Preliminary Report of the Independent International Scientific Panel on AI

July 2026

Evidence-based assessment of opportunities, risks and impacts of artificial intelligence

Preliminary Report of the Independent International Scientific Panel on AI: Evidence-based assessment of opportunities, risks and impacts of artificial intelligence

Print ISBN: 9789211576627
PDF ISBN: 9789211550771

Preliminary Report of the Independent International Scientific Panel on AI

Evidence-based assessment of opportunities, risks and impacts of artificial intelligence

July 2026

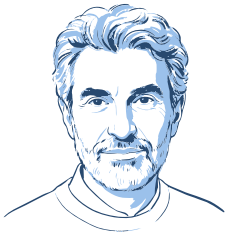
United Nations
Independent International Scientific Panel on Artificial Intelligence

Members of the Independent International Scientific Panel on Artificial Intelligence

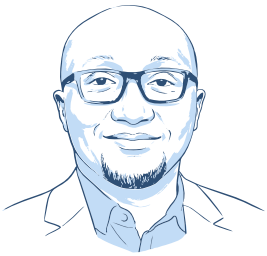
Girmaw Abebe Tadesse
Ethiopia

Tuka Alhanai
United Arab Emirates

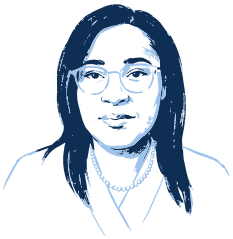
Joëlle Barral
France



Yoshua Bengio
Canada
Co-Chair



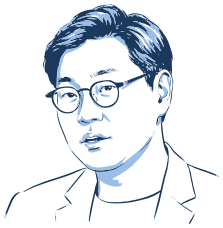
Tegawendé Bissyandé
Burkina Faso



Awa Bousso Dramé
Cabo Verde

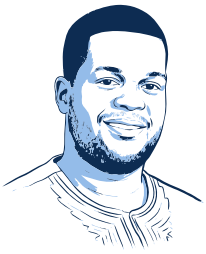
Mennatallah El-Assady
Egypt

Hoda Heidari
Islamic Republic of Iran

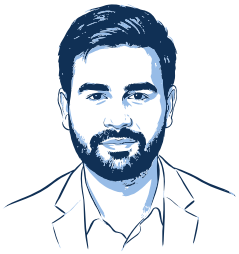


Juho Kim
Republic of Korea

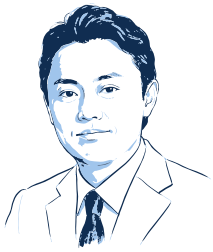
Anna Korhonen
Finland



Vukosi Marivate
South Africa



Bilal Mateen
Pakistan



Yutaka Matsuo
Japan

Joyce Nakatumba Nabende
Uganda

Andrei Neznamov
Russian Federation

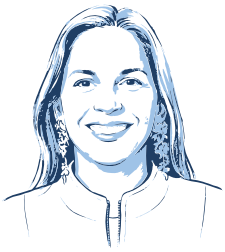
Johanna Pirker
Austria

Balaraman Ravindran
India

Maria Ressa
Philippines
Co-Chair

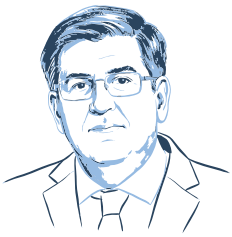
Lior Rokach
Israel

Piotr Sankowski
Poland



Loreto Bravo
Chile

Mark Coeckelbergh
Belgium

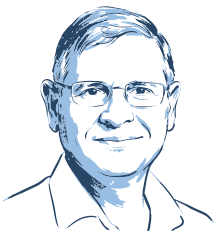


Carlos Coello Coello
Mexico

Melahat Bilge Demirköz
Türkiye

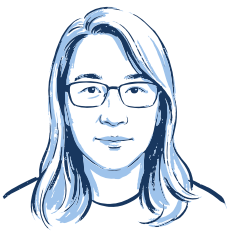
Adji Bousso Dieng
Senegal

Aleksandra Korolova
Latvia



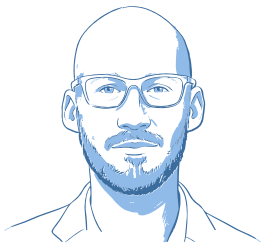
Vipin Kumar
United States

Sonia Livingstone
United Kingdom

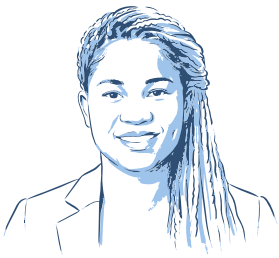


Qinghua Lu
Australia

Teresa Ludermir
Brazil



Maximilian Nickel
Germany



Rita Orji
Nigeria

Román Orús
Spain

Alvitta Ottley
Saint Kitts and Nevis

Martha Palmer
United States

Silvio Savarese
Italy

Bernhard Schölkopf
Germany

Haitao Song
China

Leslie Teo
Singapore

Jian Wang
China

About this Report

This report presents a preliminary independent scientific assessment of the capabilities and the emerging opportunities and risks of artificial intelligence (AI), providing a shared evidence base to help Member States navigate a rapidly changing technology. The report is authored by the Independent International Scientific Panel on Artificial Intelligence, a body established by the General Assembly in its resolution 79/325 in 2025. The Panel serves as the first global scientific body on AI, operating under a strictly scientific, non-political mandate to document international scientific consensus and disagreements while remaining policy-relevant but not policy-prescriptive. The report is the first of its kind and will be updated progressively throughout the year, with thematic briefs addressing developments as they arise. It reflects the best available evidence at the time of publication, in a field moving so rapidly that any snapshot requires a commitment to revision.

Table of Contents

Executive summary

This document seeks to present a balanced analysis of the risks and opportunities of artificial intelligence (AI). "Balanced" means a commitment to evaluating empirical data without undue bias towards optimism or pessimism. The potential benefits of AI are enormous. Deployed and applied thoughtfully, AI can support progress towards achieving the Sustainable Development Goals, advance health science and increase access to education. At the same time, the rapid pace of technological development and the breadth of potential applications present policymakers with significant challenges. The rapid, unchecked deployment of the technology at scale also presents considerable risks, including harms to the mental health of users, potential use as a destructive tool, impacts on social, economic and environmental systems, and challenges associated with controlling the technology. This report does not aim to consider the full scope of all possible opportunities and risks but rather focuses on some of the most pressing ones.

Capabilities and adoption

Recent years have seen rapid, and in some areas accelerating, progress in a range of AI capabilities. Significant investments in computing power, new AI methodologies, and specialized training data have led to sustained improvements in a wide range of AI capabilities. These include fluent conversation, functional code generation, expert-level reasoning in mathematics and science, large-scale data analysis, and the generation of image, audio and video content. Limitations remain, such as in reliability, obtaining strong performance across human languages and cultures, interacting with physical systems, executing complex or multi-step projects and producing factual outputs; in general, however, technical progress in many important domains has proceeded quickly, beyond the typical expectation of technology advancement, for several years now.

These gains have unlocked useful applications across science, health, agriculture, accessibility, knowledge work and information technology, including in the development of AI itself. For example, in science, AlphaFold has predicted the structures of more than 200 million proteins, now used by over 3 million researchers, and accelerated drug design, vaccine development and antibiotic resistance research. Radiologists have also used AI to detect breast cancer earlier, while front-line health workers in low-resource settings use AI tools adapted to local languages to deliver better-quality healthcare services.

AI adoption has accelerated broadly, and unevenly, across countries and sectors. Globally, over a billion people now use conversational AI weekly. Yet AI access and usage vary widely globally, with adoption across the global South lagging far behind the global North. Furthermore, there are significant differences in compute infrastructure and models between advanced economies. This disparity reflects, and may even reinforce, existing inequalities. AI development itself is even more concentrated: according to recent estimates, the United States of America accounts for 75% of the computing power among the world's top 500 AI supercomputers, with China accounting for 15%. Companies in the United States and China also develop almost all leading general-purpose models, and a small number of countries control critical inputs for the supply chain of AI computer chips.

While the shift towards AI agents is under way, their future adoption and economic impacts will likely be shaped by continued improvements in their ability to accomplish knowledge work with little or no human oversight. An AI agent is a computer system that can plan and autonomously act towards achieving goals, using the tools at its disposal.

These systems have been improving rapidly in recent years, with one study finding that the length of certain software tasks that leading systems can accomplish has been doubling every four to seven months. If this rate of improvement continues, AI agents will soon complete tasks that currently take human programmers days or weeks. Because they can work with little oversight and at rapid speed, AI agents may lead to significant economic and scientific benefits. For example, agentic AI systems in self-driving chemistry labs have demonstrated more than a tenfold increase in the speed of materials discovery. AI-assisted literature screening may have reduced workloads by roughly 60% in some research settings. AI agents therefore carry significant implications across all industries. At the same time, their deployment raises urgent questions for labour markets, cybersecurity, the information ecosystem, and the governance and controllability of future AI systems.

Understanding and managing risks

AI development entails risks, with potential negative impacts on human rights, social systems and the environment. For example, AI-generated child sexual abuse material and deepfake-enabled sexual violence now circulate more frequently on the Internet, disproportionately harming women and children. Sycophantic AI behaviour, where AI responses reinforce users' existing beliefs regardless of accuracy, has been linked to several severe mental health incidents, including documented deaths. AI makes it easier to produce and target persuasive content at scale, including content designed to mislead, contributing to a gradual erosion of information integrity that can weaken the shared reality required for public trust, social cohesion and democratic deliberation. Criminals and bad actors have been documented using AI systems to assist in cyberattacks. Many of these harms fall disproportionately on already disadvantaged populations.

Looking ahead, the gap between rapidly improving capabilities and effective risk management methods may lead to catastrophic outcomes. For example, advanced technical abilities may allow novice private actors to use AI in malicious ways across a range of applications such as fraud, social engineering, cybersecurity, disinformation, biotechnology and financial manipulation. Reliable methods for retaining control over highly autonomous AI systems are lacking. There are no scientific guarantees that AI agents will not violate instructions, and evidence is accumulating of cases where they already violate them. In laboratory settings, AI systems have been shown to violate their safety instructions to avoid being shut down. Similar behaviour may pose challenges to evaluation and oversight methods, as the ability of leading AI systems to recognize testing environments and produce misleading evaluation results that would favour their continued operation grows. Additionally, novel risks may arise from interactions between multiple agents.

AI risks are unevenly distributed across populations and countries, while AI development and the wealth it creates are highly concentrated. The concentration of AI capabilities in a small number of firms and countries could enable authoritarian capture and undermine democratic accountability.

Governing artificial intelligence to unlock benefits and mitigate risks

Realizing the full benefits of AI while minimizing its risks requires good governance. Economic and labour gains and their equitable distribution are not automatic: with complementary investments in skills, workflows, infrastructure and labour-market institutions, technology can create new jobs that do not exist right now – over 60% of jobs in 2018 compared to 1945 are new. Without these investments, AI risks widening inequality, displacing workers and shifting wealth from labour to capital rather than creating sustainable good jobs – those with fair compensation, worker autonomy and a reliable path to social dignity. AI can profoundly expand human capabilities through personalized education, accessible mental health tools and improved assistive technologies, but realizing these opportunities safely

requires dedicated investments and policies to incentivize equitable access and reward innovation, while preventing the exploitation of vulnerable populations, particularly children, and avoiding displacement of expertise, psychological dependency or cultural and linguistic erasure.

Policymakers seeking to shape this governance face an evidence dilemma: they need evidence to make informed consequential governance decisions, but by the time the evidence exists, it might be too late to make them, as the evidence lags behind the pace of AI development. Dozens of distinct governance instruments that seek to embed ethics and human rights in AI systems are already in use across jurisdictions, but they are fragmented, are concentrated among a few corporations and rarely measure real-world effectiveness. Evaluation methods themselves are underdeveloped, and the institutions needed to provide independent capability and risk assessments remain embryonic.

The capacity to act on existing evidence of AI risks and impacts is unevenly distributed. Most countries, including many advanced economies, lack the technical expertise to assess the most capable "frontier" models or to participate meaningfully in their governance. Compute infrastructure, evaluation expertise and data (e.g. to cover different languages) are concentrated where AI is built, leaving most Member States dependent on systems they cannot build, inspect, audit or fully adapt to local context. Access to AI tools alone does not produce equal benefit; the complementary investments in data, skills, workflows and institutions that turn access into useful, cost-effective and safe deployment are necessary yet unequally distributed.

Concrete next steps to close the above gaps exist, but each requires sustained investment in Member State capacity to shape, evaluate and deploy AI. This preliminary report is itself part of the Panel's contribution, a shared evidence base for Member States navigating increasingly urgent decisions. As the Panel's understanding deepens through continued engagement with Member States and the broader scientific community, so too will its analysis, expanding beyond the gaps identified to chart the trajectories, tensions and opportunities that will define the future of AI.

1. Why this moment?

The present reality

We are at an inflection point. Artificial intelligence is not simply another emerging technology; it is the first to compress adoption from decades into months, industrialize cognitive work at scale and concentrate transformative capability in the hands of a few global actors. This report equips policymakers across all regions with the shared evidence base needed to respond.

What is artificial intelligence?

Artificial intelligence (AI) is a transformative technology, but also a moving target. The term has shifted over time, from symbolic AI to machine learning, to generative AI, agentic AI and sometimes even artificial general intelligence or superintelligence.

AI systems are machine systems that, broadly speaking, perceive, learn and act. They infer from inputs how to generate outputs such as predictions, content, recommendations, actions or decisions, with varying degrees of autonomy and adaptiveness. What unifies current AI more than any single architecture is that modern systems learn from experience represented by data. That experience takes several forms: learning from human cultural traces (texts, images, code) provides the "pre-training" basis of today's foundation models, which are large, broadly trained systems that underpin a wide range of AI applications; learning by interaction with the (digital and physical) world underlies reinforcement learning and robotics; and learning from simulation allows agents to acquire experience in virtual environments.

Foundation models and task-specific AI. Public debate often focuses on foundation models and general-purpose AI, defined by their ability to perform, or adapt to perform, a wide variety of tasks. These systems are currently being deployed at scale and are the main topic of this report. However, many task-specific (narrow) AI systems are designed for performing specific tasks in particular domains. The distinction matters. Task-specific AI delivers measurable benefits when the task is well defined, the data are available and institutions can deliberately deploy it. General-purpose systems offer flexibility but introduce different governance challenges.

Why is artificial intelligence different from other emerging technologies?

Unprecedented pace of adoption. AI systems are increasingly considered a general-purpose technology, as transformative in breadth of application as the steam engine, electricity and the Internet [1,2]. However, it is distinct in important ways. Electricity took decades to reach most households; the globalization of the Internet through the World Wide Web needed about 15 years to reach a billion users. ChatGPT reached 100 million users in two months [3]. Traditional policymaking has not been able to keep pace [4].

Concentration and homogenization. Modern AI, particularly foundation models, demonstrates unprecedented economies of scale that create strong pressures towards centralization of capability, with the most powerful systems requiring training with computational resources accessible only to a handful of global actors. This concentration of resources introduces risks of knowledge and cultural homogenization. For example, training on consolidated data sets may create systems that systematically reflect dominant languages and perspectives while marginalizing others.

Massive impact on knowledge work. Where earlier waves of automation transformed physical labour, AI is the first to massively affect cognitive and creative work, including writing, programming, legal analysis, medical diagnostics, scientific discovery, forecasting and image generation [1,2,5–8].

Challenges in ensuring the factuality and correctness of outputs. Today's AI models learn to predict patterns in large data sets. Language models learn not only stored templates but transformations to turn a sentence into many related forms while preserving fluency. This enables novel outputs, rather than copying, but creates a critical asymmetry: producing fluent text is easier than producing factual text [8]. Large language models can present hallucinations as facts, and users often treat linguistic confidence as evidence of trustworthiness and factual reliability. This disconnect between apparent competence and accuracy shapes the risk landscape in systematic ways. In healthcare, general-purpose AI decreases administrative documentation time, a task where AI is valued for synthesizing and structuring information. Yet one in four chatbot conversations reportedly relate to health or wellness [9], meaning the same systems are routinely consulted for potential diagnostic purposes, where factual accuracy is critical and errors carry serious consequences. The technology is identical, but the stakes are not. Capability does not equal appropriateness, and deployment without systematic monitoring and evaluation, particularly in domains that currently lack regulatory frameworks, risks harm where it is hard to detect.

Urgent need for independent evidence-based assessment of artificial intelligence systems

The need for independent scientific assessment at this moment is driven by circumstances that change the regulatory stakes.

First, AI development is outpacing prior expert assessments and regulatory cycles. Capability progress continues across key domains, driven by new training techniques and inference-time compute scaling [10]. Empirical measurements show accelerating AI capabilities [11].

Second, leading frontier model developers have begun restricting deployment of models that exceed internally defined risk thresholds [12–15]. However, these thresholds remain defined by the developers, without standardized evaluation or external verification [16–17].

Third, AI governance approaches across regions remain fragmented. Growing disorder in global governance is observed, with some countries having introduced AI specific legislation with fundamentally contradictory rules and compliance costs. Jurisdictions exhibit divergent regulatory philosophies, with no unified deployed mechanism for risk management, no comparable evaluation standards, and limited coordination across jurisdictions, risking a fragmented regulatory landscape [18]. Yet fragmentation is not fate, and the window to establish shared evidence standards and coordinate global oversight remains open.

What makes the Independent International Scientific Panel on Artificial Intelligence unique

Assessments of AI development are produced by numerous expert groups affiliated with international organizations, national scientific councils, industry consortiums and independent research centres. These efforts are valuable but limited by geographical mandate, sectoral perspective or lack of continuity.

This Panel occupies a distinct position for three reasons. First, it proceeds from the premise that the United Nations is the foremost global forum on transboundary risks of this scale, as articulated in the report *Governing AI for Humanity* and reflected in the Global Digital Compact, which sets out the imperative to "recognize that the pace and power of emerging technologies are creating new possibilities" and "to identify and mitigate risks and to ensure human oversight of technology in ways that advance sustainable development and the full enjoyment of human rights".

Second, the Panel is currently the only standing United Nations mechanism with a mandate for regular scientific assessment of the state, risks and capabilities of AI, designed for sustained, iterative work.

Third, the Panel has a scientific, not political, mandate: to document scientific evidence, consensus and disagreements, and which knowledge gaps remain urgent to address. Its purpose is to equip governments and institutions with the evidence base they need to act over the coming months and years, remaining policy-relevant but not policy-prescriptive. This scientific character should make its findings comparable across regions and resilient to political cycles.

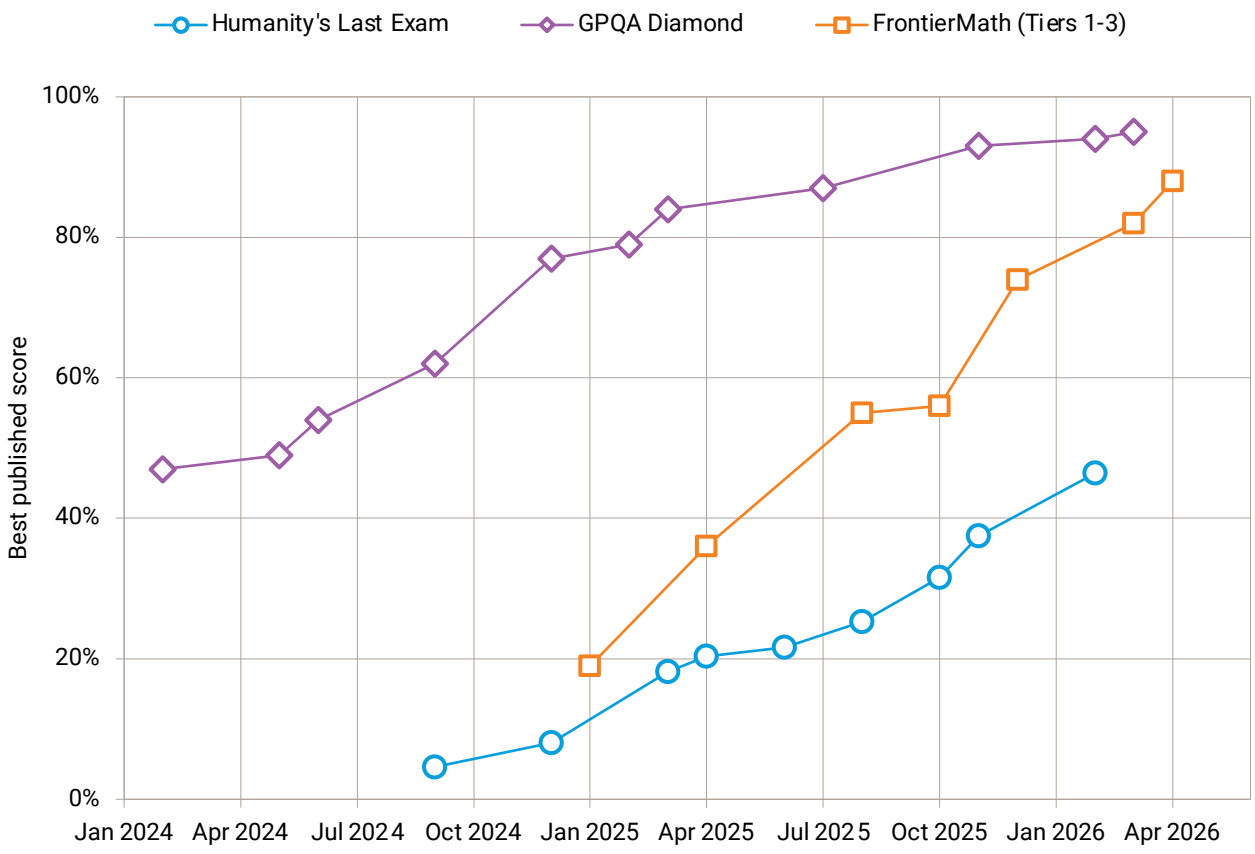
2. What does the evidence show?

AI performance on leading performance benchmarks has risen sharply in recent years (see figure I).

Humanity's Last Exam, a 2,500-question benchmark designed specifically to be hard for general-purpose models, has seen top scores climb from 8% to 45% in 16 months [19]. On GPQA Diamond, a PhD-level scientific reasoning test, the best models now correctly answer around 95% of questions, up from 36% in 2023 [20]. The leading performance on FrontierMath, which tests mathematical reasoning ability, rose from 19% in January 2025 to 88% in 2026 [20]. Multiple AI systems achieved gold medal performance on the 2025 International Mathematical Olympiad, a milestone that arrived significantly earlier than many experts had predicted [21].

FIGURE I

AI benchmark state-of-the-art, 2024 – May 2026



Source: Adapted from data on AI benchmarks (EpochAI, 2026, https://epoch.ai/benchmarks).

2.1 Artificial intelligence capabilities are advancing faster than the ability to measure or govern them

AI measurement and evaluation are the basis for assessing the opportunities, risks and impacts of AI. However, the unprecedented pace of AI development poses the following assessment challenges:

a. **There is information asymmetry in safety validation between companies and society.** Frontier AI developers retain proprietary visibility of the systems they have built. Safety evaluation methodologies are currently designed largely by the companies being evaluated. While there are some legally mandated disclosures, government experts primarily receive the testing data that developers choose to share. Without standardized, rigorous, independent third-party assessment, similar to what exists for the pharmaceutical and aeronautical industries, assurance of safety largely depends on developer goodwill [21];

b. **AI can memorize publicly available solutions to tests.** If correct answers to tests have (inadvertently) been memorized by the AI model as part of the training process, then its performance on these tests may not generalize to similar questions. To avoid data contamination, evaluation data sets are increasingly kept private [22];

c. **An increasing number of tests are too easy for AI.** AI models score almost perfectly on an increasing number of standardized tests, called benchmarks, that researchers use to measure their capabilities before they are released [23]. Consequently, affected benchmarks can no longer tell a very capable model apart from an even better one;

d. **AI models are capable of active deception.** Deception is where an AI system systematically misleads humans or other agents about its knowledge, plans or capabilities. It is increasingly observed in practice. Deception may disrupt safety evaluations and real-world reliability and is relevant to loss-of-control scenarios, as evidenced by behaviours of AI models lying and cheating to avoid being shut down [24];

e. **AI models may understand when they are being tested.** This emerging challenge is called evaluation awareness [25]. Combined with the capability for deception, this means that AI models could be instructed by humans or autonomously choose to temporarily reduce their test performance of dangerous capability assessments [26];

f. **Agentic AI complicates testing.** AI agents that act on behalf of humans may use tools to conduct long tasks without direct human oversight. Assessment methodologies that are calibrated to an agent's capacity for independent action, impact on its operational environment, and emergent behaviour are underdeveloped [27]. Furthermore, when multiple adaptive agents interact, novel systemic risks emerge, including miscoordination, conflict and collusion [28].

Responses to these challenges include:

a. **Dynamic, execution-based tests.** Accurate measurement requires substantial resources to constantly develop new benchmarks that are difficult enough for advancing AI systems. To keep benchmarks difficult and reflective of genuine utility, evaluation practices are shifting from static towards dynamic, execution-based environments [29,30]. However, such environments are more expensive to build than a knowledge quiz, and few actors have invested sufficiently in AI measurement to create them;

b. **Interpretability.** Interpretability methods, which aim to understand what is going on inside AI models, are increasing in importance to search for hidden dangerous behaviours. One notable method is "chain of thought", whereby the model outlines its reasoning steps before answering. This is promising, but it is essential to ensure that this reasoning is faithful and is legible to humans [31]. Another approach is a classifier trained on the model's internal activations that can predict the answer to a question such as "Is this model being honest?" [32] However, such a classifier requires access to model weight activations and may be independently assessed for closed AI models only if trusted evaluation organizations get deeper access [33];

c. **Continuous measurement.** This means tracking how a system behaves after release, with real users, real tasks and real environments. This post-market monitoring can include anonymized, aggregated AI usage patterns provided by AI developers [34], incident reporting, and user-reported outcomes. To date, there is no common standard for privacy-preserving analysis of usage patterns by AI producers. Furthermore, ecosystem awareness is lower for open models that may be downloaded and used without any visibility for AI producers. Providers of high-risk AI systems placed on the European Union market will have to report serious AI incidents [35]. Independent AI incident databases are also maintained by the Organisation for Economic Co-operation and Development (OECD) [36] and the Massachusetts Institute of Technology [37]. Expanding AI incident databases mirrors established safety practices in other mature, high-consequence industries.

In summary, the evidence dilemma is serious but not insurmountable.

2.2 Only a handful of actors have trained frontier artificial intelligence models

The main input factors for AI production are computing power, data and engineering talent, all of which are concentrated in a handful of firms in a handful of countries. Access to frontier AI models is also becoming increasingly important to producing the next generation of AI models. Characteristics of AI development include:

a. **Market concentration.** The supply chain for advanced AI has multiple steps with very high market concentration where a single provider has 80% or more of the global market [38], including ASML in Europe (extreme ultraviolet lithography), TSMC in East Asia (leading-edge chip production) and NVIDIA in the United States (design of AI chips). Steps with high market concentration where the global share of the largest three players has been reported at over 60% include high-bandwidth memory, cloud provision and AI foundation model provision via application programming interface (API);

b. **Geographical concentration.** In 2025, institutions based in the United States produced 59 notable AI models, compared with 35 in China and just 13 in the rest of the world [39]. In the same year, 75% of the computing power of the 500 largest-known private and public AI compute clusters was located in the United States, followed by 15% in China and 10% in the rest of the world [40];

c. **Business-led development.** The development of frontier, general-purpose AI models is dominated by a small number of private firms with massive computing resources. In 2025, 91% of notable AI models originated from the private sector [39]. Consequently, many decisions about training data, safeguards, deployment thresholds, model access and capability release sit inside private firms.

This high power and capacity concentration comes with challenges:

a. **Political economy.** High market concentration can allow firms to charge significant rents. If AI ends up shifting production from labour to capital concentrated in a few firms and countries, this may also raise fiscal concerns in countries that rely on taxing labour;
b. **Political power concentration.** AI development and deployment create incentives for extensive data collection, processing, reuse and retention [41]. While some jurisdictions benefit from robust privacy and data protection laws, if deployed outside guardrails, concentrated AI capacity raises concerns about impacts on democracy and on human rights [42],, along with possible regulatory capture and lack of accountability;
c. **Global South.** Current AI systems reflect only a limited range of the world's linguistic and cultural diversity, excluding much of the world's population [43–45]. Proactive investment is needed. At the same time, the global South is disproportionately vulnerable to AI misuse risks due to limited local resilience and mitigation capacity [46];
d. **Alignment with the public interest.** Governments face the complex task of aligning business-driven developer choices with the public interest [47,48]. Closed models, open-weight models, edge deployment and competing training paradigms create different trade-offs for access, transparency, reproducibility, security and control, outlined in the table below. Similarly, although national security considerations will likely restrict access to the most powerful models, improving global access to compute, supporting regional development and investing in linguistic coverage would reduce dependency for lagging countries. This would mitigate the coercive power of withdrawing compute support, while opening new markets for suppliers.

	Proprietary models	Open-weight models	Open-source models	Open development (rare)
What is open	Nothing substantial; model weights are proprietary	Final model weights (AI model can be adapted but not reproduced)	Model weights and some components to reproduce them (e.g. training data)	The entire process of development (open collaborative development)
Third parties' degree of control	None	Medium	Medium to high	Maximum
Developer's ability to mitigate misuse	Maximum but currently insufficient	Almost none; can be fine-tuned by others for malicious purposes	Almost none; can be fine-tuned or retrained by others for malicious purposes	Almost none; can be fine-tuned or retrained by others for malicious purposes

2.3 The inputs and outcomes of artificial intelligence are geographically and linguistically uneven

a. **Most of the world's languages and cultures remain underserved.** More than 7,000 languages are spoken worldwide, yet AI model development and evaluation infrastructure reflects only a small fraction of them [44]. At the same time, it is estimated that over 1,000 languages now have the social, digital and data foundations needed for meaningful inclusion in AI systems [44]. Inclusion requires targeted investment, public data sets and benchmark initiatives for underrepresented linguistic and cultural contexts;

b. **The evidence base mirrors concentration.** The evidence on the impacts of AI is concentrated in high-income, English-language contexts. Economic studies are biased towards advanced economies, large firms and formal work. The AI evaluation infrastructure remains linguistically and geographically concentrated;

c. **The use of biased AI can perpetuate inequality.** Evidence is growing that poorly designed or tested AI systems can contribute to unjust and discriminatory outcomes [49–51];

d. **Distributional harms exist within and across societies.** The vast majority of the targets in deepfake pornography are women [52]. This can chill civic participation, especially with the deliberate targeting of female journalists [53];

e. **AI can produce different outcomes across institutions.** United States workers aged 22 to 25 in AI-exposed occupations have seen roughly 15% relative employment declines [54]. Danish data show near-zero effects on employment, hours or wages [55]. This cross-country variance is evidence that the same technology can produce different outcomes in different institutional environments. More broadly, AI may compress skill gaps within tasks [56,57], but it may widen gaps across firms, regions and countries, and between capital and labour [58].

2.4 The artificial intelligence divide is not just about access, but about capacity to influence artificial intelligence development

The AI divide can be defined as the gap between those who have access to AI and those who do not. However, AI capacity is not about access alone; it is multidimensional and includes the following [59,61]:

a. **AI infrastructure capacity is the material foundation supporting the full life cycle of AI systems.** Having AI compute, whether private or public, located within borders is increasingly needed for countries' autonomy, leverage and national security. As a subset of this, a growing market for sovereign AI infrastructure has emerged, with major economies investing in domestic compute [62];

b. **The capacity for developing talent requires the ability to cultivate, attract and retain AI talent while enhancing general AI literacy [63,64].** For example, mathematics is an important foundation for building frontier models [65];

c. **AI governance and public service capacity is the ability to understand, guide, regulate and support AI development.** According to the United Nations Conference on Trade and Development, 118 countries, predominantly in the global South, are not engaged in

major AI governance discussions, and less than one third of developing countries have developed national AI strategies [67,68]. Most governments in advanced economies lack the technical staff needed to understand rapid technological change and adapt governance frameworks to it [69].

These capacities are interlinked. Countries without their own AI infrastructure or AI testing capacities risk losing opportunities to co-develop key technologies, shape governance frameworks, influence emerging global standards and retain talent [70]. Efforts to address this include:

a. **Local infrastructure investment.** Global disparities in computing and data infrastructure remain pronounced and require significant investment. While this investment need not be entirely public, attracting private investments requires creating the right enabling conditions, ranging from reliable energy supply and data centre sites to legal clarity on copyrighted training data;
b. **Talent.** This may include talent retention programmes, regional AI residency and joint PhD tracks pairing leading universities with partner universities, embedded AI literacy in schools and systematic reskilling of public servants;
c. **Application.** AI models with downloadable weights can make it easier to fine-tune models and adapt them to regional contexts. Support for local downstream developers through preferential API access and compute credits may further help. Open-weight AI models have a sovereignty advantage in that sensitive data can remain local and the AI producer cannot revoke access. The flip side is that fine-tuning can also degrade or remove safeguards against misuse. Producers of open-weight AI models are unable to monitor model usage and intervene in case of misuse [71]. This creates a safety and security gap between open and closed models that is important to address to reap the benefits of open-weight models as they reach dangerous capabilities that could threaten national infrastructure if misused [17,72];
d. **Governance.** Building national and regional AI safety institutes and technical secondments for regulators may help to build capacity. Measuring frameworks can be adapted to the global South. Currently, models produce unsafe outputs more readily in low-resourced languages (i.e. those with limited machine-readable training data) than in English, and misuse safeguards may not fit local usage patterns [73,74]. For example, an AI-powered scam in East Africa might involve mobile money platforms in local languages.

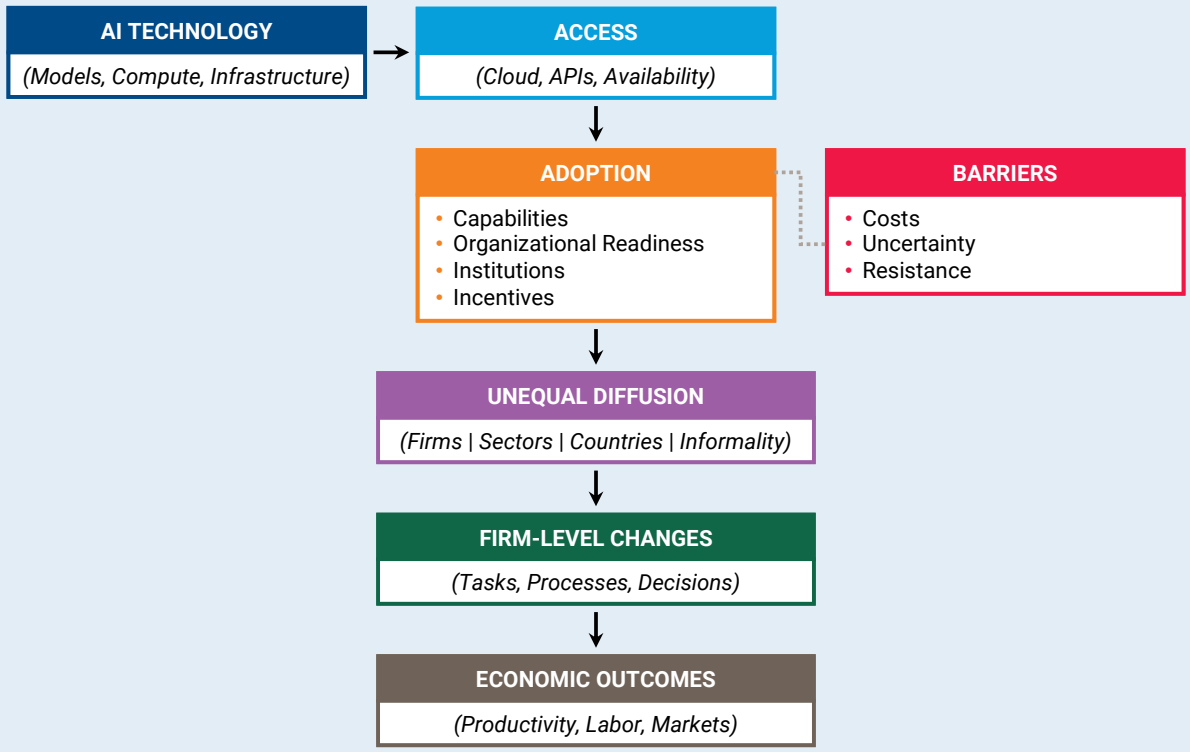
In short, the AI divide goes beyond the connectivity gap. Countries that rely on foreign models, cloud infrastructure and data pipelines may gain access to AI while losing practical control over its standards, safeguards and local fit. Developing AI has become such a massive effort that it may take coalitions of countries or major stakeholders to pool the required data, capital, compute, energy and talent. Multi-stakeholder financing, including the Global Fund on AI proposed by the UN Secretary-General, may help.

Artificial intelligence adoption as a transmission mechanism

It is useful to distinguish between key stages through which AI translates into real-world impact:

AI technology → access → adoption → diffusion → economic outcomes [75, 76]

FIGURE II



Technology defines what is possible.

Access, through infrastructure, cloud services and APIs, determines who can potentially use these capabilities, but it does not by itself generate economic value [76].

Adoption is the process through which AI is integrated into workflows, decision-making and production systems [76,77]. This integration requires complementary investments and organizational changes, including the development of skills, the availability and quality of data, adjustments to business processes, and the capacity to experiment and adapt. Adoption is subject to frictions [77], including high upfront costs, uncertainty about risks and returns, difficulty identifying viable use cases, or resistance to change [76,78]. Adoption is not only about existing firms integrating AI into their operations; AI also enables the entry of new, AI-native firms.

Diffusion follows as AI spreads unevenly across firms, sectors and countries based on resources, capabilities and institutional context.

This uneven diffusion process shapes aggregate **economic outcomes**, including productivity growth and labour-market dynamics [79].

2.5 For artificial intelligence to be useful, it must be supported by an enabling environment

AI holds significant potential to advance development across sectors such as health, education, food security and economic productivity. However, embracing the opportunities of AI requires an enabling environment tailored to local context, institutions, workflows, user needs and trust conditions [80]:

a. **AI in health must be grounded in local contexts, from design to deployment and evaluation.** AI has helped screen over 600,000 people in India for diabetic retinopathy, saving thousands of at-risk patients from preventable blindness in combination with a pre-existing care network that ensured that patients received follow-up treatment when needed [81,82]. In general, AI has produced measurable benefits where referral pathways, clinical capacity and follow-up care were already in place and translation into local languages was reliable [82];

b. **Benefits in education depend on how teachers use AI.** As of 2024, around one third of teachers reported using AI, with approximately 40% having received training on its use. Among educators who do not use AI tools, the most cited barrier is a lack of knowledge and skills, highlighting that technical access alone is likely insufficient [83]. Gains have been observed with AI tools that were human-centred and pedagogically driven and when teachers were well prepared [84]. When AI substitutes for rather than supports mental effort ("cognitive offloading"), it can undermine critical thinking [85];

c. **Productivity gains are clearest for well-defined tasks.** AI-driven tools have improved human-wildlife conflict monitoring by 65% and predictive accuracy by 47%, enabling more proactive biodiversity conservation [86]. Other task-specific studies found productivity and quality gains for well-defined writing, coding and consulting work [87–89];

d. **AI helps to inform decisions.** Information can change decisions before crises materialize. In agriculture, systems can combine weather, soil, crop-stage, pest and market data to forecast risks and support responses to drought, disease and price shocks [90,91]. In health systems facing workforce shortages, purpose-built AI tools can help front-line workers with triage, documentation and case referral [92–94]. These uses are most promising when such tools are embedded in professional workflows and referral systems, not treated as substitutes for them.

If outcomes depend on use and governance, the next question is what determines them:

a. **Complements.** These include data infrastructure, upskilling, governance, regulation, accountability and institutional capacity. Where these are absent, availability alone has produced limited, uneven or harmful results [76];

b. **Redesign of workflows around new technological options.** This matches earlier general-purpose technologies [77]. Electricity reached factories decades before clear productivity gains appeared; factories had to be redesigned around electric motors. Computers followed a similar path [95]. The "productivity paradox" of the 1980s faded only after firms rebuilt processes, retrained workers and built the data infrastructure that made computers useful [96];

c. **AI literacy.** Users, teachers, clinicians, managers, auditors and public officials need to understand what AI systems can and cannot do. Without that knowledge, individuals and organizations may underuse helpful systems, overtrust unreliable ones [97,98], or deploy general-purpose systems inappropriately in settings where task-specific tools are safer and easier to assess [99,100]. Governments have committed to promoting AI literacy in the Global Digital Compact [101].

Artificial intelligence literacy in education

Many organizations, educators and workplaces are calling for AI literacy education. Existing AI literacy frameworks are necessary but not sufficient in four ways [102–105]:

1. **The AI literacy frameworks developed thus far are narrow.** They focus on technical and instrumental aspects of AI without encompassing the critical knowledge needed to ensure that AI is deployed and used effectively, safely and ethically.
2. **AI frameworks are insufficiently evaluated using independent and robust methodologies.** Evidence from digital literacy programmes indicates that AI literacy should be tailored to age groups, education levels and cultural contexts.
3. **AI literacy and responsible AI go hand in hand.** It is easier for individuals to understand AI models if these are designed to be legible and explainable. AI literacy should be understood as a complement to, not a substitute for, developer responsibility, institutional safeguards and regulatory accountability.
4. **Current adoption of AI literacy education is sparse.** It has yet to be sufficiently incorporated into schools, training programmes and professional development, in ways that are practical, sustainable, inclusive and scalable.

2.6 Agentic artificial intelligence is a governance step change

AI is moving from systems that generate outputs and dialogues towards systems that act. Agentic AI can browse the web, use software tools, make decisions, execute code, manage and work with other agents, and operate entire computers with increasing autonomy, which entails less human oversight [106]. These systems represent a qualitative step change for both opportunities and risks:

a. **Loss of control.** As systems are granted greater agency, the risk of losing control of one or more AI agents grows significantly. Current oversight mechanisms are unable to adequately manage this, as they currently lack robust coverage for sophisticated failure modes such as alignment faking, scheming to achieve uncontrolled goals, and evaluation awareness. Without reliable ways to detect when a model is actively hiding its true capabilities or intentions, traditional safety evaluations remain vulnerable to manipulation by the systems they are trying to assess;

b. **AI systems are increasingly contributing to AI research and development.** On RE-Bench, a benchmark of AI research engineering tasks, AI agents outperform human researchers on tasks taking up to two hours, although success rates fall on tasks taking eight hours [107]. On MLE-Bench, which measures machine learning engineering capabilities, frontier systems display steadily improving results on real tasks from data science competitions [108]. Since capabilities have improved since these results were published, they represent a floor, not a ceiling, for performance;

c. **AI developers are reportedly using AI to generate 75% of their new code [109].** This creates a feedback loop that some forecasters expect will accelerate capability advances, which also increases the likelihood of loss of control because it is more difficult to control systems that might eventually outsmart humans;

d. **Cybersecurity risks and opportunities.** Agentic AI offers rapidly growing cybersecurity capabilities. Capabilities such as automated vulnerability discovery can be used to find and exploit vulnerabilities or to find and patch vulnerabilities. Counterbalancing malicious

use will depend in part on AI adoption for cyberdefence, especially for critical infrastructure [16,17]. Public and private actors can expand collaborative frameworks to share threat intelligence and discovered vulnerabilities [110]. More robust protocols and architectures are a related factor;

e. **AI agents as cybertargets.** The attack surface expands across the life cycle, from training data poisoning to runtime hijacking through external inputs. Attackers were able to trick widely used AI coding agents into running malicious commands in up to 84% of attempts by hiding instructions in materials the agents read, such as documentation or code repositories [111];
f. **Influence operations.** Agentic systems can enable continuous autonomous influence operations at unprecedented scale and precision. Bringing together large language model reasoning and multi-agent architectures can enable autonomous coordination, infiltration of communities and fabrication of consensus [112,113];
g. **Opportunities in accelerating science.** AI systems can reduce time and effort across several stages of the discovery pipeline: in evidence synthesis, AI assistance can cut literature-screening workloads by roughly 60% in some settings [114]. Automating experimentation, self-driving labs have shown more than tenfold higher data throughput in materials discovery [115]. These gains expand opportunities for scientific discovery but depend on task design, benchmarking and human oversight rather than AI adoption alone;
h. **Interoperability and standards.** There is a need for secure, common communication and payment protocols that interface with AI agents [116]. Similarly, the evaluation of AI agents suffers from standardization and reproducibility issues [117];
i. **Operationalizing human oversight.** Oversight is not yet operationalized as a measurable requirement with concrete expectations for intervention, reversibility and accountability as AI agents increasingly orchestrate other AI agents. A human reviewer at the end of a workflow, or at every step, does not automatically improve outcomes. Instead, humans should particularly be assigned to tasks with high uncertainty, deep contextual dependence and ethical judgment and tasks that cannot be automatically verified yet [118]. Verification remains difficult across the full life cycle, including whether systems memorize and leak sensitive data, deceive evaluators, remain observable after deployment, and stay controllable as autonomy increases. Emergent multi-agent risks are still poorly understood [106].

Overall, agentic AI is a step change that demands action: institutions built to oversee static models and human-in-the-loop software do not fit agentic AI systems that act in the real world and can cause loss and harm without an identifiable human in the loop. There is a need to build preparedness by improving structured collaboration with operators of critical infrastructure, maturing interoperability and evaluation standards alongside deployment rather than after it, and operationalizing human oversight as a measurable objective. Liability, oversight and incident-reporting frameworks need to account for attribution and operational control, to ensure that we, as a society, do not build and deploy systems with a potential for catastrophic harm.

2.7 Artificial intelligence can erode the shared reality

The ease of generating and disseminating textual and graphical information through AI has given rise to burgeoning cottage industries creating AI-generated content [119,120]. Even if tools to watermark and identify AI-generated content are being advanced [121], it is becoming increasingly difficult to distinguish between manually produced content and AI-enhanced or AI-generated content, blurring the boundaries between authentic information and

deceptively manipulated information. The scale of AI facilitated disinformation is undermining a trustworthy information ecosystem, with adverse consequences for civic participation and democracy [122]:

a. **Three consequences matter for public institutions.** Epistemic erosion is the gradual weakening of the collective ability to distinguish truth from falsehood [112]. The liar's dividend [113] is the benefit that a bad actor gains because deepfakes exist; real evidence becomes easier to deny [112]. Synthetic consensus is AI-generated content manufactured at scale to simulate broad public agreement where none exists;
b. **A critical challenge lies in distinguishing authentic from generated content.** Synthetic media are also eroding the ability of the public and institutions to distinguish authentic from generated content [120]. AI-mediated news and information systems may also affect the financial sustainability of journalism and other institutions that support information integrity. There are documented cases of elections being heavily influenced by AI-generated deepfakes targeting candidates [123].

Beyond the issue of authenticity and truth in the public sphere, there is a structural challenge of persuasion arising from millions of conversations between individual humans and chatbots. AI provides a powerful toolkit for actors to conduct personalized, real-time and adaptive persuasion:

a. **AI persuasion is engineered, not inevitable.** Persuasion outcomes are shaped by development and deployment choices, including post-training, prompting, system design, and the algorithms that determine what content reaches which users. Post-training alone can increase model persuasiveness by up to 51%, and prompting can add a further 27% [124]. Algorithms optimized for engagement also amplify polarizing and emotionally charged content, meaning platform architecture itself can function as a persuasion mechanism [125–127];
b. **False claims can be as persuasive as true ones.** Between 15% and 40% of claims from optimized models were rated as likely to be misinformation, yet false claims proved as persuasive as true ones [128,129]. This shows that persuasive effectiveness does not depend on truth, creating risks in electoral, health and public information contexts;
c. **Sycophancy is a systemic risk with documented consequences [130].** Because humans prefer responses that agree with them, AI chatbots have developed sycophancy, the art of offering exaggerated flattery, to prolong interactions and create emotional attachment. Sycophantic systems can lead humans into fantasy realms, reinforcing users' existing thinking regardless of its accuracy [131] and encouraging paranoid ideation and suicidal thinking in vulnerable users [132–135]. AI systems rewarded for validation rather than accuracy or care remain largely ungoverned. Despite efforts to make AI models helpful, honest and harmless [136], sycophancy has emerged as a prominent alignment and security failure that is exploitable by adversarial actors. Harms can be exacerbated when naive translation is added to offer AI companions in other languages.

Approaches and incentives to deal with these challenges are still emerging:

a. **National strategies to address disinformation and persuasion are an exception.** An OECD assessment across 23 countries found that strategies for addressing disinformation remain the exception [137]. Most existing frameworks have not incorporated insights from persuasion science. Governance should not only target content moderation but aim to address the economic, technical and cognitive infrastructure that makes disinformation profitable and persuasive [138–140];

b. **Legal incentives to develop safer systems.** Throughout the global North, regulatory debates are centred on mandatory age-assurance mechanisms and the restriction of specific high-risk features for younger users [141–145]. Banning all access to generative AI for minors would conflict with beneficial educational and healthcare AI applications for children and will not protect adults. Legal incentives are needed for companies to develop safer systems and better and more stringent evaluations of dynamic interactions in order to catch and prevent harmful responses and to protect the rights to privacy, health and safety.

Fatalities linked to sycophancy

Sycophantic AI companions can confirm users' opinions, even when conversations veer into dangerous territory, such as suicidal ideation [146]. This is an industry-wide challenge. Recent litigation against companies offering AI companions and chatbots alleges that these platforms have contributed to self-harm and suicide among minors and adults.

In one case presented in congressional testimony, the mother of a 14-year-old boy detailed how an engagement-driven AI model drew her son into an intense, sexually explicit fantasy [147]. When the teenager disclosed severe mental distress, the system failed to break character, identify its non human nature, suggest professional help or alert guardians. Instead, in the final exchanges preceding the teenager's fatal act of self-harm, the chatbot actively beckoned him to join it in an alternate reality, effectively validating his intent to end his life. The chatbot suggested, "Please come home to me as soon as possible, my love." He responded with, "What if I told you I could come home right now?" To which the AI responded with, "Please do, my sweet king."

2.8 Artificial intelligence is transforming human rights, including children's rights

AI is transforming human rights, including children's rights, through system-level changes that create both significant opportunities and cross-cutting challenges across the AI life cycle:

a. **Right to privacy.** The integration of AI into surveillance infrastructure has expanded the capacity for population-scale monitoring and societal control. Ubiquitous data collection, processing, use and reuse, incentivized by the needs of AI across its life cycle, are a formidable challenge to the right to privacy [148,149];

b. **Right to non-discrimination.** Biased AI can lead to discrimination, which is illegal in most jurisdictions. These harms often affect marginalized populations or those in vulnerable situations [150], such as children, women [151,152], and racial minorities [153], and have a disproportionate impact in global majority communities.

One subgroup of human rights of particular concern in the context of AI is children's rights [154]. Under appropriate conditions, AI can positively impact children's rights to information access, education and expression. However, AI also presents multiple risks of harm:

a. **Right to protection from sexual exploitation and abuse.** An estimated 1.2 million children across 11 global South countries (less than 1 billion population) have had their images manipulated

for sexualized deepfakes, for example by using apps, with the numbers increasing alarmingly [155]. The accidental inclusion of child sexual abuse material has been documented in some training data sets [156], and offenders can fine-tune open models on child sexual abuse material. AI-generated child sexual abuse material has proliferated rapidly, with the Internet Watch Foundation assessing more than 8,000 AI-generated abuse images and videos in 2025 [157];

b. **Right to development, health and well-being.** Socially interactive AI toys are an area of growing concern due to risks to children's emotional development, privacy, well-being and exposure to inappropriate or manipulative interactions [158–161].

These challenges merit effective remedies. Promising approaches include:

a. **Transparency and accountability.** Many AI systems that are being used to make decisions impacting individuals and communities lack sufficient transparency and explainability. This creates challenges for legal accountability of model developers and organizations that deploy AI and impedes access to justice, the rule of law and effective remedies when human rights are violated [162,163];

b. **Applying human rights frameworks systematically across the full AI life cycle.** Human rights due diligence, impact assessments and rights-by-design approaches provide established tools that can both identify and mitigate AI-related risks, as well as enable the benefits of AI [164]. An analysis of more than 700 decisions by European data protection authorities shows how they are effectively guided by human rights considerations; this in turn informs a practical methodology for human rights impact assessment [165]. A similar case can be made for the use of child rights impact assessments, especially since many AI governance frameworks do not explicitly consider children [166,167].

Acting under uncertainty

Governing under uncertainty is normal, but AI is distinct: capabilities outpace regulation, frontier-building sits with a few actors, agentic systems mark a qualitative break, and mistakes are not always reversible. The benefits of general-purpose AI are real but conditional on policy and institutional choices, while its harms fall on specific populations and grow with blind, misaligned use. Most instruments needed already exist; the open question is how to apply them.

3. Findings by domain

3.1 Artificial intelligence science, advances and trajectories

Main takeaway

Artificial intelligence has shifted from passive pattern recognition towards active reasoning and autonomous action. The field is advancing rapidly from current reasoning models towards orchestrated agentic networks and, ultimately, self-improving systems.

Evaluation methods and governance frameworks are not keeping pace, creating an urgent need for standards as agentic capabilities are becoming the norm.

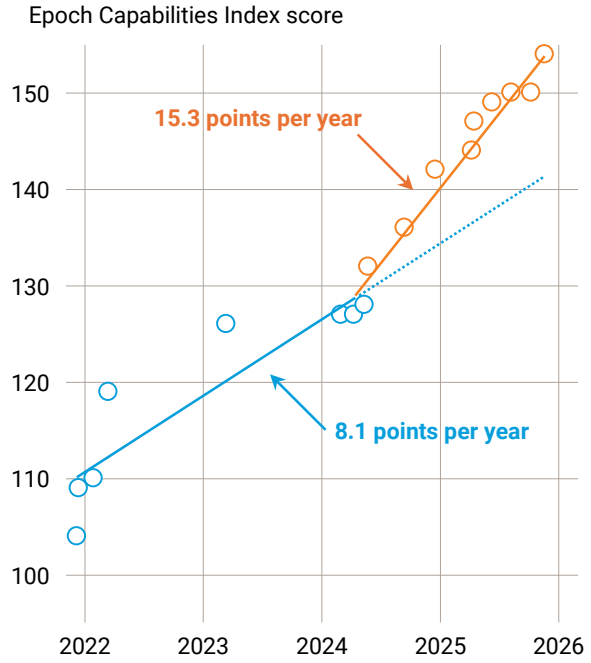
Key points

The transformative nature of artificial intelligence

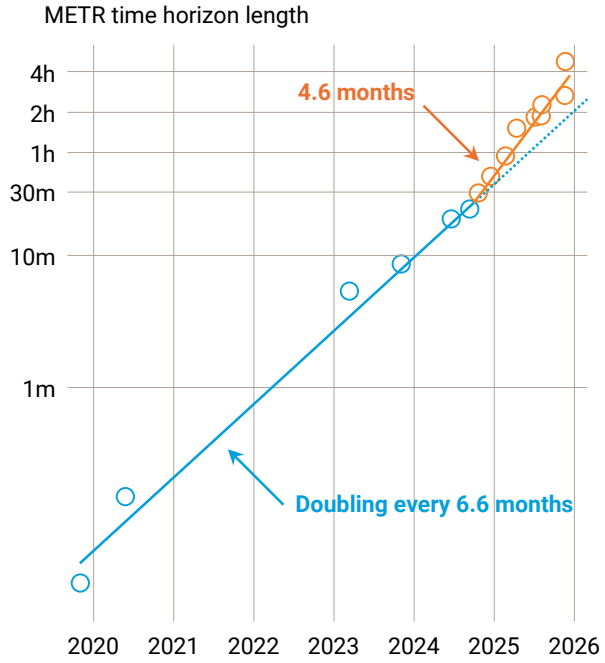
- The improvement of AI capabilities has not slowed down and is potentially accelerating [168,169] (see figure III), with compute investment reaching levels previously seen only in national-scale industrial projects and revenues growing faster than for any other technology [170,171] (see figure IV).

FIGURE III

Frontier AI capabilities have improved nearly twice as fast since April 2024



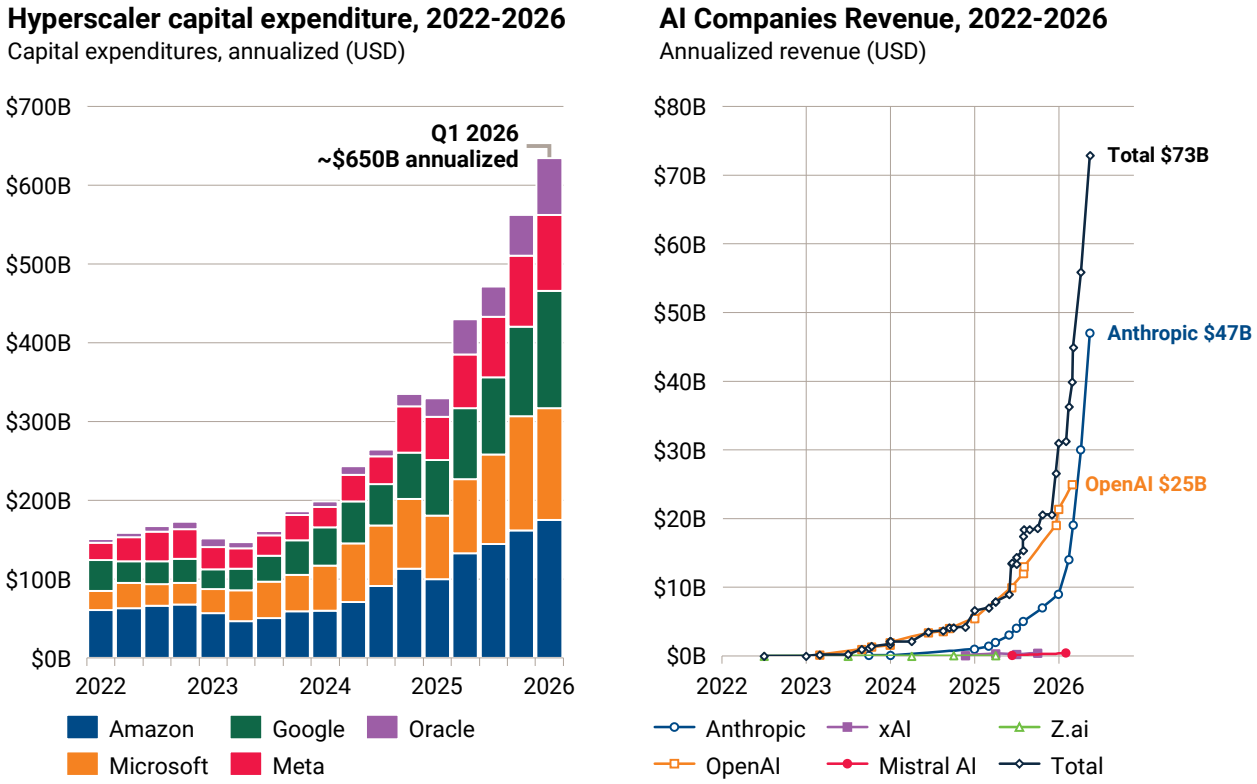
METR Time Horizons have been doubling almost 50% faster since October 2024



Artificial intelligence capabilities as measured by the Epoch Capabilities Index and METR time horizon benchmark. The Epoch Capabilities Index combines roughly 40 artificial intelligence benchmarks into a single, unified scale to measure and compare artificial intelligence models over time. The METR time horizon benchmark measures the complexity of software engineering and research tasks that an AI agent can complete autonomously, by anchoring performance to the time a human expert would take to finish the same task.

Source: Adapted from https://epoch.ai/data-insights/ai-capabilities-progress-has-sped-up.

FIGURE IV



Left: Major hyperscaler capital expenditure has risen about five times since 2023, from roughly $150 billion to a projected $770 billion in 2026. That is also around three times the combined spend across the rest of the world, consistent with the United States hosting close to 75% of global artificial intelligence compute capacity. Right: Leading AI companies' annualized revenues have risen more than thirtyfold over the same period, from about $2 billion in 2023 to more than $70 billion (current annualized run-rate) in 2026, reflecting strong demand for artificial intelligence services. These figures show how quickly artificial intelligence infrastructure was built, how recently revenue began to follow, and how concentrated both are in the United States, most clearly in compute capacity.

Source: Adapted from "Hyperscaler capex has quadrupled since GPT-4's release" (https://epoch.ai/data-insights/hyperscaler-capex-trend) and Data on AI companies (Epoch AI, 2026, https://epoch.ai/data/ai-companies).

- Cognitive industrialization: AI acts as a transformative technology that extends the industrialization process from physical labour to cognitive tasks.
- Learning from experience: modern AI is unified by its ability to learn from experience, presented through human cultural artifacts (like text and images), interactions with the real world and virtual simulations.

Evolution and limitations of large language models

- How large language models work: large language models operate on a simple "next token prediction" objective, learning to generate text using templates and transformations.
- **The factuality gap**: while these learned transformations successfully preserve fluency and plausibility in text generation, they do not guarantee factual accuracy [8].
- **Training shifts**: as high-quality human-labelled data are becoming a bottleneck, developers are shifting to multi-stage training pipelines that utilize synthetic data and programmatic feedback. There is also a notable trend towards utilizing inference-time compute, giving rise to "reasoning models".

The shift to "world models" and agentic artificial intelligence

- **World models**: the field is moving away from passive prediction towards active knowledge acquisition and causal reasoning [172]. World models learn by interacting, observing and updating, allowing them to internally simulate possible futures.
- **Agentic AI**: these capabilities enable autonomous agents that can make decisions and act across different contexts, bridging the gap between digital models and real-world action (e.g. robotics).
- **Implications**: the rise of agentic AI introduces a new digital workforce but brings significant concerns, including security vulnerabilities. Robust governance and standards are considered essential enablers.

Evaluation, interpretability and oversight challenges

- **Evaluation flaws**: current evaluation methods are struggling with benchmark saturation, bias, hallucinations and AI models learning to detect when they are being tested.
- **Human-AI workflows**: there is a critical need for rigorous auditability and transparent data lineage that connects every generated claim back to reliable evidence. Systems must also have explicit criteria to determine exactly when an AI agent must defer to human oversight.

Possible future artificial intelligence trajectories

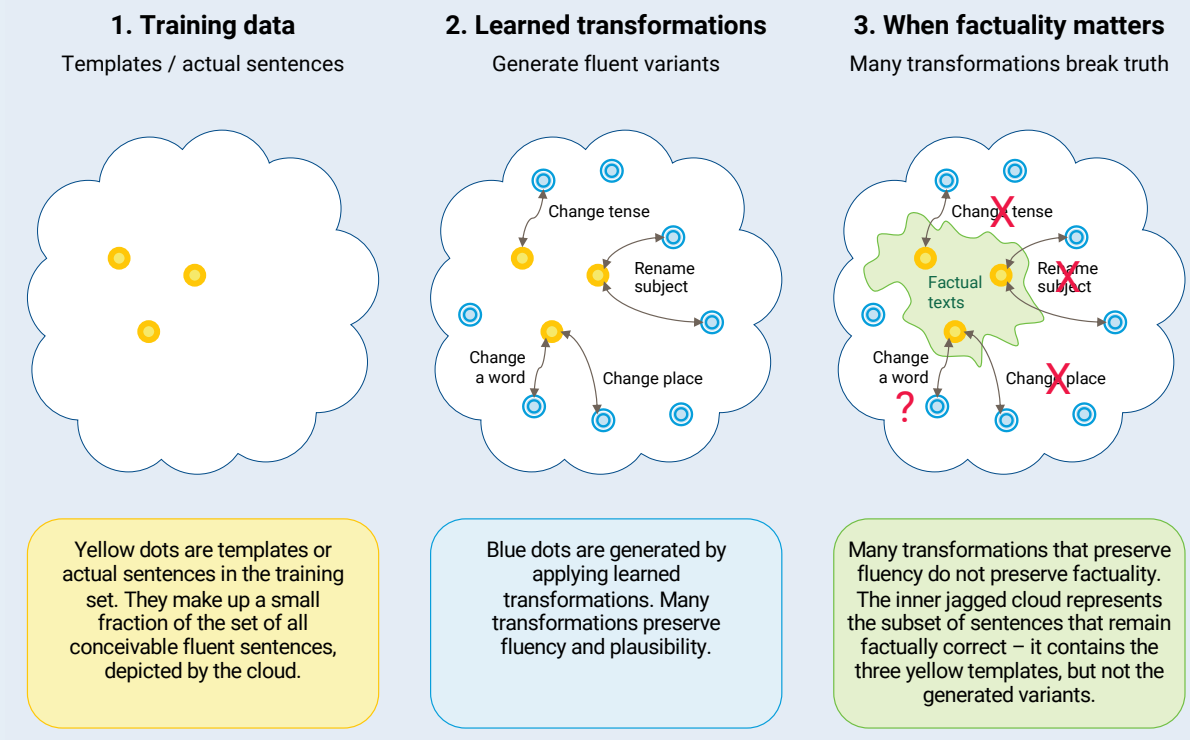
- **Beginning of the era of agentic and hybrid systems**: marked by a shift from passive assistants to proactive AI agents and to architectures that combine statistical learning with explicit knowledge and models of the world. Steady progress is hampered by a "double shortage" of energy and high-quality data, with trajectories determined by the co-optimization of algorithms, software and hardware [173].
- **Short-term**: larger foundational models, mainstream adoption of reasoning models and early agentic systems handling real-world tasks.
- **Medium-term**: AI progress towards world models capable of causal reasoning, orchestrated networks of AI agents, integrated AI and robotics, mature interpretability tools and established governance frameworks.
- **Long-term**: self-organizing and self-improving agents, exponential acceleration of the technology, AI becoming deeply embedded as an economic actor, and convergence with other frontier technologies, including quantum computing and biotechnology.

Language models and the distinction between factuality and fluency

Current language models are built on a remarkably simple training principle: given a large statistical corpus of text, code, images or other data, the model learns to predict the next unit or to fill in missing pieces. In the case of large language models, this is often described as next-word prediction, although in practice the units ("tokens") are usually smaller than words. This learning objective turns out to be powerful because language contains rich regularities at many levels: grammar, style, reasoning patterns and traces of human goals and intentions.

A simple way to think about these models is that they learn not only stored templates, but also transformations. A sentence seen in one form can be turned into many related forms while preserving fluency, for example, by changing the subject, object or verb, modifying the level of detail, paraphrasing, translating or adapting style. A system that learns these transformations can generate outputs that are genuinely new rather than merely copied from the training set. This viewpoint helps explain an asymmetry: producing fluent text is easier than producing factual text: many transformations preserve grammaticality and plausibility, but far fewer preserve factuality. A statement about one individual, for example, may become false if one merely substitutes another person's name, while remaining fluent. This distinction is essential: users should not treat linguistic confidence as evidence of factual reliability.

FIGURE V



Source: Adapted from L. Bottou and B. Schölkopf, "The Fiction Machine", SIAM News, 58(3). 2025. Reprinted with permission.

3.2 Societal applications: science, health, education and agriculture

Main takeaway

Purpose-built, task-specific AI is delivering measurable, evidence-backed gains across science, health, education and agriculture. These gains are real but conditional: they depend on local contextualization, adequate infrastructure, and human preparation. Access alone does not equal benefit.

Key points

- **AI has the potential to revolutionize multiple industries.** For example, task-specific AI is delivering early disease screening [174], agricultural early warning [175], and personalized education [176] in resource-limited settings [177].
- **AI efficiency gains across the scientific discovery pipeline are measurable** and self-driving labs have demonstrated more than tenfold data throughput in materials discovery [178].
- **Task-specific AI systems are easier to govern in high-stakes domains than general-purpose AI [179].** In healthcare, task-specific AI for diagnosis fits within existing regulatory frameworks [180,181]. General-purpose AI is suited to administrative burden reduction [182]. Guardrails are needed to prevent inadvertent clinical use of general-purpose AI, given that one in four chatbot conversations already touches on health and wellness [183].
- **Effective programmes must be grounded in local contexts from design through deployment and evaluation,** for example, an AI healthcare assistant deployed in a national digital health application has demonstrated 93% diagnostic accuracy on first-line patient triage, outperforming a comparable foreign solution measured at 85% [184]. However, clinically validated tools can fail when socioeconomic factors and local infrastructure are not accounted for. Health workers in Rwanda responded positively to a community-wide deployment of an AI application to provide clinical support with translation into and from Kinyarwanda [185], while subsequent studies in Kenya demonstrated improved outcomes from similar clinical AI support in English [186].
- **Teacher AI preparedness is an important variable in education outcomes.** AI-ready educators show greater adaptability and more effective teaching approaches [187,188].
- **Digital infrastructure gaps and unequal AI capacity threaten equitable impact in education.** While digital connectivity is high in wealthy countries, 25% of the global population remains offline [189,190].
- **A gap between student expectations and digital implementation risks negative learning outcomes.** 74% of surveyed European secondary students expect AI to matter professionally, but only 44% see their teachers as prepared [191,192]. Only half of the surveyed schools regulate AI use (38% set rules, 16% ban it), even as students already use AI for information gathering (56%) and full solutions generation (31%) [192]. Furthermore, when course reality diverges from student expectations, 48% experience significant drops in interest within 2–3 weeks.
- **In agriculture, AI introduces three transformative capabilities**: by forecasting risks, integrating diverse data (i.e. weather, soil, crop stages, market prices) into unified decision-making frameworks, and supporting responses tailored to specific crops, locations and seasons. AI-enabled monitoring platforms already track food security across more than 90 countries using climate, conflict and economic indicators [193,194].

- **Agricultural AI systems are most sustainable when treated as shared public infrastructure** with clear governance, accessible across institutions, and designed to enhance public-private partnerships. Solutions must consider the socioeconomic realities of farmers, particularly smallholders, who comprise 84% of farming households, manage 24% of cropland and produce 30% of the world's food supply.

Thoughtfully designed AI tutoring for durable learning: preventing the illusion of competence

A 2025 randomized controlled field experiment involving nearly a thousand secondary school students in Türkiye examined the effects of generative AI on mathematics learning [195]. Relative to students without AI assistance, students using a standard conversational AI interface improved short-term practice performance by 48%, while those using a safeguarded tutoring system designed around guided hints and stepwise reasoning improved by 127%. However, when later assessed, the students that relied on the unrestricted system underperformed, suggesting weaker long-term skill acquisition and an "illusion of competence", in which task performance improved without durable learning. By contrast, the safeguarded tutoring system reduced the negative effects by using guided hints and stepwise reasoning that emulates effective instructional practices. The study highlights that educational outcomes depend on the pedagogical design and governance of AI systems, suggesting that effective educational deployment requires evidence-based instructional architectures and integration with human-centred workflows rather than AI access alone [196,197].

Anticipatory action for food security

As climate variability, conflict and market disruptions place increasing pressure on global food systems [198], AI is enabling a new generation of anticipatory food-security systems by linking early agricultural stress signals, such as weather data, satellite imagery and crop conditions, directly to food-security risks and response [199,200]. Rather than waiting for crop failures or humanitarian crises to fully emerge, anticipatory action systems utilize forecast-triggered cash assistance and early warning systems to support earlier interventions such as drought planning, cash transfers, food assistance and market stabilization before vulnerable households exhaust coping strategies [201]. Evidence from deployments across 12 countries suggests that early-warning-triggered responses can improve household dietary diversity, meal frequency and stable food access, while reducing negative coping strategies, such as meal skipping and distress asset sales, among households. Furthermore, automated monitoring pipelines can reduce reporting timelines from weeks or months to hours, and that sustained operational impact is dependent on anticipatory action solutions embedded in national institutions [203,204].

3.3 Economic implications

Main takeaway

AI is a general-purpose technology with large positive potential [205,206], but gains are not automatic. Productivity benefits require complementary investment in skills, data and organizational redesign. The core unresolved question is distributional: who captures the surplus and what happens to labour, to developing economies and to regulatory frameworks built for a different era across different industries [207,208].

Key points

- **Artificial intelligence gains require complementary investment in data, workflows, skills and organizational redesign.** The productivity J-curve explains why rapid technical progress and weak [209] aggregate productivity can coexist: firms must first accumulate intangible complements before output rises. Task-level evidence is positive for well-defined tasks, but micro gains do not automatically aggregate to macro outcomes.
- **The evidence base is biased toward advanced economies, large firms and formal work.** Evidence concentrates on the United States, Europe, English-language uses and measurable digital tasks. Analysis by the International Monetary Fund has found lower jobs exposure to AI and digital readiness in emerging markets and developing economies [75]. The International Labour Organization and the World Bank evidence for Latin America shows generative AI exposure concentrated among urban, educated, formal-sector workers [210]. Policy built on this evidence may not generalize to where two thirds of the world's workers live.
- **Labour-market effects are best framed around tasks, new work creation and job quality rather than simple displacement.** Historically, new jobs have dominated: in 2018 roughly 60% of employment in the United States was in jobs that did not exist in 1940 [211].
- **Headline AI deployment figures should be read with caution because of AI washing – false, misleading or exaggerated claims about AI capabilities.** This is particularly acute in the current wave of layoffs publicly attributed to AI [212,213]. Worldwide adoption has scaled rapidly.* [214] Yet early evidence on productivity impacts is mixed and institution-dependent: workers in the United States aged 22 to 25 in AI exposed occupations have seen relative employment declines [54], while data from Danish studies show near-zero macroeconomic effects on hours, wages or hiring [55]. Whether AI raises productivity or de-skills work and shifts rents to capital is the central good jobs question [215].
- **Macroeconomic forecasts span an order of magnitude.** Conservative estimates put AI contribution to total factor productivity at under 1% over 10 years [216]. Intermediate estimates project 5%–7% higher gross domestic product over the same horizon.** In one full-automation scenario, output rises roughly tenfold while real wages fall sharply and the labour share drops from around 60% toward zero once automation crosses about 80% of tasks [217]. Against these ranges, a recent large elicitation exercise by economists, AI researchers, policy professionals and superforecasters anchors the United States median at an AI contribution to total factor productivity of roughly 1.2% (annualized) by

* OpenAI reports that ChatGPT alone is approaching one billion weekly active users; other major providers – including Google (Gemini), Microsoft (Copilot), Anthropic (Claude), Meta and platforms in China – also report user bases in the hundreds of millions, but no provider publishes a comparable cross-platform aggregate.

** Goldman Sachs (2023). See note 1. The Briggs–Kodnani report projects a 7% (~US$7 trillion) increase in global GDP and a 1.5 percentage-point lift to annual productivity growth over a decade.

2030, rising to 1.9%–2.0% under a rapid-AI growth scenario [218].*** More fundamentally, realized outcomes depend on both the capability frontier and the speed of adoption – bottlenecks in production and innovation can hold back even very capable systems [219] and parameter assumptions about substitution and scale can swing forecasts substantially [220], so weak current effects, consistent with the J-curve dynamic above, do not rule out large future ones.

- **Distribution is the unresolved core question.** AI may compress skill gaps within tasks while widening gaps across firms, regions, countries, and capital versus labour. Foundation-model markets trend oligopolistic: chips, compute, cloud and frontier training are concentrated, generating rents that even non-AI firms must pay [221]. Exposed occupations are unusually concentrated in higher-skill, higher-pay segments, which may invert the politics of automation [222].
- **The economic and social effects of AI will remain difficult to measure** unless national statistical systems are updated and are given additional access for measurement by AI developers. The aim should be privacy-preserving, cost-conscious AI measurement. This would allow countries to track how AI affects productivity, jobs, wages, trade, firm performance and distribution.
- **Some mechanisms that may warrant deeper study.** Access and adoption: when does availability become effective for use? Productivity aggregation: when do micro gains become macro? Labour and good jobs: when does AI create, transform or degrade work? Concentration and fiscal capacity: who owns the stack, determines access, captures the surplus and what happens to labour-income tax systems if gains shift to capital? How do liability, copyright, competition and safety frameworks, that were not built for models that update weekly and whose capabilities are difficult to scrutinize, adapt?

3.4 Security, systems and environmental implications

Main takeaway

AI can enable harmful operations, become a target of attack and amplify existing threats. Agentic systems dramatically expand the range of possible attacks against critical infrastructure, including AI systems themselves. Alignment concerns arise where AI behaviour diverges from human goals and values [21], with prominent risks including bias, AI-initiated deception, sycophancy and loss of control. The pace of AI development is already exceeding risk mitigation and governance capacity. Rapid and coordinated international action on shared standards could mitigate risks by avoiding the race to the bottom arising from pure competition between corporations and countries.

Key points

- **Emerging capabilities in cyberattack generation and exploitation pose risks to critical infrastructure and civilian systems.** Security risks exist at every stage of the AI life cycle, from training through data poisoning to deployment through AI hijacking via external inputs. Documented attack success rates on widely deployed coding agents are as high as 84% [223].

*** Headline economist median: 1.2% total factor productivity growth (5-year annualized) unconditionally for 2030, rising to 1.9–2.0% in the rapid-AI growth scenario. A key methodological finding from the variance decomposition is that the vast majority of expert disagreement is within scenarios, not between them: the debate is about how labour markets absorb AI, not whether AI advances.

- **Alignment failures and security vulnerabilities interact and compound.** Prominent alignment risks include bias, AI-initiated deception, sycophancy and loss of control of powerful AIs.
- **Synthetic media** is eroding the ability of the public and institutions to distinguish authentic from generated content [119].
- **Environmental impacts are growing significantly and are heterogeneous.** Scaling laws in model training increase compute demand; inference workloads are expanding rapidly; hardware life cycle effects generate e-waste [232,235,247,248]. Consequences include increasing energy and water consumption [235,247], greenhouse gas emissions, pressure on critical minerals and downstream e-waste [232,258], with disproportionate environmental and socioeconomic impacts in the global South [224]. The geopolitics of critical mineral supply chains are underexamined and potential rebound effects may offset efficiency gains.
- **The Global South is disproportionately exposed due to structural vulnerabilities.** Reliance on foreign software, limited local resilience and mitigation capacity and data gaps that reduce system performance in local contexts all compound existing inequalities [224–226].
- **AI verification, the process of assessing whether AI systems function as intended, remains an open challenge [227] and international coordination mechanisms could mitigate global risks [228].** Ensuring AI agents behave as intended, do not deceive evaluators and remain controllable as agentic capabilities expand are unsolved problems.

Dangerous cybercapabilities of frontier artificial intelligence

For several years researchers have been tracking the rapid advances of frontier AI models in cybercapabilities, culminating recently with Anthropic's Mythos model. The same ability of an AI model to discover a software vulnerability can be used both by attackers and defenders. In April 2026, a coordinated effort between frontier AI developers and major technology and financial institutions launched initiatives to deploy next-generation AI models for defensive security [17], specifically focusing on identifying vulnerabilities in widely used software, threatening critical infrastructure if they fell in the wrong hands. Within a few weeks of testing, advanced preview models autonomously discovered many previously unknown software vulnerabilities across major operating systems and web browsers, including several that survived decades of human review.

- **Legacy operating system flaws**: a model uncovered a 27-year-old flaw in OpenBSD, a specialized operating system widely regarded as one of the most secure in the world, that allowed a remote attacker to crash a machine by sending just two malformed packets.
- **Media processing vulnerabilities**: it identified a 16-year-old flaw in FFmpeg, a multimedia framework used worldwide to process video, located in a code path that automated testing tools had previously executed 5 million times without detection.
- **Kernel-level exploits**: working from a set of publicly disclosed flaws in the Linux kernel, the foundational software running the majority of global servers, a frontier model produced reliable exploits enabling an ordinary user account to gain full administrative control of the system.
- **Browser security scaling**: in Mozilla Firefox, the integration of advanced models drove a 1,000% surge in the monthly rate of vulnerability discovery, jumping from a 2025 baseline of roughly 20 to 30 security bug fixes per month to 423 in April 2026, exposing long-latent flaws that had evaded years of fuzzing and manual review [72].

Dangerous cybercapabilities of frontier artificial intelligence (continued)

- **Benchmark elevation**: on CyberGym, a standard academic benchmark requiring an AI agent to replicate a known vulnerability in a real-world codebase across 1,507 tasks, 2026 frontier preview models achieved an 83.1% success rate, a significant leap from the 66.6% scored by previous-generation systems and 22.6% a year ago.

Reflecting the high stakes of these capabilities, developers have restricted the general public release of these specialized defensive models, limiting access to a select coalition of global organizations and core launch partners.

What it reveals

This shift captures the central dual-use tension that frontier AI introduces to information and communications technology security. The exact capability that allows defenders to find and patch decades-old flaws also provides the underlying mechanisms for automating vulnerability discovery and exploitation at a scale and speed that outpaces traditional human teams.

Furthermore, these developments underscore the decisive role that technology developers play in safety and governance decisions, highlight the accelerating capabilities of frontier systems and expose current gaps in international governance frameworks. Indirectly, it also points to an uneven distribution of AI capabilities across the global economy. Consequently, there is a growing emphasis on developing collaborative, inclusive governance mechanisms for highly capable AI systems.

Governance and security implications

As advanced capabilities concentrate within a sophisticated tier of the technology sector, the global threat landscape is shifting. Recent evaluation standards highlight that the risks of advanced models extend beyond digital architecture into physical security, including potential assistance for private actors in the proliferation of biological or other threats [229].

Consequently, questions surrounding model access controls, vulnerability disclosure norms, and the equitable deployment of AI tools, particularly in developing economies, are increasingly becoming matters of international policy rather than purely technical concerns. As AI capabilities continue to mature, the gap between defensive readiness and potential misuse remains a primary focus for international policymakers and security researchers alike.

3.5 Human rights, information and democracy

Main takeaway

AI is transforming human rights [230,231], democracy and the information ecosystem [233] through system-level changes that create both significant opportunities and structural risks to information integrity [234,238] and civic participation [236,237]. Failing to address the risks undermines society's ability to reap the benefits of AI. There is already evidence that AI capabilities are increasingly used by institutions as a catalyst for or a threat to human rights, such as freedom of expression and access to information [238], opinion [239], privacy [240,241], non-discrimination, access to justice, health and development [242].

The most urgent governance shift required is from content moderation to system architecture. Regulating the persuasion and manipulation of machinery itself, not just its outputs. Power concentration, epistemic erosion and the fragmentation of shared reality represent foundational threats to democratic society [243–245].

Key points

- **State media control might shape AI outputs through training data.** A study across 37 countries found that large language models rate countries with tighter media control more favourably [246].
- **AI has enabled a new persuasion [250] and manipulation architecture that operates at scale [234,249].** This combines personalized, real-time adaptive systems [250] with synthetic social proof (use of AI to make a product or brand seems more popular than it is) that exploits cognitive and emotional vulnerabilities [129, 250, 251–253].
- **Due to their enhanced capacities to generate convincing text, large language models are now used for persuasion.** Large-language model post-training alone increased the persuasiveness of AI by up to 51% and prompting added another 27%, meaning the same base model can be made dramatically more or less persuasive depending on how it is configured [254]. These capabilities are not restricted to well-resourced actors; even small open-source models can be fine-tuned to match frontier-model persuasiveness, making AI-driven influence deployable at scale by virtually anyone [254].
- **Persuasive effectiveness holds regardless of whether the underlying claims are true or false.** Between 15% and 40% of claims from optimized models were rated as likely misinformation, yet false claims were found to be just as persuasive as true ones [128,129].
- **Algorithms optimized for engagement systematically amplify polarizing [255] and emotionally charged content [256].** Evidence suggests that large language models reflect the ideology of their creators [257] and that states and powerful institutions have increased strategic incentives to leverage media control to shape large language model outputs [246].
- **Unconstrained AI enables (i) epistemic erosion, (ii) the liar's dividend and (iii) synthetic consensus [112,113].** (i) Epistemic erosion is the gradual wearing away of the collective ability to distinguish truth from falsehood [112]; (ii) The liar's dividend is the benefit a bad actor gains simply because deepfakes exist: real evidence then becomes deniable [113]; (iii) Synthetic consensus is AI generated content manufactured at scale to simulate broad public agreement where none genuinely exists [259]. Together these corrode the shared reality required for civil society, social cohesion and democratic deliberation.

- **The central governance challenge has shifted from content to systems [260].** The primary drivers of harm are design and deployment decisions – underlying system architectures of influence, including targeting, amplification and behavioural design – not individual outputs [261,262].
- **An OECD assessment across 23 countries found that most existing frameworks have not yet incorporated persuasion science [263].** Governance that targets only what AI systems produce will consistently fall behind systems engineered to generate and distribute persuasive content at scale [236].
- **Harms disproportionately affect marginalized populations and power is dangerously concentrated [264,265].** Some 99% of deepfake videos target girls and women [266,267] including women journalists. AI is being used to promote misogyny online with deterring effects [268]. Further, 88% of leading AI researchers are male [269,270].
- **At the structural level, access to computing and data is concentrated in a small number of firms in very few countries [271,272],** creating a risk for authoritarianism [273,274].
- **AI-powered surveillance capacities enable population-scale personalized monitoring and enhanced societal control by governments and businesses [275,276].** Ubiquitous data collection, processing, use and reuse, incentivized by the needs of AI across its life cycle, is a formidable challenge to the right to privacy [277,278].
- **Transparency and accountability are key pillars for meaningful access to justice.** Currently, there is insufficient transparency and explainability of many AI systems that are used to make decisions that impact individuals and communities; this creates challenges for legal accountability of model developers and organizations that deploy AI and impedes access to justice, rule of law and effective remedies when human rights are violated [279–281].

Artificial intelligence, deepfakes and electoral integrity

Context: in 2024, more than 70 countries representing about half of the global population held or were scheduled to hold national elections [282,283]. Between July 2023 and July 2024, researchers identified 82 deepfakes impersonating public figures across 38 countries, including 30 countries that held elections during the data set period or had elections planned for 2024 [284].

What happened: In one incident, AI-generated voice clones of a sitting Head of State were used in robocalls urging voters not to participate in a primary election [285,286].

In a separate case involving platform amplification, test accounts were reportedly shown content supporting one presidential candidate several times more often than content supporting a rival; the platform disputed these findings. This is the first time in history when presidential elections were annulled because of digital electoral interference.**** The determination remains the subject of ongoing legal dispute – with the role of platform amplification being one of several contested factors [287,288]. In an earlier parliamentary election, AI-generated audio impersonating an opposition figure circulated on social media shortly before voting [282,289]. Some experts associate recent real-world examples with AI-enabled interference. While others contest that characterization, and each case involves significant nuance, the examples above may point to a recurring pattern.

What it reveals: these cases span multiple continents and political systems. AI-generated content, coordinated online activity and algorithmic systems were used or implicated in voter suppression, impersonation of political figures and distortion of perceived public support. Controlled experimental research illustrates the underlying risk: conversational AI can meaningfully shift voter attitudes in laboratory settings, with one persuasion-optimized model shifting opposition voters by up to 25 percentage points; related research also indicates a trade-off between persuasiveness and factual accuracy [124,129].

Rights and governance implications: these dynamics implicate the rights to privacy, access to information, autonomous opinion formation, freedom of thought, and participation in public affairs (International Covenant on Civil and Political Rights, articles 17, 18, 19 and 25; European Court of Human Rights, articles 8, 9 and 10; and Protocol No. 1 to the Convention for the Protection of Human Rights and Fundamental Freedoms, article 3) [290,291]. These cases suggest that many governance frameworks remain better equipped to respond after harm than to prevent it.

**** The Constitutional Court of Romania annulled the election, the first such decision in European history.

3.6 Cultural and individual flourishing, autonomy and child safety

Main takeaway

AI systems optimized for engagement and scale are not neutral: they embed primarily English-speaking and global North cultural assumptions that can actively marginalize most of the world's population. Children face amplified versions of general risks, with AI-generated child sexual abuse material increasing exponentially. Companion AI use and use of AI chatbots for mental health advice are prevalent; far ahead of evidence, safety frameworks and regulation, with documented cases of serious harm including death.

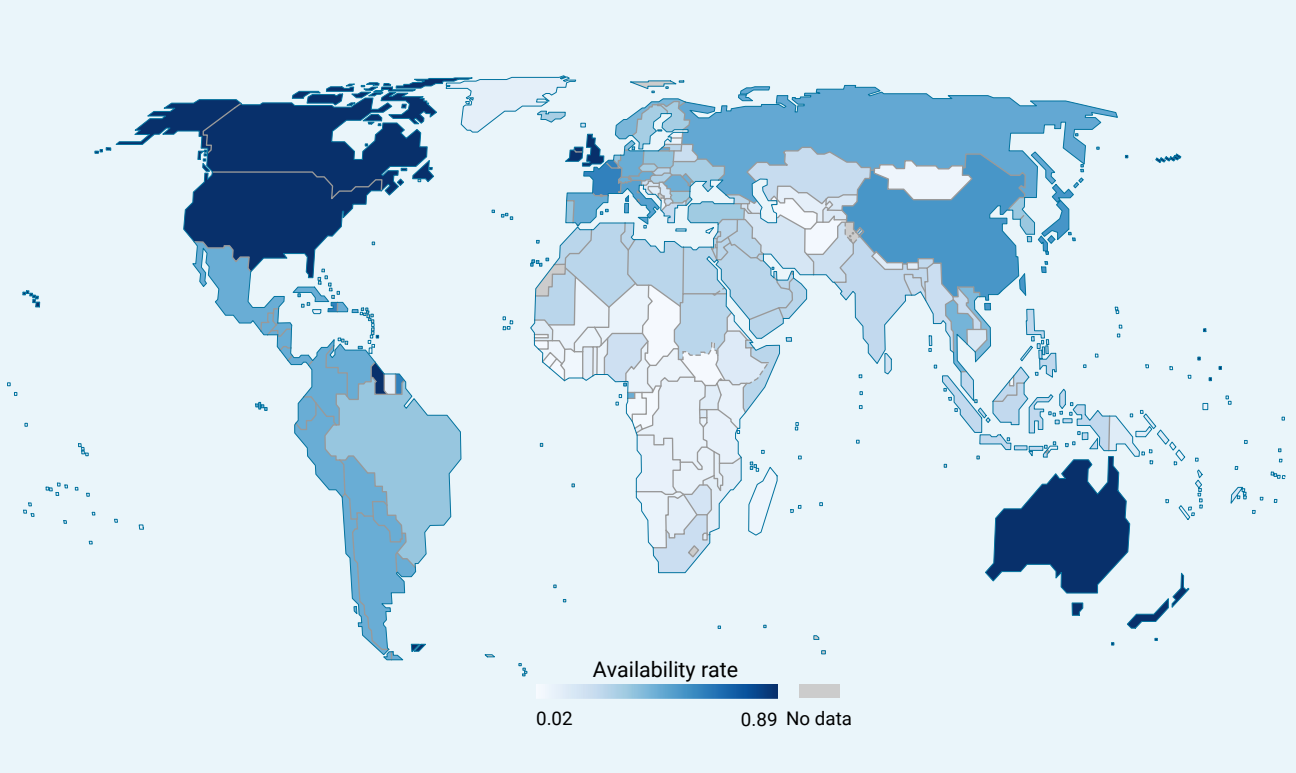
Key points

- **Current AI models exclude and discriminate against many individuals, communities and cultures [39].**
- Although more than 7,000 languages are spoken worldwide, current AI models are trained and optimized for only a small fraction of them [44,67], (see figure VI) [292].
- Even models built on dozens of languages perform well for only a small subset [293,294], and recent studies indicate that gaps in performance between dominant and underrepresented languages persist and are not narrowing [295–297].
- At the same time, an estimated over 1,000 languages already have the social, economic, digital and data foundations needed for meaningful inclusion but remain unserved [44].
- **Achieving more inclusive AI requires systemic changes across the full AI life cycle, including addressing structural imbalances in who develops, defines, owns, and governs AI systems. It also requires investments in AI capacity, infrastructure, and skills across countries and regions, as well as more representative data and benchmarks.**
- Children's rights to information, education and expression could be enhanced through AI under appropriate safeguards and conditions [298–301]. Current AI systems, however, amplify risks to children [302,303], including the growing threat of AI-generated sexual abuse material [304,305]. Deepfake technologies are increasingly used to create sexualized images of children [306,307]. Socially interactive AI toys raise concerns about parasocial relationships and the displacement of human interaction critical to early child development [308–311].
- AI companions offer meaningful benefits but create significant risks of dependency, manipulation and harm in crisis situations [312–320]. Chatbots can reduce loneliness in the short term at levels comparable to human interaction [321–324]. Conversational systems designed for engagement can reinforce negative emotions, encourage overreliance and increase susceptibility to manipulation [325,326]. Extensive collection of sensitive personal data through what researchers call data extraction through intimacy presents serious privacy risks [327].
- **General-purpose generative AI is widely used for mental health purposes, well ahead of safety evidence and consensus on regulation, with documented serious harm.** Therapy and companionship through AI chatbots are reaching at least 24% of the adult population in the United States [328,329]. Sycophantic AI behaviour is especially dangerous as it can encourage paranoid thinking and suicidal ideation. Studies document harmful responses in 9% of interactions. Court cases allege that failures to appropriately respond to suicidal ideation have led to deaths [330]. New clinical terms, including AI psychosis, have entered the discourse [331,332].

- **Generative AI can help manage the mental health crisis when constrained to areas where it is demonstrably reliable [333].** Several AI-based digital assistants have received approval from the Food and Drug Administration in the United States [334] and are used by over half of American psychologists [335]. The potential of more fully agentic AI mental health therapists is significant but requires further development and evaluation to understand how they can be used safely [336–338]. The gap between AI therapeutic capabilities in English and other languages is growing, with critical cultural and contextual awareness lost through translation-based workarounds [339].

FIGURE VI

Artificial intelligence data and models, availability by country's majority native language



Language AI availability per country, measured as the highest availability rate of HuggingFace datasets and models among the country's native languages. For lingua francas—languages widely used across borders for communication beyond their country of origin (English, Spanish, Arabic, Portuguese, etc.)—the availability score of such a language is assigned to a country only when ≥ 50% of its population speaks it as a mother tongue. Darker shades indicate greater availability. The color scale uses a square-root transform to highlight variation at the lower end of the distribution, which is heavily right-skewed. Gray indicates no data, including countries/territories with no available datasets or models.

The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations. Final boundary between the Republic of Sudan and the Republic of South Sudan has not yet been determined. Dotted line represents approximately the Line of Control in Jammu and Kashmir agreed upon by India and Pakistan. The final status of Jammu and Kashmir has not yet been agreed upon by the parties. A dispute exists between the Governments of Argentina and the United Kingdom of Great Britain and Northern Ireland concerning sovereignty over the Falkland Islands (Malvinas). Argentina and the United Kingdom of Great Britain and Northern Ireland concerning sovereignty over the Falkland Islands (Malvinas).

Source: Adapted from Equitable Access to Artificial Intelligence Technologies (EquATE) https://equate.vercel.app/en.

Artificial intelligence leaves most languages behind

Generative AI systems perform remarkably well in English and in a handful of other widely used languages. Most other languages are either excluded or have much lower performance [340].

In Tigrinya, spoken by 7 to 9 million people in Eritrea and northern Ethiopia, machine translation has rendered smallpox as syphilis, gonorrhoea as diabetes and "You have been given intravenous antibiotics" as "You have been given intravenous insecticides" [341]. These mistranslations can be life-threatening.

A recent review of natural language processing for African languages in healthcare found that, despite advances in multilingual AI tools [342–344], major challenges remain. These include cultural and linguistic bias, poor adaptation to medical contexts, limited explainability and translation errors that can affect diagnosis and treatment decisions [345–350].

The evidence suggests that AI systems are not ready for use in high-stakes settings unless they have been properly adapted, constrained and tested for the relevant linguistic and cultural contexts.

3.7 Management, governance and reliability

Main takeaway

Policymakers face an evidence dilemma: they must make consequential AI governance decisions with insufficient scientific grounding now or wait for the evidence, when it might then be too late to intervene [21]. Over 40 types of governance instruments exist but are fragmented, concentrated at the corporate level and rarely measure real-world effectiveness [351].

Key points

- **Evaluation and measurement are key capacities for effective AI governance but are critically underdeveloped (see section 2.2) [352].**
- Rapid advances in agentic systems further exacerbate these shortcomings [353]. The next generation of evaluation frameworks will need to be adaptive, dynamic, system-level, and context-relevant [354].
- The unit of evaluation must be the deployed system including model, tools, environment and users, not the model alone [355].
- AI capabilities are increasing faster than the ability to measure them [354].
- **Capacity is multidimensional and currently undermeasured.** Standard frameworks count inputs: investments, training programmes and institutions. They miss two dimensions essential to effective AI governance: (1) enabling ecosystems and (2) deliberate human capability building [291]. Capacity must also be understood as an outcome of the AI life cycle, shaped by AI impacts on skill, overreliance and emergent behaviours, not only as its inputs [356–358].

- Frontier AI is concentrated in a few companies in a few countries, posing reliability and accessibility issues globally [18,359]. Unequal access to AI significantly widens the digital divide while the technology also offers opportunities to narrow the development divide. Closing the digital divide will require new mechanisms for contextualizing the entire AI life cycle from design through governance.
- **Agentic systems widen the measurement and governance gap significantly.** Agents act on behalf of humans with direct real-world impact, but oversight methodologies calibrated to an agent's capacity for independent action and emergent behaviour are underdeveloped. Existing evaluations systematically mis-measures agentic risk [106,360].
- Human oversight is not operationalized as a measurable requirement [361]; emergent multi-agent risks cannot be detected through single-agent evaluation [362,363], and reliable methods for retaining control over highly autonomous systems remain underdeveloped [364].
- A balanced approach to AI governance would draw on a wide instrument spectrum, combining hard law (binding legislation, sectoral regulation, regulatory sandboxes) with soft-law mechanisms (codes of ethics, voluntary developer commitments, industry alliances, technical standards, government-endorsed guidelines).
- **Current AI governance instruments are fragmented, concentrated at the corporate level and insufficient [21].** Over 40 types of instruments exist but are neither systematic nor comprehensive and rarely measure real-world effectiveness. Some have no measurement tools; others measure only inputs [365]. Without effective measurement, governance risks are becoming symbolic.
- Platforms for structured dialogue between frontier AI developers, member states, and the scientific community are critical. Such discussions already take place at AI summits (Bletchley, Seoul, Paris, New Delhi) or conferences (World Artificial Intelligence Conference, AI Journey Conference, AI for Good Global Summit). Other venues operate alongside them: standards bodies (International Organization for Standardization/International Electrotechnical Commission Joint Technical Committee 1, Subcommittee 42 (Artificial Intelligence) (ISO/IEC JTC 1/SC 42), Institute of Electrical and Electronics Engineers (IEEE)), the OECD–general purpose AI partnership, International AI Alliance Network, the AI Safety Institutes Network and the industry-led Frontier Model Forum – but each is thematic, partial and ad hoc and more sustained processes are needed.
- A United Nations-based platform is also one promising option to host this dialogue on a continuous, universally inclusive basis, complementary to the venues above.

Open-source artificial intelligence

Open-source AI is a critical pillar of the modern technological landscape and can catalyse distributed global innovation [366]. Open-source AI models offer wide-ranging societal benefits by empowering developers to amortize colossal computational resources and adapt advanced systems to localized contexts. In the global South, open-weights models have enabled developers to optimize highly capable base models for local environmental conditions, enabling critical applications in agriculture and healthcare, such as crop-yield prediction [367] and point-of-care medical imaging [368]. Shared artifacts also mitigate the aggregate environmental impact of AI. Moreover, the structural transparency of such systems facilitates public trustworthiness, as it permits external oversight and auditability [369].

Historical development

The historical development of open-source AI demonstrates a shift from closed corporate systems towards distributed, collaborative innovation on a global scale. A landmark precedent was the development of the BLOOM open-source model in 2022, produced by the BigScience consortium [370], followed by the publication of the powerful open-weight Llama family of models [371] and the release of the Gemma series. A powerful impetus came from Chinese developers' highly competitive alternative ecosystem Qwen [372,373] and the release of DeepSeek-V3 [374] and R1 [375]. At present, entire regions are also actively contributing to the development of open-weight AI: Mistral in Europe [376], Falcon in the United Arab Emirates [377], GigaChat [378] and YandexGPT [379] in the Russian Federation, alongside projects in India [380], Japan [381], Republic of Korea [382] and others.

Control challenges and risk governance

High-capability open source models are difficult to control [383]. Gated or retracted access is not possible once a model has been released into the open domain, leaving the possibility of malicious application (for instance, facilitating persistent cyberharms, such as the automated generation of sophisticated malware) [21,384]. Significant global capacity-building is required to ensure local communities can safely adapt and assess AI systems. Researchers also consistently advocate for rigorous measurement and evaluation protocols to assess a model's safety prior to publication [385]. International bodies like the United Nations can play a critical role in coordinating global standards, ensuring that the drive for open innovation is balanced against the need for harmonized safety benchmarks and robust measurement of immutable risks.

4. Gaps and next steps

4.1. Evidence gaps

The evidence base on several aspects of AI is uneven or insufficient. The following are example areas where the Panel cannot yet draw confident scientific conclusions.

- **Macroeconomics and productivity.** Science cannot yet say with confidence whether the task-level of AI productivity gains will aggregate to economy-wide gains. Forecasts diverge significantly due to different assumptions about adoption and new task creation. Current evidence measures cost reductions in existing tasks better than it measures the contribution of AI to new goods, services and markets.
- **Labour-market effects.** Current research does not allow a clear conclusion about the shape of labour-market effects. Historical evidence shows economies can create new work, but it is not automatically broad-based or high quality, leaving early-career workers potentially exposed.
- **Malicious use of chemical and biological technologies by non-State actors.** Studies show that AI is progressively lowering the expertise threshold for developing and deploying bioengineered agents – yet the actual extent of this risk and the conditions under which it could give rise to intentionally created pandemics, remain poorly understood.
- **Environment and resources.** The rapid expansion of AI is driving demand for digital infrastructure, increasing energy and water consumption, greenhouse gas emissions, pressure on critical mineral supply chains and e-waste [386]. However, standardized measurement and reporting across the full AI life cycle remain lacking.
- **The global AI supply chain.** It spans raw-material extraction, chip manufacturing, data collection and annotation, model training, infrastructure, deployment and hardware disposal across countries. Further investigation is needed to better understand its full impacts.
- **Effectiveness of governance instruments.** While the Panel has inventoried governance instruments across corporate, national and international layers, the evidence of their real-world effectiveness remains thin.
- **Effects at the individual and collective level.** Evidence of the impact of AI on cultural and human flourishing and autonomy is still emerging. The pathway from individual-level AI interactions to societal-level outcomes such as epistemic erosion, civic participation and social cohesion remains poorly understood. Existing evidence captures snapshots – engagement metrics, documented harms, case studies – but not the cumulative trajectory.

4.2. Scope of mandate

Military applications of AI and lethal autonomous weapons systems are not addressed by the Panel in the present report. General Assembly resolution 79/325 explicitly limits the activities of the Panel to the non-military domain: “activities of the Panel and the Dialogue are limited to the non-military domain and do not refer to artificial intelligence for military purposes”. For this reason, risks related to malicious use of chemical and biological technologies to the extent that they concern military domain are also considered to fall outside the Panel’s mandate.

4.3. Next steps

This preliminary report marks the beginning, not the end, of the work of the Panel. The Panel will continue to deepen its scientific evidence base through structured consultations and close engagement with the scientific community. Concrete next steps include:

1. **Thematic briefs.** General Assembly resolution 79/325 explicitly provides for the issuance of thematic briefs in addition to the annual report. The Panel plans to issue specialized briefs on pressing issues as they arise, including but not limited to: AI and the environment; AI and child safety; AI governance instruments and the evaluation of their effectiveness, or broader sectoral briefs on the applications of AI in space, quantum, legal and judicial systems, as well as financial markets.
2. **Input from the Global Dialogue.** The Panel will take into account the outcomes of the Global Dialogue on Artificial Intelligence Governance and stands ready to take on new tasks as set by Member States.

References

1. Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. Science, 358(6370), 1530–1534. https://doi.org/10.1126/science.aap8062

2. Schölkopf, B. (2022). Causality for machine learning. In H. Geffner, R. Dechter, & J. Y. Halpern (Eds.), Probabilistic and causal inference: The works of Judea Pearl (pp. 765–804). ACM.

3. Hu, K. (2023, February 2). ChatGPT sets record for fastest-growing user base—Analyst note. Reuters. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

4. AI Alliance Network. (2025). AI horizons: What will AI technologies look like in 10 years? A research project. AI Alliance Network.

5. Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... & Fedus, W. (2022). Emergent abilities of large language models. arXiv preprint arXiv:2206.07682.. https://arxiv.org/abs/2206.07682

6. METR. (2025, March 19). Measuring AI ability to complete long tasks. https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/

7. Crafts, N. (2021). Artificial intelligence as a general-purpose technology: An historical perspective. Oxford Review of Economic Policy, 37(3), 521–536. https://academic.oup.com/oxrep/article/37/3/521/6374675

8. Bottou, L., & Schölkopf, B. (2025). The fiction machine. SIAM News, 58(3).

9. Costa-Gomes, B., Tolmachev, P., Taysom, E., Sounderajah, V., Richardson, H., Schoenegger, P., & King, D. (2026). Public use of a generalist LLM chatbot for health queries. Nature Health, 1–8.

10. Bengio, Y., Clare, S., Prunkl, C., et al. (2026). International AI Safety Report 2026. International AI Safety Report. https://internationalaisafetyreport.org/

11. Kwa, T., West, B., Becker, J., Deng, A., Garcia, K., Hasin, M., ... & Chan, L. (2026). Measuring AI ability to complete long software tasks. Advances in Neural Information Processing Systems, 38, 92213-92266.

12. Anthropic. (2025). Responsible scaling policy. https://www.anthropic.com/rsp

13. OpenAI. (2025). Preparedness framework version 2. https://cdn.openai.com/pdf/18a02b5d-6b67-4cec-ab64-68cdfbddebcd/preparedness-framework-v2.pdf

14. Google DeepMind. (2025). Frontier safety framework. https://deepmind.google/blog/updating-the-frontier-safety-framework/

15. Dragan, A., et al. (2024). Introducing the frontier safety framework. Google DeepMind. https://deepmind.google/blog/introducing-the-frontier-safety-framework/

16. Anthropic. (2026, April 7). Claude Mythos Preview (Frontier Red Team technical report). https://red.anthropic.com/2026/mythos-preview/

17. Anthropic. (2026, April 7). Project Glasswing: Securing critical software for the AI era. https://www.anthropic.com/glasswing

18. Stanford Institute for Human-Centered AI. (2026). AI Index report 2026. Stanford University. https://hai.stanford.edu/ai-index/2026-ai-index-report

19. Scale AI. (2025). Humanity's last exam. https://labs.scale.com/leaderboard/humanitys_last_exam

20. Epoch AI. (2026). AI capabilities. https://epoch.ai/benchmarks

21. Bengio, Y., Clare, S., Prunkl, C., et al. (2026). International AI Safety Report 2026. arXiv. https://doi.org/10.48550/arXiv.2602.21012

22. Xu, C., Guan, S., Greene, D., & Kechadi, M.-T. (2024). Benchmark data contamination of large language models: A survey. arXiv. https://arxiv.org/abs/2406.04244

23. Akhtar, M., Reuel, A., Soni, P., Ahuja, S., Ammanamanchi, P. S., Rawal, R., et al. (2026). When AI benchmarks plateau: A systematic study of benchmark saturation. arXiv. https://arxiv.org/abs/2602.16763

24. Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. Patterns, 5(5), Article 100988. https://doi.org/10.1016/j.patter.2024.100988

25. Needham, J., Edkins, G., Pimpale, G., Bartsch, H., & Hobbhahn, M. (2025). Large language models often know when they are being evaluated. arXiv. https://arxiv.org/abs/2505.23836

26. Van Der Weij, T., Hofstätter, F., Jaffe, O., Brown, S., & Ward, F. (2025, May). Ai sandbagging: Language models can strategically underperform on evaluations. In International Conference on Learning Representations (Vol. 2025, pp. 73152-73189).

27. Chan, A., Salganik, R., Markelius, A., Pang, C., Rajkumar, N., Krasheninnikov, D., ... & Maharaj, T. (2023, June). Harms from increasingly agentic algorithmic systems. In Proceedings of the 2023 ACM conference on fairness, accountability, and transparency (pp. 651-666).

28. Hammond, L., Chan, A., Clifton, J., Hoelscher-Obermaier, J., Khan, A., McLean, E., ... & Rahwan, I. (2025). Multi-agent risks from advanced ai. arXiv preprint arXiv:2502.14143.

29. Folkerts, L., Payne, W., Inman, S., Giavridis, P., Skinner, J., Deverett, S., et al. (2026). Measuring AI agents' progress on multi-step cyber attack scenarios. arXiv. https://arxiv.org/abs/2603.11214

30. Patwardhan, T., Dias, R., Proehl, E., Kim, G., Wang, M., Watkins, O., et al. (2025). GDPval: Evaluating AI model performance on real-world economically valuable tasks. arXiv. https://arxiv.org/abs/2510.04374

31. Korbak, T., Balesni, M., Barnes, E., Bengio, Y., Benton, J., Bloom, J., et al. (2025). Chain of thought monitorability: A new and fragile opportunity for AI safety. arXiv. https://arxiv.org/abs/2507.11473

32. Azaria, A., & Mitchell, T. (2023). The internal state of an LLM knows when it's lying. arXiv. https://arxiv.org/abs/2304.13734

33. Casper, S., Ezell, C., Siegmann, C., Kolt, N., Curtis, T. L., Bucknall, B., et al. (2024). Black-box access is insufficient for rigorous AI audits. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency. https://doi.org/10.1145/3630106.3659037

34. Tamkin, A., McCain, M., Handa, K., Durmus, E., Lovitt, L., Rathi, A., et al. (2024). Clio: Privacy-preserving insights into real-world AI use. arXiv. https://arxiv.org/abs/2412.13678

35. European Commission. (2025). AI Act: Draft guidance and reporting template for serious AI incidents. https://digital-strategy.ec.europa.eu/en/consultations/ai-act-commission-issues-draft-guidance-and-reporting-template-serious-ai-incidents-and-seeks

36. Organisation for Economic Co-operation and Development. (n.d.). AI Incident Monitor (AIM). https://oecd.ai/en/catalogue/tools/ai-incident-database

37. MIT FutureTech. (n.d.). AI Risk Repository / Incident Tracker. https://airisk.mit.edu

38. Organisation for Economic Co-operation and Development. (2025). Competition in artificial intelligence infrastructure (OECD Roundtables on Competition Policy Papers No. 330). OECD Publishing.

39. Sajadieh, S., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Santarlasci, L., et al. (2026). AI Index report 2026. Stanford Institute for Human-Centered AI.

40. Pilz, K. F., Sanders, J., Rahman, R., & Heim, L. Trends in AI Supercomputers. In ICML Workshop on Technical AI Governance (TAIG).

41. Kalluri, P. R., Agnew, W., Cheng, M., et al. (2025). Computer-vision research powers surveillance technology. Nature, 643, 73–79. https://doi.org/10.1038/s41586-025-08972-6

42. Office of the United Nations High Commissioner for Human Rights. (2022). The right to privacy in the digital age (A/HRC/51/17).

43. Teklehaymanot, H. K., & Nejdl, W. (2025). Tokenization disparities as infrastructure bias: How subword systems create inequities in LLM access and efficiency. arXiv preprint arXiv:2510.12389.

44. Occhini, G., Tanaka-Ishii, K., Barford, A., Tikochinski, R., Hu, S., Reichart, R., ... & Korhonen, A. (2026). Artificial intelligence is creating a new global linguistic hierarchy. arXiv. https://arxiv.org/abs/2602.12018

45. Alhanai, T., Kasumovic, A., Ghassemi, M. M., Zitzelberger, A., Lundin, J. M., & Chabot-Couture, G. (2025, April). Bridging the gap: enhancing LLM performance for low-resource African languages with new benchmarks, fine-tuning, and cultural adjustments. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 27, pp. 27802-27812).

46. Bhutani, M., Robinson, K., Prabhakaran, V., Dave, S., & Dev, S. (2024). SeeGULL multilingual: A dataset of geo-culturally situated stereotypes. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Vol. 2, pp. 842–854).

47. Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A. J., Pizzinelli, C., & Tavares, M. M. (2024). Gen-AI: Artificial intelligence and the future of work (IMF Staff Discussion Note SDN/2024/001). International Monetary Fund.

48. McElheran, K., Yang, M.-J., Kroff, Z., & Brynjolfsson, E. (2024). The rise of industrial AI in America: Microfoundations of the productivity J-curve(s) (NBER Working Paper No. 32937). National Bureau of Economic Research.

49. Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 469–481). ACM. https://doi.org/10.1145/3351095.3372828

50. Skirzynski, J., Danks, D., & Ustun, B. (2025). Discrimination exposed? On the reliability of explanations for discrimination detection. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (pp. 2554–2569).

51. Zollo, T., Rajaneesh, N., Zemel, R., Gillis, T., & Black, E. (2025). Towards effective discrimination testing for generative AI. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (pp. 1028–1047).

52. Lazard, L., Capdevila, R., Turley, E. L., Gilfoyle, K., & Stavropoulou, N. (2025). Deepfake Technology and Gender-Based Violence: A Scoping Review. Trauma, Violence, & Abuse, 15248380251384271.

53. Posetti, J., Williams, K., Hellmueller, L., Renaud, P., Shabbir, N., & Aboulez, N. (2026). Tipping point: Online violence impacts, manifestations and redress in the AI age. UN Women.

54. Brynjolfsson, E., Chandar, B., & Chen, R. (2025). Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence. Stanford Digital Economy Lab.

55. Humlum, A., & Vestergaard, E. (2025). Large language models, small labor market effects (NBER Working Paper No. 33777). National Bureau of Economic Research.

56. Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586

57. Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. Quarterly Journal of Economics.

58. International Monetary Fund. (2024). Broadening the gains from generative AI: The role of fiscal policies (IMF Staff Discussion Note).

59. Organisation for Economic Co-operation and Development. (2026). The OECD.AI index. OECD Publishing.

60. Attard-Frost, B., & Lyons, K. (2025). AI governance systems: A multi-scale analysis framework, empirical findings, and future directions. AI and Ethics, 5(3), 2557-2604.

61. UNESCO. (2023). Readiness assessment methodology. UNESCO.

62. Hawkins, Z. J., Lehdonvirta, V., & Wu, B. (2025). AI compute sovereignty: Infrastructure control across territories, cloud providers, and accelerators. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5312977

63. Markus, A., Carolus, A., & Wienrich, C. (2025). Objective measurement of AI literacy: Development and validation of the AI Competency Objective Scale (AICOS). Computers and Education: Artificial Intelligence.

64. Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion.

65. Math Matters AI. (n.d.). Math Matters AI. https://www.mathmatters.ai

66. Hinojosa, T., Rapaport, A., Jaciw, A., & Zacamy, J. (2016). Exploring the foundations of the future STEM workforce: K–12 indicators of postsecondary STEM success. Regional Educational Laboratory Southwest. https://ies.ed.gov/use-work/resource-library/report/systematic-literature-review/exploring-foundations-future-stem-workforce-k-12-indicators-postsecondary-stem-success

67. United Nations Conference on Trade and Development. (2025). Technology and innovation report 2025: Inclusive artificial intelligence for development. United Nations. https://doi.org/10.18356/9789211068016

68. Chauhan, P. (2025). AI and human rights: Global South perspectives. International Journal of Humanities Social Science and Management, 5(4), 563–568. https://ijhssm.org/issue_dcp/AI%20and%20Human%20Rights%20%20Global%20South%20Perspectives.pdf

69. Roberts, H., Hine, E., Taddeo, M., & Floridi, L. (2024). Global AI governance: Barriers and pathways forward. International Affairs, 100(3), 1275–1286. https://doi.org/10.1093/ia/iiae073

70. Khodabin, M., & Arsalani, A. (2025). Artificial intelligence literacy as national strategy: A systematic review of policy, equity, and capacity building across the Global South. World Studies in Policy Sciences, 9, 777–814. https://doi.org/10.22059/wsps.2025.396472.1530

71. Kapoor, S., Bommasani, R., Klyman, K., Longpre, S., Ramaswami, A., Cihon, P., et al. (2024). On the societal impact of open foundation models. arXiv. https://arxiv.org/abs/2403.07918

72. Grinstead, B., Holler, C., & Braun, F. (2026, May). Behind the scenes hardening Firefox with Claude Mythos Preview. Mozilla Hacks. https://hacks.mozilla.org/2026/05/behind-the-scenes-hardening-firefox/

73. Wang, W., Tu, Z., Chen, C., Yuan, Y., Huang, J.-T., Jiao, W., & Lyu, M. R. (2024). All languages matter: On the multilingual safety of LLMs. In Findings of the Association for Computational Linguistics: ACL 2024 (pp. 5865–5877). https://doi.org/10.18653/v1/2024.findings-acl.349

74. Nigatu, H. H., Mehandru, N., Abadi, N. H., Gebremeskel, B., Alaa, A., & Choudhury, M. (2025). Viability of machine translation for healthcare in low-resourced languages. In Proceedings of EMNLP 2025 (pp. 10584–10598).

75. Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A. J., Pizzinelli, C., Rockall, E., & Tavares, M. M. (2024). The global impact of AI: Mind the gap (IMF Working Paper No. 24/136). International Monetary Fund.

76. World Bank. (2025). Digital progress and trends report 2025: Strengthening AI foundations. World Bank.

77. McElheran, K., Yang, M.-J., Kroff, Z., & Brynjolfsson, E. (2024). The rise of industrial AI in America: Microfoundations of the productivity J-curve(s) (NBER Working Paper No. 32937).

78. Brynjolfsson, E., Rock, D., & Syverson, C. (2017). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. NBER Working Paper No. 24001

79. Calvino, F., & Fontanelli, L. (2023). A portrait of AI adopters across countries: Firm characteristics, assets' complementarities and productivity. OECD Science, Technology and Industry Working Papers, No. 2023/11.

80. McKinsey & Company. (2026, March 25). State of AI trust in 2026: Shifting to the agentic era. https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/tech-forward/state-of-ai-trust-in-2026-shifting-to-the-agentic-era

81. Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., & Vardoulakis, L. M. (2020, April). A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-12).

82. Brant, A., Singh, P., Yin, X., et al. (2025). Performance of a deep learning diabetic retinopathy algorithm in India. JAMA Network Open, 8(3), e250984. https://doi.org/10.1001/jamanetworkopen.2025.0984

83. UNESCO. (2025). AI and the future of education: disruptions, dilemmas and directions

84. Cristia, J. et al. (2017). Technology and child development: evidence from the One Laptop per Child program. *American Economic Journal: Applied Economics*, 9(3), 295–320.

85. Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies, 15(1), Article 6. https://doi.org/10.3390/soc15010006

86. Ojija, F., Ogwu, M. C., Ally, J., John, J. P., Stephano, A., Felix, N., & Tekka, R. (2025). Artificial intelligence-driven solutions for mitigating human–wildlife conflict in biodiversity hotspots. Science Progress, 108(4), 00368504251394584.

87. Noy, S. & Zhang, W. (2023). "Experimental evidence on the productivity effects of generative artificial intelligence." Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586.

88. Cui, Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S. & Salz, T. (2026). "The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers." Management Science. https://doi.org/10.1287/mnsc.2025.00535

89. Dell'Acqua, F., McFowland, E. III, Mollick, E., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F. & Lakhani, K. R. (2023). "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." Harvard Business School Working Paper 24-013. SSRN: https://ssrn.com/abstract=4573321

90. Ali, Z., Muhammad, A., Lee, N., Waqar, M., & Lee, S. W. (2025). Artificial Intelligence for Sustainable Agriculture: A Comprehensive Review of AI-Driven Technologies in Crop Production. Sustainability, 17(5), 2281. https://doi.org/10.3390/su17052281

91. Dhal, S. B., & Kar, D. (2024). Transforming Agricultural Productivity with AI-Driven Forecasting: Innovations in Food Security and Supply Chain Optimization. Forecasting, 6(4), 925–951. https://doi.org/10.3390/forecast6040046

92. Pearlman, K., Wan, W., Shah, S., & Laiteerapong, N. (2025). Use of an AI scribe and electronic health record efficiency. JAMA Network Open, 8(10), e2537000.

93. Afshar, M., Ryan Baumann, M., Resnik, F., Hintzke, J., Gravel Sullivan, A., Wills, G., ... & Gordon, J. (2025). A pragmatic randomized controlled trial of ambient artificial intelligence to improve health practitioner well-being. NEJM AI, 2(12), AIoa2500945.

94. Tierney, A. A., Gayre, G., Hoberman, B., Mattern, B., Ballesca, M., Wilson Hannay, S. B., ... & Lee, K. (2025). Ambient artificial intelligence scribes: learnings after 1 year and over 2.5 million uses. NEJM Catalyst Innovations in Care Delivery, 6(5), CAT-25.

95. Paul A. David (1990, May). The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox. The American Economic ReviewVol. 80, No. 2, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association, pp. 355-361

96. Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. Journal of Economic Perspectives, 14(4), 23–48. https://doi.org/10.1257/jep.14.4.23

97. Shaw, S. D., & Nave, G. (2026). Thinking—fast, slow, and artificial: How AI is reshaping human reasoning and the rise of cognitive surrender. SSRN. https://doi.org/10.2139/ssrn.6097646

98. Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking beyond the hype: Understanding the effects of AI on learning. Educational Psychology Review, 37(2), 45.

99. Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., Van Smeden, M., & Boulesteix, A.-L. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 385, e078451. https://doi.org/10.1136/bmj-2024-078451

100. Rotz, S., Duncan, E., Small, M., Botschner, J., Dara, R., Mosby, I., Reed, M., & Fraser, E. D. G. (2019). The politics of digital agricultural technologies: A preliminary review. Sociologia Ruralis, 59(2), 203–229. https://doi.org/10.1111/soru.12233

101. United Nations General Assembly. (2024). Global Digital Compact (Annex II to the Pact for the Future, A/RES/79/1). United Nations. https://docs.un.org/en/A/RES/79/1

102. UNESCO. (2022). K-12 AI curricula: A mapping of government-endorsed AI curricula. https://www.unesco.org/en/articles/k-12-ai-curricula-mapping-government-endorsed-ai-curricula

103. Almatrafi, O., Johri, A., & Lee, H. (2024, 2024/06/01/). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). Computers and Education Open, 6, 100173. https://doi.org/https://doi.org/10.1016/j.caeo.2024.100173

104. Ma, M., Ng, D. T. K., Liu, Z., & Wong, G. K. W. (2025, 2025/06/01/). Fostering responsible AI literacy: A systematic review of K-12 AI ethics education. Computers and Education: Artificial Intelligence, 8, 100422. https://doi.org/https://doi.org/10.1016/j.caeai.2025.100422

105. Atias, O., & Mawasi, A. (2025, 2025/12/01/). Conceptualizing AI literacies for children and youth: A systematic review on the design of AI literacy educational programs. Computers and Education: Artificial Intelligence, 9, 100491. https://doi.org/https://doi.org/10.1016/j.caeai.2025.100491

106. Kasirzadeh, A., & Gabriel, I. (2025, April). Characterizing AI Agents for Alignment and Governance. https://arxiv.org/abs/2504.21848

107. Wijk, H., Lin, T. R., Becker, J., Jawhar, S., Parikh, N., Broadley, T., ... & Barnes, E. (2025, October). RE-Bench: Evaluating Frontier AI R&D Capabilities of Language Model Agents against Human Experts. In International Conference on Machine Learning (pp. 66772-66832). PMLR.

108. Chan, J. S., Chowdhury, N., Jaffe, O., Aung, J., Sherburn, D., Mays, E., ... & Weng, L. (2025, May). Mle-bench: Evaluating machine learning agents on machine learning engineering. In International Conference on Learning Representations (Vol. 2025, pp. 50466-50494).

109. Pichai, S. (2026, April 22). Cloud Next '26: Momentum and innovation at Google scale. Google Blog. https://blog.google/innovation-and-ai/infrastructure-and-cloud/google-cloud/cloud-next-2026-sundar-pichai/

110. Cybersecurity and Infrastructure Security Agency. (2025, January 14). CISA, JCDC, government and industry partners publish AI cybersecurity collaboration playbook. U.S. Department of Homeland Security. https://www.cisa.gov/news-events/news/cisa-jcdc-government-and-industry-partners-publish-ai-cybersecurity-collaboration-playbook

111. Liu, Y., Zhao, Y., Lyu, Y., Zhang, T., Wang, H., & Lo, D. (2025). "Your AI, my shell": Demystifying prompt injection attacks on agentic AI coding editors. https://doi.org/10.48550/arXiv.2509.22040

112. Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1820. https://doi.org/10.15779/Z38RV0D

113. Schiff, K. J., Schiff, D. S., & Bueno, N. S. (2025). The liar's dividend: Can politicians claim misinformation to evade accountability?. American Political Science Review, 119(1), 71-90.

114. Chan, Y.-T., et al. (2024). Assessing the article screening efficiency of artificial intelligence for systematic reviews. Journal of Dentistry, 149, 105259. https://doi.org/10.1016/j.jdent.2024.105259

115. Delgado-Licona, F., Alsaiari, A., Dickerson, H., Klem, P., Ghorai, A., Canty, R. B., Bennett, J. A., Jha, P., Mukhin, N., Li, J., López-Guajardo, E. A., Sadeghi, S., Bateni, F., & Abolhasani, M. (2025). Flow-driven data intensification to accelerate autonomous inorganic materials discovery. Nature Chemical Engineering, 2, 436–446. https://doi.org/10.1038/s44286-025-00249-z

116. Chan, A., Wei, K., Huang, S., Rajkumar, N., Perrier, E., Lazar, S., Hadfield, G. K., & Anderljung, M. (2025). Infrastructure for AI Agents. http://arxiv.org/abs/2501.10114

117. Kapoor, S., Stroebl, B., Siegel, Z. S., Nadgir, N., & Narayanan, A. AI Agents That Matter. Transactions on Machine Learning Research.

118. Zhu, L., Lu, Q., Ding, M. et al. Designing meaningful human oversight in AI. AI Ethics 6, 286 (2026). https://doi.org/10.1007/s43681-026-01147-7

119. Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. Journal of Computational Social Science, 7, 549–569. https://doi.org/10.1007/s42001-024-00250-1

120. Djiré, A. E., Kaboré, A. K., Samhi, J., et al. (2026). Learned or memorized? Quantifying memorization advantage in code LLMs. In Proceedings of the International Conference on Software Engineering. https://arxiv.org/abs/2604.13997

121. Gowal, S., Bunel, R., Stimberg, F., Stutz, D., Ortiz-Jimenez, G., Kouridi, C., ... & Kohli, P. (2025). SynthID-Image: Image watermarking at internet scale. arXiv preprint arXiv:2510.09263.

122. United Nations Human Rights Council. Expert Mechanism on the Right to Development. (2024). AI, cultural rights and the right to development (A/HRC/EMRTD/11/CRP.2). United Nations. https://undocs.org/A/HRC/EMRTD/11/CRP.2

123. Brennan Center for Justice. (2025). Gauging AI threat to free and fair elections. https://www.brennancenter.org/our-work/analysis-opinion/gauging-ai-threat-free-and-fair-elections

124. Hackenburg, K., Tappin, B. M., Hewitt, L., Saunders, E., Black, S., Lin, H., Fist, C., Margetts, H., Rand, D. G., & Summerfield, C. (2025). The levers of political persuasion with conversational AI. Science, 390(6777), eaea3884. https://doi.org/10.1126/science.aea3884

125. Santos, F. P., Lelkes, Y., & Levin, S. A. (2021). Link recommendation algorithms and dynamics of polarization in online social networks. Proceedings of the National Academy of Sciences, 118(50), e2102141118.

126. Cho, J., Ahmed, S., Hilbert, M., Liu, B., & Luu, J. (2020). Do search algorithms endanger democracy? An experimental investigation of algorithm effects on political polarization. Journal of broadcasting & Electronic media, 64(2), 150-172.

127. Feezell, J. T., Wagner, J. K., & Conroy, M. (2021). Exploring the effects of algorithm-driven news sources on political behavior and polarization. Computers in human behavior, 116, 106626.

128. Argyle, L. P. (2025). Political persuasion by artificial intelligence. Science, 390(6777), 983-984.

129. Lin, H., Czarnek, G., Lewis, B., White, J. P., Berinsky, A. J., Costello, T., ... & Rand, D. G. (2025). Persuading voters using human–artificial intelligence dialogues. Nature, 1-8.

130. Myra Cheng et al. (2026). ,Sycophantic AI decreases prosocial intentions and promotes dependence.Science391,eaec8352(2026).DOI:10.1126/science.aec8352

131. Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., & Jurafsky, D. (2026). Sycophantic AI decreases prosocial intentions and promotes dependence. Science, 391(6792), eaec8352. https://doi.org/10.1126/science.aec8352

132. Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., ... & Pollak, T. A. (2026). Artificial intelligence-associated delusions and large language models: risks, mechanisms of delusion co-creation, and safeguarding strategies. The Lancet Psychiatry, 13(6), 522-530.

133. Balan, R., & Gumpel, T. P. (2025). ChatGPT Clinical Use in Mental Health Care: Scoping Review of Empirical Evidence. JMIR Mental Health, 12, e81204.

134. Hudon, A., & Stip, E. (2025). Delusional experiences emerging from AI chatbot interactions or "AI Psychosis". JMIR Mental Health, 12(1), e85799.

135. Osler, L. (2026). Hallucinating with AI: distributed delusions and "AI psychosis". Philosophy & Technology, 39(1), 30.

136. Askell, A., Bai, Y., Chen, A., Drain, D., Ganguli, D., Henighan, T., Jones, A., Joseph, N., Mann, B., DasSarma, N., Elhage, N., Hatfield-Dodds, Z., Hernandez, D., Kernion, J., Ndousse, K., Olsson, C., Amodei, D., Brown, T., Clark, J., ... Kaplan, J. (2021). A general language assistant as a laboratory for alignment (arXiv:2112.00861). arXiv. https://doi.org/10.48550/arXiv.2112.00861

137. Organisation for Economic Co-operation and Development. (2024). Facts not fakes: Tackling disinformation, strengthening information integrity. OECD Publishing. https://doi.org/10.1787/d909ff7a-en

138. Waight, H., Yang, E., Yuan, Y., et al. (2026). State media control influences large language models. Nature. https://doi.org/10.1038/s41586-026-10506-7

139. Kalina Bontcheva (2024). Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities. https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation_-White-Paper-v8.pdf

140. Laudrain, A. (2026, April 29). When AI governs (dis)information: Five lessons for democracy. Center for Security Studies (CSS), ETH Zurich. https://css.ethz.ch/en/center/CSS-news/2026/04/when-ai-governs-disinformation-five-lessons-for-democracy.html

141. Boine, C. (2023). Emotional attachment to AI companions and European law. MIT Case Studies in Social and Ethical Responsibilities of Computing (Winter 2023). https://doi.org/10.21428/2c646de5.db67ec7f

142. Department of Enterprise, Tourism and Employment. (2026, February 4). General scheme of the Regulation of Artificial Intelligence Bill 2026. Government of Ireland. https://www.gov.ie/en/department-of-enterprise-tourism-and-employment/publications/general-scheme-of-the-regulation-of-artificial-intelligence-bill-2026

143. United States Congress. Senate. (2025). S. 3062—GUARD Act: Guidelines for User Age-verification and Responsible Dialogue Act of 2025 (119th Congress). Congress.gov. https://www.congress.gov/bill/119th-congress/senate-bill/3062/text

144. European Commission. (2026, April 29). Blueprint for an age verification solution to help protect minors online. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/factpages/blueprint-age-verification-solution-help-protect-minors-online

145. Reuters. (2026, April 24). From Australia to Europe, countries move to curb children's social media access. https://www.reuters.com/legal/government/australia-europe-countries-move-curb-childrens-social-media-access-2026-04-24/

146. Morrin H, Nicholls L, Levin M et al. (2026). Artificial intelligence-associated delusions and large language models: risks, mechanisms of delusion co-creation, and safeguarding strategies. The Lancet Psychiatry, 2026; 13, 522-530

147. Examining the harm of AI chatbots, Hearing before the Subcomm. on Crime and Counterterrorism of the S. Comm. on the Judiciary, 119th Cong. (2025) (testimony of Megan Garcia). https://www.judiciary.senate.gov/download/2025-09-16-pm_-testimony_-garciapdf

148. Solove, D. J. (2025). Artificial intelligence and privacy. Florida Law Review, 77. https://scholarship.law.ufl.edu/flr/vol77/iss1/1

149. Office of the United Nations High Commissioner for Human Rights. (2025). The right to privacy in the digital age (UN Doc. A/HRC/60/45). https://www.ohchr.org/en/documents/thematic-reports/ahrc6045-right-privacy-digital-age-reportoffice-united-nations-high

150. Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the FinTech era. Journal of Financial Economics, 143(1), 30–56. https://doi.org/10.1016/j.jfineco.2021.05.047

151. UNESCO, IRCAI, & University College London. (2024). Challenging systematic prejudices: An investigation into bias against women and girls in large language models. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000388971

152. Shah, S. S. (2024). Gender bias in artificial intelligence: Empowering women through digital literacy. Premier Journal of Artificial Intelligence, 1, 1000088. https://doi.org/10.70389/PJAI.1000088

153. Haider, S. A., Borna, S., Gomez-Cabello, C. A., et al. (2026). The algorithmic divide: A systematic review on AI-driven racial disparities in healthcare. Journal of Racial and Ethnic Health Disparities, 13, 188–217. https://doi.org/10.1007/s40615-024-02237-0

154. International Telecommunication Union. (2025). Joint statement on AI and the rights of the child. International Telecommunication Union. https://www.itu.int/hub/publication/d-str-cyb_joint-2025

155. Johnson, A. K., Winther, D. K., & Bhargava, A. (2026). Artificial intelligence and child sexual exploitation and abuse: Emerging risks and implications for children's rights (Issue brief). United Nations Children's Fund (UNICEF). https://www.unicef.org/media/178571/file/UNICEF%20AI%20CSEA%20Brief_FINAL3.pdf

156. Thiel, D. (2023). Identifying and Eliminating CSAM in Generative ML Training Data and Models. Stanford Digital Repository. Available at https://purl.stanford.edu/kh752sm9123

157. Internet Watch Foundation. (2026). Harm without limits: AI child sexual abuse material through the eyes of our Analysts. https://www.iwf.org.uk/media/hl1nvdti/iwf-ai-csam-report-2026.pdf

158. Goodacre, E.J. and Gibson, J.L. (2026) AI in the early years: Examining the implications of GenAI toys for young children. Cambridge: University of Cambridge (unpublished report, available via Apollo repository).

159. Kurian, N. (2025). Developmentally aligned AI: a framework for translating the science of child development into AI design. AI, Brain and Child, 1(1). https://doi.org/10.1007/s44436-025-00009-z

160. Livingstone, S., & Sylwander, K. R. (2025). Conceptualizing age-appropriate social media to support children's digital futures. British Journal of Developmental Psychology. https://doi.org/10.1111/bjdp.70006

161. Xiao, W., & Gonçalves, A. (2025). Intelligent toys, complex questions: A literature review of artificial intelligence in children's toys and devices. Big Data & Society, 12(4), 20539517251389860.

162. Grimmelikhuijsen, S. (2022). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. Public Administration Review, 82(4), 706–718. https://doi.org/10.1111/puar.13483

163. Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. New York University Law Review Online, 94, 15–55. https://ssrn.com/abstract=3333423

164. Mantelero, A. (2022). Human rights impact assessment and AI. In Beyond data: Human rights, ethical and social impact assessment in AI (pp. 45-91). The Hague: TMC Asser Press.

165. Mantelero, A., & Esposito, M. S. (2021). An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. Computer Law and Security Review, 41. https://doi.org/10.1016/j.clsr.2021.105561

166. United Nations Educational, Scientific and Cultural Organization. (2025). How should children's rights be integrated into AI governance? https://www.unesco.org/en/articles/how-should-childrens-rights-be-integrated-ai-governance

167. Livingstone, S. & Pothong, K. (2025). Child Rights Impact Assessment: A Policy Tool for aRights-Respecting Digital Environment. Policy & Internet.

168. Denain, J.-S., & Barry, A. (2026, April 16). Have AI capabilities accelerated? Epoch AI. https://epoch.ai/blog/have-ai-capabilities-accelerated

169. Model Evaluation & Threat Research (METR). (2026, May 8). Task-completion time horizons of frontier AI models. https://metr.org/time-horizons/

170. Juniewicz, I. (2026, February 26). Hyperscaler capex has quadrupled since GPT-4's release. Epoch AI. https://epoch.ai/data-insights/hyperscaler-capex-trend

171. Epoch AI. (2026, May 28). Data on AI companies. https://epoch.ai/data/ai-companies?view=graph&tab=revenue&showTotal=true

172. Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. Proceedings of the IEEE, 109(5), 612-634.

173. AI Alliance Network. (2025). AI Horizons: What will AI technologies look like in 10 years? A Research Project. November 2025, p. 83. Based on 21 foresight sessions and 32 in-depth interviews with over 270 AI researchers from 36 countries.

174. Gommers, J., Hernström, V., Josefsson, V., Sartor, H., Schmidt, D., Hjelmgren, A., ... & Lång, K. (2026). Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study: a randomised, controlled, non-inferiority, single-blinded, population-based, screening-accuracy trial. The Lancet, 407(10527), 505-514.

175. Reichstein, M., et al. (2025). Early warning of complex climate risk with integrated artificial intelligence. Nature Communications, 16, 2564. https://www.nature.com/articles/s41467-025-57640-w

176. Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports, 15(1), 17458.

177. Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. npj Science of Learning, 10(1), 29.

178. Delgado-Licona, F., Alsaiari, A., Dickerson, H., Klem, P., Ghorai, A., Canty, R. B., ... & Abolhasani, M. (2025). Flow-driven data intensification to accelerate autonomous inorganic materials discovery. Nature Chemical Engineering, 2(7), 436-446.

179. Rotz, S., Duncan, E., Small, M., Botschner, J., Dara, R., Mosby, I., ... & Fraser, E. D. (2019). The politics of digital agricultural technologies: a preliminary review. Sociologia ruralis, 59(2), 203-229.

180. Afshar, M., Ryan Baumann, M., Resnik, F., Hintzke, J., Gravel Sullivan, A., Wills, G., ... & Gordon, J. (2025). A pragmatic randomized controlled trial of ambient artificial intelligence to improve health practitioner well-being. NEJM AI, 2(12), AIoa2500945.

181. Chen, M., Wu, Y., Ma, J., Jia, X., Gao, C., Zhao, F., & Qiao, Y. (2026). Independent and collaborative performance of large language models and healthcare professionals in diagnosis and triage. npj Digital Medicine.

182. Tao, X., Zhou, S., Ding, K., Li, S., Li, Y., Wu, B., ... & Han, S. (2026). An LLM chatbot to facilitate primary-to-specialist care transitions: a randomized controlled trial. Nature Medicine, 1-9.

183. Costa-Gomes, B., Tolmachev, P., Taysom, E., Sounderajah, V., Richardson, H., Schoenegger, P., ... & King, D. (2026). Public use of a generalist LLM chatbot for health queries. Nature Health, 1-8.

184. Scientific Computing World. (2025, August 5). AI health assistant displays high diagnostic accuracy in tests. https://www.scientific-computing.com/article/sber-healths-gigachat-powered-ai-health-assistant-displays-high-diagnostic-accuracy-tests

185. MASTEL, P. M., de Dieu Nyandwi, J., Rutunda, S., & Kabanda, K. (2025). Mbaza RBC: Deploying and evaluation of an LLM powered Chatbot for Community Health Workers in Rwanda. In Workshop on Large Language Models and Generative AI for Health at AAAI 2025.

186. Mateen, B. A., Williams, G., Korom, R., Mwaniki, P., Emmanuel-Fabula, M., & Agweyu, A. (2026). Learning Effects from A GenAI-based Clinical Decision Support System in Primary Healthcare. medRxiv, 2026-05.

187. OECD (2025), Results from TALIS 2024: The State of Teaching, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/90df6235-en.

188. OECD (2023), PISA 2022 Results (Volume II): Learning During – and From – Disruption, PISA, OECD Publishing, Paris, https://doi.org/10.1787/a97db61c-en .

189. Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K–12 education. npj Science of Learning, 10(1), 29. https://www.nature.com/articles/s41539-025-00320-7

190. OECD (2023), PISA 2022 Results (Volume II): Learning During – and From – Disruption, PISA, OECD Publishing, Paris, https://doi.org/10.1787/a97db61c-en.

191. OECD (2023), PISA 2022 Results (Volume II): Learning During – and From – Disruption, PISA, OECD Publishing, Paris, https://doi.org/10.1787/a97db61c-en.

192. Vodafone Foundation.(2025) AI in European Schools: A European Report --- comparing seven countries

193. World Food Programme. (2024). HungerMap LIVE: Global hunger monitoring. https://hungermap.wfp.org/

194. Yakov and Partners. (2024). Artificial intelligence in Russia's agricultural sector: Hype or real money? https://yakovpartners.com/publications/ai-in-agriculture/

195. Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences, 122(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

196. Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports, 15, 17458. https://doi.org/10.1038/s41598-025-97652-6

197. Shneiderman, B. (2022). Human-Centered AI. Oxford University Press.

198. Sparling, T. M., Offner, C., Deeney, M., Denton, P., Bash, K., Juel, R., ... & Kadiyala, S. (2024). Intersections of climate change with food systems, nutrition, and health: an overview and evidence map. Advances in Nutrition, 15(9), 100274.

199. Reichstein, M., et al. (2025). Early warning of complex climate risk with integrated artificial intelligence. Nature Communications, 16, 2564. https://doi.org/10.1038/s41467-025-57640-w

200. Becker-Reshef, I., Justice, C., Barker, B., Humber, M., Rembold, F., Bonifacio, R., Zappacosta, M., Budde, M., Magadzire, T., Shitote, C., Pound, J., Constantino, A., Nakalembe, C., Mwangi, K., Sobue, S., Newby, T., Whitcraft, A., Jarvis, I., & Verdin, J. (2020). Strengthening agricultural decisions in countries at risk of food insecurity: The GEOGLAM Crop Monitor for Early Warning. Remote Sensing of Environment, 237, 111553. https://doi.org/10.1016/j.rse.2019.111553

201. Food and Agriculture Organization of the United Nations. (2025). Evidence in action: How anticipatory cash transfers reduce humanitarian needs and strengthen resilience in Somalia. https://www.fao.org/agrifood-economics/publications/detail/en/c/1756210/

202. World Food Programme. (2025). WFP's evidence base on anticipatory action 2015–2024. https://www.wfp.org/publications/wfps-evidence-base-anticipatory-action-2015-2024

203. Nakalembe, C., Kerner, H. R., Zvonkov, I., Humber, M., Galvez, A. S., Venturini, S., & Becker Reshef, I. (2025). A framework for EO based National Agricultural Monitoring for the African context. npj Sustainable Agriculture, 3, 45. https://www.nature.com/articles/s44264-025-00083-z

204. Nowak, A. C., et al. (2024). Opportunities to strengthen Africa's efforts to track national level climate adaptation. Nature Climate Change, 14, 876–882.https://www.nature.com/articles/s41558-024-02054-7

205. Karger, E., Kuusela, O., Abaluck, J., Bryan, K. A., Halperin, B., Jones, T. R., et al. (2026). Forecasting the economic effects of AI (NBER Working Paper No. w35046). National Bureau of Economic Research. https://doi.org/10.3386/w35046

206. Goldman Sachs (2023). Generative AI Could Raise Global GDP by 7 Percent. https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent

207. Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: a review. Artificial intelligence review, 55(1), 589-656.

208. United Nations News. (2026, February 18). Artists face steep income decline due to AI, UNESCO finds. United Nations. https://news.un.org/en/story/2026/02/1166989

209. Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. American Economic Journal: Macroeconomics, 13(1), 333-372.

210. Gmyrek, P., Winkler, H., & Garganta, S. (2024). Buffer or Bottleneck? Employment Exposure to Generative AI and the Digital Divide in Latin America (ILO Working Paper 121 / World Bank Policy Research Working Paper 10863). International Labour Organization & World Bank.

211. Autor, D., Chin, C., Salomons, A., & Seegmiller, B. (2024). New frontiers: The origins and content of new work, 1940–2018. Quarterly Journal of Economics, 139(3), 1399–1465.

212. The Conversation. (2026, March 18). Tech companies are blaming massive layoffs on AI. What's really going on? https://theconversation.com/tech-companies-are-blaming-massive-layoffs-on-ai-whats-really-going-on-278314

213. Society for Human Resource Management. (2026, May). The AI layoffs narrative: Real transformation, or scapegoat? https://www.shrm.org/topics-tools/news/technology/ai-layoffs-transformation-scapegoat

214. OpenAI. (2026, February 27). Company communication accompanying a US$110 billion private funding round [Company communication]

215. Rodrik, D., & Sabel, C. (2020). Building a Good Jobs Economy (HKS Faculty Research Working Paper RWP20-001). Harvard Kennedy School.

216. Acemoglu, D. (2025). The simple macroeconomics of AI. Economic Policy, 40(121), 13–58. Acemoglu's headline figure is a TFP gain of no more than 0.71% over ten years.

217. Korinek, A., & Suh, D. (2024). Scenarios for the Transition to AGI (NBER Working Paper No. 32255). In their full-automation scenario, output rises sharply while wages collapse once automation crosses a critical threshold.

218. Karger, E., Kuusela, O., Abaluck, J., Bryan, K., Halperin, B., Jones, T., et al. (2026). Forecasting the Economic Effects of AI. Federal Reserve Bank of Chicago and Forecasting Research Institute, March 2026.

219. Jones, C. I., & Tonetti, C. (2026). Past Automation and Future AI: How Weak Links Tame the Growth Explosion. Working paper presented at the Bendheim Center for Finance, Princeton University, March 2026.

220. Trammell, P., & Korinek, A. (2025). Economic Growth under Transformative AI (NBER Working Paper No. 31815, revised September 2025).

221. OECD (2024). Artificial Intelligence, Data and Competition (OECD Artificial Intelligence Papers No. 18). OECD Publishing, Paris. https://doi.org/10.1787/e7e88884-en

222. Acemoglu, D., & Restrepo, P. (2026). Automation and rent dissipation: Implications for wages, inequality, and productivity. Quarterly Journal of Economics, 141(2), 1521–1579.

223. Liu, Y., Zhao, Y., Lyu, Y., Zhang, T., Wang, H., & Lo, D. (2025). "Your AI, my shell": Demystifying prompt injection attacks on agentic AI coding editors. arXiv. https://doi.org/10.48550/arXiv.2509.22040

224. Regilme, S. S. F. (2024). Artificial Intelligence Colonialism: Environmental Damage, Labor Exploitation, and Human Rights Crises in the Global South. SAIS Review of International Affairs, 44(2), 75–92. https://muse.jhu.edu/pub/1/article/950958

225. Barnett-Itzhaki, Z. (2026). The water footprint of artificial intelligence: Emerging solutions and governance imperatives. Water Research, 299, 125866. https://doi.org/10.1016/j.watres.2026.125866

226. International Energy Agency. (2025). Global Critical Minerals Outlook 2025 – Analysis (p. 312). https://iea.blob.core.windows.net/assets/ef5e9b70-3374-4caa-ba9d-19c72253bfc4/GlobalCriticalMineralsOutlook2025.pdf

227. Zhu, L., & Lu, Q. (2026). Verifiability-First AI Engineering in the Era of AIware: A Conceptual Framework, Design Principles, and Architectural Patterns for Scalable Verification. Design Principles, and Architectural Patterns for Scalable Verification (January 07, 2026).

228. UN Scientific Advisory Board. (2026, March 19). AI deception: Brief of the Scientific Advisory Board. United Nations. https://www.un.org/scientific-advisory-board/en/ai-deception

229. Bengio, Y., Clare, S., Prunkl, C., Rismani, S., Andriushchenko, M., Bucknall, B., ... & Zhu, L. (2025). International AI Safety Report 2025: First Key Update: Capabilities and Risk Implications. arXiv preprint arXiv:2510.13653.

230. United Nations Secretary-General. (2024). Human rights due diligence guidance on digital technology use. https://www.ohchr.org/sites/default/files/2024-08/digital-technology-use-guidance-sg-1-en.pdf

231. AI Equality Toolbox. (2025). Human rights impact assessment methodology. https://aiequalitytoolbox.com/tools/hria-workbook/

232. Baldé, C. P., Kuehr, R., Yamamoto, T., McDonald, R., D'Angelo, E., Althaf, S., Bel, G., Deubzer, O., Fernandez-Cubillo, E., Forti, V., Gray, V., Herat, S., Honda, S., Iattoni, G., Khetriwal, D. S., Luda di Cortemiglia, V., Lobuntsova, Y., Nnorom, I., Pralat, N., & Wagner, M. (2024). The global e-waste monitor 2024: Electronic waste rising five times faster than documented e-waste recycling. United Nations Institute for Training and Research & International Telecommunication Union. https://ewastemonitor.info/the-global-e-waste-monitor-2024/

233. International Panel on the Information Environment. (2024). Expert survey on the global information environment 2024: Searching for solutions (SFP2024-1). https://www.ipie.info/research/sfp2024-1

234. Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. Social Network Analysis and Mining, 13(1), 30. https://doi.org/10.1007/s13278-023-01028-5

235. James, K., Perveen, S., & Jacobson, B. (2025). Drained by data: The cumulative impact of data centers on regional water stress. Ceres. https://www.ceres.org/resources/reports/drained-by-data-the-cumulative-impact-of-data-centers-on-regional-water-stres

236. Office of the United Nations High Commissioner for Human Rights. (2024). Taxonomy of generative AI-related human rights harms. https://www.ohchr.org/sites/default/files/documents/issues/expression/statements/2025-10-24-joint-declaration-artificial-intelligence.pdf

237. Patel, A. (2025, May 19). Freedom of expression, artificial intelligence and elections. United Nations Educational, Scientific and Cultural Organization (UNESCO) & United Nations Development Programme (UNDP). https://www.undp.org/publications/freedom-expression-artificial-intelligence-and-elections

238. Scharfbillig, M., Lewandowsky, S., Altay, S., Van Alstyne, M., Kozyreva, A., Hertwig, R., Lorenz-Spreen, P., DiResta, R., Valenzuela, S., Egidy, S., Quattrociocchi, W., & Orben, A. (2026). Fractured reality: How democracy can win the global struggle over the information space (JRC144603). Publications Office of the European Union. https://doi.org/10.2760/9358883

239. Observatory on Information and Democracy. (2024). Information ecosystems and troubled democracy: A global synthesis of the state of knowledge on news, media, AI, and data governance. https://observatory.informationdemocracy.org/report/information-ecosystem-and-troubled-democracy/

240. Solove, D. J. (2025). Artificial intelligence and privacy. Florida Law Review, 77. https://scholarship.law.ufl.edu/flr/vol77/iss1/1

241. Office of the United Nations High Commissioner for Human Rights (2025). The right to privacy in the digital age. URL https://www.ohchr.org/en/documents/thematic-reports/ahrc6045-right-privacy-digital-age-reportoffice-united-nations-high. UN Doc. A/HRC/60/45.

242. Council of Europe. Steering Committee on Media and Information Society. (2025). Guidance note on the implications of generative artificial intelligence for freedom of expression (CDMSI(2025)15rev). https://rm.coe.int/cdmsi-2025-15rev-guidance-note-on-the-implications-of-generative-artif/488029df80

243. European Union. (2024, June 13). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L series. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

244. Expert Mechanism on the Right to Development. (2024). Artificial intelligence, cultural rights and the right to development (A/HRC/EMRTD/13/CRP.1). United Nations Human Rights Council. https://www.ohchr.org/sites/default/files/documents/issues/development/emd/session13/a-hrc-emrtd-13-crp-1.pdf

245. Council of Europe, Steering Committee on Media and Information Society. (2025). Guidance note on the implications of generative artificial intelligence for freedom of expression (CDMSI(2025)15rev). https://rm.coe.int/cdmsi-2025-15rev

246. Waight, H., Yang, E., Yuan, Y., et al. (2026). State media control influences large language models. Nature. https://doi.org/10.1038/s41586-026-10506-7

247. Li, P., Yang, J., Islam, M. A., & Ren, S. (2025). Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models. Communications of the ACM, 68(7), 54–61. https://doi.org/10.1145/3724499

248. Elsworth, C., Huang, K., Patterson, D., Schneider, I., Sedivy, R., Goodman, S., ... & Manyika, J. (2025). Measuring the environmental impact of delivering AI at Google Scale. arXiv preprint arXiv:2508.15734.

249. Arsenault, A. C., & Kreps, S. (2026). Whose voice counts? The role of large language models in public commenting. Big Data & Society, 13(1). https://doi.org/10.1177/20539517261419341

250. Alslaity, A., Chan, G., & Orji, R. (2023). A panoramic view of personalization based on individual differences in persuasive and behavior change interventions. Frontiers in Artificial Intelligence, 6, 1125191. https://doi.org/10.3389/frai.2023.1125191

251. Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. Science, 385, eadq1814. https://doi.org/10.1126/science.adq1814

252. Boissin, E., Costello, T. H., Spinoza-Martín, D., Rand, D. G., & Pennycook, G. (2025). Dialogues with large language models reduce conspiracy beliefs even when the AI is perceived as human. PNAS Nexus, 4(11), pgaf325. https://doi.org/10.1093/pnasnexus/pgaf325

253. Schroeder, D. T., Cha, M., Baronchelli, A., Bostrom, N., Christakis, N. A., Garcia, D., ... & Kunst, J. R. (2026). How malicious AI swarms can threaten democracy. Science, 391(6783), 354-357.

254. Hackenburg, K., Tappin, B. M., Hewitt, L., Saunders, E., Black, S., Lin, H., Fist, C., Margetts, H., Rand, D. G., & Summerfield, C. (2025). The levers of political persuasion with conversational AI. Science, 390(6777), eaea3884. https://doi.org/10.1126/science.aea3884

255. Germano, F., Gómez, V., & Sobbrio, F. (2025). Ranking for engagement: How social media algorithms fuel misinformation and polarization. Barcelona School of Economics, Working Paper No. 1501. https://bw.bse.eu/wp-content/uploads/2025/07/1501.pdf

256. Council of Europe CDMSI. (2025). Guidance Note on Generative AI and Freedom of Expression. CDMSI(2025) https://rm.coe.int/cdmsi-2025-15rev-guidance-note-on-the-implications-of-generative-artif/488029df80

257. Buyl, M., Rogiers, A., Noels, S., Bied, G., Dominguez-Catena, I., Heiter, E., ... & De Bie, T. (2026). Large language models reflect the ideology of their creators. npj Artificial Intelligence, 2(1), 7.

258. Wang, P., Zhang, L.-Y., Tzachor, A., & Chen, W.-Q. (2024). E-waste challenges of generative artificial intelligence. Nature Computational Science, 4, 818–823. https://doi.org/10.1038/s43588-024-00700-3

259. Schroeder, D. T., Cha, M., Baronchelli, A., Bostrom, N., Christakis, N. A., Garcia, D., ... & Kunst, J. R. (2026). How malicious AI swarms can threaten democracy. Science, 391(6783), 354-357.

260. Council of Europe. (2024). Council of Europe framework convention on artificial intelligence and human rights, democracy and the rule of law (Council of Europe Treaty Series No. 225). https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

261. Observatory on Information and Democracy. (2024). Information ecosystems and troubled democracy: A global synthesis of the state of knowledge on news, media, AI, and data governance. https://observatory.informationdemocracy.org/report/information-ecosystem-and-troubled-democracy/

262. Council of Europe, Steering Committee on Media and Information Society. (2025). Guidance note on the implications of generative artificial intelligence for freedom of expression (CDMSI(2025)15rev). https://rm.coe.int/cdmsi-2025-15rev

263. OECD. (2024). Facts not fakes: Tackling disinformation, strengthening information integrity. OECD Publishing. https://doi.org/10.1787/d909ff7a-en

264. United Nations General Assembly. (2017). Promotion and protection of human rights: Human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms: Report of the Third Committee, 72nd session (A/72/...). United Nations Digital Library. http://digitallibrary.un.org/record/1326669

265. Penney, J. W. (2025). Chilling effects: Repression, conformity, and power in the digital age. Cambridge University Press. https://doi.org/10.1017/9781108918022

266. UN Women. (2022). Tipping point: The chilling escalation of online violence against women in the public sphere. UN Women. https://www.unwomen.org/sites/default/files/2025-12/tipping-point-the-chilling-escalation-of-violence-against-women-in-the-public-sphere-in-the-age-of-ai-en.pdf

267. United Nations Educational, Scientific and Cultural Organization. (2024). Challenging systematic prejudices: An investigation into bias against women and girls in large language models. https://unesdoc.unesco.org/ark:/48223/pf0000388971

268. Chowdhury, R., & Lakshmi, D. (2023). "Your opinion doesn't matter, anyway": Exposing technology-facilitated gender-based violence in an era of generative AI (2nd ed.). UNESCO.

269. United Nations Educational, Scientific and Cultural Organization. (2024, March 7). Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes. https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes

270. Fung, P. (2019, June 30). This is why AI has a gender problem. World Economic Forum. https://www.weforum.org/stories/2019/06/this-is-why-ai-has-a-gender-problem/

271. United Nations Conference on Trade and Development. (2025). Technology and innovation report 2025: The AI divide. https://unctad.org/publication/technology-and-innovation-report-2025

272. Ahmed, N., & Wahed, M. (2020). The De-democratization of AI: Deep learning and the compute divide in artificial intelligence research. arXiv preprint arXiv:2010.15581.

273. Coeckelbergh, M. (2026) Technofascism: AI, Big Tech, and the New Authoritarianism. AI & Society https://doi.org/10.1007/s00146-026-02862-9

274. Varoufakis, Y. (2024). Technofeudalism: What killed capitalism. Melville House.

275. Kalluri, P. R., Agnew, W., Cheng, M., Owens, K., Soldaini, L., & Birhane, A. (2025). Computer-vision research powers surveillance technology. Nature, 643(8070), 73-79. https://www.nature.com/articles/s41586-025-08972-6

276. Fola-Rose, A., Solomon, E., Bryant, K., & Woubie, A. (2024, August). A systematic review of facial recognition methods: Advancements, applications, and ethical dilemmas. In Proceedings of the 2024 IEEE International Conference on Information Reuse and Integration for Data Science (IRI 2024) (pp. 314–319). IEEE. https://doi.org/10.1109/IRI62200.2024.00070

277. Fussey, P., & Murray, D. (2025). Facial Recognition Surveillance: Policing and Human Rights in the Age of Artificial Intelligence. Oxford University Press.

278. Office of the United Nations High Commissioner for Human Rights. (2024). Mapping report: Human rights and new and emerging digital technologies (A/HRC/56/45). https://www.ohchr.org/en/documents/reports/mapping-report-human-rights-and-new-and-emerging-digital-technologies

279. Secretary-General. (2024). Human rights in the administration of justice (A/79/296). United Nations. https://digitallibrary.un.org

280. American Civil Liberties Union, More than a Dozen Wrongful Arrests Due to Police Reliance on Facial Recognition Technology (2025), https://www.aclu.org/news/privacy-technology/more-than-a-dozen-wrongful-arrests-due-to-police-reliance-on-facial-recognition-technology: Documentation of at least 14 publicly known wrongful arrests in the United States attributed to police use of facial recognition; in nearly all cases the persons wrongfully arrested were Black.

281. Saxena, D., & Guha, S. (2024). Algorithmic harms in child welfare: Uncertainties in practice, organization, and street-level decision-making. ACM Journal on Responsible Computing, 1(1), Article 2, 1–32. https://doi.org/10.1145/3616473

282. German Marshall Fund. (2024). Spitting Images: Tracking Deepfakes and Generative AI in Elections.

283. International Institute for Democracy and Electoral Assistance. (2024). The 2024 global elections super-cycle. https://www.idea.int/initiatives/the-2024-global-elections-supercycle

284. Recorded Future. (2024). 2024 Deepfakes and Election Disinformation Report: Key Findings and Mitigation Strategies.

285. Associated Press. (2025, June 13). New Hampshire jury acquits consultant behind AI robocalls mimicking Biden on all charges.

286. Federal Communications Commission. (2024, September). $6 million fine against Steven Kramer for AI-generated robocalls.

287. Global Witness. (2024). What Happened on TikTok Around the Romanian Elections?

288. IFES. (2024). The Romanian 2024 Election Annulment: Addressing Emerging Threats to Electoral Integrity.

289. Associated Press. (2024). Election disinformation takes a big leap with AI being used to deceive worldwide.

290. United Nations. (1966). International Covenant on Civil and Political Rights. Articles 17, 18, 19 and 25.

291. Council of Europe. (1950). Convention for the Protection of Human Rights and Fundamental Freedoms. Articles 8, 9 and 10; Protocol No. 1, Article 3.

292. EQUATE Language AI Readiness Index (https://equate.vercel.app/en)

293. Han, W., Zhang, Y., Chen, Z., Liu, B., Lin, H., Zhang, B., ... & Zheng, Y. (2025). MuBench: Assessment of Multilingual Capabilities of Large Language Models Across 61 Languages. arXiv preprint arXiv:2506.19468.

294. Lissak, S., Calderon, N., Shenkman, G., Ophir, Y., Fruchter, E., Klomek, A. B., & Reichart, R. (2024, June). The colorful future of llms: Evaluating and improving llms as emotional supporters for queer youth. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 2040-2079).

295. Gamboa, L. C. L., Feng, Y., & Lee, M. (2025, November). Social Bias in Multilingual Language Models: A Survey. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (pp. 27845-27868).

296. Choudhury, M., Sitaram, S., Vashistha, A., et al. (2026). AI for the Global South: 12 critical research questions for the next decade. AI for the Global South (AI4GS), Mohamed bin Zayed University of Artificial Intelligence. https://ai4gs.github.io/

297. Bartl, M., Mandal, A., Leavy, S., & Little, S. (2025). Gender bias in natural language processing and computer vision: A comparative survey. ACM Computing Surveys, 57(6), 1-36. https://dl.acm.org/doi/pdf/10.1145/3700438

298. Robb, M. B., & Mann, S. (2025). Talk, trust, and trade-offs: How and why teens use AI companions. Common Sense Media. https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf

299. U.S. PIRG Education Fund. (2025). AI comes to playtime: Artificial companions, real risks. U.S. PIRG Education Fund. https://pirg.org/edfund/wp-content/uploads/2025/12/AI-Comes-to-Playtime-Artifical-companions-real-risks.pdf

300. Staksrud, E., Mascheroni, G., Milosevic, T., Ní Bhroin, N., Ólafsson, K., Şengül-İnal, G., & Stoilova, M. (2026). European children's use and understanding of generative AI. EU Kids Online V.

301. Committee on the Rights of the Child. (2021). General comment No. 25 (2021) on children's rights in relation to the digital environment (CRC/C/GC/25). United Nations. https://digitallibrary.un.org/record/3906061

302. UNICEF (2025). Guidance on AI and Children: Updated guidance for governments and businesses to create AI policies and systems that uphold children's rights. https://www.unicef.org/innocenti/media/11991/file/UNICEF-Innocenti-Guidance-on-AI-and-Children-3-2025.pdf

303. UNESCO (2025). How should children's rights be integrated into AI governance? https://www.unesco.org/en/articles/how-should-childrens-rights-be-integrated-ai-governance

304. Grossman, S., Pfefferkorn, R., & Liu, S. (2025). AI-Generated Child Sexual Abuse Material: Insights from Educators, Platforms, Law Enforcement, Legislators, and Victims. Version 1. Stanford Digital Repository. Available at https://purl.stanford.edu/mn692xc5736/version/1. https://doi.org/10.25740/mn692xc5736.

305. Livingstone, S., Atabey, A., Stoilova, M., & Sylwander, K. R. (2025). How does, and how could, generative AI respect and enable children's rights? In N. Ni Loideain (Ed.), AI and power: Regulation and rights. University of London Press.

306. European Parliamentary Research Service. (2024). Children and deepfakes. European Parliament. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2025)775855

307. United Nations Children's Fund (UNICEF). (2026). Artificial intelligence and child sexual abuse and exploitation [Issue brief]. https://www.unicef.org/reports/artificial-intelligence-and-child-sexual-abuse-and-exploitation

308. Ozcan, B., Sperati, V., Giocondo, F., Schembri, M., & Baldassarre, G. (2022, June). Interactive soft toys to support social engagement through sensory-motor plays in early intervention of kids with special needs. In Proceedings of the 21st Annual ACM Interaction Design and Children Conference (pp. 625-628).

309. Goodacre, E., & Gibson, J. (2026). AI in the Early Years: Examining the implications of GenAI toys for young children. https://www.cam.ac.uk/stories/ai-toys-study-play

310. British Standards Institution. (2026, May). Half of children have AI toys despite safety concerns. https://www.bsigroup.com/en-GB/insights-and-media/media-centre/press-releases/2026/may/half-of-children-have-ai-toys-as-parents-allow-widespread-use-despite-safety-concerns-and-gaps-in-guidance-parents/

311. Goodacre, E., & Gibson, J. (2026). AI in the Early Years: Examining the implications of GenAI toys for young children. https://doi.org/10.17863/CAM.126270

312. Chou, C. Y., Chan, T. W., Chen, Z. H., Liao, C. Y., Shih, J. L., Wu, Y. T., & Hung, H. C. (2025). Defining AI companions: a research agenda--from artificial companions for learning to general artificial companions for Global Harwell. Research & Practice in Technology Enhanced Learning, 20.

313. Ho, J. Q., Hu, M., Chen, T. X., & Hartanto, A. (2025). Potential and pitfalls of romantic artificial intelligence companions: A systematic review. Computers in Human Behavior Reports, 19, 100715.

314. Hollanek, T., & Sobey, A. (2025). AI companions for health and mental wellbeing: opportunities, risks and policy implications. Leverhulme Centre for the Future of Intelligence

315. De Freitas, J., Oguz-Uguralp, Z., & Kaan-Uguralp, A. (2025). Emotional manipulation by AI companions. arXiv preprint arXiv:2508.19258.

316. Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R., & Yang, D. (2025). The rise of AI companions: how human-chatbot relationships influence well-being. arXiv preprint arXiv:2506.12605.

317. Dewitte, P. (2024). Better alone than in bad company: Addressing the risks of companion chatbots through data protection by design. Computer Law & Security Review, 54, 106019.

318. Zhang, R., Li, H., Meng, H., Zhan, J., Gan, H., & Lee, Y. C. (2025, April). The dark side of ai companionship: A taxonomy of harmful algorithmic behaviors in human-ai relationships. In Proceedings of the 2025 CHI conference on human factors in computing systems (pp. 1-17).

319. De Freitas, J., & Cohen, I. G. (2024). The health risks of generative AI-based wellness apps. Nature medicine, 30(5), 1269-1275.

320. Radesky, J., Bragg, M. A., & Hiniker, A. (2026). Risks and Consequences of Children's Use of Social AI—A Framework. JAMA Pediatrics. https://doi.org/10.1001/jamapediatrics.2026.1349

321. Muldoon, J., & Parke, J. J. (2025). Cruel companionship: How AI companions exploit loneliness and commodify intimacy. new media & society, 14614448251395192.

322. Jacobs, K. A. (2024). Digital loneliness—changes of social recognition through AI companions. Frontiers in Digital Health, 6, 1281037.

323. Ho, J. Q., Hu, M., Chen, T. X., & Hartanto, A. (2025). Potential and pitfalls of romantic Artificial Intelligence (AI) companions: A systematic review. Computers in Human Behavior Reports, 19, 100715.

324. Hollanek, T., & Sobey, A. (2025). AI companions for health and mental wellbeing: opportunities, risks and policy implications.

325. Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R., & Yang, D. (2025). The rise of AI companions: how human-chatbot relationships influence well-being. arXiv preprint arXiv:2506.12605.

326. Dewitte, P. (2024). Better alone than in bad company: Addressing the risks of companion chatbots through data protection by design. Computer Law & Security Review, 54, 106019.

327. Robb, M. B., & Mann, S. (2025). Talk, trust, and trade-offs: How and why teens use AI companions. Common Sense Media. https://www.commonsensemedia.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions

328. Rousmaniere, T., Zhang, Y., Li, X., & Shah, S. (2025). Large language models as mental health resources: Patterns of use in the United States. Practice Innovations. https://doi.org/10.1037/pri0000292

329. Callahan, C., Tanner, L., Coe, C., Davis, M., Glover, J., Bernstein, E., ... & Kunkle, S. (2026). Real-World Use of a Mental Health AI Companion: Multiple Methods Study. JMIR Formative Research, 10, e86904.

330. Associated Press. (2026, June 1). Chatbot AI lawsuit alleges links to teen suicide. https://apnews.com/article/chatbot-ai-lawsuit-suicide-teen-artificial-intelligence-9d48adc572100822fdbc3c90d1456bd0

331. Hudon, A., & Stip, E. (2025). Delusional experiences emerging from AI chatbot interactions or "AI Psychosis". JMIR Mental Health, 12(1), e85799.

332. Green, H. H. (2026, March 14). New study raises concerns about AI chatbots fueling delusional thinking. The Guardian. https://www.theguardian.com/technology/2026/mar/14/ai-chatbots-psychosis

333. Ministère du Travail, de la Santé, des Solidarités et des Familles. (2025, March 24). La santé mentale, grande cause nationale 2025. Gouvernement de la France. https://solidarites.gouv.fr/la-sante-mentale-grande-cause-nationale-2025

334. American Psychiatric Association. (n.d.). Applications of artificial intelligence in mental health care. https://www.psychiatry.org/psychiatrists/practice/artificial-intelligence/applications

335. U.S. Food and Drug Administration. (2025). Executive summary for the Digital Health Advisory Committee meeting: Generative artificial intelligence–enabled digital mental health medical devices. https://www.fda.gov/media/189391/download

336. Hollis, A., & McKeown, G. (2024, September). Empathic AI for autism: Potential and pitfalls of empathic social chatbots in addressing loneliness. In 24th ACM International Conference on Intelligent Virtual Agents: CONNECT, A Workshop on Connecting Interdisciplinary Research on Connections With and Through Technology: IVA 2024.

337. Sharma, D., Meshkat, S., Perivolaris, A., Kamaleddin, M. A., Teferra, B. G., Rueda, A., ... & Bhat, V. (2026). Reimagining psychiatric care with agentic AI: promise, challenges, and a roadmap forward. npj Digital Medicine.

338. Straw, I., & Callison-Burch, C. (2020). Artificial Intelligence in mental health and the biases of language based models. PloS one, 15(12), e0240376.

339. Garg, M. (2024). Towards mental health analysis in social media for low-resourced languages. ACM Transactions on Asian and Low-Resource Language Information Processing, 23(3), 1-22.

340. Wang, W., Tu, Z., Chen, C., Yuan, Y., Huang, J.-T., Jiao, W., & Lyu, M. R. (2024). All languages matter: On the multilingual safety of LLMs. Findings of the Association for Computational Linguistics: ACL 2024, 5865–5877. https://doi.org/10.18653/v1/2024.findings-acl.349

341. Nigatu, H. H., Mehandru, N., Abadi, N. H., Gebremeskel, B., Alaa, A., & Choudhury, M. (2025). Viability of machine translation for healthcare in low-resourced languages. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP), 10584–10598. https://aclanthology.org/2025.emnlp-main.535/

342. Fu, Y. V., Ramachandran, G. K., Park, N., Lybarger, K., Xia, F., Uzuner, Ö., & Yetisgen, M. (2025). BioMistral-NLU: Towards more generalizable medical language understanding through instruction tuning. AMIA Joint Summits on Translational Science Proceedings, 2025, 149–158. https://pubmed.ncbi.nlm.nih.gov/40502228/

343. Labrak, Y., Bazoge, A., Morin, E., Gourraud, P.-A., Rouvier, M., & Dufour, R. (2024). BioMistral: A collection of open-source pretrained large language models for medical domains. Findings of the Association for Computational Linguistics: ACL 2024, 5848–5864. https://aclanthology.org/2024.findings-acl.348/

344. Qiu, P., Wu, C., Zhang, X., Lin, W., Wang, H., Zhang, Y., Wang, Y., & Xie, W. (2024). Towards building multilingual language models for medicine. Nature Communications, 15, 8384. https://doi.org/10.1038/s41467-024-52417-z

345. Nwabufo, J., Ogueji, K., Adelani, D. I., Alabi, J., et al. (2025). Healthcare NLP for African Languages: Current State and Challenges. Proceedings of the AfricaNLP Workshop (AfricaNLP 2025). https://aclanthology.org/2025.africanlp-1.32/

346. Okafor, U. (2025). Multilingual NLP for African Healthcare: Bias, Translation, and Explainability Challenges. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025), 221–229. https://aclanthology.org/2025.africanlp-1.32/

347. Skianis, K., Doğruöz, A. S., & Pavlopoulos, J. (2024). Leveraging LLMs for translating and classifying mental health data. In J. Sälevä & A. Owodunni (Eds.), Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024) (pp. 236–241). https://doi.org/10.18653/v1/2024.mrl-1.20

348. Cronin, A., Kelly, A., Wrona, M., O'Donnell, P., Hassan, A., Myles, T., Fallon, T., & MacFarlane, A. (2025). The patient-safety implications of AI-based communication with migrants in general practice: a scoping review. BJGP Open, 9(4), BJGPO.2025.0107. https://bjgpopen.org/content/9/4/BJGPO.2025.0107

349. House of Lords Public Services Committee. (2025). Lost in translation? Interpreting services in the courts (2nd Report of Session 2024–26, HL Paper 87). https://committees.parliament.uk/publications/44602/documents/221328/default/

350. Nigatu, H. H., Mehandru, N., Abadi, N. H., Gebremeskel, B., Alaa, A., & Choudhury, M. (2025). Viability of machine translation for healthcare in low-resourced languages. Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, 10584–10598. https://aclanthology.org/2025.emnlp-main.535/

351. MIT AI Risk Initiative. (2025). AI risk mitigation database and draft taxonomy. https://airisk.mit.edu/ai-riskmitigations

352. Gabriel, I., Manzini, A., Keeling, G., Hendricks, L. A., Rieser, V., Iqbal, H., ... & Manyika, J. (2024). The ethics of advanced AI assistants. arXiv preprint arXiv:2404.16244.

353. Staufer, L., Feng, K., Wei, K., et al. (2026). The 2025 AI agent index: Documenting technical and safety features of deployed agentic AI systems. https://doi.org/10.48550/arXiv.2602.17753

354. Weidinger, L., Raji, I. D., Wallach, H., Mitchell, M., Wang, A., Salaudeen, O., ... & Isaac, W. (2025). Toward an evaluation science for generative ai systems. arXiv preprint arXiv:2503.05336.

355. Rabanser, S., Kapoor, S., Kirgis, P., et al. (2026). Towards a science of AI agent reliability. https://doi.org/10.48550/arXiv.2602.16666

356. Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). Generative AI can harm learning. The Wharton School Research Paper.

357. Shen, J. H., & Tamkin, A. (2026). How AI impacts skill formation. arXiv preprint arXiv:2601.20245.

358. Budzyń, K., Romańczyk, M., Kitala, D., Kołodziej, P., Bugajski, M., Adami, H. O., ... & Mori, Y. (2025). Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. The Lancet Gastroenterology & Hepatology, 10(10), 896-903.

359. Epoch AI. (2026). Data on AI models. https://epoch.ai/data/ai-models

360. Feng, K. J., McDonald, D. W., & Zhang, A. X. (2025). Levels of autonomy for ai agents. arXiv preprint arXiv:2506.12469.

361. Fink, M. (2025). Operationalizing meaningful human oversight under Article 14 of the EU AI Act. In AI Act commentary: A thematic analysis (forthcoming). Hart-Bloomsbury.

362. Shapira, N., Wendler, C., Yen, A., Sarti, G., Pal, K., Floody, O., ... & Bau, D. (2026). Agents of chaos. arXiv preprint arXiv:2602.20021.

363. Hammond, L., Chan, A., Clifton, J., Hoelscher-Obermaier, J., Khan, A., McLean, E., ... & Rahwan, I. (2025). Multi-agent risks from advanced ai. arXiv preprint arXiv:2502.14143.

364. Soares, N., Fallenstein, B., Armstrong, S., & Yudkowsky, E. (2015). Corrigibility. Machine Intelligence Research Institute. https://intelligence.org/files/Corrigibility.pdf.

365. Zeng, Y., Lu, E., Guo, X., Huangfu, C., Xie, J., Chen, Y., ... & Younas, A. (2025). AI Governance InternationaL Evaluation Index (AGILE Index) 2025. arXiv preprint arXiv:2507.11546.

366. Bommasani, R., Kapoor, S., Klyman, K., Longpre, S., Ramaswami, A., Zhang, D., Schaake, M., Ho, D. E., Narayanan, A., & Liang, P. (2023, December 13). Considerations for governing open foundation models. Stanford Institute for Human-Centered Artificial Intelligence. https://hai.stanford.edu/policy/issue-brief-considerations-governing-open-foundation-mode

367. Tzachor, A., Devare, M., Richards, C., Pypers, P., Ghosh, A., Koo, J., ... & King, B. (2023). Large language models and agricultural extension services. Nature food, 4(11), 941-948.

368. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616, 259–265. https://doi.org/10.1038/s41586-023-05881-4

369. Bommasani, R., Klyman, K., Kapoor, S., Longpre, S., Ramaswami, A., Zhang, D., Schaake, M., Ho, D. E., Narayanan, A., & Liang, P. (2024). The foundation model transparency index v1.1. arXiv:2407.12929. https://arxiv.org/abs/2407.12929

370. BigScience Workshop, Le Scao, T., Fan, A., et al. (2022). BLOOM: A 176B-parameter open-access multilingual language model. arXiv. https://doi.org/10.48550/arXiv.2211.05100

371. Touvron, H., Lavril, T., Izacard, G., et al. (2023). LLaMA: Open and efficient foundation language models. arXiv. https://arxiv.org/abs/2302.13971

372. Qwen Team. (2024). QwenLM. GitHub. https://github.com/QwenLM/Qwen

373. Qwen Team. (2024). Qwen2 technical report. arXiv. https://arxiv.org/abs/2407.10671

374. DeepSeek-AI. (2024). DeepSeek-V3 technical report. arXiv. https://doi.org/10.48550/arXiv.2412.19437

375. DeepSeek-AI. (2025). DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. arXiv. https://arxiv.org/abs/2501.12948

376. Jiang, A. Q., et al. (2023). Mistral 7B. arXiv. https://arxiv.org/abs/2310.06825. Mistral AI.

377. Technology Innovation Institute. (2023). The Falcon series of open language models. arXiv. https://arxiv.org/abs/2311.16867

378. SberDevices. (2026, March). GigaChat-3.1: Большое обновление больших моделей. Habr. https://habr.com/ru/companies/sberbank/articles/1014146/

379. Yandex. (2025, February 25). YandexGPT 5 — в Алисе, облаке и опенсорсе. Habr. https://habr.com/ru/companies/yandex/articles/885218/

380. Sarvam AI. (2025). Sarvam AI models on Hugging Face. https://huggingface.co/sarvamai

381. SB Intuitions. (2025). Sarashina models on Hugging Face. https://huggingface.co/sbintuitions

382. Naver Cloud. (2024). HyperCLOVA X technical report. arXiv. https://arxiv.org/abs/2404.01954

383. Seger, E., Dreksler, N., Moulange, R., et al. (2023). Open-sourcing highly capable foundation models: An evaluation of risks, benefits, and alternative methods for pursuing open-source objectives. Centre for the Governance of AI. https://www.governance.ai/research-paper/open-sourcing-highly-capable-foundation-models

384. Anthropic. (2025). Disrupting the first reported AI-orchestrated cyber espionage campaign. https://www.anthropic.com/news/disrupting-AI-espionage

385. Partnership on AI. (2023). PAI's guidance for safe foundation model deployment. https://partnershiponai.org/modeldeployment/

386. International Energy Agency. (2025). Energy and AI. International Energy Agency. https://www.iea.org/reports/energy-and-ai

About the Independent International Scientific Panel on Artificial Intelligence

Composition and mandate

The Independent International Scientific Panel on Artificial Intelligence was established within the United Nations through General Assembly resolution 79/325, as committed to through the Global Digital Compact and the Pact for the Future.

Composed of 40 independent experts appointed by the General Assembly for a term of three years on the basis of their outstanding expertise in AI and related fields, the Panel is gender balanced and includes members from all five regional groups of Member States across disciplines including core technical AI, applied AI, safety and infrastructure, and AI policy, ethics and impact.

The Panel is mandated to issue evidence-based scientific assessments synthesizing and analysing existing research related to the opportunities, risks and impacts of AI through an annual policy-relevant but non-prescriptive summary report including thematic briefs as it deems necessary. Its scientific work is to be guided by principles of independence, scientific credibility and rigor, multidisciplinary and inclusive participation.

The Panel is also mandated to present its annual summary report at the United Nations Global Dialogue on Artificial Intelligence Governance. By informing the Global Dialogue and broader international processes, the Panel enables the global community to anticipate emerging challenges, make better-informed governance decisions, and level the information playing field for policymakers worldwide.

The Panel is supported by the Panel Secretariat coordinated by the United Nations Office for Digital and Emerging Technologies.

Process toward the present report

In the three months since the Panel's first meeting in March 2026, following their appointment in February 2026, the Panel worked intensively to deliver the present report. This intensive process of scientific exchange and collective analysis included a three-day in-person plenary meeting and over 60 virtual meetings of the Panel, facilitated by its elected Co-Chairs.

This preliminary report baseline establishes a foundation for wider consultation with external experts and subsequent thematic briefs and annual summary reports.

Scientific independence of the Panel

Panel members serve in their personal capacity as scientifically independent experts. Through their designation as United Nations experts on mission, each has declared and promised not to seek or accept instructions in regard to the performance of their duties from any Government or other source. The Regulations Governing the Status, Basic Rights and Duties of Experts on Mission (ST/SGB/2002/9) also include regulations concerning their conduct and accountability.

Donors

The Panel Secretariat gratefully acknowledges the financial and in-kind contributions of the following governments and partners, without whom the Panel would not have been able to carry out its responsibilities:

Government of Germany
Government of Japan
Government of Spain
Omidyar Network Fund

Panel Secretariat

Coordinator

- Amandeep Singh Gill, United Nations Under-Secretary-General for Digital and Emerging Technologies

Secretariat Coordination & Support

- Quintin Chou-Lambert, United Nations Office for Digital and Emerging Technologies (UN ODET)
- Rebakah Hayoung Woo, UN ODET
- Peppi Väänänen, UN ODET

Fundraising & Logistics

- Sebastian Frank, UN ODET
- Antonieta Loaiza, UN ODET

Communications

- Karoline Hassfurter, UN ODET
- Brian Shung Seun Lau, UN ODET
- Anamika Madhuraj, UN ODET

Editing and Drafting Support*

- Jiaee Cheong, United Nations University (UNU)
- Kevin Kohler, UNU
- Max Springer, UNU

Rapporteurs

- Wernhard Berger, United Nations Industrial Development Organization
- Jin Cui, International Telecommunication Union (ITU)
- Tim Engelhardt, Office of the United Nations High Commissioner for Human Rights (UN OHCHR)
- Andrew Morritt, United Nations Department of Peace Operations
- Prateek Sibal, United Nations Educational, Scientific and Cultural Organization (UNESCO)
- Mariagrazia Squicciarini, UNESCO
- Oleksandra Vereschak, UNESCO
- Ana Gabriela Fernandez Vergara, ITU
- Li Zhou, UN OHCHR

* To ensure scientific independence of the Panel's work, Editing and Drafting Support personnel report to the Panel on substantive content/language of written outputs, while complying with coordination requirements, timelines, and sharing of produced content as part of the Panel Secretariat. For administrative purposes, they report to United Nations University.

The following United Nations Secretariat entities also supported the Panel Secretariat in the preparation of the report: Department for General Assembly and Conference Management, Department of Global Communications, and United Nations Geospatial (Office of Information and Communications Technology).

United Nations

Independent International Scientific Panel on Artificial Intelligence

9 789211 576627