OECD publishing

G7 HIROSHIMA PROCESS ON GENERATIVE ARTIFICIAL INTELLIGENCE (AI)

TOWARDS A G7 COMMON UNDERSTANDING ON GENERATIVE AI

REPORT PREPARED FOR THE 2023 JAPANESE G7 PRESIDENCY AND THE G7 DIGITAL AND TECH WORKING GROUP

7 September 2023

OECD
BETTER POLICIES FOR BETTER LIVES

G7 leaders identified topics for discussion in the Hiroshima process and called for an early stock taking of opportunities and challenges related to generative AI. This report presents the results of a questionnaire developed to support the stocktaking to help guide G7 discussions on common policy priorities with regard to generative AI. It also provides a brief overview of the development of generative AI over time and across countries. The report and questionnaire results should be understood as representing a snapshot in time: they are indicative of trends identified in summer 2023 in a rapidly evolving area of technology. The report helped inform and structure discussions of the G7 Hiroshima AI Process.

This document was prepared by the Organisation for Economic Co-operation and Development (OECD) Directorate for Science Technology and Innovation (STI) for the 2023 Japanese G7 Presidency and the G7 Digital and Tech Working Group, to inform discussions during the G7 Hiroshima Artificial Intelligence Process and the related interim virtual Ministers' Meeting on generative artificial intelligence on 7 September 2023. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the member countries of the OECD or the G7.

This work is published under the responsibility of the Secretary-General of the OECD.

EXECUTIVE SUMMARY

Generative AI has rapidly entered public discourse.

Generative AI has entered into public consciousness and is increasingly present in peoples' everyday conversations worldwide. The number of news articles and tweets related to 'generative AI' grew eight-fold over just six months.

Growth in generative AI research, including its open-source code development, preceded the surge in investments.

The widespread awareness and rapid uptake of generative AI have been enabled by steady, incremental progress in both research and code development. Fundamental innovations such as the 'Transformers' architectures, contributions of the open-source community, alongside improvement in computing power have paved the way for the proliferation of large language models as well as other type of generative AI models. Scientific publications and open-source code development on generative AI have grown remarkably since 2017, and this trend accelerated in 2023. Venture capital investments in generative AI have skyrocketed and were estimated at USD 12 billion globally in the first half of 2023 alone. Scientific publications and software, including open-source code, related to generative AI have seen a parallel remarkable surge since 2017, with this trend further accelerating in 2023.

Rapid advances in generative AI are driven by its expected potential to drive productivity gains and to promote innovation and entrepreneurship, as well as to unlock solutions to global challenges.

In a questionnaire administered in Q3 2023, G7 members unanimously saw productivity gains, promoting innovation and entrepreneurship and unlocking solutions to global challenges, as some of the greatest opportunities of AI technologies worldwide, including for emerging and developing economies. G7 members also emphasised generative AI's potential role to help address pressing societal challenges, such as improving healthcare and helping to solve the climate crisis, and to support progress towards achieving the Sustainable Development Goals (SDGs).

Yet, generative AI's potential benefits come with risks.

The capacity of generative AI to exacerbate the challenges of disinformation and manipulation of opinions is considered by G7 members as one of the major threats stemming from generative AI, alongside risks of intellectual property rights infringement and privacy breaches. Early efforts to track AI incidents found one thousand distinct incidents and hazards related to generative AI, based on roughly 5 600 news articles dated from January to July 2023.

As these risks evolve rapidly, their management and mitigation is at the top of the agenda for G7 governments.

Responsible use of generative AI, addressing disinformation, safeguarding intellectual property rights, and governing generative AI are among the top priorities for G7 policymakers and require international cooperation with like-minded partners. Other urgent and important issues emphasised by G7 members include privacy and data governance, transparency, fairness and bias, human and fundamental rights, security and robustness of AI systems, and impacts on the functioning of democracy.

G7 jurisdictions are evaluating their respective responses to generative AI, as well as the policy gaps.

Countries are leveraging existing and forthcoming legal and policy frameworks and developing guidelines or regulation to address risks related to generative AI. National initiatives are also being strengthened to seize its opportunities. New issues raised by generative AI appear to affect specific sectors in particular, such as education, media, and the workplace.

G7 members are aligned on the need to provide effective tools for safety, quality control, and capacity and trust building for generative AI.

Safety, quality control, capacity and trust building for generative AI were seen as among the most urgent and important international action the G7 could undertake. Engaging in dialogue was also considered to be most *urgent*, and developing voluntary codes of conduct was identified as among the most *important* actions.

Table of contents

FIGURES

TABLES

INTRODUCTION

In April 2023, Japan hosted the G7 Digital and Technology Ministers' Meeting in Takasaki. The Ministers agreed on the Ministerial Declaration, which emphasises the importance of international discussions on the interoperability between different AI governance frameworks and stock-taking of the opportunities and challenges brought by generative AI.

In their Declaration of April 2023, the G7 Digital and Technology Ministers recognised *"the need to take stock in the near term of the opportunities and challenges of [generative AI technologies] and to continue promoting safety and trust as these technologies develop" and therefore undertook "to convene future G7 discussions on generative AI which could include topics such as governance, how to safeguard intellectual property rights including copyright, promote transparency, address disinformation, including foreign information manipulation, and how to responsibly utilise these technologies."* (Paragraph 47).

The Ministerial discussion on AI was escalated to the Leaders' discussion at the G7 Summit meeting in May, hosted in Hiroshima. The Leaders agreed to task their Ministers to establish the "Hiroshima AI process", where G7 members continue the discussion on generative AI in an inclusive manner.

Generative AI can be understood as a form of AI model specifically intended to produce new digital material as an output (including text, images, audio, video, software code), including when such AI models are used in applications and their user interfaces. These are typically constructed as machine learning systems that have been trained on massive amounts of data. They work by predicting words, pixels, waveforms, data points, etc. that would resemble the models' training data, often in response to prompts (OECD, 2023[1]), (OECD, Forthcoming[2]).

To support the G7 Hiroshima AI Process launched by G7 Leaders, Japan circulated a questionnaire in June 2023 to G7 members. The questionnaire aimed at taking stock of G7 members' existing and planned policy initiatives and considerations on the main opportunities and risks associated with generative AI. It was organised in the following four inter-related sections:

1. Scoping opportunities and risks
2. Priorities in terms of values-based principles
3. Potential collective international approach
4. National and regional initiatives

The questionnaire was composed mostly of closed questions, which provided respondents with selected options to rank or to choose from, i.e. a list of opportunities or risks related to generative AI, priorities among the five areas identified in the G7 Leaders' Statement and among the OECD AI Principles, and a selection of possible policy actions that G7 members could recommend. Open questions allowed respondents to report on national or regional initiatives in G7 jurisdictions pertaining to generative AI.

This report presents advancements in generative AI based on data from the OECD.AI Policy Observatory and the OECD AI Incident Monitor (section 1), and analyses responses to the questionnaire circulated to G7 members (section 2).

1. SETTING THE SCENE

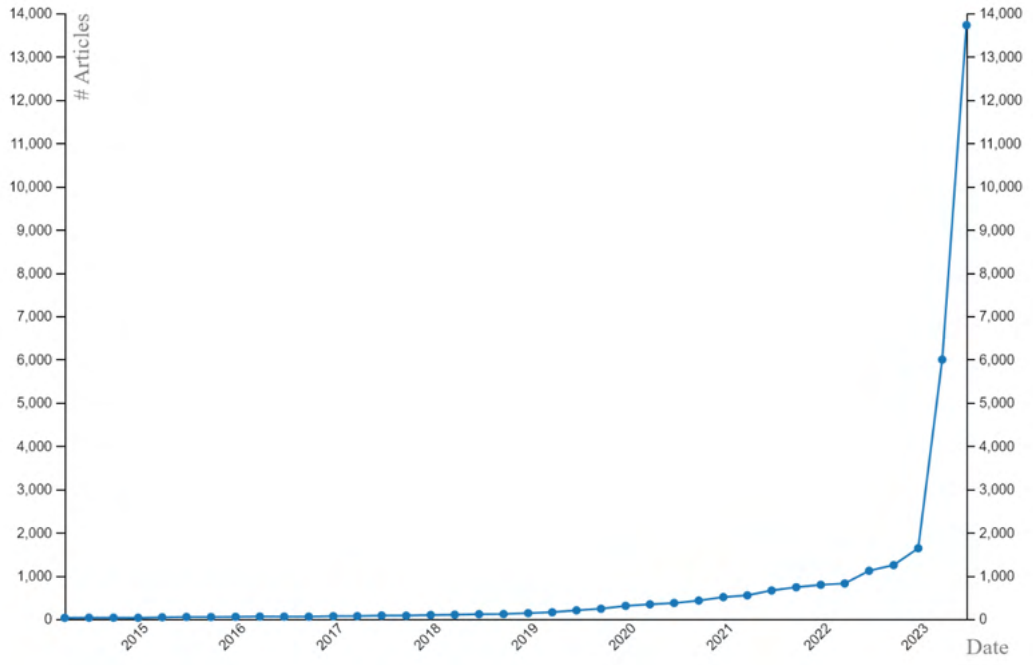
1.1. GENERATIVE AI TRENDS

Generative AI has taken centre stage in the public, academic, and political discussions surrounding AI. It is predicted to create significant economic value. Companies have begun to adopt the technology to create new business opportunities, and start-ups are competing for venture capital.

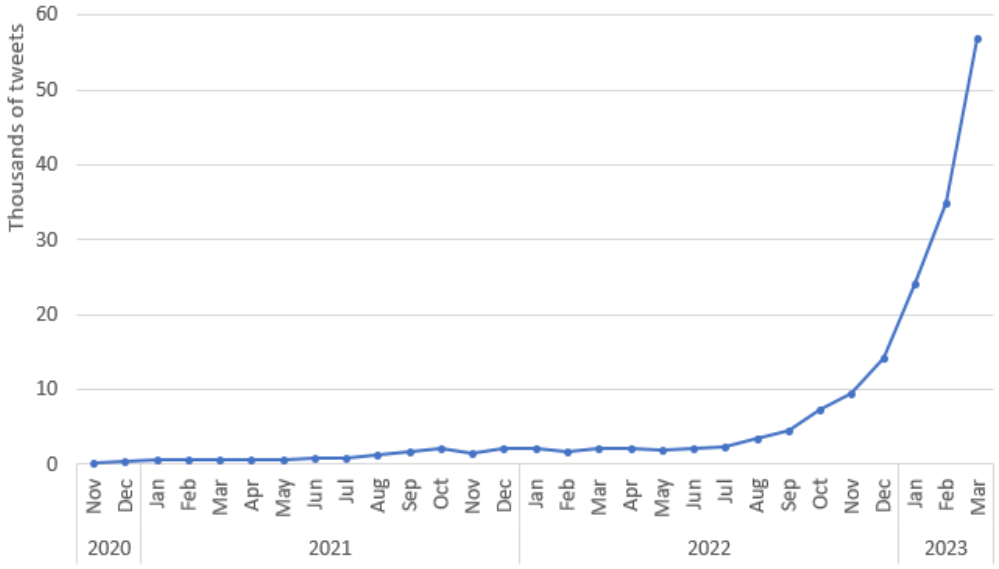
Generative AI has entered into public awareness and is increasingly present in everyday conversations worldwide, as evidenced by the surge in related news articles and tweets. Both indicators show an eight-fold increase in a mere six-month period, with new articles on generative AI increasing from 1.6 thousand in the last quarter of 2022 to almost 14 thousand in the second quarter of 2023, and tweets about generative AI reaching 57 thousand in March 2023, up from an initial 7 thousand in October 2022 (FIGURE 1.1, panels a and b, respectively).

FIGURE 1.1. GENERATIVE AI HAS RAPIDLY ENTERED PUBLIC DISCOURSE

a) Number of news articles globally on generative AI and related topics



b) Number of tweets globally on generative AI and related topics

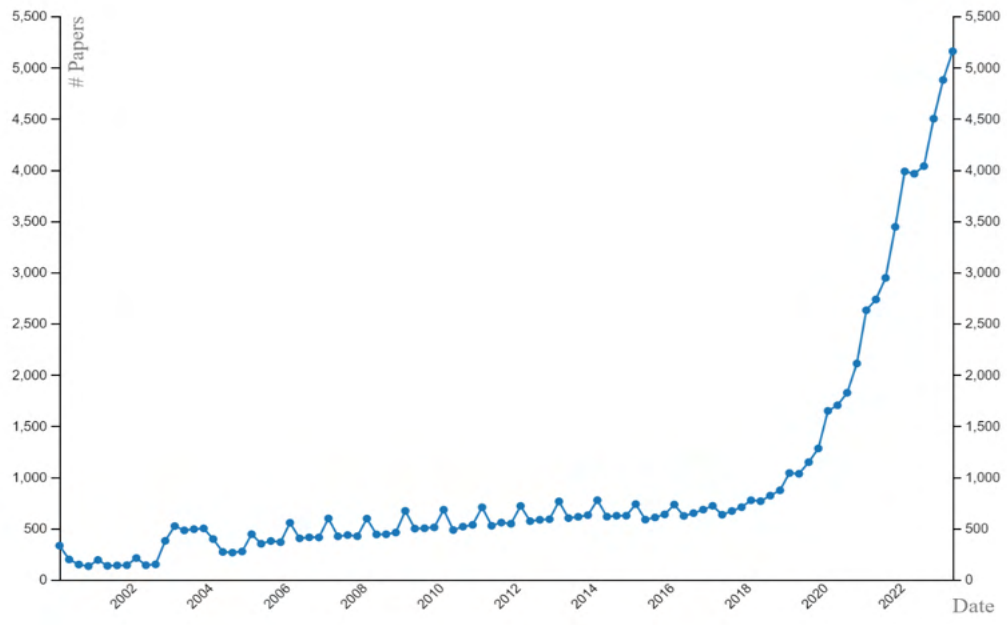


Note: News articles related to generative AI are a subset of AI-related articles that include the following wikidata concepts and all of their descendants: generative AI, transformer models, language model, and generative model. Generative AI tweets contain "generative AI" or #generativeAI. News articles data has been adjusted using a technique called quarterly smoothing to make it easier to compare and understand the trends over time.
Source: OECD.AI, using data from Event Registry for news articles and from X for tweets.

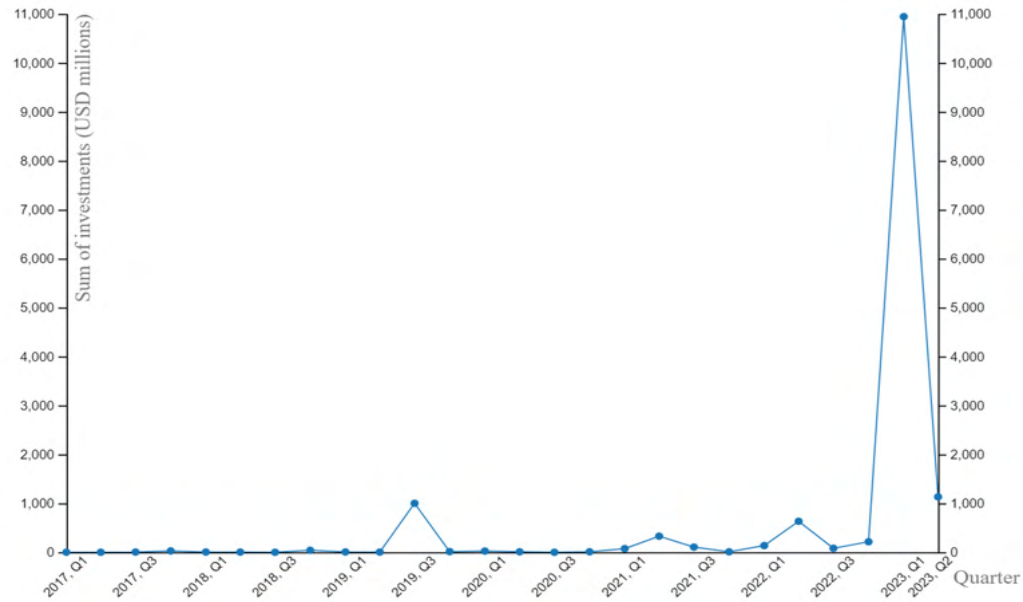
Research and venture capital investments into generative AI development have also seen substantial increase. Scientific publications pertaining to generative AI have grown fivefold since 2019, which can be attributed to heightened interest in fundamental innovations such as transformer models and advancements in computing power, and which paved the way for the proliferation of large language models (Figure 1.2, panel a). Venture capital investments in generative AI in the first half of 2023 reached a total of USD 12 billion globally (Figure 1.2, panel b). Peaks in venture capital investments in 2019 and 2023 reflect Microsoft's USD 1 billion and USD 10 billion investments in OpenAI, respectively.

FIGURE 1.2. THE GROWTH OF GENERATIVE AI RESEARCH PRECEDES THE SURGE IN INVESTMENTS

a) Number of scientific publications globally on generative AI and related topics



b) Sum of global venture capital investments on generative AI startups

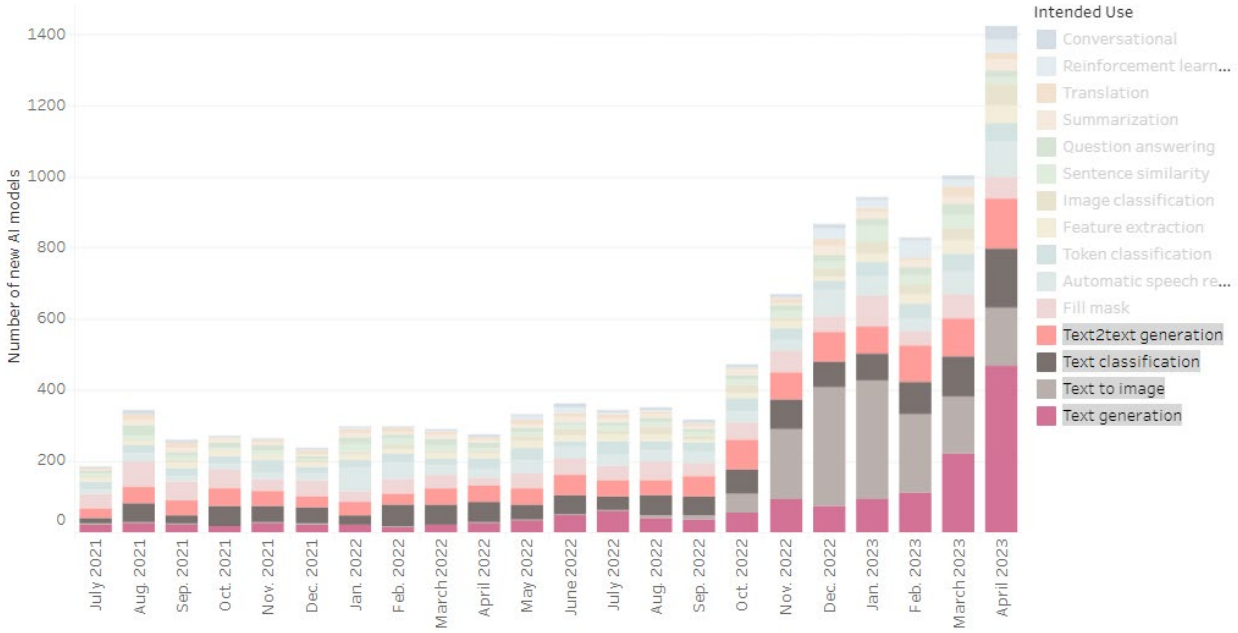


Note: Scientific publications related to generative AI are a subset of AI-related publications that include the following wikidata concepts and all of their descendants: generative AI, transformer models, language model, and generative model. VC investments related to generative AI capture startups that include concepts like generative AI, generative adversarial network, text generation, image generation, audio generation, and generative model in their company descriptions. Quarterly data smoothing is applied to both datasets to remove noise.
Source: OECD.AI, using data from OpenAlex for research publications and from Preqin for venture capital investments.

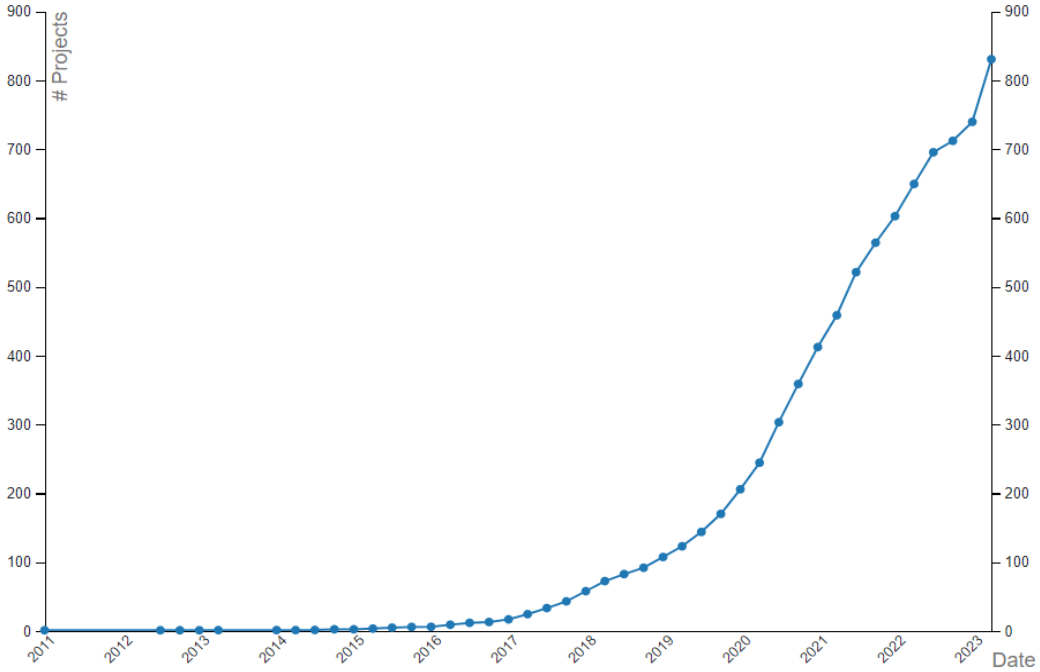
The open-source community has traditionally been a driving force behind AI advancements. This trend appears to continue in the context of generative AI. Since October 2022, there has been a significant increase in the availability and development of open-source AI models dedicated to generative AI systems, as shown by the rising number of text generation models uploaded to the Hugging Face repository in the recent months (Figure 1.3, panel a). In contrast, the upswing in open-source code development pertaining to generative AI on GitHub starts in 2017 and follows a more gradual trajectory. This nuanced pattern suggests that the progress leading to generative AI has been the result of steady and incremental advancements in code development (Figure 1.3, panel b).

FIGURE 1.3. GENERATIVE AI OPEN-SOURCE CODE HAS SEEN MORE GRADUAL GROWTH THAN OPEN-SOURCE MODELS

a) Upsurge in the number of open-source generative AI models on Hugging Face



b) Growth in the number of open-source generative AI code development projects on GitHub



Note: Open-source code related to generative AI are a subset of AI-related GitHub repositories that include the following wikidata concepts and all of their descendants: generative AI, transformer models, language model, and generative model. Quarterly data smoothing is applied to GitHub data to remove noise.
Source: OECD.AI, using data from Hugging Face for open-source models and from GitHub for open-source code.

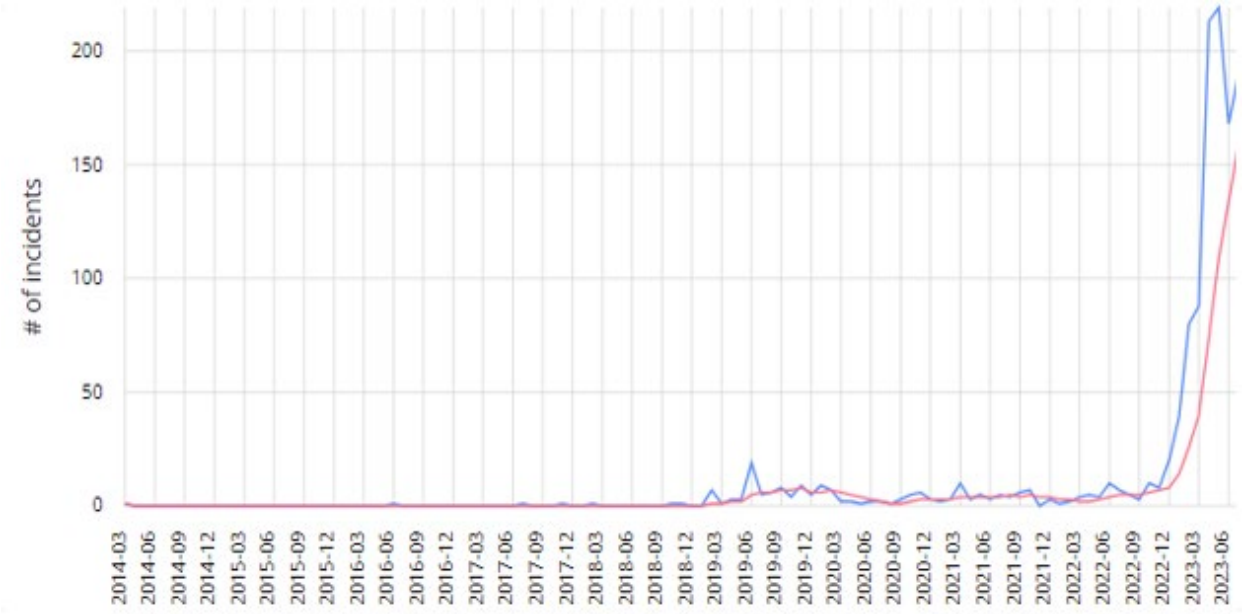
1.2. INCIDENTS AND HAZARDS RELATED TO GENERATIVE AI

2. While generative AI has the potential to revolutionise industry and society in positive ways, the use of the technology also poses risks to individuals and societies. For example, generative AI can be exploited for malicious purposes, leading to serious negative consequences such as the propagation of disinformation and the creation of manipulated content like deepfakes. The recognition of this multi-purpose nature of AI technologies and how they can be deployed – including generative AI – has prompted the OECD to develop a global AI Incidents Monitor, designed to furnish real-time evidence on AI risks to inform policy decisions. This is achieved through the scrutiny of real-world incidents and hazards in real time as reported by reputable news outlets.

The term "incident" encompasses a collection of one or more news articles covering the same event.1 Over the period from January to July 2023, approximately one thousand incidents and hazards related to generative AI were reported across roughly 5 600 news articles (Figure 1.4). While the Monitor is still under development (its release is expected in November 2023), the initial findings shed some light on the potential risks posed by generative AI systems and can help contribute to shaping a safer AI landscape for the future.

1 The OECD.AI expert group on AI incidents is currently discussing a working definition of AI incidents and hazards. For the purposes of this report, the term “incidents” is used as umbrella term to include also “hazards”.

FIGURE 1.4. GENERATIVE AI-RELATED INCIDENTS AND HAZARDS REPORTED BY REPUTABLE NEWS OUTLETS HAVE GROWN EXPONENTIALLY



Note: The blue line shows the real count of incidents and hazards reported each month. The red line displays the same data but adjusted using quarterly smoothing. The peak in 2019 relates to a surge in the reporting of incidents and hazards related to deepfake technology.
Source: OECD.AI, AI Incidents Monitor (forthcoming), using data from Event Registry.

2. GENERATIVE AI FROM A G7 PERSPECTIVE

This section presents the results of the questionnaire developed to support a stocktaking to help guide G7 discussions on common policy priorities with regard to generative AI. As most questions provided a list of options to rank or choose from, the rankings shown do not suggest e.g., most important to least important priorities, but a snapshot of country responses on the top priorities outlined in the questionnaire at a given time.

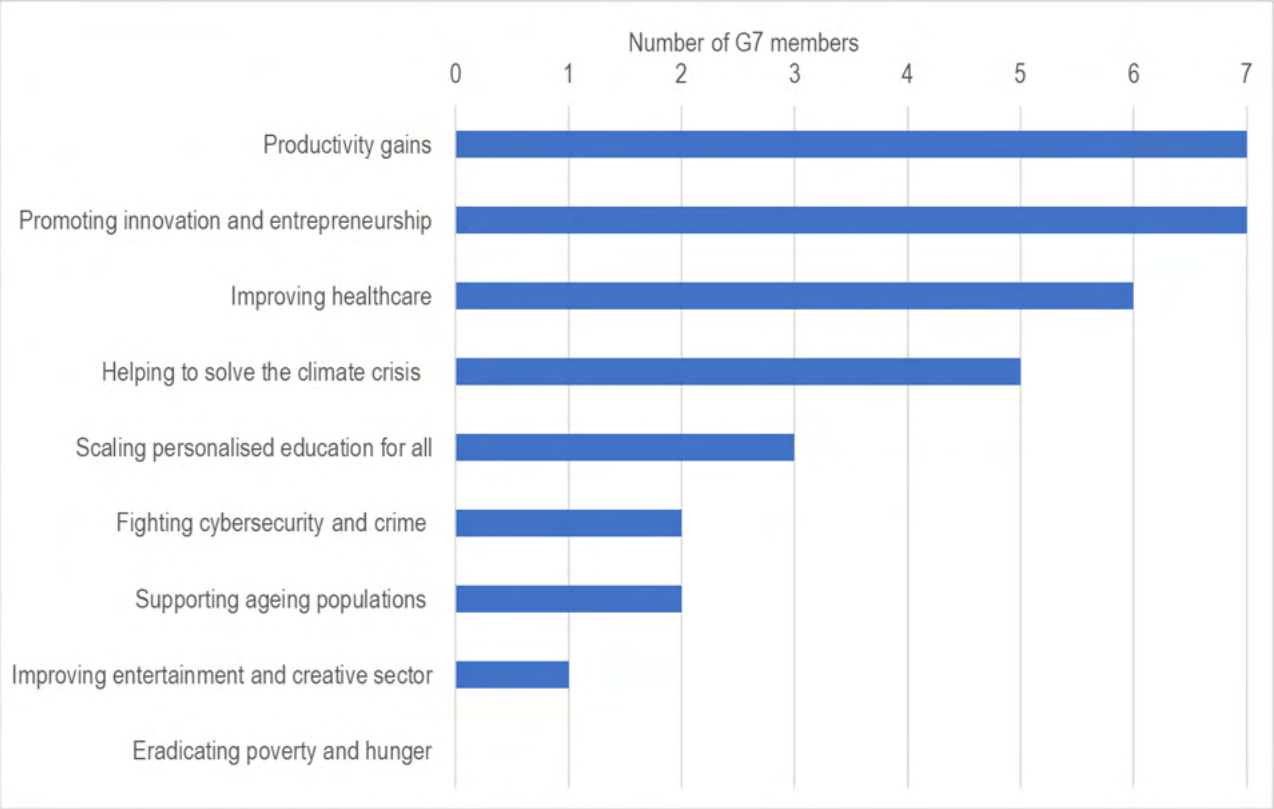
2.1. OPPORTUNITIES AND RISKS FOR G7 MEMBERS

Productivity gains and promoting innovation and entrepreneurship were viewed by all respondents as among the major opportunities made possible by generative AI, among opportunities outlined in the questionnaire. Improving healthcare followed closely, as did helping to solve the climate crisis (FIGURE 2.1).

Strengthening the traceability and the transparency of democratic processes and improving citizens' access to public services were also mentioned as other opportunities.

FIGURE 2.1. TOP FIVE *OPPORTUNITIES* OF GENERATIVE AI TO HELP ACHIEVE NATIONAL AND REGIONAL GOALS

Number of G7 members that selected (five) specific opportunities from a pre-populated drop-down list



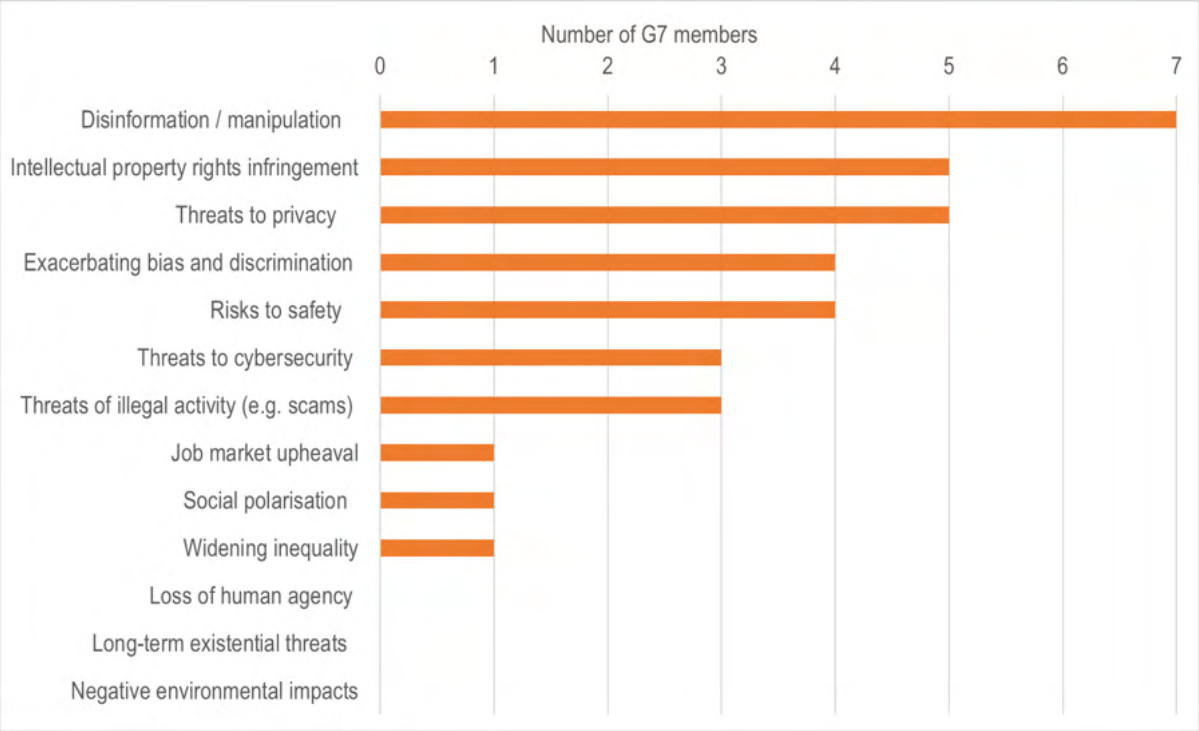
Note: The figure aggregates responses from seven respondents to the question: "*From your country or region's perspective, what are the top five opportunities generative AI presents to help achieve national and regional goals? (Please select five options)*".

Disinformation and the associated manipulation of opinions were viewed by all respondents as the dominant risk posed by generative AI, among risks outlined in the questionnaire. Most G7 members also considered intellectual property right infringement as well as threats to privacy as major risks (Figure 2.2).

Threats to security (including cybersecurity); manipulation and improper use of data; and threats to human rights were also highlighted as additional risks.

FIGURE 2.2. TOP FIVE *RISKS* PRESENTED BY GENERATIVE AI IN ACHIEVING NATIONAL AND REGIONAL GOALS

Number of G7 members that selected (five) specific risks from a pre-populated drop-down list



Note: The figure aggregates responses from seven respondents to the question: "*From your country or region's perspective, what are the top five risks generative AI presents to achieving national and regional goals? (Please select five options)*".

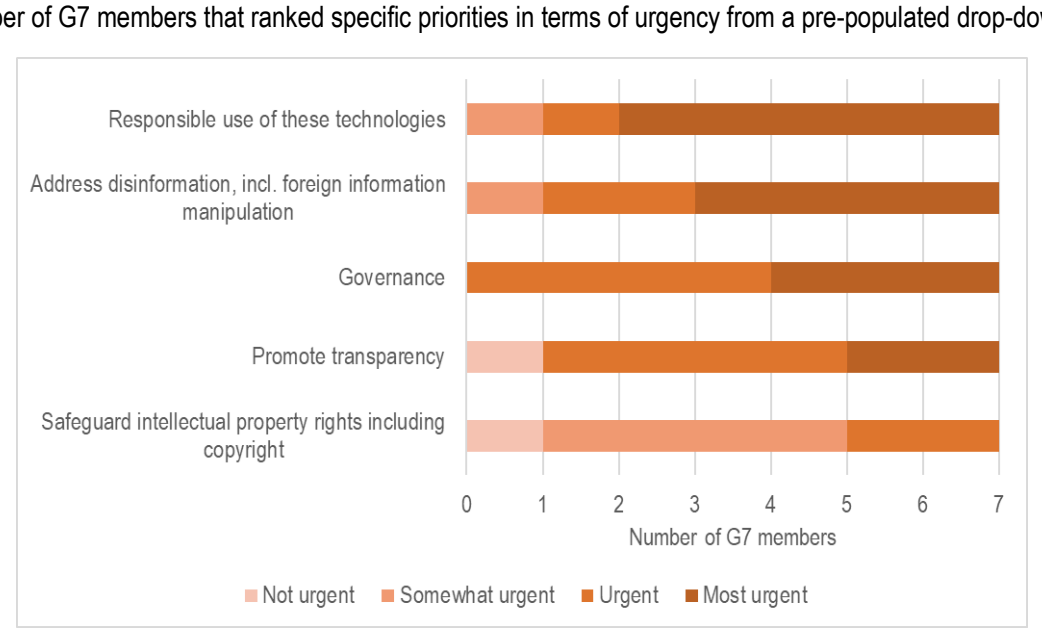
2.2. PRIORITIES IN TERMS OF VALUES-BASED PRINCIPLES FOR G7 MEMBERS

MOST *URGENT* AND *IMPORTANT* PRIORITIES REGARDING GENERATIVE AI FOR G7 MEMBERS

The 'responsible' use of generative AI technologies was widely viewed as the most "urgent" priority for policy among the priorities highlighted in the G7 statement. This was followed closely by addressing disinformation and by governing generative AI appropriately (Figure 2.3).

Threats to cybersecurity and biosecurity were also indicated among the most urgent priorities regarding generative AI.

FIGURE 2.3. MOST *URGENT* PRIORITIES REGARDING GENERATIVE AI AMONG THE FIVE PRIORITIES HIGHLIGHTED IN THE G7 LEADER'S STATEMENT



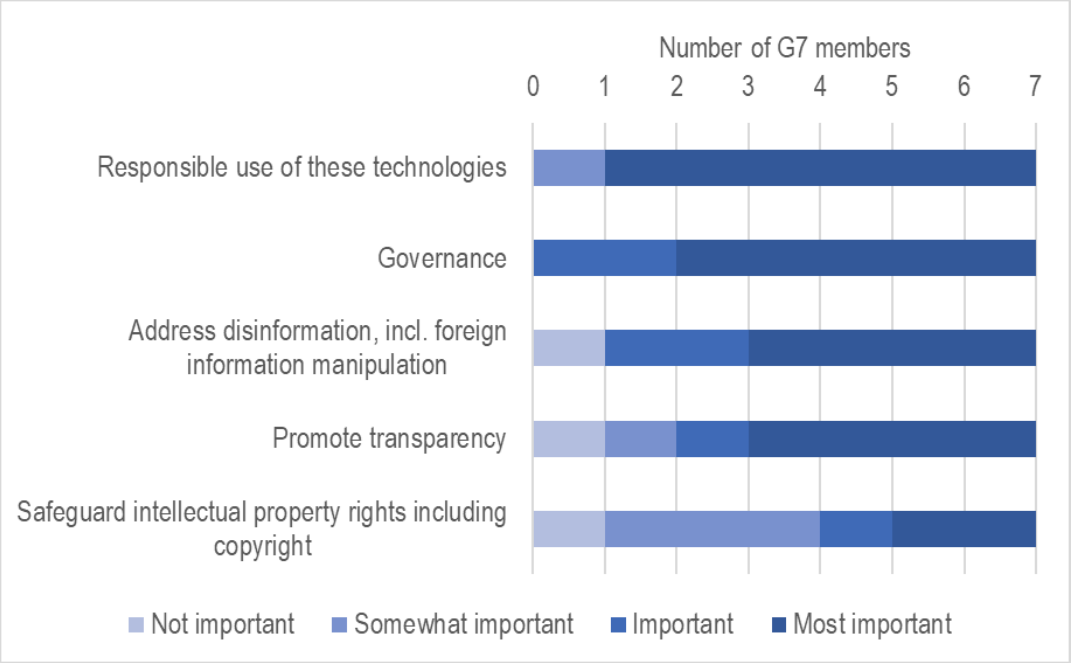
Note: The figure aggregates responses from seven respondents to the question: "*From a policy perspective what do you see as the most urgent and the most important priorities regarding generative AI? (Please rank the concepts below by order of urgency and importance. Different concepts can have the same priority level)*".

The 'responsible' use of generative AI technologies was also viewed as the most "important" priority for policy, followed by governance and by addressing disinformation. While "importance" and "urgency" of issues were ranked slightly differently, they highlighted the same overall priorities (Figure 2.4).

The threat to cybersecurity was also indicated as an additional important priority, and the threat to biosecurity as one of the most important priorities in the field of generative AI.

FIGURE 2.4. MOST *IMPORTANT* PRIORITIES REGARDING GENERATIVE AI AMONG THE FIVE PRIORITIES HIGHLIGHTED IN THE G7 LEADER'S STATEMENT

Number of G7 members that ranked specific priorities in terms of importance from a pre-populated drop-down list

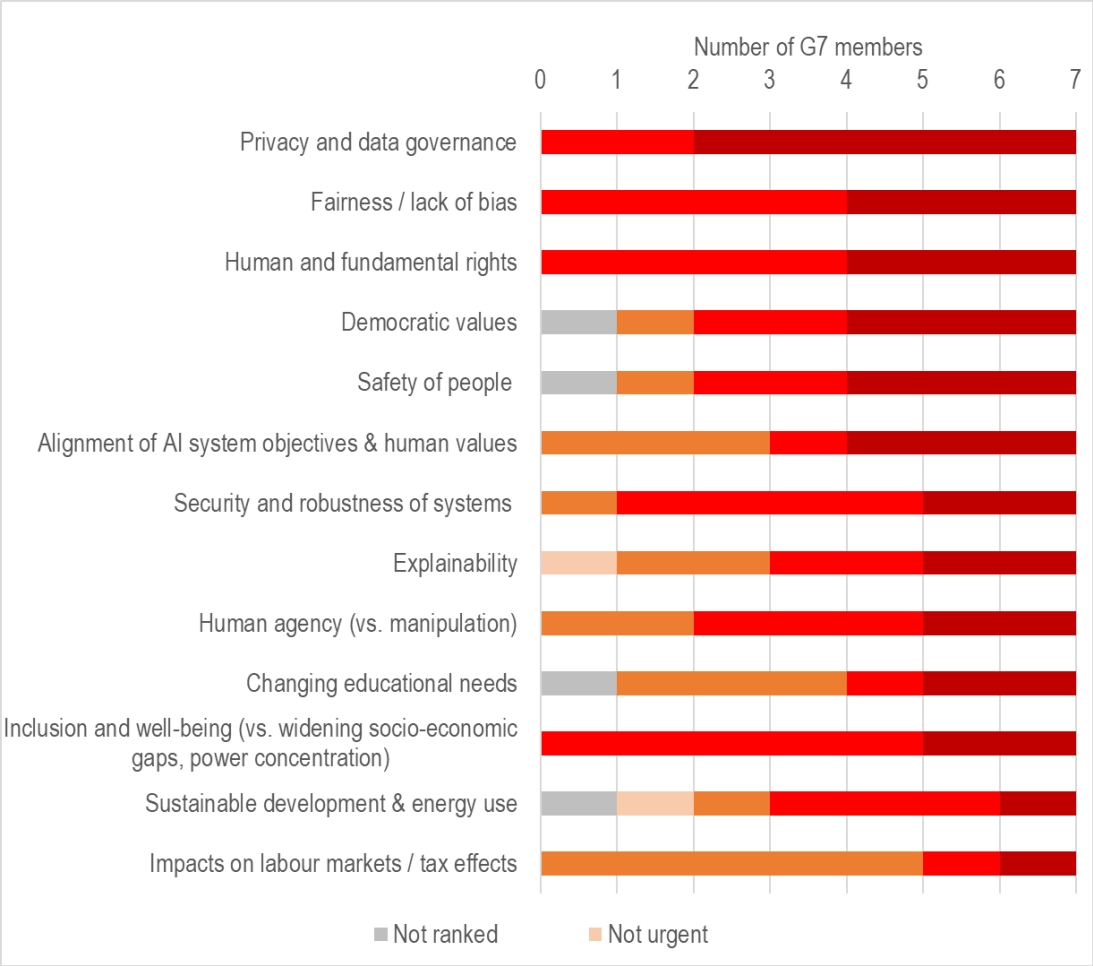


Note: The figure aggregates response from seven respondents to the question: "*From a policy perspective what do you see as the most urgent and the most important priorities regarding generative AI? (Please rank the concepts below by order of urgency and importance. Different concepts can have the same priority level)*".

In addition, privacy and data governance were prioritised as the most urgent 'other' issues to address outlined in the questionnaire, followed by fairness and bias, and human and fundamental rights (Figure 2.5).

FIGURE 2.5. MOST *URGENT* PRIORITIES REGARDING GENERATIVE AI: OTHER PRIORITIES HIGHLIGHTED IN THE OECD AI PRINCIPLES

Number of G7 members that ranked specific priorities in terms of urgency from a pre-populated drop-down list

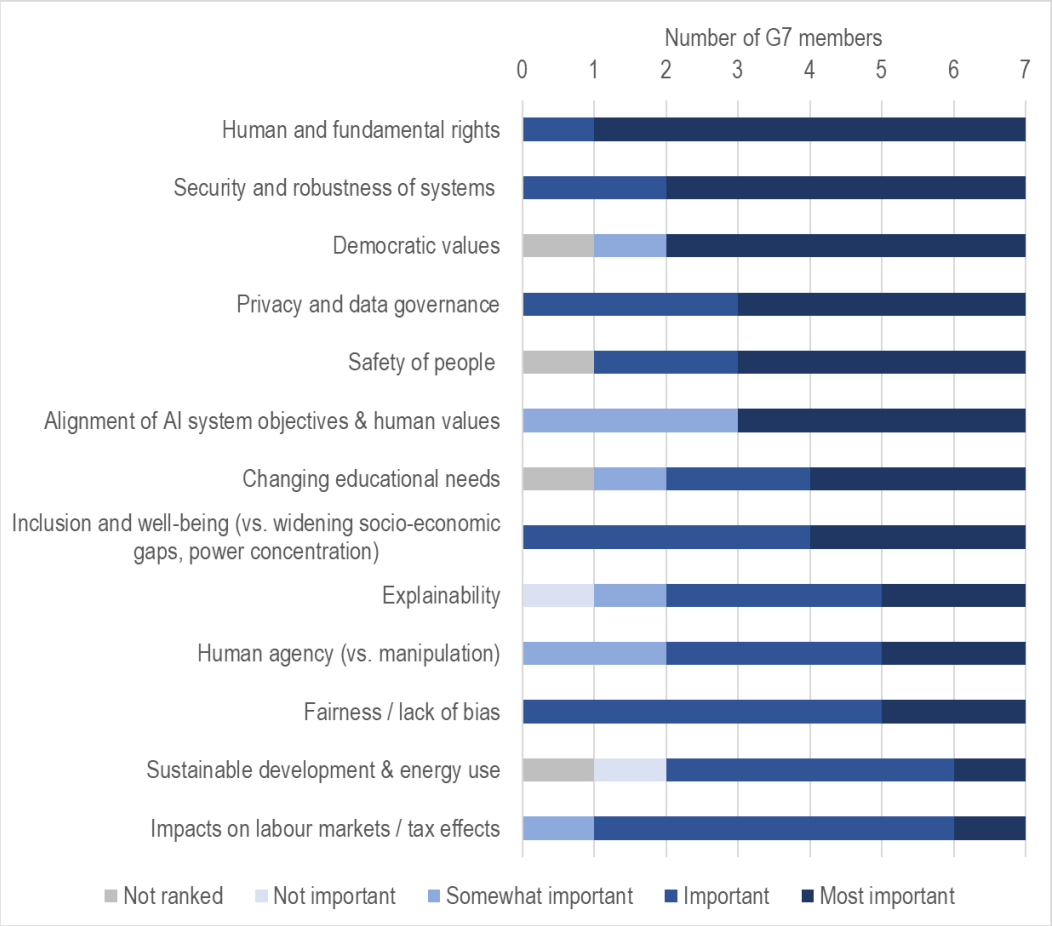


Note: The figure aggregates responses from seven respondents to the question: "*From a policy perspective what do you see as the most urgent and the most important priorities regarding generative AI? (Please rank the concepts below by order of urgency and importance. Different concepts can have the same priority level)*".

Human and fundamental rights, security and robustness of AI systems, democratic values, and privacy and data governance were viewed as the most important 'other' priorities outlined in the questionnaire (Figure 2.6).

FIGURE 2.6. MOST *IMPORTANT* PRIORITIES REGARDING GENERATIVE AI: OTHER PRIORITIES HIGHLIGHTED IN THE OECD AI PRINCIPLES

Number of G7 members that ranked specific priorities in terms of importance from a pre-populated drop-down list

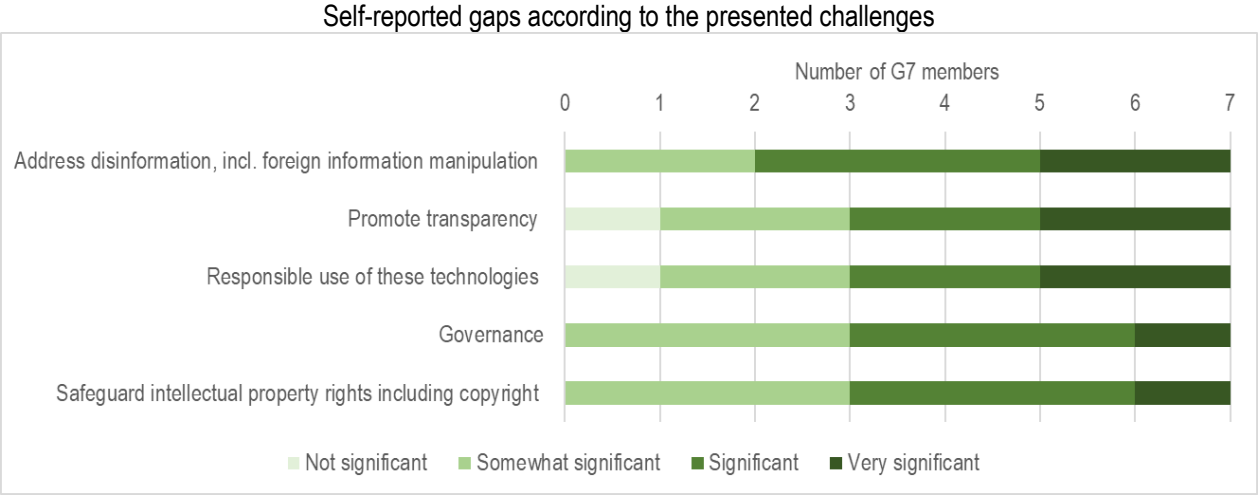


Note: The figure aggregates responses from seven respondents to the question: "*From a policy perspective what do you see as the most urgent and the most important priorities regarding generative AI? (Please rank the concepts below by order of urgency and importance. Different concepts can have the same priority level)*".

MOST SIGNIFICANT GAPS IN EXISTING POLICIES OR POLICIES UNDERWAY TO ADDRESS THE CHALLENGES OF GENERATIVE AI IN G7 JURISDICTIONS

Different G7 members' perceived policy gaps were diverse, with different G7 members highlighting disinformation, transparency, and responsible use as their most significant gaps among those outlined in the questionnaire (Figure 2.7).

FIGURE 2.7. GAPS IN G7 MEMBERS' EXISTING OR UNDERWAY POLICIES TO SUCCESSFULLY ADDRESS CHALLENGES OF GENERATIVE AI



Note: The figure aggregates responses from seven respondents to the question: "*How significant are the gaps you see in your policies (existing or underway) to successfully address the following challenges of generative AI?*".

MEASURES THAT G7 MEMBERS ARE TAKING TO FILL THE IDENTIFIED POLICY GAPS

Applying existing (or forthcoming) legal frameworks, and evaluating policy gaps

Respondent countries are leveraging existing as well as forthcoming legal frameworks. They are also assessing the scale and extent of gaps pertaining to generative AI challenges.

- **Canada**, under the proposed Artificial Intelligence and Data Act (AIDA), sets out a risk-based regulatory framework for the responsible design, development, and use of AI systems in the private sector, including generative AI systems.
- **France** is analysing the new challenges posed by generative AI regarding existing legislation (such as GDPR and the EU Copyright Directive) and noted that the EU AI Act is expected to address some of the issues, such as governance and responsible use of AI systems.
- **Germany** is working on closer strategic alignment across government entities and on developing relevant procedures.
- **Italy** In the light of the advent of generative AI (e.g. ChatGPT), provisional agreement of the European Parliament on the AI Act and increased ethical concerns and risks regarding Generative AI, in early July 2023, the Department for Digital Transformation - Presidency of the Council of Ministers, presented a proposal for revision of the "Strategic Plan for Artificial Intelligence 2021". The proposed revision is nearing completion and will be subject to a public consultation procedure starting as of 30 September. The entry into force of the new "Strategic Plan for AI" is scheduled for December 31, 2023. Also, the government established a Permanent Committee on AI within the Inter-Ministerial Committee on Digital Transition; the committee includes experts from universities, research centres and the associations of Italian companies. In March 2023, the Italian Data Protection Authority - DPA imposed an immediate temporary limitation on the processing of Italian users' data by OpenAI. In its order, DPA highlighted that no information was provided to users and data subjects whose data were collected by Open AI; more importantly, it underlined that there was no legal basis underpinning the massive collection and processing of personal data in order to 'train' the algorithms on which the platform relies. Furthermore, there was no age verification for minors.

- **Japan** indicated that it is applying existing legal frameworks (e.g., the Penal Code and the enforcement system) and existing guidelines (e.g., the AI R&D Guidelines, the AI Utilization Guidelines, or the Governance Guidelines for Implementation of AI Principles), although these might be challenged by new issues from generative AI. As new issues arise with the emergence of generative AI, it considers integrating and revising the guidelines into a unified and easy to understand guideline for developers, providers, users, and other business.
- The **United Kingdom** is working across government to assess scale and extent of gaps in existing mitigation measures and is exploring further measures at every stage of the AI supply chain**.**
- Similarly, in the **United States**, pursuant to an existing executive order mandating certain principles for federal AI activities, the Office of Management and Budget is developing guidance that will establish specific policies that federal departments and agencies must follow to strengthen AI governance, advance AI procurement, and manage algorithmic risk to safeguard American people's rights and safety. The U.S. Department of Commerce's National Institute of Standards and Technology released an AI Risk Management Framework (AI RMF) in January 2023 and in June 2023 launched a Generative AI Public Working Group to develop a profile of AI RMF for generative AI systems.
- The proposed **European Union's Artificial Intelligence Act (EU AI Act)** aims at promoting the development and uptake of AI while addressing potential risks certain AI systems, including generative AI, can pose to safety and fundamental rights. In addition, the EU will further enhance its regulatory toolbox, with the revision of existing legislation, like for example the recently adopted Machinery Regulation, as well as the proposed (and currently in the legislative process) revision of the Product Liability Directive and the proposal for the Cyber Resilience Act. In addition to regulatory measures, the EU is also pursuing the EU Coordinated Plan on AI with Member States and working on the AI Pact. The Pact would encourage companies to voluntarily communicate the processes and practices they are putting in place to prepare for compliance with the EU AI Act and ensure that the design, development and use of AI is trustworthy.

Developing guidelines, and establishing new guidance as well as governance bodies

The Treasury Board of **Canada**'s Secretariat (TBS) plans to issue guidelines on the use of generative AI in the federal government. This 'guide' will provide federal institutions with guidance on the use of these tools. It provides an overview of generative AI, identifies challenges and concerns relating to its use, puts forward principles for using it responsibly, and offers policy considerations and best practices. Additionally, in early August, Canada launched roundtable sessions to seek stakeholder feedback on a proposed Canadian code of practice for generative AI. The code will provide voluntary guidance to companies developing and using generative AI systems, and it will help them to prepare their processes and products before formal regulation takes effect.

Similarly, the **United States** Office of Management and Budget is developing guidance that will establish specific policies that federal departments and agencies must follow to strengthen AI governance, advance AI procurement, and manage algorithmic risk to safeguard American people's rights and safety.

Germany reported undertaking several specific actions including a) setting up an advisory centre for the use of AI in the public sector to address competence building and networking; and b) establishing an AI Quality and Innovation Centre.

The **European Union** in the EU AI Act would also envisage certain governance structure and specific policies to ensure that design, development and use of AI technologies, including generative AI, is trustworthy.

Highlighting the need for international governance

Where current gaps cannot be filled by resorting to current legal systems, G7 members refer to the need to explore further mitigation measures, to take inspiration from other countries, and to coordinate measures at the international level to develop comprehensive and consistent approaches among countries. A G7 member referred to international mechanisms that seek to counter foreign disinformation, such as the OECD Mis/Dis Information Hub, the G7 Rapid Response Mechanism, and the Summit for Democracy (S4D) Information Integrity cohort.

EXAMPLES OF AREAS AND SECTORS IN WHICH EXISTING LAWS AND POLICIES ARE BEING APPLIED TO, OR ARE BEING CHALLENGED BY, NEW ISSUES RAISED BY GENERATIVE AI

Existing laws and policies, including consumer protection and privacy laws, apply to generative AI. Several G7 members reported that legislation in various sectors is applicable to, but also challenged by, generative AI. G7 members highlighted in particular the challenges posed to privacy, and intellectual property, including copyright. Some jurisdictions noted that compliance is already required by existing data protection legislation or intellectual property regimes. Others highlighted that AI has raised new legal questions about ownership of content created wholly or in part with AI, such as images and texts, as well as questions about how rights associated with training data affect the legal status of models' output, and that these questions are being investigated at national level.

G7 members also provided examples of sectors particularly challenged by generative AI. These include creative industries, knowledge work, law, cybersecurity, health and medical devices technologies, the financial sector, and federal public services.

Issues raised by generative AI seem to be particularly pressing for the following sectors, given that these were highlighted by several G7 members:

- **Education**: Several G7 members noted that the education sector requires particular attention as it is already highly affected by generative AI, and that this is expected to increase in the future. A G7 member is convening experts to work with the education sector to share and identify best practice cases as well as opportunities.

- **Workplace:** Two G7 members highlighted that generative AI affects workplace related matters. Generative AI can speed up recruitment processes (e.g. through generative AI-powered chatbots), but if data is biased, this may negatively impact the fairness of recruitment processes. Furthermore, they noted that in the workplace, employees may use generative AI without appropriate guidance or regulation and expose sensitive or confidential corporate data and/or personal information to third parties outside the company.

- **Communication and media:** Two G7 members noted that journalism and information other areas in which laws and policies are being applied to or are challenged by generative AI, stressing to the risk for generative AI to create misinformation and deepfakes.

SPECIFIC CHARACTERISTICS OF GENERATIVE AI THAT CHALLENGE REGULATION / GOVERNANCE

Unpredictability, adaptivity, autonomy, and multi-purpose nature of generative AI

Several G7 members indicated unpredictability of impact, adaptivity, autonomy, and a multi-purpose nature as characteristics of generative AI that challenge its regulation and governance.

- Generative AI creates new **disruptive innovation**, impacting a broad range of inexperienced users and developers. The technology's wide field of application, increasing interactions between generative AI systems, and rapid technical developments cause high uncertainty and unpredictability for society.
- Many generative AI applications today have little **'autonomy'**, i.e., little capacity to make decisions or take actions on their own, without human oversight or direction. They often produce content when instructed based on prompts. However, the output of generative AI can lead to autonomous or partly autonomous action programmed in traditional automation software whereby the output of the generative AI is acted upon. Looking to the future however, as generative AI like large language models are increasingly used as autonomous generative 'agents' with plugins to connect them to third-party applications, they are becoming more autonomous. For example, plugins enable language models to operate on recent data, including real-time information, such as stock prices or news articles, and to assist users in new ways, such as through autonomous ordering and booking.
- The **'adaptivity'** of generative AI comes with difficulties for developers to understand the intent or logic that leads to systems' outcomes. This is because AI systems are 'trained' by inferring patterns and connections in data which are often not easily discernible to human programmers. This mechanism allows generative AI systems to develop the ability to perform new tasks or forms of reasoning that the developers did not expect - a powerful source of capabilities, but also a barrier to intentionally designing or even fully understanding model capabilities. This gap between the developers' knowledge and intent and the system's capabilities can complicate assigning responsibility for outcomes.
- The adaptivity and multi-purpose nature of generative AI, and its ability to develop and push out content much more rapidly than has been the case before, **may exacerbate bias and other risks** in more contexts than previous systems, promoting or reinforcing stereotypical or harmful representations.
- The diversity of use cases and contexts to which generative AI can be applied is also challenging in its own right, as each of use case and context can potentially have different regulatory or governance requirements.
- Opacity, complexity, and continuous adaptation of AI systems can generate or exacerbate existing risks to health, safety, and fundamental rights.

Lack of transparency

Lack of transparency of generative AI, both in the development stage (i.e. developers being transparent about how they developed the system) and use (i.e. users being transparent about the fact that they are using a system), was raised by one G7 member as an issue that challenges regulation/governance. One G7 member is developing legislation which provides transparency obligations, including for certain generative AI systems, which was given as an example seeking to address this concern.

Misinformation and disinformation (including foreign disinformation campaigns)

A G7 member stressed generative AI models' ability to create synthetic content (e.g., deepfakes) at scale with little resources or expertise. In particular, they noted that the next generation of interactive generative media will leverage targeted influence content that is highly personalised, localised, and conversational. Another G7 member expressed concerns about the capacity of AI-generated content to influence human behaviour, expression, and emotion at scale, as well as of content reflecting or promoting misinformation. Moreover, the member also warned of incorrect or fabricated content that is presented as a fact (i.e., "confidently wrong" or "hallucinated" output).

Low levels of digital literacy may further accelerate the spread and exacerbate the impact of misinformation, calling for improved digital literacy and assessment tools for content authenticity.

Vast amounts of data pose regulatory challenges

Two G7 members stated that training generative AI relies on vast amounts of publicly available data, which can be used without permission and may therefore not comply with data protection and copyright laws. Training data thus gives rise to copyright issues when outputs resemble the original sources. Another G7 member further noted the lack of clarity regarding what kind of training data is used as well as regarding the implications for consumer protection, and intellectual property (IP) protection and enforcement, particularly given the difficulty of constraining models from reproducing copyrighted content. The member also stressed that regulation and governance regimes are not keeping pace with rapid advances in AI capabilities.

2.3. POTENTIAL COLLECTIVE INTERNATIONAL APPROACH

COMMON CHALLENGES POSED BY GENERATIVE AI THAT REQUIRE INTERNATIONAL ALIGNMENT AND COLLABORATION

Regulatory frameworks and interoperability

G7 members responded that there is a need to establish appropriate regulation and oversight.

- One G7 member recalled that existing frameworks like the OECD AI principles require international norms, standards, and assessment processes. The member hence called for international alignment and collaboration, including with developing countries.
- Another G7 member stressed the importance of international governance (e.g., establishing common governance frameworks and international standards on the reliability of generative AI, i.e., quality control) and the need for international institutions to facilitate rapid, coordinated action among nations. This is to allow them to respond to currently unknown threats. The same G7 member noted that different areas of global governance – from climate change to international trade – have benefited from codifying and institutionalising cooperation among nations.
- Another G7 member also highlighted the need for precise and detailed principles to enable their implementation in G7 countries. The member further suggested that principles could be applied through general agreements with generative AI system providers or through binding legislation if agreements cannot be reached.
- Others stressed the need for greater interoperability of regulatory frameworks in different jurisdictions. Another G7 member also stressed the need for common guidelines to promote responsible AI development and use.

Risks

Several G7 members stressed the importance of preventing the use of generative AI to create chemical or biological threats (e.g. viruses), or massive disinformation/misinformation (including from foreign actors), highlighting the need for international cooperation on these common challenges. Cybersecurity concerns were also identified as a challenge posed by generative AI that requires international alignment and collaboration, calling for addressing international AI cyber security risks on a global level. Others also mentioned the risk of undermining social stability. Others pointed out that the rapid pace of technological developments makes it difficult for policy makers to keep up.

Personal data and intellectual property rights

A G7 member emphasised the need to take measures regarding the use of personal data as well as material protected by intellectual property rights. These measures should consider, on the one hand, the difference between data used in the process of training a model, and on the other hand, data used when the system interacts with the end user and generates content. Others also called for finding common ground at the international level on managing the data provided to train AI systems. A G7 member suggested that data provided by human operators should be traceable and withdrawable by the original provider.

Transparency

The need for transparency in both the development and use of generative AI was stressed by several G7 members. Some respondents mentioned the work led by the European Union (the EU AI Act) regarding responsibilities between actors within AI value chains. In particular, these G7 members argued that foundational model providers should provide sufficient transparency for the providers of final "products" to place them on the market safely. They also recalled that the proposed EU AI Act contains the following transparency provisions for generative AI systems: (a) a generative AI chatbot would be subject to transparency obligations; (b) generative AI system that can be directly used for high-risk applications (e.g., the evaluation of job candidates), would have to fulfil the corresponding requirements for high-risk AI systems, which include transparency provisions.

Ethics

Some G7 members mentioned the need to implement common policies on AI ethics to address issues of biases in AI systems. One G7 member added that a code of ethics is required to reduce the general misuse of generative AI, but also to govern its use in military contexts.

Seizing the benefits of AI for the common good

A G7 member highlighted that international collaboration is needed for governments to demonstrate responsible leadership in deploying and using these technologies for the public good in various sectors, including health (to enhance medical diagnostics), and the public administration (to make government more effective/efficient). Others also stated that their objective is to ensure that AI is oriented towards the common good. A G7 member pointed out that the rapid development of generative AI creates new opportunities for emerging and developing countries and that international cooperation to foster more local AI innovation should focus on the promotion of representative datasets, AI know-how, policy frameworks for responsible AI and appropriate data protection.

Unequal opportunities (within societies, and for the Global South)

One G7 member noted that while rapid advancements of generative AI create new opportunities for countries in the Global South, they also bear certain risks. In particular, the member stressed that countries in Asia and

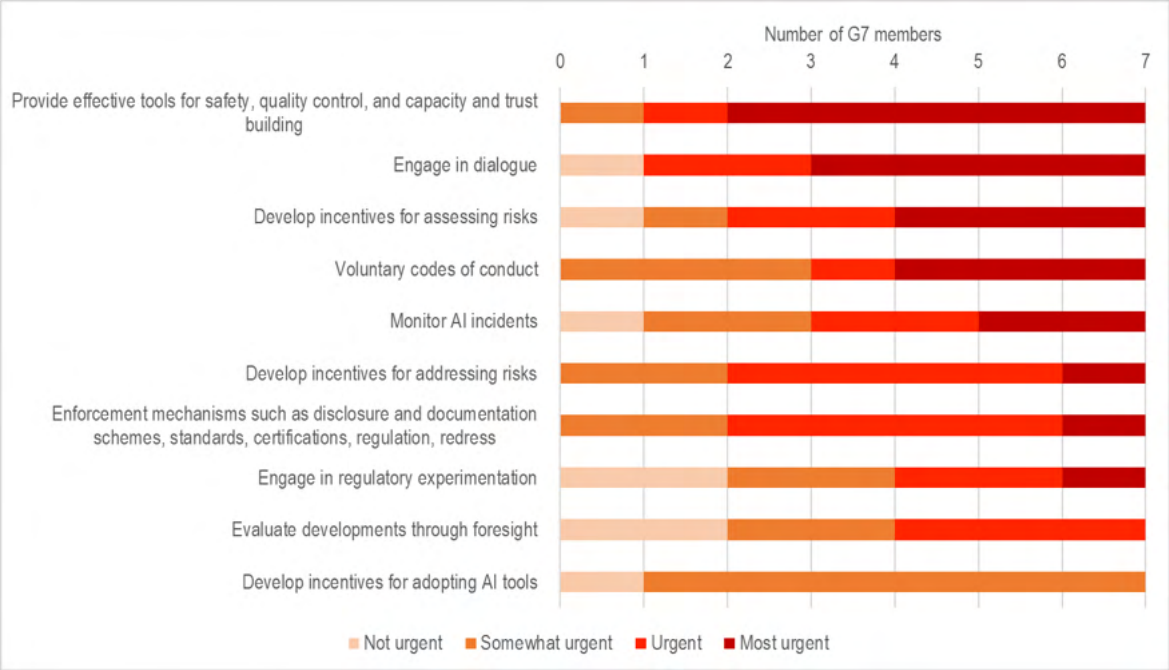
Africa are lagging behind when it comes to inclusive policy frameworks for values-based AI. Generative AI also has the potential to deepen digital divides: because generative AI is trained on existing data that may be biased, it could contribute to greater polarisation worldwide. This is why a G7 member called for involving a diverse range of stakeholders in AI governance, including governments, private companies, and civil society. Others echoed this point as they stressed the need for governments to be more responsive to rapid changes/advancements in this technology, including by engaging with a diversity of stakeholders to understand the impacts of these technologies on different parts of society and different sectors. Similarly, another G7 member pointed out that international alignment and collaboration is necessary to ensure both the trustworthiness of these technologies and a common approach towards governing them. This is particularly pertinent given that AI technologies are developed and subsequently deployed globally not just in democratic, but also in non-democratic states.

TYPES OF POSSIBLE POLICY ACTIONS THAT THE G7 COULD RECOMMEND

Providing effective tools for safety, quality control, and capacity / trust building, as well as engaging in dialogue were viewed by respondents as the most urgent actions the G7 could recommend among those outlined in the questionnaire (Figure 2.8).

Regarding "develop incentives", a G7 member noted that in many cases, the underlying tools to assess or address risks may not yet exist. Therefore, research on and development of better risk analysis tools remain important and will be needed before it is possible to incentivise use.

FIGURE 2.8. MOST URGENT TYPES OF ACTIONS THE G7 COULD RECOMMEND



Note: The figure aggregates responses from seven respondents to the question: "*What type of actions could the G7 recommend to collectively evaluate developments and how to harness the opportunities and address the risks posed by generative AI?*".

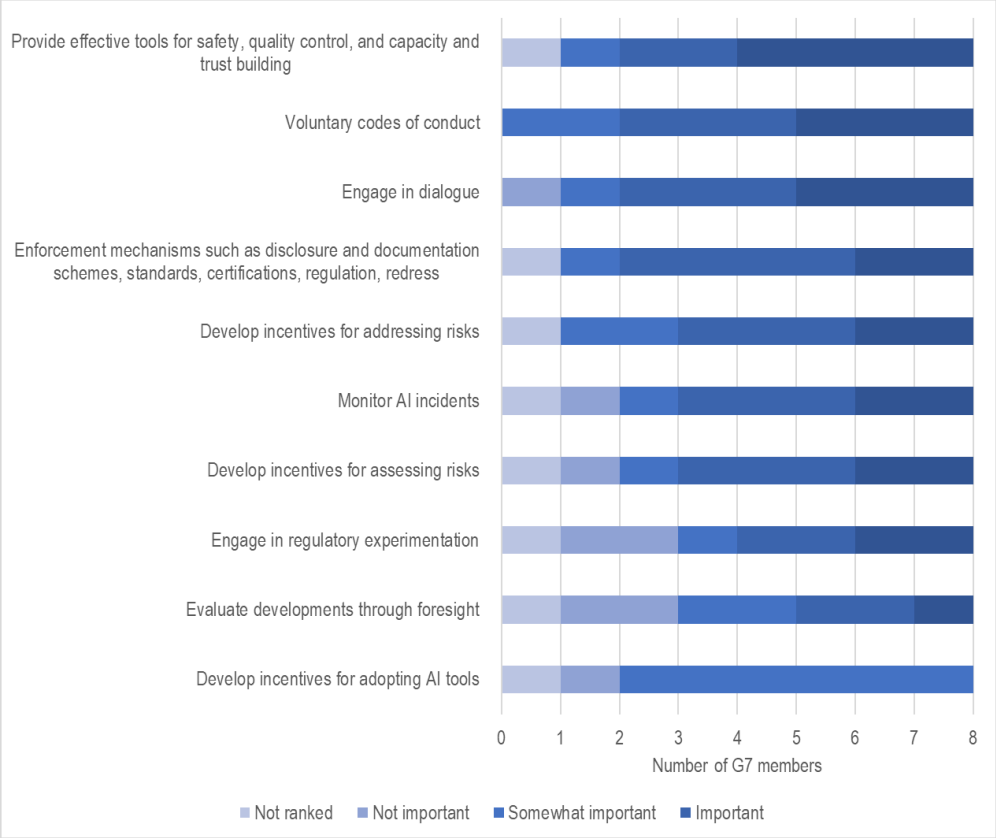
Providing effective tools for safety, quality control, and capacity / trust building, and voluntary codes of conduct were viewed by respondents as the most important actions the G7 could recommend among those outlined in the questionnaire (Figure 2.9).

A G7 member stated that working on concrete projects with AI experts to address risks and harness opportunities represents both an urgent and important type of action the G7 could recommend. According to other G7 members, other urgent and important types of action are awareness raising, sharing of knowledge and best practices, and engaging in dialogue with all relevant national and international stakeholders. Furthermore, respondents mentioned engaging in regulatory approaches as among the most important action the G7 could recommend to collectively evaluate developments and the risks posed by generative AI.

Another suggestion was tracking policy initiatives focused on generative AI through the OECD AI Policy Observatory to help grow an evidence base for policy discussions. While countries are already reporting initiatives related to generative AI to the OECD AI Policy Observatory, there is a plan to make them better identifiable by creating a specific tag and section for those initiatives.

Regarding tools for safety, quality control, capacity and trust building, the OECD launched in April 2023 the Catalogue of Tools and Metrics for Trustworthy AI. The catalogue is a platform where AI practitioners from all over the world can share and compare tools and build upon each other's efforts to create global best practices and speed up the process of implementing the OECD AI Principles. Tools can be of procedural, technical or educational nature, and are classified by their objective (that is, the OECD AI Principle they are designed to address). Currently, the catalogue counts 585 tools for trustworthy AI, out of which 28 tools specifically address generative AI.

FIGURE 2.9. MOST IMPORTANT TYPES OF ACTIONS THE G7 COULD RECOMMEND



Note: The figure aggregates responses from eight respondents to the question: "*What type of actions could the G7 recommend to collectively evaluate developments and how to harness the opportunities and address the risks posed by generative AI?*".

2.4. NATIONAL AND REGIONAL INITIATIVES IN G7 JURISDICTIONS

G7 members are leveraging opportunities (Table 2.1), or addressing challenges of generative AI (Table 2.2) either through existing initiatives or by launching new initiatives

The **United States** responded that many of their existing policies, executive orders, and approaches (e.g., the NIST AI Risk Management Framework) still apply to generative AI. Several of the voluntary commitments for leading AI companies include commitments on issues specific to generative AI, such as a commitment to develop watermarking systems.

Canada mentioned that some departments and agencies are currently exploring the possibility of launching new initiatives to support research and development in the field of generative AI and will share more information about these as they advance.

The **European Union**'s existing legislation and policies, including on data protection, misinformation, unfair commercial practices, copyright protection, and digital services fully apply to generative AI. Moreover, within the framework of the ongoing legislative negotiations on the EU AI Act, discussions are ongoing to strengthen provisions specifically focused on generative AI.

TABLE 2.1. OVERVIEW OF NATIONAL AND REGIONAL INITIATIVES TO LEVERAGE OPPORTUNITIES OF GENERATIVE AI

Member	Existing initiative	Description/Actions addressing generative AI	Type of generative AI	New initiative	Description	Type of generative AI
Canada	Pan-Canadian AI Strategy	Strategy focused on investments in talent, research capacity, commercialization and standardization.		Guidance on the use of generative AI in the federal public service		
	Directive on the Use of Automated Decision-Making	Directive accompanied by the Algorithmic Impact Assessment, which governs the use of AI systems within the federal public service		Proposed Canadian code of practice for generative AI	Canada is hosting roundtable sessions to seek stakeholder feedback on a proposed Canadian code of practice for generative AI. The code will provide voluntary guidance to companies developing and using AI systems, and will help them to prepare their processes and products before formal regulation takes effect.	
EU	EU AI Act	Transparency obligations, regulatory sandboxes	Multi-modal	Testing and Experimentation Facilities (TEFs)	EU Commission is co-funding the TEFs to support AI developers.	Open to all AI technologies
	Coordinated Plan on AI	Strategy on AI and priority areas for action, developed in cooperation with the Member States. Include actions to faciliate access to computing power, microelectronics, TEFs, digital innovation hubs		European Digital Infrastructure Consortium		
	The European High Performance Computing Joint Undertaking (EuroHPC JU)					
France	National AI Strategy	Access to computing power for start-ups	Multi-modal	EDIC for NLP	European initiative to develop an international consortium for NLP (currently being extended towards multimodal generative AI).	Multi-modal/text
		Call for projects for the creation of digital commons for generative AI in French	Multi-modal			
		Global generative AI challenge to evaluate the state of the art among LLM models in the world	Text			
Germany	National AI Strategy	Framework for a holistic policy on the future development and application of AI	Multi-modal	Large European AI Models (LEAM)	Initiative from the German AI Association to foster LLM development in the EU and Germany	Text
	FAIR Forward	Development initiative for a more open and sustainable application of AI in developing and emerging economies		F13	LLM based on the Luminous model used in the government of the federal state of Baden-Württemberg	Text

Member	Existing initiative	Description/Actions addressing generative AI	Type of generative AI	New initiative	Description	Type of generative AI
	Guideline for the promotion of AI for the common good	Strengthen the use of data and technologies using AI for the common good		BEKI	Advisory centre for AI in the public sector	
	Centres of excellence for AI research	Six national competence centres for AI research to strengthen excellence in AI	Multi-modal			
	AI quality and innovation center	Development of practical AI testing approaches to ensure standards and quality in AI applications	Multi-modal			
	Civic Coding – KI für das Gemeinwohl	Structures promoting the emergence of social innovations with AI				
	Learning systems – Germany's platform for AI	Forum for exchange and cooperation	Multi-modal			
Japan	AI strategy 2022	National AI Strategy presenting a comprehensive policy package related to AI		AI Strategy Team and AI Strategy Council	Government and experts' bodies examinig a wide range of issues related to generative AI.	Multi-modal
				Tentative summary of AI issues	Summary of issues related to AI (mainly generative AI), by the AI Strategic Council	Multi-modal
				Use of AI	Agreement on the business use of generative AI within the government.	Multi-modal
				Strengthening of AI development capability	Support to the development of computing resources for private companies.	Multi-modal
Italy	Fund for the development of AI, blockchain and internet of things (IoT) technologies and applications.	Fund for enterprises, including SMEs. It supports projects involving the implementation of industrial research and innovation for developing AI, blockchain, and internet of things technologies applications	Multi-modal	Revision of the Italian Strategic Plan for AI 2021 (to be completed and out for open consultation at the end of Sep. 2023)	Updating the national AI strategy in view of the AI Act and the development of Generative AI	Multimodal
	Italian National PhD Program in Artificial Intelligence.	Five federated PhD courses that bring together 61 universities and research institutions	Multi-modal	New governance: National Authority on AI within the National Agency for Digital Italy (AGID) under the supervision of the Department for Digital Transformation	The Authority will coordinate all the new policy initiatives with a special focus on Generative AI	
				Corporate Venture Capital Fund for AI start-ups to develop solutions for the public sector	Implementing the "National AI Strategy 2021," Goal n. 5: "Develop AI-driven policies and services in the public sector by boosting public sector innovation"	Multimodal

Member	Existing initiative	Description/Actions addressing generative AI	Type of generative AI	New initiative	Description	Type of generative AI
				Personalised Virtual assistant - National Institute for Social Security (INPS)	Virtual assistant based on Generative AI. The virtual assistant helps users to improve the search experience on the INPS website. Also, in its beta version, the virtual assistant can help users to submit specific questions relating to "Opzione Donna" ("Option Woman"), a social security measure for women only	Multimodal
				Certified e-mails (PEC) automatic classification and sorting - National Institute for Social Security (INPS)	Optimise the communication flow of millions of certified emails that are received weekly	Text
				Future Artificial Intelligence Research (FAIR) Centre	Research network to address research aspects, methodologies, models, technologies, and ethical and legal rules for building human centric AI systems. It includes four research institutions and 14 universities	
United Kingdom	AI Regulation White Paper	Context-based, proportionate, and adaptable approach to regulating AI. It draws on expertise of existing regulators	Multi-modal	Foundation Model Taskforce	Taskforce will lead AI safety research to drive forward safe and reliable development of Foundation Models	Multi-modal
	Centre for Data Ethics and Innovation (CDEI) portfolio of AI assurance techniques	Case studies of AI assurance techniques being applied by organisations across a range of sectors	Multi-modal	UK Global Summit on AI Safety	The summit will consider risks of AI, and discuss how they can be mitigated through internationally coordinated action	Multi-modal
	AI Standards Hub	Practical tools and information to improve AI standards adoption and development	Multi-modal			
United States				Voluntary Commitments	These include a commitment from leading AI companies to develop and deploy advanced AI systems to help address society's greatest challenges.	Multi-modal
				NIST Generative AI Public Working Group	Developing a profile of the NIST AI Risk Management Framework for generative AI systems	Multi-modal

Member	Existing initiative	Description/Actions addressing generative AI	Type of generative AI	New initiative	Description	Type of generative AI
				President's Council of Advisors on Science and Technology (PCAST) working group on generative AI	Developing recommendations for the President on how best to ensure that these technologies are developed and deployed as equitably, responsibly, and safely as possible.	Multi-modal

Table 2.2. OVERVIEW OF NATIONAL AND REGIONAL INITIATIVES TO ADDRESS RISKS RELATED TO GENERATIVE AI

Member	Initiative	Description	Type of generative AI
Canada	Artificial Intelligence and Data Act (AIDA)	Proposes the development of a risk-based national regulatory framework for the responsible design, development, and use of AI in Canada's private sector	Multi-modal
	Proposed Canadian code of practice for generative AI	Canada is hosting roundtable sessions to seek stakeholder feedback on a proposed Canadian code of practice for generative AI. The code will provide voluntary guidance to companies developing and using AI systems, and will help them to prepare their processes and products before formal regulation takes effect.	Mutli-modal
	TBS guide on the use of generative AI in the Government of Canada	Provides guidance to federal institutions on their use of generative AI tools	Multi-modal
EU	Artificial Intelligence Act	EU legislation currently in the legislative process	Multi-modal
Japan	Tentative summary of AI issues	Summary of issues related to AI (mainly generative AI), by the AI Strategic Council	Multi-modal
	Response to risks of AI	Review of uniform guidelines for business players, summary of issues on intellectual property rights, establishment of guidelines for the use of generative AI in education, etc.	Multi-modal
Italy	Policy Paper on Risk Assessment, Auditing and Management	Various regulatory actors engaged to produce a policy paper with insights on risk assessment, auditing, and risk management	Multi-modal
United Kingdom	AI Regulation White Paper	Context-based, proportionate, and adaptable approach to regulating AI. It draws on expertise of existing regulators	Multi-modal
	Centre for Data Ethics and Innovation (CDEI) portfolio of AI assurance techniques	Case studies of AI assurance techniques being applied by organisations across a range of sectors	Multi-modal
	UK Global Summit on AI Safety	The summit will consider risks of AI, and discuss how they can be mitigated through internationally coordinated action	Multi-modal
	AI Standards Hub	Practical tools and information to improve AI standards adoption and development	Multi-modal
United States	Voluntary Commitments	Voluntary commitments from leading AI companies to help move toward safe, secure, and transparent development of AI technology	Multi-modal
	NIST Generative AI Public Working Group	Developing a profile of the NIST AI Risk Management Framework for generative AI systems	Multi-modal
	President's Council of Advisors on Science and Technology (PCAST) working group on generative AI	Developing recommendations for the President on how best to ensure that these technologies are developed and deployed as equitably, responsibly, and safely as possible.	Multi-modal

NATIONAL AND REGIONAL INITIATIVES ON GENERATIVE AI BY G7 MEMBER

Canada

- Artificial Intelligence and Data Act (AIDA): The AIDA was tabled in Canada's Parliament in June 2022 and proposes the development of a risk-based national regulatory framework for the responsible design, development, and use of AI in Canada's private sector.
- Proposed Canadian code of practice for generative AI: Canada is hosting roundtable sessions to seek stakeholder feedback on a proposed Canadian code of practice for generative AI. The code will provide voluntary guidance to companies developing and using AI systems, and it will help them to prepare their processes and products before formal regulation takes effect.
- TBS guide on the use of generative AI in the Government of Canada: The TBS guide, intended as an evergreen document to be refined over time, is intended to provide guidance to federal institutions on their use of generative AI tools. It would provide an overview of generative AI, identifies challenges and concerns relating to its use, puts forward principles for using it responsibly, and offers policy considerations and best practices.

EU

- Testing and Experimentation Facilities (TEFs): In cooperation with the EU Member States, the European Commission is co-funding the TEFs in order to support AI developers to bring trustworthy AI to the market in a more efficient way and to facilitate its uptake in Europe.
- Artificial Intelligence Act (AI Act): Since AI systems in general and generative AI systems in particular could potentially challenge and impact a broad spectrum of areas and use cases, the European Commission proposed the AI Act in 2021. The EU AI Act is a legal framework, which aims at making sure that people can trust what AI has to offer.

France

- National Strategy on AI: It includes the support of the emergence of generative AI initiatives in France, via an easier access to computing power for start-ups, a call for projects for the creation of digital commons for generative AI in French (databases, AI application models, etc.) and the launch of a global generative AI challenge in order to evaluate the state-of-the-art among Large Language Models (LLM) around the world.
- EDIC for NLP: European initiative to develop an international consortium for NLP (currently being extended towards multimodal generative AI). This will allow to develop a European set of databases, share the European computing power for promising projects, an incubator for start-ups, financing for research projects, etc.

Germany

- National AI Strategy: Framework for a holistic policy on the future development and application of AI.
- FAIR Forward: Development initiative working towards a more open and sustainable application of AI that involves developing and emerging economies.
- Large European AI Models (LEAM) initiative: Initiative from the German AI Association to foster LLM development in the EU and Germany.
- Centres of excellence for AI research: The Federal Ministry of Education and Research has established six national competence centres for AI research to strengthen excellence and competitiveness as well as to become a leading centre for AI research.
- Civic Coding – KI für das Gemeinwohl: Innovation Network to create structures that promote the emergence of social innovations and the social appropriation of AI on a broader basis.
- Guideline for the promotion of AI for the common good: Funding research, implementation, and model projects to strengthen the use of data and technologies using AI for the common good.
- F13: LLM based on the Luminous model. It is used in the government of the federal state of Baden-Württemberg.
- Learning systems – Germany's platform for AI: A forum for exchange and cooperation which brings together expertise from science, industry, and society for fostering Germany's position as an international technology leader.
- BEKI: Advisory centre for AI in the public sector.

Japan

- AI Strategy 2022: Launched in April 2022 and based on the principles Dignity for People, Diversity and Sustainability, this strategy is to contribute to the resolution of global issues through

the realisation of Society 5.0, and to present a comprehensive policy package related to AI for overcoming Japan's own social issues and improving industrial competitiveness.

- The AI Strategy Team and the AI Strategy Council: In April 2023, the AI Strategy Team - consisting of working-level officials from related ministries and agencies - was established to study how to address a wide range of issues related to generative AI. Furthermore, in May 2023, the AI Strategy Council - consisting of experts - was established to examine not only technology but also the legal system and ethics from a wide range of perspectives.
- Tentative summary of AI issues: Based on recent rapid changes in technology and the 2023 G7 Hiroshima Summit, members of the AI Strategic Council summarised issues related to AI, mainly generative AI, as of the end of May 2023. Currently, Japan is in the process of promoting discussions on response to risks of AI, use of AI, strengthening of AI development capability, as well as international discussions, including the "Hiroshima AI Process" as the G7 chair country.
- Response to risks of AI: Review of uniform guidelines for business players, summary of issues on intellectual property rights, establishment of guidelines for the use of generative AI in education, etc.
- Use of AI: Agreement on the business use of generative AI within the government.
- Strengthening of AI development capability: Support to the development of computing resources for private companies.

Italy

- Implementing the National AI Strategy 2021: Goal n. 5 "Develop AI-driven policies and services in the public sector by boosting public sector innovation".
- Establishment of the Corporate Venture Capital Fund for AI start-ups to develop solutions for the public administration: The Fund (operating in 2024) will invest in start-ups with the potential to develop breakthrough technologies to automate public institutions' processes. The investment amounts to EUR 600 million. The Fund includes an institutional mechanism to foster the dialogue between start-ups and the public administrations to better understand the public sector's needs. The Fund is under the responsibility of the National Cybersecurity Agency and the Department for Digital Transformation.
- Establishment of the National Authority on AI within the National Agency for Digital Italy (AGID), under the supervision of the Department for Digital Transformation. The Authority will coordinate all the new policy initiatives with a special focus on Generative AI.
- Fund for the development of AI, blockchain and internet of things (IoT) technologies and applications: A dedicated fund for enterprises, including SMEs. It supports projects involving the implementation of industrial research and innovation for developing AI, blockchain, and IoT technologies applications.
- The National Institute for Social Security (INPS)'s Personalized virtual assistant to help users to navigate the many services offered on the INPS portal.
- The National Institute for Social Security (INPS) - Certified e-mails automatic classification and sorting: The project aims to optimise the communication flow of the millions of certified emails sent to INPS through a system that can automatically understand the content of the email received and direct it to the proper official in charge of that particular response.
- Italian National PhD Program in Artificial Intelligence: The Italian National PhD Program in Artificial Intelligence is made of five federated PhD courses that bring together 61 universities and research institutions. The 5 PhD courses share a common basis in the foundations and developments of AI, and each one has an area of specialisation in a strategic sector of AI

application. Each PhD course is organised by a lead university, in collaboration with the National Research Council (CNR).

- Future Artificial Intelligence Research (FAIR) Centre: The Future AI Intelligence Research (FAIR) project aims to help address the research questions, methodologies, models, technologies, and even ethical and legal rules for building human centric AI systems. FAIR constitutes a research network spread over the country and includes four research institutions (CNR, Fondazione Bruno Kessler, INFN, and IIT), 14 universities (Politecnico di Milano, Politecnico di Torino, Sapienza, Scuola Normale Superiore, SISSA, Università Bocconi, Università Campus Biomedico di Roma, Università della Calabria, Università di Bari, Università di Bologna, Università di Catania, Università di Napoli "Federico II," Università di Pisa, Università di Trento) and seven companies (Bracco, Deloitte, Expert. ai, Intesa Sanpaolo, Leonardo, Lutech, STMicroelectronics).
- Policy Paper on Risk Assessment, Auditing and Management: Various regulatory actors engaged in a discussion to produce a policy paper to produce insights on the risk assessment, auditing, and risk management, with the aim of understanding the position of start-ups, trade association, and the third sector.

United Kingdom

- AI Regulation White Paper: The White Paper sets out the UK's context-based, proportionate, and adaptable approach to regulate AI, and draws on expertise of existing regulators; encouraging them to consider how best to govern AI in their own sectors. It will enable the UK to achieve the right balance between responding to risks and maximising opportunities afforded by AI.
- Centre for Data Ethics and Innovation (CDEI) portfolio of AI assurance techniques: The portfolio features case studies of AI assurance techniques being applied by organisations using cutting-edge technologies across a range of sectors. This will act as a valuable resource for those developing and procuring AI systems to understand how AI assurance techniques can help them measure, evaluate, and communicate trustworthiness of AI systems, as well as how techniques align with proposed regulatory principles identified in the UK's AI Regulation White Paper.
- Foundational Model Taskforce: The taskforce will lead vital AI safety research as part of driving forward safe and reliable development of Foundation Models while seizing extraordinary opportunities they present. The taskforce is backed with initial GBP100 million of government funding.
- UK Global Summit on AI Safety: The summit will consider risks of AI, including frontier systems, and discuss how they can be mitigated through internationally coordinated action. It will also provide a platform for countries to work together on further developing a shared approach to mitigate risks.
- AI Standards Hub: It aims to improve AI standards adoption and development by providing businesses, regulators, and civil society organisations in the UK with practical tools and information. Moreover, they need to apply AI standards effectively and contribute to their development. AI Standards Hub is part of the National AI Strategy and ultimately aims to increase the UK's contribution to the development of global AI technical standards.

United States

- Voluntary Commitments: To make the most of AI's potential, the United States is encouraging the AI industry to uphold the highest standards to ensure that innovation does not come at the expense of Americans' rights and safety. The White House secured voluntary commitments from leading AI companies to help move toward safe, secure, and transparent development of AI

technology. The commitments underscore three principles fundamental to the future of AI – safety, security, and trust – and mark a critical step toward developing responsible AI.

- NIST Generative AI Public Working Group: This working group builds on the success of the NIST AI Risk Management Framework to address rapidly advancing AI technologies. The Public Working Group on Generative AI will help address the opportunities and challenges associated with AI that can generate content, such as code, text, images, videos and music. The public working group will also help NIST develop key guidance to help organizations address the special risks associated with generative AI technologies.
- PCAST Generative AI Working Group: The President's Council of Advisors on Science and Technology (PCAST) has launched a working group on generative artificial intelligence (AI) to help assess key opportunities and risks and provide input on how best to ensure that these technologies are developed and deployed as equitably, responsibly, and safely as possible. The PCAST Working Group on Generative AI aims to build upon existing efforts by identifying additional needs and opportunities and making recommendations to the President for how best to address them.

References

OECD (2023), *AI language models: Technological, socio-economic and policy considerations*, OECD Publishing, https://doi.org/10.1787/13d38f92-en. [1]

OECD (Forthcoming), *Generative Artificial Intelligence. Select policy considerations, Part 1*. [2]