atau berdimensi tinggi dapat menghasilkan *output* yang kurang optimal. Oleh karena itu, ketika penyedia sistem AI memberikan penjelasan mengenai suatu *output*, penyedia sistem AI dapat mempertimbangkan untuk menyajikan–dengan cara yang jelas, sederhana, dan sesuai dengan konteks–faktor yang menentukan, data, logika, atau algoritma di balik *output* tertentu, atau menjelaskan mengapa situasi yang tampak serupa menghasilkan hasil yang berbeda. Hal ini harus dilakukan dengan cara yang memungkinkan nasabah memahami dan menguji hasil tersebut, sambil tetap mematuhi kewajiban perlindungan data pribadi.

This includes providing clear and easily understandable information so that data owners can test the results. However, in some cases, explainability requirements may negatively impact the accuracy and performance of AI systems. Simplifying variables (to make them easier to understand) to address complex or high-dimensional issues may result in suboptimal output. Therefore, when AI system providers offer explanations regarding an output, they should consider presenting–clearly, simply, and contextually–the determining factors, data, logic, or algorithms behind a specific output or explaining why seemingly similar situations yield different results. This must be done in a way that allows customers to understand and test the results while still complying with personal data protection obligations.

b. *Security and Resilience* (Aman dan Tangguh), penggunaan AI dalam aktivitas sehari-hari selain memberikan kemudahan juga memperluas ruang yang dapat menjadi celah bagi pelaku serangan siber. Oleh karena itu, penyelenggaraan sistem AI wajib dilengkapi dengan strategi mitigasi agar memliki sistem keamanan dan ketahanan yang baik sehingga sistem AI dapat dipercaya oleh konsumen. Keamanan mencakup protokol untuk menghindari,

b. Security and Resilience, the use of AI in daily activities not only provides convenience but also expands the space that can become a gap for cyber attackers. Therefore, the implementation of AI systems must be equipped with mitigation strategies to ensure good security and resilience so that the AI system can be trusted by consumers. Security includes protocols to avoid, protect against,

melindungi, merespons, atau pulih dari serangan. Ketahanan berkaitan dengan ketangguhan sistem, termasuk data, dalam hal terjadi penyalahgunaan model atau data.

Strategi keamanan AI yang tangguh harus mencakup serangkaian kerangka tata kelola yang kuat dan komprehensif. Hal ini termasuk menetapkan kebijakan yang jelas dan *feasible*, memperhatikan kecukupan dari sisi sumber daya, serta melakukan peninjauan dan pemeliharaan rutin terhadap sistem AI yang digunakan untuk memastikan model yang digunakan tetap sesuai dengan tujuan dan tidak usang serta rentan. Selain itu, kebijakan dalam pengelolaan data juga berperan penting. Pelindungan data yang menyeluruh, keamanan yang kuat, pemantauan berkelanjutan, pengujian rutin, praktik pengumpulan data yang bertanggung jawab, serta mitigasi bias algoritma merupakan elemen penting dalam membangun sistem AI

respond to, or recover from attacks. Resilience relates to the robustness of the system, including data, in cases of model or data misuse.

A robust AI security strategy should encompass a comprehensive and strong governance framework. This includes establishing clear and feasible policies, considering resource adequacy, as well as conducting regular reviews and maintenance of the AI systems used to ensure that the models remain aligned with their objectives and are not outdated or vulnerable. Additionally, policies regarding data management also play a crucial role. Comprehensive data protection, strong security measures, continuous monitoring, routine testing, responsible data collection practices, and algorithm bias mitigation are essential elements in building resilient AI systems. These elements ensure resistance to attacks,

yang tangguh. Elemen-elemen ini memastikan ketahanan terhadap serangan, kemampuan beradaptasi terhadap dinamika lingkungan, serta kinerja yang tetap optimal meskipun menghadapi gangguan atau disrupsi yang tidak terduga.

2. *Accountability* (Akuntabilitas)

Akuntabilitas dalam sistem AI adalah memastikan penyelenggara dapat mempertanggungjawabkan bahwa sistem AI yang dikembangkan dan dijalankan berfungsi dengan wajar. Sistem AI yang diselenggarakan dapat dipercaya, bermanfaat, adil, menghormati hak asasi manusia, transparan dan dapat dijelaskan, serta kuat dan aman. Keseluruhan proses yang dilakukan telah transparan (*transparency*), selain agar dapat memenuhi kebutuhan *explainable*, juga untuk kepentingan audit dan pertanggungjawaban, serta diharapkan pula berjalannya mekanisme pengelolaan data yang baik dan benar (*data privacy*). Prinsip ini memastikan bahwa sistem AI tidak hanya adil dan etis tetapi juga terdapat jalur tanggung jawab yang jelas untuk mengelola risiko dan mengatasi dampak negatif dari penggunaan AI, sehingga dapat meningkatkan kepercayaan dan keyakinan nasabah dan pemangku kepentingan.

adaptability to environmental dynamics, and optimal performance even when facing unexpected disruptions or disturbances.

2. Accountability

Accountability in AI systems ensures that the organizers can be held responsible for the proper functioning of the developed and operated AI systems. The AI systems provided should be trustworthy, beneficial, fair, respect human rights, transparent and explainable, as well as robust and secure. The entire process carried out has been transparent (transparency), not only to meet the requirement of being of explainable but also for audit and accountability purposes, while also ensuring that good and proper data management mechanisms (data privacy) are in place. This principle ensures that AI systems are not only fair and ethical but also have clear accountability pathways to manage risks and address negative impacts from AI usage, thereby enhancing trust and confidence among customers and stakeholders.

a. *Transparency* (Transparansi), istilah transparansi dalam konsep AI dapat mencakup beberapa konteks makna yaitu transparansi berupa praktik pengungkapan keterlibatan AI dalam sebuah mekanisme (dengan output yang dapat berupa prediksi, rekomendasi atau keputusan, dan lain-lain). Di sisi lain, transparansi juga dapat bermakna memberikan kejelasan dan keterbukaan tentang bagaimana sebuah sistem AI dikembangkan, dioperasikan hingga menghasilkan sebuah output, dan bagaimana sistem tersebut memproses data. Hal ini merupakan landasan penerapan AI yang bertanggung jawab, dengan memastikan bahwa sistem AI dapat dipahami, akuntabel, dan selaras dengan nilai-nilai etika dan kemasyarakatan. Interpretabilitas ini juga dapat memungkinkan pengguna untuk memahami arsitektur model, fitur-fitur yang digunakan serta bagaimana sistem AI tersebut menggabungkannya untuk menghasilkan output yang mungkin memengaruhi keputusan yang berkaitan dengan diri seseorang serta dampak yang dapat ditimbulkannya. Tujuan prinsip ini ialah agar bank tetap mengedepankan pertanggungjawaban dan komitmen atas *disclosure* mengenai sistem AI yang telah dilakukan. Penyediaan informasi yang bermanfaat, sesuai dengan konteks, serta bersifat *real-time* dengan maksud menumbuhkan pemahaman umum mengenai kemampuan dan keterbatasan atas sistem AI yang digunakan, serta memastikan tersedianya informasi yang jelas dan mudah dipahami mengenai

a. Transparency, the term transparency in the context of AI can encompass several meanings, including transparency as a practice of disclosing AI involvement in a mechanism (with outputs that may include predictions, recommendations, or decisions, among others). On the other hand, transparency can also mean providing clarity and openness about how an AI system is developed, operated to produce an output, and how the system processes data. This serves as a foundation for the responsible application of AI by ensuring that AI systems are understandable, accountable, and aligned with ethical and societal values. This interpretability also allows users to understand the model architecture, the features used, and how these features are combined by the AI system to generate outputs that may influence decisions related to individuals and their potential impacts. The goal of this principle is for banks to prioritize accountability and commitment regarding disclosures about their implemented AI systems. Providing useful information that is contextually relevant and real-time aims to foster a general understanding of both the capabilities and limitations of the utilized AI systems while ensuring clear and easily understandable information regarding data sources/inputs, factors involved, and process stages leading to predictions, recommendations or decisions for

sumber data/input, faktor, tahapan proses yang menghasilkan prediksi, rekomendasi, atau keputusan untuk para pihak yang terdampak oleh sebuah sistem AI. Terdapat kewajiban transparansi tambahan untuk jenis AI tertentu. Misalnya:

1. Sistem AI yang dimaksudkan untuk berinteraksi langsung dengan individu harus dirancang untuk memberi tahu pengguna bahwa mereka sedang berinteraksi dengan sistem AI, kecuali jika hal ini terlihat jelas oleh individu dari konteksnya. *Chatbot*, misalnya, harus dirancang untuk memberi tahu pengguna bahwa itu adalah *chatbot*.
2. Sistem AI yang menghasilkan teks, gambar, atau konten tertentu lainnya harus menggunakan format yang dapat dibaca mesin untuk menandai *output* dihasilkan atau dimanipulasi oleh AI. Hal ini mencakup, misalnya, AI yang menghasilkan *deepfake*–gambar atau video yang diubah untuk menunjukkan bahwa seseorang melakukan atau mengatakan sesuatu yang tidak mereka lakukan atau katakan.

   Dengan demikian, bank perlu memastikan komunikasi yang jelas dan terbuka mengenai bagaimana sistem AI digunakan, bagaimana sebuah *output* dihasilkan, serta bagaimana data diproses. Transparansi juga memastikan bahwa nasabah dan para pemangku kepentingan memahami peran AI dalam operasional perbankan, serta dapat mempercayai sistem AI yang digunakan.

parties affected by an AI system. There are additional transparency obligations for certain types of AIs. For example:

1. AI systems intended to interact directly with individuals must be designed to inform users that they are interacting with an AI system, unless this is clearly evident to the individual from the context. Chatbots, for example, should be designed to inform users that they are chatbots.
2. AI systems that generate text, images, or other specific content must use machine-readable formats to label outputs as generated or manipulated by AI. This includes, for example, AI that produces deepfakes–images or videos altered to show someone doing or saying something they did not do or say.

   Thus, the bank needs to ensure clear and open communication regarding how AI systems are used, how an output is generated, and how data is processed. Transparency also ensures that customers and stakeholders understand the role of AI in banking operations and can trust the AI systems being used.

b. Data Privacy (Pelindungan Data), data memiliki peran yang sangat penting dalam seluruh tahapan *lifecycle* AI, mulai dari tahap perencanaan, pengembangan hingga *deployment* yang hasilnya akan dipengaruhi dari kualitas *training data* yang digunakan. Pengelolaan data dalam konteks kerangka sistem AI mengacu pada proses dan mekanisme yang digunakan untuk mengumpulkan, menyimpan, mengatur, dan menyiapkan data dalam jumlah besar yang secara khusus untuk digunakan melatih dan menjalankan sebuah sistem AI, memastikan bahwa data yang dikumpulkan dapat diakses, bersih, terintegrasi, dan diatur sesuai dengan karakteristik dan kebutuhan dari model AI yang akan dijalankan. Pengelolaan data dalam seluruh tahapan *lifecycle* AI ini krusial karena akan berpengaruh terhadap performa dan keakuratan *output* dari sebuah sistem AI. Praktik pengelolaan data yang tepat dapat membantu mengidentifikasi dan mengatasi potensi bias dalam kumpulan data, sehingga menghasilkan *output* AI yang obyektif. Privasi AI mengacu pada perlindungan informasi pribadi atau sensitif yang dikumpulkan, digunakan, dibagikan, atau disimpan oleh AI. Privasi AI terkait erat dengan privasi data.

Privasi data, juga dikenal sebagai privasi informasi, adalah prinsip bahwa seseorang harus memiliki kendali atas data pribadinya. Pengelolaan data dalam penggunaan AI juga wajib dilakukan

b. Data Privacy, Data plays a vital role in all stages of the AI lifecycle, from planning and development to deployment, with the results being influenced by the quality of the training data used. Data management in the context of an AI system framework refers to the processes and mechanisms used to collect, store, organize, and prepare large amounts of data specifically for training and running an AI system, ensuring that the collected data is accessible, clean, integrated, and organized according to the characteristics and needs of the AI model being deployed. Data management throughout all stages of the AI lifecycle is crucial as it affects performance and accuracy of outputs from an AI system. Proper data management practices can help identify and address potential biases in datasets, resulting in objective AI outputs. AI privacy refers to protecting personal or sensitive information collected, used, shared, or stored by AI. AI privacy is closely related to data privacy.

Data privacy, also known as information privacy, is based on the principle that individuals should have control over their personal data. Data management in using AI must also comply with

dengan patuh terhadap ketentuan terkait pelindungan data yang ada serta hanya menggunakan data untuk tujuan yang telah dinyatakan dan disepakati oleh pemilik. Selain itu, pemilik data dalam hal ini adalah konsumen wajib diberikan tingkat kontrol yang diperlukan atas data mereka, termasuk pilihan untuk ikut serta atau tidak mengikutsertakan data mereka serta menyediakan media atau saluran pengaduan dalam hal terjadi kekhawatiran konsumen terhadap keamanan data yang digunakan. Indonesia saat ini telah menerbitkan Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi (UU PDP) yang dapat digunakan sebagai pedoman bagi Bank menyusun kebijakan pengelolaan data pribadi nasabah yang digunakan dalam sistem AI. Mengelola data pribadi sesuai UU PDP di Indonesia mencakup beberapa rangkaian langkah, antara lain:

existing regulations regarding data protection while only using data for purposes that have been stated and agreed upon by its owners. Additionally, in this case where owners are consumers must be provided with necessary levels of control over their data including options to opt-in or opt-out regarding their participation as well as providing media or channels for complaints if there are consumer concerns about security related to their utilized data. Indonesia has currently enacted Law Number 27 Year 2022 concerning Personal Data Protection (PDP Law) which can serve as a guideline for banks in formulating policies on managing customers' personal data used within AI systems. Managing personal data according to PDP Law in Indonesia includes several steps, such as:

1. Mulai tahap awal identifikasi dan kategorisasi jenis data yang akan dikumpul dan diolah;
2. Memastikan nasabah pemilik data paham dan menyetujui tujuan dan pengelolaan yang akan dilakukan terhadap data pribadi mereka, serta penggunaan data sesuai dengan tujuan awal;
3. Memastikan pengolahan data dilakukan dengan prinsip yang ditetapkan antara lain keterbukaan, keabsahan, kualitas data, dan keamanan data;
4. Memastikan keamanan terhadap data pribadi yang dikumpulkan;
5. Memfasilitasi hak-hak pemilik data terhadap data mereka (dalam hal ini hak mengakses, memperbaiki, dan menghapus data pribadi mereka);
6. Melakukan dokumentasi atas seluruh proses tahapan pengolahan data;
7. Memiliki kebijakan/SOP dalam hal terjadi keadaan darurat;

1. Etarting with the initial stage of identifying and categorizing the types of data to be collected and processed;
2. Ensuring that customers, as data owners, understand and agree to the purposes and management of their personal data, as well as the use of data in accordance with the initial objectives;
3. Ensuring that data processing is carried out according to established principles, including transparency, legality, data quality, and data security;
4. Ensuring the security of the collected personal data;
5. Facilitating the rights of data owners regarding their data (in this case, the rights to access, correct, and delete their personal data);
6. Documenting all processes involved in data processing stages;
7. Having policies/SOPs in place for emergency situations;

8. Secara berkala melakukan pengembangan kompetensi SDM; dan

9. Melakukan evaluasi berkala terhadap kebijakan pengolahan data yang telah dijalankan dan tindaklanjutnya.

Selain itu, bank juga dapat melakukan langkah seperti menjaga agar semua informasi yang digunakan untuk mengembangkan, melatih, menerapkan, mengelola, dan mengatur sistem AI bersifat pribadi dan aman. Hal ini termasuk ketika vendor pihak ketiga dan mitra bisnis terlibat. Pastikan data yang digunakan untuk melatih dan mengembangkan sistem AI didasarkan pada standar data yang baik dan telah dianalisis secara menyeluruh serta disesuaikan untuk mengetahui adanya bias, data buruk, dan data hilang.

3. *Human Oversight* (Pengawasan Manusia)

Human Oversight atau Pengawasan Manusia menjadi nilai krusial dalam mewujudkan *"Trustworthy AI"* atau AI yang dapat diandalkan. Dalam hal ini proses adopsi sistem AI dalam seluruh proses tahapan AI (*AI Lifecycle*) tetap membutuhkan intervensi, kontrol, dan pertimbangan dari manusia baik dari mulai *input* hingga *output*. Intervensi manusia dilakukan secara tepat waktu dan tepat sasaran sebagai langkah antisipasi jika sistem AI membuat keputusan sendiri, terjadinya bias yang potensial dan tidak disadari, serta untuk mencegah AI menghasilkan hasil yang tidak sesuai dengan nilai keadilan, tidak sesuai tujuan, melanggar ketentuan, dan tidak sesuai dengan etika kemanusiaan. Dalam dokumen *The White House Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* (2022), konsep intervensi manusia pada adopsi AI dijelaskan secara lebih spesifik menjadi beberapa langkah yaitu:

1. Menyediakan mekanisme untuk memilih keluar dari sistem AI untuk digantikan dengan alternatif manusia

Alternatif manusia tersedia bila diperlukan **Availability of human alternatives**	Masyarakat/konsumen yang terdampak oleh AI harus diberi pemberitahuan singkat dan jelas bahwa mereka berhak memilih untuk tidak diproses menggunakan AI, beserta petunjuk yang jelas tentang bagaimana jika memilih tidak menggunakan AI. Affected individuals/customers must be given a brief and clear notification that they have the right to opt out of being processed using AI, along with clear instructions on how to choose not to use AI.
Alternatif manusia yang tepat waktu dan tidak memberatkan **Timely and effortless human alternatives**	Pilihan untuk tidak menggunakan sistem AI harus tepat waktu dan tidak membebani secara tidak wajar. The option to opt out of using the AI system must be timely and not impose an unreasonable burden.
Pemberitahuan dan instruksi yang singkat, jelas, dan mudah diakses **Brief, clear, and easy to access notifications and instructions**	Petunjuk keluar dari sistem AI harus diberikan dalam bentuk yang mudah diakses dan mudah ditemukan oleh pihak yang terdampak oleh sistem AI. Instructions for opting out of the AI system must be provided in a format that is easily accessible and readily discoverable by those affected by the AI system.

8. Periodically developing human resources competency; and

9. Conducting regular evaluations of the data processing policies that have been implemented and their follow-up actions.

In addition, the bank can also take steps such as ensuring that all information used to develop, train, deploy, manage, and regulate AI systems remains private and secure. This includes situations where third-party vendors and business partners are involved. Ensure that the data used to train and develop AI systems is based on good data standards and has been thoroughly analyzed while being adjusted to identify any biases, poor data, and missing data.

3. Human Oversight

Human oversight is a crucial value in realizing "Trustworthy AI" or reliable AI. In this regard, the adoption process of AI systems throughout the entire AI lifecycle still requires human intervention, control, and consideration from input to output. Human intervention should be conducted in a timely and targeted manner as a precautionary measure in case the AI system makes decisions independently, potential biases arise unconsciously, and to prevent AI from producing outcomes that are inconsistent with values of fairness, deviate from objectives, violate regulations, and are not aligned with humanitarian ethics. In the document The White House Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People (2022), the concept of human intervention in AI adoption is explained more specifically into several steps, namely:

1. Providing mechanisms to opt out of AI systems in favor of human alternatives

2. Memberikan pertimbangan manusia yang tepat waktu dan perbaikan melalui sistem *fallback* (rencana alternatif yang dapat digunakan dalam keadaan darurat) dan eskalasi jika sistem AI mengalami kegagalan. Pertimbangan manusia/sistem *fallback* tersebut perlu memenuhi asas berikut:

2. Providing timely human consideration and remediation through a fallback system (an alternative plan that can be deployed in emergencies) and escalation if the AI system fails. The human considerations/ fallback systems need to adhere to the following principles:

Proporsional	Ketersediaan pertimbangan manusia dan *fallback* harus proporsional dengan potensi AI.
Proportional	The availability of human consideration and fallback must be proportional to the potential impact of the AI.
Aksesibel	Mekanisme untuk pertimbangan manusia dan *fallback* harus mudah ditemukan.
Easy to Access	Mechanisms for human consideration and fallback must be easily found.
Nyaman	Mekanisme untuk pertimbangan manusia dan *fallback* tidak boleh terlalu memberatkan dibandingkan dengan AI yang setara.
Convenient	Mechanisms for human consideration and fallback should not be overly burdensome compared to equivalent AI systems.
Adil	Pertimbangan harus diberikan untuk memastikan hasil dari *fallback* dan sistem eskalasi dilakukan dengan adil.
Fair	Consideration must be given to ensure that the outcomes from fallback and escalation systems are carried out fairly.
Tepat waktu	Pertimbangan manusia dan *fallback* hanya berguna jika dilakukan dan diselesaikan tepat waktu.
Timely	Human consideration and fallback are only effective if they are executed and resolved in a timely manner.
Efektif	Struktur organisasi yang melingkupi proses pertimbangan dan *fallback* harus dirancang sedemikian rupa sehingga jika manusia yang berwenang mengambil keputusan memutuskan bahwa keputusan tersebut harus dibatalkan, keputusan baru akan diberlakukan secara efektif.
Effective	The organizational structure encompassing the consideration and fallback processes must be designed in such a way that if an authorized human decision-maker determines that a decision should be revoked, a new decision will be implemented effectively.
Terpelihara	Proses pertimbangan manusia dan *fallback* dan setiap proses otomatis terkait harus dipertahankan dan didukung selama AI yang relevan terus digunakan.
Maintained	The human consideration and fallback processes, along with any related automated processes, must be maintained and supported as long as the relevant AI is in use.

3. Menetapkan pelatihan, penilaian, dan pengawasan untuk mencegah bias otomasi pada sistem AI dan memastikan semua komponen sistem berbasis manusia berjalan efektif

3. Establishing training, assessment, and supervision to prevent automation bias in AI systems and ensuring that all human-based system components operate effectively

Pelatihan dan penilaian	Siapa pun yang mengelola, melakukan interaksi, atau menafsirkan *output* AI harus menerima pelatihan terkait sistem tersebut.
Training and Assessment	Anyone who manages, interacts with, or interprets AI outputs must receive training related to the system.
Pengawasan	Sistem berbasis manusia berpotensi menimbulkan bias. Hasil penilaian dari efikasi (efektivitas) dan potensi bias harus diawasi oleh struktur tata kelola untuk memperbarui pengoperasian sistem berbasis manusia guna mengurangi dampak tersebut.
Supervision	Human-based systems have the potential to introduce bias. The results of efficacy (effectiveness) assessments and potential biases must be overseen by a governance structure to update the operation of human-based systems in order to mitigate such impacts.

4. Menerapkan tambahan pengawasan dan perlindungan manusia untuk AI yang terkait dengan domain sensitif. Beberapa aspek yang perlu diperhatikan antara lain:

4. The implementation of additional human oversight and protections for AI is tied to sensitive domains. Several aspects that need to be considered include:

Data dengan cakupan yang sempit **Data with limited scope**	Pengawasan manusia harus memastikan bahwa AI dalam domain sensitif memiliki cakupan yang sempit untuk mencapai tujuan yang ditentukan. Human oversight must ensure that AI in sensitive domains has a limited scope to achieve the specified objectives.
Disesuaikan dengan situasi **Adjusted to the situation**	Pengawasan manusia harus memastikan bahwa AI dalam domain sensitif disesuaikan dengan *use-case* spesifik dan skenario penerapan di dunia nyata. Human oversight must ensure that AI in sensitive domains is tailored to specific use cases and real-world application scenarios.
Pertimbangan manusia sebelum mengambil keputusan berisiko tinggi **Human consideration before making high-risk decisions**	Sistem AI tidak boleh diizinkan untuk campur tangan secara langsung dalam situasi berisiko tinggi, tanpa pertimbangan manusia. AI systems must not be allowed to intervene directly in high-risk situations without human consideration.
Akses untuk memeriksa sistem **Access to inspect the system**	Perancang, pengembang, dan pengguna AI harus mempertimbangkan pengabaian kerahasiaan terbatas yang diperlukan untuk memberikan pengawasan terhadap sistem yang digunakan dalam domain sensitif. Designers, developers, and users of AI must consider the limited confidentiality waiver necessary to provide oversight of systems used in sensitive domains.

5. Menunjukkan akses ke alternatif manusia, pertimbangan, dan *fallback*

5. Demonstrating access to human alternatives, considerations, and fallback

Pelaporan **Reporting**	Pelaporan tentang aksesibilitas, ketepatan waktu, dan efektivitas pertimbangan manusia dan *fallback* harus dipublikasikan secara berkala selama sistem tersebut digunakan. Reporting on the accessibility, timeliness, and effectiveness of human consideration and fallback must be published regularly while the system is in use.

Dalam panduan ini, nilai *Human Oversight* tidak hanya meliputi intervensi dan pengawasan manusia atas sistem AI tetapi juga mencakup bagaimana memastikan sistem AI memenuhi prinsip *ethical & fairness* (etika dan keadilan), *sustainable* (keberlanjutan), dan *inclusive* (inklusif).

In this guideline, the value of Human Oversight not only includes human intervention and oversight of AI systems but also encompasses how to ensure that AI systems adhere to the principles of ethics and fairness, sustainable, and inclusive.

a. *Inclusivity* (Inklusif), prinsip ini menegaskan bahwa pengembangan sistem AI harus dapat mengakomodasi berbagai karakter dan keragaman masyarakat. Sistem AI memiliki potensi yang besar untuk menghasilkan keputusan yang bias. Hal ini bisa dicegah jika pengembangan sistem AI telah mempertimbangkan faktor inklusivitas di setiap siklus pengembangannya. Sistem AI tidak boleh menolak masyarakat yang memenuhi syarat berdasarkan identitas mereka. Sistem AI tidak boleh memperdalam perpecahan historis dan sosial yang merugikan berdasarkan agama, ras, kasta, jenis kelamin,

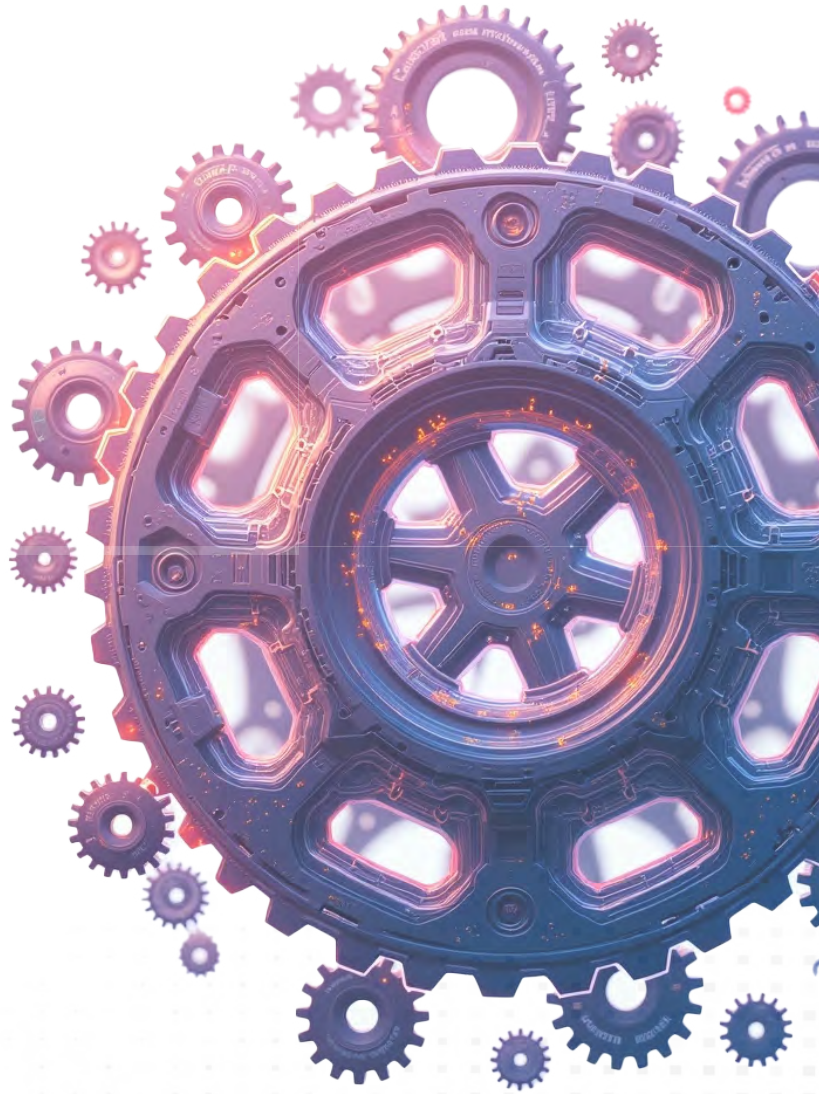
a. Inclusivity, these principles emphasize that the development of AI systems must accommodate various characters and the diversity of society. AI systems have significant potential to produce biased decisions. This can be prevented if the development of AI systems has considered inclusivity factors at every stage of their development. AI systems must not exclude qualified individuals based on their identity. They should not exacerbate historical and social divides that harm individuals based on religion, race, caste, gender, descent, place of birth or residence in terms of education,

keturunan, tempat lahir atau tempat tinggal dalam hal pendidikan, pekerjaan, akses ke ruang publik, dan lainnya. Dalam konteks perbankan, bank perlu memastikan bahwa sistem AI yang digunakan tidak melakukan pengecualian layanan atau manfaat yang tidak adil. Jika terjadi keputusan yang merugikan nasabah atau calon nasabah, mekanisme penyelesaian keluhan yang tepat harus dirancang dengan cara yang mudah dan dapat diakses oleh nasabah atau calon nasabah tersebut.

b. *Sustainability* (Keberlanjutan), pengembangan dan penggunaan sistem AI bagi manusia harus bertujuan untuk mencapai kesejahteraan dan memberikan manfaat bagi masyarakat. *Trustworthy AI* berperan penting dalam meningkatkan pertumbuhan yang inklusif dan berkelanjutan baik bagi *stakeholders* maupun masyarakat. Dalam konteks yang lebih luas, pengembangan AI dapat secara substansial berkontribusi dalam pencapaian *Sustainable Development Goals* (SDGs). Hal ini menunjukkan bahwa penggunaan AI diharapkan dapat memberikan keadilan bagi semua lapisan masyarakat, termasuk kelompok masyarakat rentan (*vulnerable society*) dan kurang terwakili (*underrepresented population*), seperti etnis minoritas, wanita, anak-anak, lansia, dan masyarakat kurang berpendidikan atau kurang terampil, termasuk masyarakat dengan disabilitas sehingga penggunaan AI dapat mewujudkan pertumbuhan yang berkelanjutan secara optimal, serta memberikan manfaat bagi keberlangsungan usaha bagi institusi atau lembaga yang mengadopsi sistem AI dalam kegiatan usahanya.

employment, access to public spaces, and more. In the context of banking, banks need to ensure that the AI systems used do not unfairly exclude services or benefits. If decisions are made that adversely affect customers or potential customers, appropriate complaint resolution mechanisms must be designed in a way that is easy and accessible for those customers or potential customers.

b. Sustainability, The development and use of AI systems for humans must aim to achieve well-being and provide benefits to society. Trustworthy AI plays a crucial role in enhancing inclusive and sustainable growth for both stakeholders and the community. In a broader context, AI development can substantially contribute to achieving the Sustainable Development Goals (SDGs). This indicates that the use of AI is expected to deliver justice for all segments of society, including vulnerable groups and underrepresented populations, such as ethnic minorities, women, children, the elderly, and less educated or less skilled individuals, including those with disabilities such that the use of AI can optimally realize sustainable growth while providing benefits for the continuity of businesses for institutions or organizations that adopt AI systems in their operations.



Dalam ruang lingkup perbankan, jika bank mengadopsi sistem AI dalam kegiatan bisnisnya, sistem AI harus dipantau secara berkala dalam setiap tahapan pengembangannya dan dinilai dampaknya terhadap keberlanjutan usaha bank dan target keberlanjutan lain termasuk memastikan bahwa sistem AI bertindak adil dalam memproses produk dan layanan bagi nasabah/calon nasabah yang beragam, memberikan manfaat jangka panjang baik bagi bank maupun nasabah, dan menjaga kepentingan *stakeholders* untuk mendukung kegiatan usaha bank yang terus berkembang.

Within the scope of banking sector, if a bank adopts AI systems in their business activities, these systems must be monitored regularly at every stage of their development and assessed for their impact on business sustainability goals as well as other sustainability targets which includes ensuring that AI systems act fairly in processing products and services for diverse customers or potential customers while providing long-term benefits both for banks and customers alike, and maintaining stakeholder interest to support the continously evolving banking operations.

c. *Ethics & Fairness* (Etika dan Keadilan), penggunaan AI harus selaras dengan standar etika, nilai, dan kode etik yang berlaku di masyarakat dan bank serta keputusan yang dibuat oleh AI harus memenuhi standar etika yang sama dengan keputusan yang dibuat oleh manusia (bertanggung jawab, adil, dan tidak melanggar nilai-nilai kemanusiaan), serta keputusan tersebut tidak boleh merugikan individu atau kelompok kecuali jika keputusan tersebut dapat dibenarkan dan dapat dijelaskan. Bank perlu memastikan bahwa sistem AI yang diadopsi telah menghilangkan bias dan diskriminasi yang tidak adil dan merugikan terhadap individu atau kelompok tertentu berdasarkan ras, jenis kelamin, asal negara, usia, pendapat politik, agama, dan sebagainya.

c. Ethics & Fairness, the use of AI must align with the ethical standards, values, and codes of conduct that are applicable in society and within the bank while decisions made by AI should meet the same ethical standards as those made by humans (responsible, fair, and not violating humanitarian values), and such decisions must not harm individuals or groups unless they are justifiable and explainable. Banks need to ensure that the adopted AI systems have eliminated unfair biases and discrimination against specific individuals or groups based on race, gender, nationality, age, political opinions, religion, and so forth.

B. Integrasi Prinsip Tata Kelola AI Perbankan Indonesia dengan Elemen Bisnis Perbankan

Penerapan prinsip-prinsip di atas bertujuan agar bank dapat memitigasi potensi risiko, membangun kepercayaan pemangku kepentingan, dan memanfaatkan potensi AI sepenuhnya dengan cara yang aman dan berkelanjutan. Dalam implementasinya, prinsip-prinsip tata kelola AI tersebut harus mengintegrasikan tiga elemen berikut:

a. *People* (Sumber Daya Manusia), merujuk pada pentingnya pengembangan keterampilan dan kompetensi yang tepat, seperti pelatihan keahlian dalam AI untuk pegawai bank dan untuk memastikan bahwa teknologi diterapkan dengan bijaksana.

b. *Process* (Proses), mengacu pada pengembangan dan penerapan prosedur serta kebijakan tata kelola yang jelas, termasuk manajemen risiko dan pengawasan yang berkelanjutan untuk memastikan kepatuhan terhadap standar etika dan regulasi.

c. *Technology* (Teknologi dan Infrastruktur), mencakup penerapan alat dan sistem AI yang andal, transparan, dan aman, dengan pemantauan yang terus-menerus untuk mendeteksi potensi risiko dan bias.

B. Integrating the AI Governance Principles of Indonesian Banking with Elements of the Banking Business

The purpose of implementing the principles above are for banks to mitigate potential risks, build stakeholder trust, and fully leverage the potential of AI in a safe and sustainable manner. In its implementation, the principles of AI governance must integrate the following three elements:

a. People, which refers to the importance of developing appropriate skills and competencies, such as training in AI expertise for bank employees, and ensuring that technology is applied wisely.

b. Process, which refers to the development and implementation of clear governance procedures and policies, including risk management and ongoing oversight to ensure compliance with ethical standards and regulations.

c. Technology, including the implementation of reliable, transparent, and secure AI tools and systems, with continuous monitoring to detect potential risks and biases.

Adapun integrasi prinsip nilai acuan di dalam masing-masing elemen adalah sebagai berikut.

The integration of benchmark value principles for each element is as follows.

Prinsip 1. *Reliability*

Principle No. 1 Reliability

Reliability **(Keandalan)**	1. ***People:*** Pegawai bank harus dilatih untuk memahami bagaimana bias dapat muncul dalam model AI dan mereka harus dilatih cara memastikan keandalan sistem AI. Pelatihan harus berfokus pada cara mendeteksi dan memitigasi bias dalam keputusan yang didorong oleh AI. 2. ***Process:*** Prosesnya harus mencakup audit rutin dan penilaian model AI untuk memastikan keandalannya. Setiap bias yang ditemukan selama audit harus diperbaiki untuk memastikan bahwa sistem tidak mendiskriminasi kelompok tertentu. 3. ***Technology:*** Sistem AI harus dirancang untuk memastikan keandalan dengan menggunakan data dan algoritma yang tidak memihak. Teknologi harus memungkinkan pemantauan berkelanjutan untuk keandalan sistem dan memungkinkan dilakukannya penyesuaian ketika bias terdeteksi. 1. **People:** Bank employees must be trained to understand how bias can arise in AI models, and they should be trained on how to ensure the reliability of AI systems. Training should focus on ways to detect and mitigate bias in AI-driven decisions. 2. **Process:** The process must include routine audits and assessments of AI models to ensure their reliability. Any biases found during the audits should be corrected to ensure that the system does not discriminate against specific groups. 3. **Technology:** AI systems must be designed to ensure reliability by using unbiased data and algorithms. The technology should enable continuous monitoring for system reliability and allow for adjustments to be made when biases are detected.
Explainability **(Dapat Dijelaskan)**	1. ***People:*** SDM bank memiliki kemampuan untuk mendeskripsikan model AI, perkiraan dampak, potensi bias, dan memahami bagaimana model AI dapat dijelaskan dalam melakukan mengambil keputusan. 1. **People:** Banks' human resources must have the capability to describe AI models, estimate their impacts, identify potential biases, and understand how AI models can be explained in the decision-making process.
Explainability **(Dapat Dijelaskan)**	2. ***Process:*** Memastikan AI dapat menjelaskan *dataset*, model yang digunakan dan bagaimana data dilatih, yang didukung dengan perencanaan dalam membuat algoritma AI yang bertanggung jawab. 3. ***Technology:*** Penggunaan sistem AI yang akurat, dapat ditelusuri, termasuk dimungkinkan untuk melakukan penyesuaian. 2. **Process:** Ensuring that AI can explain the dataset, the models used, and how the data is trained, supported by planning in creating responsible AI algorithms. 3. **Technology:** The use of AI systems must be accurate, traceable, and allow for adjustments to be made as necessary.
Security and Resilience **(Keamanan dan Ketahanan)**	1. ***People:*** Pegawai bank harus dilatih untuk mengenali dan memitigasi risiko termasuk keamanan siber yang terkait dengan sistem AI. Hal ini termasuk memastikan bahwa sistem AI aman dari serangan eksternal dan akses dibatasi hanya untuk personel yang berwenang. 2. ***Process:*** Proses harus memastikan bahwa penilaian keamanan rutin dilakukan pada sistem AI, dengan adanya protokol untuk mengatasi kerentanan. Rencana tanggap insiden juga harus didefinisikan dengan jelas dan mencakup langkah-langkah spesifik untuk mengatasi pelanggaran keamanan. 3. ***Technology:*** Sistem AI harus menyertakan fitur keamanan yang kuat, seperti enkripsi, penyimpanan data yang aman, dan pemantauan berkelanjutan terhadap ancaman dunia maya. Teknologi tersebut harus tahan terhadap serangan dan memastikan bahwa data nasabah dan bank selalu aman. 1. **People:** Bank employees must be trained to recognize and mitigate risks, including cybersecurity threats associated with AI systems. This includes ensuring that AI systems are secure from external attacks and that access is restricted only to authorized personnel. 2. **Process:** The process must ensure that routine security assessments are conducted on AI systems, with protocols in place to address vulnerabilities. An incident response plan should also be clearly defined and include specific steps to address security breaches. 3. **Technology:** AI systems must include robust security features, such as encryption, secure data storage, and continuous monitoring for cyber threats. The technology should be resilient against attacks and ensure that customer and bank data is always protected.

Prinsip 2. *Accountability*

Principle No. 2 Accountability

Accountability **(Akuntabilitas)**	1. ***People:*** Peran dan tanggung jawab yang jelas harus ditetapkan untuk memastikan akuntabilitas dalam penggunaan dan pengelolaan AI. Pegawai bank harus bertanggung jawab atas penerapan dan pemantauan sistem AI dengan benar, memastikan kepatuhan terhadap standar etika dan peraturan. 2. ***Process:*** Harus ada protokol yang secara jelas mendefinisikan siapa yang bertanggung jawab pada setiap tahap pengembangan, penerapan, dan pemantauan AI. Hal ini termasuk menjaga pengawasan untuk memastikan sistem memenuhi persyaratan yang ditetapkan. 3. ***Technology:*** Sistem AI harus menggabungkan mekanisme untuk melacak dan mencatat keputusan. Hal ini memastikan bahwa, jika terjadi kesalahan atau masalah, tanggung jawab dapat ditelusuri kembali ke individu atau proses yang benar, sehingga membantu mencegah dan mengatasi penyalahgunaan. 1. **People:** Clear roles and responsibilities must be established to ensure accountability in the use and management of AI. Bank employees should be responsible for the proper implementation and monitoring of AI systems, ensuring compliance with ethical standards and regulations. 2. **Process:** There needs to be protocols that clearly define who is responsible at each stage of the development, implementation, and monitoring of AI. This includes maintaining oversight to ensure that the systems meet the established requirements 3. **Technology:** AI systems must incorporate mechanisms to track and record decisions. This ensures that, in the event of an error or issue, accountability can be traced back to the appropriate individual or process, thereby helping to prevent and address misuse.
Transparency **(Transparansi)**	1. ***People:*** Pegawai bank harus dilatih untuk memahami sistem AI yang digunakan, sehingga memastikan kejelasan dalam proses pengambilan keputusan. Komunikasi yang jelas tentang cara kerja model AI diperlukan untuk membangun kepercayaan nasabah dan regulator. 1. **People:** Bank employees must be trained to understand the AI systems in use, ensuring clarity in the decision-making process. Clear communication about how AI models work is essential to build trust among customers and regulators.

Transparency **(Transparansi)**	2. ***Process:*** Prosedur tata kelola harus transparan, mendokumentasikan bagaimana model AI dibuat, digunakan, dan diaudit. Harus ada jalur pengambilan keputusan yang jelas dan dapat diakses oleh semua pemangku kepentingan. 3. ***Technology:*** Sistem AI harus dirancang dengan mempertimbangkan transparansi, memastikan bahwa operasi dan hasilnya dapat dijelaskan. Sistem AI yang transparan memungkinkan pemangku kepentingan (termasuk regulator dan nasabah) memahami alasan di balik keputusan. 2. **Process:** Governance procedures must be transparent, documenting how AI models are created, used, and audited. There should be a clear decision-making pathway that is accessible to all stakeholders. 3. **Technology:** AI systems must be designed with transparency in mind, ensuring that their operations and outcomes can be explained. Transparent AI systems allow stakeholders (including regulators and customers) to understand the reasoning behind decisions.
Data Privacy **(Pelindungan Data)**	1. ***People:*** Pegawai bank harus dilatih tentang UU PDP dan penanganan data nasabah. Mereka harus memiliki *awareness* untuk menjaga kerahasiaan dan risiko yang terkait dengan pelanggaran atau penyalahgunaan data. 2. ***Process:*** Prosesnya harus mencakup kebijakan tata kelola data yang ketat, memastikan bahwa data pelanggan ditangani secara bertanggung jawab. Prosedur ini harus menentukan bagaimana data harus disimpan, dibagikan, dan dimusnahkan, sesuai dengan UU PDP. 3. ***Technology:*** Teknologi AI harus menerapkan prinsip privasi, memastikan bahwa data sensitif pelanggan dienkripsi, dianonimkan, dan dilindungi di setiap tahapan *lifecycle* AI. Pengkinian langkah-langkah privasi perlu dilakukan untuk memenuhi peraturan yang terus berkembang. 1. **People:** Bank employees must be trained on the Personal Data Protection Law (UU PDP) and the handling of customer data. They should have awareness of maintaining confidentiality and the risks associated with data breaches or misuse. 2. **Process:** The process must include strict data governance policies, ensuring that customer data is handled responsibly. These procedures should specify how data should be stored, shared, and disposed of in accordance with the Personal Data Protection Law (UU PDP). 3. **Technology:** AI technology must implement privacy principles, ensuring that sensitive customer data is encrypted, anonymized, and protected at every stage of the AI lifecycle. Updates to privacy measures need to be conducted to comply with evolving regulations.

Prinsip 3. Pengawasan Manusia

Principle No. 3. Human Oversight

Human Oversight **(Pengawasan Manusia)**	1. ***People:*** Keterlibatan dan peran dari SDM bank dan pihak terkait lain yang kompeten untuk memastikan AI bekerja sesuai tujuan dan etika yang diharapkan, dengan didukung pelatihan dan keterampilan, pengambilan keputusan dan etika yang memadai. 2. ***Process:*** Bank memiliki kerangka kerja dan praktik yang memastikan AI beroperasi dalam batas yang aman dan terkendali, yang didukung dengan penerapan tata kelola, manajemen risiko, dan evaluasi dan validasi secara ketat dan berkelanjutan. 3. ***Technology:*** Teknologi yang digunakan bank dalam penerapan AI harus mendukung transparansi, akuntabilitas, dan kontrol manusia, antara lain terkait transparansi algoritma, kontrol manual, pengamanan data, dan peningkatan kinerja AI secara berkelanjutan melalui pengkinian teknologi dan pengawasan secara otomatis. 1. **People:** The involvement and roles of the bank's human resources and other competent stakeholders are essential to ensure that AI operates in accordance with the intended goals and ethical standards. This should be supported by adequate training, skills development, decision-making capabilities, and ethical considerations. 2. **Process:** The bank has a framework and practices that ensure AI operates within safe and controlled boundaries, supported by the implementation of governance, risk management, and rigorous and continuous evaluation and validation processes. 3. **Technology:** The technology used by the bank in the implementation of AI must support transparency, accountability, and human control, this includes aspects such as algorithm transparency, manual controls, data security, and continuous performance enhancement of AI through technology updates and automated monitoring.
Inclusivity **(Inklusif)**	1. ***People:*** Program pelatihan AI harus inklusif, memberikan keterampilan kepada pegawai bank yang berasal dari berbagai latar belakang untuk menggunakan dan mengelola AI. Hal ini memastikan bahwa sistem AI dapat diakses oleh semua pegawai. 1. **People:** AI training programs must be inclusive, encompassing bank employees from various skill backgrounds to use and manage AI effectively. This is essential to ensure that the AI systems are accessible to all employees.
Inclusivity **(Inklusif)**	2. ***Process:*** Proses harus dirancang untuk memastikan bahwa penerapan AI adil dan dapat diakses oleh semua segmen nasabah, terutama kelompok yang kurang terlayani. Kebijakan harus ada untuk memastikan bahwa manfaat AI dirasakan secara inklusif. 3. ***Technology:*** Sistem AI harus dibangun secara inklusif, memastikan bahwa sistem tersebut berfungsi dengan baik untuk beragam pengguna. Hal ini mencakup penggabungan fitur aksesibilitas (seperti opsi bahasa dan penyesuaian antarmuka) untuk memenuhi berbagai demografi dan kebutuhan nasabah. 2. **Process:** The process must be designed to ensure that the implementation of AI is fair and accessible to all customer segments, particularly underserved groups. Policies should be in place to ensure that the benefits of AI are experienced inclusively. 3. **Technology:** AI systems must be built inclusively, ensuring that they function effectively for a diverse range of users. This includes incorporating accessibility features (such as language options and interface customization) to meet the various demographics and needs of customers.
Sustainability **(Keberlanjutan)**	1. ***People:*** SDM bank dan pihak terkait lain memiliki pemahaman dalam melakukan implementasi dan pengawasan dalam melakukan penilaian dampak penggunaan AI terhadap keberlanjutan usaha bank dan target keberlanjutan lain, memberikan manfaat jangka panjang, dan menjaga kepentingan *stakeholders* untuk mendukung kegiatan usaha bank yang terus berkembang. 2. ***Process:*** Bank mengadopsi proses yang mendukung keberlanjutan usaha bank dan target keberlanjutan lain dalam setiap tahap implementasi dan siklus hidup AI, yang memberikan manfaat bagi nasabah dan *stakeholders* terkait secara jangka panjang. 1. **People:** The bank's human resources and other relevant parties must have an understanding of how to implement and oversee the assessment of the impact of AI usage on the sustainability of the bank's operations and other sustainability targets. This ensures long-term benefits while safeguarding stakeholder interests to support the bank's ongoing business activities. 2. **Process:** The bank adopts processes that support the sustainability of its operations and other sustainability targets at every stage of AI implementation and lifecycle. This approach provides long-term benefits for customers and relevant stakeholders.

Sustainability **(Keberlanjutan)**	3. ***Technology:*** Pengembangan teknologi AI untuk mendukung inovasi, efisiensi, transparansi, dan pengurangan dampak negatif terhadap lingkungan, dalam mendukung peningkatan layanan dan operasional bank. 3. **Technology**: The development of AI technology should support innovation, efficiency, transparency, and the reduction of negative environmental impacts, thereby enhancing the services and operations of the bank.
Ethics & Fairness **(Beretika)**	1. ***People:*** SDM bank memahami implikasi etis dari penerapan AI dan didukung mekanisme pengendalian dalam menilai dan memantau apakah AI bekerja secara etis. 2. ***Process:*** Penerapan proses yang harus mendukung penerapan etika dalam seluruh siklus hidup AI antara lain melalui analisi dampak, mengintegrasikan standar etika, nilai, dan kode etik yang berlaku pada bank untuk memastikan bahwa keputusan yang dihasilkan AI tidak menyimpang dan menerapkan mekanisme penanganan kesalahan/ pelanggaran. 3. ***Technology:*** Penggunaan sistem AI yang dirancang untuk mematuhi standar etika, nilai, dan kode etik serta meminimalkan risiko kesalahan/ pelanggaran. 1. **People:** Bank's human resources understand the ethical implications of AI implementation and are supported by control mechanisms to assess and monitor whether AI operates ethically. 2. **Process:** The implementation of processes must support the application of ethics throughout the entire AI lifecycle, including impact analysis, integrating ethical standards, values, and applicable codes of conduct within the bank to ensure that decisions generated by AI are not deviant and to implement mechanisms for handling errors/violations. 3. **Technology:** The use of AI systems designed to comply with ethical standards, values, and codes of conduct while minimizing the risk of errors/violations.

HALAMAN INI SENGAJA DIKOSONGKAN

THIS PAGE IS INTENTIONALLY LEFT BLANK

Bab 5
Manajemen Risiko dan Tata Kelola

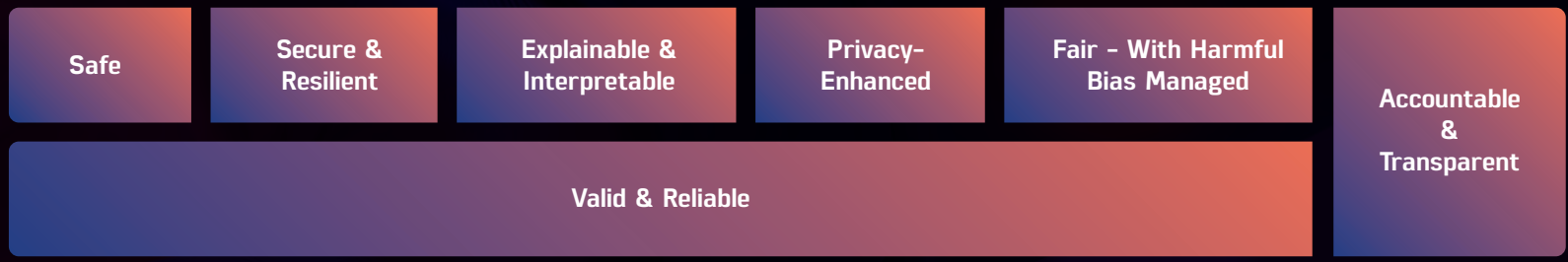
Chapter 5
Risk Management and Governance

A. Manajemen Risiko Kecerdasan Artifisial

A. Artificial Intelligence Risk Management

Gambar 10. Karakteristik AI yang Dapat Dipercaya

Figure 10. Trustworthy AI Characteristics



Source: NIST (2023)

National Institute of Standards and Technology (NIST) menyusun dan menerbitkan Artificial Intelligence Risk Management Framework (AI RMF) pada tanggal 26 Januari 2023. AI RMF dimaksudkan untuk meningkatkan kemampuan organisasi dalam mengelola risiko yang ditimbulkan oleh penggunaan AI dengan memasukkan unsur *trustworthiness* ke dalam *lifecycle* AI. Agar dapat dipercaya, sistem AI harus memenuhi kriteria yang bermanfaat bagi *stakeholders*, dan mengurangi potensi risiko yang timbul dari penggunaan AI. Kerangka ini mengartikulasikan karakteristik-karakteristik AI yang dapat dipercaya tersebut.

National Institute of Standards and Technology (NIST) have compiled and published the Artificial Intelligence Risk Management Framework (AI RMF) on January 26, 2023. The AI RMF was intended to enhance organizations' ability to manage risks arising from the use of AI by incorporating trustworthiness elements into the AI lifecycle. For the AI system to be trustworthy, it must meet criteria that are beneficial for stakeholders and reduce potential risks associated with AI usage. This framework articulates the characteristics of trustworthy AI.

Valid & Reliable adalah kondisi yang diperlukan untuk dapat dipercaya dan ditunjukkan sebagai dasar awal bagi karakteristik "AI yang dapat dipercaya" lainnya. Sedangkan untuk *Accountable & Transparent* ditampilkan dalam kotak vertikal karena berkaitan dengan seluruh karakteristik lainnya.

a. Valid & Reliable

Validasi adalah "konfirmasi, melalui penyediaan bukti obyektif, bahwa persyaratan untuk tujuan penggunaan atau penerapan tertentu telah dipenuhi" (Sumber: ISO 9000:2015). Penerapan sistem AI yang tidak akurat dan tidak dapat diandalkan akan menciptakan dan meningkatkan risiko yang ditimbulkan oleh

Valid & Reliable are conditions necessary for trustworthiness and serve as a foundational basis for other characteristics of "trustworthy AI." Meanwhile *Accountable & Transparent* is displayed in a vertical box because it relates to all other characteristics.

a. Valid & Reliable

Validation is the "confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled" (Source: ISO 9000:2015). The implementation of inaccurate and unreliable AI systems will create and increase risks and reduce trust in

sistem AII dan mengurangi kepercayaan terhadap sistem AI tersebut. Keandalan didefinisikan sebagai "kemampuan suatu sistem untuk bekerja sesuai kebutuhan, tanpa kegagalan, untuk interval waktu tertentu, dalam kondisi tertentu" (Sumber: ISO/IEC TS 5723:2022). Keandalan merupakan tujuan utama dari pengoperasian sistem AI yang sesuai dengan persyaratan penggunaan yang diharapkan dan selama periode waktu tertentu, termasuk selama masa pakai sistem AI.

b. Safe

Sistem AI tidak boleh membahayakan kehidupan manusia, kesehatan, harta benda, atau lingkungan. Pengoperasian sistem AI yang aman ditingkatkan melalui:

1. Praktik desain, pengembangan, dan penerapan yang bertanggung jawab.
2. Penyampaian informasi yang jelas kepada *deployers* mengenai penggunaan sistem AI yang bertanggung jawab.
3. Pengambilan keputusan yang bertanggung jawab oleh *deployers* dan pengguna akhir.
4. Penjelasan dan dokumentasi risiko berdasarkan bukti empiris kejadian.

Berbagai jenis risiko AI mungkin memerlukan pendekatan manajemen risiko yang disesuaikan berdasarkan konteks dan tingkat dampak potensi risiko yang ditimbulkan. Risiko AI yang berpotensi menimbulkan masalah serius atau memiliki dampak yang besar memerlukan prioritas yang paling mendesak dan proses manajemen risiko yang paling menyeluruh. Memastikan sejak awal bahwa sistem AI aman untuk digunakan (mulai dari tahap perencanaan dan desain) dapat mencegah kegagalan atau risiko yang membahayakan sistem AI. Cara praktis untuk menjaga keamanan AI meliputi pengujian dan simulasi yang teliti, pemantauan secara langsung, serta memastikan pengawas (*human oversight*) dapat mengubah, atau mengendalikan sistem AI jika terjadi penyimpangan dari fungsi yang diharapkan.

that system. Reliability is defined as the "ability of an item to perform as required, without failure, for a given time interval, under given conditions" (Source: ISO/IEC TS 5723:2022). Reliability is a primary goal of operating AI systems in accordance with expected usage requirements over a certain period, including during the lifespan of the AI system.

b. Safe

AI systems must not endanger human life, health, property, or the environment. The safe operation of AI systems is enhanced through:

1. Responsible design, development, and deployment practices.
2. Clear information to deployers regarding the responsible use of AI systems.
3. Responsible decision-making by deployers and end-users.
4. Explanations and documentation of risks supported by empirical evidence of incidents.

Various types of AI risks may require a tailored AI risk management approach based on the context and potential impact level of the risks posed. AI risks that could lead to serious issues or have significant impacts require the most urgent priority and the most comprehensive risk management processes. Establishing safety measures for AI systems from the early stages of planning and design helps prevent potential failures or risks that could jeopardize the system. Practical ways to maintain AI safety include thorough testing and simulation, real-time monitoring, as well as ensuring that supervisors (human oversight) are able to intervene, modify, or control the AI system in the event of deviations from its intended functions.

c. Secure & Resilient

Sistem AI harus dapat bertahan terhadap dampak buruk atau perubahan yang tidak terduga dalam penggunaannya atau dapat mempertahankan fungsi dan strukturnya dalam menghadapi perubahan internal dan eksternal. Sistem AI yang *secure* adalah sistem AI yang dapat menjaga kerahasiaan dan integritas melalui mekanisme perlindungan yang mencegah akses dan penggunaan yang tidak sah (*unauthorized*). Sementara resiliensi (*reslient*) adalah kemampuan sistem AI untuk kembali ke fungsi normal setelah kejadian buruk yang tidak terduga, dan mencakup protokol untuk menghindari, melindungi, merespon, atau pulih dari serangan.

Sebagai bagian dari pendekatan *secure & resilient*, diperlukan penerapan *red-teaming* dalam pengujian sistem AI. *Red-teaming* melibatkan simulasi serangan oleh tim independen untuk mengidentifikasi kelemahan keamanan, bias, dan potensi penyalahgunaan sistem. Strategi ini membantu meningkatkan daya tahan sistem AI terhadap ancaman dunia nyata serta memastikan kepatuhan terhadap standar keamanan dan regulasi yang berlaku.

c. Secure & Resilient

AI systems must be able to withstand adverse impacts or unexpected changes in their usage, as well as maintain their functions and structures in the face of internal and external changes. A secure AI system is one that can preserve confidentiality and integrity through protective mechanisms that prevent unauthorized access and use. Meanwhile, resilient refers to an AI system's capacity to restore normal functioning following unexpected adverse events, including mechanisms for prevention, protection, response, and recovery from potential attacks.

As part of the secure & resilient approach, the implementation of red-teaming in AI system testing is necessary. Red-teaming involves simulating attacks by an independent team to identify security vulnerabilities, biases, and potential misuse of the system. This strategy helps enhance the resilience of AI systems against real-world threats and ensures compliance with applicable security standards and regulations.

d. Accountable & Transparent

Sistem AI yang dapat dipercaya bergantung pada akuntabilitas. Akuntabilitas mengandalkan transparansi. Transparansi mencerminkan sejauh mana informasi tentang sistem AI dan *output* yang dihasilkan tersedia bagi individu yang berinteraksi dengan sistem tersebut, baik mereka menyadari interaksi tersebut ataupun tidak. Transparansi yang bermakna memberikan akses terhadap tingkat informasi yang sesuai tahap siklus hidup AI dan disesuaikan dengan peran atau pengetahuan individu yang berinteraksi dengan atau menggunakan sistem AI. Transparansi dapat mendorong tingkat pemahaman yang lebih tinggi sehingga meningkatkan kepercayaan terhadap sistem AI.

e. Explainable & Interpretable

Explainability mengacu pada mekanisme yang mendasari pengoperasian sistem AI. Sedangkan *Interpretability* mengacu pada makna dari *output* sistem AI. Secara bersama-sama, *explainability* dan *interpretability* membantu individu-individu yang mengoperasikan atau

d. Accountable & Transparent

Trustworthy AI systems rely on accountability. Accountability depends on transparency. Transparency reflects the extent to which information about the AI system and its outputs is available to individuals interacting with the system, regardless of whether they are aware that they are doing so. Meaningful transparency provides access to an appropriate level of information at each stage of the AI lifecycle and is tailored to the roles or knowledge of individuals interacting with or using the AI system. Transparency can encourage a higher level of understanding, thereby increasing trust in AI systems.

e. Explainable & Interpretable

Explainability refers to the mechanisms underlying the operation of AI systems. Meanwhile, Interpretability pertains to the meaning of the outputs generated by these systems. Together, explainability and interpretability help individuals who operate or oversee AI systems, as

mengawasi sistem AI, serta pengguna sistem AI, untuk mendapatkan wawasan yang lebih mendalam mengenai fungsionalitas dari sistem, termasuk *output* yang dihasilkan. *Explainability* dapat menjawab pertanyaan "bagaimana" suatu keputusan dibuat dalam sistem. *Interpretability* dapat menjawab pertanyaan "mengapa" suatu keputusan dibuat oleh sistem dan makna atau konteksnya bagi pengguna.

f. Privacy-Enhanced

Privasi secara umum mengacu pada norma dan praktik yang membantu menjaga identitas. Norma dan praktik ini umumnya mengatur agar individu terlindungi dari intervensi, memiliki batasan atas pengawasan, serta berhak memberikan persetujuan terkait pengungkapan atau pengendalian aspek identitas pribadinya. Nilai-nilai privasi seperti anonimitas, kerahasiaan, dan kontrol harus memandu pilihan untuk desain, pengembangan, dan penerapan sistem AI. Sistem AI juga dapat menghadirkan risiko baru terhadap privasi dengan memungkinkan inferensi mengenai informasi yang bersifat pribadi tentang individu.

g. Fair – with Harmful Bias Managed

Keadilan dalam AI mencakup kepedulian terhadap kesetaraan dengan mengatasi isu-isu seperti bias dan diskriminasi yang merugikan pihak/kelompok tertentu. Standar keadilan bisa jadi rumit dan sulit untuk didefinisikan karena persepsi keadilan berbeda antar budaya dan dapat berubah tergantung penerapannya. Upaya manajemen risiko AI harus mempertimbangkan perbedaan-perbedaan ini.

NIST telah mengidentifikasi 3 (tiga) kategori utama bias AI yang harus dipertimbangkan dan dikelola: sistemik, komputasi dan statistik, serta kognitif manusia. Masing-masing hal ini dapat

well as users of these systems, gain deeper insights into the functionality of the system, including its outputs. Explainability can answer questions about "how" a decision is made within the system. Interpretability can address questions about "why" a decision is made by the system and what that decision means or its context for users.

f. Privacy-Enhanced

Privacy generally refers to the norms and practices that help to safeguard identity. These norms and practices typically govern freedom from interference, restrictions on observation, or individuals' rights to consent to the disclosure or control of aspects of their identity. Privacy values such as anonymity, confidentiality, and control should guide choices for the design, development, and implementation of AI systems. AI systems can also present new risks to privacy by enabling inferences about personal information regarding individuals.

g. Fair – with Harmful Bias Managed

Fairness in AI encompasses a concern for equality by addressing issues such as bias and discrimination that harm certain parties/groups. Standards of fairness can be complex and difficult to define because perceptions of fairness vary across cultures and may change depending on their application. AI risk management efforts must take these differences into account.

NIST has identified 3 (three) main categories of AI bias that must be considered and managed: systemic, computational, and statistical, and human cognitive. Each of these can

Gambar 11. Kerangka Manajemen Risiko Kecerdasan Artifisial NIST

Figure 11. The Artificial Intelligence Risk Management Framework Developed by NIST

Source: NIST (2023)

terjadi tanpa adanya keberpihakan, atau niat diskriminatif. Bias sistemik terjadi ketika prasangka atau ketidakadilan yang sudah ada dalam masyarakat, proses, atau data yang digunakan untuk melatih sistem AI terintegrasi ke dalam cara kerja teknologi tersebut. Bias ini dapat muncul di berbagai tahap siklus hidup AI, seperti dalam pengumpulan data, desain model, pengujian, hingga implementasi di dunia nyata. Bias komputasi dan statistik adalah bentuk bias yang terjadi karena ketidaksempurnaan data, metode statistik, atau algoritma yang digunakan. Bias ini sering muncul ketika data atau metode pengolahan data tidak merepresentasikan populasi atau tujuan sebenarnya. Bias kognitif manusia berkaitan dengan kecenderungan manusia untuk memproses informasi atau membuat keputusan dengan cara yang dipengaruhi oleh keyakinan, pengalaman, atau pola pikir tertentu. Bias ini tidak hanya memengaruhi bagaimana manusia memahami atau menggunakan AI, tetapi juga cara mereka mempercayai, menafsirkan, dan berinteraksi dengan sistem tersebut.

Untuk meningkatkan kemampuan organisasi dalam mengelola risiko yang muncul dari penggunaan AI, NIST juga menyusun kerangka manajemen risiko AI dengan 4 (empat) fungsi berikut:

occur without any bias or discriminatory intent. Systemic bias occurs when existing prejudices or injustices in society, processes, or data used to train AI systems are integrated into the functioning of the technology. This bias can arise at various stages of the AI lifecycle, such as in data collection, model design, testing, and implementation in the real world. Computational and statistical bias is a form of bias that occurs due to imperfections in the data, statistical methods, or algorithms used. This bias often arises when the data or data processing methods do not represent the actual population or objectives. Human cognitive bias is related to the tendency of humans to process information or make decisions in ways influenced by specific beliefs, experiences, or mindsets. This bias not only affects how humans understand or use AI but also how they trust, interpret, and interact with the system.

To enhance organizations' ability to manage risks arising from the use of AI, NIST has also developed an AI risk management framework with the following 4 (four) functions:

a. Govern

Mengimplementasikan budaya manajemen risiko di organisasi dalam mendesain, mengembangkan, dan mengevaluasi sistem AI. Bank dapat menguraikan fungsi *govern* dalam beberapa kategori dengan memberikan contoh tindakan yang perlu dilakukan untuk menjalankan fungsi ini.

Contoh:

1. Kategori: Karakteristik AI yang dapat dipercaya diintegrasikan ke dalam kebijakan, proses, dan prosedur organisasi.
2. Tindakan yang dapat dilakukan:
   a. Menetapkan istilah dan konsep utama yang terkait dengan sistem AI dan ruang lingkup tujuan dan penggunaannya.
   b. Menghubungkan tata kelola AI dengan tata kelola dan pengendalian risiko organisasi yang telah ada.

b. Map

Fungsi ini dimaksudkan untuk meningkatkan kemampuan organisasi untuk mengidentifikasi risiko AI dan faktor-faktornya. Bank dapat menguraikan fungsi *map* dalam beberapa kategori dengan memberikan contoh tindakan yang perlu dilakukan untuk menjalankan fungsi ini.

a. Govern

To implement a risk management culture within the organization when designing, developing, and evaluating AI systems. Banks can outline governance functions into several categories by providing examples of actions that need to be taken to carry out this function.

Examples:

1. Category: Characteristics of trustworthy AI are integrated into the organization's policies, processes, and procedures.
2. Actions that can be taken:
   a. Establishing key terms and concepts related to AI systems and the scope of their objectives and usage.
   b. Connecting AI governance with the existing organizational governance and risk control.

b. Map

This function is intended to enhance the organization's ability to identify AI risks and their factors. Banks can outline the map function into several categories by providing examples of actions that need to be taken to carry out this function.

Contoh:

1. Kategori: Misi organisasi dan tujuan yang relevan untuk teknologi AI dipahami dan didokumentasikan.
2. Tindakan yang dapat dilakukan:
   a. Membangun praktik yang transparan ke dalam proses pengembangan sistem AI.
   b. Evaluasi tujuan sistem AI dengan mempertimbangkan potensi risiko, nilai-nilai sosial, dan prinsip-prinsip organisasi.

c. Measure

Menggunakan metode kualitatif, kuantitatif, atau kombinasi dari keduanya untuk menganalisis, menilai, *benchmark*, dan memantau risiko AI dan dampak yang ditimbulkan. Bank dapat menguraikan fungsi *measure* dalam beberapa kategori dengan memberikan contoh tindakan yang perlu dilakukan untuk menjalankan fungsi ini.

Contoh:

1. Kategori: Kesesuaian metrik AI dan efektivitas pengendalian dievaluasi dan diperbarui secara berkala termasuk laporan yang memuat *error* dan dampaknya terhadap masyarakat.
2. Tindakan yang dapat dilakukan:
   a. Menilai efektivitas metrik dan pengendalian yang diterapkan secara berkala di seluruh siklus hidup sistem AI.
   b. Mengembangkan metrik baru ketika metrik yang ada tidak mencukupi atau tidak efektif dalam rangka perbaikan dan peningkatan metrik.

d. Manage

Mengalokasikan sumber daya untuk memetakan dan mengukur risiko secara berkala. Hal ini terdiri dari rencana untuk merespon, memulihkan, dan mengomunikasikan suatu insiden. Bank dapat menguraikan fungsi *manage* dalam beberapa kategori dengan memberikan contoh tindakan yang perlu dilakukan untuk menjalankan fungsi ini.

Contoh:

1. Kategori: Mendokumentasikan risiko AI berdasarkan dampak, kemungkinan, atau sumber daya atau metode yang tersedia.
2. Tindakan yang dapat dilakukan:
   a. Sistem AI dengan toleransi risiko yang lebih rendah mendapatkan pengawasan, mitigasi, dan sumber daya manajemen yang lebih besar; serta

Examples:

1. Category: The organization's mission and relevant objectives for AI technology are understood and documented.
2. Actions that can be taken:
   a. Building transparent practices into the AI system development process.
   b. Evaluating the objectives of AI systems while considering potential risks, social values, and organizational principles.

c. Measure

Using qualitative, quantitative, or a combination of both methods to analyze, assess, benchmark, and monitor AI risks and their impacts. Banks can outline the measurement function into several categories by providing examples of actions that need to be taken to carry out this function.

Examples:

1. Category: The suitability of AI metrics and the effectiveness of controls are assessed and updated regularly, including reports that contain errors and their impacts on society.
2. Actions that can be taken:
   a. Assessing the effectiveness of metrics and controls applied regularly throughout the AI system lifecycle.
   b. Developing new metrics when existing metrics are insufficient or ineffective for the purpose of improvement and enhancement of metrics.

d. Manage

Allocating resources to map and measure risks regularly. This includes plans for responding to, recovering from, and communicating about an incident. Banks can outline the management function into several categories by providing examples of actions that need to be taken to carry out this function.

Examples:

1. Category: Documenting AI risks based on impact, likelihood, or available resources or methods.
2. Actions that can be taken:
   a. AI systems with lower risk tolerance receive greater oversight, mitigation, and management resources.

b. Meninjau dan mengkalibrasi ulang toleransi risiko, sesuai kebutuhan, dan informasi dari pemantauan dan penilaian terhadap sistem AI.

Implementasi manajemen risiko di atas harus melibatkan pelaku AI yang beragam dari berbagai disiplin ilmu, latar belakang, dan pengalaman untuk memastikan bahwa sistem AI bekerja secara adil, transparan, dan bertanggung jawab di seluruh siklus hidupnya.

Kerangka ini mencakup 6 (enam) tahapan penting dalam siklus hidup AI, yaitu:

a. Application Context

1. Tahap: Perencanaan dan desain
2. Kegiatan: Pada tahap ini, sistem AI dirancang dengan mendokumentasikan konsep, tujuan, asumsi dasar, serta konteksnya dengan mempertimbangkan persyaratan hukum, peraturan, dan etika. Selain itu, proses ini mencakup audit dan penilaian dampak.
3. Pelaku representatif: Operator sistem, *end users,* perancang AI, auditor, pakar regulasi, pakar TEVV (*Testing, Evaluation, Verification, and Validation*), manajer produk, pakar tata kelola, dan komunitas terdampak.
4. Tujuan: Memastikan bahwa pengembangan AI dimulai dengan pemahaman yang jelas tentang kebutuhan dan potensi risiko.

b. Data & Input

1. Tahap: Pengumpulan dan pemrosesan data
2. Kegiatan: Data dikumpulkan, divalidasi, dan diproses untuk memastikan metadata serta karakteristik *dataset* sesuai dengan tujuan, regulasi, dan pertimbangan etis. Validasi internal dan eksternal dilakukan pada tahap ini.

b. Reviewing and recalibrating risk tolerance as needed, based on information from monitoring and assessment of AI systems.

The implementation of the above risk management must involve diverse AI stakeholders from various disciplines, backgrounds, and experiences to ensure that AI systems operate fairly, transparently, and responsibly throughout their lifecycle.

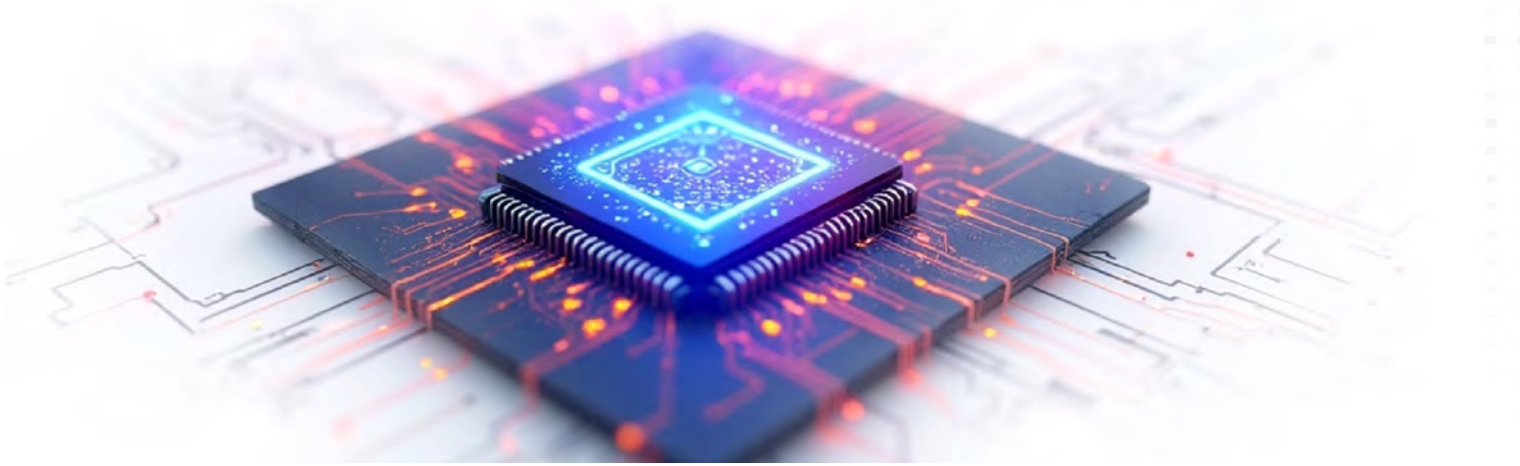
This framework includes 6 (six) key stages in the AI lifecycle, namely:

a. Application Context

1. Stage: Plan and Design
2. Activity: At this stage, the AI system is designed by documenting concepts, objectives, underlying assumptions, and context while considering legal, regulatory, and ethical requirements. Additionally, this process includes audits and impact assessments.
3. Representative Actors: System operators, end users, AI designers, auditors, regulatory experts, TEVV (Testing, Evaluation, Verification, and Validation) specialists, product managers, governance experts, and affected communities.
4. Purpose: Ensuring that AI development begins with a clear understanding of needs and potential risks.

b. Data & Input

1. Stage: Collect and Process Data
2. Activity: Data is collected, validated, and processed to ensure that the metadata and characteristics of the dataset align with objectives, regulations, and ethical considerations. Internal and external validation is conducted at this stage.



3. Pelaku representatif: *Data scientist*, *data engineers*, penyedia data, pakar domain, analis sosial-budaya, dan pakar TEVV.
4. Tujuan: Mengumpulkan data yang representatif dan bebas bias untuk menghindari hasil yang tidak adil atau diskriminatif.

c. AI Model

1. Tahap: Pembuatan dan pengujian model.
2. Kegiatan: Pada tahap ini, algoritma dipilih atau dibuat, kemudian dilatih menjadi model AI. Tahap ini juga melibatkan pengujian model.
3. Pelaku representatif: Pengembang model, *model engineers*, *data scientists*, pakar domain, dengan berkonsultasi dengan analis sosial-budaya dan pakar TEVV.
4. Tujuan: Membuat model AI yang akurat, andal, dan sesuai dengan kebutuhan.

3. Representative actors: Data scientists, data engineers, data providers, domain experts, socio-cultural analysts, and TEVV specialists.
4. Purpose: Collecting representative and bias-free data to avoid unfair or discriminatory outcomes.

c. AI Model

1. Stage: Build and Use Model.
2. Activity: At this stage, algorithms are selected or created and then trained into an AI model. This phase also involves testing the model.
3. Representative actors: Model developers, model engineers, data scientists, domain experts, in consultation with socio-cultural analysts and TEVV.
4. Purpose: Creating AI models that are accurate, reliable, and aligned with needs.

d. AI Model

1. Tahap: Verifikasi dan validasi model.
2. Kegiatan: Model yang dikembangkan diverifikasi, dikalibrasi, dan output yang dihasilkan diinterpretasikan untuk memastikan kualitas dan keandalan. Proses ini juga mencakup pengujian lebih lanjut.
3. Pelaku representatif: Pengembang model, *model engineers*, *data scientists*, pakar domain, dengan

d. AI Model

1. Stage: Verify and validate model.
2. Activity: The developed model is verified, calibrated, and its outputs are interpreted to ensure quality and reliability. This process also includes further testing.
3. Representative actors: Model developers, model engineers, data scientists, domain experts,

berkonsultasi dengan analis sosial-budaya dan pakar TEVV.

4. Tujuan: Memastikan model berfungsi dengan baik dan sesuai dengan standar yang telah ditetapkan.

c. *Task & Output*

1. Tahap: Penerapan dan penggunaan model.
2. Kegiatan: AI diimplementasikan dan diuji kompatibilitasnya dengan sistem lama, kesesuaiannya dengan peraturan, serta efektivitasnya dalam organisasi. Evaluasi pengalaman pengguna menjadi fokus utama pada tahap ini.
3. Pelaku representatif: Integrator sistem, pengembang sistem dan perangkat lunak, pakar domain, pakar pengadaan, *third-party suppliers*, dengan berkonsultasi dengan analis sosial-budaya, pakar tata kelola dan pakar TEVV.
4. Tujuan: Mengimplementasikan AI dengan aman dan efektif, serta memantau dampak penggunaannya.

d. *Application Context*

1. Tahap: Operasi dan pemantauan sistem.
2. Kegiatan: Sistem AI dioperasikan sambil terus memantau dampaknya, baik yang diharapkan maupun yang tidak diharapkan. Tahap ini memastikan kepatuhan terhadap hukum, regulasi, dan etika.
3. Pelaku representatif: Operator sistem, *end users*, auditor, praktisi, pakar domain, pendesain AI, penilai dampak, manajer produk, pakar regulasi dan tata kelola, dan komunitas terdampak.
4. Tujuan: Menjamin sistem AI terus berjalan sesuai dengan tujuan awal dan mempertimbangkan dampak jangka panjang.

e. *People & Planet*

1. Tahap: Penggunaan dan dampak
2. Kegiatan: Sistem AI harus digunakan dengan pendekatan yang inklusif (tidak menciptakan ketidakadilan), memantau dampaknya terhadap pengguna dan lingkungan, dan tetap melindungi hak-hak terkait.
3. Pelaku representatif: *end users*, operator dan praktisi, komunitas terdampak, pembuat kebijakan, organisasi advokasi, *environmental groups*, dan peneliti.
4. Tujuan: Menggunakan AI untuk memberikan manfaat kepada pengguna, sambil meminimalkan risiko yang tidak diinginkan, termasuk terhadap lingkungan.

in consultation with socio-cultural analysts and TEVV specialists.

4. Purpose: Ensuring that the model functions well and meets established standards.

c. Task & Output

1. Stage: Deploy and use model.
2. Activity: AI is implemented and tested for compatibility with legacy systems, compliance with regulations, and effectiveness within the organization. User experience evaluation becomes a primary focus at this stage.
3. Representative actors: System integrators, system and software developers, domain experts, procurement specialists, third-party suppliers, in consultation with socio-cultural analysts, governance experts, and TEVV specialists.
4. Purpose: Implementing AI safely and effectively, while monitoring its impact.

d. Application Context

1. Stage: Operate and monitor system.
2. Activity: The AI system is operated while continuously monitoring its impacts, both expected and unexpected. This stage ensures compliance with laws, regulations, and ethical standards.
3. Representative actors: System operators, end users, auditors, practitioners, domain experts, AI designers, impact assessors, product managers, regulatory and governance experts, and affected communities.
4. Purpose: Ensuring that the AI system continues to operate according to its original objectives while considering long-term impacts.

e. People & Planet

1. Stage: Use and Impact
2. Activity: The AI system must be used with an inclusive approach (does not create injustices), monitoring its impact on users and the environment, and advocating for related rights.
3. Representative actors: end users, operators, and practitioners, affected communities, policymakers, advocacy organizations, environmental groups, and researchers.
4. Purpose: Using AI to provide benefits to users while minimizing unwanted risks, including those to the environment.

Gambar 12. Tahapan dalam Siklus Hidup AI

Figure 12. Stages in the AI Lifecycle

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compilance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objective, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for right.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analyst familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impacted individuals/communities; evaluators.	End users, operators, and practitioners; impacted individuals/ communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Source: NIST (2023)

Disamping itu, khusus untuk risiko yang berkaitan dengan GenAI, NIST melakukan *GenAI Public Working Group* yang menghasilkan beberapa pertimbangan utama terkait GenAI yang dapat digunakan oleh organisasi dalam merancang, mengembangkan, dan menggunakan GenAI.

1. Organizational Governance

Organisasi perlu meninjau kembali dan menyesuaikan kerangka tata kelola mereka untuk menangani risiko unik yang ditimbulkan oleh GenAI. Ini mencakup:

a. Menyesuaikan tingkat risiko sistem GenAI agar selaras dengan prioritas risiko organisasi.

b. Menetapkan peran dan tanggung jawab yang jelas dalam tata kelola, pengawasan, dan kepatuhan terkait GenAI.

2. Third-Party Considerations

Penggunaan sistem dan model GenAI dari pihak ketiga dapat menimbulkan risiko terkait hak kekayaan intelektual, privasi data, dan keamanan. Organisasi disarankan untuk:

a. Memperbarui proses uji kelayakan dalam pengadaan GenAI.

b. Mewajibkan persyaratan mengenai prinsip transparansi untuk vendor GenAI.

c. Menyusun kebijakan yang jelas untuk manajemen risiko terkait penggunaan pihak ketiga, termasuk pemantauan berkelanjutan dan rencana kontingensi jika terjadi kegagalan.

3. Pre-Deployment Testing

Sistem GenAI harus menjalani pengujian, evaluasi, validasi, dan verifikasi (*test, evaluation, validation, and verification*/ TEVV) yang ketat sebelum digunakan. Beberapa langkah penting terkait TEVV meliputi:

a. Menggunakan metodologi pengujian terstruktur, bukan hanya berdasarkan pengalaman atau anekdot.

b. Mengidentifikasi bias, kerentanan keamanan, dan masalah keandalan.

c. Melakukan pengujian dengan validitas eksternal (*external validity*) untuk menilai dampak dalam kondisi dunia nyata.

4. Structured Public Feedback

Mekanisme umpan balik dari pemangku kepentingan dapat membantu meningkatkan tata kelola GenAI. Metode yang direkomendasikan meliputi:

a. Keterlibatan partisipatif, seperti: diskusi kelompok, studi pengguna, dan survei.

In addition, specifically for risks related to Generative AI (GenAI), NIST has established the GenAI Public Working Group, which has produced several key considerations regarding GenAI that organizations can use in designing, developing, and utilizing GenAI.

1. Organizational Governance

Organizations need to review and adjust their governance frameworks to address the unique risks posed by Generative AI (GenAI). This includes:

a. Adjusting the risk levels of GenAI systems to align with the organization's risk priorities.

b. Establishing clear roles and responsibilities in governance, oversight, and compliance related to GenAI.

2. Third-Party Considerations

The use of third-party GenAI systems and models can pose risks related to intellectual property rights, data privacy, and security. Organizations are advised to:

a. Updating the feasibility testing processes in GenAI procurement.

b. Mandating transparency principles for GenAI vendors.

c. Formulating clear policies for risk management related to third-party usage, including continuous monitoring and contingency plans in case of failures.

3. Pre-Deployment Testing

GenAI systems must undergo rigorous testing, evaluation, validation, and verification (TEVV) before use. Some important steps related to TEVV include:

a. Using structured testing methodologies, rather than relying solely on experience or anecdotes.

b. Identifying biases, security vulnerabilities, and reliability issues.

c. Conducting testing with external validity to assess impact in real-world conditions.

4. Structured Public Feedback

Feedback mechanisms from stakeholders can help improve GenAI governance. Recommended methods include:

a. Participatory engagement: Focus groups, user studies, and surveys.

b. Uji coba, mengamati bagaimana pengguna berinteraksi dengan konten yang dihasilkan AI dalam kondisi nyata.

c. AI *Red-Teaming*, yaitu pengujian terstruktur untuk mengidentifikasi risiko, bias, dan kerentanan keamanan.

5. Content Provenance

Untuk meningkatkan kepercayaan dan transparansi, organisasi dapat melacak asal-usul dan riwayat konten yang dihasilkan AI dengan teknik seperti:

a. Pelacakan sumber data (*provenance data tracking*) guna membedakan konten asli dari yang diubah oleh AI.

b. Digital watermarking dan pelacakan metadata (pengembang model GenAI atau pembuat konten GenAI, tanggal/waktu pembuatan, lokasi, modifikasi, dan sumber) untuk memastikan integritas konten.

c. Metode verifikasi manusia untuk membedakan konten yang dibuat oleh manusia dari yang dihasilkan AI.

6. Enhancing Content Provenance through Structured Public Feedback

Organisasi dapat mengintegrasikan umpan balik (sebelum dan sesudah penerapan) dari pemangku kepentingan dalam upaya provenansi konten (pencatatan asal dan jejak perubahan suatu konten) dengan:

a. Memanfaatkan pendekatan umpan balik eksternal (AI *red-teaming*, masukan komunitas) untuk meningkatkan metode provenansi.

b. Memantau reaksi pengguna terhadap langkah-langkah transparansi seperti pemberian label pada konten buatan AI.

c. Melacak serta mendokumentasikan provenansi *dataset* untuk mengidentifikasi apakah data yang dihasilkan AI menjadi penyebab utama permasalahan yang terjadi dalam sistem GenAI.

7. Incident Disclosure

Insiden AI didefinisikan sebagai suatu peristiwa, keadaan, atau rangkaian peristiwa di mana pengembangan, penggunaan, atau malfungsi satu atau lebih sistem AI secara langsung atau tidak langsung menyebabkan dampak negatif. Dampak ini dapat berupa:

a. Bahaya terhadap kesehatan individu atau kelompok, termasuk dampak psikologis dan kesehatan mental.

b. Gangguan terhadap pengelolaan dan operasional infrastruktur vital.

c. Pelanggaran hak asasi manusia atau hukum yang melindungi hak-hak fundamental, ketenagakerjaan, dan kekayaan intelektual.

d. Kerugian terhadap lingkungan.

b. Testing: Observing how users interact with AI-generated content in real conditions.

c. AI Red-Teaming: Structured testing to identify risks, biases, and security vulnerabilities

5. Content Provenance

To enhance trust and transparency, organizations can track the origins and history of AI-generated content using techniques such as:

a. Data provenance tracking to distinguish original content from ones that are altered by AI.

b. Digital watermarking and metadata tracking (including the GenAI model developer or content creator, creation date/time, location, modifications, and source) to ensure content integrity.

c. Human verification methods to distinguish between content created by humans and ones that are generated by AI

6. Enhancing Content Provenance through Structured Public Feedback

Organizations can integrate feedback (before and after implementation) from stakeholders in content provenance efforts (recording the origin and change history of content) by:

a. Using external feedback approaches (AI red-teaming, community input) to enhance provenance methods.

b. Monitoring user reactions to transparency measures such as labeling AI-generated content.

c. Tracking and documenting the provenance of datasets to identify whether AI-generated data is the primary cause of issues occurring within GenAI systems.

7. Incident Disclosure

AI incidents are defined as an event, condition, or series of events in which the development, use, or malfunction of one or more AI systems directly or indirectly causes negative impacts. These impacts can include:

a. Hazards to the health of individuals or groups, including psychological impacts and mental health issues.

b. Disruptions to the management and operation of vital infrastructure.

c. Violations of human rights or laws protecting fundamental rights, employment, and intellectual property.

d. Environmental harm or damage.

Untuk menangani dan mengelola insiden AI tersebut, organisasi dapat:

a. Mengembangkan saluran formal yang terstandarisasi untuk mendokumentasikan dan melaporkan insiden AI, termasuk GenAI.

b. Menyusun pedoman pelaporan insiden AI bagi seluruh pemangku kepentingan dalam siklus hidup AI untuk memastikan aktor yang terlibat dalam sistem AI memahami peran dan tanggung jawab mereka dalam mendeteksi, melaporkan, dan menangani insiden AI tersebut.

c. Melakukan pencatatan serta analisis terhadap insiden AI secara berkala.

d. Memfasilitasi pertukaran informasi mengenai insiden AI dengan pemangku kepentingan terkait.

Selain NIST, OECD juga menerbitkan kerangka manajemen risiko kecerdasan buatan. Dalam mengevaluasi kerangka manajemen risiko, OECD menemukan keselarasan antara *interoperability framework* dan beberapa standar kerangka manajemen risiko AI. Meskipun urutan operasi, target audiens, cakupan risiko, segmen dalam siklus hidup sistem AI, dan terminologi yang digunakan mungkin berbeda, semua kerangka kerja manajemen risiko AI pada dasarnya bertujuan untuk mencapai hasil yang sama–yaitu AI yang bertanggung jawab, etis, dan terpercaya–melalui proses manajemen risiko yang kurang lebih serupa, serta mencakup langkah-langkah berikut:

1. MENENTUKAN (*DEFINE*) cakupan, konteks, dan kriteria, termasuk prinsip, pemangku kepentingan, dan pelaku AI yang relevan untuk setiap fase siklus hidup sistem AI dan untuk siklus hidup itu sendiri.

To address and manage such AI incidents, organizations can:

a. Developing standardized formal channels for documenting and reporting AI incidents, including those related to GenAI.

b. Establishing incident reporting guidelines for all stakeholders throughout the AI lifecycle to ensure that actors involved in the AI system understand their roles and responsibilities in detecting, reporting, and addressing AI incidents.

c. Conducting regular documentation and analysis of AI incidents.

d. Facilitating information exchange regarding AI incidents with relevant stakeholders.

In addition to NIST, OECD has also published a framework for artificial intelligence risk management. In evaluating the risk management framework, OECD found alignment between interoperability frameworks and several AI risk management standards. Although the order of operations, target audience, scope of risks, segments in the AI system lifecycle, and terminology used may differ, all AI risk management frameworks fundamentally aim to achieve the same outcomes–namely responsible, ethical, and trustworthy AI–through a similar risk management process that includes the following steps:

1. DEFINE the scope, context, and criteria, including principles, stakeholders, and relevant AI actors for each phase of the AI system lifecycle and for the lifecycle itself.

2. MENILAI (*ASSESS*) risiko terhadap AI dengan mengidentifikasi dan menganalisis masalah pada tingkat individu, agregat, dan masyarakat serta mengevaluasi kemungkinan dan tingkat dampaknya.
3. MENGATASI (*TREAT*) risiko untuk menghentikan, mencegah, atau mengurangi dampak buruk, sesuai dengan kemungkinan dan tingkat dampak risiko.
4. MENGATUR (*GOVERN*) proses manajemen risiko dengan menanamkan dan menumbuhkan budaya manajemen risiko dalam organisasi, memantau dan mereviu proses secara berkelanjutan, serta mendokumentasikan, mengomunikasikan, mengkonsultasikan tentang proses dan hasil.

2. ASSESS risks to AI by identifying and analyzing issues at the individual, aggregate, and societal levels, as well as evaluating their likelihood and impact severity.
3. TREAT risks to stop, prevent, or mitigate adverse impacts, in accordance with the likelihood and severity of the risk.
4. GOVERN the risk management process by: instilling and fostering a risk management culture within the organization; continuously monitoring and reviewing the process; and documenting, communicating, and consulting on the process and outcomes.

Sementara itu, beberapa yurisdiksi dan otoritas lain juga mengeluarkan beberapa kerangka manajemen risiko. Perbedaan utama antara standar-standar ini adalah target utama dari penerapannya. *OECD Due Diligence Guidance* dan standar ISO terutama ditujukan untuk perubahan pada tingkat dewan direksi atau organisasi guna memungkinkan manajemen risiko

Meanwhile, several jurisdictions and other authorities have also issued various risk management frameworks. The main difference between these standards is the primary target of their implementation. The OECD Due Diligence Guidance and ISO standards are primarily aimed at changes at the board or organizational level to enable effective risk management.

Tabel 6. Kerangka Manajemen Risiko Kecerdasan Artifisial

Table 6. Artificial Intelligence Risk Management Frameworks

OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT	
	Monitor & review	Communicate	Consult	Document	Embed				
OECD DDG	TRACK	COMMUNICATE	EMBED			IDENTIFY & ASSESS		CEASE, PREVENT & MITIGATE	REMEDIATION
ISO 31000	MONITORING & REVIEW	COMMUNICATION & CONSULTATION		RECORDING & REPORTING	LEADERSHIP & COMMITMENT	SCOPE, CONTEX & CRITERIA	RISK ASSESSMENT	RISK TREATMENT	
NIST AI RMF	GOVERN					MAP	MEASURE	MANAGE	
EU AI ACT	Post-market monitoring system and regular systematic updating	Communication of residual risks, accuracy, conformity, serious incident	N/A	Documentation, record keeping, traceability	Quality management system	Identify, analyse and evaluate known and foreseeable risks, test system		Eliminate, reduce, mitigate, and control any risks	

OECD INTEROPERABILITY FRAMEWORK	GOVERN					DEFINE	ASSESS	TREAT
	Monitor & review	Communicate	Consult	Document	Embed			
AIDA	Monitor compliance with mitigation measures, record keeping	Publication system description, notification of material harm	N/A	Keeping general and additional records	Establish compliance measures	Identify and assess risks		Implement and monitor measures to mitigate or cease risks and compliance orders
HUDERIA	Iterative requirements	N/A	Stakeholder engagement process (SEP)	N/A	N/A	Context-Based Risk Analysis (COBRA)	Human Rights, Democracy and the Rule of the Law Impact Assessment (HUDERIA)	Impact Mitigation Plan (IMP)
IEEE 7000–21	N/A	Transparency management process	Ethical values elicitation and prioritisation	N/A	N/A	Concept of operations and context exploration	Ethical values elicitation and prioritisation	Ethical requirements definition and ethical risk-based design
ISO/IEC Guide 51	Validation & documentation	N/A	N/A	Validation & documentation	N/A	Identify user, intended use and reasonably foreseeable misuse/ Hazard identification	Estimation/ Evaluation of risk	Risk reduction

Source: OECD Artificial Intelligence Papers No. 5, Common Guideposts to Promote Interoperability in AI Risk Management (2023)

diterapkan secara efektif. Sementara standar lainnya juga memberikan rekomendasi pada level pengurus, namun implementasi utamanya lebih berfokus pada tingkat teknis (misalnya, mengidentifikasi dan mengatasi risiko dalam desain sistem AI dan sepanjang siklus hidup sistem AI).

Masing-masing standar kerangka manajemen risiko AI berikut memberikan gambaran umum mengenai tujuan, proses manajemen risiko, diikuti dengan analisis singkat tentang persamaan dan kesenjangan antara standar dan *interoperability framework*.

While other standards also provide recommendations at the governance level, their main implementation focus is more on the technical level (for example, identifying and addressing risks in AI system design and throughout the AI system lifecycle).

Each of the following AI risk management framework standards provides an overview of objectives, risk management processes, followed by a brief analysis of the similarities and gaps between the standards and interoperability frameworks.

a. OECD Due Diligence Guidance for Responsible Business Conduct (OECD DDG)

1. Tujuan: Membantu bisnis mengidentifikasi dan mengatasi dampak negatif yang mungkin terjadi dalam operasional, rantai pasok, atau hubungan bisnis.

a. OECD Due Diligence Guidance for Responsible Business Conduct (OECD DDG)

1. Purpose: Helping businesses identify and address potential negative impacts that may occur in operations, supply chains, or business relationships.

2. Struktur: 6 (Enam) langkah utama, yaitu: menanamkan tanggung jawab bisnis dalam kebijakan dan sistem manajemen; mengidentifikasi dan menilai dampak negatif; menghentikan, mencegah, atau mengurangi dampak buruk; memantau implementasi dan hasilnya; mengkomunikasikan bagaimana dampak tersebut ditangani; dan menyediakan mekanisme pemulihan (jika diperlukan).

3. Keselarasan: Hampir sepenuhnya sesuai dengan *interoperability framework*, tetapi memiliki kategori tambahan untuk pemulihan (*remediation*) dalam proses manajemen risiko.

b. ISO 31000:2018 Risk Management – Guidelines (ISO 31000) and ISO/IEC 23894:2023

1. Tujuan: Memberikan prinsip dan panduan umum mengenai penerapan manajemen risiko di berbagai sektor dan organisasi, atau berlaku dalam berbagai keadaan sehingga memungkinkan untuk disesuaikan dengan organisasi mana pun dan konteks spesifiknya.

2. Struktur: Menentukan ruang lingkup risiko, menilai risiko, menangani risiko, dan menanamkan manajemen risiko ke dalam kebijakan perusahaan. ISO/IEC 23894:2023 menambahkan panduan khusus untuk AI, termasuk tata kelola, keterlibatan pemangku kepentingan, pemantauan, dan penilaian ulang sistem manajemen risiko untuk peningkatan berkelanjutan.

3. Keselarasan: Sejalan dengan *interoperability framework*, tetapi lebih fokus pada nilai organisasi dibandingkan dampak eksternal bagi masyarakat.

c. NIST AI Risk Management Framework (NIST AI RMF)

1. Tujuan: Kerangka kerja sukarela untuk membantu organisasi mengelola risiko AI sepanjang siklus hidupnya.

2. Structure: The 6 (six) main steps are: instilling business responsibility in policies and management systems, identifying and assessing negative impacts; stopping, preventing, or mitigating adverse impacts; monitoring implementation and outcomes; communicating how those impacts are addressed; and providing recovery mechanisms (if necessary).

3. Alignment: Almost entirely aligned with the interoperability framework, but includes an additional category for remediation in the risk management process.

b. ISO 31000:2018 Risk Management – Guidelines (ISO 31000) and ISO/IEC 23894:2023

1. Purpose: Providing principles and general guidelines regarding the application of risk management across various sectors and organizations, or applicable in different circumstances, allowing for adaptation to any organization and its specific context.

2. Structure: Defining the scope of risks, assessing risks, addressing risks, and embedding risk management into corporate policies. ISO/IEC 23894:2023 adds specific guidance for AI, including governance, stakeholder engagement, monitoring, and reassessment of the risk management system for continuous improvement.

3. Alignment: In line with the interoperability framework, but with a greater focus on organizational value rather than external impacts on society.

c. NIST AI Risk Management Framework (NIST AI RMF)

1. Purpose: A voluntary framework to assist organizations in managing AI risks throughout its lifecycle.

2. Struktur: Empat fungsi utama–*GOVERN* (kebijakan dan akuntabilitas), *MAP* (memahami sistem AI dan risikonya), *MEASURE* (menilai keandalan AI), dan *MANAGE* (mengalokasikan sumber daya untuk mitigasi risiko).

3. Keselarasan: Sebagian besar selaras dengan *interoperability framework*, tetapi elemen *GOVERN* (pemantauan, dokumentasi, komunikasi, dan konsultasi) terintegrasi di berbagai langkah, bukan sebagai fungsi terpisah.

d. European Union proposal for a regulation laying down harmonised rules on AI (EU AIA)

1. Tujuan: Regulasi yang mengkategorikan AI berdasarkan:
   a. Risiko yang tidak dapat diterima (misalnya, AI untuk manipulasi bawah sadar atau penilaian sosial yang merugikan).
   b. Risiko tinggi (memerlukan pengelolaan risiko dan kepatuhan ketat).
   c. Risiko minimal (hanya memerlukan beberapa kepatuhan).

4. Struktur: Sistem AI berisiko tinggi wajib menjalani evaluasi kepatuhan sebelum digunakan dan harus dilakukan pemantauan berkelanjutan.

2. Structure: Four main functions–GOVERN (policies and accountability), MAP (understanding the AI system and its risks), MEASURE (assessing AI reliability), and MANAGE (allocating resources for risk mitigation).

3. Alignment: Mostly aligned with the interoperability framework, but the GOVERN elements (monitoring, documentation, communication, and consultation) are integrated across various steps rather than as separate functions.

d. European Union proposal for a regulation laying down harmonised rules on AI (EU AIA)

1. Purpose: Regulations that categorize AI based on:
   a. Unacceptable risks (for example, AI for subconscious manipulation or harmful social scoring).
   b. High risks (requiring strict risk management and compliance).
   c. Minimal risks (requiring only minimal compliance).

4. Structure: High-risk AI systems are required to undergo compliance evaluations before use and must be subject to continuous monitoring.

5. Keselarasan: Sejalan dengan *DEFINE*, *ASSESS*, dan *TREAT*, namun belum maksimal dalam aspek *GOVERN*, yaitu konsultasi dengan pemangku kepentingan internal dan eksternal, serta tidak secara eksplisit mengintegrasikan budaya manajemen risiko dalam organisasi.

e. Proposed Canada Artificial Intelligence and Data Act (AIDA)

1. Tujuan: Mengatur desain, pengembangan, dan penerapan AI berbasis risiko.

2. Struktur: Meliputi proses perancangan, pengembangan, peluncuran, dan operasional AI dengan kewajiban untuk mengidentifikasi dan mengatasi risiko sistem AI berisiko tinggi.

3. Keselarasan: Mirip dengan *EU AI Act*, sejalan dengan *DEFINE*, *ASSESS*, dan *TREAT*, tetapi tidak secara eksplisit mengintegrasikan manajemen risiko pada setiap tingkat pengambilan keputusan dan pengembangan AI.

f. Draft Council of Europe Human Rights, Democracy and the Rule of Law Risk and Impact Assessment (HUDERIA)

1. Tujuan: Menilai risiko AI terhadap hak asasi manusia, demokrasi, dan supremasi hukum.

5. Alignment: In line with DEFINE, ASSESS, and TREAT, but lacking in the GOVERN aspect, which includes consultation with internal and external stakeholders, and does not explicitly integrate a risk management culture within the organization.

e. Proposed Canada Artificial Intelligence and Data Act (AIDA)

1. Purpose: Regulating the design, development, and application of risk-based AI.

2. Structure: Covering the processes of designing, developing, launching, and operating AI with the obligation to identify and address risks associated with high-risk AI systems.

3. Alignment: Similar to the EU AI Act, aligned with DEFINE, ASSESS, and TREAT, but does not explicitly integrate risk management at every level of decision-making and AI development.

f. Draft Council of Europe Human Rights, Democracy and the Rule of Law Risk and Impact Assessment (HUDERIA)

1. Purpose: Assessing AI risks to human rights, democracy, and the rule of law.

2. Struktur: Meliputi analisis risiko berbasis konteks, keterlibatan pemangku kepentingan, penilaian dampak, dan mitigasi risiko.

3. Keselarasan: Sejalan dengan *DEFINE*, *ASSESS*, dan *TREAT*, namun belum maksimal dalam aspek *GOVERN*, seperti kurangnya komunikasi publik mengenai kepatuhan sistem AI terhadap regulasi, tata kelola, dan standar etika setelah dilakukan penilaian terhadap sistem AI, serta keterlibatan pimpinan untuk mengintegrasikan manajemen risiko ke dalam struktur organisasi.

g. IEEE 7000-21 Standard Model Process for Addressing Ethical Concerns during System Design (IEEE 7000-21)

1. Tujuan: Membantu mengintegrasikan pertimbangan standar etika dan pandangan pemangku kepentingan ke dalam desain AI.

2. Struktur: Menentukan pemangku kepentingan, mengantisipasi risiko, dan mengintegrasikan hasil konsultasi ke dalam desain produk atau layanan.

3. Keselarasan: Memetakan beberapa fungsi *GOVERN dari interoperability framework* ke IEEE 7000-21 tidak mudah. Pemantauan tidak didefinisikan sebagai proses terpisah, tetapi menjadi bagian integral dari semua langkah IEEE 7000-21. Dokumentasi proses manajemen risiko dan integrasinya ke dalam budaya organisasi disebutkan di berbagai bagian IEEE 7000-21, namun bukan fokus utama.

2. Structure: Covers context-based risk analysis, stakeholder engagement, impact assessment, and risk mitigation.

3. Alignment: In line with DEFINE, ASSESS, and TREAT, but lacking in the GOVERN aspect, such as insufficient public communication regarding AI system compliance with regulations, governance, and ethical standards after assessment of the AI system, as well as leadership involvement to integrate risk management into the organizational structure.

g. IEEE 7000-21 Standard Model Process for Addressing Ethical Concerns during System Design (IEEE 7000-21)

1. Purpose: Helping to integrate stakeholder ethical standards considerations and perspectives into AI design.

2. Structure: Identifying stakeholders, anticipating risks, and integrating consultation outcomes into the design of products or services.

3. Alignment: Mapping several GOVERN functions from the interoperability framework to IEEE 7000-21 will not be straightforward. Monitoring is not defined as a separate process but is an integral part of all steps in IEEE 7000-21. Documentation of the risk management process and its integration into organizational culture is mentioned in various sections of IEEE 7000-21, but it is not a primary focus.

h. ISO/IEC Guide 51:2014 3rd edition (ISO/IEC Guide 51)

1. Tujuan: Memberikan panduan dalam pengembangan standar keselamatan produk dengan fokus pada identifikasi, penilaian, dan mitigasi risiko.

2. Struktur: Berfokus pada mitigasi risiko dalam desain dan pengembangan produk.

3. Keselarasan: Panduan ISO/IEC 51 tidak memasukkan sebagian besar sub elemen *GOVERN*, seperti menanamkan kebijakan manajemen risiko, berkonsultasi dengan pemangku kepentingan dan mengomunikasikan upaya manajemen risiko.

h. ISO/IEC Guide 51:2014 3rd edition (ISO/IEC Guide 51)

1. Purpose: Providing guidance in the development of product safety standards with a focus on identification, assessment, and mitigation of risks.

2. Structure: Focusing on risk mitigation in the design and development of products.

3. Alignment: The ISO/IEC 51 guidelines do not include most of the GOVERN sub-elements, such as embedding risk management policies, consulting with stakeholders, and communicating risk management efforts.

B. Klasifikasi Risiko Sistem AI

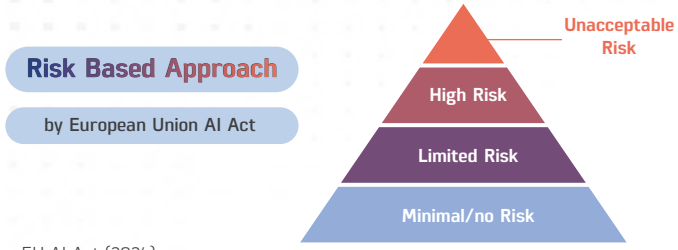
Dalam hal penerapan manajemen risiko AI, *European Union* (EU) *AI Act* menggunakan pendekatan *risk-based approach* dan mengklasifikasikan sistem AI berdasarkan tingkat risiko yang ditimbulkannya. Klasifikasi ini terdiri dari 4 (empat) kategori:

B. AI System Risk Classifications

In terms of the application of AI risk management, the European Union (EU) AI Act employs a risk-based approach and classifies AI systems based on the level of risk they pose. This classification consists of 4 (four) categories:

Gambar 13. Klasifikasi Risiko Sistem AI

Figure 13. AI System Risk Classifications



Source: EU AI Act (2024)

1. Risiko Tidak Dapat Diterima (*Unacceptable Risk*)

Sistem AI yang termasuk dalam kategori ini dianggap sangat berbahaya sehingga dilarang penggunaannya karena dapat mengancam hak asasi manusia, keselamatan, atau keamanan masyarakat. Contoh sistem dengan risiko yang tidak dapat diterima:

a. Manipulasi: AI yang memanipulasi perilaku manusia, terutama kelompok rentan, seperti mempromosikan disinformasi atau konten adiktif.

b. Penilaian sosial: Sistem yang digunakan untuk menilai individu berdasarkan perilaku atau karakteristik mereka, seperti sistem kredit sosial.

c. Pengawasan biometrik *real-time*: AI yang digunakan untuk pengawasan publik secara *real-time* tanpa dasar hukum yang jelas, seperti pengenalan wajah untuk pelacakan massal.

1. Unacceptable Risk

AI systems that fall into this category are considered highly dangerous and are prohibited from use as they may threaten human rights, safety, or public security. Examples of systems with unacceptable risks include:

a. Manipulation: AI that manipulates human behavior, particularly of vulnerable groups, such as promoting misinformation or addictive content.

b. Social scoring: Systems used to assess individuals based on their behavior or characteristics, such as social credit systems.

c. Real-time biometric surveillance: AI used for public surveillance in real-time without a clear legal basis, such as facial recognition for mass tracking.

2. Risiko Tinggi (*High Risk*)

Sistem AI dikategorikan sebagai risiko tinggi jika:

a. Sistem AI dimaksudkan untuk digunakan sebagai komponen keselamatan suatu produk, atau sistem AI itu sendiri merupakan produk.

b. Memerlukan penilaian dari pihak ketiga sebelum dapat dipasarkan atau digunakan.

c. Melakukan *profiling* individu, yaitu pemrosesan otomatis data pribadi untuk menilai berbagai aspek kehidupan seseorang, seperti kinerja kerja, situasi ekonomi, kesehatan, preferensi, minat, keandalan, perilaku, atau lokasi. Contoh sistem AI dengan risiko tinggi:

1. AI yang mendeteksi transaksi mencurigakan secara *real-time* dan otomatis memblokir atau meminta verifikasi tambahan jika ada potensi penipuan. Namun, jika sistem ini gagal, nasabah bisa mengalami kehilangan dana, dan bank bisa terkena risiko hukum serta reputasi.

2. AI yang menilai kelayakan kredit berdasarkan data historis dan pola perilaku peminjam dapat mencegah bank untuk memberikan kredit kepada individu atau bisnis

2. High Risk

AI systems are categorized as high risk if:

a. The AI system is intended to be used as a safety component of a product, or the AI system itself is a product.

b. Requires assessment by a third party before it can be marketed or used.

c. Additionally, AI systems are always considered high risk if they perform individual profiling, which is the automated processing of personal data to evaluate various aspects of a person's life, such as work performance, economic situation, health, preferences, interests, reliability, behavior, or location. Examples of high-risk AI systems include:

1. AI that detects suspicious transactions in real-time and automatically blocks or requests additional verification if there is potential fraud. However, if this system fails, customers may experience financial loss, and the bank could face legal and reputational risks.

2. AI that assesses creditworthiness based on historical data and borrower behavior patterns can prevent banks from extending credit to high-risk individuals or

yang berisiko tinggi. Namun, jika sistem ini tidak akurat, bank bisa mengalami kerugian finansial.

Berdasarkan *EU AI Act*, penyedia sistem AI dengan risiko tinggi harus:

1. Membangun sistem manajemen risiko di sepanjang siklus hidup sistem AI.
2. Melaksanakan tata kelola data, memastikan bahwa *dataset* pelatihan, validasi, dan pengujian relevan, cukup representatif, serta tidak terdapat kesalahan dan data lengkap sesuai dengan tujuan yang dimaksudkan.
3. Menyusun dokumentasi teknis untuk menunjukkan kepatuhan dan memberikan informasi kepada otoritas guna menilai kepatuhan tersebut.
4. Merancang sistem AI berisiko tinggi dengan mekanisme pencatatan untuk secara otomatis merekam peristiwa yang relevan dalam mengidentifikasi risiko.
5. Menyediakan instruksi penggunaan bagi pihak yang menggunakan sistem AI untuk memastikan pemenuhan kepatuhan.
6. Merancang sistem AI berisiko tinggi agar memungkinkan penggunanya menerapkan pengawasan oleh manusia.
7. Merancang sistem AI berisiko tinggi guna mencapai tingkat akurasi, ketahanan, dan keamanan siber yang sesuai.
8. Membangun sistem manajemen mutu untuk memastikan kepatuhan.

businesses. However, if this system is inaccurate, the bank may incur financial losses.

According to the EU AI Act, providers of high-risk AI systems must:

1. Establish a risk management system throughout the AI system lifecycle.
2. Implement data governance, ensuring that training, validation, and testing datasets are relevant, sufficiently representative, and free from errors while being complete according to the intended purpose.
3. Prepare technical documentation to demonstrate compliance and provide information to authorities for assessing that compliance.
4. Design high-risk AI systems with logging mechanisms to automatically record relevant events in identifying risks.
5. Provide usage instructions for parties using the AI system to ensure compliance is met.
6. Design high-risk AI systems to allow users to implement human oversight.
7. Design high-risk AI systems to achieve appropriate levels of accuracy, robustness, and cybersecurity.
8. Establish a quality management system to ensure compliance.

3. Risiko Terbatas (*Limited Risk*)

Sistem AI dengan risiko terbatas melibatkan masalah transparansi. Oleh karena itu, *EU AI Act* mewajibkan penyedia sistem AI untuk menerapkan prinsip transparansi guna membangun kepercayaan pengguna terhadap sistem AI tersebut. Contoh sistem AI dengan risiko terbatas:

a. Bank menggunakan *chatbot* berbasis AI untuk memberikan layanan pelanggan. Jika *chatbot* tidak mengungkapkan dengan jelas bahwa pengguna sedang berbicara dengan AI, pelanggan dapat mengira mereka berinteraksi dengan manusia.

b. Aplikasi perbankan yang menggunakan AI untuk memberi saran otomatis tentang pengelolaan keuangan pribadi berdasarkan pola pengeluaran pengguna. Tanpa prinsip transparansi, nasabah mungkin menganggap rekomendasi AI sebagai saran keuangan profesional, padahal AI

3. Limited Risk

AI systems with limited risk involve transparency issues. Therefore, the EU AI Act requires providers of AI systems to implement transparency principles to build user trust in these AI systems. Examples of AI systems with limited risk include:

a. Banks use AI-based chatbots to provide customer service. If the chatbot does not clearly disclose that users are interacting with AI, customers may mistakenly believe they are interacting with a human.

b. Banking applications that use AI to provide automated advice on personal financial management based on users' spending patterns. Without transparency principles, customers may perceive AI recommendations as professional financial advice, whereas the AI is solely based on data patterns

hanya berdasarkan pola data tanpa mempertimbangkan kondisi unik pengguna secara holistik.

Adapun contoh tindakan yang harus diambil:

a. Pengguna harus diberitahu bahwa mereka sedang berinteraksi dengan sistem AI.

b. Konten yang dihasilkan oleh AI harus dapat diidentifikasi dengan jelas sebagai konten buatan AI, misalnya dengan pemberian label.

4. Risiko Minimal/Tidak Ada (Minimal/*No Risk*)

Kategori ini mencakup sebagian besar sistem AI yang telah digunakan saat ini dan biasanya tidak diperlukan regulasi tertentu. Sistem ini memiliki dampak rendah yang tidak memengaruhi keselamatan atau hak dasar pengguna. Contohnya adalah:

a. *Spam filter* atau aplikasi video *game*.

b. AI yang secara otomatis mengategorikan transaksi pengguna (misalnya, belanja, tagihan, transportasi) untuk mempermudah pelacakan pengeluaran.

without considering the unique circumstances of the user holistically.

The following are examples of actions that must be taken:

a. Users must be informed that they are interacting with an AI system.

b. Content generated by AI must be clearly identifiable as AI-generated content, for example, by providing labels.

4. Minimal/No Risk

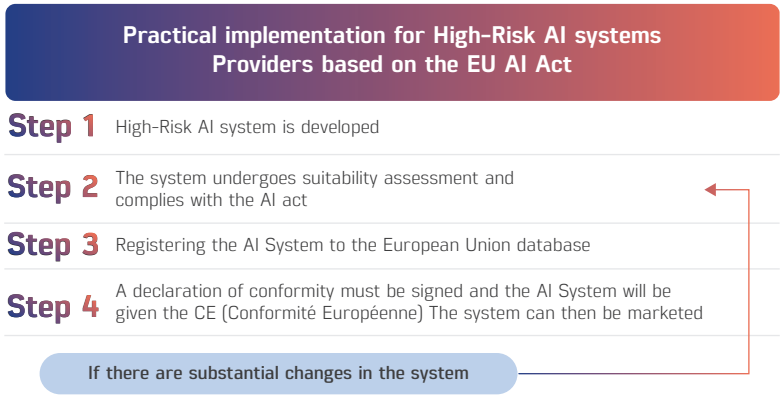
This category includes most AI systems that are currently in use and typically do not require specific regulation. These systems have a low impact that does not affect the safety or fundamental rights of users. Examples include:

a. Spam filter or video game applications.

b. AI that automatically categorizes user transactions (e.g., shopping, bills, transportation) to facilitate expense tracking.

Gambar 14. Langkah Implementasi bagi Penyedia Sistem AI Berisiko Tinggi

Figure 14. Implementation Steps for Providers of High-Risk AI Systems



Source: EU AI Act (2024)

Untuk sistem AI berisiko tinggi (*high-risk*), *EU AI Act* menjelaskan langkah-langkah implementasi praktik manajemen risikonya sebagai berikut:

1. Langkah 1: Sistem AI dikembangkan dan dirancang untuk tujuan tertentu yang berada di bawah kategori berisiko tinggi, seperti AI yang digunakan dalam layanan keuangan, perawatan kesehatan, transportasi, dan sebagainya.

2. Langkah 2: Sistem tersebut harus menjalani penilaian kesesuaian untuk memastikan sistem memenuhi persyaratan yang ditetapkan dalam *EU AI Act*.

For high-risk AI systems, the EU AI Act outlines the implementation steps for risk management practices as follows:

1. Step 1: AI systems are developed and designed for specific purposes that fall under the high-risk category, such as AI used in financial services, healthcare, transportation, etc.

2. Step 2: These systems must undergo a suitability assessment to ensure that they meet the requirements established in the EU AI Act.

3. Langkah 3: Setelah memenuhi persyaratan dimaksud, sistem AI harus didaftarkan dalam "*database*" resmi Uni Eropa untuk memastikan pencatatan dan pengawasannya.

4. Langkah 4: Deklarasi kesesuaian harus ditandatangani yang menegaskan bahwa sistem AI mematuhi semua persyaratan peraturan, dan diberikan tanda CE (*Conformité Européenne*) yang menunjukkan bahwa sistem telah memenuhi standar EU AI Act.

3. Step 3: After meeting the specified requirements, the AI system must be registered in an official European Union database to ensure proper documentation and oversight.

4. Step 4: A declaration of conformity must be signed, affirming that the AI system complies with all regulatory requirements, and a CE (Conformité Européenne) mark must be provided to indicate that the system meets the standards of the EU AI Act.

Jika terdapat perubahan/modifikasi substansial pada sistem AI (misalnya perubahan algoritma, set data, atau fungsionalitas), penyedia sistem AI harus kembali ke Langkah 2 dan menilai kembali kesesuaian untuk memastikan kepatuhannya terhadap peraturan.

If there are substantial changes/ modifications to the AI system (e.g., changes in algorithms, datasets, or functionalities), the AI system provider must return to step 2 and reassess suitability to ensure compliance with regulations.

C. Peran Direksi dan Komisaris dalam Tata Kelola AI

Direksi dan Dewan Komisaris memiliki peranan penting dan substantif dalam Tata Kelola AI untuk mewujudkan "*Trustworthy AI*" di masing-masing bank.

1. Memastikan pemilik risiko dan peran serta tanggung jawab terkait AI didefinisikan dengan jelas dan bahwa individu tersebut memiliki keahlian dan sumber daya yang sesuai untuk menjalankan peran tersebut dengan benar.

2. Memahami sistem AI yang kritikal dan berisiko tinggi yang digunakan dan diterapkan di seluruh bisnis, beserta karakteristik, sumber, dan keandalan data yang digunakan untuk melatih sistem berisiko tinggi.

3. Memahami strategi AI perusahaan dan keselarasannya dengan strategi bisnis yang lebih luas.

C. Roles of the Board of Directors and Commissioners in AI Governance

The Board of Directors and the Board of Commissioners have an important and substantive role in AI Governance to realize "Trustworthy AI" in each bank.

1. Ensuring that risk owners and roles, as well as responsibilities related to AI, are clearly defined and that individuals have the appropriate expertise and resources to perform those roles correctly.

2. Understanding the critical and high-risk AI systems used and implemented across the business, along with the characteristics, sources, and reliability of the data used to train these high-risk systems.

3. Understanding the company's AI strategy and its alignment with the broader business strategy.

4. Memahami bagaimana bisnis memastikan bahwa masalah etika yang terlibat dalam penggunaan AI diidentifikasi dan ditangani, terutama bias dan diskriminasi.
5. Memastikan kepatuhan terhadap program manajemen risiko AI diaudit oleh fungsi audit sesuai dengan peran lini ketiga.
6. Memastikan pemilik risiko AI secara teratur meninjau efektivitas program dan kebijakan manajemen risiko AI.
7. Menetapkan atau menyetujui toleransi risiko terhadap sistem AI.
8. Memahami profil risiko AI Perusahaan.
9. Memahami mekanisme penilaian risiko terhadap sistem AI dan penggunaannya–termasuk kriteria serta proses evaluasi–serta mengetahui jenis penggunaan yang dilarang beserta alasannya.
10. Memastikan ada proses bagi manajemen untuk meningkatkan dan memberi pengarahan kepada Dewan tentang setiap insiden AI, termasuk tanggapan organisasi, dampak, status investigasi dan pembelajaran apa pun yang diidentifikasi sebagai bagian dari tinjauan pasca insiden.

4. Understanding how the business ensures that ethical issues involved in the use of AI are identified and addressed, particularly bias and discrimination.
5. Ensuring that compliance with the AI risk management program is audited by the audit function in accordance with the third line of defense role.
6. Ensuring that AI risk owners regularly review the effectiveness of the AI risk management program and policies.
7. Establishing or approving risk tolerance for AI systems.
8. Understanding the company's AI risk profile.
9. Understanding how AI systems and use cases are assessed for risk (i.e., assessment criteria and assessment processes), and which ones have been prohibited and why.
10. Ensuring that there is a process for management to escalate and brief the board on any AI incidents, including the organization's response, any impacts, the status of investigations, and any learnings identified as part of the post-incident review.
11. Memahami legalitas penggunaan dan penerapan AI, termasuk pengumpulan dan penggunaan data pelatihan, di seluruh bisnis.
12. Memahami *trade-off* yang dibuat dalam keputusan yang melibatkan AI (misalnya akurasi vs. keadilan, interpretabilitas vs. privasi, akurasi vs. privasi, akurasi vs. adaptabilitas).
13. Memastikan AI menjadi agenda rapat Dewan Komisaris dan Direksi secara berkala, baik pada rapat Dewan Komisaris dan Direksi penuh maupun rapat komite risiko, dan bahwa Dewan Komisaris dan Direksi memiliki akses yang memadai terhadap keahlian AI.

Dalam konteks Indonesia, peran Direksi dan Dewan Komisaris, harus diselaraskan dengan penerapan fungsi Direksi dan Dewan Komisaris sebagaimana peraturan perundang-undangan yang berlaku dan penerapan tata kelola sebagaimana diatur dalam Peraturan Otoritas Jasa Keuangan (POJK) No. 17 Tahun 2023 tentang Penerapan Tata Kelola Bagi Bank Umum dan POJK No. 11/POJK.03/2022 tentang Penyelenggaraan Teknologi Informasi oleh Bank Umum. Selain peran Dewan Komisaris dan Direksi, pelaksanaan tata kelola AI perlu didukung oleh Komite

11. Understanding the legality of the use and application of AI, including the collection and use of training data, across the business.
12. Understanding the trade-offs made in decisions involving AI (e.g., accuracy vs. fairness, interpretability vs. privacy, accuracy vs. privacy, accuracy vs. adaptability).
13. Ensuring that AI is a regular agenda item for the Board of Commissioners and the Board of Directors, both in full board meetings and risk committee meetings, and that the Board of Commissioners and the Board of Directors have adequate access to AI expertise.

In the context of Indonesia, the roles of the Board of Directors and the Board of Commissioners must be aligned with the implementation of their functions as stipulated by applicable laws and regulations, as well as governance practices outlined in Financial Services Authority (OJK) Regulation No. 17 of 2023 concerning Governance Implementation for Commercial Banks and OJK Regulation No. 11/POJK.03/2022 regarding Information Technology Implementation by Commercial Banks. In addition to the roles of the Board of Commissioners and Directors, AI

AI. Komite AI dapat beranggotakan perwakilan dari bagian/divisi fungsional utama termasuk Hukum, Kepatuhan, Risiko, Pengembangan Produk, Pengadaan, *Data Science*, Keamanan Siber, Pemasaran, dan Layanan Pelanggan. Komite AI memiliki peran antara lain:

1. Mengawasi desain dan peluncuran kerangka tata kelola AI perusahaan.
2. Mendefinisikan peran dan tanggung jawab utama terkait pengawasan, desain, pengembangan, dan penggunaan AI di seluruh bisnis.
3. Membuat prinsip panduan AI.
4. Mendefinisikan dan mendokumentasikan cakupan program tata kelola AI (termasuk jenis model, algoritma, dan sistem mana yang termasuk dan tidak termasuk dalam cakupan beserta alasannya, serta membangun skala risiko untuk penggunaan dalam cakupan AI).
5. Mengidentifikasi dan mengawasi kebijakan, proses, dan pelatihan untuk memungkinkan desain, penggunaan, dan pengawasan AI secara bertanggung jawab.
6. Mengidentifikasi area yang memerlukan *reviu* atau pengawasan manusia, termasuk untuk mengidentifikasi ketidakakuratan, mengidentifikasi bias, dan melakukan *quality assurance* (jaminan kualitas).
7. Mengembangkan proses untuk menilai dan mengeskalasi kasus penggunaan AI berisiko tinggi.
8. Melaporkan kepada manajemen perusahaan (dewan dan manajemen senior).
9. Membantu dalam mengelola insiden yang berkaitan dengan penggunaan AI.

Dalam konteks perbankan Indonesia, Komite AI dapat menjadi bagian dari fungsi Komite Pengarah TI sebagaimana yang telah diatur dalam POJK No. 11/POJK.03/2022 tentang Penyelenggaraan Teknologi Informasi oleh Bank Umum atau dibentuk menjadi komite tersendiri yang khusus menangani teknologi AI. Hal ini disesuaikan dengan tingkat kompleksitas adopsi AI di bank, dalam hal bank melakukan adopsi AI untuk fungsi-fungsi yang kritikal dengan rasio adopsi yang tinggi maka bank dapat membentuk Komite AI secara tersendiri.

governance needs to be supported by an AI Committee. The AI Committee may consist of representatives from key functional areas including Legal, Compliance, Risk, Product Development, Procurement, Data Science, Cybersecurity, Marketing, and Customer Service. The AI Committee has several roles including:

1. Supervising the design and launch of the company's AI governance framework.
2. Defining key roles and responsibilities related to the oversight, design, development, and use of AI across the business.
3. Creating guiding principles for AI.
4. Defining and documenting the scope of the AI governance program (including which types of models, algorithms, and systems are included or excluded from scope along with justifications, as well as establishing a risk scale for use within AI scope).
5. Identifying and overseeing policies, processes, and training to enable responsible design, use, and oversight of AI.
6. Identifying areas that require human review or oversight, including identifying inaccuracies, detecting bias, and conducting quality assurance (QA).
7. Developing processes to assess and escalate high-risk AI use cases.
8. Reporting to company management (the board and senior management).
9. Assisting in managing incidents related to the use of AI.

In the context of Indonesian banking, the AI Committee can be part of the IT Steering Committee function as regulated in OJK Regulation No. 11/POJK.03/2022 regarding Information Technology Implementation by Commercial Banks or can be established as a separate committee specifically handling AI technology. This is adjusted according to the level of complexity in AI adoption at the bank; if the bank adopts AI for critical functions with a high adoption ratio, then it may establish a separate AI Committee.

HALAMAN INI SENGAJA DIKOSONGKAN

THIS PAGE IS INTENTIONALLY LEFT BLANK

Bab 6
Panduan Implementasi Tata Kelola Kecerdasan Artifisial

Chapter 6
Artificial Intelligence Governance Implementation Guidelines

Kecerdasan artifisial (AI) sebagai salah satu *emerging technology* dan disebut juga sebagai *transformative technology,* telah menjadi perhatian khusus dari berbagai pihak termasuk pemerintah, industri, akademisi, dan masyarakat luas, sehubungan dengan potensi dan peluang yang ditawarkan AI serta risiko yang menyertainya. Sistem AI, termasuk sistem AI dengan model yang paling terdepan (*advanced AI systems*) telah menjadi *tools* dan katalisator yang menjadi fokus dari berbagai institusi/organisasi termasuk bank dalam upaya untuk mengoptimalkan kinerja bisnis dan operasionalnya serta mempertahankan eksistensi di lingkungan bisnis yang semakin kompleks.

Pada 30 Januari 2025, HLB International Limited merilis *Survey of Business Leaders 2025.* Survei secara kuantitatif dan kualitatif tersebut dilakukan pada periode bulan September dan November 2024 kepada 1.242 *business leaders* yang tersebar di 55 negara. Hasil survei menunjukkan bahwa organisasi yang berkinerja tinggi (*high-performing organisations* atau disebut sebagai *Profit Accelerators*) di tahun 2025 akan berfokus pada tiga aspek untuk mencapai profitabilitas maksimal, yakni:

Artificial Intelligence (AI), as one of the emerging technologies and also referred to as transformative technology, has garnered special attention from various parties including the government, industry, academia, and the general public, in relation to the potential and opportunities offered by AI as well as the accompanying risks. AI systems, including advanced AI systems, have become tools and catalysts that are a focus for various institutions/organizations including banks in their efforts to optimize business performance and operations while maintaining their existence in an increasingly complex business environment.

On January 30, 2025, HLB International Limited released the Survey of Business Leaders 2025. This quantitative and qualitative survey was conducted between September and November 2024 with 1,242 business leaders across 55 countries. The survey results indicate that high-performing organizations (referred to as Profit Accelerators) in 2025 will focus on three aspects to achieve maximum profitability, namely:

- Efisiensi operasional, 65% *Profit Accelerators* memprioritaskan transformasi operasional skala besar seperti memodernisasi sistem teknologi (52%) dan menyederhanakan proses operasional (45%).
- Inovasi, 67% pemimpin mengidentifikasi AI sebagai teknologi terpenting untuk lima tahun ke depan. *Profit Accelerators* telah memanfaatkan AI dalam analisis prediktif, *rapid prototyping*, dan personalisasi pelanggan berbasis data.
- Investasi SDM, 60% *Profit Accelerator* berfokus pada pembelajaran dan pengembangan (*learning and development*) untuk meningkatkan efektivitas tenaga kerja.

- Operational efficiency, 65% of Profit Accelerators prioritize large-scale operational transformation such as modernizing technology systems (52%) and streamlining operational processes (45%).
- Innovation, 67% of leaders identify AI as the most important technology for the next five years. Profit Accelerators have leveraged AI in predictive analytics, rapid prototyping, and data-driven customer personalization.
- Human Resource Investment, 60% of Profit Accelerators focus on learning and development to enhance workforce effectiveness.

Sementara itu, KPMG merilis *KPMG AI Quarterly Pulse Survey* (2024) dimana survei dilaksanakan pada 7 November-9 Desember 2024 kepada 100 eksekutif dan pemimpin terkemuka di AS. Hasil survei ini menunjukkan bahwa AI mengubah industri dan menjadi prioritas utama bagi berbagai perusahaan. Organisasi terus

Meanwhile, KPMG released the KPMG AI Quarterly Pulse Survey (2024), which was conducted from November 7 to December 9, 2024, with 100 executives and leading leaders in the U.S. The results of this survey indicate that AI is transforming industries and becoming a top priority for various companies. Organizations continue to adjust their

menyesuaikan strategi serta investasi yang terus meningkat dalam *Generative AI* (GenAI). Disamping itu, tantangan seperti faktor makroekonomi, privasi data, dan kualitas data tetap menjadi perhatian utama dalam implementasi AI ke depan. Secara detail, survei tersebut menunjukkan bahwa:

1. AI akan mengubah bisnis secara fundamental: 56% eksekutif percaya AI akan mengubah bisnis mereka dalam 1 tahun ke depan, meningkat menjadi 67% dalam dua tahun ke depan.
2. Investasi AI semakin besar: 68% pemimpin bisnis akan menginvestasikan $50-$250 juta dalam GenAI dalam 12 bulan ke depan, meningkat dari 45% pada Q1-2024.
3. Faktor makroekonomi berpengaruh besar: 88% pemimpin bisnis menyebut faktor makroekonomi sebagai tantangan utama dalam strategi AI.
4. *AI Agents* mulai diuji coba: 51% perusahaan sedang mengeksplorasi penggunaan *AI Agents*, sementara 37% melakukan uji coba (*piloting*). Saat ini, hanya 12% organisasi yang telah menerapkan AI *Agents* dalam operasional mereka.

strategies and increase investments in Generative AI (GenAI). Additionally, challenges such as macroeconomic factors, data privacy, and data quality remain key concerns in the future implementation of AI. Specifically, the survey shows that:

1. AI will fundamentally change business: 56% of executives believe AI will transform their business within the next year, increasing to 67% in the next two years.
2. AI investments are increasing: 68% of business leaders will invest $50-$250 million in GenAI in the next 12 months, up from 45% in Q1-2024.
3. Macroeconomic factors have a significant impact: 88% of business leaders cite macroeconomic factors as a major challenge in their AI strategy.
4. AI Agents are beginning to be tested: 51% of companies are exploring the use of AI Agents, while 37% are conducting pilot tests. Currently, only 12% of organizations have implemented AI Agents in their operations.

5. Peran CIO dalam AI meningkat: 71% *Chief Information Officers* (CIO) kini memimpin inisiatif AI, diikuti oleh CEO (17%) dan *Chief Innovation Officer* (10%). Perubahan ini cukup signifikan dibandingkan awal tahun 2024, di mana 49% CEO masih menjadi pemimpin utama dalam AI.

Di sektor perbankan, kombinasi dari *tools* AI (*Predictive AI, Generative AI* maupun *advanced AI systems*, dengan penggunaan *machine learning/ deep learning*) berpotensi mengubah industri perbankan dengan mendorong inovasi, memberdayakan pengambilan keputusan yang lebih cerdas serta menciptakan pengalaman yang lebih personal dan menarik. Namun demikian, pengembangan dan penerapannya harus dilakukan secara etis dan bertanggung jawab untuk memaksimalkan manfaatnya sekaligus memitigasi potensi risiko.

Implementasi AI yang bertanggung jawab sangat diperlukan, termasuk pada sektor jasa keuangan khususnya pada industri perbankan. *Financial Stability Board* (FSB) pada 14 November 2024 menerbitkan laporan mengenai *The Financial Stability Implications of Artificial Intelligence*, yang menguraikan perkembangan terkini terkait penerapan kecerdasan artifisial di sektor keuangan dan implikasinya terhadap stabilitas keuangan. Dalam laporan ini diuraikan bahwa:

1. AI menawarkan manfaat dari peningkatan efisiensi operasional, kepatuhan terhadap peraturan, produk keuangan yang dipersonalisasi, dan analisis data tingkat lanjut.
2. Namun demikian, AI juga dapat menambah kerentanan sektor keuangan tertentu karena berpotensi untuk meningkatkan risiko sistemik yang akan menimbulkan risiko terhadap stabilitas keuangan.
3. Beberapa kerentanan terkait AI a.l. ketergantungan pihak ketiga dan konsentrasi penyedia layanan, korelasi pasar, risiko siber, dan risiko model, kualitas data dan tata kelola.
4. Selain itu, *Generative AI* dapat meningkatkan penipuan keuangan dan disinformasi di pasar keuangan. Sistem AI yang tidak selaras dan tidak dikalibrasi untuk beroperasi dalam batasan hukum, peraturan, dan etika juga dapat berimplikasi mengganggu stabilitas keuangan.

5. The role of the CIO in AI is increasing: 71% of Chief Information Officers (CIOs) are now leading AI initiatives, followed by CEOs at 17% and Chief Innovation Officers at 10%. This change is quite significant compared to early 2024, when 49% of CEOs were still the primary leaders in AI.

In the banking sector, the combination of AI tools (Predictive AI, Generative AI, and advanced AI systems, along with the use of machine learning/deep learning) has the potential to transform the banking industry by driving innovation, empowering smarter decision-making, and creating more personalized and engaging experiences. However, development and implementation must be conducted ethically and responsibly to maximize benefits while mitigating potential risks.

Responsible AI implementation is essential, particularly in the financial services sector, especially within the banking industry. The Financial Stability Board (FSB) published a report on November 14, 2024, titled "The Financial Stability Implications of Artificial Intelligence," which outlines recent developments related to the application of AI in the financial sector and its implications for financial stability. This report details that:

1. AI offers benefits such as enhanced operational efficiency, regulatory compliance, personalized financial products, and advanced data analytics.
2. However, AI can also increase vulnerabilities in certain financial sectors as it has the potential to elevate systemic risks that could pose threats to financial stability.
3. Several vulnerabilities related to AI include third-party dependency and service provider concentration, market correlation, cybersecurity risks, model risk, data quality issues, and governance challenges.
4. Additionally, Generative AI can increase financial fraud and misinformation in the financial markets. Misaligned and uncalibrated AI systems that operate outside legal, regulatory, and ethical boundaries can also disrupt financial stability.

5. Dari perspektif jangka panjang, penggunaan AI dapat mendorong perubahan dalam struktur pasar, kondisi ekonomi makro dan penggunaan energi, yang mungkin berimplikasi terhadap pasar dan lembaga keuangan.

6. Laporan tersebut mencatat bahwa kerangka kerja peraturan dan pengawasan yang ada mengatasi banyak kerentanan yang terkait dengan adopsi AI. Namun diperlukan upaya untuk memastikan bahwa kerangka kerja yang ada cukup komprehensif. Karenanya, laporan tersebut menghimbau FSB, *standard-setting bodies*, dan otoritas nasional untuk:
   - Mempertimbangkan cara mengatasi kesenjangan data dan informasi agar dapat memantau adopsi AI dengan lebih baik dan menilai implikasinya terhadap stabilitas keuangan.
   - Menilai apakah kerangka kebijakan keuangan saat ini cukup untuk mengatasi kerentanan terkait AI baik di tingkat domestik maupun internasional.
   - Meningkatkan kemampuan regulasi dan pengawasan, misalnya dengan berbagi informasi dan praktik terbaik lintas batas dan lintas sektor, serta memanfaatkan perangkat yang didukung AI.

5. From a long-term perspective, the use of AI can drive changes in market structure, macroeconomic conditions, and energy usage, which may have implications for markets and financial institutions.

6. The report notes that the existing regulatory and supervisory framework addresses many vulnerabilities associated with AI adoption. However, efforts are needed to ensure that the existing framework is sufficiently comprehensive. Therefore, the report calls on the FSB, standard-setting bodies, and national authorities to:
   - Consider ways to address data and information gaps to better monitor AI adoption and assess its implications for financial stability.
   - Assess whether the current financial policy framework is adequate to address AI-related vulnerabilities at both domestic and international levels.
   - Enhance regulatory and supervisory capabilities, for example by sharing information and best practices across borders and sectors, as well as leveraging AI-supported tools.

European Central Bank pada Mei 2024 juga menegaskan bahwa penerapan AI di seluruh sistem keuangan perlu dipantau secara ketat seiring dengan perkembangan teknologi. Inisiatif regulasi mungkin perlu dipertimbangkan untuk antisipasi kegagalan pasar dan tidak dapat diatasi oleh kerangka kehati-hatian saat ini. Selain dapat memberikan manfaat dan risiko di tingkat lembaga keuangan serta untuk seluruh sistem keuangan, AI bermanfaat untuk mendorong kemajuan ekonomi yang menguntungkan konsumen, bisnis, dan ekonomi secara keseluruhan, serta juga meningkatkan efisiensi intermediasi keuangan melalui pemrosesan informasi yang lebih cepat dan komprehensif yang mendukung pengambilan keputusan, yang dapat memperkuat sistem keuangan dan berkontribusi pada stabilitas keuangan. Namun demikian, AI juga memiliki potensi risiko, antara lain:

- Meningkatkan risiko yang terkait dengan bias, halusinasi, penyalahgunaan, yang dapat mendistorsi hasil pasar keuangan, merusak ketahanan kerangka operasional, atau secara sistematis membiaskan pemrosesan informasi dan manajemen risiko atau pengambilan keputusan lembaga.

The European Central Bank in May 2024 also emphasized that the implementation of AI across the financial system needs to be closely monitored as technology evolves. Regulatory initiatives may need to be considered to anticipate market failures that cannot be addressed by the current prudential framework. AI can provide benefits and risks at both the financial institution level and for the entire financial system. AI is beneficial for driving economic progress that benefits consumers, businesses, and the economy as a whole while also enhancing financial intermediation efficiency through faster and more comprehensive information processing that supports decision-making, which can strengthen the financial system and contribute to financial stability. However, AI also has potential risks, including:

- Increasing risks associated with bias, hallucinations, and misuse that can distort financial market outcomes, undermine the resilience of operational frameworks, or systematically bias information processing and risk management or decision-making within institutions.

- Implikasi sistemik AI akan bergantung pada tingkat penetrasi teknologi dan konsentrasi pemasok, yang sulit diprediksi.
- Teknologi AI dan penggunaannya di sektor keuangan masih terus berkembang. Lebih jauh, pertimbangan seperti dampak AI yang lebih luas terhadap ekonomi makro dan iklim serta aspek moral dan etika dari penyalahgunaan AI, perlu dieksplorasi lebih lanjut.
- Dapat berdampak pada kepercayaan publik terhadap intermediasi keuangan, yang merupakan landasan stabilitas keuangan.

Untuk kondisi di Indonesia, berdasarkan hasil riset yang melibatkan Kadin Indonesia pada Agustus 2024 diperoleh kondisi bahwa 61% institusi keuangan di Indonesia yakin bahwa mereka memiliki kesiapan teknologi yang kuat serta dengan data dan teknologi yang telah mapan untuk mengimplementasikan *Generative* AI (GenAI). Disamping itu, diyakini bahwa potensi GenAI di sektor keuangan Indonesia sangat luas, dimana teknologi ini antara lain dapat memperluas akses keuangan, meningkatkan pengalaman pelanggan, memperluas skala layanan dengan cepat. Hasil riset tersebut juga menunjukkan bahwa baik institusi keuangan besar maupun *startup fintech* telah dengan cepat mengadopsi teknologi GenAI, namun banyak inisiatif tersebut masih dalam tahap uji coba dan belum memberikan nilai bisnis yang signifikan dalam skala besar. Manfaat utama yang diharapkan dari GenAI adalah peningkatan efisiensi operasional, seperti otomatisasi tugas-tugas dasar.

Untuk dapat mewujudkan penerapan AI yang bertanggung jawab di sektor perbankan Indonesia, diperlukan panduan yang dapat menjadi acuan minimal bagi sektor perbankan dalam melakukan implementasi sistem AI termasuk *advanced AI systems,* sehingga dapat memberikan manfaat yang luas dan dengan risiko yang terkendali.

- The systemic implications of AI will depend on the level of technology penetration and supplier concentration, which are difficult to predict.
- AI technology and its use in the financial sector continue to evolve. Furthermore, considerations such as the broader impact of AI on macroeconomics and climate, as well as the moral and ethical aspects of AI misuse, need to be further explored.
- It can impact public trust in financial intermediation, which is the foundation of financial stability.

For the situation in Indonesia, based on research involving Kadin Indonesia in August 2024, it was found that 61% of financial institutions in Indonesia believe they have strong technological readiness along with established data and technology to implement Generative AI (GenAI). Furthermore, it is believed that the potential of GenAI in the Indonesian financial sector is vast; this technology can expand access to finance, enhance customer experience, and rapidly scale services. The research results also indicate that both large financial institutions and fintech startups have quickly adopted GenAI technology; however, many of these initiatives are still in the pilot stage and have not yet provided significant business value on a large scale. The main benefits expected from GenAI are increased operational efficiency such as automating basic tasks.

To realize the responsible implementation of AI in the Indonesian banking sector, guidelines are needed that can serve as a minimum reference for the banking sector in implementing AI systems, including advanced AI systems, so that they can provide broad benefits while keeping risks under control.

A. Motivasi Impelementasi AI

Secara umum aspek-aspek yang menjadi motivasi dan alasan bagi organisasi untuk mengimplementasikan AI adalah karena pertimbangan sebagai berikut (Lee, dkk, 2023).

a. Efisiensi, produktivitas, dan proses (mengoptimalkan proses internal dan bisnis, meningkatkan efisiensi, memungkinkan karyawan untuk melakukan pengendalian, meningkatkan produktivitas).

b. Pengambilan keputusan (mendapatkan keputusan yang lebih baik, membuat keputusan lebih cepat dan lebih informatif, menghasilkan prediksi yang lebih akurat untuk mendukung pengambilan keputusan).

c. Layanan pelanggan dan engagement (keterikatan) (meningkatkan pengalaman pelanggan/ *customer experience*, meningkatkan hubungan pelanggan/*customer relations*, meningkatkan kunjungan web perusahaan, menghemat waktu pelanggan, meningkatkan layanan).

d. Optimalisasi biaya dan sumber daya (mengurangi biaya dengan menggantikan tenaga ahli manusia yang mahal, mengkonfigurasi/ mengkonfigurasi ulang dan mengoptimalkan sumber daya perusahaan, mengurangi dan mengendalikan biaya serta jumlah karyawan).

e. Kapasitas pemrosesan dan analisis data (memproses jumlah data yang terus bertambah, meningkatkan kapasitas untuk analisis data, memanfaatkan semua jenis data dalam dan antar organisasi).

f. Otomasi (mengotomatisasi analisis, operasi, proses, prosedur, dan antarmuka pelanggan).

g. Manajemen pengetahuan dan informasi (mendukung dan mendorong akuisisi,

A. Motivations for AI Impelementation

In general, the aspects that motivate and justify organizations to implement AI are due to the following considerations (Lee et al., 2023).

a. Efficiency, productivity, and processes (optimizing internal and business processes, enhancing efficiency, enabling employees to exercise control, increasing productivity).

b. Decision-making (achieving better decisions, making decisions faster and more informatively, generating more accurate predictions to support decision-making).

c. Customer service and engagement (increasing customer experience, enhancing customer relations, boosting company website visits, saving customers' time, improving services).

d. Cost and resource optimization (reducing costs by replacing expensive human experts, configuring/reconfiguring and optimizing company resources, reducing and controlling costs as well as the number of employees).

e. Data processing and analysis capacity (processing the growing amount of data, enhancing capacity for data analysis, leveraging all types of data within and across organizations).

f. Automation (automating analysis, operations, processes, procedures, and customer interfaces).

g. Knowledge and information management (supporting and promoting the acquisition,

pengelolaan, penerapan, dan penyerapaan pengetahuan, serta mengoptimalkan pengumpulan, penyimpanan, pemrosesan, dan penyebaran informasi).

h. Pengembangan produk (meningkatkan produk, menciptakan produk baru).

i. Kualitas dan akurasi (meningkatkan kualitas *insight* yang dihasilkan, menghilangkan bias manusia, mengurangi kesalahan).

j. Pendapatan dan keuntungan (meningkatkan pendapatan, meningkatkan profitabilitas).

k. Dukungan pasar (menghadapi tantangan dari pendatang baru yang mengubah *status quo*, mencari pasar baru).

l. Regulasi (memenuhi peraturan baru).

m. Dukungan terhadap karyawan (mengurangi beban pekerjaan manual, mempermudah dan membantu karyawan dalam mengambil keputusan atau melaksanakan tugas dengan lebih efektif).

n. Meningkatkan pekerjaan dan kesesuaian keterampilan staf (memberikan kesempatan kepada karyawan untuk pekerjaan yang lebih maju, kreatif, dan bernilai tinggi, mengatasi ketidakcocokan personel).

management, application, and absorption of knowledge while optimizing the collection, storage, processing, and dissemination of information).

h. Product development (enhancing existing products, creating new products).

i. Quality and accuracy (improving the quality of generated insights, eliminating human bias, reducing errors).

j. Revenue and profit (increasing revenue, enhancing profitability).

k. Market support (facing challenges from newcomers that disrupt the status quo, seeking new markets).

l. Regulation (complying with new regulations).

m. Employee support (reducing the burden of manual work, facilitating and assisting employees in making decisions or performing tasks more effectively).

n. Enhancing jobs and staff skill alignment (providing opportunities for employees to engage in more advanced, creative, and high-value work, addressing personnel mismatches).

B. Tantangan Implementasi AI

Tantangan utama dalam implementasi AI diidentifikasi dan dikategorikan ke dalam empat dimensi yakni organisasi, sistem informasi, teknologi, dan sumber daya manusia (Lee, dkk, 2023).

a. Organisasi, terkait dengan budaya (adanya resistensi di berbagai tingkatan, kesalahpahaman pemimpin, kurangnya kelincahan), keuangan (biaya dan investasi yang tinggi, dana yang tidak mencukupi), pengetahuan (kurangnya pengetahuan tentang waktu, teknologi, data, kemampuan, dan tingkat penggunaan), perberbedaan ekspektasi di berbagai tingkat organisasi, kurangnya kepemimpinan, tidak tersedianya mitra teknologi yang sesuai, kesulitan dalam mendefinisikan masalah yang harus diselesaikan.

b. Sistem Informasi, terkait etika dan hukum (kendala etika dan hukum, tantangan dalam menangani privasi dan regulasi, bias dan keadilan, kerahasiaan, risiko keamanan, kurangnya alat audit,

B. Challenges to Implement AI

The main challenges in implementing AI are identified and categorized into four dimensions: organization, information systems, technology, and human resources (Lee et al., 2023).

a. Organization-related challenges include culture (resistance at various levels, misunderstandings among leaders, lack of agility), finance (high costs and investments, insufficient funding), knowledge (lack of understanding about timing, technology, data capabilities and usage levels), differing expectations at various organizational levels; lack of leadership; unavailability of suitable technology partners; difficulties in defining the problems that need to be solved.

b. Information Systems related challenges include ethics and legal issues (ethical and legal constraints; challenges in handling privacy and regulations; bias and fairness; confidentiality; security risks; lack

standar atau panduan), data (ukuran, akses, ketersediaan, sumber, kualitas, berbagi data, dan pelabelan data), kolaborasi AI-manusia dalam membuat AI dan kecerdasan manusia bekerja bersama, mekanisme pengendalian AI, pemeliharaan (pemeliharaan, perbaikan, kerusakan, masalah daya, mengikuti perubahan teknologi), mengelola pelatihan model, mengevaluasi kinerja sistem AI.

c. Teknologi, sehubungan keterbatasan teknologi dimana teknologi AI yang masih berkembang dan tidak stabil menghasilkan hasil yang tidak memuaskan, teknologi yang terlalu dipromosikan di pasar, model *black-box,* masalah multitugas dimana satu model hanya untuk satu tugas, integrasi dengan teknologi, sistem, dan proses yang ada.

d. Sumber Daya Manusia, terkait kurangnya keahlian dan karyawan yang terlatih, kondisi karyawan (tantangan dalam penempatan kembali staf, penurunan keterampilan staf, dampak negatif potensial akibat pengurangan tenaga kerja manusia, terjadinya isolasi manusia, pengetahuan dan keterampilan baru yang dibutuhkan staf), pelanggan (keinginan pelanggan akan sentuhan manusia, ketakutan dan ketidakpercayaan terhadap AI).

of auditing tools or standards/ guidelines), data (size; access; availability; sources; quality of data sharing and labeling), collaboration between AI and humans in making AI work alongside human intelligence; mechanisms for controlling AI systems maintenance (maintenance repairs damage power issues following technological changes); managing model training evaluating the performance of AI systems.

c. Technology-related challenges include limitations where emerging and unstable AI technology produces unsatisfactory results; overhyped technology in the market; black-box models; multitasking issues where one model is only for a single task; integration with existing technologies, systems, and processes.

d. Human Resources-related challenges include the lack of expertise and trained employees; employee conditions (challenges in redeploying staff; decline in staff skills; potential negative impacts from reducing human labor; occurrence of human isolation; new knowledge and skills required by staff); customers (customer desire for a human touch; fear and distrust towards AI).

C. Panduan Implementasi AI

Dalam konteks penerapan sistem AI pada bank, regulasi OJK dan ketentuan lain terkait tugas, tanggung jawab, serta wewenang Direksi dan Dewan Komisaris serta pelaksanaan fungsi dan aspek terkait lainnya dari tiga lini pertahanan (*three lines of defense*) yang mencakup lini manajemen bisnis, lini manajemen risiko dan kepatuhan, serta lini audit internal, tetap berlaku. Disamping itu, regulasi OJK yang terkait dengan aspek penyelenggaraan TI, layanan digital pada bank serta regulasi yang terkait lainnya, juga tetap menjadi pedoman. Hal ini dengan pertimbangan bahwa sistem AI merupakan salah satu adopsi dari teknologi informasi (sebagai *tools*) yang dilakukan bank.

Panduan ini menjadi acuan minimal bagi bank yang menerapkan sistem AI dalam organisasi bank dan proses bisnisnya. Bank dapat mengacu pada berbagai standar yang ada untuk memperkuat penerapan AI, sehingga penerapan AI yang mengedepankan integritas

C. Guideline to Implement AI

In the context of implementing AI systems in banks, OJK regulations and other provisions related to the duties, responsibilities, and authorities of the Board of Directors and Board of Commissioners as well as the implementation of functions and other related aspects from the three lines of defense, which include business management line; risk management and compliance line; and internal audit line remain applicable. In addition to this, OJK regulations related to IT governance aspects; digital services in banks; as well as other relevant regulations also continue to serve as guidelines. This is based on the consideration that AI systems are one form of adoption from information technology (as tools) carried out by banks.

This guideline serves as a minimum reference for banks implementing AI systems within their organizations and business processes. Banks can refer to various existing standards to strengthen the implementation of AI so that the application of AI

dan nilai-nilai etika dilakukan dengan menggunakan struktur dan proses tata kelola yang baik, risiko yang di mitigasi, serta memastikan kepatuhan terhadap hukum, peraturan, pedoman, termasuk standar yang berlaku, antara lain:

a. Undang-Undang No.4 Tahun 2023 tentang Pengembangan dan Penguatan Sektor Keuangan (UUP2SK).

b. Undang-Undang No. 27 tahun 2022 tentang Pelindungan Data Pribadi (UU PDP).

c. Panduan/regulasi OJK yang terkait, antara lain:
   1. Cetak Biru Transformasi Digital Perbankan.
   2. Panduan Resiliensi Digital (*Digital Resilience*).

prioritizes integrity and ethical values while utilizing good governance structures and processes; mitigating risks; and ensuring compliance with laws, regulations, guidelines including applicable standards among others:

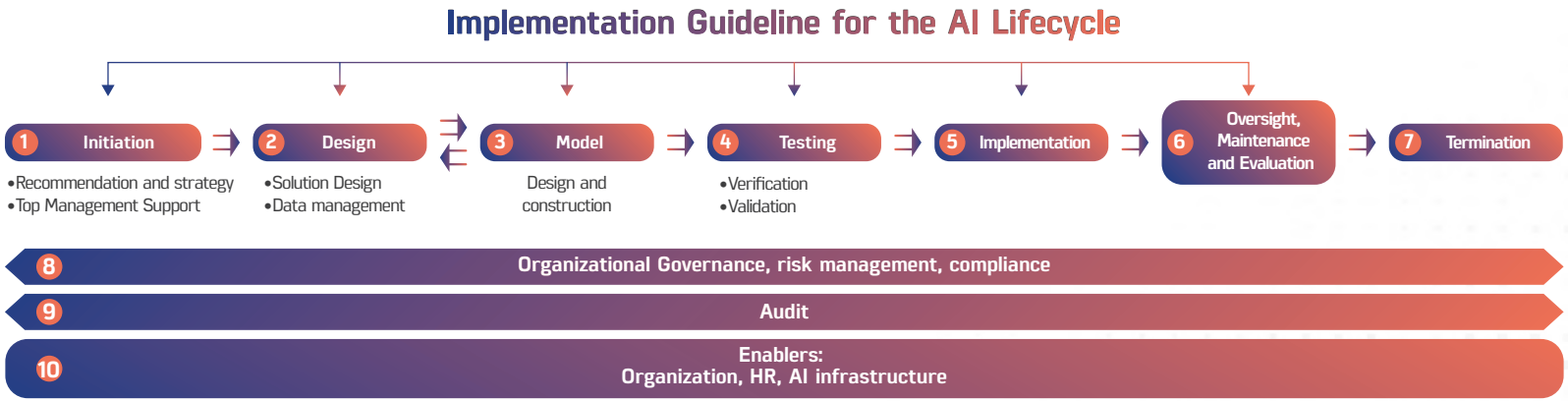
a. Law No. 4 of 2023 on the Development and Strengthening of the Financial Sector (UUP2SK).

b. Law No. 27 of 2022 on Personal Data Protection (UU PDP).

c. OJK guidelines/regulations related to this include, among others:
   1. Blueprint for Digital Transformation in Banking.
   2. Digital Resilience Guideline.

3. Ketentuan OJK mengenai:
   a. Penyelenggaraan Teknologi Informasi oleh Bank Umum.
   b. Layanan Digital oleh Bank Umum.
   c. Tingkat Kematangan Digital Bank Umum (DMAB).
   d. Ketahanan dan Keamanan Siber bagi Bank Umum.
   e. Penerapan Manajemen Risiko dalam Penggunaan Teknologi Informasi oleh Bank Umum.
   f. Penerapan Tata Kelola bagi Bank Umum, dan Penerapan Tata Kelola Syariah bagi Bank Umum Syariah dan Unit Usaha Syariah.
   g. Pelindungan Konsumen.

3. OJK Provisions regarding:
   a. Provision of Information Technology by Commercial Banks.
   b. Digital Services by Commercial Bank.
   c. Digital Maturity Level of Commercial Banks (DMAB).
   d. Cyber Resilience and Security for Commercial Banks.
   e. Implementation of Risk Management in the Use of Information Technology by Commercial Banks.
   f. Implementation of Governance for Commercial Banks, and Implementation of Sharia Governance for Sharia Commercial Banks and Sharia Business Units.
   g. Consumer Protection.



Source: Otoritas Jasa Keuangan (2025)

C.1. Inisiasi

Pada tahapan ini penting untuk memastikan pengawasan, akuntabilitas serta pengelolaan risiko dari penerapan sistem AI. Aspek-aspek yang perlu menjadi perhatian antara lain:

1. Sasaran dan strategi yang jelas.

   Bank menetapkan sasaran dan strategi yang jelas dalam penerapan sistem AI, untuk memastikan hasil yang diinginkan tercapai (*digital strategy*).

   Menyelaraskan sasaran dan strategi bank merupakan titik awal dalam kesuksesan penerapan AI, karena AI tidak hanya tentang teknologi tetapi juga terkait dengan aktivitas, batasan, nilai, dan tujuan bank.

   Strategi harus mencakup analisis *return on investment* dan rencana proyek AI dengan perspektif jangka panjang. Memilih fungsi sistem AI yang tepat yang sesuai dengan nilai strategis bank sangat penting untuk keberhasilan akhir dari tujuan bank.

   Strategi juga harus menjelaskan bagaimana bank akan mengadopsi dan mengimplementasikan AI dan mendapatkan manfaatnya yang selaras dengan tujuan bank. Strategi AI tidak hanya menyatakan apa yang ingin dicapai bank dengan implementasi AI, tetapi juga menyediakan proses, rencana, dan kerangka waktu tertentu untuk mengaktualisasikan tujuan tersebut. Selain itu, strategi AI dapat saja memerlukan penyesuaian dalam struktur organisasi bank, mekanisme kolaborasi antar unit, serta bagaimana alur data di seluruh unit organisasi pada bank.

   Seiring dengan terus berkembangnya teknologi AI, penting bagi Bank untuk dapat mengikuti berbagai perubahan ke depan, dan seiring dengan meningkatnya pengenalan teknologi AI perlu didukung dengan sumber daya manusia yang memiliki kompetensi digital.

C.1. Initiation

At this stage, it is important to ensure oversight, accountability, and risk management in the implementation of AI systems. Aspects that need to be considered include:

1. Clear objectives and strategies.

   The bank establishes clear objectives and strategies in the implementation of AI systems to ensure that the desired outcomes are achieved (digital strategy).

   Aligning the bank's objectives and strategies is the starting point for successful AI implementation because AI is not just about technology but also relates to the activities, constraints, values, and goals of the bank.

   The strategy must include a return on investment analysis and a long-term perspective project plan for AI. Choosing the right functions of the AI system that align with the strategic values of the bank is crucial for the ultimate success of the bank's objectives.

   The strategy must also explain how the bank will adopt and implement AI and derive benefits that align with the bank's objectives. The AI strategy not only states what the bank aims to achieve with the implementation of AI but also provides specific processes, plans, and timelines to actualize those goals. Additionally, the AI strategy may require adjustments in the organizational structure of the bank, collaboration mechanisms between units, as well as how data flows across organizational units within the bank.

   As technology continues to evolve, it is important for banks to keep up with various future changes, and as the recognition of AI technology increases, it must be supported by human resources with digital competencies.

AAI (2023) menyatakan bahwa strategi AI yang komprehensif terdiri dari empat bagian:

1. Visi AI, yakni mendefinisikan visi bank dan bagaimana AI dapat membantu mencapai tujuan bank, sehingga dapat ditetapkan tujuan penerapan AI secara jelas. Visi AI mencakup bagaimana bank memiliki keunggulan kompetitif di masa depan dengan dukungan AI dan bagaimana hal itu akan memengaruhi produk dan proses di bank. Selanjutnya, identifikasi area di mana AI bisa memberikan manfaat terbesar (*fields of action*) dan menciptakan nilai bisnis, baik melalui penjualan produk atau layanan tertentu, peningkatan efisiensi operasional, atau keduanya. Terakhir, pastikan ada komitmen yang jelas. Tetapkan tujuan yang terukur dan alokasikan sumber daya yang diperlukan untuk menerapkan AI dengan sukses.
2. AI *use cases*. Visi perlu diterjemahkan ke dalam portofolio *use case* AI. Untuk membangun portofolio ini, perlu mengidentifikasi, menentukan, mengevaluasi, dan memprioritaskan *use case* yang relevan. AI *use cases* menggambarkan bagaimana solusi AI dapat digunakan dalam konteks nyata untuk mencapai tujuan tertentu (menyelesaikan masalah, meningkatkan efisiensi, atau menciptakan nilai). Aspek yang perlu menjadi perhatian:

AAI (2023) states that a comprehensive AI strategy consists of four parts:

1. The vision for AI involves defining the bank's vision and how AI can help achieve the bank's objectives so that clear goals for the implementation of AI can be established. The vision for AI includes how the bank will have a competitive advantage in the future with the support of AI and how this will impact products and processes within the bank. Furthermore, identify areas where AI can provide the greatest benefits (fields of action) and create business value either through selling specific products or services or enhancing operational efficiency, or both. Finally, ensure there is a clear commitment by setting measurable goals and allocating necessary resources to successfully implement AI.
2. AI use cases. The vision needs to be translated into an AI use case portfolio. To build this portfolio, it is necessary to identify, define, evaluate, and prioritize relevant use cases. AI use cases describe how AI solutions can be used in real-world contexts to achieve specific objectives (solving problems, enhancing efficiency or creating value). Aspects that need attention:

a. Untuk menemukan, mengidentifikasi dan mengevaluasi *use cases* yang relevan, bank dapat menggunakan empat langkah pendekatan sistematis:

1. *Use cases* potensial diidentifikasi, baik berdasarkan tujuan strategis (berorientasi pada peluang) atau kekuatan yang ada (berorientasi pada aset dan kapabilitas).
2. Semua *use cases* potensial perlu didefinisikan dengan jelas dan komprehensif.
3. *Use cases* potensial harus diprioritaskan berdasarkan nilai dan peluang potensialnya, a.l. memulai dengan *use cases* yang relatif mudah diimplementasikan dan memberikan nilai tinggi.
4. *Use cases* dapat dieksekusi. Bank harus mengembangkan strategi pelaksanaan, divalidasi, diproduksi dan dipelihara untuk memberikan nilai berkelanjutan.

b. Mengembangkan pemahaman yang jelas tentang kapan harus membangun sendiri (membuat) atau membeli *use cases* AI. Terdapat enam faktor yang perlu menjadi pertimbangan:

1. Nilai strategis, mengacu pada potensi nilai *use cases* sehubungan dengan keunggulan kompetitif. Sumber nilai strategis dapat berupa peningkatan

a. To discover, identify, and evaluate relevant use cases, banks can use a four-step systematic approach:

1. Potential use cases are identified based on strategic objectives (opportunity-oriented) or existing strengths (asset and capability-oriented).
2. All potential use cases need to be clearly and comprehensively defined.
3. Potential use cases should be prioritized based on their value and potential opportunities, starting with use cases that are relatively easy to implement and provide high value.
4. Use cases can be executed. The bank must develop an implementation strategy, validate, produce, and maintain them to provide sustainable value.

b. Developing a clear understanding of when to build (create) or buy AI use cases. There are six factors that need to be considered:

1. Strategic value refers to the potential value of use cases in relation to competitive advantage. Sources of strategic value can include increased

efisiensi, pengurangan biaya, atau pengembangan fitur/ layanan produk berbasis AI.

2. Pentingnya kepemilikan dan kendali model *machine learning* (ML), dengan pertimbangan keunggulan kompetitif, keamanan, atau masalah regulasi.
3. Potensi pembelajaran, sejauh mana *use cases* menawarkan pembelajaran dari pengembangan internal.
4. Keunggulan kompetitif yang sulit ditiru oleh pesaing dalam pembuatan AI. Sumber keunggulan ini bisa berupa keterampilan khusus atau data eksklusif.
5. Kinerja solusi eksternal, mempertimbangkan kinerja (minimum) yang dapat diberikan oleh mitra eksternal yang memenuhi persyaratan bank terkait performa teknis, kualitas dan kecepatan pengiriman, atau kemampuan penyesuaian dan performa jangka panjang.
6. Total biaya kepemilikan, mencakup semua biaya yang diperlukan untuk pengembangan, penerapan, dan pemeliharaan *use cases* selama masa penggunaannya (misalnya, biaya komputasi, biaya pengembang, atau biaya untuk mengakses data pihak ketiga).

efficiency, cost reduction, or the development of AI-based product features/services.

2. The importance of ownership and control of machine learning (ML) models, considering competitive advantage, security, or regulatory issues.
3. Learning potential, to what extent use cases offer learning from internal development.
4. Competitive advantages that are difficult for competitors to replicate in AI development. Sources of this advantage can include specialized skills or exclusive data.
5. Performance of external solutions, considering the (minimum) performance that can be provided by external partners who meet the bank's requirements regarding technical performance, quality and delivery speed, or adaptability and long-term performance.
6. Total cost of ownership includes all costs required for the development, implementation, and maintenance of use cases during their usage period (e.g., computing costs, developer costs, or costs to access third-party data).

c. Mengelola Portofolio *Use Case* AI

Setelah menemukan berbagai kemungkinan penerapan AI (*use case*), langkah selanjutnya adalah mengelolanya dengan baik, termasuk mengevaluasi dan menentukan prioritas berdasarkan dampak, kelayakan, dan kesesuaiannya dengan tujuan bisnis, memantau faktor eksternal dan internal yang dapat memengaruhi keberhasilan, dan mengidentifikasi risiko dan ketergantungan antar proyek sehingga langkah antisipatif bisa diambil.

3. Faktor pendukung (*enablers*) yang diperlukan. Untuk menjalankan *use case*, diperlukan sejumlah faktor pendukung yang berkaitan dengan:
   - Organisasi, dimana AI hanya akan berdampak besar jika bank menyesuaikan struktur dan cara kerjanya.
   - Keahlian, dimana bank perlu memiliki tim yang tepat untuk menerapkan AI yang memerlukan strategi dalam merekrut, mengembangkan, dan mempertahankan talenta AI, serta menciptakan budaya kerja yang mendukung kolaborasi.
   - Budaya, dimana keberhasilan AI bergantung pada penerimaan dan kolaborasi antara manusia dan teknologi. Bank harus

c. Managing the AI Use Case Portfolio

After identifying various potential applications of AI (use cases), the next step is to manage them effectively, including evaluating and prioritizing based on impact, feasibility, and alignment with business objectives; monitoring external and internal factors that may affect success; as well as identifying risks and dependencies between projects so that anticipatory measures can be taken.

3. Supporting factors (enablers) required. To execute the use cases, a number of supporting factors related to are needed:
   - Organization, where AI will only have a significant impact if the bank adjusts its structure and ways of working.
   - Expertise, where the bank needs to have the right team to implement AI which requires a strategy for recruiting, developing, and retaining AI talent, as well as creating a work culture that supports collaboration.
   - Culture, where the success of AI depends on the acceptance and collaboration between humans and technology. The

menyesuaikan struktur, pola kerja, dan strategi agar AI dapat diterima dengan baik.

- Data, dimana AI bergantung pada data berkualitas tinggi. Bank harus memiliki infrastruktur data yang jelas, termasuk sumber data, *pipeline* data, pembersihan dan persiapan data, serta sistem manajemen data yang memudahkan akses dan pemanfaatan informasi.
- Teknologi, dimana infrastruktur IT yang tepat sangat penting, termasuk pemilihan antara *server* sendiri atau *cloud*, serta *software* yang digunakan untuk membangun dan mengembangkan AI. Keputusan ini harus mempertimbangkan faktor keamanan, biaya, dan fleksibilitas jangka panjang.
- Ekosistem. AI tidak berkembang sendiri, melainkan dalam ekosistem yang melibatkan laboratorium riset, komunitas teknologi, *start up*, universitas, dan perusahaan besar lain. Bank harus membangun strategi kemitraan dengan pihak-pihak yang dapat mendukung implementasi AI secara efektif.

4. Eksekusi

Eksekusi pengembangan AI dapat dibagi menjadi tiga fase utama:

a. Riset dan eksplorasi, berfokus pada penentuan tujuan, standar evaluasi, serta asumsi dasar

bank must adjust its structure, work patterns, and strategies to ensure that AI is well accepted.

- Data, where AI relies on high-quality data. The bank must have a clear data infrastructure, including data sources, data pipelines, data cleaning and preparation processes, as well as a data management system that facilitates access to and utilization of information.
- Technology, where the right IT infrastructure is crucial, including the choice between on-premises servers or cloud solutions, as well as the software used to build and develop AI. These decisions must consider security factors, costs, and long-term flexibility.
- Ecosystem. AI does not develop in isolation but within an ecosystem that involves research laboratories, technology communities, startups, universities, and other large companies. The bank must build partnership strategies with parties that can effectively support the implementation of AI.

4. Execution

The execution of AI development can be divided into three main phases:

a. Research and exploration, focusing on defining objectives, evaluation standards, and

untuk membangun *proof of concept*. Perlu diperhatikan durasi untuk menghindari pemborosan sumber daya pada proyek yang tidak layak.

b. Pengembangan dan validasi, dengan membangun prototipe awal (*minimal viable product*) untuk menguji nilai guna AI. Di samping itu, model dilatih, diuji, dan disempurnakan dengan berbagai arsitektur guna meningkatkan akurasi.

c. Operasionalisasi dan pemeliharaan. Setelah model mencapai tingkat akurasi yang memadai, model dipindahkan dari lingkungan uji ke dunia nyata. Model juga harus dipantau secara berkala untuk mengatasi penurunan kinerja akibat data *drift* (perubahan dalam distribusi data sehingga model AI atau *machine learning* mengalami penurunan kinerja) atau perubahan lingkungan. Sistem AI yang bergantung pada data yang terus berubah membutuhkan pemeliharaan berkelanjutan. Aspek yang harus menjadi perhatian adalah kualitas data, integritas, bias, keandalan layanan, dan respons terhadap data yang tidak biasa (*outliers*).

foundational assumptions to build a proof of concept. It is important to pay attention to the duration to avoid wasting resources on unfeasible projects.

b. Development and validation, by building an initial prototype (minimal viable product) to test the utility of AI. In addition, the model is trained, tested, and refined with various architectures to improve accuracy.

c. Operationalization and maintenance. Once the model reaches an adequate level of accuracy, it is transitioned from the testing environment to the real world. The model must also be monitored regularly to address performance degradation due to data drift (changes in data distribution that cause AI or machine learning models to underperform) or environmental changes. AI systems that rely on continuously changing data require ongoing maintenance. Aspects that need attention include data quality, integrity, bias, service reliability, and response to unusual data (outliers).

2. Dukungan dari *top management*.

   Penerapan AI memerlukan investasi, dukungan sumber daya, komitmen, dan dedikasi manajemen di semua tingkatan dari atas ke bawah. Direksi dan didukung oleh semua pejabat secara aktif, termasuk pemilik bisnis (PSP) bank perlu mendorong perubahan yang diperlukan untuk menerapkan AI, menjadi *role model* dan memfasilitasi adopsi dan insentif untuk perubahan.

   Para pengurus bank perlu menyadari pentingnya penggunaan AI yang selaras dengan tujuan dan kinerja bisnis, serta memiliki pemahaman yang mendalam mengenai strategi membangun, mengembangkan, dan mengimplementasikan sistem AI serta rencana pelaksanaan proyek-proyek berbasis AI. Karenanya perlu ditentukan prioritas penerapan sistem AI melalui diskusi dan kolaborasi dengan berbagai pihak termasuk tim

2. Support from top management.

   The implementation of AI requires investment, resource support, commitment, and management dedication at all levels from top to bottom. The board of directors and all officials actively supported by business owners (controlling shareholders) in the bank need to drive the necessary changes for implementing AI, serve as role models, and facilitate adoption as well as incentives for change.

   Bank executives need to recognize the importance of using AI in alignment with business goals and performance, as well as having a deep understanding of the strategies for building, developing, and implementing AI systems along with project execution plans for AI-based initiatives. Therefore, it is essential to establish priorities for the implementation of AI systems through discussions and collaboration with various stakeholders, including the AI project team to enhance

proyek AI, untuk dapat meningkatkan peluang sukses dalam transformasi proses bisnis terkait penerapan AI.

Terdapat tiga prinsip utama yang berlaku untuk semua organisasi dan dapat membantu pemimpin dalam mencapai kesuksesan dengan AI (AAI, 2023):

a. Perlakukan solusi AI sebagai produk, bukan proyek

   Hal ini mempertimbangkan bahwa solusi AI tidak pernah benar-benar "selesai.", dimana AI memerlukan pemeliharaan berkelanjutan, termasuk pembaruan model dan penyesuaian terhadap data yang terus berkembang. Oleh sebab itu, Bank perlu membentuk tim yang bertanggung jawab atas pengembangan dan integrasi AI secara penuh dalam organisasi. Oleh karena itu, Bank harus memperlakukan AI sebagai produk yang terus berkembang, bukan sekadar proyek sementara.

b. Menyeimbangkan koordinasi pusat dan desentralisasi

   Terdapat pandangan bahwa AI lebih baik dikelola secara terpusat, untuk mencegah duplikasi dan memastikan kolaborasi antar tim. Namun, hanya mengandalkan sentralisasi juga bukan solusi terbaik. Ketika tanggung jawab AI terlalu terpusat, akan terdapat silo antara dengan fungsi bisnis. Oleh karena itu, perlu diseimbangkan antara pendekatan pusat dan desentralisasi. Salah satu pendekatan terbaik yang sering diterapkan adalah model *hybrid*, dimana tim AI pusat (sering disebut *Center of Excellence* (CoE)) mengelola fungsi-fungsi utama dengan tetap terhubung dengan unit-unit yang terdesentralisasi di Bank. Tim ini juga dapat dilengkapi oleh tim integrasi yang bertanggung jawab atas pengadopsian dan skala solusi AI.

c. Kepemimpinan yang kuat dan komitmen dari jajaran *C-Level* (direksi) yang memiliki edukasi tentang AI

   Pemimpin yang ideal adalah seseorang yang memahami AI secara teknis, memiliki wawasan bisnis, dan mampu mengadvokasi perubahan dalam model bisnis dan budaya bank. Karenanya, edukasi tentang AI bagi *top management* sangat penting.

the chances of success in transforming business processes related to the application of AI.

There are three key principles that apply to all organizations and can help leaders achieve success with AI (AAI, 2023):

a. Treat AI solutions as products, not projects

   This takes into account that AI solutions are never truly "finished.", as they require ongoing maintenance, including model updates and adjustments to continuously evolving data. Therefore, the bank needs to establish a team responsible for the full development and integration of AI within the organization. Thus, the bank should treat AI as a continuously evolving product rather than just a temporary project.

b. Balancing central coordination and decentralization

   There is a view that AI is better managed centrally to prevent duplication and ensure collaboration among teams. However, relying solely on centralization is not the best solution either. When the responsibility for AI is too centralized, silos can develop between business functions. Therefore, it is necessary to balance between central and decentralized approaches. One of the best approaches often applied is the hybrid model, where a central AI team (often referred to as a Center of Excellence (CoE)) manages key functions while remaining connected with decentralized units in the bank. This team can also be complemented by an integration team responsible for adopting and scaling AI solutions.

c. Strong leadership and commitment from the C-Level (board of directors) who have education about AI

   The ideal leader is someone who understands AI technically, has business insights, and can advocate for changes in the bank's business model and culture. Therefore, education about AI for top management is very important.

C.2. Desain

1. Perancangan solusi

   Bagaimana solusi yang dirancang dan dibangun mengacu pada nilai-nilai utama yang terdapat pada Bab 4 buku ini. Agar sistem AI yang dikembangkan sesuai dengan tujuan yang diharapkan bank, antara lain memperhatikan:

   a. Penyelarasan tujuan

      Penting untuk terlebih dahulu mengidentifikasi permasalahan bisnis yang dapat diselesaikan oleh

C.2. Design

1. Solution design

   How the solutions designed and built refer to the core values outlined in Chapter 4 of this book. To ensure that the AI systems developed align with the bank's expected objectives, among other things, it is necessary to consider:

   a. Alignment of objectives

      It is important to first identify the business problems that can be solved by AI and analyze how

AI dan menganalisis bagaimana kemampuan bank saat ini untuk dapat menindaklanjutinya. Salah satu tantangan utama dalam penerapan AI adalah integrasinya dengan infrastruktur yang sudah ada, yang sering kali memerlukan penyesuaian di seluruh unit organisasi bank. Memulai inisiatif AI tanpa tujuan yang jelas dapat menyebabkan kegagalan implementasi. Oleh karena itu, penyelarasan tujuan antara inisiatif AI dan kebutuhan bisnis bank menjadi langkah awal yang sangat penting untuk keberhasilan penerapan AI.

b. Kompatibilitas dalam adopsi AI

Kompatibilitas mengacu pada kesesuaian antara teknologi AI dan kebutuhan bisnis. Semakin sesuai teknologi dengan tugas yang diharapkan, semakin tinggi tingkat adopsinya. Kompatibilitas ini terbagi menjadi dua aspek utama, yakni proses bisnis dan kasus bisnis (*business case*).

Untuk memastikan keberhasilan adopsi AI, bank harus memiliki *business case* yang jelas, yang mencakup masalah yang ingin diselesaikan serta manfaat AI dalam meningkatkan eksekusi dan hasil bisnis. Selain itu, adopsi AI sering kali memunculkan kebutuhan baru, sehingga perlu menyesuaikan proses bisnis agar tetap selaras dengan persyaratan teknologi yang diterapkan.

the bank's current capabilities can follow up on them. One of the main challenges in implementing AI is its integration with existing infrastructure, which often requires adjustments across all units of the bank's organization. Starting an AI initiative without clear objectives can lead to implementation failure. Therefore, aligning the objectives between the AI initiatives and the business needs of the bank becomes a crucial first step for successful implementation of AI.

b. Compatibility in AI adoption

Compatibility refers to the alignment between AI technology and business needs. The more aligned the technology is with the expected tasks, the higher the adoption rate will be. This compatibility is divided into two main aspects: business processes and business cases.

To ensure the successful adoption of AI, banks must have a clear business case that includes the problems they want to solve as well as the benefits of AI in enhancing execution and business outcomes. Additionally, the adoption of AI often creates new needs, necessitating adjustments to business processes to remain aligned with the requirements of the implemented technology.

c. Kemampuan teknologi

Kemampuan teknologi membangun fondasi yang penting dan fleksibel untuk penerapan sistem AI secara efektif. Kemampuan teknologi yang kuat terutama analitik, data, dan sistem teknologi, yang dapat menyederhanakan kompleksitas integrasi AI karena memungkinkan fungsi yang terkait untuk menerapkan teknologi AI dengan cepat dan efisien.

d. Infrastruktur sistem AI

Untuk implementasi AI yang sukses, bank harus berinvestasi dalam infrastruktur teknologi yang tepat untuk mengadopsi AI yang mencakup *hardware* dan *software*.

Sistem AI sangat bergantung pada infrastruktur TI untuk pengembangan dan operasinya, sehingga infrastruktur ini menjadi faktor krusial untuk keberhasilan penerapan AI. Infrastruktur TI mendukung integrasi AI dan memastikan aliran informasi yang lancar, pengelolaan volume dan kualitas data yang maksimal, yang diperlukan untuk komunikasi antar sistem dan algoritma AI.

Untuk mengadopsi AI, bank juga memerlukan akses ke solusi berbasis *cloud* atau memiliki *hardware* komputasi yang tepat untuk memfasilitasi penggunaan AI.

c. Technological capabilities

Technological capabilities build an important and flexible foundation for the effective implementation of AI systems. Strong technological capabilities, particularly in analytics, data, and technology systems, can simplify the complexities of AI integration as they enable related functions to implement AI technology quickly and efficiently.

d. AI system infrastructure

For successful AI implementation, banks must invest in the right technology infrastructure to adopt AI that includes hardware and software.

AI systems heavily rely on IT infrastructure for their development and operations, making this infrastructure a crucial factor for the success of AI deployment. IT infrastructure supports the integration of AI and ensures smooth information flow, optimal management of data volume and quality necessary for communication between AI system and algorithm.

To adopt AI, banks also need access to cloud-based solutions or have the appropriate computing hardware to facilitate the use of AI.

Software AI mencakup paket dan algoritma yang sesuai untuk proyek AI. Bekerjasama dengan konsultan/ pakar dengan pengetahuan domain bisnis dan teknologi AI serta pengalaman proyek adalah satu cara terbaik dalam implementasi AI yang sukses.

Menggunakan pendekatan validasi silang (*cross-validation approaches*) berbasis waktu perlu dilakukan di mana model diuji terhadap data yang lebih baru. Selain itu, sistem AI secara rutin perlu diuji dan diperbarui untuk memastikan kinerjanya tetap optimal. Strategi pengujian yang kuat, termasuk pengujian positif (pengujian yang dilakukan untuk memverifikasi bahwa sistem AI dapat menghasilkan output yang benar atau diharapkan ketika diberikan input yang valid dan sesuai dengan kondisi normal) dan pengujian negatif (pengujian ini dilakukan untuk memeriksa bagaimana sistem AI menangani input yang tidak valid, tidak terduga, atau di luar kondisi normal) sangat penting untuk memastikan keandalan sistem AI. Pengujian kinerja juga perlu dilakukan untuk memverifikasi bahwa infrastruktur yang ada memenuhi persyaratan sistem AI.

Dengan infrastruktur dan strategi pengujian yang tepat, bank dapat memaksimalkan potensi AI untuk mendukung tujuan bisnisnya.

AI software includes packages and algorithms suitable for the AI project. Collaborating with consultants/ experts who have domain knowledge in business and technology as well as project experience is one of the best ways to ensure successful implementation of AI.

Using time-based cross-validation approaches is necessary where the model is tested against newer data. Additionally, AI systems need to be routinely tested and updated to ensure their performance remains optimal. A strong testing strategy, including positive testing (testing conducted to verify that the AI system can produce correct or expected outputs when given valid inputs under normal conditions) and negative testing (this testing is done to check how the AI system handles invalid, unexpected inputs or those outside of normal conditions), is crucial for ensuring the reliability of the AI system. Performance testing also needs to be conducted to verify that the existing infrastructure complies to the AI system requirements.

With the right infrastructure and testing strategies, banks can maximize the potential of AI to support their business objectives.

e. Tata kelola pada tingkat sistem

Tata kelola pada tingkat sistem membantu entitas memastikan bahwa sistem AI memenuhi persyaratan kinerja. GAO (2021), mengidentifikasi tata kelola pada tingkat sistem terkait aspek:

- Spesifikasi, yakni menetapkan dan mendokumentasikan spesifikasi teknis untuk memastikan sistem AI memenuhi tujuan yang diharapkan.
- Kepatuhan, memastikan sistem AI mematuhi hukum, peraturan, standar, dan panduan yang relevan.
- Transparansi, sesuai dengan batasan sebagaimana regulasi, memungkinkan pemangku

e. Governance at the system level

Governance at the system level helps entities ensure that AI systems meet performance requirements. GAO (2021) identifies governance at the system level related to aspects:

- Specifications, which involve establishing and documenting technical specifications to ensure that the AI system meets the expected objectives.
- Compliance, ensuring that the AI system adheres to relevant laws, regulations, standards, and guidelines.
- Transparency, in accordance with regulatory constraints, allows external stakeholders to

kepentingan eksternal mengakses informasi tentang desain, operasi, dan batasan sistem AI, untuk meningkatkan transparansi.

f. Keamanan layanan dan pelindungan data pribadi

Saat membangun sistem AI, selain fokus pada otomatisasi juga berfokus pada keamanan layanan karena transaksi diproses dengan kecepatan tinggi. Semakin krusial suatu proses bisnis, semakin besar pula upaya yang diperlukan untuk memastikan keandalan dan keamanannya.

Salah satu aspek penting dalam penanganan data pribadi adalah de-identifikasi (*de-identify*), yaitu proses yang membuat data sulit dikaitkan dengan individu tertentu. Hal ini dilakukan dengan menghapus atau mengganti informasi yang dapat mengidentifikasi seseorang. Data yang telah dide-identifikasi sesuai dengan pedoman yang berlaku dianggap bukan lagi sebagai data pribadi, sehingga dapat digunakan atau dibagikan ke pihak ketiga tanpa memerlukan persetujuan dari pemilik data. Namun demikian, masih terdapat risiko re-identifikasi jika muncul teknologi baru atau informasi tambahan yang memungkinkan data tersebut dikaitkan kembali dengan individu tertentu, sehingga penting untuk memastikan bahwa data yang telah dide-identifikasi tidak dapat diubah kembali menjadi data pribadi selama proses integrasi atau analisis dan menghindari pelanggaran terhadap peraturan hukum yang berlaku.

access information about the design, operation, and limitations of the AI system to enhance transparency.

f. Service security and personal data protection

When building AI systems, in addition to focusing on automation, there is also a focus on service security due to the high-speed processing of transactions. The more critical a business process is, the greater the effort required to ensure its reliability and security.

One important aspect of handling personal data is de-identification, which is the process that makes it difficult to associate data with specific individuals. This is done by removing or replacing information that can identify a person. Data that has been de-identified in accordance with applicable guidelines is no longer considered personal data, allowing it to be used or shared with third parties without requiring consent from the data owner. However, there remains a risk of re-identification if new technologies or additional information emerge that allow the data to be linked back to specific individuals. Therefore, it is crucial to ensure that de-identified data cannot be reverted to personal data during integration or analysis processes and to avoid violations of applicable legal regulations.

g. Proses digital

Tingkat digitalisasi pada bank menjadi salah satu faktor untuk keberhasilan penerapan AI, karena digitalisasi memungkinkan pengumpulan data secara *real-time* dari berbagai proses dan

g. Digital processes

The level of digitization in banks is one of the factors for the successful implementation of AI because digitization enables real-time data collection from various processes and products. Without

produk. Tanpa infrastruktur digital, sistem AI tidak akan memiliki data pelatihan yang penting untuk melatih algoritma, sehingga tidak dapat secara efektif dalam implementasi AI. Peran proses digital ini tidak hanya menjadi dasar, tetapi juga sangat penting untuk memungkinkan bank memanfaatkan AI secara efektif, yang menekankan perlunya pendekatan terintegrasi dalam transformasi digital untuk penerapan AI. Karenanya, upaya untuk mencapai kesuksesan dalam penerapan transformasi digital perbankan termasuk adopsi AI, juga harus mengacu sebagaimana ketentuan OJK mengenai Penilaian Tingkat Maturitas Digital Bank Umum.

a digital infrastructure, AI systems will lack the essential training data needed to train algorithms effectively for implementing AI. The role of these digital processes is not only foundational but also crucial for enabling banks to leverage AI effectively, emphasizing the need for an integrated approach in the digital transformation for implementing AI. Therefore, efforts to achieve success in banking's digital transformation initiatives including the adoption of AI must also refer to OJK regulations regarding the Assessment of Digital Maturity Levels in Commercial Banks.

h. Dukungan dari pengguna akhir (nasabah)

Dukungan dari pengguna akhir adalah salah satu faktor kunci keberhasilan dalam implementasi AI. Pengguna akhir harus dilibatkan dalam proyek AI untuk memahami kebutuhan mereka. Sistem AI harus mudah digunakan, andal, dan tersedia dari perspektif pengguna akhir. Dengan melibatkan pengguna akhir, memastikan sistem AI mudah digunakan serta didukung dengan edukasi yang memadai termasuk dengan menyediakan panduan pengguna yang spesifik, implementasi AI dapat lebih sukses dan memberikan nilai tambah bagi bisnis.

h. Support from end-users (customers)

Support from end users is a key factor for success in AI implementation. End users must be involved in the AI project to understand their needs. The AI system should be user-friendly, reliable, and accessible from the end user's perspective. By involving end users and ensuring that the system is easy to use while being supported by adequate education–including providing specific user guides–AI implementation can be more successful and add value to the business.

i. Inklusif

Menghindari desain yang tidak inklusif, dimana sistem AI tidak dapat diakses dan digunakan secara merata oleh nasabah/calon nasabah atau kelompok nasabah/calon nasabah dengan kriteria tertentu (suku, ras, jenis kelamin, disabilitas).

2. Pengelolaan data

Secara umum, dalam pelaksanaan pengelolaan data bank agar memperhatikan ketentuan yang terkait dengan pelindungan data pribadi, sebagaimana Undang-Undang No. 27 tahun 2022 tentang Pelindungan Data Pribadi serta ketentuan OJK yang terkait. Pengawasan dari bank (*human oversight*) untuk menilai kecukupan dan kualitas pengelolaan data yang akan digunakan pada sistem AI juga perlu menjadi perhatian penting oleh bank.

a. Tata kelola data

Penerapan AI dihadapkan pada berbagai tantangan, termasuk aspek etika, politik, hukum, serta kebijakan, yang mencakup isu aliran data, privasi, dan aksesibilitas. Oleh karena itu, diperlukan sistem tata kelola yang kuat untuk memastikan bahwa teknologi AI dan protokol data mematuhi regulasi serta prinsip etis yang berlaku. Dengan adanya tata kelola yang baik, risiko yang terkait dengan penggunaan AI dapat diminimalkan.

i. Inclusive

Avoiding non-inclusive design, where the AI system cannot be accessed and used equitably by customers/prospective customers or groups of customers/prospective customers based on specific criteria (such as ethnicity, race, gender, disability), is crucial.

2. Data management

In general, in the implementation of data management, banks should pay attention to regulations related to personal data protection as stipulated in Law No. 27 of 2022 on Personal Data Protection and relevant OJK regulations. Oversight from the bank (human oversight) to assess the adequacy and quality of the data management that will be used in AI systems is also an important consideration for banks.

a. Data governance

The implementation of AI faces various challenges including ethical, political, legal aspects as well as policies that encompass issues of data flow, privacy, and accessibility. Therefore, a strong governance system is needed to ensure that AI technology and data protocols comply with applicable regulations and ethical principles. With good governance in place, the risks associated with the use of AI can be minimized.

Tata kelola yang efektif di seluruh unit organisasi bank tidak hanya membantu mengurangi bias dalam AI, tetapi juga meningkatkan kepercayaan terhadap *insight* yang dihasilkan oleh algoritma AI, seperti pola, tren, prediksi, serta rekomendasi berbasis data. Oleh sebab itu, tata kelola AI dan data memainkan peran strategis dalam menentukan keberhasilan implementasi serta skala AI pada bank.

Terhadap data yang digunakan untuk mengembangkan Model AI, bank harus mendokumentasikan sumber dan asal data, memastikan keandalan data, serta menilai atribut data, variabel, dan augmentasi/ *enhancement* data untuk kesesuaian. Disamping itu, terhadap data yang digunakan untuk mengoperasikan sistem AI, bank harus menilai keterhubungan dan ketergantungan aliran data yang mengoperasikan sistem AI, mengidentifikasi potensi bias, dan menilai keamanan serta privasi data. Standarisasi proses tata kelola data penting untuk dilakukan agar pengelolaan data dapat berjalan dengan efektif.

b. Kualitas dan kuantitas data

Inti dari AI adalah data. Data merupakan *enabler* penting dari adopsi AI termasuk adopsi AI pada bank. Algoritma AI sangat bergantung pada data, sehingga ketersediaan *series* data yang luas dan berkualitas tinggi menjadi syarat utama. Data pelatihan yang memadai diperlukan untuk memastikan proses pembelajaran AI berjalan efektif, sementara kurangnya data pelatihan dapat mengganggu kinerja dan keandalan algoritma.

Kualitas data sangat penting untuk memberikan prediksi yang andal. Jika data pelatihan memiliki kualitas rendah, wawasan yang dihasilkan oleh AI juga akan berkualitas rendah dan tidak berguna dalam konteks organisasi (*garbage-in, garbage-out*). Aspek penting dari kualitas data berkaitan kelengkapan, entri yang tepat, dan fitur yang jelas, serta penggunaan data yang bebas dari bias dan mengikuti prinsip-prinsip yang bertanggung jawab dan dapat dipercaya.

Dengan memberikan perhatian pada kualitas dan kuantitas data, termasuk pemeliharaan yang dilakukan secara konsisten, bank dapat memenuhi harapan terhadap AI sekaligus menciptakan solusi yang andal dan dapat dipercaya.

c. Keamanan data

Dalam proses transformasi digital, bank membutuhkan infrastruktur TI yang solid sebagai dasar utama untuk mendukung proses bisnis inti dan penerapan AI, termasuk untuk melindungi bank dari risiko khususnya serangan siber yang dapat mengancam aset informasi

Effective governance across all units of the banking organization not only helps reduce bias in AI but also enhances trust in the insights generated by AI algorithms, such as patterns, trends, predictions, and data-driven recommendations. Therefore, governance of AI and data plays a strategic role in determining the success of implementation and scaling of AI within banks.

For the data used to develop AI models, banks must document the sources and origins of the data, ensure its reliability, and assess the attributes of the data along with variables and any augmentation/ enhancement for suitability. Additionally, for the data used to operate AI systems, banks should evaluate the interconnections and dependencies of the data flows that operate these systems while identifying potential biases and assessing security as well as privacy concerns. Standardization of governance processes is crucial in this context.

b. Data quality and quantity

The core of AI is data. Data is a crucial enabler of AI adoption, including in banks. AI algorithms heavily rely on data; therefore, the availability of extensive and high-quality datasets is a primary requirement. Adequate training data is necessary to ensure that the learning process for AI operates effectively, while a lack of training data can disrupt the performance and reliability of algorithms.

Data quality is crucial for providing reliable predictions. If the training data is of low quality, the insights generated by AI will also be of low quality and useless in the organizational context (garbage-in, garbage-out). Important aspects of data quality relate to completeness, accurate entry, and clear features, as well as using data that is free from bias and adheres to responsible and trustworthy principles.

By paying attention to the quality and quantity of data, including consistent maintenance, banks can meet expectations for AI while also creating reliable and trustworthy solutions.

c. Data security

In the digital transformation process, banks require a solid IT infrastructure as a primary foundation to support core business processes and the implementation of AI, this is essential for protecting banks from risks, particularly cyberattacks that can threaten

yang menjadi dasar operasional AI. Jika aspek keamanan ini diabaikan, konsekuensi yang timbul bisa sangat merugikan, seperti kerusakan sistem, kebocoran data, hingga berkurangnya kepercayaan nasabah. Oleh sebab itu, penerapan protokol keamanan siber yang kuat menjadi elemen krusial untuk menjamin keberhasilan implementasi AI sekaligus melindungi keandalan dan integritas sistem AI.

Untuk membantu entitas menggunakan data yang sesuai dengan tujuan penggunaan setiap sistem AI, GAO (2021) mengidentifikasi praktik utama untuk memastikan data berkualitas tinggi, andal, dan representatif, sebagai berikut:

a. Data yang digunakan untuk pengembangan model:

- Mendokumentasikan sumber dan asal data yang digunakan untuk mengembangkan model yang mendukung sistem AI.
- Melakukan evaluasi keandalan data yang digunakan dalam pengembangan model.
- Meninjau atribut yang digunakan untuk klasifikasi data.
- Evaluasi variabel data yang digunakan dalam model AI.
- Meninjau penggunaan data sintetis, imputasi, dan/atau *augmented data* (data yang telah dimodifikasi atau ditingkatkan).

the information assets underlying AI operations. If security aspects are neglected, the consequences can be very detrimental, ranging from system damage and data breaches to a decline in customer trust. Therefore, the implementation of strong cybersecurity protocols is a crucial element to ensure the successful implementation of AI while protecting the reliability and integrity of AI systems.

To assist entities in using data that aligns with the intended purpose of each AI system, GAO (2021) identifies key practices to ensure high-quality, reliable, and representative data as follows:

a. Data used for model development:

- Documenting the sources and origins of the data used to develop models that support AI systems.
- Evaluating the reliability of the data used in model development.
- Reviewing the attributes used for data classification.
- Evaluation of the data variables used in the AI model.
- Reviewing the use of synthetic data, imputation, and/or augmented data (data that has been modified or enhanced).

b. Data yang digunakan untuk operasi sistem

- Melakukan evaluasi keterkaitan dan ketergantungan aliran data yang menjalankan sistem AI.
- Meninjau keandalan, kualitas, dan representasi semua data yang digunakan dalam operasi sistem, termasuk potensi bias, ketidakadilan, dan dampak sosial lainnya.
- Melakukan evaluasi keamanan dan privasi data dalam sistem AI.

Aspek-aspek yang terkait data perlu menjadi perhatian bank dalam mengimplementasikan AI. Aspek yang perlu menjadi perhatian antara lain:

b. Data used for system operations

- Evaluating the interconnections and dependencies of the data flows that operate the AI system.
- Reviewing the reliability, quality, and representation of all data used in system operations, including potential biases, unfairness, and other social impacts.
- Evaluating the security and privacy of data in the AI system.

Data-related aspects need to be a concern for banks when implementing AI. The aspects that need attention include:

1. Integrasi data, yang mencakup pengelolaan volume data, memindahkan data dari lokal (*on-premise*) ke *cloud*, memungkinkan akses *real-time*, dan mengelola perubahan dalam data.
2. Tata kelola dan keamanan data harus menjadi perhatian utama sejak awal strategi AI, untuk memastikan bahwa data digunakan dan diakses dengan cara yang benar dan aman.
3. Kualitas dan kesiapan data. Proses sentralisasi data yang dibutuhkan untuk proyek AI memilki volume yang sangat besar dan sering berubah dengan cepat, berasal dari berbagai sumber, memiliki tingkat kebersihan dan kelengkapan yang berbeda-beda, serta bisa berbentuk terstruktur, semi-terstruktur, atau tidak terstruktur.
4. Kebersihan data atau *data cleaning*, yang merupakan proses untuk memperbaiki data yang tidak akurat, tidak lengkap, atau tidak teratur.
5. Budaya organisasi terkait data (*data culture*), yang merupakan keyakinan dan perilaku kolektif sumber daya manusia pada bank untuk memanfaatkan data demi peningkatan kinerja bisnis serta mendorong transformasi organisasi dalam mengatasi tantangan bisnis yang semakin

1. Data integration, which includes managing data volume, migrating data from on-premise to the cloud, enabling real-time access, and managing changes in the data.
2. Data governance and security must be a primary concern from the outset of the AI strategy to ensure that data is used and accessed correctly and securely.
3. Data quality and readiness. The data centralization process required for AI projects has a very large volume and often changes rapidly, coming from various sources with varying levels of cleanliness and completeness, and can be structured, semi-structured, or unstructured.
4. Data cleaning, which is the process of correcting inaccurate, incomplete, or irregular data.
5. Organizational culture related to data (data culture), which is the collective beliefs and behaviors of human resources in banks to leverage data for improving business performance and driving organizational transformation in addressing increasingly complex business

kompleks. *Data culture* yang kuat antara lain akan dapat meningkatkan produktivitas, inovasi, layanan, pengambilan keputusan yang lebih baik dan otimalisasi biaya. BARC (2023) mendefinisikan *data culture* dalam enam aspek yaitu *data leadership, data strategy, data governance, data literacy, data access,* dan *data communication.*

6. Fondasi data (*data foundation*) yang tahan lama diperlukan agar dapat mendukung teknologi saat ini dan di masa depan. Tiga prinsip utama untuk kesiapan data bagi AI yang berkelanjutan:

a. Fondasi data harus dibangun sebelum menerapkan AI

Bank perlu mengatasi aspek teknis dalam infrastruktur data (*technical debt in data infrastructure*) sebelum dapat mengembangkan aplikasi AI yang berdampak besar. *Technical debt in data infrastructure* antara lain disebabkan oleh:

- Adanya silo data, dimana data tersebar di berbagai sistem yang tidak terhubung, sehingga sulit diintegrasikan.
- Teknologi yang sudah usang (*legacy systems*), dimana sistem tidak kompatibel dengan solusi AI modern.

challenges. A strong data culture can enhance productivity, innovation, service quality, better decision-making, and cost optimization. BARC (2023) defines data culture in six aspects: data leadership, data strategy, data governance, data literacy, data access, and data communication.

6. A durable data foundation is necessary to support current and future technologies. Three key principles for data readiness for sustainable AI:

a. The data foundation must be established before implementing AI

Banks need to address technical aspects of data infrastructure (technical debt in data infrastructure) before they can develop impactful AI applications. Technical debt in data infrastructure is caused by:

- The existence of data silos, where data is scattered across various unconnected systems, making integration difficult.
- Outdated technology (legacy systems), where the systems are incompatible with modern AI solutions.

- Standarisasi data yang minim, dimana data memiliki struktur, format, atau pola yang tidak konsisten.
- *Pipeline* data yang tidak efisien, yakni proses ekstraksi, transformasi, dan pemuatan data (*extract, transform, and load* (ETL)) lambat atau tidak fleksibel.
- Keamanan dan kepatuhan yang lemah, dimana sistem tidak didesain dengan tata kelola dan perlindungan data yang kuat sejak awal.

Technical debt in data infrastructure akan berdampak:

- Memperlambat adopsi AI karena data sulit diakses atau digunakan secara efektif.
- Biaya tinggi untuk perbaikan karena bank harus melakukan *rekonstruksi kode program* secara signifikan.
- Kinerja buruk akibat infrastruktur yang tidak optimal dalam menangani volume data besar.

Karenanya *technical debt in data infrastructure* perlu diselesaikan karena AI memerlukan data yang bersih, terstruktur, dan dapat diakses dengan cepat, antara lain melalui:

- Melakukan migrasi ke arsitektur data modern seperti *cloud data lakes* atau *data mesh*.

- Minimal data standardization, where the data has inconsistent structures, formats, or patterns.
- Inefficient data pipelines, where the extract, transform, and load (ETL) processes are slow or inflexible.
- Weak security and compliance, where systems are not designed with strong governance and data protection from the outset.

The impacts of technical debt in data infrastructure:

- Slowing down AI adoption because data is difficult to access or use effectively.
- High costs for repairs because banks must significantly reconstruct the program code.
- Poor performance due to infrastructure that is not optimized for handling large data volumes.

Therefore, technical debt in data infrastructure needs to be addressed because AI requires clean, structured data that can be accessed quickly, including through:

- Migrating to modern data architectures such as cloud data lakes or data mesh.

- Mengadopsi *pipeline* data otomatis dan *real-time* untuk mempercepat aliran data.
- Menerapkan tata kelola data sejak awal agar integritas dan keamanan data tetap terjaga.

b. Terhadap data yang digunakan untuk mengembangkan model AI, bank harus mendokumentasikan sumber dan asal data, memastikan keandalan data, serta menilai atribut data, variabel, dan augmentasi/ *enhancement* (memperbaiki, memperluas, atau meningkatkan data agar lebih bermanfaat, relevan, atau berkualitas tinggi) data untuk kesesuaian.

c. Terhadap data yang digunakan untuk mengoperasikan sistem AI, bank harus menilai keterhubungan dan ketergantungan aliran data yang mengoperasikan sistem AI, mengidentifikasi potensi bias, dan menilai keamanan serta privasi data.

7. Berbagai jenis bias dan faktor lain yang memengaruhi kewajaran yang mungkin terjadi terkait data, dimana data dapat memberikan gambaran yang tidak akurat tentang kondisi nyata atau merugikan kelompok atau aspek tertentu, antara lain (OECD, 2019):

- Bias pelaporan, dimana informasi yang tersedia tidak sepenuhnya dilaporkan.
- Adopting automated and real-time data pipelines to accelerate data flow.
- Implementing data governance from the outset to maintain data integrity and security

b. For the data used to develop AI models, banks must document the sources and origins of the data, ensure its reliability, and assess the attributes of the data, variables, and augmentation/ enhancement (improving, expanding or enhancing the data to make it more useful, relevant or high-quality) for suitability.

c. For the data used to operate AI systems, banks must assess the interconnections and dependencies of the data flows that operate the AI systems, identify potential biases, and evaluate the security and privacy of the data.

7. Various types of biases and other factors that may affect fairness related to data, where the data can provide an inaccurate representation of real conditions or harm certain groups or aspects, including (OECD, 2019):

- Reporting bias, where the available information is not fully reported.
- Bias seleksi, dimana data yang digunakan lebih banyak mewakili satu kelompok dibandingkan kelompok lain, sehingga sistem AI bekerja lebih baik untuk satu kelompok tetapi kurang akurat untuk yang lain. Kondisi ini bisa terjadi karena cakupan data yang tidak merata atau metode pengambilan sampel yang kurang inklusif.
- Bias homogenitas luar kelompok, dimana cenderung melihat individu di luar kelompok sendiri sebagai lebih seragam dalam sikap, nilai, atau karakteristik dibandingkan dengan anggota kelompok sendiri.

C.3. Model

Perancangan dan pembangunan model AI merupakan tahap yang sangat penting dalam mengembangkan sistem AI, di mana pilihan arsitektur secara signifikan memengaruhi kinerja dan skala sistem AI. Sebagai inti dari sistem AI, model merepresentasikan sebagian atau keseluruhan lingkungan eksternal yang menggambarkan struktur dan dinamika sistem tersebut (OECD, 2019). Model dapat dibangun berdasarkan pada data dan/ atau pengetahuan ahli, yang dibuat oleh manusia dan/atau oleh alat otomatis seperti algoritma *machine learning*. Interpretasi model memungkinkan manusia atau sistem otomatis untuk memahami dan memanfaatkan keluaran (*output*) yang dihasilkan oleh model tersebut.

- Selection bias, where the data used represents one group more than another, causing the AI system to perform better for one group but less accurately for others. This condition can occur due to uneven data coverage or less inclusive sampling methods.
- Outgroup homogeneity bias, where individuals outside one's own group are perceived as more uniform in attitudes, values, or characteristics compared to members of one's own group.

C.3. Model

The design and development of AI models is a crucial stage in developing AI systems, where architectural choices significantly affect the performance and scale of the AI system. As the core of the AI system, models represent part or all of the external environment that describes the structure and dynamics of that system (OECD, 2019). Models can be built based on data and/or expert knowledge, created by humans and/ or by automated tools such as machine learning algorithms. The interpretation of the model allows humans or automated systems to understand and utilize the outputs generated by the model.

Dalam pembangunan model, penting untuk memperhatikan tujuan yang ingin dicapai dan ukuran kinerja yang mencakup akurasi, efisiensi dalam proses pelatihan, *dataset* yang memadai. Aspek yang perlu menjadi perhatian dalam perencanaan dan pengembangan AI secara efektif, antara lain:

1. Manajemen model secara terorganisir, dengan membangun dan memperbarui inventaris model secara menyeluruh pada bank untuk memastikan keterlacakan (*audit trail*) dan efisiensi dalam penerapan AI.
2. Validasi dan kualitas data pelatihan (*training data*), dengan memastikan integritas data melalui evaluasi secara ketat terhadap proses pengumpulan data, serta memastikan bahwa data pelatihan (*training data*) secara akurat mencerminkan populasi dimana AI akan digunakan. Jika terdapat kesenjangan dalam data, penggunaan model dapat dibatasi untuk meminimalkan bias dan meningkatkan keandalan.
3. Menetapkan tujuan yang tepat bagi sistem AI serta memastikan bahwa setiap komponen mulai dari algoritma hingga input data memiliki keterkaitan yang kuat dengan tujuan tersebut.
4. Mengkaji dampak AI terhadap bank serta terhadap konsumen yang menggunakan layanan atau produk berbasis AI, serta mengidentifikasi dan mengukur potensi risiko kesalahan yang mungkin dihasilkan oleh AI untuk memastikan keandalan dan akuntabilitas sistem.

In model development, it is important to consider the objectives to be achieved and performance metrics that include accuracy, efficiency in the training process, and adequate datasets. Aspects that need to be considered in effective AI planning and development include:

1. Organized model management, by building and updating a comprehensive inventory of models within the bank to ensure traceability (audit trail) and efficiency in AI implementation.
2. Validation and quality of training data, by ensuring data integrity through rigorous evaluation of the data collection process, as well as ensuring that the training data accurately reflects the population where AI will be used. If there are gaps in the data, model usage may be limited to minimize bias and enhance reliability.
3. Establishing appropriate goals for the AI system and ensuring that every component, from algorithms to data inputs, has a strong connection to those goals.
4. Assessing the impact of AI on banks and on consumers using AI-based services or products, as well as identifying and measuring potential risks of errors that may arise from AI to ensure the reliability and accountability of the system.

5. Sebelum sistem AI (termasuk algoritma AI) diterapkan dan selama penerapan sistem AI (termasuk algoritma AI), bank:

   a. Mendokumentasikan posisi sistem AI dalam arsitektur TI bank, serta mengelola interaksi AI dengan sistem lain untuk memitigasi risiko dan dampak potensial.

   b. Menetapkan ukuran kinerja yang jelas dan terdokumentasi:
      - Yang harus dipenuhi sebelum sistem AI diterapkan atau diperbarui, sesuai dengan sumber daya dan toleransi risiko bank.
      - Sesuai dengan sumber daya dan toleransi risiko bank, selama sistem AI diterapkan.
      - Untuk menjaga kinerja sistem AI selama penerapan.

   c. Memiliki sistem pengendalian dan tata kelola yang memadai, dalam hal terhadap sistem AI dilakukan penyesuaian dan pembaruan selama penerapannya.

   d. Menetapkan mekanisme dan ukuran dalam pemantauan kinerja sistem AI secara terstruktur dan sesuai dengan ukuran yang ditetapkan bank, untuk memastikan sistem AI mempertahankan tingkat kinerja yang diinginkan.

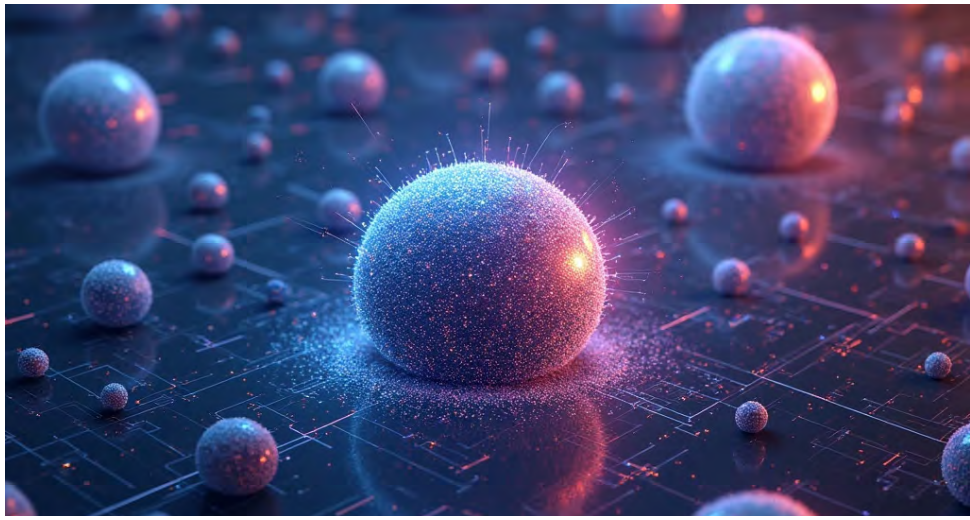
5. Before the AI system (including AI algorithms) is implemented and during the implementation of the AI system (including AI algorithms), banks:

   a. Documents the position of the AI system within the bank's IT architecture, as well as managing the interactions of AI with other systems to mitigate potential risks and impacts.

   b. Establishes clear and documented performance metrics:
      - That must be met before the AI system is implemented or updated, in accordance with the bank's resources and risk tolerance.
      - In line with the bank's resources and risk tolerance, during the implementation of the AI system.
      - To maintain the performance of the AI system during implementation.

   c. Having adequate control and governance systems, in terms of making adjustments and updates to the AI system during its implementation.

   d. Establishing mechanisms and metrics for structured performance monitoring of the AI system in accordance with the metrics set by the bank, to ensure that the AI system maintains the desired level of performance.



   e. Menetapkan mekanisme evaluasi kinerja sistem AI secara berkala yang mencakup aspek penggunaan, operasional, dan pengelolaan risiko, yang dapat dipengaruhi dari lingkungan internal dan eksternal.

6. Rekayasa model, kalibrasi, dan interpretasi yang mungkin dapat mengandung bias yang memengaruhi kewajaran sistem AI, seperti bias eksperimen, yaitu ketika model secara tidak sadar dipengaruhi oleh keyakinan atau asumsi awal dari perancangnya (OECD, 2019).

   e. Establishing mechanisms for periodic performance evaluation of the AI system that includes aspects of usage, operations, and risk management, which can be influenced by internal and external environments.

6. Model engineering, calibration, and interpretation that may contain biases affecting the fairness of the AI system, such as experimental bias, which occurs when the model is unconsciously influenced by the beliefs or initial assumptions of its designers (OECD, 2019).

C.4. Pengujian

Dalam tahap pengujian, dilakukan verifikasi dan validasi sistem AI yang bertujuan untuk mereviu dan menilai metode serta prosedur untuk menilai apakah kinerja model dan sistem AI pada berbagai dimensi dan pertimbangan berfungsi sebagaimana mestinya sehingga keputusan yang dibuat oleh sistem AI atau output yang dihasilkan dapat dipercaya.

Verifikasi merupakan proses yang bertujuan untuk memastikan bahwa sistem AI telah dibangun dan dikembangkan sesuai dengan spesifikasi desain dan persyaratan yang ditentukan serta memiliki basis pengetahuan yang lengkap dan konsisten secara internal. Disamping itu, validasi merupakan proses yang bertujuan untuk memastikan bahwa keluaran (*output*) yang dihasilkan oleh sistem AI adalah benar dan memenuhi kebutuhan pengguna dimana keluaran (*output*) yang dihasilkan oleh sistem AI sebanding dengan keluaran (*output*) yang dihasilkan oleh manusia, ketika diberikan input yang sama.

Agar sistem AI mampu beroperasi dalam tingkat kinerja yang memadai dan konsisten dari waktu ke waktu, serta dapat menghindari bias dan faktor lain yang memengaruhi kewajaran sistem AI:

1. Bank memastikan untuk memiliki dan mengembangkan metode verifikasi dan validasi yang sesuai untuk memastikan kinerja sistem AI beroperasi secara memadai.
2. Sebelum memutuskan untuk menerapkan sistem AI atau pembaruan dari sistem AI, bank meriviu dan memastikan bahwa dampak sistem AI dapat diterima, sesuai etika dan memenuhi target kinerja untuk penerapan, yang didukung dengan dokumentasi yang memadai.
3. Bank memiliki dan mengembangkan metode verifikasi dan validasi yang sesuai untuk memastikan kinerja algoritma secara memadai, mengingat kinerja algoritma merupakan aspek penting dalam pengembangan dan pengendalian kualitas sistem AI.
4. Dalam verifikasi dan validasi, bank memiliki ukuran dan dokumentasi memadai terkait dengan kualitas data yang digunakan oleh sistem AI, termasuk pemantauan terkait kualitas data untuk memastikan risiko terkait data dapat diterima sesuai toleransi risiko bank dan berperan optimal dalam kinerja sistem AI.
5. Bank memiliki ukuran dan dokumentasi serta melakukan pemantauan secara memadai terhadap dampak sistem AI, dalam upaya memastikan dampak dari sistem AI yang dimiliki dan digunakan bank masih tetap dapat diterima dan dilakukan secara bertanggung jawab.

C.4. Testing

In the testing phase, verification and validation of the AI system are conducted to review and assess the methods and procedures to determine whether the performance of the model and AI system functions as intended across various dimensions and considerations, ensuring that decisions made by the AI system or outputs generated can be trusted.

Verification is the process aimed at ensuring that the AI system has been built and developed according to specified design specifications and requirements, and that it has a complete and internally consistent knowledge base. Additionally, validation is the process aimed at ensuring that the outputs generated by the AI system are correct and meet user needs, where the outputs produced by the AI system are comparable to those produced by humans when given the same inputs.

To ensure that the AI system can operate at an adequate and consistent level of performance over time, while avoiding biases and other factors that affect the fairness of the AI system:

1. The bank ensures to have and develop appropriate verification and validation methods to ensure that the performance of the AI system operates adequately.
2. Before deciding to implement an AI system or updates to the AI system, the bank reviews and ensures that the impact of the AI system is acceptable, ethical, and meets performance targets for implementation, supported by adequate documentation.
3. The bank posesses and develops appropriate verification and validation methods to ensure the adequate performance of algorithms, considering that algorithm performance is a crucial aspect in the development and quality control of AI systems.
4. In verification and validation, the bank has adequate metrics and documentation related to the quality of the data used by the AI system, including monitoring of data quality to ensure that risks associated with the data are acceptable according to the bank's risk tolerance and optimally contribute to the performance of the AI system.
5. The bank has metrics and documentation and conducts adequate monitoring of the impact of the AI system, in an effort to ensure that the impact of the AI systems owned and used by the bank remains acceptable and is carried out responsibly.

C.5. Implementasi

Implementasi (penerapan) sistem AI dalam operasional bank (*live production*) mencakup uji coba (*piloting*), pengecekan kompatibilitas dengan sistem lama, memastikan kepatuhan terhadap regulasi, mengelola perubahan dalam organisasi, serta mengevaluasi pengalaman pengguna (OECD 2019), agar sistem AI dapat memiliki kinerja yang dapat dipercaya dan dapat diandalkan.

Aspek-aspek dimaksud perlu menjadi perhatian bank agar implementasi sistem AI tidak memiliki risiko negatif. Risiko negatif AI dapat terjadi dari implementasi sistem AI yang tidak akurat, tidak dapat diandalkan, kurang mampu beradaptasi dengan data dan situasi di luar pelatihannya atau penerapan yang tidak menyeluruh dimana beberapa kelompok pemangku kepentingan mungkin tidak dapat menggunakan atau merasakan manfaat dari sistem AI yang telah diterapkan, yang berdampak berkurangnya kepercayaan terhadap output yang dihasilkan sistem AI. Disamping itu, bank perlu menetapkan batasan kinerja AI sesuai kebutuhan, tujuan, dan lingkungan bisnis yang akan dilakukan bank.

C.5. Implementation

The implementation of the AI system in the bank's operations (live production) includes piloting, checking compatibility with legacy systems, ensuring compliance with regulations, managing organizational change, and evaluating user experience (OECD 2019), so that the AI system can achieve trustworthy and reliable performance.

These aspects need to be a concern for the bank to ensure that the implementation of the AI system does not have negative risks. Negative risks of AI can arise from the implementation of an inaccurate or unreliable AI system that is less capable of adapting to data and situations outside its training or incomplete application where some stakeholder groups may not be able to use or benefit from the implemented AI system. This impacts a decrease in trust in the outputs generated by the AI system. Additionally, banks need to establish performance limits for their AIs according to their needs, objectives, and business environment they will operate in.

Dalam implementasi sistem AI pada bank, aspek yang perlu menjadi perhatian antara lain:

a. Konsumen (Pengguna/Nasabah)

   Aspek yang perlu menjadi perhatian bank antara lain:

   1. Memiliki kebijakan dan prosedur yang terdokumentasi mengenai pengungkapan, pemberitahuan privasi, dan syarat penggunaan secara jelas mengenai penggunaan AI agar tidak menyesatkan pengguna.
   2. Analisis umpan balik konsumen

      Melakukan analisis umpan balik pengguna, antara lain terhadap aspek berikut:

      - Pemahaman dan persepsi, dengan mengumpulkan umpan balik pengguna dari berbagai sumber (survei, ulasan, media sosial) untuk memahami sentimen dan tren.
      - Identifikasi dan analisis, melalui analisa umpan balik untuk mengidentifikasi masalah dan area perbaikan.
      - *Iterative review* dengan terus memantau reaksi konsumen terhadap perubahan yang diterapkan, mengumpulkan umpan balik baru untuk menilai efektivitas tindakan yang diambil.

In implementing AI systems in banks, these are the aspects that needs to be considered:

a. Consumers (Users/Customers)

   Aspects that need to be considered by the bank include:

   1. Having documented policies and procedures regarding disclosure, privacy notices, and terms of use clearly related to the use of AI to avoid misleading users.
   2. Consumer feedback analysis

      Conducting user feedback analysis, including the following aspects:

      - Und erstanding and perception, by collecting user feedback from various sources (surveys, reviews, social media) to understand sentiment and trends.
      - Identification and analysis, through feedback analysis to identify issues and areas for improvement.
      - Iterative review by continuously monitoring consumer reactions to implemented changes, collecting new feedback to assess the effectiveness of actions taken.

3. Pelindungan data pribadi

   Bank harus memperhatikan regulasi terkait pelindungan data pribadi dalam penerapan sistem AI, baik mengacu kepada UU PDP serta panduan/regulasi OJK yang terkait.

   Dalam pemenuhan UU PDP dimaksud dan berbagai peraturan terkait, bank harus menyiapkan roadmap untuk mempersiapkan diri, yang mencakup:

   a. Strategi, dalam menentukan arah, tingkat dan preferensi risiko bank dalam membangun tata kelola privasi.
   b. Struktur Organisasi dan Tanggung Jawab, dalam menerapkan strategi privasi secara efektif, dan melibatkan berbagai disiplin ilmu, serta mencakup pembentukan tim privasi serta Petugas Perlindungan Data untuk implementasi secara efektif.
   c. Pengelolaan Data serta Kebijakan dan Prosedur Transfer Data, melalaui kolaborasi dengan unit bisnis guna memastikan data terlindungi, teratur, terkelola, dan dimanfaatkan secara optimal sesuai dengan strategi bank, termasuk mencakup penyelesaian tantangan teknis seperti permintaan akses data, penyimpanan data, hak untuk dihapus, notifikasi pelanggaran, serta transfer data lintas batas (negara) dan kepada pihak ketiga.
   d. Pelatihan, Komunikasi, dan Kesadaran terkait pelindungan data pribadi melalui peningkatan pemahaman dan kepatuhan ketentuan bagi seluruh karyawan dalam organisasi.
   e. Audit Penilaian Dampak Privasi dan Sertifikasi *Privacy by Design*.
   f. Mengintegrasikan prinsip privasi ke dalam setiap aktivitas bank dengan panduan yang jelas dan praktis selama pengembangan produk

3. Personal data protection

   The bank must pay attention to regulations related to personal data protection in the implementation of AI systems, both referring to the Personal Data Protection Law (UU PDP) and relevant OJK guidelines/regulations.

   In fulfilling the aforementioned Personal Data Protection Law and various related regulations, the bank must prepare a roadmap for readiness, which includes:

   a. Strategy, in determining the direction, level, and risk preferences of the bank in establishing privacy governance.
   b. Organizational Structure and Responsibilities, in effectively implementing privacy strategies, involving various disciplines, and including the establishment of a privacy team as well as a Data Protection Officer for effective implementation.
   c. Data Management and Data Transfer Policies and Procedures, through collaboration with business units to ensure that data is protected, organized, managed, and utilized optimally in accordance with the bank's strategy, including addressing technical challenges such as data access requests, data storage, the right to be forgotten, breach notifications, as well as cross-border data transfers (to other countries) and to third parties.
   d. Training, Communication, and Awareness related to personal data protection through enhancing understanding and compliance with regulations for all employees within the organization.
   e. Privacy Impact Assessment Audits and Privacy by Design Certification.
   f. Integrating privacy principles into every bank activity with clear and practical guidelines during the development of new products or

atau layanan baru (*privacy by design*), serta mengevaluasi sistem yang sudah ada atau baru dengan metode penilaian dampak privasi (*privacy impact assessment*) yang telah ditetapkan.

g. Inventarisasi Pemrosesan Data, sebagai fondasi utama dari program privasi dan menjadi kewajiban hukum sebagaimana diatur dalam UU PDP.

Perbankan harus mempersiapkan diri secara optimal dalam menindaklanjuti regulasi mengenai pelindungan data pribadi dimaksud, terlebih jika bank mengimplementasikan sistem AI (*privacy by design*) yakni dengan melakukan pendekatan desain sistem yang mengintegrasikan perlindungan privasi ke dalam setiap tahap pengembangan dan operasi sistem AI, mulai dari konsep awal hingga implementasi dan pemeliharaan. Konsep ini bertujuan untuk memastikan bahwa privasi pengguna menjadi prioritas utama, bukan sekadar tambahan atau perbaikan setelah sistem sudah berjalan.

b. Mitra dengan Pihak Ketiga

Dari hasil survei yang dilakukan di berbagai negara, bank bermitra dengan pihak ketiga dalam mengimplementasikan teknologi AI. Saat mengimplementasikan teknologi AI, meski menggunakan pihak ketiga, bank harus tetap mengikuti pendekatan implementasi yang telah distandardisasi. Pendekatan standarisasi implementasi ketika bank menggunakan pihak ketiga dalam implementasi AI:

1. Fase 1: Strategi

   Strategi merupakan hal penting untuk menyelaraskan tujuan dan target hasil dari pengimplementasian AI. Strategi dalam hal ini dapat berupa penetapan cakupan proyek, menentukan peran dan tanggung jawab, serta menentukan keputusan pengembangan teknologi AI.

2. Fase 2: Design/Rancangan

   Rancangan dalam model operasi bank dalam fase ini melibatkan:

   - Meninjau kembali tata kelola data (*data governance*) termasuk tinjauan mengenai kualitas data, keabsahan data, serta skema/proses operasional penggunaan AI yang telah ditentukan.
   - Evaluasi proses/skema pengaplikasian AI untuk mengidentifikasi kemungkinan "*re-design*" atau "*re-engineering*" skema yang lebih efektif.
   - Menilai apakah bank memiliki tingkat pengetahuan, keterampilan, dan teknologi yang tepat untuk dapat mempertahankan dan mengevaluasi model AI pasca implementasi.

services (privacy by design), as well as evaluating existing or new systems with established privacy impact assessment methods.

g. Data Processing Inventory, as a fundamental foundation of the privacy program and a legal obligation as stipulated in the Personal Data Protection Law.

Banks must optimally prepare themselves to follow up on regulations regarding personal data protection, especially if the bank implements AI systems (privacy by design) by adopting a system design approach that integrates privacy protection into every stage of AI system development and operation, from the initial concept to implementation and maintenance. This concept aims to ensure that user privacy is a top priority, not just an addition or improvement after the system is already operational.

b. Third-Party Partnerships

From the survey results conducted in various countries, banks partner with third parties in implementing AI technology. When implementing AI technology, even when using third parties, banks must still adhere to a standardized implementation approach. The standardized implementation approach when banks use third parties in AI implementation includes:

1. Phase 1: Strategy

   Strategy is essential to align the objectives and target outcomes of AI implementation. The strategy in this case may include defining the project scope, determining roles and responsibilities, as well as making decisions regarding AI technology development.

2. Phase 2: Design

   Designing in the bank's operating model in this phase involves:

   - Reviewing data governance, including an assessment of data quality, data validity, and the operational schemes/processes for using AI that have been established.
   - Evaluating the processes/schemes of AI application to identify potential "re-design" or "re-engineering" of more effective schemes.
   - Assessing whether the bank has the appropriate level of knowledge, skills, and technology to maintain and evaluate the AI model post-implementation.

- Bekerja sama dengan vendor untuk menyesuaikan solusi AI agar memenuhi persyaratan bisnis atau peraturan yang relevan.

Fase kedua ini memerlukan diskusi dan komunikasi ekstensif dengan berbagai pihak serta semua keputusan desain harus didokumentasikan dengan baik.

3. Fase 3: Model

   Pada fase ini, teknologi model AI akan dibangun/dilatih dengan langkah-langkah, yaitu 1) mengklasifikasikan domain, 2) mengklasifikasikan tujuan dari program AI, dan 3) melatih AI dalam merekognisi suatu pola yang relevan dengan tugas yang diberikan. Tim harus dapat membuat metodologi dan prosedur pelatihan sesuai dengan kerangka tata kelola AI, seperti memeriksa keselarasan dengan kebijakan internal bank, hingga persyaratan bisnis dan regulasi penggunaan AI.

4. Fase 4: Evaluasi/ *Testing*

   Fase ini merupakan fase "*quality control*" untuk mengevaluasi tingkat akurasi dan presisi kinerja teknologi AI serta mengevaluasi segala hal yang berkaitan dengan keamanan siber untuk meminimalkan potensi kelemahan AI yang dapat menurunkan tingkat kepercayaan kepada teknologi AI yang digunakan.

- Collaborating with vendors to customize AI solutions to meet relevant business or regulatory requirements.

This second phase requires extensive discussion and communication with various parties, and all design decisions must be well documented.

3. Phase 3: Model

   In this phase, the AI model technology will be built/trained with the following steps: 1) classifying the domain, 2) classifying the objectives of the AI program, and 3) training the AI to recognize patterns relevant to the assigned tasks. The team must be able to create methodologies and training procedures in accordance with the AI governance framework, such as checking alignment with internal bank policies, as well as business requirements and regulations for using AI.

4. Phase 4: Evaluation/Testing

   This phase is the "quality control" phase to evaluate the accuracy and precision of AI technology performance, as well as to assess all aspects related to cybersecurity in order to minimize potential vulnerabilities in AI that could undermine trust in the technology being used.

Proses evaluasi/tes harus melalui tinjauan pemangku kepentingan, tinjauan teknologi, evaluasi oleh pakar di bidangnya, tinjauan ahli teknologi, evaluasi para ahli dibidang terkait dan dilakukan dengan proses yang sesuai standar bank.

Hasil evaluasi/tes harus berupa keputusan mengenai kesesuaian AI (apakah model AI sudah sesuai atau perlu di "*re-model*").

5. Fase 5: Penerapan & Pengembangan

   Implementasi AI harus mengikuti rencana implementasi yang telah ditetapkan pada fase-fase sebelumnya.

   Pasca implementasi, teknologi AI perlu di tinjau selama periode tertentu guna menyesuaikan teknologi AI dengan kemungkinan adanya perubahan persyaratan dari regulator.

The evaluation/testing process must undergo stakeholder reviews, technology assessments, evaluations by subject matter experts, expert technology reviews, and evaluations by specialists in related fields, all conducted in accordance with the bank's established standards.

The results of the evaluation/testing must yield a decision regarding the suitability of the AI (whether the AI model is appropriate or needs to be "re-modeled").

5. Phase 5: Implementation & Development

   The implementation of AI must follow the implementation plan established in the previous phases.

   Post-implementation, the AI technology needs to be reviewed over a certain period to adjust the AI technology in response to potential changes in regulatory requirements;

Teknologi AI yang telah diterapkan juga harus terus dipantau kinerjanya, didukung dengan respons insiden yang memadai, serta memungkinkan umpan balik dari pengguna untuk perbaikan berkelanjutan.

Bank juga harus melaksanakan manajemen risiko terhadap pihak ketiga, antara lain dengan:

- Melaksanakan uji tuntas berbasis risiko terhadap vendor dengan menilai berbagai aspek, seperti pengalaman dalam bisnis, rencana strategis, manajemen risiko dan pengendalian, keamanan informasi, kondisi keuangan. Kepatuhan terhadap hukum dan regulasi, serta ketahanan operasional, termasuk rencana kesinambungan bisnis.
- Mengharuskan vendor pihak ketiga untuk memberikan penjelasan yang transparan dan rinci mengenai cara kerja algoritma, kriteria yang digunakan, serta informasi lain yang diperlukan untuk memverifikasi hasil *tools* AI dan kepatuhannya terhadap regulasi.
- Melakukan inspeksi serta pengujian awal dan berkala terhadap a *tools* lat AI.
- Meninjau batasan penggunaan pada perangkat lunak pihak ketiga yang berlisensi.

The implemented AI technology must also be continuously monitored for performance, supported by adequate incident response mechanisms, and allow for user feedback to facilitate ongoing improvements.

The bank must also implement risk management for third parties, including:

- Conducting risk-based due diligence on vendors by assessing various aspects, such as business experience, strategic plans, risk management and controls, information security, financial condition, compliance with laws and regulations, as well as operational resilience, including business continuity plans.
- Requiring third-party vendors to provide transparent and detailed explanations regarding how the algorithms work, the criteria used, as well as other information necessary to verify the results of AI tools and their compliance with regulations.
- Conducting inspections and initial as well as periodic testing of AI tools.
- Reviewing usage limitations on licensed third-party software.

- Memastikan vendor pihak ketiga memberikan jaminan bahwa alat AI mereka mematuhi hukum dan regulasi yang berlaku.
- Memastikan terhadap pelindungan konsumen dan keamanan data.

c. Lingkungan

Aspek yang perlu menjadi perhatian bank antara lain:

1. *Ecosystem partners* (mitra ekosistem).

   Keberhasilan implementasi AI sangat dipengaruhi oleh kolaborasi dalam ekosistem bisnis. Dukungan yang dibutuhkan tidak hanya berasal dari mitra di sektor teknologi, tetapi juga dari berbagai elemen dalam ekosistem bisnis yang lebih luas, yang mencakup pemasok, pesaing, pelanggan, serta mitra aliansi lintas industri, yang bersama-sama berkontribusi dalam memastikan penerapan AI berjalan secara optimal dan efektif.

2. Aspek Etika dan Moral

   Aspek etika dan moral merupakan aspek penting dalam penerapan AI. bank harus memastikan bahwa aplikasi AI telah dikembangkan berdasarkan prinsip-prinsip etika dan tidak mengandung bias yang tidak diketahui. Etika AI didefinisikan serangkaian nilai, prinsip, dan teknik yang menggunakan standar yang diterima secara luas tentang benar dan salah untuk memandu perilaku moral dalam pengembangan dan penggunaan teknologi AI. Etika AI dapat membantu bank memastikan bahwa penggunaan teknologi mereka selaras dengan nilai-nilai pada bank. Transparansi, bias, dan diskriminasi merupakan beberapa tantangan yang muncul saat mengembangkan sistem AI. Beberapa negara dan lembaga telah mendefinisikan prinsip-prinsip utama yang harus mendasari penggunaan AI yang antara lain terkait dengan keadilan, transparansi, keselamatan dan keamanan, akuntabilitas, privasi dan pelindungan data, pengawasan manusia.

3. Peraturan

   Regulasi dari pemerintah maupun regulator menunjukkan perhatian terhadap isu etika dan moral, serta memberikan arahan yang membentuk cara pengembangan aplikasi AI, antara lain terkait

- Ensuring that third-party vendors provide assurances that their AI tools comply with applicable laws and regulations.
- Ensuring consumer protection and data security.

c. Environment

Aspects that banks need to take into consideration includes:

1. Ecosystem partners.

   The success of AI implementation is greatly influenced by collaboration within the business ecosystem. The support needed comes not only from partners in the technology sector but also from various elements within the broader business ecosystem, which includes suppliers, competitors, customers, and cross-industry alliance partners, all of whom contribute to ensuring that AI deployment runs optimally and effectively.

2. Ethical and moral aspects

   Ethical and moral aspects are important in the implementation of AI. Banks must ensure that AI applications have been developed based on ethical principles and do not contain undisclosed biases. AI ethics is defined as a set of values, principles, and techniques that use widely accepted standards of right and wrong to guide moral behavior in the development and use of AI technology. AI ethics can help banks ensure that their use of technology aligns with the bank's values. Transparency, bias, and discrimination are some of the challenges that arise when developing AI systems. Several countries and institutions have defined key principles that should underlie the use of AI, which include fairness, transparency, safety and security, accountability, privacy and data protection, and human oversight.

3. Regulation

   Regulations from the government and regulators indicate a focus on ethical and moral issues, as well as providing guidance that shapes the development of AI applications, including aspects related to

pelindungan data pribadi, kekayaan intelektual, reputasi dan berbagai persyaratan lain, yang perlu menjadi perhatian bank dalam mengadopsi AI.

4. Tekanan Lingkungan

   Faktor penting lain yang mendorong bank dalam adopsi AI adalah tekanan kompetitif (*competitive pressure*), yang mengacu bagaimana bank dipengaruhi oleh pesaingnya dan merespons tindakan pesaing untuk mempertahankan atau memperoleh keunggulan. Ancaman kehilangan daya saing memotivasi bank untuk terus beradaptasi dan mengadopsi inovasi seperti AI. Selain itu, permintaan nasabah (konsumen) juga menjadi pendorong kuat, dimana nasabah mengharapkan layanan yang lebih personal, sehingga bank terdorong untuk menerapkan AI demi memenuhi dan melampaui ekspektasi nasabah.

d. Bank yang Beroperasi Lintas Yurisdiksi

   Bagi bank yang memiliki kantor cabang di luar negeri, harus memperhatikan regulasi terkait AI pada yurisdiksi setempat agar adopsi AI di kantor cabang di luar negeri memenuhi ketentuan.

personal data protection, intellectual property, reputation, and various other requirements that banks need to consider when adopting AI.

4. Environmental Pressure

   Another important factor driving banks in the adoption of AI is competitive pressure, which refers to how banks are influenced by their competitors and respond to competitors' actions to maintain or gain an advantage. The threat of losing competitiveness motivates banks to continuously adapt and adopt innovations such as AI. In addition, customer demand also serves as a strong driver, where customers expect more personalized services, prompting banks to implement AI in order to meet and exceed customer expectations.

d. Banks Operating Across Jurisdictions

   For banks with branches abroad, it is essential to pay attention to regulations related to AI in the local jurisdiction so that the adoption of AI in overseas branches complies with the applicable provisions.

C.6. Pemantauan, Pemeliharaan dan Evaluasi

Melibatkan pengoperasian sistem AI serta evaluasi terus-menerus terhadap rekomendasi dan dampaknya, baik yang diharapkan maupun yang tidak, dengan mempertimbangkan tujuan serta aspek etika. Fase ini bertujuan untuk mengidentifikasi masalah dan melakukan penyesuaian, baik dengan kembali ke fase sebelumnya maupun, dan jika diperlukan dilakukan pemusnahan (pengakhiran atau penghentian) penggunaan sistem AI.

Pemantauan dan pemeliharaan secara rutin diperlukan untuk menilai output dan dampak dari sistem AI, karena kinerja model dapat menurun seiring waktu akibat perubahan data (*data drift*) atau lingkungan. Sistem yang beroperasi dengan data atau lingkungan yang terus berubah memerlukan penyesuaian berkelanjutan dan membutuhkan biaya yang dapat memengaruhi kelayakan penggunaan AI dalam berbagai kasus (AAI, 2023). Beberapa aspek yang perlu dipantau antara lain meliputi kesehatan layanan (*service health*), kualitas dan

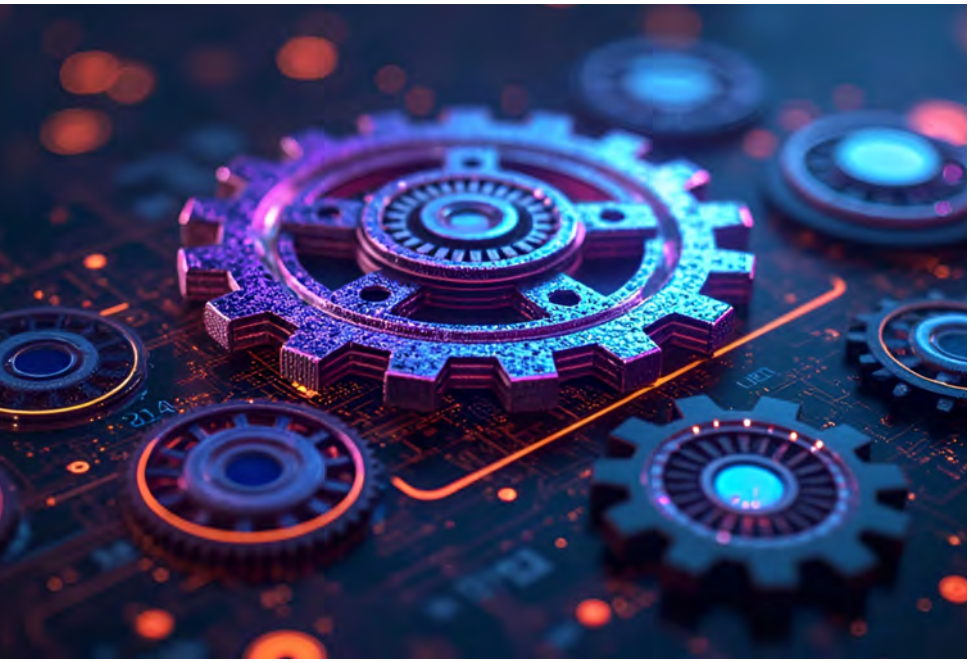
C.6. Oversight, Maintenance and Evaluation

This covers the operation of AI systems as well as continuous evaluation of recommendations and their impacts, both expected and unexpected, while considering objectives and ethical aspects. This phase aims to identify issues and make adjustments, either by reverting to previous phases or, if necessary, terminating (discontinuing) the use of the AI system.

Routine monitoring and maintenance are necessary to assess the output and impact of AI systems, as model performance can decline over time due to data changes (data drift) or environmental shifts. Systems operating with continuously changing data or environments require ongoing adjustments and incur costs that may affect the feasibility of using AI in various cases (AAI, 2023). Several aspects that need to be monitored include: service health, data quality and integrity, data drift, overall model performance and by

integritas data, pergeseran/perubahan data (*data drift*), kinerja model secara umum dan per segmen, bias dan keadilan, skala penerapan, *interpretability* (sejauh mana manusia dapat memahami dan menjelaskan bagaimana sebuah model AI mengambil keputusan atau membuat prediksi), peringatan untuk kejadian tertentu (*alerting for events*), pelaporan, manajemen infrastruktur, serta bagaimana sistem menangani data yang tidak biasa (*outliers*).

segment, bias and fairness, scalability of implementation, interpretability (the extent to which humans can understand and explain how an AI model makes decisions or predictions), alerting for events, reporting, infrastructure management, as well as how the system handles unusual data (outliers).



Pemantauan langsung (*live monitoring*) dilakukan bank dengan melakukan pengawasan oleh manusia (secara *high level* dan *day to* day) berbasis risiko, serta menerapkan pencatatan otomatis dan *audit trail*. Memastikan *human-in-the-loop* berjalan, dimana manusia tetap terlibat dalam proses pengambilan keputusan atau pelatihan model AI guna memastikan bahwa sistem AI tidak bekerja dengan sepenuhnya otomatis, tetapi tetap mendapatkan intervensi, validasi, atau koreksi dari manusia.

Live monitoring is conducted by banks through human oversight (at a high level and day-to-day) based on risk, as well as implementing automatic logging and audit trails. Ensuring that human-in-the-loop processes are in place, where humans remain involved in the decision-making process or training of AI models to ensure that the AI system does not operate fully automatically but still receives intervention, validation, or correction from humans.

Upaya perbaikan yang dilakukan bank didokumentasikan dengan tepat dalam inventaris model (*model inventory*). Agar sistem AI tetap andal dan relevan seiring waktu, GAO (2021) mengidentifikasi praktik utama untuk memantau kinerja serta menilai keberlanjutan dan potensi pengembangan AI.

a. Pemantauan Kinerja Secara Berkelanjutan

- Perencanaan Pemantauan: Menyusun rencana untuk pemantauan berkala atau berkelanjutan terhadap sistem AI guna memastikan sistem AI berfungsi sesuai yang diharapkan.
- *Drift* (Perubahan Data/Model): Menentukan batas perubahan data dan model yang masih dapat diterima agar AI tetap menghasilkan output yang diinginkan.

The improvement efforts made by the bank are properly documented in the model inventory. To ensure that AI systems remain reliable and relevant over time, GAO (2021) identifies key practices for monitoring performance as well as assessing the sustainability and potential development of AI.

a. Continous Performance Oversight

- Monitoring Planning: Developing a plan for periodic or continuous monitoring of AI systems to ensure that the AI system functions as expected.
- Drift (Data/Model Changes): Defining the acceptable limits of data and model changes to ensure that AI continues to produce the desired output.

- Keterlacakan: Mendokumentasikan hasil kegiatan pemantauan serta tindakan korektif yang dilakukan untuk meningkatkan transparansi dan akuntabilitas.

b. Penilaian Keberlanjutan dan Pengembangan AI

- Evaluasi Berkelanjutan: Menilai kegunaan sistem AI secara rutin untuk memastikan tetap relevan dengan kondisi terkini.
- Skala: Mengidentifikasi kondisi yang memungkinkan sistem AI diperluas atau dikembangkan lebih lanjut di luar penggunaan saat ini.

Untuk memastikan sistem AI menghasilkan output yang sesuai dengan tujuan program, GAO (2021) mengidentifikasi praktik utama dalam memastikan AI memenuhi tujuannya:

a. Evaluasi Kinerja pada Tingkat Komponen

- Dokumentasi: Mendokumentasikan komponen model dan non-model, termasuk spesifikasi dan parameter operasionalnya.
- Metrik/Ukuran Kinerja: Menentukan metrik kinerja yang presisi, konsisten, dan dapat direproduksi.
- Penilaian Kinerja: Mengevaluasi setiap komponen berdasarkan metrik yang telah ditetapkan untuk memastikan fungsinya sesuai dengan tujuan program.

- Traceability: Documenting the results of monitoring activities and corrective actions taken to enhance transparency and accountability.

b. Penilaian Keberlanjutan dan Pengembangan AI

- Continuous Evaluation: Regularly assessing the usefulness of AI systems to ensure they remain relevant to current conditions.
- Scale: Identifying conditions that allow the AI system to be expanded or further developed beyond its current use.

To ensure that AI systems produce outputs aligned with program objectives, GAO (2021) identifies key practices for ensuring that AI meets its goals:

a. Performance Evaluation at the Component Level

- Documentation: Documenting model and non-model components, including specifications and operational parameters.
- Performance Metrics/Measures: Defining precise, consistent, and reproducible performance metrics.
- Performance Assessment: Evaluating each component based on established metrics to ensure its function aligns with program objectives.

- Evaluasi Output: Menilai apakah output dari setiap komponen sesuai dengan konteks operasional sistem AI.

b. Evaluasi Kinerja pada Tingkat Sistem

- Dokumentasi: Mendokumentasikan metode jelas, metrik kinerja, dan hasil dari sistem AI untuk memberikan transparansi dalam kinerjanya.
- Metrik Kinerja: Menentukan metrik kinerja yang presisi, konsisten, dan dapat direproduksi.
- Penilaian Kinerja: Mengevaluasi sistem berdasarkan metrik yang telah ditentukan untuk memastikan AI berfungsi sebagaimana mestinya dan berdaya (tangguh).

- Output Evaluation: Assessing whether the output from each component aligns with the operational context of the AI system.

b. Performance Evaluation at the System Level

- Documentation: Documenting clear methods, performance metrics, and results of the AI system to provide transparency in its performance.
- Performance Metrics: Defining precise, consistent, and reproducible performance metrics.
- Performance Assessment: Evaluating the system based on established metrics to ensure that AI functions as intended and is resilient.

- Identifikasi Bias: Mengidentifikasi potensi bias, ketidakadilan, dan dampak sosial lainnya yang mungkin timbul dari sistem AI.
- Pengawasan Manusia: Menentukan dan mengembangkan prosedur untuk pengawasan manusia terhadap sistem AI guna memastikan akuntabilitas.

C.7. Pemusnahan

Pemusnahan (pengakhiran) dilakukan dengan menghentikan penggunaan sistem AI secara permanen (diakhiri siklus hidupnya) karena tidak lagi efektif, usang, atau berisiko, dan upaya perbaikan dan pembaruan tidak lagi cukup untuk memenuhi persyaratan baru. Dalam tahapan ini, sistem AI dapat dihentikan (dinonaktifkan) atau digantikan dengan sistem AI yang baru.

Pemusnahan (pengakhiran) juga termasuk penghentian terhadap algoritma atau data, sehingga algoritma atau data tidak lagi digunakan dan tidak lagi tersedia untuk digunakan atau agar tidak disalahgunakan.

Aspek-aspek yang menjadi perhatian bank dalam pemusnahan (pengakhiran atau penghentian) sistem AI agar pengelolaan siklus hidup sistem AI berjalan dengan efektif, mengurangi risiko, dan memaksimalkan manfaat dari investasi teknologi yang dilakukan, antara lain:

1. Pemusnahan sistem AI oleh bank dilakukan dengan cara yang meminimalkan potensi gangguan dan risiko terhadap bank. Karenanya, keputusan pemusnahan harus didasarkan tata kelola yang baik.
2. Bank memiliki prosedur dan mekanisme internal terkait:
   a. Keputusan pemusnahan sistem AI termasuk kebijakan *retention schedule* (kebijakan waktu penyimpanan data sebelum akhirnya diarsipkan/dihapus).
   b. Penghapusan data secara permanen dari media penyimpanan atau didasarkan pada periode penyimpanan data yang telah ditetapkan.
   c. Penghapusan akses pengguna, lisensi yang terkait, dan sebagainya.
3. Melakukan penilaian terhadap penggunaan sistem AI sebelum dilakukan pemusnahan termasuk antara lain kinerja sistem AI berdasarkan ukuran tertentu, kepatuhan terhadap regulasi, keterlibatan dan kepuasan nasabah.
4. Melakukan pembaruan inventaris algoritma, sehingga algoritma yang tidak lagi digunakan tidak lagi tersedia

- Bias Identification: Identifying potential biases, injustices, and other social impacts that may arise from the AI system.
- Human Oversight: Defining and developing procedures for human oversight of the AI system to ensure accountability.

C.7. Termination

Termination (discontinuation) is carried out by permanently stopping the use of the AI system (ending its lifecycle) because it is no longer effective, outdated, or poses risks, and efforts for improvement and updates are no longer sufficient to meet new requirements. In this phase, the AI system may be shut down (deactivated) or replaced with a new AI system.

Termination (discontinuation) also includes the cessation of algorithms or data, ensuring that the algorithms or data are no longer used and are not available for use to prevent misuse.

Aspects of concern for banks in the termination (discontinuation or cessation) of AI systems to ensure effective management of the AI system lifecycle, reduce risks, and maximize the benefits of technology investments include:

1. The termination of AI systems by banks is carried out in a manner that minimizes potential disruptions and risks to the bank. Therefore, the decision to terminate must be based on good governance.
2. The bank has internal procedures and mechanisms related to:
   a. The decision to terminate the AI system includes a retention schedule policy (the policy regarding the duration of data storage before it is ultimately archived or deleted).
   b. Permanent deletion of data from storage media or based on the established data retention period.
   c. Deletion of user access, associated licenses, and other related elements.
3. Conducting an assessment of the use of the AI system before termination, including among others system performance based on certain metrics, compliance with regulations, customer engagement and satisfaction.
4. Updating the inventory of algorithms, so that algorithms that are no longer in use are not available for use or

untuk digunakan atau berpotensi disalahgunakan, dengan melakukan dokumentasi yang memadai antara lain terkait ketergantungan algoritma lain, tidak adanya redundansi data dan versi algoritma serta data diadministrasikan untuk tujuan audit.

5. Memastikan migrasi data yang relevan ke sistem AI yang baru dilakukan secara aman, atau dilakukan pengarsipan/ penyimapanan data dengan tepat. Aspek ini antara lain mencakup identifikasi terhadap data yang perlu dimigrasikan atau disimpan, menjaga integritas dan keamanan data.

6. Komunikasi dan informasi kepada pengguna/nasabah secara aktif dan memadai selama proses pemusnahan sistem AI, antara lain terkait jadwal pemusnahan sistem, perubahan terkait alur kerja, dan opsi dukungan yang tersedia oleh bank selama transisi.

7. Melakukan review/tinjauan paska pemusnahan, dimana dilakukan evaluasi setelah sistem AI dihentikan untuk mengidentifikasi pembelajaran yang bisa diperoleh dari proses pemusnahan , masukan dari pengguna/ nasabah terkait periode trransisi atau rekomendasi lain yang relevan.

8. Mengadministrasikan secara tertib proses dan dokumentasi yang terkait dengan sistem AI yang telah dilakukan pemusnahan, untuk kepentingan audit atau pembelajaran ke depan.

potential misuse, by conducting adequate documentation including dependencies on other algorithms, absence of data redundancy, and versions of algorithms and data administered for audit purposes.

5. Ensuring that the migration of relevant data to the new AI system is carried out securely, or that data archiving/ storage is done appropriately. This aspect includes identifying the data that needs to be migrated or stored, maintaining data integrity and security.

6. Active and adequate communication and information to users/customers during the AI system termination process, including details related to the system termination schedule, changes in workflow, and support options available from the bank during the transition.

7. Conducting a post-termination review, where an evaluation is performed after the AI system has been stopped to identify lessons learned from the termination process, feedback from users/customers regarding the transition period, or other relevant recommendations.

8. Administering the process and documentation related to the AI system that has been terminated in an orderly manner, for audit purposes or future learning.

C.8. Tata Kelola Organisasi, Manajemen Risiko, Kepatuhan

Penerapan tata kelola organisasi, manajemen risiko dan kepatuhan pada bank dalam impelementasi AI di sepanjang siklus hidup AI.

a. Tata Kelola Organisasi

Dalam pengembangan dan penerapan sistem AI pada bank, tetap harus mengacu pada penerapan tata kelola yang baik pada bank yang paling sedikit mencakup prinsip keterbukaan, akuntabilitas, tanggung jawab, independensi dan kewajaran, serta mengikuti perkembangan dinamika industri untuk mendorong penerapan tata kelola yang baik pada bank. Karenanya, bank perlu memastikan pengurus, manajemen dan SDM yang terkait pada bank memahami penggunaan, risiko, dan tanggung jawab dari implementasi sistem AI.

Tata kelola yang baik pada bank merupakan struktur, proses, dan mekanisme pengelolaan bank untuk pencapaian penyelenggaraan kegiatan usaha bank yang memperhatikan kepentingan seluruh pemangku kepentingan yang terkait, menciptakan dan mengoptimalkan nilai perusahaan pada bank secara berkelanjutan, serta berlandaskan ketentuan peraturan perundang-undangan, standar, nilai etika, prinsip, dan praktik yang berlaku umum.

C.8. Organizational Governance, Risk Management, Compliance

The application of organizational governance, risk management, and compliance in banks during the implementation of AI throughout the AI lifecycle.

a. Organizational Governance

In the development and implementation of AI systems in banks, it is essential to adhere to good governance practices that at a minimum encompass principles of transparency, accountability, responsibility, independence, and fairness, as well as keeping up with industry dynamics to promote good governance in banks. Therefore, banks need to ensure that the board of directors, management, and relevant human resources understand the use, risks, and responsibilities associated with the implementation of AI systems.

Good governance in banks is the structure, processes, and management mechanisms of the bank aimed at achieving the conduct of banking business activities that consider the interests of all relevant stakeholders, creating and optimizing sustainable corporate value for the bank, and based on applicable laws and regulations, standards, ethical values, principles, and generally accepted practices.

Disamping itu, seluruh aktivitas yang terkait dengan implementasi sistem AI pada bank perlu didukung dengan kebijakan dan prosedur formal untuk penggunaan sistem AI, termasuk prosedur eskalasi, antara lain berupa surat keputusan, manual, kebijakan atau pedoman bank (*standard operating procedure*), dokumen operasional bank lain, yang disusun sesuai dengan ketentuan perundang-undangan yang berlaku dan sesuai dengan proses bisnis dan mekanisme persetujuan pada bank.

b. Manajemen Risiko

Pihak-pihak yang terlibat dalam pengembangan, penerapan, dan pengelolaan sistem AI (*AI actors*) perlu menerapkan strategi manajemen risiko di sepanjang siklus hidup AI, antara lain

1. Menerapkan manajemen risiko termasuk manajemen risiko khusus AI untuk mengidentifikasi, mengukur, memantau, dan mengendalikan risiko dalam adopsi sistem AI secara sistematis.
2. Manajemen risiko AI melibatkan orang yang berbeda dari tim yang berbeda dengan tanggung jawab yang berbeda. Jika tanggung jawab ini tidak dikoordinasikan secara memadai, kesenjangan dalam cakupan risiko dapat terjadi. Jika risiko tertentu tidak tercakup secara memadai oleh sistem manajemen risiko, risiko tersebut tidak dapat diidentifikasi, yang dapat mengakibatkan penilaian risiko yang salah (misalnya, total risiko sistem AI yang tidak aman dinilai dapat diterima) dan respons risiko yang tidak memadai (misalnya, sistem AI yang tidak aman diterapkan tanpa tindakan pencegahan keselamatan yang memadai). Bank harus mampu mencegah kondisi tersebut dengan mengidentifikasi dan menutup kesenjangan dalam cakupan risiko antara lain dengan pelaksanaan fungsi dan tanggung jawab dalam setiap *lines of defense* secara tertib.
3. Banyak praktik manajemen risiko terlihat baik dalam teori tetapi tidak efektif dalam praktik, misalnya dalam mengidentifikasi risiko yang relevan, menilai dampaknya, atau

In addition, all activities related to the implementation of AI systems in banks need to be supported by formal policies and procedures for the use of AI systems, including escalation procedures, such as decisions in writing, manuals, bank policies or guidelines (standard operating procedures), and other operational documents of the bank that are prepared in accordance with applicable laws and regulations and aligned with business processes and approval mechanisms within the bank.

b. Risk Management

The parties involved in the development, implementation, and management of AI systems (AI actors) need to apply risk management strategies throughout the AI lifecycle, including

1. Implementing risk management, including specific AI risk management, to systematically identify, measure, monitor, and control risks in the adoption of AI systems.
2. AI risk management involves different people from various teams with distinct responsibilities. If these responsibilities are not adequately coordinated, gaps in risk coverage may occur. If certain risks are not sufficiently covered by the risk management system, those risks cannot be identified, which can lead to incorrect risk assessments (e.g., the total risk of an unsafe AI system being deemed acceptable) and inadequate risk responses (e.g., an unsafe AI system being implemented without adequate safety precautions). Banks must be able to prevent such conditions by identifying and closing gaps in risk coverage, including through the orderly implementation of functions and responsibilities within each line of defense.
3. Many risk management practices may appear sound in theory but are ineffective in practice, for example, in identifying relevant risks, assessing their impacts, or

mengantisipasi perubahan lanskap risiko. Penyebabnya beragam, mulai dari ketergantungan pada satu metode, bias kognitif, hingga kesalahan manusia. Evaluasi risiko perlu mencermati berbagai kelemahan dalam praktik yang diterapkan, seperti kurangnya kesiapan menghadapi kesalahan manusia atau upaya untuk melewati sistem keamanan. Diperlukan juga fungsi audit internal yang benar-benar independen dan dengan standar audit yang tepat, agar penerapan manajemen risiko berjalan efektif.

4. *AI impact assessments (*penilaian dampak AI)

   Berdasarkan NIST (2023), *AI impact assessments* merupakan aktivitas yang meliputi penilaian dan evaluasi persyaratan akuntabilitas sistem AI, penanganan bias yang merugikan, pemeriksaan dampak sistem AI, keamanan produk, tanggung jawab, keamanan, dan lain sebagainya.

   Pihak-pihak yang terlibat dalam pengembangan, penerapan, dan pengelolaan sistem AI (*AI actors*) perlu menerapkan strategi manajemen risiko untuk menghindari atau mengurangi berbagai bias di sepanjang siklus hidup AI.

   Penilaian dampak AI menjadi bagian dari program tata kelola AI yang melibatkan berbagai tim internal (sebagai *reviewers*) yang antara lain menangani privasi, risiko, hukum, produk, SDM, *engineering*, dan pemasaran.

   Future of Privacy Forum (2024) mengidentifikasi empat langkah utama dalam proses penilaian dampak AI:

   » Langkah 1: Memulai Penilaian Dampak AI

     Keadaan yang memicu penilaian dampak AI akan bervariasi berdasarkan berbagai faktor, seperti hukum yang berlaku, pengembangan produk atau layanan, penggunaan teknologi baru, dan peran bank dalam ekosistem AI (misalnya, pengembang atau pengguna).

     Insentif untuk melakukan penilaian dampak AI ini dapat muncul pada berbagai titik dalam siklus pengembangan dan penerapan AI. Misalnya,

anticipating changes in the risk landscape. The causes are varied, ranging from reliance on a single method, cognitive biases, to human errors. Risk evaluation needs to pay attention to various weaknesses in the practices applied, such as a lack of preparedness for human errors or attempts to bypass security systems. An independent internal audit function with appropriate audit standards is also required to ensure that risk management implementation is effective.

4. AI impact assessments

   According to NIST (2023), AI impact assessments are activities that include the assessment and evaluation of accountability requirements for AI systems, addressing harmful biases, examining the impacts of AI systems, product security, liability, safety, and others.

   The parties involved in the development, implementation, and management of AI systems (AI actors) need to apply risk management strategies to avoid or mitigate various biases throughout the AI lifecycle.

   AI impact assessments are part of the AI governance program that involves various internal teams (as reviewers) that handle privacy, risk, legal, product, human resources, engineering, and marketing.

   Future of Privacy Forum (2024) Identifying four main steps in the AI impact assessment process:

   » Step 1: Starting the AI Impact Assessment

     The circumstances that trigger an AI impact assessment will vary based on various factors, such as applicable laws, product or service development, the use of new technologies, and the bank's role in the AI ecosystem (e.g., developer or user).

     Incentives to conduct this AI impact assessment may arise at various points in the AI development and implementation cycle. For example, banks may conduct an initial assessment

bank dapat melakukan penilaian awal selama fase pengembangan dan setelah perubahan substansial dilakukan pada model atau sistem. Penilaian tambahan mungkin diperlukan saat diterapkan dalam konteks yang berbeda.

» Langkah 2: Mengumpulkan Informasi tentang Model dan Sistem

Agar dapat menilai risiko yang dapat ditimbulkan oleh sistem AI, bank harus memiliki informasi dasar yang relevan termasuk memahami cara kerja dan cara pembuatan sistem tersebut. Bank yang memperoleh model atau sistem dari pihak ketiga dapat memfokuskan penyelidikannya pada kematangan kerangka tata kelola AI pihak ketiga dan apakah mereka telah menerapkan kontrol untuk meminimalkan risiko tertentu.

Contoh informasi yang relevan antara lain:

- Bagaimana model dilatih.
- Apa saja kasus penggunaan (*use case*) model atau sistem tersebut.
- Siapa saja pengguna akhir, dan di mana sistem akan diterapkan.
- Kasus penggunaan potensial untuk sistem dan mengelompokkannya berdasarkan kategori (misalnya, penggunaan yang diharapkan dan tidak diharapkan).

during the development phase and after substantial changes are made to the model or system. Additional assessments may be required when applied in different contexts.

» Step 2: Collecting Information about the Model and System

To assess the risks that may arise from the AI system, banks must have relevant baseline information, including understanding how the system works and how it was created. Banks that obtain models or systems from third parties can focus their investigation on the maturity of the third-party AI governance framework and whether they have implemented controls to minimize specific risks.

Examples of relevant information include:

- How the model is trained.
- What are the use cases of the model or system.
- Who the end users are, and where the system will be applied.
- Potential use cases for the system and categorizing them based on categories (e.g., expected and unexpected uses).

» Langkah 3: Menilai Risiko dan Manfaat

Reviewers mempertimbangkan informasi tentang model atau sistem untuk menentukan risiko dan manfaat model atau sistem AI untuk kasus penggunaan tertentu. Jenis dan tingkat risiko yang ada akan memengaruhi strategi manajemen risiko yang diterapkan oleh bank untuk mengurangi risiko hingga berada dalam batas yang dapat diterima. Informasi tentang manfaat juga akan memengaruhi keputusan bank tentang bagaimana melanjutkan proyek AI. Sebagai bagian dari langkah ini, bank dapat mempertimbangkan:

- Potensi bahaya risiko dan manfaat yang terkait dengan kasus penggunaan tertentu.
- Bagaimana daftar risiko sistem yang dihasilkan dibandingkan dengan taksonomi risiko perusahaan (dibagi ke dalam kategori seperti tinggi, menengah, dan rendah).
- Jika penilaian dilakukan selama fase pengembangan atau paska-pengembangan, maka penilaian mencakup penyempurnaan selama penerapan, keberadaan dan konteks risiko dalam pengujian sistem.

» Step 3: Assessing Risks and Benefits

Reviewers consider information about the model or system to determine the risks and benefits of the AI model or system for specific use cases. The types and levels of existing risks will influence the risk management strategies applied by banks to reduce risks to acceptable levels. Information about benefits will also affect the bank's decisions on how to proceed with the AI project. As part of this step, banks may consider:

- The potential hazards of risks and benefits associated with specific use cases.
- How the list of system risks generated compares with the enterprise risk taxonomy (divided into categories such as high, medium, and low).
- If the assessment is conducted during the development or post-development phase, then the assessment includes refinements during implementation, the existence and context of risks in system testing.

» Langkah 4: Mengidentifikasi dan Menguji Strategi Manajemen Risiko

- Bank memilih strategi manajemen risiko berdasarkan respons mereka terhadap risiko tertentu yang telah diidentifikasi. Sebagai contoh, jika sebuah bank mengidentifikasi halusinasi sebagai suatu risiko, bank tersebut akan menyesuaikan responsnya terhadap risiko ini, seperti dengan menyesuaikan implementasi AI untuk mengurangi halusinasi dan memastikan pemantauan berkelanjutan terhadap output untuk mendeteksi halusinasi.
- Setelah bank mengidentifikasi strategi manajemen risiko, mereka dapat menguji efektivitasnya selama sistem beroperasi dan menyeimbangkan risiko residual dengan manfaat untuk menentukan langkah-langkah yang tepat untuk ke depannya. Bank kemudian dapat mencatat dan mengoperasionalkan keputusan akhir, seperti memajukan proyek AI ke tahap berikutnya dalam siklus hidupnya.

c. Kepatuhan

Aspek-aspek yang perlu menjadi perhatian bank di sepanjang siklus hidup AI antara lain:

1. Bank memahami regulasi terkait sistem AI dan regulasi lain terkait adopsi TI oleh bank, serta memastikan kepatuhannya terhadap regulasi yang relevan:
   - Berbagai instrumen regulasi mempengaruhi pilihan desain, membatasi fungsi, atau melarang penerapan desain atau use case tertentu.
   - Mengetahui batasan-batasan yang dapat memengaruhi desain dan operasi sistem AI serta implikasi dan opsi yang dapat dilakukan.
   - Mencegah investasi pada sistem AI agar tepat sasaran dan tidak salah arah.
2. Regulasi yang terkait dalam pengembangan dan penerapan sistem AI dikomunikasikan dan dipahami oleh semua pihak pada bank yang terlibat dalam pengembangan dan penerapan sistem AI.
3. Memastikan bank selalu *update* dan mengikuti perubahan regulasi yang relevan dengan sistem AI dalam upaya memastikan fitur AI atau versi baru dari sistem AI tidak melanggar regulasi.
4. Bank melakukan pemantauan secara berkala dan berkelanjutan untuk memastikan kepatuhan sistem

» Step 4: Identifying and Testing Risk Management Strategies

- Banks choose risk management strategies based on their responses to specific identified risks. For example, if a bank identifies hallucination as a risk, it will adjust its response to this risk, such as by modifying the AI implementation to reduce hallucinations and ensuring continuous monitoring of outputs to detect hallucinations.
- After the bank identifies risk management strategies, they can test their effectiveness while the system is operational and balance residual risks with benefits to determine appropriate next steps. The bank can then document and operationalize the final decisions, such as advancing the AI project to the next stage in its lifecycle.

c. Compliance

Aspects that banks need to pay attention to throughout the AI lifecycle include:

1. Banks understand regulations related to AI systems and other regulations concerning IT adoption by banks, as well as ensuring compliance with relevant regulations:
   - Various regulatory instruments influence design choices, limit functions, or prohibit the implementation of certain designs or use cases.
   - Understanding the limitations that may affect the design and operation of AI systems, as well as the implications and options available.
   - Preventing investments in AI systems from being misdirected and ensuring they are targeted effectively.
2. Regulations related to the development and implementation of AI systems are communicated and understood by all parties within the bank involved in the development and implementation of AI systems.
3. Ensuring that the bank stays updated and follows changes in regulations relevant to AI systems in an effort to ensure that AI features or new versions of AI systems do not violate regulations.
4. The bank conducts regular and ongoing monitoring to ensure that the AI system's compliance remains

AI tetap sesuai dengan regulasi yang berlaku dan sesuai dengan toleransi risiko bank.

5. Bank memiliki prosedur internal dan unit kerja yang melakukan penilaian dan pemantauan secara berkala terhadap kepatuhan pemenuhan regulasi dari pengembangan dan penerapan sistem AI, serta mekanisme koordinasi dan tindak lanjut yang diperlukan dalam hal terdapat isu atau pelanggaran regulasi, atau dari penerapan sistem AI.

C.9. Audit Sistem AI

Direksi dan dewan komisaris sebagai fungsi pengawasan pada bank perlu didukung dengan informasi yang independen dan objektif tentang praktik penerapan dan manajemen risiko bank dalam implementasi sistem AI, yang diperoleh dari hasil pelaksanaan audit yang dilakukan secara berkala terhadap seluruh aktivitas dalam siklus hidup AI.

a. Audit Intern Sistem AI

Audit intern bank dapat dilengkapi dengan tim audit spesialisasi AI. Audit intern yang independen pada bank perlu menerapkan *best practice* (Schuett, 2024) bagi pelaksanaan fungsi audit intern AI:

in accordance with applicable regulations and aligns with the bank's risk tolerance.

5. The bank has internal procedures and work units that conduct regular assessments and monitoring of compliance with regulations regarding the development and implementation of AI systems, as well as necessary coordination mechanisms and follow-up actions in case of issues or regulatory violations related to the implementation of AI systems.

C.9. AI System Audit

The board of directors and the board of commissioners, as oversight functions within the bank, need to be supported with independent and objective information about the bank's practices in implementing and managing risks related to AI system implementation. This information should be obtained from periodic audits conducted on all activities throughout the AI lifecycle.

a. AI System Internal Audit

The bank's internal audit can be complemented by a specialized AI audit team. An independent internal audit within the bank needs to implement best practices (Schuett, 2024) for the execution of the AI internal audit function:

1. Tim audit intern AI harus menggunakan metodologi audit yang khusus untuk AI, yang didasarkan pada praktik terbaik dalam keamanan dan tata kelola AI. Tim audit intern AI harus memiliki waktu yang cukup untuk melakukan evaluasi mendalam serta meninjau praktik keamanan dan tata kelola yang diterapkan oleh pengembang (misalnya, model ancaman (*threat models*), evaluasi model, dan teknik penyelarasan).
2. Tim audit intern AI harus mendapatkan dukungan dari manajemen senior, terutama eksekutif yang bertanggung jawab atas penelitian AI, pengembangan produk, dan manajemen risiko.
3. Tim yang bekerja pada keamanan dan tata kelola AI harus merespons dengan cepat setiap permintaan dari tim intern AI internal. Pengembang AI juga harus mendorong kerja sama dan membangun kepercayaan diantara berbagai tim.
4. Tim audit intern AI harus mendapatkan dukungan untuk beradaptasi dengan perubahan dalam bank (misalnya, dengan merekrut lebih banyak auditor dan memberikan waktu yang cukup bagi mereka untuk melakukan tugasnya).
5. Lingkup dan tujuan audit AI harus didefinisikan dengan jelas.

1. The AI internal audit team must use an audit methodology specifically designed for AI, based on best practices in AI security and governance. The AI internal audit team should have sufficient time to conduct in-depth evaluations and review the security and governance practices implemented by developers (e.g., threat models, model evaluations, and alignment techniques).
2. The AI internal audit team must receive support from senior management, particularly executives responsible for AI research, product development, and risk management.
3. Teams working on AI security and governance must respond quickly to any requests from the internal AI audit team. AI developers should also promote collaboration and build trust among various teams.
4. The AI internal audit team must receive support to adapt to changes within the bank (e.g., by recruiting more auditors and providing them with sufficient time to perform their tasks).
5. The scope and objectives of the AI audit must be clearly defined.

b. Audit Eksternal Sistem AI

Audit eksternal sistem AI dilakukan oleh auditor ekstern yakni akuntan publik (AP) dan kantor akuntan publik (KAP) yang independen, terdaftar pada OJK, tercatat dalam daftar AP dan KAP yang aktif pada OJK serta memiliki kompetensi dalam pelaksanaan audit terkait sistem AI.

Dalam hal terdapat keterbatasan kemampuan satuan kerja audit intern bank, pelaksanaan fungsi audit intern sistem AI dapat dilakukan oleh auditor ekstern. Penggunaan auditor ekstern untuk melaksanakan

b. AI System External Audit

External audits of AI systems are conducted by external auditors, namely public accountants and independent public accounting who are registered with the Financial Services Authority (OJK), listed in the active Public Accountants and Public Accounting Firms registry at OJK, and possess the necessary competencies to perform audits related to AI systems.

In cases where there are limitations in the capabilities of the bank's internal audit unit, the internal audit function for AI systems can be performed by external auditors. The use of external auditors to carry out the internal

fungsi audit intern atas sistem AI tidak mengurangi tanggung jawab pimpinan satuan kerja audit intern. Selain itu, penggunaan auditor ekstern harus mempertimbangkan ukuran dan kompleksitas usaha bank serta memperhatikan ketentuan peraturan perundang-undangan terkait auditor ekstern, dan pelaksanaannya dilakukan sesuai standar dan prosedur audit TI Bank. Pelaksanaan fungsi audit intern AI oleh auditor ekstern tetap memperhatikan aspek kompetensi (antara lain pengetahuan dan pengalaman yang memadai) dan independensi serta didasari dengan suatu perjanjian kerja sama. Disamping itu, bank secara berkala melakukan kaji ulang terhadap fungsi audit intern sistem AI oleh pihak ekstern yang independen agar pelaksanaan fungsi audit Sistem AI dapat berjalan efektif.

Beberapa hal yang perlu diperhatikan oleh bank dalam melakukan audit dan pengawasan AI secara lebih detil dapat mengacu pada Bab 7 buku ini.

C.10. Faktor Pendukung (*Enabler*)

a. Organisasi

1. Kesiapan organisasi

   Kesiapan organisasi mengacu pada ketersediaan kelengkapan sumber daya bank yang diperlukan untuk adopsi AI, antara lain sumber daya keuangan serta keterampilan sumber daya manusia (SDM) bank. Adopsi teknologi baru menimbulkan persyaratan keterampilan baru yang membutuhkan karyawan dengan keterampilan teknis dalam mengadopsi sistem AI.

   Bank juga memerlukan dan dapat menggunakan *domain expert* (ahli di bidang tertentu yang membantu memastikan implementasi AI dan teknologi sesuai dengan kebutuhan spesifik industri) yang memahami tugas, alur kerja, dan logika proses bisnis yang ada serta memiliki kemampuan untuk mempertimbangkan bagaimana sistem AI dapat meningkatkan bisnis bank.

   Mengevaluasi ketersediaan keahlian internal diperlukan untuk memastikan bahwa SDM tidak hanya mengetahui cara memanfaatkan alat dan teknologi AI, tetapi juga aspek bisnis apa yang harus menjadi sasaran mereka.

   Kesiapan organisasi juga mencakup ukuran keberhasilan proyek AI (*Key Performance Indicators*) yang tidak hanya terkait performa model namun juga keberhasilan proyek secara keseluruhan dan dampak terhadap kinerja bisnis.

2. Pengorganisasian, peran dan tanggung jawab.

   Bank menetapkan pengorganisasian yang jelas yang menguraikan peran, tanggung jawab, dan wewenang

audit function on AI systems does not diminish the responsibility of the leadership of the internal audit unit. Furthermore, the use of external auditors must consider the size and complexity of the bank's operations and adhere to relevant legal regulations regarding external auditors, with implementation conducted according to banking IT audit standards and procedures. The execution of the internal audit function for AI by external auditors must still consider aspects of competence (including adequate knowledge and experience) and independence, and should be based on a cooperation agreement. In addition, the bank periodically conducts reviews of the internal audit function for AI systems by independent external parties to ensure that the implementation of the AI system audit function operates effectively.

Several aspects that banks need to consider when conducting audits and oversight of AI in more detail can refer to Chapter 7 of this book.

C.10. Supporting Factors (Enabler)

a. Organization

1. Organization readiness

   Organizational readiness refers to the availability of the necessary resources within the bank for AI adoption, including financial resources and human resource (HR) skills. The adoption of new technologies creates new skill requirements, necessitating employees with technical skills to implement AI systems.

   The bank also requires and can utilize domain experts (specialists in specific fields who help ensure that AI implementation and technology meet the specific needs of the industry) who understand existing tasks, workflows, and business process logic, as well as have the ability to consider how AI systems can enhance the bank's business.

   Evaluating the availability of internal expertise is necessary to ensure that human resources not only know how to utilize AI tools and technologies but also understand which business aspects should be their targets.

   Organizational readiness also includes the measures of success for AI projects (Key Performance Indicators) that are not only related to model performance but also to the overall success of the project and its impact on business performance.

2. Organization, Roles, and Responsibilities.

   The bank establishes a clear organization that outlines roles,

dalam penerapan sistem AI untuk seluruh siklus hidup AI, guna memastikan operasi yang efektif dalam implementasi strategi digital (sistem AI), memfasilitasi kolaborasi antar unit dan fungsi yang multidisiplin, koreksi yang tepat waktu, pengawasan, dan inovasi yang berkelanjutan.

Seluruh anggota direksi setidaknya memiliki pemahaman dasar tentang konsep AI, bagaimana algoritma pembelajaran mesin bekerja, dan metode/teknik analisis data. Paling sedikit satu anggota direksi memiliki pemahaman yang memadai terkait AI dan penerapannya di sektor perbankan.

Organisasi yang mendukung penerapan AI yang bertanggung jawab, perlu didukung dengan komite yang bertanggung jawab kepada Direksi untuk bertugas memastikan inisiatif AI selaras dengan tujuan bank serta mengelola potensi risiko dan aspek terkait lainnya. Komite AI pada Bank bersifat lintas fungsi dan terdiri dari anggota dengan keterampilan, latar belakang, dan pengalaman yang beragam. Komite AI dapat dibentuk secara tersendiri yang diketuai oleh direktur yang membawahkan satuan kerja penyelenggara TI atau direktur yang membawahkan satuan kerja manajemen risiko. Adapun ruang lingkup dari fungsi Komite AI adalah memberikan rekomendasi kepada Direksi antara lain terkait hal-hal berikut:

responsibilities, and authorities in the implementation of AI systems for the entire AI lifecycle, to ensure effective operations in implementing digital strategies (AI systems), facilitate collaboration among multidisciplinary units and functions, timely corrections, oversight, and continuous innovation.

All board members should have at least a basic understanding of AI concepts, how machine learning algorithms work, and data analysis methods/techniques. At least one board member should have adequate knowledge related to AI and its application in the banking sector.

Organizations that support the responsible implementation of AI need to be backed by a committee accountable to the Board of Directors, tasked with ensuring that AI initiatives align with the bank's objectives and manage potential risks and related aspects. The AI committee at the bank is cross-functional and consists of members with diverse skills, backgrounds, and experiences. The AI committee can be formed independently, chaired by a director overseeing the IT unit or a director overseeing the risk management unit. The scope of the AI Committee's functions includes providing recommendations to the Board of Directors regarding, among other things, the following matters:

a. Keselarasan proyek AI yang akan dilakukan bank dengan rencana strategis bank, standar etika, dan regulasi.
b. Desain dan kerangka tata kelola AI bank beserta panduan yang menjadi pedoman internal bank.
c. Peran dan tanggung utama terkait pengawasan, desain, pengembangan, dan penggunaan AI pada bank.
d. Tata kelola data dalam pengelolaan data yang terpadu antar pemilik data dan pemrosesan data pada berbagai unit di bank.
e. Potensi risiko yang terkait dengan proyek AI, termasuk masalah privasi data, bias algoritma, dan pelindungan nasabah.
f. Cakupan program tata kelola AI (termasuk jenis model, algoritma, dan sistem mana yang termasuk dan tidak termasuk dalam cakupan beserta alasannya, serta skala risiko terkait penggunaan).
g. Kebijakan, proses, dan pelatihan untuk memungkinkan desain, penggunaan, dan pengawasan AI secara bertanggung jawab dan beretika.

a. Alignment of AI projects to be undertaken by the bank with the bank's strategic plans, ethical standards, and regulations.
b. The design and framework of the bank's AI governance along with guidelines that serve as internal directives for the bank.
c. The main roles and responsibilities related to the oversight, design, development, and use of AI in the bank.
d. Data governance in the management of integrated data among data owners and data processing across various units in the bank.
e. Potential risks associated with AI projects, including data privacy issues, algorithmic bias, and customer protection.
f. Scope of the AI governance program (including which types of models, algorithms, and systems are included and excluded from the scope along with the reasons, as well as the risk scale associated with their use).
g. Policies, processes, and training to enable the responsible and ethical design, use, and oversight of AI.

h. Area yang memerlukan revieu atau pengawasan manusia, termasuk untuk mengidentifikasi ketidakakuratan, mengidentifikasi bias, dan melakukan *quality assurance* (jaminan kualitas).

i. Menilai dan mengeskalasi kasus penggunaan AI berisiko tinggi

j. Pengelolaan insiden yang berkaitan dengan penggunaan AI.

k. Memfasilitasi diskusi, membangun kesepakatan di antara anggota komite, dan memandu proses pengambilan keputusan.

h. Areas that require human review or oversight, including identifying inaccuracies, detecting bias, and conducting quality assurance).

i. Assessing and escalating high-risk AI use cases.

j. Incident management related to the use of AI.

k. Facilitating discussions, building consensus among committee members, and guiding the decision-making process.

l. Berkomunikasi dengan berbagai pemangku kepentingan mengenai inisiatif AI, termasuk direksi, dewan komisaris, manajemen senior, tim teknis, tim hukum, dan nasabah sebagai pengguna akhir.

m. Mendorong pemahaman tentang teknologi AI dan pertimbangan etika dalam bank melalui program pelatihan.

n. Memantau kinerja sistem AI dan memastikan kepatuhan terhadap pedoman yang ditetapkan.

Pelaksanaan fungsi Komite AI dapat menjadi bagian dari aktivitas Komite Pengarah TI sebagaimana ketentuan OJK yang mengatur mengenai Penyelenggaraan Teknologi Informasi oleh Bank Umum (POJK No. 11/POJK.03/2022). Komite AI sedikitnya beranggotakan:

- Direktur yang membawahkan satuan kerja penyelenggara TI
- Direktur yang membawahkan satuan kerja manajemen risiko
- Pejabat eksekutif yang memimpin satuan kerja penyelenggara TI
- Pejabat eksekutif yang memimpin satuan kerja pengguna TI.
- Pejabat eksekutif yang memimpin satuan kerja manajemen risiko.

l. Communicating with various stakeholders regarding AI initiatives, including the board of directors, board of commissioners, senior management, technical teams, legal teams, and customers as end users.

m. Promoting understanding of AI technology and ethical considerations within the bank through training program.

n. Monitoring the performance of AI systems and ensuring compliance with established guidelines.

The implementation of the AI Committee's functions can be part of the activities of the IT Steering Committee as stipulated by OJK regulations governing the Implementation of Information Technology by Commercial Banks (POJK No. 11/POJK.03/2022). The AI Committee must consist of at least:

- A director overseeing the IT unit
- A director overseeing the risk management unit
- An executive officer leading the IT unit
- An executive officer leading the IT users unit.
- An executive officer leading the risk management unit.

- Pejabat eksekutif yang memimpin satuan kerja kepatuhan.
- Pejabat eksekutif yang memimpin satuan kerja hukum.
- SDM bank yang memiliki keahlian yang sangat memadai terkait sistem AI, jika diperlukan.
- Pihak eksternal yang *expert* di bidang AI, jika diperlukan dan dengan tetap memperhatikan aspek kehati-hatian.

3. Membangun *Center of Excellence* (CoE)

   AI CoE adalah penggerak utama dan titik kontak utama untuk semua topik AI dalam bank . AI CoE mengkonsolidasikan keahlian, mengimplementasikan proyek strategis, dan mendukung adopsi AI dalam bank, serta melakukan sinergi dan koordinasi lintas departemen. Tanggung jawab AI CoE sangat luas. CoE yang sukses diklasifikasikan dalam tiga tugas utam (AAI, 2023):

   a. Menentukan strategi dan prioritas *use case* AI
   - Mendefinisikan strategi AI
   - Mendefinisikan kerangka kerja AI
   - Memprioritaskan *use case* AI

   b. Mengembangkan solusi AI
   - Desain produk AI
   - Implementasi solusi AI

- An executive officer leading the compliance unit.
- An executive officer leading the legal unit.
- Bank personnel with highly adequate expertise related to AI systems, if necessary.
- External parties who are experts in the field of AI, if necessary, while still being mindful of the prudential aspect.

3. Developing a Center of Excellence (CoE)

   AI CoE is the main driver and primary contact point for all AI topics within the bank. The AI CoE consolidates expertise, implements strategic projects, and supports AI adoption in the bank, as well as performing synergy and cross-departmental coordination. The responsibilities of the AI CoE are very broad. A successful CoE is classified into three main tasks (AAI, 2023):

   a. Determining the strategy and priorities for AI use cases
   - Defining AI strategy
   - Defining AI framework
   - Prioritizing AI use case

   b. Developing AI solutions
   - Designing AI products
   - Implementing AI solutions

c. Mendukung bank dalam Adopsi AI

- Mencari dan mengembangkan talenta AI
- Membangun budaya AI
- Meningkatkan kesadaran dan komunikasi
- Mendorong kolaborasi dan alih pengetahuan
- Mengelola infrastruktur AI
- Bertanggung jawab atas data
- Mengembangkan ekosistem AI

Dengan menerapkan struktur ini, bank dapat memastikan bahwa AI tidak hanya menjadi eksperimen teknologi, tetapi juga berkontribusi secara nyata terhadap pertumbuhan dan inovasi bisnis bank.

4. Nilai

Adanya komitmen bank untuk menetapkan dan menerapkan nilai dan prinsip untuk membangun kepercayaan publik dalam penggunaan sistem AI yang bertanggung jawab.

5. Budaya organisasi

Implementasi AI bukan hanya tentang adopsi teknologi. Meskipun teknologi dan talenta organisasi hal yang sangat penting, menyelaraskan budaya organisasi pada bank untuk mendukung adopsi AI pada bank juga sama pentingnya.

c. Supporting banks in AI Adoption

- Seeking and builing AI talents
- Building AI culture
- Increasing awareness and communication
- Promoting collaboration and knowledge transfer
- Managing AI infrastructure
- Responsibility over data
- Developing AI ecosystem

By implementing this structure, the bank can ensure that AI is not just a technological experiment but also contributes significantly to business growth and innovation.

4. Value

The bank is committed to establishing and implementing values and principles to build public trust in the responsible use of AI systems. The bank is committed to establishing and implementing values and principles to build public trust in the responsible use of AI systems.

5. Organization culture

The implementation of AI is not just about technology adoption. While technology and organizational talent are very important, aligning the organizational culture within the bank to support AI adoption is equally important.

AI dapat dilihat sebagai teknologi inovatif, yang akan mengubah model bisnis dan sistem pada bank, dan karenanya bank harus mampu menanggapi perubahan ini termasuk memiliki karyawan yang bersedia menggunakan teknologi baru dalam jangka panjang.

Budaya inovatif mendorong SDM memiliki kemauan untuk mengeksploitasi ide-ide baru dan terus menerus belajar dan berinovasi untuk mendukung penerapan dan penggunaan aplikasi AI serta mampu mengidentifikasi dan memanfaatkan peluang baru untuk aplikasi AI.

Bank dengan budaya inovatif memiliki modal yang lebih baik untuk mengintegrasikan AI dalam setiap lini kerja yang dimiliki. Budaya organisasi pada bank yang dipimpin oleh direksi untuk mendorong inovasi dan kolaborasi lintas fungsi (antar unit) pada bank akan memfasilitasi implementasi AI yang lebih efektif.

6. *Agility*

Agility (kelincahan) merupakan kemampuan bank untuk merespon dan beradaptasi terhadap perubahan secara cepat dan efektif. *Agility* penting dalam implementasi AI pada bank, yang membantu transisi dari proses lama ke proses baru untuk memungkinkan implementasi AI. Dalam organisasi

AI can be seen as an innovative technology that will transform business models and systems within the bank, and therefore the bank must be able to respond to these changes, including having employees who are willing to use new technologies in the long term.

An innovative culture encourages human resources to have the willingness to exploit new ideas and continuously learn and innovate to support the implementation and use of AI applications, as well as being able to identify and leverage new opportunities for AI applications.

Banks with an innovative culture have a better foundation for integrating AI into every line of work they possess. An organizational culture within the bank, led by the board of directors to encourage innovation and cross-functional collaboration (between units), will facilitate a more effective implementation of AI.

6. Agility

Agility is the bank's ability to respond and adapt to changes quickly and effectively. Agility is important in the implementation of AI within the bank, as it helps transition from old processes to new ones to enable AI implementation. In modern organizations, agility becomes a crucial aspect of organizational

modern, *agility* menjadi aspek penting dalam transformasi organisasi yang didukung oleh teknologi dan kapabilitas digital dalam upaya untuk meningkatkan proses dan mendorong model bisnis baru.

7. Keterlibatan pemangku kepentingan

   Berkoordinasi dan bekerjasama untuk mendapatkan perspektif yang beragam dari berbagai komunitas dan pemangku kepentingan pada sepanjang siklus hidup AI untuk mendapatkan berbagai referensi dan mengurangi risiko, dengan tetap menjaga aspek kehati-hatian.

b. Sumber Daya Manusia

Aspek yang perlu menjadi perhatian bank antara lain:

1. Kompetensi AI

   Bank memiliki SDM dengan pengetahuan, keahlian dan pengalaman multidisiplin yang memadai terkait inovasi, desain, pengembangan, penerapan, penilaian, dan pemantauan sistem AI.

   Tanpa organisasi yang tepat dan diisi dengan sumber daya manusia yang tepat, implementasi AI tidak mungkin dilakukan dengan baik. Sebelum meluncurkan proyek AI, bank harus menyiapkan organisasi AI dan sumber daya manusia yang merupakan tim spesialis AI yang memiliki pengetahuan domain bisnis dan latar belakang teknologi AI yang memadai.

   Bank juga memberikan peluang untuk peningkatan kompetensi SDM di bidang teknis, nonteknis, dan kepemimpinan, untuk menghasilkan penerapan AI yang bertanggung jawab.

   Bank membangun kompetensi AI khususnya terhadap unit-unit yang terkait dalam implementasi dan pengguna sistem AI, termasuk membangun pengetahuan tentang bagaimana AI berkembang dan bagaimana AI dapat menguntungkan bisnis. Karenanya, mengetahui di mana dan bagaimana AI dapat diimplementasikan merupakan faktor kunci keberhasilan. Bank harus mengembangkan dan mempertahankan spesialis AI secara internal dengan visi dan pandangan perspektif jangka panjang. Selain itu, bank dapat mengoperasikan tim manajemen perubahan proyek (*change management office*) untuk meningkatkan dan memperluas proyek AI yang bekerja sama dengan pengguna akhir bisnis.

   Kompetensi khusus juga diperlukan untuk mendukung penerapa AI secara optimal, seperti kompetensi di bidang data (analisis data, arsitektur data, integrasi data).

transformation supported by technology and digital capabilities in efforts to enhance processes and drive new business models.

7. Stakeholder engagement

   Coordinating and collaborating to gain diverse perspectives from various communities and stakeholders throughout the AI lifecycle to obtain various references and mitigate risks, while maintaining aspects of prudentiality.

b. Human Resources

Aspects that banks need to take into consideration includes:

1. AI competence

   The bank has human resources with adequate knowledge, skills, and multidisciplinary experience related to the innovation, design, development, implementation, evaluation, and monitoring of AI systems.

   Without the right organization and filled with the right human resources, AI implementation cannot be carried out effectively. Before launching an AI project, the bank must prepare an AI organization and human resources that consist of a team of AI specialists with adequate business domain knowledge and a background in AI technology.

   The bank also provides opportunities for enhancing the competencies of human resources in technical, non-technical, and leadership areas to produce responsible AI implementations.

   The bank builds AI competencies especially for units involved in the implementation and use of AI systems, including building knowledge about how AI evolves and how it can benefit the business. Therefore, knowing where and how AI can be implemented is a key success factor. The bank must develop and maintain AI specialists internally with a long-term vision and perspective. Additionally, the bank can operate a project change management team (change management office) to enhance and expand AI projects in collaboration with business end users.

   Specialized competencies are also required to optimally support AI implementation, such as competencies in the field of data (data analysis, data architecture, data integration).

2. Penerimaan (*receptivity*)

   Receptivity atau kesediaan menerima ide-ide atau gagasan diperlukan dalam hal bank akan membangun kapabilitas AI. Bank harus memprioritaskan strategi yang ditujukan untuk mempengaruhi budaya dan struktur organisasi guna meningkatkan penerimaan terhadap solusi AI. Bank dengan sumber daya yang memadai namun gagal mengatasi faktor-faktor resistensi tidak mungkin memperoleh nilai dari investasi dalam AI.

3. Kepercayaan Karyawan terhadap AI (*Employee-AI Trust*)

   Sistem AI dapat mereplikasi kognisi manusia (dhi. mencoba meniru cara manusia berpikir dan mengambil keputusan) dan otomasi tugas manual, yang berdampak pada perubahan peran karyawan pada bank. Implementasi AI dapat menyebabkan peran yang ada perlu disesuaikan atau bahkan menciptakan peran baru. Oleh karena itu, karyawan harus memahami tujuan AI, cara kerjanya, serta bagaimana AI akan memengaruhi tugas dan tanggung jawab mereka. Kepercayaan karyawan terhadap keputusan penggunaan AI dan output yang dihasilkan ini menjadi faktor kunci dalam kolaborasi antara manusia dan mesin. Sumber daya yang terlibat dalam AI juga perlu memastikan bahwa AI beroperasi sesuai dengan desain yang diinginkan. Kesediaan karyawan untuk mempercayai AI sangat bergantung pada pemahaman mereka terhadap teknologi tersebut.

4. Kolaborasi multidisiplin

   Pada dasarnya AI adalah teknologi lintas fungsi yang memengaruhi berbagai aspek pada bank. Dengan demikian, kolaborasi multidisiplin sangat penting untuk mendapatkan perspektif yang komprehensif dalam implementasi AI. Bank harus mampu menyatukan keahlian dari berbagai unit kerja termasuk menyelaraskan AI dengan peluang dan tantangan di berbagai tingkat organisasi.

5. Pekerjaan yang dibantu AI (*AI augmented work*)

   SDM bank perlu beradaptasi dengan peran yang didukung oleh teknologi AI. Keterampilan yang mereka miliki saat ini akan ditingkatkan dengan kehadiran AI. Oleh karena itu,

2. Receptivity

   Receptivity or the willingness to accept ideas is necessary when the bank aims to build AI capabilities. The bank must prioritize strategies aimed at influencing culture and organizational structure to enhance acceptance of AI solutions. A bank with adequate resources but failing to address resistance factors is unlikely to derive value from its investment in AI.

3. Employee-AI Trust

   AI systems can replicate human cognition (i.e., attempting to mimic how humans think and make decisions) and automate manual tasks, which impacts the changing roles of employees within the bank. The implementation of AI may require existing roles to be adjusted or even create new roles. Therefore, employees must understand the objectives of AI, how it works, and how AI will affect their tasks and responsibilities. Employee trust in the decisions made by AI and the outputs it generates is a key factor in the collaboration between humans and machines. The resources involved in AI also need to ensure that AI operates according to the desired design. Employees' willingness to trust AI greatly depends on their understanding of the technology.

4. Multidisciplinary collaboration

   Essentially, AI is a cross-functional technology that influences various aspects of the bank. Therefore, multidisciplinary collaboration is crucial to obtain a comprehensive perspective in AI implementation. The bank must be able to unite expertise from various work units, including aligning AI with opportunities and challenges at different levels of the organization.

5. AI augmented work

   Bank human resources need to adapt to roles supported by AI technology. The skills they currently possess will be enhanced by the presence of AI. Therefore,

penting untuk memahami dampak penerapan AI pada bank dan bagaimana menciptakan kolaborasi yang efektif antara kecerdasan AI dan kecerdasan manusia.

c. Infrastruktur AI

Infrastruktur AI merupakan tulang punggung berbagai aplikasi AI dan *machine learning* dengan menyediakan daya komputasi serta sumber daya untuk memproses *dataset* dalam jumlah besar, terdiri dari perangkat keras (*hardware*) dan perangkat lunak (*software*) yang dirancang untuk mendukung beban kerja sistem AI dan pembelajaran mesin (*machine learning*).

Berbeda dengan infrastruktur TI tradisional, infrastruktur AI dioptimalkan untuk menangani kebutuhan komputasi tinggi dan pemrosesan data dalam jumlah besar yang diperlukan oleh algoritma AI. Infrastruktur AI memungkinkan bank memanfaatkan AI secara optimal dengan akses ke daya komputasi tinggi, pemrosesan data yang cepat, dan solusi penyimpanan yang dapat diubah sesuai kebutuhan.

Infrastruktur AI diprediksi akan terus berkembang seiring dengan kemajuan teknologi, memungkinkan bank untuk mengoptimalkan penggunaan AI dalam berbagai bidang secara lebih efektif. Karenanya, bank perlu memiliki strategi pemenuhan dan investasi infrastruktur AI yang efektif dan efisien, didukung dengan anggaran dan pemeliharaan yang memadai.

Aspek-aspek yang perlu menjadi perhatian bank terkait infrastruktur AI pada bank:

1. Infrastruktur AI bank harus dirancang dengan langkah-langkah keamanan yang kuat, termasuk enkripsi, kontrol akses, dan pemenuhan regulasi seperti pelindungan data pribadi untuk menjaga privasi.
2. Infrastruktur AI bank harus dapat berkembang seiring bertambahnya ukuran model dan *dataset* untuk memenuhi peningkatan kebutuhan (skalabilitas), serta fleksibilitas untuk

it is important to understand the impact of AI implementation on the bank and how to create effective collaboration between AI intelligence and human intelligence.

c. AI infrastructure

AI infrastructure is the backbone of various AI and machine learning applications, providing computational power and resources to process large datasets. It consists of hardware and software designed to support the workloads of AI systems and machine learning.

Unlike traditional IT infrastructure, AI infrastructure is optimized to handle high computational needs and large-scale data processing required by AI algorithms. AI infrastructure enables banks to optimally leverage AI with access to high computing power, fast data processing, and scalable storage solutions.

AI infrastructure is predicted to continue evolving alongside technological advancements, allowing banks to optimize the use of AI in various fields more effectively. Therefore, banks need to have effective and efficient strategies for fulfilling and investing in AI infrastructure, supported by adequate budgeting and maintenance.

Aspects that banks need to take into considerations regarding AI infrastructure include:

1. The bank's AI infrastructure must be designed with strong security measures, including encryption, access controls, and compliance with regulations such as personal data protection to maintain privacy.
2. The bank's AI infrastructure must be able to evolve as the size of models and datasets increases to meet growing demands (scalability), as well as provide flexibility to support

mendukung dan berdaptasi dengan berbagai model dan algoritma AI yang terus berkembang (misalnya, penggunaan infrastruktur berbasis *cloud* dalam pemenuhan skala yang dibutuhkan).

3. Bank memastikan alur kerja AI beroperasi secara optimal dalam siklus hidup AI sehingga infrastruktur AI dapat berperan secara optimal serta dapat meningkatkan produktivitas secara keseluruhan.
4. Integrasi antara infrastruktur AI dengan infrastruktur/sistem TI yang sudah ada pada bank sangat diperlukan untuk memanfaatkan data yang tersedia, memastikan transisi menuju kemampuan sistem AI bank yang lebih terdepan.
5. Bank memberikan perhatian penting kepada sistem penyimpanan, mengingat sistem penyimpanan memiliki peran krusial dalam infrastruktur AI yang memengaruhi kinerja dan biaya.
6. Pilihan infrastruktur AI yang dilakukan bank harus mampu beradaptasi dengan perkembangan teknologi yang cepat, mudah diperbarui sesuai tren terkini, serta mampu membangun budaya inovasi dan pembelajaran berkelanjutan pada bank.

and adapt to various evolving AI models and algorithms (e.g., the use of cloud-based infrastructure to meet required scale).

3. The bank ensures that the AI workflow operates optimally throughout the AI lifecycle so that the AI infrastructure can play its role effectively and enhance overall productivity.
4. Integration between AI infrastructure and the existing IT infrastructure/systems within the bank is essential to leverage available data, ensuring a transition towards more advanced AI capabilities for the bank.
5. The bank places significant attention on storage systems, considering that storage systems play a crucial role in AI infrastructure, affecting performance and costs.
6. The choice of AI infrastructure made by the bank must be able to adapt to rapid technological developments, be easily updated according to current trends, and foster a culture of innovation and continuous learning within the bank.
7. Investasi infrastruktur yang sesuai memungkinkan pengelolaan data yang lebih baik yang penting untuk akurasi model AI, mendukung fleksibilitas, meningkatkan daya saing, efisiensi operasional, inovasi, serta mampu mendukung strategi dan model bisnis dan menciptakan peluang pasar baru.

d. Dampak Implementasi

Mengintegrasikan AI ke dalam operasional bank bukan sekadar perubahan teknologi, tetapi transformasi menyeluruh pada struktur, budaya, dan *mindset* organisasi. Dengan menerapkan kerangka kerja strategis, mendorong budaya inovasi, dan menangani isu etika, bank dapat memanfaatkan potensi AI, mengutamakan strategi adaptif, serta mengoptimalkan sumber daya dalam menghadapi berbagai tantangan perubahan.

Karenanya, implementasi AI pada bank memerlukan dukungan dari seluruh pihak yang terlibat (*AI actors*) baik internal dan eksternal Bank dalam seluruh siklus hidup AI. Disampng itu, alokasi sumber daya yang tepat sangat diperlukan agar pengembangan dan penerapan AI sesuai dengan tujuan yang diharapkan bank dan sejalan dengan regulasi yang ada.

7. Investing in appropriate infrastructure enables better data management, which is crucial for the accuracy of AI models, supports flexibility, enhances competitiveness, operational efficiency, innovation, and can support strategies and business models while creating new market opportunities.

d. Implementation Impacts

Mengintegrasikan AI ke dalam operasional bank bukan sekadar perubahan teknologi, tetapi transformasi menyeluruh pada struktur, budaya, dan *mindset* organisasi. Dengan menerapkan kerangka kerja strategis, mendorong budaya inovasi, dan menangani isu etika, bank dapat memanfaatkan potensi AI, mengutamakan strategi adaptif, serta mengoptimalkan sumber daya dalam menghadapi berbagai tantangan perubahan.

Karenanya, implementasi AI pada bank memerlukan dukungan dari seluruh pihak yang terlibat (*AI actors*) baik internal dan eksternal Bank dalam seluruh siklus hidup AI. Disampng itu, alokasi sumber daya yang tepat sangat diperlukan agar pengembangan dan penerapan AI sesuai dengan tujuan yang diharapkan bank dan sejalan dengan regulasi yang ada.

Bab 7
Pengawasan dan Audit

Chapter 7
Supervision and Audit

Bab ini membahas bagaimana pendekatan yang dapat dilakukan bank ketika akan melakukan audit AI. Perlu dipahami konteks yang dimaksud adalah pemeriksaan atau audit terhadap AI yang digunakan pada operasional perbankan. Hal ini perlu diketahui mengingat audit AI memiliki konteks yang berbeda dengan proses audit menggunakan AI atau AI sebagai *tools* atau alat bantu bagi auditor ketika melakukan proses pemeriksaan.

This chapter discusses the approaches that banks can take when conducting AI audits. It is important to understand that the context here refers to the examination or audit of AI used in banking operations. This needs to be understood considering AI audits have a different context from audit processes using AI or AI as a tool to assist auditors during the examination process.

Pelaksanaan audit AI merupakan bagian dari pelaksanaan audit teknologi informasi (TI) sehingga bank perlu memperhatikan ketentuan dan panduan terkait yang telah dikeluarkan oleh OJK. Ketentuan OJK yang mengatur terkait pelaksanaan audit internal TI terdapat pada POJK No. 1/POJK.03/2019 tentang Penerapan Fungsi Audit Internal pada Bank Umum, POJK No. 11/POJK.03/2022 tentang Penyelenggaraan Teknologi Informasi oleh Bank Umum, POJK No. 17 Tahun 2023 tentang Penerapan Tata Kelola bagi Bank umum, dan SEOJK No. 21/SEOJK.03/2017 tentang Penerapan Manajemen Risiko Dalam Penggunaan Teknologi Informasi oleh Bank Umum.

The implementation of AI audits is part of the implementation of information technology (IT) audits, so the bank needs to pay attention to the related provisions and guidelines issued by OJK. The OJK regulations governing the implementation of internal IT audits are contained in POJK No. 1/POJK.03/2019 concerning the Implementation of Internal Audit Functions in Commercial Banks, POJK No. 11/POJK.03/2022 concerning Information Technology Implementation by Commercial Banks, POJK No. 17 of 2023 concerning Governance Implementation for Commercial Banks, and SEOJK No. 21/SEOJK.03/2017 concerning Risk Management Implementation in the Use of Information Technology by Commercial Banks.

Audit AI adalah rangkaian kegiatan dalam satu proses pemeriksaan untuk mendapatkan keyakinan yang memadai bahwa AI yang di implementasikan oleh bank telah dilatih menggunakan *input* data yang representatif dan kredibel, berlangsung pada suatu proses pengolahan yang transparan, dengan *output* yang dapat dijelaskan (kondisi *black box* yang minimal) serta hasil yang secara umum dapat diprediksi atau sesuai ekspektasi. Melakukan audit AI merupakan tantangan bagi auditor bank maupun pihak eksternal yang ditunjuk oleh bank. Terlepas dari tantangan yang muncul, auditor tidak perlu terlalu khawatir karena mereka telah memiliki pengalaman menghadapi tantangan serupa, seperti saat bank pertama kali mengadopsi teknologi baru, misalnya cloud dan keamanan siber. Dahulu seorang auditor TI memiliki tanggung jawab termasuk melakukan audit algoritma. Namun saat ini guna memastikan bahwa algoritma AI telah sejalan dengan nilai-nilai sesuai panduan ini maka audit algoritma dapat dilakukan oleh seorang auditor spesialis model/algoritma.

AI audit is a series of activities within an examination process to obtain reasonable assurance that the AI implemented by the bank has been trained using representative and credible input data, conducted through a transparent processing procedure, with outputs that can be explained (minimizing black box conditions), and results that are generally predictable or as expected. Conducting AI audits presents challenges for both bank auditors and external parties appointed by the bank. Despite emerging challenges, auditors need not worry too much since they have experienced facing similar challenges, such as when banks first adopted new technologies like cloud computing and cybersecurity. Previously, an IT auditor had responsibilities including auditing algorithms. However, currently to ensure that AI algorithms align with values according to these guidelines, algorithm audits can be conducted by a specialist model/algorithm auditor.

Dalam melakukan audit AI, auditor bank diarahkan untuk fokus pada proses tata kelola, manajemen risiko AI dan integrasinya ke dalam sistem bank. Dalam melakukan pemeriksaan seorang pemeriksa memastikan apakah lingkungan pengendalian dan struktur tata kelola yang sudah ditetapkan dapat berjalan dengan efektif. Auditor internal

In conducting AI audits, bank auditors are directed to focus on governance processes, AI risk management, and its integration into the bank's systems. During the examination, an auditor ensures whether the established control environment and governance structure operate effectively. Internal bank auditors can focus on developing a

bank dapat fokus mengembangkan sikap positif dan memiliki kerangka berpikir yang sistematis. Auditor tidak diharapkan menjadi ahli AI namun memiliki disiplin, pendekatan yang fokus pada pola berpikir kritis terkait risiko. Memiliki pengetahuan tentang AI itu penting namun mengetahui segala aspek teknis seperti menguraikan algoritma AI tentu tidak dapat dilakukan oleh semua orang sehingga mungkin diperlukan asistensi dari auditor eksternal/pihak ketiga di luar bank.

Untuk memastikan pelaksanaan audit internal yang efektif dan menyeluruh sebagaimana dimaksud dalam Pasal 53 POJK No. 11/POJK.03/2022 tentang Penyelenggaraan Teknologi Informasi oleh Bank Umum, bank wajib memastikan ketersediaan jejak audit atas seluruh kegiatan penyelenggaraan AI untuk keperluan pengawasan, penegakan hukum, penyelesaian sengketa, verifikasi, pengujian, dan pemeriksaan lain.

positive attitude and having a systematic mindset. Auditors are not expected to be AI experts but should have discipline and an approach focused on critical thinking related to risks. Having knowledge about AI is important, but understanding all technical aspects, such as explaining AI algorithms, cannot be done by everyone; therefore, assistance from external auditors or third parties outside the bank may be required.

To ensure the effective and comprehensive implementation of internal audits as referred to in Article 53 of POJK No. 11/POJK.03/2022 concerning Information Technology Implementation by Commercial Banks, the bank is required to ensure the availability of audit trails for all AI implementation activities for purposes of supervision, law enforcement, dispute resolution, verification, testing, and other examinations.

A. Tantangan Melakukan Audit AI

Dalam melakukan pemeriksaan AI beberapa aspek terkait menimbulkan tantangan bagi auditor, antara lain:

1. AI (atau khususnya, audit algoritma) pada dasarnya adalah hal yang sangat kompleks dan memerlukan pendekatan multidimensi dan sering kali memerlukan keahlian khusus.
2. Risiko-risiko terkait AI saat ini terus muncul seiring dengan perkembangan teknologi/AI yang cepat.
3. Konteks nilai utama yang digunakan dan relevan dapat berbeda tergantung pada karakteristik industri.
4. AI sebagai topik audit saat ini merupakan sesuatu yang masih terus berkembang sehingga alat bantu dan pendekatan audit yang diadopsi secara luas juga masih terbatas.
5. Kesempatan pelatihan bagi auditor internal yang tersedia untuk meningkatkan kompetensi audit AI yang terbatas.

A. Challenge in performing AI Audit

In conducting AI examinations, several related aspects pose challenges for auditors, including:

1. AI (or specifically, algorithm audits) is inherently very complex and requires a multidimensional approach and often demands specialized expertise.
2. Risks related to AI continue to emerge alongside the rapid development of technology/AI.
3. The context of the main values used and their relevance can vary depending on the characteristics of the industry.
4. AI as an audit topic is currently still evolving, so widely adopted audit tools and approaches are also still limited.
5. Training opportunities available for internal auditors to enhance AI audit competencies are limited.

B. Siklus Hidup dan Audit AI

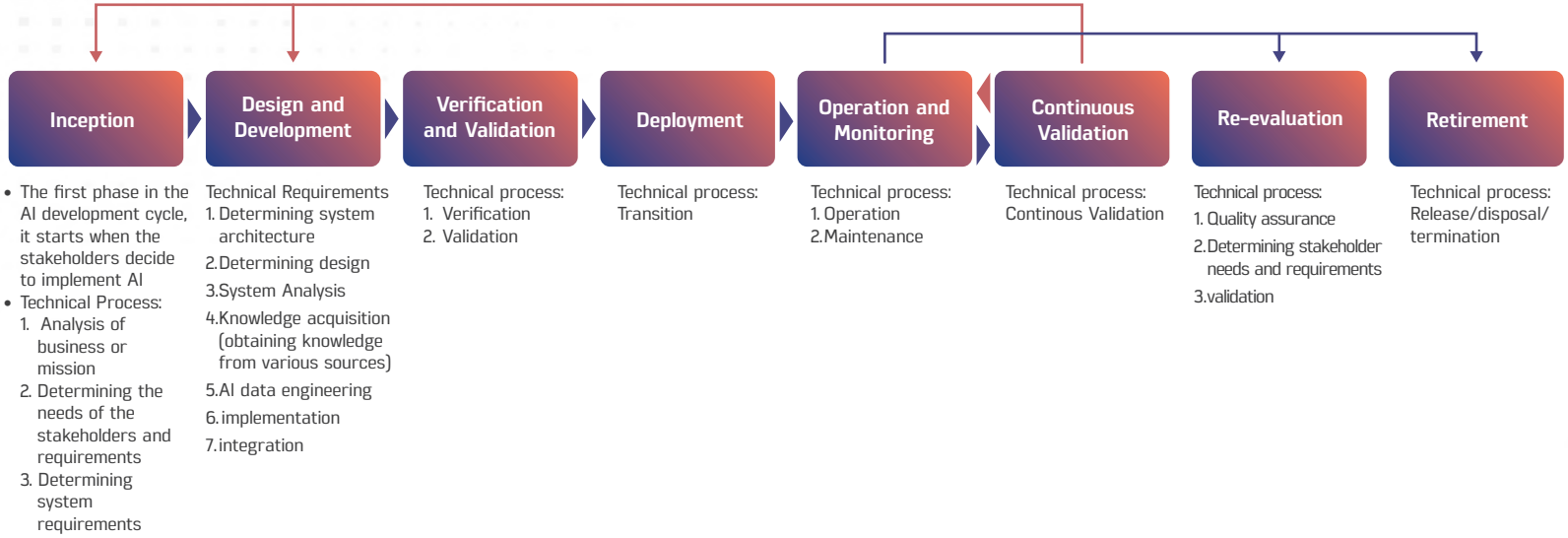
Dalam melakukan pemeriksaan AI seorang auditor dapat juga mengaitkan tahapan proses audit dengan siklus hidup sistem teknologi informasi yang juga mencakup AI.

B. AI Life Cycle and Audit

In conducting AI examinations, an auditor can also relate the audit process stages to the information technology system lifecycle, which also includes AI.

Gambar 15. Siklus Hidup AI

Figure 15. AI Life Cycle



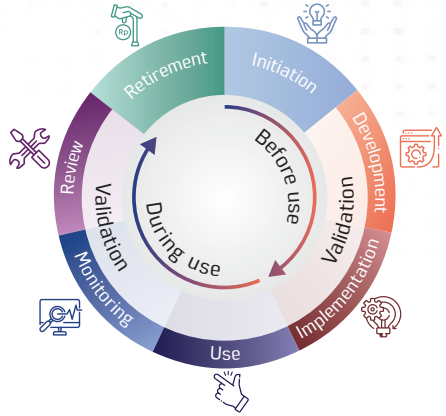
Source: ISO/IEC 5338

Beberapa peneliti dan riset mengusulkan model penilaian tersebut. Luliana Sandu, Menno Wiersma, Daphne Manichand (2022), melakukan pendekatan yang terinspirasi dari model manajemen risiko (*three lines system*) untuk kebutuhan audit algoritma, khususnya bagi audit internal, yang disebut sebagai *Life Cycle Framework for Auditing AI Algorithms*. Metode ini sebagai salah satu contoh dan tentu bisa dikembangkan oleh bank ketika akan melakukan pemeriksaan AI.

Several researchers and studies have proposed this assessment model. Lulana Sandu, Menno Wiersma, Daphne Manichand (2022) took an approach inspired by the risk management model (three lines system) for algorithm audit needs, especially for internal audits, called the Life Cycle Framework for Auditing AI Algorithms. This method serves as one example and can certainly be developed by the bank when conducting AI examinations.

Gambar 16. *The Life Cycle Framework for The AI Algorithm Audit*

Figure 16. The Life Cycle Framework for The AI Algorithm Audit



Source: Lulana Sandu, Menno Wiersma, Daphne Manichand, "Time to audit your AI algorithms", Maandblad voor Accountancy en Bedrijfseconomie, Amsterdam University Press, September 2022.

Tabel 7. *The Life Cycle Framework for The AI Algorithm Audit*

Table 7. The Life Cycle Framework for The AI Algorithm Audit

Lifecycle	Risk	Normative Statement
Initiation	Algorithms developed without adequate governance, insufficient support, or with excessively high risks can lead to unnecessary resource usage or create unwarranted costs.	1. Proof of adequate governance
		1.1 Relevant stakeholders (owners, validators, users), clearly explained.
		1.2 The role of stakeholders is explained (for example, owners approve the use of algorithms, validators act as a "second set of eyes" for the developer function, users provide feedback on algorithm usage).
		1.3 A clear separation between roles that need to be separated (for example, developers and validators) is implemented.
		1.4 Diversity in AI development teams includes race, gender, sexual orientation, age, economic status, etc., depending on potential biases.
		1.5 An organizational culture that involves "experts" for independent feedback on algorithms at all phases of their lifecycle (for example, domain experts involved in the algorithm usage phase).
		1.6 Algorithms are correctly listed in the algorithm inventory.
		2. Initial objectives, regulatory environment, and risk description
		2.1 The purpose of the algorithm is explained and shared with users.
		2.2 Regulatory risk categories (high or low) are determined and established.
		2.3 Regulatory requirements are listed (for example, the obligation to conduct conformity assessments according to laws) and used in algorithm development (for example, version control is applied during algorithm development).

Lifecycle	Risk	Normative Statement
Development	Inadequate skills are used for development and/or carried out without sufficient understanding of the context, resulting in data or algorithm reliability not meeting standards, which jeopardizes algorithm outcomes.	3. Documentation requirements, containing:
		3.1 Internal guidelines (for example, codes of ethics used, company ESG values) are applied and documented..
		3.2 Input and output data of the algorithm are explained (for example, datasheets for datasets that are used).
		3.3 The model's functioning is adequately documented (for example, model cards or method cards that are used).
		3.4 A risk registry (list of risks) containing all potential hazards caused by the algorithm is available.
		3.5 For all new algorithms, an initial impact assessment is conducted, including documentation of all potential risks, including ethical risks.
		4. Algorithm reliability, documenting:
		4.1 The selection and configuration of the algorithm (for example, hyperparameters) are appropriate and based on theoretical results (for example, compared with previous algorithm usage), ensuring that the algorithm does not inadvertently identify irrelevant relationships in the data or expose data containing noise.
		4.2 The selection of the algorithm is appropriate to the context in which it is applied (for example, the chosen algorithm aligns with business objectives; algorithm settings are configured according to hyperparameters or other settings that correspond with available data–for instance, more complex algorithms may require larger data samples).
		4.3 Approaches to improve the algorithm (for example, regularization, activation functions, optimization) are correctly applied.
		4.4 External tools used (for example, text parsers that extract features from text data) are well understood.
		4.5 Mitigation steps are applied to risks and used conservatively (for example, even if there is only a potential privacy risk, privacy constraints are enforced on the algorithm during development).
		4.6 Testing (for example, performance accuracy per sub-group, sensitivity/scenario analysis, statistical fairness tests, overfitting detection) is conducted to validate the algorithm's performance.
		4.7 The results of the algorithm are compared with expert opinions or other benchmarks (for example, results from other platforms or algorithms).
		4.8 If necessary, expert opinions on the algorithm or data are sought and utilized.
		4.9 If applicable, officially recorded expert opinions are documented and justified.
		4.10 Assumptions and limitations of the algorithm are explained (for example, through an algorithm model card) describing the scenarios in which the algorithm will be used).
		4.11 The results of the algorithm align with the company's ESG values (for example, the algorithm's outcomes do not conflict with the company's core values).
		4.12 For high-risk applications, decisions made by the algorithm are explainable and understandable by humans.
		5. Data quality is adequately described in the documentation:
		5.1 High-quality data: complete (for example, data is unbiased so it does not misrepresent protected groups), consistent, unique, timely, accurate, valid, complete, representative of the population where the algorithm will be used.
		5.2 Data transformation (for example, scaling, missing data imputation, feature engineering) is done correctly.

Lifecycle	Risk	Normative Statement
Implementation	Implementation unaligned with the developed algorithm, bad input data, or allowing misuse of the algorithm, which jeopardizes results during usage.	6. Implementation documentation complies with requirements and aligns with development:
		6.1 The implementation process is documented (for example, implementation can be carried out through randomized controlled experiments).
		6.2 The algorithm design is specified (for example, in a method card).
		6.3 Changes to the algorithm or data are explained and documented.
		6.4 There is documentation of functional testing and user acceptance, especially for external tools.
		6.5 Technical roles and permissions, defined.
		7. Implementation results (algorithm and output) align with the design:
		7.1 The algorithm prototype (code, data, model, output) aligns with the implementation.
		7.2 Testing is conducted to identify vulnerabilities (for example, robustness testing).
Use	The use of the algorithm does not align with the design, or vice versa, causing the algorithm to produce incorrect results or results that are not aligned with its objectives.	8. The use of the algorithm is documented, aligned with best practices, and consistent with the algorithm's objectives:
		8.1 There is documentation regarding the use of the algorithm.
		8.2 The use of the algorithm is aligned with its objectives and documentation.
		9. Staff training:
		9.1 Staff have knowledge of how to use the algorithm.
		10. Evidence of potential feedback and actual feedback from users:
		10.1 Implementation of user feedback logs.
Monitoring	Algorithm monitoring is untimely or does not track the correct indicators, making it difficult to ensure whether the model continues to perform as expected	11. Monitoring documentation complies with requirements and includes indicators with thresholds that signify model performance:
		11.1 Performance metrics (for example, performance accuracy) and acceptable thresholds are defined.
		11.2 Monitoring frequency is adequate and followed up (for example, monitoring may be ongoing for self-learning algorithms).
		11.3 Assumptions and limitations of the algorithm are described for the stated purposes and use of the algorithm.
		11.4 Conditional approval (for example, the algorithm is approved for direct use with additional bias screening) is monitored.
Review	Algorithm review is not aligned with its intended purpose or is not conducted in a timely manner, making it impossible to ensure that the implemented algorithm remains adequate and consistent with its intended objectives	12. Documentation complies with requirements and contains an analysis of goal alignment, as well as conclusions for implementing 're-parameterization,' 'improvement,' or 'redevelopment' according to the requirements:
		12.1 There is an established review frequency that is followed up accordingly.
		12.2 The review contains an analysis of goal alignment based on monitoring criteria (for example, whether the algorithm's use remains consistent with its intended purpose, and whether there is still sufficient knowledge and understanding of the algorithm).
		12.3 The review provides descriptions and timelines for planned improvements and changes in accordance with findings or identified weaknesses.

Lifecycle	Risk	Normative Statement
		12.4 If concluded through the review, re-parameterization is performed (for example, dynamic calibration of the algorithm where hyperparameters are automatically recalibrated, which may be necessary for self-learning algorithms).
		12.5 If concluded through the review, improvements are carried out.
		12.6 If concluded through the review, redevelopment is carried out.
		12.7 Previous issues have been resolved, findings and recommendations have been implemented according to plan (for example, risk mitigation by a specified date).
Retirement	Algorithms and data that are no longer used are not deleted, which hinders inventory management or leads to the misuse of algorithms without proper maintenance, or previously unused algorithms are reactivated, causing procedural failures.	13. Deletion procedures comply with these requirements:
		13.1 Dependence on other algorithms, addressed.
		13.2 There is no data redundancy.
		13.3 Algorithms are accurately reflected in the algorithm inventory.
		13.4 Algorithm and data versions are stored for audit purposes.
Validation	Validation is not in accordance with the process, or lacks adequate skills, resulting in insufficient challenge to development, implementation, and use, which jeopardizes algorithm quality.	14. There are efficient controls to ensure proper model implementation:
		14.1 Internal guidelines are followed and documented (for example, code of ethics).
		14.2 There are relevant risk evaluations (for example, including ESG risks) and classification (for example, high-risk algorithms are correctly identified).
		14.3 Algorithm implementation can be replicated from the documentation.
		15. There is fair challenge to the quality of the algorithm and data:
		15.1 There is an evaluation of whether the development process aligns with the underlying problem used by the algorithm (for example, the development team is sufficiently diverse, ensuring conceptual quality in algorithm selection).
		15.2 There is an evaluation of algorithm performance (for example, using statistical testing, k-fold cross-validation, under/overfitting analysis, sensitivity analysis, backtesting).
		15.3 Assumptions and limitations of the algorithm, challenged.
		15.4 There is evaluation of data quality.
		15.5 There is an evaluation of the Extraction-Transformation-Processing process to identify potential issues in how data is collected (for example, potential bias introduced/collected during the data collection stage).
		16. Findings and recommendations correspond to the identified weaknesses and comply with these requirements:
		16.1 Validation provides findings and recommendations in a timely manner.
		16.2 Validation provides risk levels (for example, non-urgent, informational, low, medium, high, or critical).
		16.3 Developers and users are consulted regarding findings, recommendations, and the level of findings.
		16.4 Conclusions from the Validation stage, followed up.

C. Perencanaan Audit

Dalam melakukan perencanaan audit AI terdapat beberapa langkah yang perlu dipersiapkan yaitu:

1. Menentukan ruang lingkup
   a. Tentukan fungsi AI yang akan diperiksa; dan
   b. Jenis dan lingkup area *output* yang ingin dilihat.
2. Kumpulkan bukti
   a. Diperlukan bukti yang mendukung efektivitas pengendalian sesuai dengan hal yang ingin didapatkan penjelasannya.
   b. Tentukan bagaimana bukti tersebut dapat di evaluasi dan membandingkannya dengan tujuan yang telah ditetapkan.
3. Tentukan jenis penilaian

   Pilih metodologi yang akan digunakan untuk mengevaluasi efektivitas kontrol.
4. Jalan prosedur asesmen dan pengujian

   Gunakan teknik dan langkah-langkah pengujian untuk mengevaluasi bukti. Langkah ini dapat mencakup *review* dokumen, melakukan prosedur analisa, pengujian, dan simulasi.

C. Audit Planning

In conducting AI audit planning, there are several steps that need to be prepared, namely:

1. Determining scope
   a. Determining AI functions to examine; and
   b. The types and scope of output areas to be reviewed.
2. Evidence gathering
   a. Evidence is required to support the effectiveness of controls in accordance with what is intended to be obtained.
   b. Determine how the evidence can be evaluated and compared with the established objectives.
3. Determine the type of assessments

   Choose the methodology to be used for evaluating the effectiveness of controls.
4. Execute assessment and testing procedures

   Use testing techniques and steps to evaluate evidence. This step may include document review, performing analysis procedures, testing, and simulation.

D. Audit AI oleh Pihak Eksternal

Penggunaan jasa pihak eksternal untuk melaksanakan fungsi audit internal AI tidak mengurangi tanggung jawab bank. Selain itu, penggunaan jasa pihak eksternal harus mempertimbangkan ukuran dan kompleksitas usaha. Dalam menggunakan jasa pihak eksternal untuk melaksanakan fungsi audit internal atas AI, bank perlu memperhatikan kerahasiaan data dan/atau informasi pada bank yang akan diakses oleh pihak eksternal.

Dalam hal bank menggunakan jasa pihak eksternal dalam pelaksanaan audit internal AI, penggunaan jasa pihak eksternal dilakukan sesuai dengan POJK mengenai penerapan fungsi audit internal bagi bank umum.

Untuk melakukan pemeriksaan AI di bank, otoritas terkait dan *standards setting bodies* telah mengeluarkan beberapa panduan dalam melakukan audit AI. Beberapa panduan yang ada dan dapat dikembangkan oleh bank antara lain:

1. **Value, Criteria, Indicators, and Observable (VCIO)**

   The AI Ethics Impact Group (AIEI) pada tahun 2019 mengeluarkan kerangka multi disiplin untuk menilai

D. AI Audit by External Parties

The use of external parties to carry out the internal AI audit function does not reduce the bank's responsibility. In addition, the use of external parties must consider the size and complexity of the business. When using external parties to perform internal AI audit functions, banks need to pay attention to the confidentiality of data and/or information accessed by external parties.

In cases where banks use external parties in conducting internal AI audits, such use must comply with POJK regarding the implementation of internal audit functions for commercial banks.

To conduct AI audits in banks, relevant authorities and standards-setting bodies have issued several guidelines for performing AI audits. Some existing guidelines that can be developed by banks include:

1. **Value, Criteria, Indicators, and Observable (VCIO)**

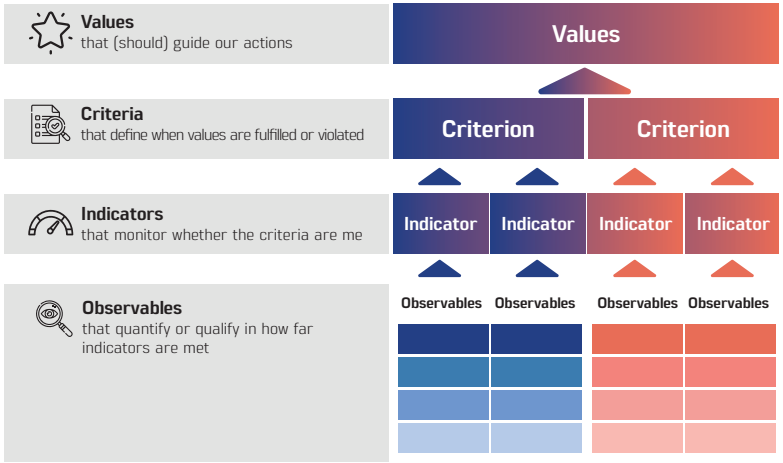
   The AI Ethics Impact Group (AIEI) in 2019 issued a multidisciplinary framework to assess AI

implementasi AI di tiap organisasi. *AI Ethics Impact Group* merupakan tim yang terdiri dari berbagai disiplin ilmu dengan berbagai latar belakang termasuk ilmu komputer, filsafat, ilmu sosial, fisika, dan teknologi. Kerangka yang disusun melakukan asesmen algoritma AI mencakup nilai (*values*), kriteria (*criteria*), Indikator (*indicators*) dan hasil pengamatan (*observables*) pada sebuah sistem AI. Model VCIO ini juga dilengkapi dengan matriks risiko untuk klasifikasi sistem AI yang berbeda.

implementation in each organization. The AI Ethics Impact Group is a team consisting of various disciplines with diverse backgrounds including computer science, philosophy, social sciences, physics, and technology. The developed framework assesses AI algorithms covering values, criteria, indicators, and observables within an AI system. This VCIO model is also equipped with a risk matrix for classifying different AI systems.

Tabel 8. *The VCIO Model*

Table 8. The VCIO Model



Source: AIEI (2019)

Gambar di atas menunjukkan terdapat 4 (empat) hierarki menggunakan pendekatan audit ini. Setiap tingkatan memiliki keterkaitan di mana pemenuhan nilai yang di bagian atas tergantung pada nilai pada tingkatan di bawah. Sebaliknya, tidak dimungkinkan pada model VCIO menghasilkan nilai pada kategori di bawahnya secara deduktif. Sebagai contoh untuk nilai terkait Keberlanjutan (*sustainability*) salah satu kriteria penilaian mungkin adalah 'Berkurangnya sumber daya' yang mungkin tidak dapat diukur seluruhnya dalam bentuk angka. Hal ini menunjukkan bahwa untuk setiap nilai/kriteria/indikator/hal-hal yang diamati tidak semuanya harus dibuktikan dengan angka/data kuantitatif dan sering diperlukan *judgment* oleh auditor secara normatif.

The image above shows 4 (four) hierarchies using this audit approach. Each level is interconnected, where fulfilling the values at the upper levels depends on the values at the lower levels. In contrast, it is not possible in the VCIO model to deductively generate values in the lower categories. For example, for a value related to Sustainability, one of the assessment criteria might be 'Reduced resources,' which may not be fully measurable in numerical form. This shows that for each value/criteria/indicator/observable, not everything must be proven with numbers or quantitative data, and often normative judgment by the auditor is required.

Pendekatan VCIO yang diusulkan oleh AIEI ini memiliki potensi terjadinya konflik atau kontradiksi antara setiap nilai/kriteria/indikator/hasil pengamatan. Untuk menghilangkan konflik tersebut maka nilai yang ada perlu dilakukan pemeringkatan sesuai dengan tingkat pentingnya nilai tersebut bagi bank. Hal ini dapat dilakukan melalui 2 (dua) pendekatan yaitu:

The VCIO approach proposed by AIEI has the potential for conflicts or contradictions between each value/criteria/indicator/observable. To eliminate these conflicts, the existing values need to be ranked according to their importance level for the bank. This can be done through 2 (two) approaches, namely:

1. *Bottom-up*: Titik awal adalah menentukan urgensi masalah yang timbul. Apakah hal ini dapat membahayakan keberlangsungan usaha bank dan jika hal ini tidak diselesaikan maka akan menghambat proses-proses selanjutnya.
2. *Top-down*: Untuk nilai tertentu dapat ditetapkan memiliki hierarki yang lebih tinggi karena nilai tersebut mempengaruhi tindakan lain. Untuk melakukan apakah nilai ini krusial dan mempengaruhi proses lain maka dapat digambarkan dalam suatu grafik ketergantungan (*path dependencies*). Untuk memberikan gambaran penggunaan metode VCIO ini ke nilai tertentu.

Bank dapat melihat contoh implementasinya ke nilai Transparansi pada tabel berikut:

1. Bottom-up: The starting point is to determine the urgency of the arising issue. Whether this can endanger the bank's business continuity and if left unresolved, it will hinder subsequent processes.
2. Top-down: For certain values, a higher hierarchy can be established because those values influence other actions. To determine whether a value is crucial and affects other processes, it can be illustrated in a dependency graph (path dependencies). This provides an overview of applying the VCIO method to specific values.

The bank can see an example of the implementation for the Transparency value in the following table:

Tabel 9. *VCIO Model* : Nilai Transparansi

Table 9. VCIO Model : Transparency Value

TRANSPARENCY						Value
Accessibility						Criteria
Are the modes of interpretability target-group-specific and have been developed with the target groups?	Who has access to information about data sets and the algorithm/ model used?	Is the operating principle comprehensible and interpretable?	Are the modes of interpretability in their target-group-specific form intelligible for the target groups?	Are the hyperparameters (parameters of learning methods) accessible?	Has a mediating authority been established to settle and regulate transparency conflicts?	Indicators
Yes	Everyone	Yes, the model itself is directly comprehensible	Yes, the modes of interpretability have been tested with target groups for intelligibility	Yes, to everyone	Yes, a competent authority has been established	Observables
Yes, but without participation of the target groups	All people directly affected	Yes, the modes of interpretability are provided with the model itself				
		No, the mode of interpretability can only be used post hoc by experts	Yes, target groups can complain or ask if they do not understand a mode of interpretability	Yes, but only to information and trust intermediaries (regulators, watchdogs, researchers, courts)	Yes, a competent authority has been established but its powers are limited	

Yes, but the modes of interpretability are only specific for one target group	Only information and trust intermediaries (regulators, watchdogs, research, courts)	No, the modes of interpretability need to be adjusted to the individual model and use by experts	No	No	No
No, the modes of interpretability[4] are not target-group-specific	Nobody	No, but the model is theoretically comprehensible			
		No, there are no known modes of interpretability			

4 "Modes of interpretability" refers to different methods to ensure or increase interpretability (use of simple model, explanation of data and model used, etc.).

TRANSPARENCY						Value
Disclosure of origin of data sets			Disclosure of properties of algorithm/model used			Criteria
Is the data's origin documented?	Is it plausible for each purpose, which data is being used?	Are the training data set's characteristics documented and disclosed? Are the corresponding data sheets comprehensive?	Has the model in question been tested and used before?	Is it possible to inspect the model so that potential weaknesses can be discovered?	Taking into account efficiency and accuracy, has the simplest and most intelligible model been used?[1]	Indicators
Yes, comprehensive logging of all training and operating data, version control of data sets etc.[2]	Yes, the use of data and the individual application are intelligible	Yes and the data sheets are comprehensive	Yes, the model is widely used and tested both in theory and practice[3]	Yes, the model can easily be inspected and tested	Yes, the model has been evaluated and the most intelligible model has been used	Observables
	Yes, it is intelligible on an abstract, not case specific level, which data is being used		Yes, the model known and tested in either theory or practice			
Yes, logging and version control through an intermediary (e.g. data supplier)		Yes, but (some) data sheets contain few or missing information		Yes, but the model can only be tested by certain people due to non-disclosure	No, but the model was evaluated regarding interpretability and this evaluation is disclosed to the public	
	No, but a summary on data usage is available		Yes, the model is known to some experts but has not been tested yet			
No logging: data used is not controlled or documented in any way		No		No	No, the model has not been evaluated	
	No		No, the model has been developed recently			

1 This indicator would require further specification regarding the balance between using an efficient and accurate model and using a model which is technically simple and thus naturally easier to comprehend and follow.
2 This observable could include further levels of logging and documentation of data sets.
3 This observable could help to determine the levels needed in other observables: if the model has been widely used and tested, it might not require additional testing.

Tabel 10. *VCIO Model*: Nilai Akuntabilitas

Table 10. VCIO Model: Accountability Value

ACCOUNTABILITY							Value
Assignment of internal organisational responsibility (prospective)					Technical measures to ensure accountability		Criteria
Has a system of central or shared responsibilities been established in the operating institution?	Are the responsibilities between different institutions clarified?	Have responsibilities been clarified with the system manufacturers during development?	Is the assignment of responsibilities regularly reviewed and updated?	In case of shared responsibility, do those responsible know their roles and duties?	Are there methods for complexity reduction of technical functions e.g. to ensure internal traceability?	Are systems with a learning component monitored in their interaction with their environment?	Indicators
Yes, there is a clearly defined contract	Yes, there is a clearly defined contract	Yes, there is a clearly defined contract	Yes, permanently	Yes, they have access to detailed documentation	Yes, techniques to casually explain outputs and to observe environmental influences on AI systems are available	Yes, techniques to casually explain outputs and to observe environmental influences on AI systems are available	Observables
Yes, the agreements are documented in another form	Yes, the agreements are documented in another form	Yes, the agreements are documented in another form	Yes, after significant changes to the application or its environment	Yes, but they are only informed of their own obligations	Yes, but monitoring and explanations are only possible with restrictions	Yes, but monitoring and explanations are only possible with restrictions	
No, but there was an oral agreement	No, but there was an oral agreement	No, but there was an oral agreement	Yes, at regular intervals	No	No	No	
No	No	No	No, does not take place				

ACCOUNTABILITY							Value
Corporate/institutional liability (retrospective)		Disclosure of internal organisational responsibilities (prospective)				Error tolerance	Criteria
Are appropriate monetary means an insurance policy and/ or other forms of compensation in place in case of liability?	Is there an ombudsperson?	Is there an institutionalised opportunity to provide anonymous information to relevant parties?	Are responsibilities defined with respect to third-parties (affected persons/ users)?	Are responsibilities for possible damage and liability cases documented?	Is there a comprehensive logging of the design process?	Is there a culture of dealing openly with mistakes within organisations?	Indicators

Yes, sufficient financial resources are available	Yes, a respective body has been established and openly announced	Yes, a respective body has been established and openly announced	Yes	Yes	Yes, comprehensive logging of all incoming training and operating data, version control of data records, etc	Yes, errors can be addressed without excessive penalty threats	Observables
Yes, funds are available for typical or probable claims, but not for less probable scenarios	Yes, but access is only possible when fulfilling certain requirements	Yes, but access is only possible with difficulties or full security is not guaranteed	No, but there are other ways to contact responsible persons	No	Yes, logging/version control from second parties (e.g. by data suppliers)	Yes, but openness to error leads to tolerance for error	
No	No	No	No, there is no office to contact		No, incoming data is not controlled or documented in any way	No, there is no sufficient focus on errors	

Selanjutnya, menggunakan contoh tersebut di atas bank dalam hal ini dapat melakukan penilaian atas setiap nilai-nilai utama perbankan Indonesia mengikuti pola yang sama.

Hasil penilaian yang didapat dari VCIO untuk setiap indikator kemudian menjadi dasar pemberian *rating* implementasi AI di bank. Jumlah *rating* dapat disesuaikan dengan sistem penilaian yang ada di masing-masing bank. Sebagai contoh mengacu pada SE OJK No. 24/SEOJK.03/2023 tentang Penilaian Tingkat Maturitas Digital Bank Umum hasilnya diklasifikasikan ke dalam 5 (lima) tingkat yaitu, Tingkat 1, Tingkat 2, Tingkat 3, Tingkat 4, dan Tingkat 5. Adapun Tingkat 1 mencerminkan kondisi pemenuhan yang paling baik dan Tingkat 5 mencerminkan tingkat yang paling rendah dan masih banyak kelemahan yang harus diperbaiki oleh bank. Sejalan dengan model VCIO tim riset AIEI mengelompokkan *rating* AI ke dalam 7 (tujuh) tingkatan dimulai dari Tingkat A sampai dengan Tingkat G. Tingkat A mencerminkan nilai terbaik yang dapat diperoleh oleh institusi yang menerapkan AI dengan pemenuhan berbagai indikator yang ada. Sebaliknya, Tingkat G adalah *rating* terendah dan masih banyak ditemukan kelemahan atau *gap* dalam sistem AI dengan nilai ideal yang ingin dicapai.

Next, referencing the example above, the bank can assess each of the main values of Indonesian banking by following the same pattern.

The assessment results obtained from VCIO for each indicator then become the basis for rating AI implementation in the bank. The number of ratings can be adjusted according to the existing evaluation system in each bank. For example, referring to OJK Circular Letter No. 24/SEOJK.03/2023 concerning the Assessment of Digital Maturity Levels of Commercial Banks, the results are classified into 5 (five) levels: Level 1, Level 2, Level 3, Level 4, and Level 5. Level 1 reflects the best fulfillment condition, while Level 5 represents the lowest level with many weaknesses that the bank must improve. In line with the VCIO model, the AIEI research team classifies AI ratings into 7 (seven) levels, ranging from Level A to Level G. Level A represents the best value that an institution implementing AI can achieve by meeting various existing indicators. Conversely, Level G is the lowest rating, indicating many weaknesses or gaps in the AI system compared to the ideal values to be achieved.

Sebagai panduan untuk menentukan *rating* sistem AI yang telah diimplementasikan di bank dapat dilakukan melalui 2 (dua) cara yaitu:

a. Untuk setiap indikator yang diamati dapat memiliki nilai metrik individu sehingga untuk nilai keseluruhan dapat dilakukan menggunakan nilai rata-rata. Sebagai ilustrasi misal untuk indikator tertentu mendapatkan nilai A atau yang paling tinggi, indikator kedua mendapatkan nilai D, dan variabel indikator yang diamati lainnya mendapatkan nilai B. Dengan menggunakan pendekatan ini maka secara rata-rata nilai yang diberikan untuk Kriteria yang membawahi indikator-indikator tersebut adalah B.

b. Bank dapat menetapkan nilai minimum yang harus dicapai untuk setiap aspek yang di observasi. Contoh asesmen menggunakan pendekatan ini seperti ketika badan resmi pemerintah menetapkan standar kualitas air yang baik untuk diminum adalah mengandung tidak lebih dari X fosfat yang dapat membahayakan orang yang meminumnya.

Untuk memberikan gambaran bagaimana proses *rating* atau pemberian nilai pada model VCIO dapat dilihat pada gambar berikut.

As a guide to determining the rating of the AI system implemented in the bank, it can be done through 2 (two) ways, namely:

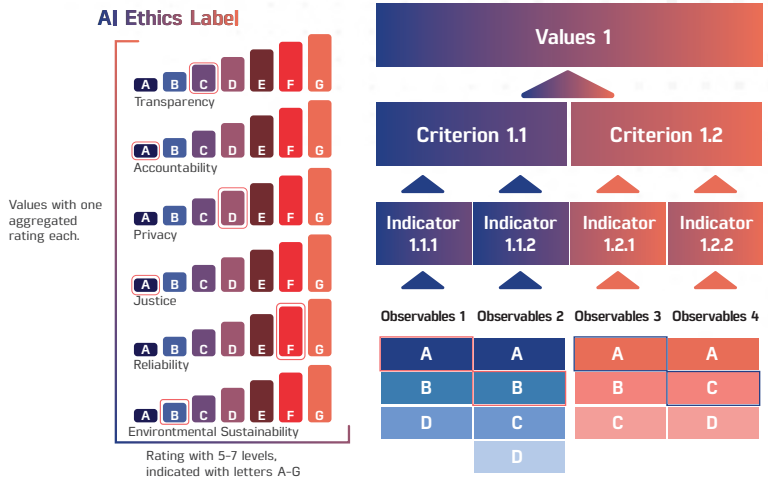
a. Each observed indicator can have an individual metric value, so the overall value can be calculated using the average. For illustration, if a certain indicator receives a rating of A (the highest), the second indicator gets a D, and other observed indicators receive a B, then using this approach, the average rating given for the criterion encompassing these indicators would be B.

b. The bank can set a minimum value that must be achieved for each observed aspect. An example of assessment using this approach is when an official government body sets a standard for good drinking water quality, such as containing no more than X amount of phosphate that could harm those who consume it.

To provide an illustration of how the rating or scoring process works in the VCIO model, please refer to the following image.

Gambar 17. Proses *Rating* dalam *VCIO Model*

Gambar 17. Rating Process in the VCIO Model



Source: AIEI

a. Konteks Klasifikasi AI dan Matriks Risiko

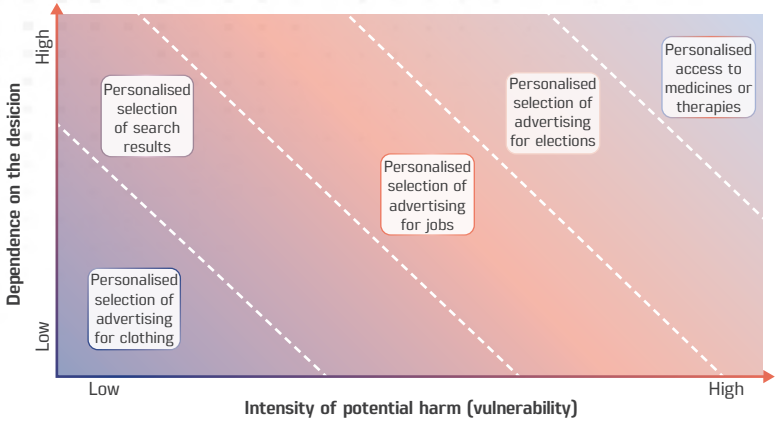
AI Ethics Impact memberikan saran agar dilakukan klasifikasi AI ke dalam matriks risiko. Untuk diketahui bahwa mengingat nilai-nilai AI yang diutamakan bagi industri dapat berbeda-beda. Misalkan untuk bidang kedokteran atau industri kesehatan nilai Transparansi sangat diutamakan karena terkait dengan hidup manusia. Untuk industri yang lain seperti Keuangan

a. Context for AI Classification and Risk Matrix

AI Ethics Impact recommends classifying AI into a risk matrix. It is important to note that the prioritized AI values can vary across industries. For example, in the medical or healthcare field, Transparency is highly prioritized because it relates to human life. In other industries such as Finance and Banking, other values like Accountability might

Gambar 18. Klasifikasi AI dan Matriks Risiko

Figure 18. AI Classification and Risk Matrix



Source: AIEI

dan Perbankan mungkin ada nilai lain yang dianggap lebih penting dibandingkan Transparansi misalnya Akuntabilitas dan seterusnya. Terlepas dari nilai apa yang menjadi prioritas klasifikasi AI harus berdasarkan risiko atau dampak yang diakibatkan dari implementasi AI di operasionalnya masing-masing.

Matriks di atas memberi panduan bagaimana sistem AI diklasifikasikan ke beberapa kelompok sesuai dengan risikonya. Garis horizontal menunjukkan potensi risiko tergantung pada intensitas dari kerugian/kerusakan yang mungkin timbul. Sedangkan untuk garis vertikal menunjukkan ketergantungan (vulnerabilitas) atas keputusan yang diambil oleh sistem AI.

1. Intensitas Potensi Risiko (Garis X)

   Untuk garis X atau horizontal merupakan evaluasi intensitas sistem AI yang berpotensi membahayakan manusia, organisasi (bank), dan lingkungan. Untuk menilai hal ini terdapat beberapa hal yang harus dipertimbangkan yaitu:

a. Dampak terhadap hak fundamental, kesetaraan, dan keadilan sosial. Apakah sistem AI berdampak negatif terhadap lingkungan/alam, hukum dan hak dasar manusia. Dampak yang juga menjadi perhatian termasuk akses ke asuransi kesehatan, pensiun, unsur demografis bahkan menyebabkan hal yang fatal bahkan kematian (contoh: penanganan kondisi pasien ICU).

b. Jumlah orang yang terdampak. semakin banyak orang yang terdampak maka makin tinggi risiko sistem AI tersebut (contoh: penilaian yang adil untuk penerimaan pekerjaan).

c. Dampak ke masyarakat. Apakah sistem AI berdampak ke masyarakat secara keseluruhan (contoh: pilihan pribadi, politik).

be considered more important than Transparency, and so on. Regardless of which value is prioritized, AI classification must be based on the risks or impacts resulting from AI implementation in their respective operations.

The matrix above provides guidance on how AI systems are classified into several groups according to their risk levels. The horizontal axis represents the potential risk based on the intensity of possible loss or damage. Meanwhile, the vertical axis indicates the dependency (vulnerability) on decisions made by the AI system.

1. Potential Risk Intensity (X-Axis)

   For the X-axis or horizontal line, it represents the evaluation of the intensity of an AI system's potential to harm humans, organizations (banks), and the environment. To assess this, several factors must be considered, including:

a. The impact on fundamental rights, equality, and social justice. Whether the AI system negatively affects the environment/nature, laws, and basic human rights. Other concerns include access to health insurance, pensions, demographic factors, and even causing fatal outcomes or death (for example: handling ICU patient conditions).

b. The number of people affected; the greater the number of people impacted, the higher the risk posed by the AI system (for example: fair assessment in job recruitment).

c. Impact on society. Whether the AI system affects society as a whole (for example: personal or political choices).

2. Ketergantungan pada Keputusan AI (Garis Y)

   Garis Y menunjukkan tingkat ketergantungan pihak yang berpotensi terdampak akibat dari keputusan algoritma AI. Ada 3 (tiga) faktor yang dapat di evaluasi dalam menilai ketergantungan atas keputusan AI yaitu:

a. Kendali (*control*) putusan sistem AI di-*filter* melalui tindakan yang sebelumnya dilakukan melalui interaksi manusia (contoh: rekomendasi pembelian di platform *e-commerce*). Hal ini tidak membutuhkan regulasi yang terlalu ketat dibandingkan dengan sistem AI yang mengambil keputusan otomatis tanpa interaksi dengan manusia seperti mematikan atau aktivasi sistem darurat reaktor energi nuklir.

b. Kemampuan untuk mengganti sistem AI dengan model AI lain dengan keputusan yang berbeda (*switch ability*). Ketergantungan pada satu sistem AI dan pasar pengembang AI yang monopolistik dapat mengakibatkan munculnya ketergantungan pada satu atau beberapa sistem AI.

c. Kemungkinan untuk mengubah keputusan algoritma AI (*redress*) dan waktu yang dibutuhkan untuk menindaklanjuti hal tersebut. Keputusan AI yang tidak dapat dikoreksi meningkatkan ketergantungan pada sistem AI. Keputusan AI butuh waktu yang lebih lama untuk dikoreksi menunjukkan risiko yang lebih tinggi dibandingkan putusan AI yang dapat di *redress* dengan cepat.

2. Dependency on AI Decisions (Y-Axis)

   The Y-axis shows the level of dependency of parties potentially affected by decisions made by AI algorithms. There are 3 (three) factors that can be evaluated in assessing the dependency on AI decisions, namely:

a. Control over AI system decisions is filtered through actions previously carried out via human interaction (for example: purchase recommendations on e-commerce platforms). This does not require strict regulation compared to AI systems that make automatic decisions without human interaction, such as shutting down or activating emergency systems in nuclear energy reactors.

b. The ability to replace an AI system with another AI model that makes different decisions (switchability). Dependency on a single AI system and a monopolistic AI developer market can result in reliance on one or a few AI systems.

c. The possibility to change AI algorithm decisions (redress) and the time required to follow up on this. AI decisions that cannot be corrected increase dependency on the AI system. AI decisions that take longer to be corrected indicate a higher risk compared to those that can be redressed quickly.

b. Deskripsi Matriks Risiko Sistem AI

a. KELAS 0 – Tidak Memerlukan *Rating*

   Intensitas dari dampak dan eksposur keputusan sangat rendah sehingga pengaturan sistem AI jenis ini tidak perlu dilakukan secara ketat.

   Mayoritas sistem AI akan berada pada klasifikasi ini dan direkomendasikan tidak diperlukan kewajiban untuk transparansi misalnya terdapat proses kontrol yang permanen. Dalam kasus terdapat keraguan akan sistem AI yang ada dapat dilakukan *post hoc* analisis.

b. KELAS 1

   Jika intensitas potensi risiko dan eksposur atas keputusan AI melebihi *threshold* tertentu dan diperlukan pengaturan/ *rating* sistem AI. Panduan AIEI merekomendasikan analisa sistem AI dari fenomena *Black Box*

b. AI System Risk Matrix Description

a. CLASS 0 – Does Not Require A Rating

   The intensity of the impact and exposure of the decision is very low, so regulation of this type of AI system does not need to be strict.

   The majority of AI systems will fall into this classification, and it is recommended that transparency obligations are not required, for example, if there is a permanent control process. In cases of doubt about the existing AI system, a post hoc analysis can be conducted.

b. CLASS 1

   If the intensity of potential risk and exposure from AI decisions exceeds a certain threshold and regulation/rating of the AI system is required, AIEI guidelines recommend analyzing the AI system regarding the Black Box

dan penjelasan bagaimana sistem mengambil keputusan secara otomatis yang berdampak sosial.

Untuk sistem AI klasifikasi ini perlu adanya kewajiban transparansi minimal termasuk *interface* untuk menganalisis sistem sebagai sebuah '*black box*' dan penjelasan bagaimana keputusan diambil oleh sistem.

c. KELAS 2

Sistem AI di kelas ini memiliki potensi risiko dan eksposur keputusan AI yang cukup tinggi. Input data yang dimasukkan ke dalam sistem harus diungkapkan secara penuh dan bagaimana keputusan diambil harus dapat diverifikasi.

Untuk AI jenis ini direkomendasikan transparansi atas nilai yang diterapkan oleh sistem dan penggunaan data. Bank diharapkan dapat menyediakan keputusan algoritma dalam satu daftar isi yang memungkinkan otoritas memonitor guna memastikan sistem bekerja sesuai dengan tujuannya.

d. KELAS 3

Sistem AI jenis ini memiliki potensi dampak yang sangat tinggi ke masyarakat/lingkungan.

phenomenon and explaining how the system makes automatic decisions that have social impacts.

For AI systems in this classification, there is a need for minimum transparency obligations, including an interface to analyze the system as a 'black box' and explanations of how decisions are made by the system.

c. CLASS 2

AI systems in this class have a fairly high potential risk and exposure from AI decisions. The input data fed into the system must be fully disclosed, and how decisions are made must be verifiable.

For this type of AI, transparency is recommended regarding the values applied by the system and data usage. Banks are expected to provide algorithmic decisions in a single index that allows authorities to monitor and ensure the system operates according to its purpose.

d. CLASS 3

This type of AI system has a very high potential impact on society and the environment. The

Sistem dapat berjalan tanpa sepengetahuan dari pihak yang terdampak atau bekerja tidak sesuai dengan ekspektasi yang diharapkan. Risiko yang ada harus diturunkan ke level yang rendah untuk menghindari keputusan sistem yang tidak sesuai.

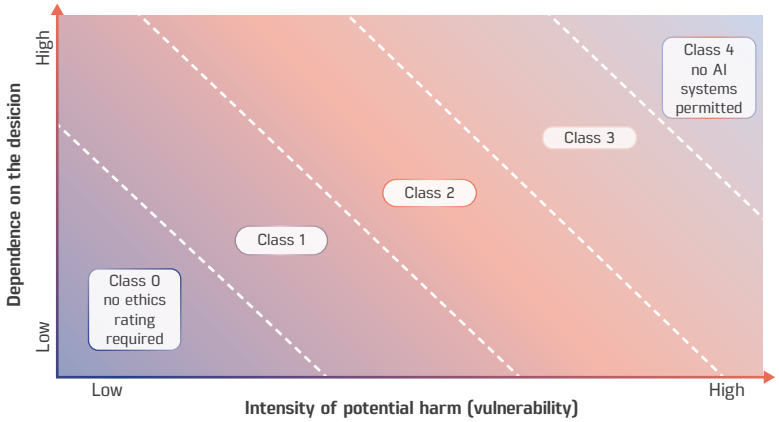
Jenis sistem AI di kelas ini perlu diawasi dan dipantau secara ketat meliputi *input* dan pelatihan data yang digunakan termasuk prosedur *machine learning*. Metode *machine learning* yang belum memenuhi unsur transparansi dan tidak dapat dijelaskan. Sebaiknya hanya sistem AI yang dapat dipahami bisa diberikan izin untuk diterapkan oleh Bank. Informasi yang terkait dengan sistem AI ini harus komprehensif dan dapat diverifikasi oleh pihak yang kompeten.

system may operate without the knowledge of affected parties or function in ways that do not meet expected outcomes. Existing risks must be reduced to a low level to avoid inappropriate system decisions.

AI systems in this class require strict supervision and monitoring, including the input and training data used, as well as machine learning procedures. Machine learning methods that lack transparency and cannot be explained should not be permitted. Only AI systems that can be understood should be authorized for implementation by the bank. Information related to these AI systems must be comprehensive and verifiable by competent parties.

Gambar 19. Matriks Risiko dengan 5 Kelas

Figure 19. Risk Matrix with 5 Classes



Source: AIEI

e. KELAS 4

Sistem AI di kelas ini memiliki potensi risiko yang sangat tinggi dan dapat membawa dampak kerusakan masal. Sistem AI yang berada pada kelas ini seharusnya tidak menggunakan proses *machine learning*. Untuk sistem AI yang berada pada kelas ini jika memang sangat dibutuhkan maka risiko dan potensi kerusakan yang dapat muncul perlu diturunkan melalui kerangka mitigasi risiko yang komprehensif sampai dengan tingkat risiko tersebut menjadi lebih rendah sehingga sistem AI dapat diklasifikasikan atau masuk ke kelas 3.

e. CLASS 4

AI systems in this class have a very high potential risk and can cause massive damage. AI systems in this class should not use machine learning processes. For AI systems in this class that are deemed absolutely necessary, the risks and potential damages must be reduced through a comprehensive risk mitigation framework until the risk level is lowered enough for the AI system to be reclassified or fall into Class 3.

Proses asesmen sistem AI yang dikembangkan oleh AI Ethics Impact Group ini bermanfaat bagi banyak *stakeholder* seperti bank yang akan mengembangkan dan menggunakan AI dalam operasionalnya, pengembang sistem, regulator, konsumen atau calon nasabah bank.

Gambaran dari proses pengawasan AI menggunakan kriteria VCIO ini secara keseluruhan dapat dilihat pada Gambar 19.

2. AI Audit Toolkit ISACA

Tidak terdapatnya standar audit AI yang baku dapat menyebabkan inkonsistensi pelaksanaan pemeriksaan sehingga laporan hasil audit menjadi sulit untuk dibandingkan. Panduan ISACA bertujuan sebagai standar metodologi yang dapat digunakan untuk mengevaluasi desain pengendalian, efektivitas operasional, alat, dan proses sistem AI. Sebagai titik awal auditor ketika melakukan audit AI dapat mengacu pada kerangka COBIT 2019 yang dikeluarkan oleh ISACA. COBIT 2019 dapat menjadi panduan oleh auditor karena mencakup deskripsi proses, *output* yang dihasilkan, praktik dasar, dan produk *output* dari setiap domain. Panduan ini berlaku umum sehingga dapat digunakan pada berbagai jenis organisasi termasuk bank yang mengimplementasikan AI.

The AI system assessment process developed by the AI Ethics Impact Group is beneficial for many stakeholders, such as banks that will develop and use AI in their operations, system developers, regulators, consumers, or prospective bank customers.

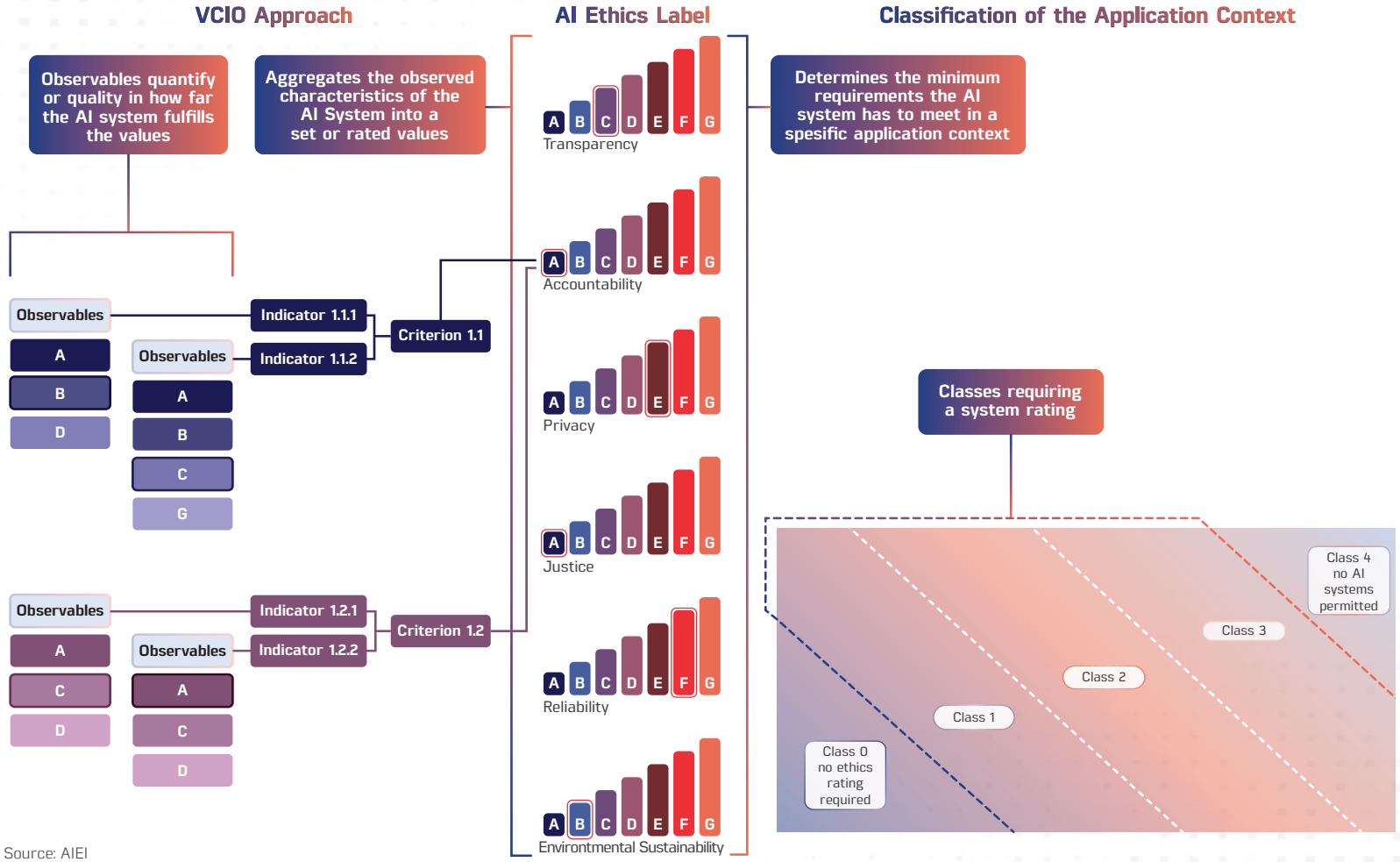
The overall overview of the AI oversight process using the VCIO criteria can be seen in Figure 19.

2. ISACA AI Audit Toolkit

The absence of standardized AI audit standards can lead to inconsistencies in the execution of inspections, making audit reports difficult to compare. The ISACA guidelines aim to serve as a methodological standard that can be used to evaluate the design of controls, operational effectiveness, tools, and processes of AI systems. As a starting point for auditors when conducting AI audits, they can refer to the COBIT 2019 framework issued by ISACA. COBIT 2019 serves as a guide for auditors because it includes process descriptions, expected outputs, best practices, and deliverables for each domain. These guidelines are general and can be applied across various types of organizations, including banks implementing AI.

Gambar 20. Ilustrasi Klasifikasi dan Pemeringkatan Risiko atas Implementasi AI

Figure 20. Illustration of Classification and Risk Rating for AI Implementation



Source: AIEI

Dilihat dari sisi strategi implementasi AI, terdapat beberapa risiko yang mungkin muncul seperti:

a. Tidak terdapat kesesuaian antara rencana TI dengan kebutuhan bisnis bank.
b. Rencana TI tidak konsisten dengan ekspektasi organisasi.
c. Penerjemahan rencana pelaksanaan TI yang keliru atau tidak sejalan dengan rencana strategis TI.
d. Struktur tata kelola tidak efektif untuk memastikan akuntabilitas proses TI yang terkait dengan AI.

From the perspective of AI implementation strategy, several potential risks may arise, such as:

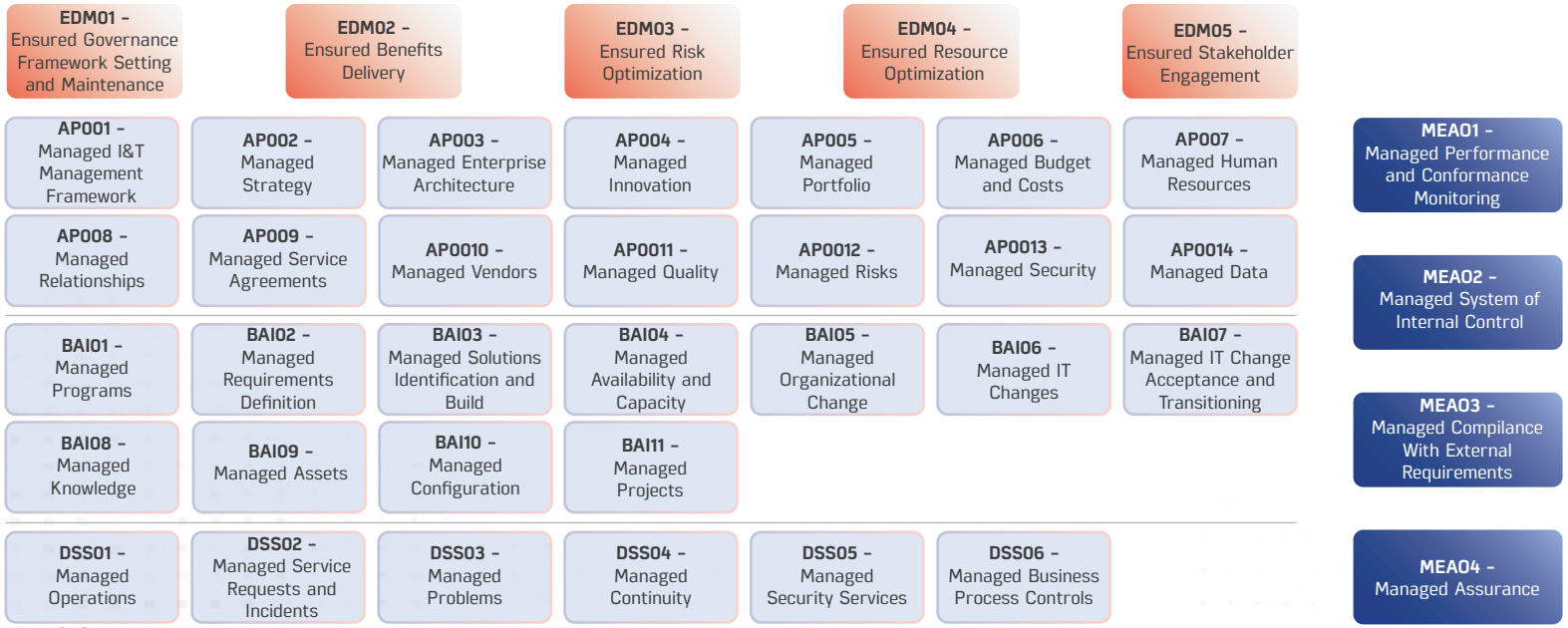
a. There is a mismatch between the IT plan and the bank's business needs.
b. The IT plan is inconsistent with the organization's expectations.
c. Incorrect or misaligned translation of the IT implementation plan with the IT strategic plan.
d. The governance structure is ineffective in ensuring accountability for IT processes related to AI.

Selain menggunakan kerangka COBIT 2019, ISACA juga menerbitkan panduan yang lebih rinci yang dilengkapi dengan kertas kerja untuk auditor dapat melakukan audit AI. Pedoman tersebut dituangkan dalam buku panduan Artificial Intelligence Audit Toolkit yang diterbitkan pada

In addition to using the COBIT 2019 framework, ISACA also published a more detailed guide equipped with working papers for auditors to conduct AI audits. This guideline is presented in the Artificial Intelligence Audit Toolkit book, published in 2024. The audit toolkit provides a

Gambar 21. Kerangka COBIT 2019

Figure 21. COBIT 2019 Framework



Source: ISACA

Tabel 11. Ruang Lingkup Audit *Toolkit* ISACA

Table 11. Scope of ISACA Audit Toolkit

Control Family	Description
Adversarial Defense & Robustness	Focuses on Strategies and techniques to protect AI systems against adversarial attacks, ensuring model integrity and realibility
AI Bias Mitigation & Fairness	Dedicated to identifying and mitigating biases in AI systems to ensure fair and unbiased decision-making processes
AI Data Privacy & Rights	Addresses the protection of data in AI systems, emphasizing personal data rights and compilance with data protection laws
AI Ecosystem Security	Focused on securing the AI ecosystem against external threats, ensuring the security of operations and data
AI Life Cycle Management	Covers the entire life cycle of AI systems, from development to maintenance, for effective management, and improvement
AI Model Governance	Focused on the responsible, ethical governance of AI models, ensuring compilance with standards and regulations
AI Operations	Involves operational aspects of AI systems, focusing on efficient and effective management
Assets Management	Focuses on managing assets in AI systems, ensuring optimal use and security
Audit & Compilace	Involves auditing AI systems for compilance wtih laws and internal policies, ensuring accountability
Business Continuity	Dedicated to maintaining and restoring business operations with AI during disruptions for continuous operation
Data Protection	Focuses on safeguarding AI system data against unauthorized access and breaches
Ethical AI Governance & Accountability	Encompasses ethical considerations and accountability mechanisms in AI systems
External Components & Supply Chain Governance	Manages and secures external components and supply chains of AI systems
Governance & Strategy	Involves overaching governance and strategic planning of AI initiatives
Human-AI Interaction & Experience	Concerned with interactions between humans and AI systems, focusing on user experience
Identity & Access Management	Includes controls for managing identities and access in AI systems
Incident Management	Involves handling incidents in AI systems, including response and recovery
Legal Regulatory & AI-Prohibited Use Cases	Addresses legal and regulatory aspects of AI, including compilance and prohibited use cases
Risk Management	Covers the identification and mitigation of risk associated with AI systems
Secure Systems Design & Development	Covers designing and developing AI systems with embedded security
Training & Awareness	Includes training and awareness initiatives for AI systems
User Privacy, Engagement, & Protection	Focuses on protecting user privacy and engagement in AI systems

Sumber: ISACA

tahun 2024. Audit *toolkit* memberikan pendekatan yang terstruktur untuk menilai kepatuhan dan pengendalian AI dengan spektrum yang luas sehingga bermanfaat untuk banyak pihak mulai dari pihak developer TI sampai dengan auditor TI. Lingkungan pengendalian sesuai panduan ini terbagi dalam beberapa kelompok untuk memudahkan penilaian.

Lingkungan pengendalian tersebut terdiri ke dalam 22 (dua puluh dua) parameter utama yang mencakup berbagai domain antara lain; Mitigasi Bias dan Keadilan, Data Privasi, Tata Kelola Model, Nilai Etika, Tata Kelola, dan seterusnya termasuk juga bagaimana ketahanan menghadapi serangan terhadap model AI dan keandalan mitigasi risikonya.

structured approach to assessing AI compliance and controls across a broad spectrum, making it useful for many parties, from IT developers to IT auditors. The control environment according to this guideline is divided into several groups to facilitate assessment.

The control environment consists of 22 (twenty-two) main parameters covering various domains, including Bias Mitigation and Fairness, Data Privacy, Model Governance, Ethical Values, Governance itself, and so on. It also includes aspects such as resilience against attacks on AI models and the reliability of risk mitigation measures.

Tabel 12. *ISACA AI Audit Toolkit Overview*

Table 12. ISACA AI Audit Toolkit Overview

Artificial Intelligence Audit Toolkit Overview								
1 Rationale	2 Responsibility	3 Data	4 Fairness	5 Safety & Performance	6 Impact	Control Fam ily	Control Category	Control Number	Control Name
						AI Bias Mitigation & Fairness	AI Fairness Certification & Standards	AD-D5-01	Standardized Procedure for Assesing High-Risk Systems
							Bias & Fairness Validation Tools	AF-BV-01	Unfairness Test Scope
								AF-BV-02	AI Systems and Discrimination
								AF-BV-03	Consideration of Adversarial Machine Learning for Bias Measurement
								AF-BV-04	Bias Detection & Connection
						AI Data Privacy & Rights	AI Data Access, Sharing, & Control	DP-AC-01	Enhanced Dataset Access
								DP-AC-02	Priorities Rights in AI
							AI Data Anomymization, Deidentification & Sensitive Data Handling	DP-AD-01	Data Privacy & Handling Protocols
								DP-AD-02	AI Data Retention & Encryption Protocols
								DP-AD-03	AI System & Data Protection - Training Data Compromise
								DP-AD-04	AI System & Data Protection - Training Data Protection Controls
								DP-AD-05	AI System & Data Protection
								DP-AD-06	Privacy-First Data Handling
								DP-AD-07	Differential Privacy
							Privacy-Enhancing Technologies	DP-PT-01	Privacy-Enhancing Technology Integration

Source: ISACA

Lingkungan pengendalian di atas kemudian dipetakan dengan mengacu pada ketentuan dan regulasi di berbagai negara sebagai:

***a. Alignment* kebijakan** – Evaluasi apakah kontrol sejalan dengan regulasi, standar industri, wilayah geografis, dan kebijakan bank.

The control environment described above is then mapped by referring to regulations and requirements in various countries as follows:

a. Policy alignment – Evaluation of whether controls align with regulations, industry standards, geographic regions, and the bank's policies.

b. Pengumpulan bukti – Mendapatkan bukti pendukung untuk memastikan kepatuhan termasuk dokumentasi, *audit logs*, pengaturan konfigurasi, dan bukti lainnya.

c. Prosedur Pengujian – Menguji prosedur yang diterapkan oleh bank untuk menguji efektifitas pengendalian risiko. Hal ini termasuk

b. Evidence gathering – Obtaining supporting evidence to ensure compliance, including documentation, audit logs, configuration settings, and other relevant proof.

c. Testing Procedure – Testing the procedures implemented by the bank to assess the effectiveness of risk controls. This includes

melakukan simulasi, penilaian, dan pengujian teknis.

d. **Dokumentasi dan Panduan** – Sebagai dokumentasi dan materi panduan untuk membantu auditor dalam melakukan asesmen pengendalian AI.

Lebih lanjut di dalam 22 (dua puluh dua) parameter utama terdapat 86 (delapan puluh enam) area kunci yang kemudian terdiri dari 251 kontrol. Sebagai contoh gambar di bawah memberikan gambaran bagaimana proses audit AI dilakukan menurut panduan dari ISACA. Elemen dasar yang dikembangkan ISACA lewat panduan ini mencakup beberapa hal:

a. Sintesis dan pemetaan pengendalian

Struktur pengendalian termasuk definisinya merupakan sintesis tim ahli ISACA menggunakan berbagai standar yang dikeluarkan oleh otoritas negara dan lembaga yang berwenang seperti The National Institute of Standards and Technology (NIST) 800-53, International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 27001. Selain itu ISACA juga telah melakukan review atas beberapa panduan AI yang spesifik mengatur hal-hal tertentu antara lain EU AI Act, Singapore's Model AI Governance Framework, MITRE Atlas, dan The Open Worldwide Application Security Project (OWASP) Machine Learning Security Top Ten.

conducting simulations, evaluations, and technical testing.

d. **Documentation and Guidance** – As documentation and guidance material to assist auditors in conducting AI control assessments.

Furthermore, within the 22 (twenty-two) main parameters, there are 86 (eighty-six) key areas which then consist of 251 controls. As an example, the image below provides an overview of how the AI audit process is conducted according to ISACA guidelines. The basic elements developed by ISACA through this guide include several aspects:

a. Synthesis and mapping of controls

The control structure, including its definition, is a synthesis by the ISACA expert team using various standards issued by national authorities and competent institutions such as The National Institute of Standards and Technology (NIST) 800-53, International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 27001. In addition, ISACA has also reviewed several AI guidelines that specifically regulate certain matters, including the EU AI Act, Singapore's Model AI Governance Framework, MITRE Atlas, and The Open Worldwide Application Security Project (OWASP) Machine Learning Security Top Ten.

b. Dapat Dijelaskan (*Explainability*)

Pendekatan ini secara umum mengakui bahwa *output* AI yang dapat dijelaskan merupakan salah satu kendali penerapan tata kelola dan manajemen proses yang baik. Kontrol yang dikembangkan mengacu pada 6 (enam) dimensi nilai Explainability yang diusulkan oleh The Information Commissioner's Office (ICO) dan The Alan Turing Institute. Hal ini mencakup Rasional, Tanggung Jawab, Data, Keadilan, Keamanan, dan Dampak. Dampak setiap kontrol tersebut kemudian di cek kontribusinya terhadap nilai transparansi, akuntabilitas, dan sehingga output dari sistem AI dapat dijelaskan.

Alternatif lain pendekatan yang dapat digunakan oleh seorang auditor AI yaitu dengan mengevaluasi *library* kontrol AI dan kesesuaiannya dengan tujuan pemeriksaan. Sebagai contoh ketika bank menggunakan AI untuk mengidentifikasi transaksi keuangan mencurigakan maka kendali deteksi anomali menjadi penting untuk

Singapore's Model AI Governance Framework, MITRE Atlas, and The Open Worldwide Application Security Project (OWASP) Machine Learning Security Top Ten.

b. Explainability

This approach generally acknowledges that explainable AI output is one of the controls for implementing good governance and process management. The controls developed refer to 6 (six) dimensions of Explainability value proposed by The Information Commissioner's Office (ICO) and The Alan Turing Institute. These include Rationale, Responsibility, Data, Fairness, Security, and Impact. The impact of each control is then checked for its contribution to transparency, accountability, and thus the explainability of AI system outputs.

Another approach that can be used by an AI auditor is to evaluate the AI control library and its alignment with the examination objectives. For example, when a bank uses AI to identify suspicious financial transactions then anomaly detection controls become important to examine their implementation.

Tabel 13. Dimensi Keterjelasan (*Explainability*) dalam ISACA AI Audit *Toolkit*

Table 13. ISACA AI Audit Toolkit Explainability Dimensions

No	Explanation Dimension	Explanation Dimension Description	Explanation Dimension Evidence/ Deliverables/Artifacts	Explanation Dimension Assessment Types	Explanation Dimension Assessment Method
1	**Rationale Explanation**	Focuses on the logical reasoning behind AI desicions; aims to make the decision-making process of AI systems transparent and comprehensible	Logic maps, decision trees, and documentation explaining the AI model's reasoning	Evaluation of the clarity and logic of the explanations provided by AI models	Reviewing the logical coherence and transparency of decision-making processes in AI models
2	**Responsibility Explanation**	Delineates the accontability framework within AI operations, clarifying who or what is responsible for specific decisions and actions	Accountability matrices, role definitions, and documentation of decision ownership	Assessment of clear lines of accountability and responsibility in AI system operations	Analyzing documentation and system design to ensure clear assignment of responsibilities
3	**Data Explanation**	Focuses on the origin, nature, and processing of data used by AI systems	Data lineage records, processing logs, and data source documentation	Verification of the authenticity. relevance, and integrity of the data used	Inspecting data handling practices, source validation, and data integrity checks
4	**Fairness Explanation**	Addresses how AI systems ensure fairness and avoid bias in their operations	Fairness metrics reports, bias detection analyses, and corrective action plans	assessment of mechanisms and practices in place to detect and mitigate bias	Evaluating the effectiveness of bias detection tools and the implementation of fairness guidelines
5	**Safety & Performance Explanation**	Elucidates the measures taken to ensure the safety and optimal performance of AI systems	Safety protocols, performance testing results, and maintenance records	Evaluation of the systems safety features and performance standards	Reviewing safety compliance documents and performance benchmarking results
6	**Impact Explanation**	Explores the broader implications of AI systems, including the societal, ethical, environmental, and regulatory impacts	Impacts assessment reports, ethical considerations documentation, sustainability analyses, and compliance and regulatory, impact assessment reports	Analysis of the AI system's impact on users, society, and the environment; analysis of the impact of legislation, regulations, and frameworks	Examining comprehensive impact assessments and sustainability strategies; identifying relevant laws and regulations, evaluating regulatory impact, and assessing, level of compliance

Source: ISACA

diperiksa implementasinya. Adapun tujuan dari asesmen, waktu, dan sumber daya yang tersedia akan menentukan jenis pengujian dan bukti yang harus dikumpulkan. Transaksi yang tidak wajar dapat diperiksa menggunakan teknik anomali transaksi dengan contoh seperti di bawah:

The purpose of the assessment, timing, and available resources will determine the type of testing and evidence that must be collected. Unusual transactions can be examined using transaction anomaly techniques with examples as shown below:

Tabel 14. Teknik Deteksi Anomali ISACA

Table 14. Anomaly Detection Techniques ISACA

Control and Description	Explainability Dimensions	Evidence/Deliverables/Artifacts
ADR-DM 01 Anomaly Detection Techniques Use anomaly detection techniques to detect abnormal behavior in the training data, such as sudden changes or data labelling issues.	**Rationale** Anomaly detection techniques are implemented in an AI system to identify unusual or unexpected patterns, behavirs, or data points. The rationale for this control is to ensure that the AI system can make informed desicions by detecting and flagging anomalies. which may be indicative of errors or malicious activity. This explanation should be delivered in a nontechnical and accessible way for various stakeholders.	1. **Rationale Explanation Report**–A document that provides a high-level overview of the need for anomaly detection techniques in the AI system and their role in ensuring data integrity and desicion making. 2. **Visual Representation**–Diagrams or infographics that illustrate how anomaly detection works in the context of the AI system, making it easier for nontechnical stakeolders to understand.
	Responsibility This dimension requires identifying the personnel and departments responsible for the design, implementation, and oversight of anomaly detection techniques in AI systems. it entails specifying who monitors and evaluates the detection of abnormal behavior in training data and who addresses the detected issues and outline the chain of command and communication for handling such abnormalities..	1. **Anomaly Detection Team Roster**–A document listing AI team members involved with the anomaly detection process and their roles and contact information. 2. **Anomaly Response Protocols**–Formal documentation of the procedures for when anomalies are detected, including the escalation process and responsible parties. 3. **Training Data Audit Trails**–Detailed logs that track all actions taken on training data, linking each action to a responsible party.
	Data This Dimention identifies and explains the presence and nature of outliers or patterns in the training data that deviate from normal behavior. This process is crucial to ensure the training data's validity and to mitigate the risk of model bias or error propagation.	1. **Anomaly Detection Reports**–Document outlining the anomalies detected and potential data quality issues. 2. **Data Quality Assessments**–Evaluations of data qualty before and after cleaning. 3. **Data Audit Logs**–Chronological records of the data examined and the anomalies found. 4. **Documentation of Remediation Steps**–Descriptions of the procedures taken to address the identified anomalies. 5. **Updated Data Sets**–The final, cleaned data sets used for model training after anomaly correction.
	Fairness This dimension clarifies how such techniques pinpoint and rectify irregular patterns in the training data that may contribute to biased outcomes. It explains the rationale behind detecting outliers that could skew decision making and endures equitable AI operations.	1. **Comprehensive Anomaly Assesment Report**–Documents the nature of each anomaly, potential for bias, corrective measures, and changes in decision pattern. 2. **Anomaly Resolution Log**–Captures the decision trait from detection to resolution of biases. 3. **Equity Impact Statements**–Summarizes how the correction of anomalies contributes to equitable outcomes.

Control and Description	Explainability Dimensions	Evidence/Deliverables/Artifacts
	Safety and Performance This dimensions details the steps and methodologies implemented within anomaly detection techniques to identify and adress adnormal behavior in training data. This ensures the AI system's decisions are based on accurate, reliable data, thereby enhancing the system's overall safetyand performance.	1. **Anomaly Detection Reports**–Detailed documentation of the detected anomalies, potential impact on AI performance, and remediation actions taken. 2. **Data Quality Assessment**–A thorough report evaluating the quality of training data before and after anomaly detection interventions. 3. **Model Performance Records**–Pre- and post-deployment performance metrics of the AI model demonstrating the efficacy of anomaly detection controls.
	Impact This dimension involves detailing how anomaly detection techniques can influence both the individual and society by identifying and addressing abnormal behaviors in AI training data, it focuses on the potential consequences if such anomalies go undetected, such as biased outcomes, misrepresentations, and the propagation of inaccuracies, which can have broad implications for fairness and trsut in AI systems.	1. **Anomaly Impact Reports**–A comprehensive document outlining potential and actual impacts identified through anomaly detection, including case studies where interventions prevented harm 2. **Algorithmic Impact Assessments (AIA)**–Assessments that evaluate the potential impacts of anomalies on individuals and society, often required by regulations or industry standards. 3. **Data Quality and Integrity Framework**–A framework outlining the procedures and standards for maintaining data quality, mitigating the impact of anomalies on the AI's decision making process.

Source: ISACA

3. The Institute of Internal Auditors (IIA)

Untuk bank di mana AI telah dikembangkan dan diterapkan, auditor internal harus berdiskusi dengan tim/bagian/divisi yang menangani teknologi. Diskusi tersebut mencakup meminta mereka menjelaskan AI/algoritma yang telah diterapkan, termasuk fungsi dan sumber data, penggunaan, batasan, risiko, dan implikasi etis. Auditor internal juga harus mulai memahami pengendalian apa yang ada untuk membantu mengelola risiko yang ditimbulkan oleh AI. Memperoleh pemahaman awal tentang desain pengendalian yang digunakan untuk mengelola risiko terkait AI merupakan langkah penting bagi auditor. Bagi bank yang tidak mengetahui dengan jelas apakah atau bagaimana AI digunakan (secara formal atau informal), maka diskusi dengan fungsi atau divisi TI bank merupakan titik awal yang baik karena bagian tersebut biasanya memiliki kecenderungan

3. The Institute of Internal Auditors (IIA)

For banks where AI has been developed and implemented, internal auditors must engage in discussions with the teams/units/divisions handling the technology. These discussions include asking them to explain the AI/algorithms applied, including functions and data sources, usage, limitations, risks, and ethical implications. Internal auditors should also begin to understand what controls exist to help manage the risks posed by AI. Obtaining an initial understanding of the control design used to manage AI-related risks is an important step for auditors. For banks that are not clearly aware of whether or how AI is used (formally or informally), discussions with the bank's IT function or division are a good starting point since this department has higher tendency to experiment with and utilize AI within their teams. If the IT team confirms that AI is being used, or if initial

yang lebih untuk bereksperimen dan memanfaatkan AI di tim mereka. Jika tim TI mengonfirmasi bahwa AI sedang digunakan, atau jika pengamatan awal menentukan AI digunakan dalam bank, maka langkah berikutnya adalah menentukan sejauh mana AI digunakan.

Ketika melakukan *field work* auditor internal bank biasanya berinteraksi dengan anggota C-level seperti dengan *Chief Financial Officer* (CFO), atau anggota eksekutif lainnya seperti *Chief Information Security Officer* (CISO) atau *Chief Information Officer* (CIO). Diskusi awal dengan anggota C-level memberikan kesempatan bagi tim audit mendalami penerapan AI di operasional bank. Pertanyaan yang dapat diajukan auditor kepada pihak eksekutif antara lain:

a. Apakah strategi AI telah ditetapkan? jika demikian, apa saja rincian strategi itu (termasuk aspek seperti penggunaan AI untuk memaksimalkan efisiensi operasi atau menggunakan AI untuk mengurangi biaya)?
b. Apakah C-level tersebut sudah menentukan siapa yang bertanggung jawab mengelola risiko terkait AI?
c. Peran apa yang dimainkan C-level dalam melibatkan Dewan Direksi (atau pengurus) untuk tata kelola AI?

Pada tahap ini, auditor internal perlu melakukan tindakan yaitu:

a. Meneliti penggunaan AI dalam organisasi mereka dan meninjau sumber daya eksternal.
b. Melakukan pertemuan dan diskusi terkait AI dengan manajemen, termasuk tim AI atau manajemen TI (atau keduanya) dan anggota eksekutif (CFO, CISO, CIO, dll.).
c. Berkolaborasi dengan manajemen dalam meninjau atau mengembangkan bagaimana AI dapat dimanfaatkan (atau direncanakan untuk digunakan di masa depan).
d. Memahami apakah tata kelola AI sudah diterapkan di bank.

Pedoman audit AI berdasarkan kerangka IIA pertama kali diterbitkan pada tahun 2017. Panduan ini terdiri atas 3 (tiga) domain yang memberikan gambaran sistematis kepada auditor proses audit yang dilakukan. Kerangka audit tersebut dikaitkan dengan model *Three Lines of Defense* yang terdiri dari Manajemen Pengendalian, Manajemen Risiko dan Kepatuhan, dan Auditor Internal dan Eksternal.

observations determine that AI is in use within the bank, then the next step is to determine the extent of AI usage.

When conducting fieldwork, internal auditors of the bank typically interact with C-level members such as the Chief Financial Officer (CFO) or other executives like the Chief Information Security Officer (CISO) or Chief Information Officer (CIO). Initial discussions with C-level members provide an opportunity for the audit team to delve into AI implementation in the bank's operations. Questions that auditors can ask executives include:

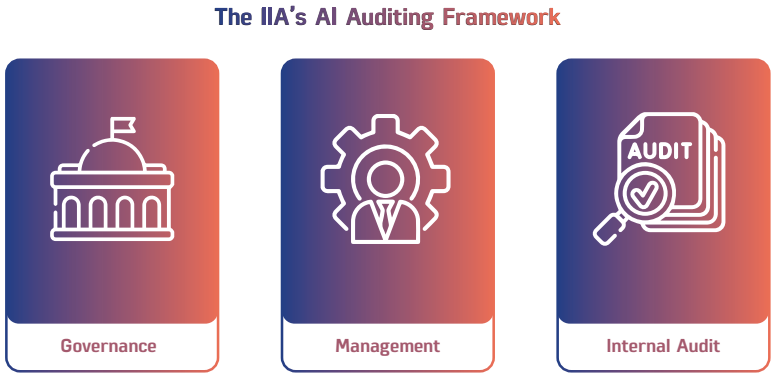
a. Has an AI strategy been established? If so, what are the details of that strategy (including aspects such as using AI to maximize operational efficiency or using AI to reduce costs)?
b. Has the C-level executive determined who is responsible for managing AI-related risks?
c. What role does the C-level executive play in involving the Board of Directors (or trustees) in AI governance?

At this stage, internal auditors need to take the following actions:

a. Investigate the use of AI within their organization and review external resources.
b. Conduct meetings and discussions related to AI with management, including the AI team or IT management (or both), and executive members (CFO, CISO, CIO, etc.).
c. Collaborate with management in reviewing or developing how AI can be utilized (or planned to be used in the future).
d. Understand whether AI governance has been implemented in the Bank.

The AI audit guidelines based on the IIA framework were first published in 2017. This guide consists of 3 (three) domains that provide a systematic overview to auditors of the audit process conducted. The audit framework is linked to the Three Lines of Defense model, which consists of Control Management, Risk and Compliance Management, and Internal and External Auditors.

Gambar 22. Kerangka Audit AI dari IIA
Figure 22. The IIA's AI Auditing Framework



Source: IIA

Tata Kelola – Domain pertama ini merupakan pendekatan strategis organisasi dalam penggunaan dan pengawasan AI. Tahap ini memberikan gambaran bagaimana AI direncanakan penggunaannya, dikendalikan dan dieksekusi oleh manajemen. Manajemen pengendalian tergantung pada informasi yang diberikan oleh fungsi internal audit bank. Auditor internal harus membuka jalur komunikasi yang baik untuk memberikan masukan kepada komite audit, anggota direksi dan komisaris, atau bagian lain di bank.

Governance – The first domain is the organization's strategic approach to the use and oversight of AI. This stage provides an overview of how AI usage is planned, controlled, and executed by management. Control management depends on information provided by the bank's internal audit function. Internal auditors must establish good communication channels to provide input to the audit committee, board members and commissioners, or other parts within the bank.

Manajemen – menjelaskan pendekatan yang akan diadopsi oleh organisasi ketika merencanakan dan menggunakan AI. Pada bagian ini juga dikaji terkait manajemen risiko bank terkait AI dan domain ini biasanya juga merupakan area audit bagi internal auditor.

Internal audit – mencakup aspek konsultasi dan audit untuk memberikan keyakinan yang memadai bagi manajemen bank. Titik awal bagi auditor internal bank ketika melakukan pemeriksaan terkait AI bisa dimulai dari mengevaluasi kerangka kebijakan terkait implementasi teknologi informasi atau AI.

Rencana strategis bagi bank untuk menerapkan AI perlu memperhatikan 2 (dua) poin penting yaitu:

a. Merencanakan strategi AI bukan hal yang dapat dilakukan dalam waktu singkat. Hal ini memerlukan proses yang berulang dan terus dilakukan secara berkala. Perlu ditentukan kesepakatan di internal bank mengenai jangka waktu yang tepat kapan akan dilakukan review kebijakan strategi AI secara berkala.
b. Strategi AI sebaiknya dilakukan melibatkan banyak fungsi/ departemen di bank mengingat

Management – explains the approach that the organization will adopt when planning and using AI. This section also reviews the bank's risk management related to AI and this domain is usually also an audit area for internal auditors.

Internal audit – includes aspects of consultation and audit to provide adequate assurance to the bank's management. The starting point for internal auditors when conducting AI-related examinations can begin by evaluating the policy framework related to information technology or AI implementation.

The strategic plan for the bank to implement AI needs to consider 2 (two) important points, namely:

a. Planning an AI strategy is not something that can be done quickly. It requires a repetitive process that is carried out regularly. An internal agreement within the bank needs to be established regarding the appropriate timeframe for periodically reviewing the AI strategy policy.
b. The AI strategy should involve many functions/departments within the bank considering the critical role of

krusialnya AI bagi operasional bank. Perencanaan seharusnya melibatkan C- level manajemen karena dampak AI dapat mengubah strategi dan operasional bank.

AI in bank operations. The planning should involve C-level management since the impact of AI can change the bank's strategy and operations.

Dalam setiap proses implementasi AI di bank perlu dipastikan bahwa hal tersebut dilakukan secara transparan, dapat dijelaskan, dilakukan secara bertanggung jawab, dan dapat dilakukan audit atas proses implementasi tersebut. Proses yang **transparan** maksudnya adalah tujuan penggunaan AI dan algoritmanya dapat dipahami. Mekanisme perhitungan, proses dan hasil output AI tersebut **dapat dijelaskan**. Penggunaan AI tersebut dilakukan secara **bertanggungjawab** mengacu pada nilai-nilai etika yang ditetapkan dalam buku panduan ini. Selanjutnya, ketika aplikasi AI yang diimplementasikan menggantikan proses yang dilakukan secara manual yang melibatkan manusia maka proses tersebut **dapat dilakukan audit** dan ditelusuri. Oleh karena itu, pengembangan AI juga seharusnya memfasilitasi adanya *log* audit dan atau informasi lainnya untuk memfasilitasi hal ini.

In every AI implementation process in the bank, it must be ensured that it is carried out transparently, in an explainable manner, responsibly, and that the implementation process can be audited. **Transparent** processes mean that the purpose of using AI and its algorithms can be understood. The calculation mechanisms, processes, and AI output results are **explainable**. The use of AI must be carried out **responsibly**, referring to the ethical values established in this guideline. Furthermore, when the implemented AI application replaces processes previously done manually involving humans, those processes must be **auditable** and traceable. Therefore, AI development should also facilitate audit logs and/or other information to support this capability.

Identifikasi risiko terkait AI terutama bagi bank yang baru akan menerapkannya pada operasional merupakan hal yang tidak mudah. Hal ini memerlukan kolaborasi antara internal audit, kepatuhan, hukum dan unit/fungsi/divisi lainnya yang dipandang terkait. Divisi/fungsi/unit yang melakukan asesmen manajemen risiko bank perlu mengidentifikasi dan mengembangkan *risk register* terkait AI. Sebagai contoh jika bank menggunakan AI yang dikembangkan oleh pihak ketiga maka perlu dipastikan terkait data dan algoritma yang digunakan untuk melatih model AI tersebut terdokumentasi dan tersedia untuk di evaluasi oleh pihak bank. Hal lainnya seperti *update* secara berkala risiko AI yang muncul kepada manajemen level atas.

Identifying AI-related risks, especially for banks that are just beginning to implement it in their operations is not an easy task. This requires collaboration between internal audit, compliance, legal, and other relevant units/functions/divisions. The division/function/unit responsible for the bank's risk management assessment needs to identify and develop a risk register related to AI. For example, if the bank uses AI developed by a third party, it must be ensured that the data and algorithms used to train the AI model are documented and available for evaluation by the bank. Other aspects include regularly updating emerging AI risks to senior management.

Asesmen risiko AI seharusnya mengikuti proses yang dilakukan oleh bank ketika melakukan penilaian risiko non AI. Dampak risiko dan probabilitas tingkat terjadi risiko tersebut harus diperhitungkan. Asesmen yang dilakukan sebaiknya

AI risk assessment should follow the same process that the bank uses when conducting non-AI risk assessments. The impact of the risk and the probability of its occurrence must be taken into account. The assessment should

juga menguantifikasi akibat yang muncul antar lain risiko hukum, reputasi, dampak keuangan dan lingkungan. Kombinasi antara dampak dan kemungkinan terjadinya risiko tersebut merupakan risiko bawaan (*inherent risk*) tanpa memperhitungkan adanya pengendalian internal untuk mitigasi risiko AI yang ada. Sebagai contoh ketika bank telah mengidentifikasi risiko keamanan AI maka serangan siber terhadap model AI merupakan salah satu ancaman. Untuk mengatasi risiko ini maka bank dapat menerapkan pengendalian serangan siber dengan beberapa teknik pengamanan yang ada untuk membuat risiko ini dapat diturunkan tingkat risikonya ke level yang dapat diterima oleh bank.

also quantify the consequences that may arise, including legal risks, reputational risks, financial impacts, and environmental effects. The combination of impact and likelihood represents the inherent risk without considering existing internal controls to mitigate AI-related risks. For example, when a bank has identified AI security risks, cyberattacks on AI models are one of the threats. To address this risk, the bank can implement cybersecurity controls using various existing security techniques to reduce the risk level to an acceptable level for the bank.

Selanjutnya, prioritas dari risiko AI perlu dilakukan pemeringkatan mulai dari yang paling penting dan membawa dampak besar ke bank. Sebagaimana diketahui organisasi sering menghadapi permasalahan terbatasnya sumber daya untuk mengatasi risiko-risiko yang ada di bank. Oleh karena itu perlu disusun skala prioritas dan perlu dilakukan pengkinian secara berkala. Setelah bank melakukan berbagai aktivitas terkait asesmen risiko dan rencana mitigasi maka perlu juga disusun mekanisme monitoring proses yang ada melalui audit internal. Tindakan yang dilakukan oleh bank di atas dapat dirangkum dalam 4 (empat) kegiatan sebagai berikut:

Next, AI risks need to be prioritized by ranking them from the most critical and impactful to the bank. As is well known, organizations often face limited resources to address existing risks within the bank. Therefore, a priority scale must be established and regularly updated. After the bank carries out various activities related to risk assessment and mitigation planning, it is also necessary to establish a monitoring mechanism for the ongoing processes through internal audit. The actions taken by the bank above can be summarized into 4 (four) activities as follows:

Tabel 15. Respon Risiko Dasar

Table 15. Basic Risk Responses

Response	Characteristics	Definition
Treat	Reduce, Mitigate, Enchance, Exploit, Leverage, Optimize	Apply controls to reduce inherent risk to an acceptable residual level or apply other measures to maximize and take advantage of potential possible variances in outcomes
Tolerate	Accept, Pursue	Determine whether portential benefits warrant taking the risk, having established measures considered necessary to mitigate or leverage likelihood and/or impact.
Transfer	Share, Spread	Spread risk either by transferring some or all of it to a third party (such as through insurance or outsorcing), or applying the resources of multiple teams to hedge against possible losses
Terminate	Avoid	Terminate or avoid risk by abandoning the planned action or elimination the goal altogether, prioritizing other goals in preferences

Source: IIA

IIA juga memberikan panduan awal yang dapat digunakan oleh auditor internal bank jika akan melakukan audit AI. Panduan tersebut dapat dilihat pada *checklist* di bawah dan dapat disesuaikan oleh masing-masing bank yang menerapkan AI dalam operasional mereka.

The IIA also provides initial guidelines that can be used by internal auditors of banks when conducting AI audits. These guidelines can be seen in the checklist below and can be adjusted by each bank implementing AI in their operations.

Tabel 16. Pertimbangan dalam Proses Audit AI

Table 16. Considerations in the AI Audit Process

Aspects or Considerations	Status / Results
Create a vision, strategy, and prioritization for AI and update frequently.	
Link AI initiative to organizational strategic objectives. (This may include revenue enhancing use cases, or internal applications to reduce cost or improve efficiencies.)	
Ensure that ethics, bias, social, and legal aspects are included in the strategy.	
Determine how to measure success of AI initiatives, including goals and ROI.	
Ensure that the AI strategic plan is consistent with the organization's risk culture.	
Ensure that the AI strategic plan is consistent with the organization's values.	
Ensure that the internal control environment is conducive for supporting AI. Consider what immediate policy changes are needed to support AI growth – adding a question about AI use in the third-party vendor management policy, for example.	
Define executive management responsible for overseeing AI initiatives.	
Establish a cross-functional AI Leadership Team to monitor all AI initiatives.	
Ensure that legal and compliance teams monitor all current and emerging regulatory requirements.	
Define the role of internal audit as an advisor and/or assurance provider.	
Ensure that Three Lines Model is in place and includes AI.	
Ensure that CISO (or equivalent) is involved in all AI initiatives.	
Ensure that third-party roles in AI initiatives are clearly defined and monitored.	
Ensure that finance/accounting tracks ROI on AI initiatives.	
Develop an AI acceptable use policy that is required for all employees.	
Develop policies and procedures for executing and maintaining AI initiatives.	
Develop policies and procedures for AI initiatives that utilize third parties.	
Ensure IT resources are sufficient to support AI initiatives and controls.	
Ensure staffing levels are sufficient to support AI initiatives and controls.	
Ensure HR recruiting has a focus on hiring practices for professionals with AI experience.	
AI leadership maintains required AI management knowledge.	
AI operational employees maintain required AI technical knowledge.	
All employees complete training regarding acceptable use and risks of AI.	
Include subject of AI in employee handbook and in new-hire orientation.	
Ensure that fair social, environmental, and economic aspects are considered in all AI-related projects.	
Ensure that AI-related data is secure, private, and confidential.	
Ensure that AI-related data is transparent, explainable, and responsible.	

Aspects or Considerations	Status / Results
Define objectives, goals, timing, and resource requirements for AI projects.	
Define operating responsibilities for all relevant employees in AI projects.	
Ensure that user access to AI is commensurate with job duties.	
Define data requirements and privacy considerations for AI projects.	
Define applicable legal and regulatory requirements for AI projects.	
Perform AI project risk assessment to identify possible threats to success.	
Define possible biases, including ethical and social considerations for AI projects.	
Define success metrics or project key performance indicators for AI projects.	
Establish reporting parameters such as frequency, content, and milestones for AI projects.	
Establish testing approach to validate AI is working as intended prior to and after going live.	
Report on achievement of metrics/KPIs to executive leadership and board.	
Ensure reporting includes disclosure of bias, ethical, or social concerns.	
Ensure reporting includes compliance with legal and regulatory requirements.	
Ensure reporting includes disclosure of any unintended or negative results.	
Ensure reporting includes disclosure of possible data loss or privacy breaches.	
Ensure that related internal controls are evaluated and reported periodically.	

Aspects or Considerations	Status / Results
Include AI as part of the enterprise risk management (ERM) process.	
Identify risks that threaten AI strategic goals and objectives.	
Identify risks that may have ethical, social, environmental, or financial implications.	
Identify risks that are related to the use of third parties for AI.	
Ensure that a process is in place to capture new or emerging risks.	
Ensure that employees with AI risk management responsibilities are properly trained.	
Perform an AI-based risk assessment and update periodically.	
Prioritize AI-related risks based on severity score (impact and likelihood).	
Ensure there is a process in place to select appropriate risk responses, including monitoring progress of responses.	
Ensure the organization is engaged with the board regarding AI strategy, goals, and objectives.	
Ensure the organization provides periodic updates to the board regarding AI in a manner that is clear and easily understandable.	
Ensure the organization engages the board regarding risk management approach for AI.	
Perform initial internal and external research of AI.	
Document if an AI strategy has been developed.	
Conduct internal discussions with established organizational relationships (such as IT and CFO) to understand how AI is currently being utilized and managed.	

Aspects or Considerations	Status / Results
Conduct initial discussions with AI/data science team (if applicable) and/or IT management.	
Create an inventory of current and planned AI uses.	
For current uses of AI, develop understanding of how it is being used, goals, and objectives.	
For planned uses of AI, develop understanding of approach, how risks are assessed, and plan for testing prior to deployment.	
Develop understanding of the following aspects of AI-related input data: • Governance. • Architecture. • User Access. • Cybersecurity Controls. • Processing Controls (integrity, accuracy, completeness). • Third-Party Considerations (SOC reports).	
Verify how AI is tested and reviewed to ensure it achieves its objectives and is free from biases, both pre-deployment and post-deployment.	
Verify that AI initiatives have clear objectives, and goals, and that projects are managed by an appropriate level of leadership.	
Verify that periodic reporting to the governing body is performed by management.	
Verify that AI is considered as a part of the enterprise risk management process, and includes risks related to: • Ethics. • Social and Economic Considerations. • Environmental Aspects. • Financial Implications. • Legal and Regulatory Violations.	
Verify that policies and procedures have been developed that outline how AI should be used and managed by the organization, including an AI acceptable use policy.	
Develop an understanding how an organization supports learning and training of AI to raise knowledge and awareness for all employees.	

Source: IIA

Pedoman audit yang dijelaskan di atas hanyalah beberapa contoh dari banyak panduan yang saat ini ada dan bisa menjadi tambahan rujukan bagi bank. Beberapa contoh pendekatan audit yang juga bisa menjadi referensi seperti *Guidance on the AI Auditing Framework* dari Information Commissioner's Office dan *Checklist for AI Auditing* yang dikeluarkan oleh The European Data Protection Board. Kedua prinsip ini bisa menjadi panduan ketika bank ingin melakukan audit khususnya terhadap pemenuhan kepatuhan AI pada UU Pelindungan Data Pribadi.

The audit guidelines described above are just a few examples among many existing references that banks can use as additional resources. Some other audit approaches that can also serve as references include the Guidance on the AI Auditing Framework from the Information Commissioner's Office and the Checklist for AI Auditing issued by The European Data Protection Board. These two principles can serve as guides when banks want to conduct audits, especially regarding compliance with AI under the Personal Data Protection Law.

Daftar Singkatan
List of Abbreviations

AI	*Artificial Intelligence*
AIDA	*Artificial Intelligence and Data Analytics*
AIEI	The AI Ethics Impact Group
AP	Akuntan Publik
BCG	The Boston Consulting Group's
BEC	*Business Email Compromise*
CAC	Cyberspace Administration of China
CBRN	*Chemical, Biological, Radiological, or Nuclear*
CCTV	*Closed Circuit Television*
CEO	*Chief Executive Officer*
CFO	*Chief Financial Officer*
ChatGPT	*Chat Generative Pre-trained Transformer*
CIO	*Chief Information Officer*
CISO	*Chief Information Security Officer*
COBIT	*Control Objectives for Information and Related Technologies*
CoE	*Center of Exellence*
CSAM	*Chile Sexual Abuse Material*
DFFT	*Data Free Flow with Trust*
DL	*Deep Learning*
DMAB	*Digital Maturity Assessment for Bank*
ECB	European Central Bank
ENISA	*European Union Agency for Cybersecurity*
ESG	*Environmental, Social, and Governance*
ETL	*Extract, Transform, and Load*
EU	*European Unions*
FEAT	*Fairness, Ethics, Accountability, and Transparency*
FLOP	*Floating Point Operations*
FSB	Financial Stability Board
G7	Group of Seven
GAN	*Generative Adversarial Networks*
GDPR	*General Data Protection Regulation*
GenAI	*Generative Artificial Intelligence*
GPAI	*General Purpose Artificial Intelligence*
HAM	Hak Asasi Manusia
ICU	*Intensive Care Unit*
ICO	The Information Commissioners Office
IIIA	The Institute of Internal Auditors
IMDA	Infocomm Media Development Authority
ISACA	Information Systems Audit and Control Association

ISO	International Organization for Standardization
IT	*Information Technology*
KAP	Kantor Akuntan Publik
KYC	*Know Your Customer*
LLM	*Large Language Model*
MAS	Monetary Authority of Singapore
METI	Ministry of Economy, Trade, and Industry
MIC	Ministry of Internal Affairs and Communications
ML	*Machine Learning*
NAIS	*The National AI Strategy*
NCII	Non-Censensual Intimate Imagery
NIST	National Institute of Standards and Technology
NITI	National Institution for Transforming India
OECD	Organisation for Economic Co-Peration and Development
OJK	Otoritas Jasa Keuangan
OWASP	*The Open Worldwide Application Security Project*
P2SK	Pengembangan dan Penguatan Sektor Keuangan
PDP	Perlindungan Data Pribadi
PII	Personaly Identifying Information
POJK	Peraturan Otoritas Jasa Keuangan
PSP	Pemegang Saham Pengendali
R&D	*Research and Development*
RMF	*Artificial Intelligence Risk Management Framework*
RPA	*Robotic Process Automation*
SDGs	*Sustainable Development Goals*
SDM	Sumber Daya Manusia
SE	Surat Edaran
SEOJK	Surat Edaran Otoritas Jasa Keuangan
TEVV	*Testing, Evaluation, Verification, and Validation*
TI	Teknologi Informasi
UE	Uni Eropa
UMKM	Usaha Mikro Kecil Menengah
UNESCO	United Nations Educational, Scientific, and Cultural Organization
USD	*United States Dollar*
UU	Undang-Undang
VCIO	*Value, Criteria, Indicators, and Observable*

Daftar Istilah
Glossary

Artificial Intelligence (AI)	Teori dan pengembangan sistem komputer yang mampu melakukan tugas-tugas yang biasanya memerlukan kecerdasan manusia. Sebagai bidang ilmu, AI telah berkembang selama bertahun-tahun. Namun, peningkatan daya komputasi dan tersedianya data dalam jumlah besar baru-baru ini menyebabkan meningkatnya minat terhadap potensi aplikasi kecerdasan artifisial. Saat ini, aplikasi AI telah digunakan untuk berbagai keperluan, seperti mendiagnosis penyakit, menerjemahkan bahasa, mengemudikan kendaraan secara otomatis, dan juga semakin luas penggunaannya dalam sektor keuangan.	The theory and development of computer systems capable of performing tasks that traditionally required human intelligence. Although AI has existed as a field for many years, recent advances in computing power and the exponential growth of data availability have sparked renewed interest in its potential applications. Today, AI is already being used to diagnose diseases, translate languages, and drive vehicles. It is also playing an increasingly important role in the financial sector.
Agentic AI	Kecerdasan artifisial yang memiliki kemampuan untuk bertindak secara mandiri dan otonom untuk mencapai tujuan tertentu atau menyelesaikan tugas tertentu tanpa perlu campur tangan manusia secara langsung.	Artificial intelligence with the ability to act independently and autonomously to achieve specific goals or complete tasks without direct human intervention.
AI regulatory sandboxes	Mekanisme pengujian sistem AI dalam lingkungan yang terkendali, memungkinkan eksperimen inovatif dengan pengawasan regulator.	Mechanisms for testing AI systems in controlled environments, allowing innovative experimentation under regulatory oversight.
Adversarial inputs	Input data yang dimodifikasi dengan sengaja agar model AI membuat kesalahan prediksi, meskipun perubahannya tidak kelihatan bagi manusia.	Input data that is deliberately altered to cause an AI model to make prediction errors, even though the modifications are imperceptible to humans.
Algorithmic bias	Ketidakseimbangan atau ketidakadilan dalam hasil sistem AI yang disebabkan oleh bias dalam data pelatihan atau algoritma.	An imbalance or unfair outcome produced by an AI system due to biases in the training data or the algorithm itself.
Black box AI	Sistem AI yang operasional internalnya tidak dapat dijelaskan atau dipahami secara mudah oleh manusia, termasuk oleh pengembangnya.	An AI system whose internal processes are difficult to explain or understand, even by its own developers.

Cyber security risk	Potensi kerugian atau dampak negatif yang mungkin terjadi pada system atau organisasi sebagai akibat dari ancaman yang mengeksploitasi kerentanannya dalam konteks teknologi informasi dan komunikasi.	The potential loss or negative impact to a system or organization resulting from a threat that exploits vulnerabilities in the context of information and communications technology.
Deep learning	Suatu bentuk pembelajaran mesin yang menggunakan algoritma yang bekerja dalam "lapisan" yang terinspirasi oleh struktur dan fungsi otak. Algoritma pembelajaran mendalam, yang strukturnya disebut jaringan saraf buatan, dapat digunakan untuk pembelajaran terbimbing, tak terbimbing, atau penguatan (yang merupakan bentuk pembelajaran mesin).	A form of machine learning that uses layered algorithms inspired by the structure and function of the human brain. These algorithms, known as artificial neural networks, are the foundation of deep learning and can be applied to supervised, unsupervised, or reinforcement learning tasks.
Deepfakes	Media sintetis yang menggantikan seseorang dalam gambar atau video yang ada dengan gambar orang lain menggunakan AI.	Synthetic media in which a person in an existing image or video is replaced with someone else's likeness using AI techniques.
Deep Neural Networks	Sebuah jenis jaringan saraf tiruan yang memiliki banyak lapisan tersembunyi antara lapisan *input* dan *output*, yang memungkinkan model ini untuk menangkap pola yang lebih kompleks dalam data.	A type of artificial neural network with multiple hidden layers between the input and output, enabling the model to learn and represent more complex patterns in the data.
Data privacy attacks	Serangan yang ditujukan untuk mengekspos, mengakses, atau mencuri informasi pribadi atau sensitif dari suatu sistem tanpa izin.	An attack designed to expose, access, or steal personal or sensitive information from a system without authorization.
Emerging threats	Ancaman baru atau berkembang yang berpotensi merusak stabilitas, keamanan, atau kesejahteraan suatu sistem, negara, atau populasi.	A new or emerging threat with the potential to undermine the stability, security, or well-being of a system, country, or population.
Emerging technology	Teknologi yang sedang dalam tahap awal pengembangan atau adopsi, dan memiliki potensi untuk mengganggu industri.	Technologies in the early stages of development or adoption that have the potential to significantly disrupt an industry.

Explainability	Kemampuan sistem AI untuk menjelaskan keputusan atau *output*-nya secara dapat dipahami oleh manusia.	The ability of an AI system to explain its decisions or outputs in a manner that is understandable to humans.
Explainable AI (XAI)	Sistem AI yang dirancang agar proses pengambilan keputusannya dapat dijelaskan oleh manusia.	An AI system designed to ensure that its decision-making process can be understood and explained by humans.
Environmental impact	Dampak pengembangan, pelatihan, dan penerapan sistem AI terhadap lingkungan, termasuk konsumsi energi dan emisi karbon.	The environmental impact of developing, training, and deploying AI systems, including factors such as energy consumption and carbon emissions.
Generative AI	AI yang menghasilkan konten baru, seperti teks, gambar, dan video, sering kali berdasarkan permintaan pengguna. AI generatif didukung oleh model dasar, seperti model bahasa yang besar.	AI that creates new content–such as text, images, or videos–often based on user prompts. Generative AI is typically powered by foundation models, including large language models
Generative Adversarial Networks (GAN)	Arsitektur AI yang terdiri dari dua jaringan saraf *generator* dan *discriminator* yang saling bersaing untuk menghasilkan data sintesis yang realistis.	An AI architecture composed of two neural networks–a generator and a discriminator–that compete with each other to produce realistic synthetic data.
General-Purpose AI (GPAI)	Sistem AI yang dapat digunakan untuk berbagai tujuan dan diterapkan dalam berbagai konteks, tidak terbatas pada satu aplikasi spesifik.	AI systems designed for multiple purposes and adaptable to various contexts, rather than being limited to a single specific application.
General AI	Kecerdasan buatan tingkat lanjut yang memiliki kemampuan kognitif setara atau melampaui manusia dalam berbagai tugas.	Advanced artificial intelligence with cognitive abilities equal to or exceeding those of humans across a wide range of tasks.
Human oversight	Keterlibatan manusia dalam pemantauan dan pengendalian AI untuk menjamin kesesuaian etika dan keamanan.	Human oversight of AI systems to ensure they operate ethically and safely, including monitoring and intervention when necessary.

Machine learning	Cabang AI yang berfokus pada pembangunan sistem yang dapat belajar dari dan membuat keputusan berdasarkan data. Alih-alih diprogram secara eksplisit untuk melakukan suatu tugas, sistem ML dilatih menggunakan sejumlah besar data, yang digunakan untuk membuat prediksi atau keputusan.	A branch of AI focused on developing systems that learn from data and make decisions or predictions without being explicitly programmed. Instead, machine learning systems are trained on large datasets to recognize patterns and improve performance over time.
Model extraction	Serangan di mana penyerang mencoba merekonstruksi model AI target dengan mengamati *input-output* model, sehingga bisa menyalin atau mengeksploitasi fungsi model.	An attack in which an adversary attempts to reconstruct a target AI model by analyzing its input-output behavior, with the goal of copying or exploiting its functionality.
Narrow AI	Jenis AI yang dirancang untuk melakukan tugas tertentu secara spesifik seperti pengenalan wajah atau deteksi penipuan. AI mulai beroperasi berdasarkan pemrograman dan aturan eksplisit.	A type of AI designed to perform a specific task, such as facial recognition or fraud detection. It operates based on explicit programming and predefined rules.
Neural Networks	Model Komputasi terinspirasi dari struktur otak manusia, terdiri dari lapisan neuron artifisial yang saling terhubung dan digunakan untuk mengenali pola dalam data.	A computational model inspired by the structure of the human brain, consisting of layers of interconnected artificial neurons, used to recognize patterns in data.
Natural Language Processing (NLP)	Bidang AI yang memungkinkan komputer untuk memahami, menafsirkan, dan menghasilkan bahasa manusia, dengan aplikasi seperti terjemahan otomatis dan *chatbots*.	A field of AI that focuses on enabling computers to understand, interpret, and generate human language, with applications such as automatic translation and chatbots.
Overreliance on AI	Ketergantungan berlebihan pada sistem AI untuk mengambil keputusan tanpa keterlibatan manusia yang memadai, yang dapat menimbulkan risiko kesalahan, bias, atau kegagalan etis.	Excessive reliance on AI systems for decision-making without adequate human involvement, which can lead to errors, biases, or ethical failures.
Predictive AI	Jenis AI yang menggunakan data historis untuk memprediksi kejadian atau hasil di masa depan.	A type of AI that analyzes historical data to predict future events or outcomes.

Robotic Process Automation (RPA)	Teknologi yang memungkinkan penggunaan perangkat lunak untuk meniru tindakan manusia dalam menjalankan proses bisnis berulang secara otomatis dilingkungan digital.	Technology that enables software to mimic human actions and automatically perform repetitive business processes in a digital environment.
Reinforcement learning	Suatu jenis paradigma pembelajaran mesin di mana suatu agen belajar membuat keputusan dengan mengambil tindakan dalam suatu lingkungan untuk mencapai suatu tujuan. Proses pembelajaran didorong oleh umpan balik yang diterima agen dari lingkungan dalam bentuk penghargaan atau hukuman.	A type of machine learning paradigm in which an agent learns to make decisions by interacting with an environment to achieve a goal, guided by feedback in the form of rewards or penalties.
Rising contenders	Teknologi, perusahaan, model AI, atau pendekatan yang menunjukkan potensi kuat untuk mendominasi di masa depan, meskipun saat ini belum menjadi pemimpin pasar.	A technology, company, AI model, or approach that demonstrates strong potential to become a future market leader, despite not currently holding a dominant position.
Red-teaming	Proses pengujian sistem atau teknologi suatu organisasi dengan mensimulasikan tindakan pihak penyerang untuk mengidentifikasi kelemahan atau kerentanan yang mungkin tidak terlihat oleh pengembang sistem.	The process of testing an organization's systems or technology by simulating the actions of an adversary to identify weaknesses or vulnerabilities that may not be evident to the system's developers.
Regulatory sandbox	Suatu kerangka kerja yang dibentuk oleh regulator yang memungkinkan pengujian inovasi secara langsung dalam skala kecil di bawah pengawasan regulator.	A framework established by a regulator that permits small-scale, live testing of innovations under regulatory oversight.
Social engineering	Teknik manipulasi psikologis yang digunakan oleh pelaku untuk mengecoh individu agar membocorkan informasi sensitif atau melakukan tindakan tertentu.	Psychological manipulation techniques used by attackers to deceive individuals into revealing sensitive information or performing specific actions.
Training data poisoning	Serangan terhadap sistem pembelajaran mesin di mana data pelatihan dimanipulasi secara sengaja untuk memengaruhi atau merusak hasil model.	An attack on a machine learning system where the training data is intentionally manipulated to influence or corrupt the model's outcomes.
Trustworthy AI	Kecerdasan artifisial yang dapat dipercaya, karena bersifat legal, etis, dan kuat secara teknis serta sosial.	Artificial intelligence that can be trusted because it is lawful, ethical, and technically as well as socially robust.

HALAMAN INI SENGAJA DIKOSONGKAN

THIS PAGE IS INTENTIONALLY LEFT BLANK

Daftar Pustaka
References

Buku/Publikasi:
Book/Publications:

Abercrombie, Cortnie. "Trustworthy AI explained with 12 principles and a framework". Articles. September 9, 2024. TechTarget. 2024.

AC Ventures, BCG, BCG X, Kadin Indonesia. "Harnessing the Power of (Gen) AI in Indonesian Financial Services". 2024.

AI Ethics Impact Group. "From Principles to Practice: An interdisciplinary framework to operationalise AI ethics". AI Ethics Impact Group. 2020.

AppliedAI Initiative GmbH. "Generative AI Agents in Action: Revolutionizing Software Development Testing". Initiative for Applied Artificial Intelligence. 2024.

AppliedAI Initiative GmbH. "Applying AI: Building the organization for scaling AI". Initiative for Applied Artificial Intelligence. 2023.

AppliedAI Initiative GmbH. "Applying AI: Elements of a Comprehensive AI Strategy 2nd Edition". Initiative for Applied Artificial Intelligence. 2023.

AppliedAI Initiative GmbH. "Enterprise Guide for Make-or-Buy Decisions". Initiative for Applied Artificial Intelligence. 2023.

Avicena Tech Corp, "The History of Artificial Intelligence", Avicena Tech Corp, 2024.

Awati, Rahul. "What is black box AI?". Articles. October, 2024. TechTarget. 2024.

BARC. "BARC Data Culture Survey 23 - How to Liberalize Data Access to Empower Data Users", Topical Survey. 2023.

Boston Consulting Group. "The AI Maturity Matrix". BCG Center for Public Economics. 2024

Brookings Institution. "Deepfakes And International Conflict". 2023.

Burke, John. "Unlocking the potential of white box machine learning algorithms?". Articles. December 7, 2022. TechTarget. 2022.

Burrell, J. "How the machine 'thinks': Understanding opacity in machine learning algorithms". 2016.

Deloitte. "Changing the game: the impact of artificial intelligence on the banking and capital markets sector". Deloitte China. 2024

Domin, Heather. "AI governance trends: How regulation, collaboration and skills demand are shaping the industry". Articles. Sep 5, 2024. World Economic Forum. 2024.

Eastgate Software. "Black Box AI: What Is It And How Does It Work?". Articles. January 26, 2024. Eastgate Software. 2024.

European Data Protection Supervisor. "TechDispatch: Explainable Artificial Intelligence". 2023.

European Union, "Proposal for a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts," 2021

European Union. "Artificial Intelligence initial strategy and deployment roadmap 2024-2025". 2024.

European Union Regulation No. 2024/1689 of June 13, 2024. "Artificial Intelligence Act". 2024.

European Confederation of Institutes of Internal Auditing. "The AI Act: Road to Compliance", A Practical Guide for Internal Auditors. 2025.

European Commission. "Ethics Guidelines for Trustworthy AI", High-Level Expert Group on Artificial Intelligence, 8 April 2019. 2019.

European Commission. "AI Watch Historical Evolution of Artificial Intelligence: Analysis of the three main paradigm shifts in AI". Publications Office of the European Union. 2020.

Faggella, Daniel. "AI Transparency in Finance – Understanding the Black Box". Articles. January 27, 2020. Emerj Artificial Intelligence Research. 2020.

Financial Services Information Sharing and Analysis Center. "Deepfakes in the Financial Sector: Understanding the Threats". Managing the Risks. 2024.

Financial Stability Board. "The Financial Stability Implications of Artificial Intelligence". FSB. 2024.

Future of Privacy Forum (FPF). "AI Governance Behind the Scenes: Emerging Practices for AI Impact Assessments". FPF. 2024.

G'sell, Florence. "Regulating Under Uncertainty: Governance Options for Generative AI". Stanford Cyber Policy Center. September, 2024.

G7 Hiroshima Summit. "Hiroshima AI Process G7 Digital & Tech Ministers' Statement". December 1, 2023. 2023.

Glover, Ellen. "What Is Black Box AI?". Articles. August 6, 2024. Built In. 2024.

Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courvile, Yoshua Bengio. "Generative Adversarial Nets". 2014.

Goswami, Kumar. "The AI/ML Revolution: Data Management Needs to Evolve". Articles. May 12, 2023. Komprise. 2023

Hiroki Habuka and David U. Socol de la Osa. "Shaping Global AI Governance Enhancements and Next Steps for the G7 Hiroshima AI Process". Centre for Strategic and International Studies. 2024.

HKMA. "Regtech Adoption Practice Guide - Artificial Intelligence-based Regtech Solutions". 2022.

HLB International. :Press Release 30 January 2025". 2025.

KPMG. "AI Q4 Pulse Survey: Key Findings, Q4 2024". AI & Digital Innovation. 2024.

Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission Singapore (PDPC). "Compendium of Use Cases: Practical Illustrations of the Model AI Governance Framework". 2020.

Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission Singapore (PDPC). "Compendium of Use Cases: Practical Illustrations of the Model AI Governance Framework (Volume 2)". 2020.

Info-communications Media Development Authority (IMDA and AI Verify Foundation. "Model AI Governance Framework for Generative AI, Fostering a Trusted Ecosystem". 2024.

International Organization for Standardization/ International Electrotechnical Commission, ISO/IEC 27001:2022. "Information security, cybersecurity and privacy protection, Information security management systems, Requirements". 2022.

Information Systems Audit and Control Association (ISACA). "COBIT 2019 Framework: Governance and Management Objectives". 2019.

Information Systems Audit and Control Association (ISACA). "Auditing Artificial Intelligence". 2019.

Information Systems Audit and Control Association (ISACA). "Artificial Intelligence Audit Toolkit". 2024.

Kelley, Kassidy. "Solving the AI black box problem through transparency?". Articles. August 16, 2021. TechTarget. 2021.

Kelompok Kerja Penyusun Strategi Nasional untuk Kecerdasan Artifisial yang dibentuk oleh Badan Pengkajian dan Penerapan Teknologi (BPPT). "Strategi Nasional Kecerdasan Artifisial Indonesia 2020-2045". 2020.

Kenton, Will. "What Is a Black Box Model? Definition, Uses, and Examples". Articles. April 2, 2024. Investopedia. 2024.

Kosinski, Matthew. "What is AI data management?". Articles. September 6, 2024. IBM. 2024.

Kosinski, Matthew. "What is black box artificial intelligence (AI)?". Articles. October 29, 2024. IBM. 2024.

Lalchand, et al. "Generative AI is expected to magnify the risk of deepfakes and other fraud in banking". Articles. May 29, 2024. Deloitte Center for Financial Services. 2024.

LeCun, Yann, Yoshua Bengio, Geoffrey Hinton. "Deep learning". Nature 521, 436-444. 2015.

Lee, Maggie C.M., Helana Scheepers, Ariel K.H. Lui, Eric W.T. Ngai. "The Implementation of Artificial Intelligence in Organizations: A Systematic Literature Review". Elsevier Journal Information & Management 60 (2023) 103816. Information & Management. 2023.

Leitner, Georg, Jaspal Singh, Anton van der Kraaij and Balázs Zsámbok - European Central Bank. "The Rise of Artificial Intelligence: Benefits and Risks for Financial Stability". The Financial Stability Review, May 2024. 2024.

Menteri Komunikasi dan Informatika RI. "Surat Edaran Menteri Komunikasi dan Informatika No. 9 Tahun 2023 tentang Etika Kecerdasan Artifisial". 2023.

Ministry of Internal Affairs and Communications Ministry of Economy, Trade and Industry, Japan. "AI Guidelines for Business Ver1.0, April 19, 2024". 2024.

Ministry of Internal Affairs and Communications Ministry of Economy, Trade and Industry, Japan. "Outline of AI Guidelines for Business Ver1.0, April 19, 2024". 2024.

Monetary Authority of Singapore. "Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector". 2024.

MIT Technology Review Insights. "AI readiness for C-suite Leaders, May 29, 2024". 2024.

National Institute for Standards and Technology. "Special Publication 800-53: Security and Privacy Controls for Information Systems and Organizations, Revision 5, USA, September 2020". 2020.

National Institute of Standards and Technology. "Artificial Intelligence Risk Management Framework (AI RMF 1.0)". U.S. Department of Commerce. 2023.

National Institute of Standards and Technology. "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1)". Juli 2024. 2024

National Institution for Transforming India (NITI Aayog). "National Strategy for Artificial Intelligence #AiforAll". 2018.

National Institution for Transforming India (NITI Aayog). "Responsible AI #AIforAll, Approach Document for India, Part 1 – Principles for Responsible AI". 2021.

National Institution for Transforming India (NITI Aayog). "Responsible AI #AIforAll, Approach Document for India: Part 2 - Operationalizing Principles for Responsible AI". 2021.

Otoritas Jasa Keuangan. "Surat Edaran Otoritas Jasa Keuangan No. 21/SEOJK.03/2017 tentang Penerapan Manajemen Risiko Dalam Penggunaan Teknologi Informasi oleh Bank Umum". 2017.

Otoritas Jasa Keuangan. "Peraturan Otoritas Jasa Keuangan No. 11/POJK.03/2022 tentang Penyelenggaraan Teknologi Informasi oleh Bank Umum". Tambahan Lembaran Negara Republik Indonesia Nomor 5/OJK. 2022.

Otoritas Jasa Keuangan. "Peraturan Otoritas Jasa Keuangan No. 1/POJK.03/2019 tentang Penerapan Fungsi Audit Intern pada Bank Umum". Tambahan Lembaran Negara Republik Indonesia Nomor 6308. 2023.

Otoritas Jasa Keuangan. "Peraturan Otoritas Jasa Keuangan No. 17 Tahun 2023 tentang Penerapan Tata Kelola bagi Bank umum". Tambahan Lembaran Negara Republik Indonesia Nomor 53/OJK. 2023.

Otoritas Jasa Keuangan dan Asosiasi Fintech di Indonesia. "Panduan Kode Etik Kecerdasan Buatan (*Artificial Intelligence*/AI) yang Bertanggung Jawab dan Terpercaya di Industri Teknologi Finansial". 2023.

Oxford Insight. "GovernmentAI Readiness Index 2023". Oxford Insight. 6 December 2023.

Personal Data Protection Commission. "Model Artificial Intelligence Governance Framework". Second Edition. 21 January 2020. 2020.

Personal Data Protection Comission. "Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems". 2024.

Republik Indonesia. Undang-Undang (UU) Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi. 2022.

Sandu, Iuliana, Menno Wiersma, Daphne Manichand. "Time to audit your AI algorithms, Maandblad voor Accountancy en Bedrijfseconomie". Amsterdam University Press. September 2022. 2022.

Schuett, Jonas."Frontier AI Developers Need an Internal Audit Function". Risk Analysis. 2024.

Surkov, et al. "Unleashing the power of machine learning models in banking through explainable artificial intelligence (XAI)". Articles. May 17, 2022. Deloitte. 2022.

Tramer, Florian, Fan Zhang, Ari Juels, Michael K. Reiter, Thomas Ristenpart. "Stealing Machine Learning Models via Prediction APIs". 25th USENIX Security Symposium. 2016.

The Government Accountability Office. "Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities". GAO. 2021.

The Institute of Internal Auditors (IIA). "The IIA's Artifical Intelligence Auditing Framework". 2024.

The Organisation for Economic Co-operation and Development. "Advancing Accountability in AI - Governing and Managing Risks Throughout the Lifecycle for Trustworthy AI". 2023.

The Organisation for Economic Co-operation and Development. "OECD Artificial Intelligence Papers No. 5, Common Guideposts to Promote Interoperability in AI Risk Management". November 2023. 2023.

The Organisation for Economic Co-operation and Development. "AI, Data Governance and Privacy - Synergies and Areas of International Co-operation". 2024.

The Organisation for Economic Co-operation and Development. "OECD AI Principles". 2024.

United Nations Educational Scientific Cultural Organization (UNESCO). "Recommendation on the Ethics of Artificial Intelligence". *Adopted on* 23 November 2021. 2022.

United Nations Educational Scientific Cultural Organization (UNESCO). "Key facts, UNESCO's Recommendation on the Ethics of Artificial Intelligence". *Adopted on* 23 November 2021. 2023.

United Nations Educational Scientific Cultural Organization (UNESCO). "Consultation Paper on AI Regulation - Emerging Approach Across the World". 16 August 2024. 2024.

United States Department of Homeland Security. "Increasing Threat of Deepfake Identities". The Analytic Exchange Program (AEP). 2021.

United States Department of the Treasury. "Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector". 2024.

United States Government - The White House Office of Science and Technology Policy. "The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People". October 2022. 2022.

Willcocks, L., Lacity, M., &Craig, A. "The IT Function and Robotic Process Automation". 2015.

World Bank Group. "Global Trends in AI Governance, Evolving Country Approaches". 2024.

World Economic Forum, Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission Singapore (PDPC). "Companion to the Model AI Governance Framework – Implementation and Self-Assessment Guide for Organizations". 2020.

World Economic Forum. "How Agentic AI will transform financial services with autonomy, efficiency and inclusion". 2024.

World Economic Forum. "Artificial Intelligence and Cybersecurity: Balancing Risks and Rewards". 2025.

Yasar, et al. "What is deepfake technology?". Articles. August 2024. TechTarget. 2024.

Zuccarelli, Eugenio. "Building trust in AI means moving beyond black-box algorithms. Here's why". Articles. Apr 2, 2024. World Economic Forum. 2024.

Artikel/Laman Web:
Articles/Websites:

Business Insider. "Winning Strategies for AI in Banking" tersedia pada https://www.businessinsider.com/intelligence/winning-strategies-for-ai-in-banking?utm_source=chatgpt.com (terakhir diakses pada tanggal 14 April 2025).

Creemers, Rogier, Graham Webster, Helen Toner. "Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022". 2022 tersedia pada https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/ (terakhir diakses pada tanggal 2 Februari 2025).

Fernandez, Miriam. "AI in Banking: AI Will Be An Incremental Game Changer". 31 Oktober 2023. S&P Global tersedia pada https://www.spglobal.com/en/research-insights/special-reports/ai-in-banking-ai-will-be-an-incremental-game-changer (terakhir diakses pada tanggal 14 April 2025).

Fortune Business Insight. "Artificial Intelligence Market Size and Future Outlook". 7 April 2025 tersedia pada https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114 (terakhir diakses pada tanggal 14 April 2025).

Government of India, Ministry of Electronics and Information Technology. "Artificial Intelligence Committees Reports". 2019 tersedia pada https://www.meity.gov.in/artificial-intelligence-committees-reports (terakhir diakses pada tanggal 2 Februari 2025).

Interim Measures for the Management of Generative Artificial Intelligence Services. 10 Juli 2023. Source of text: http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm translated by China Law Translate on 2023/07/13 tersedia pada https://www.chinalawtranslate.com/en/generative-ai-interim/ (terakhir diakses pada tanggal 2 Februari 2025).

Kamalnath, Vishnu, Larry Lerner, Jared Moon, Gökhan Sari, Vik Sohoni, and Shuo Zhang. "Capturing the full value of generative AI in banking". 5 Desember 2023. McKinsey's Financial Services Practice tersedia pada https://www.mckinsey.com/industries/financial-services/our-insights/capturing-the-full-value-of-generative-ai-in-banking (terakhir diakses pada tanggal 14 April 2025).

Kosinski, Matt. "What is the European Union Artificial Intelligence Act (EU AI Act)?". 20 September 2024 tersedia pada https://www.ibm.com/id-id/think/topics/eu-ai-act (terakhir diakses pada tanggal 31 Januari 2025).

Mitre Atlas. "ATLAS Matrix" tersedia pada https://atlas.mitre.org/matrices/ATLAS (terakhir diakses pada tanggal 14 April 2025).

OWASP®. "OWASP Machine Learning Security Top Ten" tersedia pada https://owasp.org/www-project-machine-learning-security-top-10/ (terakhir diakses pada tanggal 14 April 2025).

Personal Data Protection Commision Singapore. "Singapore's Approach to AI Governance" tersedia pada https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework (terakhir diakses pada tanggal 3 Februari 2025).

The Organisation for Economic Co-operation and Development. "Transparency and explainability (Principle 1.3)" tersedia pada https://oecd.ai/en/dashboards/ai-principles/P7 (terakhir diakses pada tanggal 5 Februari 2025).

Tim Penyusun
Editorial Team

Pengarah
Steering Committee

Dian Ediana Rae | Kepala Eksekutif Pengawas Perbankan merangkap Anggota Dewan Komisioner Otoritas Jasa Keuangan

Koordinator
Coordinators

Indah Iramadhini | Direktur Pengaturan Kelembagaan, Produk dan Aktivitas Perbankan selaku Plt. Kepala Departemen Pengaturan dan Pengembangan Perbankan

Mohamad Miftah | Direktur Pengembangan Perbankan

Tim Perumus
Drafting Team

M. Zulkifli Salim | Ihsan Ismady Putra | Ardyansah | Muhammad Radhi | Nurani Pertiwi Ekaputri | Norkolis Dwi Atmoko | Annisa Dwi Ramadhania Nasura

OTORITAS
JASA
KEUANGAN

DEPARTEMEN PENGATURAN DAN
PENGEMBANGAN PERBANKAN
DEPARTMENT OF BANKING
REGULATION AND DEVELOPMENT

MENARA RADIUS PRAWIRO
KOMPLEK PERKANTORAN BANK INDONESIA
JL. M.H. THAMRIN NO. 2
JAKARTA 10350