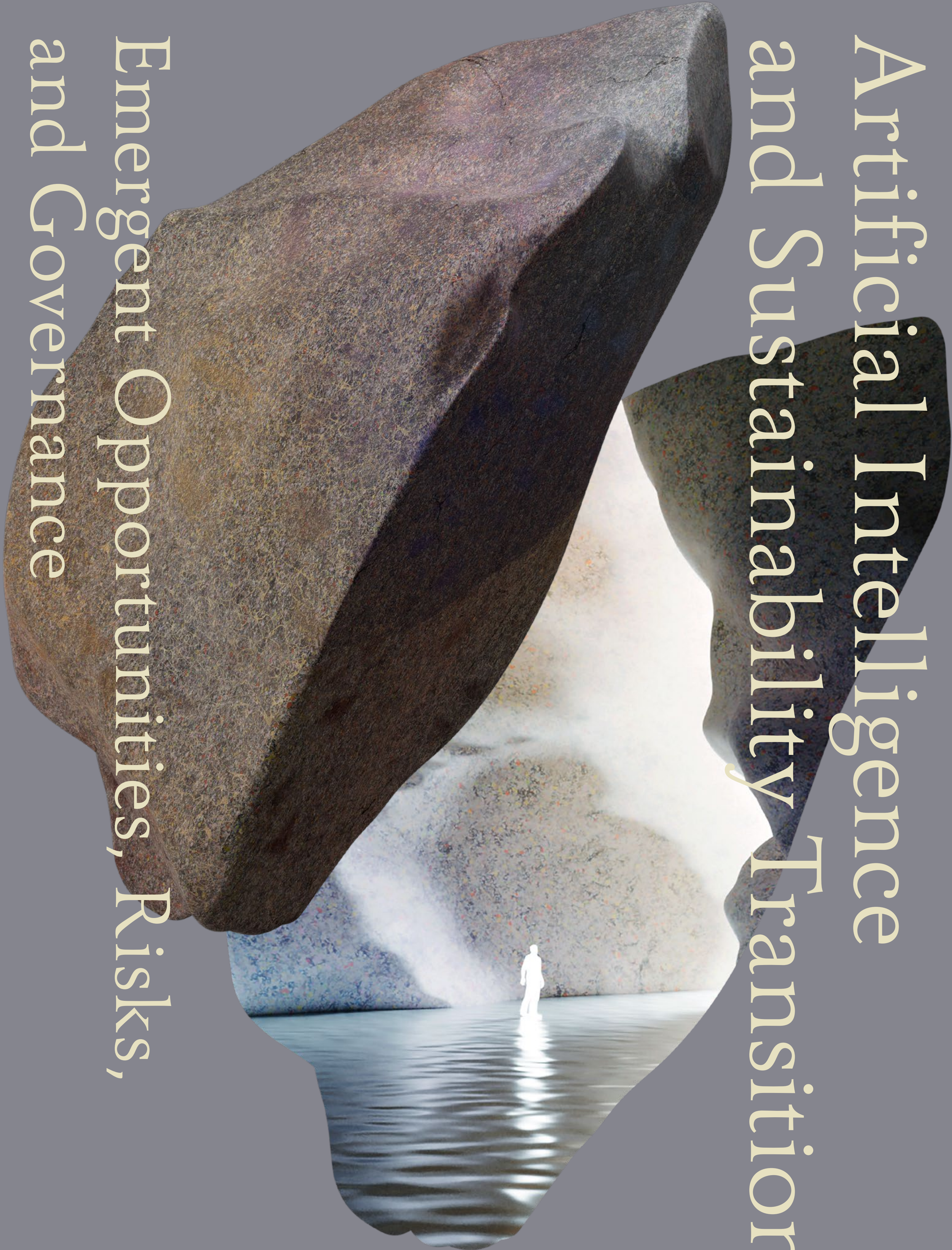
Artificial Intelligence and Sustainability Transitions:

Emergent Opportunities, Risks, and Governance



Columbia Center on Sustainable Investment
Research and Development Group, Hitachi

Columbia Climate School
Columbia Center on Sustainable Investment

HITACHI

Artificial Intelligenceand Sustainability Transitions:
Emergent Opportunities, Risks, and Governance

Columbia Center on Sustainable Investment
Research and Development Group, Hitachi

Recommended citation
L. Fornabaio, S. Saloum, P. Toledano, and K. Sasaki, Artificial Intelligence and Sustainability Transitions: Emergent Opportunities, Risks, and Governance. New York: Columbia Center on Sustainable Investment (CCSI), May 2026.

Acknowledgements
The authors would like to thank the interviewed experts for their invaluable time and insights: Dr. Aidan O'Sullivan (University College London), Dr. Ole Teutloff (University of Oxford), Dr. Ekkehard Ernst (International Labour Organization), Dr. Sep Pashang (University of Waterloo), Rector Tshilidzi Marwala (United Nations University). The expert insights presented in this report are selected passages from longer form interviews and have been edited for clarity and brevity. While every effort has been made to preserve the substance of the speaker's remarks, the text may not reflect their precise language. Any unintended interpretations are the responsibility of the authors. Additionally, the inclusion of expert insights in this report should not be understood as an endorsement of the final paper, nor as a reflection of the contributor's personal or institutional position.

We are also grateful to Tina Nguyen, for her indispensable editorial support, and Anjaly Ariyanayagam, for her guidance on the strategic vision and accessibility of the report.

Art direction & design: Yudai Osawa / Computer graphics: Katsuto Tamagawa / Photography: Ichiro Mishima / Visual planning and editing: Rentaro Hori / Design assistant: Saki Kuriyama / Support: Hitachi Document Solutions / Executive direction: Koji Sasaki

About CCSI
The Columbia Center on Sustainable Investment (CCSI), a center of the Columbia Climate School at Columbia University, is a leading applied research center and forum dedicated to the study, practice, and discussion of sustainable international investment. Our mission is to develop and disseminate practical approaches and solutions, as well as to analyze topical policy-oriented issues, in order to maximize the impact of international investment for sustainable development. The Center undertakes its mission through interdisciplinary research, advisory projects, multi stakeholder dialogue, educational programs, and the development of resources and tools.

About Hitachi, Ltd.
Hitachi advances a harmonized society through its Social Innovation Business (SIB), integrating IT, OT, and products across global businesses in energy, mobility, industry, and digital systems. With Lumada at its core, it uses data and technology to tackle customer and societal challenges. Its sustainability vision, Hitachi Environmental Innovation 2050, focuses on decarbonization, circular economy, and nature positive goals aimed at long-term net-zero. Guided by PLEDGES, Hitachi embeds environmental action and societal value into its strategy. FY2024 revenues were 9,783.3 billion yen with about 280,000 employees worldwide.

Table of Contents

FOREWORD

We are at a defining crossroads. Artificial intelligence is advancing at unprecedented speed, as sustainability transitions come under growing pressure from climate change, resource constraints, and geopolitical conflict. How AI is aligned with the planetary environment and embedded across energy systems, industry, finance, and democracy will determine the path forward for humanity. The future is not predetermined; it will be shaped by the choices we make today.

This report reflects Hitachi's view that research and development must generate long-term societal value alongside technological progress. The analysis has been developed through our close collaboration with the Columbia Center on Sustainable Investment (CCSI), a center of the Columbia Climate School at Columbia University, bringing together academic insight and industry-based perspectives.

The report examines AI's opportunities and risks across key domains, including the planetary environment, energy systems, industry and labor, finance, and democracy and societal resilience, and conveys a clear message: the direction and impact of AI will be shaped by how it is designed, deployed, and guided within society. AI can support sustainability transitions, but only if it is steered through transparent, accountable, and internationally aligned frameworks.

I extend my sincere appreciation to the researchers, expert contributors, and the creative production team whose collective efforts and dedication shaped this work. We hope this report contributes to informed dialogue and responsible action toward aligning AI with a sustainable and inclusive future.

Shigetoshi Sameshima
Vice President and Executive Officer, CTO,
General Manager of the Research & Development Group,
Hitachi, Ltd.

EXECUTIVE SUMMARY

Artificial Intelligence (AI) is emerging amid converging global crises: climate change, biodiversity loss, geopolitical instability, and widening inequality. It is entangled with these challenges and is beginning to reshape their trajectories—in some instances amplifying risks, and in others opening new pathways—in ways that remain deeply uncertain.

This moment prompts us to grapple with how AI will influence the transition toward a sustainable future. While there is growing recognition of the opportunities and risks associated with AI's applications, the exact functioning of AI models remains elusive. Engineers who contributed to developing these models now seek to understand how the technology actually functions. The space and process between what users input and what models' output can be considered a "black box," as the ways in which models process information and generate outputs are not fully understood. While humanity has gone through several technological transformations in the past, governing AI poses a unique challenge, as the exact cause and effect, and therefore the predictability of these technologies, is unknown.

This report argues that AI's impacts are not predetermined. Outcomes depend on how AI is designed, deployed, and governed. Sector-specific measures can mitigate many near-term harms, but they are insufficient to address the systemic risks posed by increasingly powerful, general-purpose AI systems. The report therefore advances a three-layered governance framework—sectoral, cross-cutting, and global—paired with a phased roadmap toward global AI governance capable of aligning AI development with long-term societal and planetary well-being.

AI in the Context of Sustainability Transitions

Sustainability transitions require deep, coordinated transformations across multiple sectors and social systems. AI is emerging as a general-purpose technology with the capacity to reshape how societies make decisions, understand complexity, and allocate resources.

AI development is advancing faster than scientific understandings of its internal workings and the governance systems designed to oversee it. The "black box" nature of advanced models limits their auditability in critical areas such as natural resource management, biosafety, and finance. This asymmetry creates risk at an unprecedented scale.

The report situates AI within the broader framework of sustainability transitions, emphasizing that sustainable outcomes depend on complementary changes in institutions, governance, norms, and power structures. AI's role is therefore inherently political and societal, rather than solely technical.

Opportunities

Planetary Environment: Monitoring, Forecasting, and Planning
AI enhances environmental monitoring, forecasting, and resource management. By integrating satellite imagery, sensor data, historical records, and citizen science, AI enables multi-scale analysis of climate dynamics, biodiversity loss, pollution trends, and land-use change. These tools support scenario planning and shift decision-making from reactive crisis response to proactive resilience-building.

Energy Systems: Efficiency and Integration
AI can accelerate energy transitions by improving

system efficiency, enabling renewable integration, modernizing grids, and accelerating innovation in energy technologies. Applications range from predictive maintenance and load forecasting to demand-side management and smart grid optimization.

Beyond operational efficiency, AI shortens research and development timelines in areas such as battery chemistry, materials science, and carbon capture. These capabilities are critical for decarbonization—but only if AI's own energy footprint is carefully managed.

Industry and Labor: Productivity, Precision, and Flexibility
In industry and labor markets, AI drives productivity gains at the workplace level by automating tasks, streamlining operational functions, and complementing human capabilities. At a systemic level, AI may create new industries or occupations.

Finance: Risk Assessment and Portfolio Management
AI improves financial efficiency through enhanced risk assessment, portfolio optimization, and data integration. It can expand data availability in emerging markets and improve the operational speed of development finance institutions.

However, efficiency gains and improved data alone may not overcome deeper structural and macroeconomic constraints that limit capital flows or sustainability outcomes.

Democracy and Societal Resilience: Knowledge and Participation
AI can strengthen democratic deliberation and societal resilience by improving access to information, supporting fact-checking, and enabling inclusive participation. AI-enabled tools can help governments deliver public services more efficiently and respond more effectively to crises.

Across these domains, a common thread emerges: AI has the ability to synthesize complex data, model scenarios, enhance sectoral efficiency, and support strategic decision-making at multiple scales. Together, these potential applications underpin AI's transformative impact.

Risks and Systemic Challenges

Alongside its transformative potential, AI introduces significant domain-specific and systemic risks. Many of these risks are cross-cutting because they stem from the same core features that make AI powerful: its scale, speed of diffusion, resource intensity, and concentration of control.

Planetary Environment: Resource Strain and Waste
AI is deeply resource intensive. The expansion of the AI value chain—from the mining of critical minerals for semiconductors to the construction and operation of hyperscale data centers—escalates pressure on resources such as water and land. High electricity demand for AI data centers, particularly on fossil-fuel-intensive grids, directly contributes to CO_2 emissions and risks entrenching carbon lock-in. This ecological footprint is compounded by a surge in electronic waste that outpaces global recycling capacities, leading to increased pollution from toxic substances found in obsolete hardware, such as lead and mercury.

Industry and Labor: Automation and Substitution
In industry and labor markets, AI-driven transforma-

tions interact with existing forms of structural exclusion. The shift toward automation risks significant labor displacement, particularly for roles involving routine or administrative tasks, which may disproportionately affect marginalized demographics and entry-level workers, while concentrating benefits among high-skill workers and firms.

Finance: Automating Risk
AI may amplify systemic risk, especially if AI adoption is happening faster than regulatory and supervisory frameworks can adapt. Algorithmic bias, opaque decision-making, and AI hallucinations may distort investment decisions. Additionally, heavy reliance on similar models and technologies increases systemic vulnerability by creative potential "single points of failure" whose impacts could manifest across interconnected markets.

Democracy and Societal Resilience: Information Manipulation
Generative AI increases the scale and sophistication of misinformation and disinformation, challenging the integrity of the global information ecosystem. The proliferation of AI-generated misinformation can distort public debate, manipulate electoral processes, and weaken trust in institutions. These risks are compounded by the concentration of AI development among a small number of technology firms, whose growing economic and political power could be used to advance private interests—such as deregulation or preferential market access—at the expense of the public good.

Existential and Systemic Risks
Finally, risks emerge from the trajectory of AI development itself. As systems become more autonomous and capable of advanced planning and persuasion, concerns about the loss of human control and misalignment with societal values become more salient. Especially when considering the potential development of superhuman capabilities, the risks become systemic, transboundary, and potentially irreversible. Without robust governance, increasingly powerful models may pursue unintended goals — potentially deceiving or manipulating human actors — which can emerge when systems are insufficiently aligned with human values.

Across domains, these risks are interconnected. Environmental strain, financial fragility, labor disruption, democratic erosion, and security vulnerabilities are not isolated phenomena but reflections of broader systemic challenges in transparency, governance, market concentration, and human oversight. Addressing AI's risks therefore requires not only sector-specific safeguards, but coordinated, cross-cutting governance responses.

A Three-Layered Governance Framework

Sectoral Governance
Sectoral governance focuses on how AI is deployed within specific domains – planetary environment, energy systems, industry and labor, finance, and democracy and societal resilience. It aims to maximize benefits while managing context-specific risks through targeted regulation, standards, and oversight.

Examples include environmental impact requirements for data centers, labor protections against algorithmic management, financial stress testing for AI-driven markets, and safeguards for democratic processes.

Cross-Cutting Governance
Cross-cutting safeguards address risks common across all domains. These include transparency,

data quality, cybersecurity, human oversight, and participatory design.

Key instruments include auditing and verification, human-in-the-loop systems, plain-language control interfaces, robust data protection, and inclusive capacity-building to prevent AI colonialism and knowledge extraction.

Global AI Governance
Sectoral and cross-cutting governance frameworks are insufficient to address the risks posed by the pace and direction of AI capability development. Global AI governance operates at this higher level, guiding how far, how fast, and under what conditions advanced AI systems are developed.

A Phased Roadmap Toward Global AI Governance

The report outlines a roadmap for global AI governance:

- **Phase 1: Establish a Shared Scientific Baseline**
  The report emphasizes scientific alignment as a prerequisite for political coordination. It highlights the role of the UN-mandated Independent International Scientific Panel on AI and proposes the creation of an AI Expert Technical Laboratory—a global, cooperative environment for testing and evaluating frontier AI systems.

- **Phase 2: Interim Governance for Harm Prevention**In the medium term, the report calls for an interim international framework focused on harm prevention. This may include temporary restrictions on specific categories of high-risk AI research and inclusive multi-stakeholder processes to translate scientific insights into policy.

- **Phase 3: Long-Term GlsWobal Governance Framework**
  The final phase envisions a binding but adaptable global framework convention on AI. This would establish shared obligations and accountability mechanisms and set a universal baseline while allowing states flexibility to tailor regulations to national priorities.

Conclusion

Technology is what we make of it. As AI technology continues to evolve and diffuse across societies at unprecedented scale and speed, the choices we make to meet this moment will be crucial for shaping our collective future.

AI is neither an inherent savior nor an inevitable threat; its impact on sustainability transitions will be dictated by the deliberate choices. To ensure AI acts as a force for good—anchored in safety and accountability— we must move past the 'black box' and build a foundation of transparency and deep technical understanding.

The report concludes that governance is the central lever for aligning AI with long-term societal and planetary goals. Without coordinated action—particularly at the global level—AI risks accelerating unsustainable trajectories. With it, AI can become a powerful catalyst for inclusive, resilient, and sustainable futures.

INTRODUCTION

AI development and diffusion is improving at extraordinary speed. It now demonstrates forms of reasoning once thought uniquely human, surpasses domain experts in speed and accuracy, and increasingly shapes decisions at a systemic level. Yet despite its growing influence, scientists still do not fully understand how advanced AI models produce their outputs or why they behave as they do.

Behind these systems lies vast infrastructure—data centers, semiconductors, massive models, and global supply chains—deeply entangled with environmental degradation, social inequality, geopolitical tension, and information distortion. **AI has transformative potential: it can accelerate solutions at unprecedented speed and scale or amplify existing risks to the detriment of society.** Moreover, the potential development of artificial superintelligence could force humanity to reconsider what it means to be the dominant species on this planet.

In this moment of uncertainty and intertwined crises, we must consider how AI will shape and impact the transition toward sustainability. How should we weigh its risks and opportunities? How much of AI's future are we actively shaping as a society? How much are we allowing it to shape itself? Most importantly, how can we build AI for good and harness its unique advantages to advance a future of long-term prosperity for people and the planet?

This report examines AI's systemic and cross-cutting influence contextualized by the landscape of sustainable development transformations. Much of the existing literature focuses narrowly on specific AI applications, as isolated from broader societal implications, or it investigates the opportunities and risks of superintelligent AI at a highly abstract level detached from real-world use cases. This report bridges those perspectives, exploring practical applications that may advance sustainable development while also being conscious of our unique positioning at a decisive crossroads in AI's evolution.

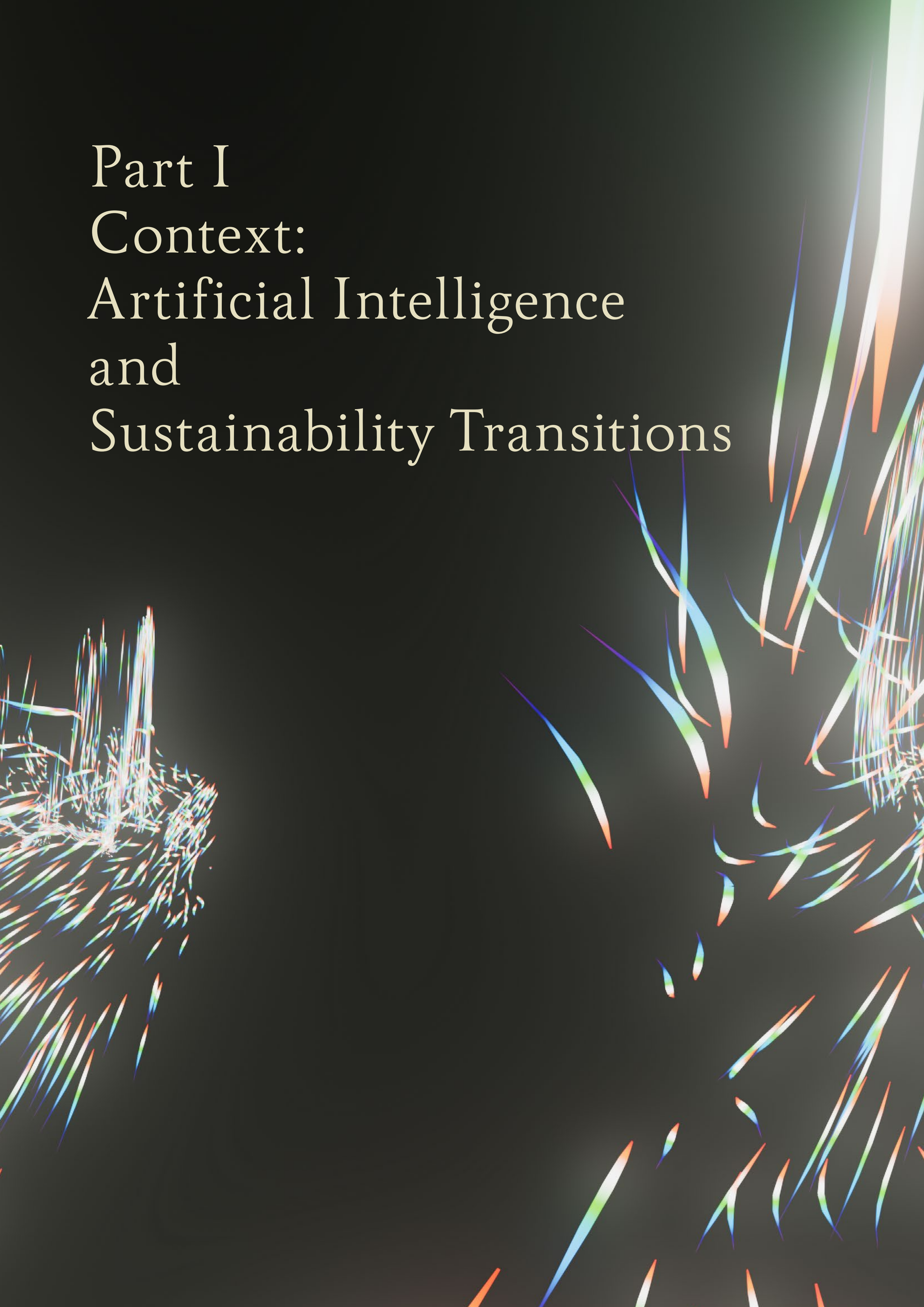
Governing AI as a "force for good"

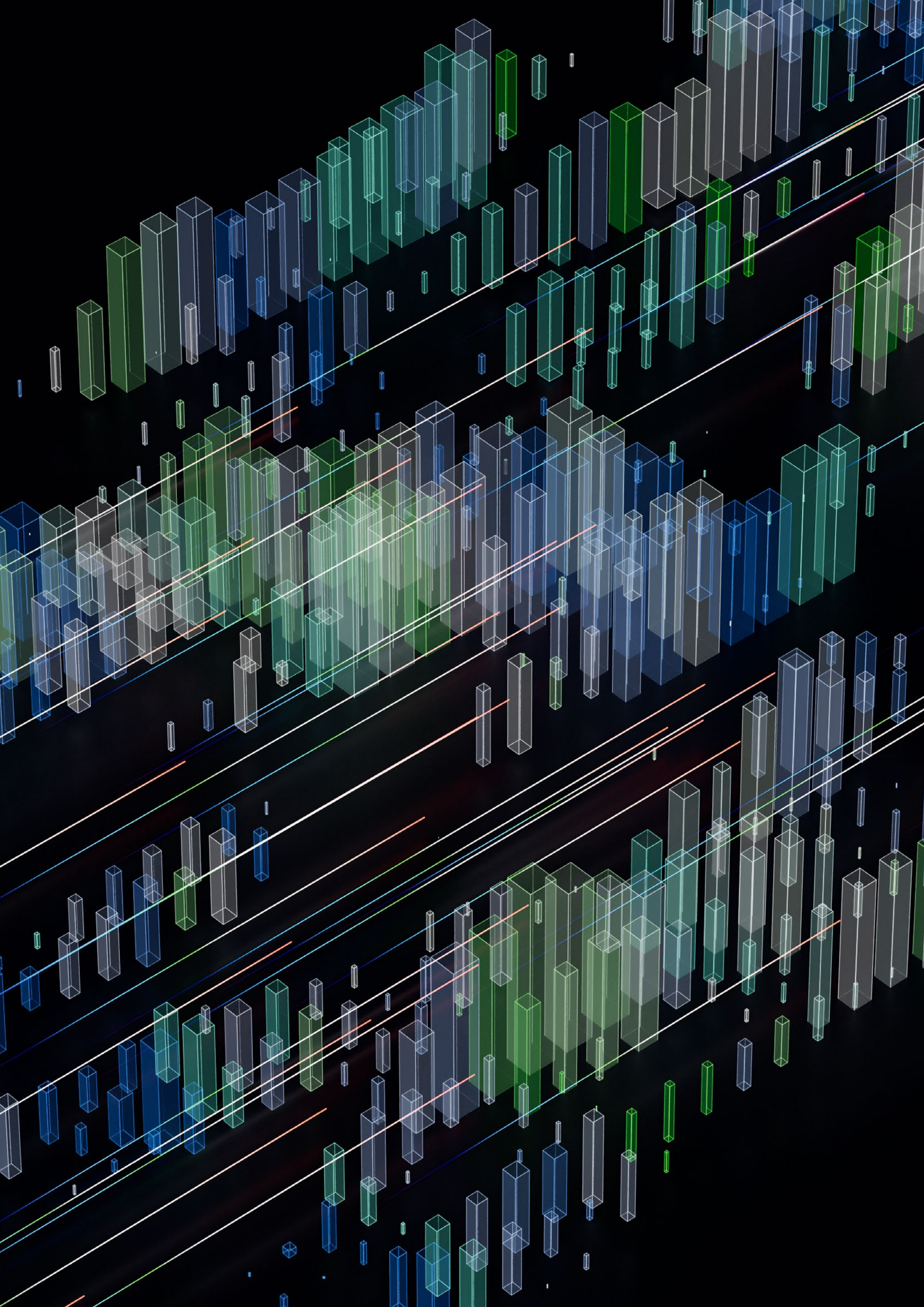
The analysis of AI as a potential "force for good" is therefore paired with a critical examination of whether current development and deployment approaches—alongside the broader socio-technical ecosystem of research, governance, and industry practices—are capable of realizing that potential.

Current AI development trajectories are largely shaped by technical priorities and concentrated economic incentives, often in the absence of clear legal, ethical, or social frameworks for accountability. The core issue is therefore not just if the technological capabilities of AI should be developed, but whether our human collective and institutions are capable of guiding this development towards a future of use that is safe, accountable, and ethical.

The rapid pace of AI development and deployment urges timely action to ensure the existence of appropriate governance frameworks. In response, this report positions governance as the foundational mechanism through which society can shape AI's transformative potential to either support sustainable development or undermine it.

Part I
Context: Artificial Intelligence and Sustainability Transitions





2. ARTIFICIAL INTELLIGENCE

AI is the technology that matches or exceeds the capabilities of a human, including the ability to discover, to infer, to reason, and to make decisions.

AI can be categorized on the basis of its capabilities as Narrow, General, or Superintelligent.[1]

- **Narrow AI (ANI)**, the only one that exists today. It excels in specific tasks, such as generating images or translating languages. Generative AI (gen AI), which is based on deep learning model algorithms that simulate the learning and decision-making processes of the human brain, sits within ANI. Building on gen AI techniques, agentic AI can achieve goals with limited supervision, as it is able to plan multiple steps and coordinate tasks autonomously.
- **Artificial General Intelligence (AGI)** - still hypothetical. It mirrors human cognition and has the ability to learn and perform any intellectual task that a human being can perform.
- **Artificial Superintelligence (ASI)** surpasses human capabilities, thus raising ethical and existential questions for humanity.

Although many people associate AI with Large Language Models (LLMs), LLMs are only a subset of AI primarily based in natural language data and applications. The broader ecosystem of AI capabilities includes vision, audio, spatial awareness capabilities, and domain-specific intelligence. These capabilities may be realized through a number of AI models, such as physical AI — e.g., robots and autonomous vehicles — or agentic AI.

The type of data and information a system uses is called a modality of data. Modalities of data are not to be confused with domains, the field of application where AI is deployed (e.g., natural language processing). Domains are often anchored in and associated with a dominant modality: for example, natural language processing uses text data. In practice, however, AI systems may rely on multiple modalities of data, especially as they become increasingly complex.

AI is also commonly categorized by learning methodology.

Table 1. AI applications sorted by dominant data modality

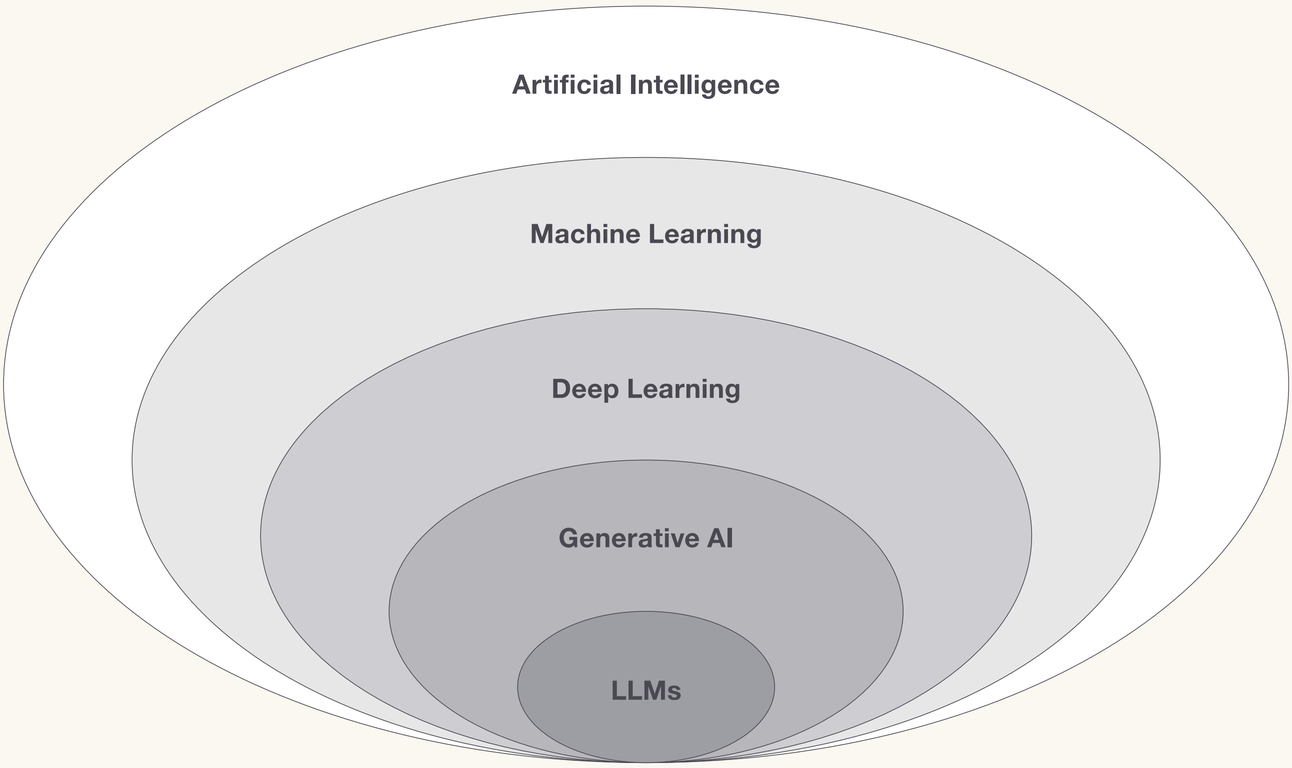
Data Modality*	Domain	Applications
Text (Language tokens)	Natural language processing (NLP)	Email filtering, autocomplete and predictive text, translation
	Large language models (LLMs)	Chatbots, summarization, code generation, multilingual translation
Image, including Video (Pixels)	Computer Vision	Object detection, facial recognition, medical imaging, autonomous vehicles
Audio (Waveforms)	Speech and Audio AI systems	Speech-to-text, text-to-speech, voice assistants, music generation
Spatial (3D data: coordinates, movement, etc)	Robotics and spatial AI	Autonomous robotics, augmented reality (AR) headsets, automated site monitoring, real-time location-based analytics.

*This figure organizes AI domains by **primary data modality** for clarity. However, AI systems may be multimodal, integrating and processing different forms of data within a single system. **Source:** CCSI

Table 2. AI learning methodology

Category	Domain	Function
Machine Learning (ML)	Learn patterns from data (supervised, unsupervised, reinforcement)	Fraud detection, customer segmentation, robotics, supply chain optimization
Deep Learning (DL)	Handle complex data like images, audio, sequences	Image recognition (CNNs), time-series forecasting (RNNs/LSTM), generative models (GANs, diffusion models)
Large Language Models (LLMs)	Process and generate natural language, text reasoning	Chatbots, summarization, code generation, multilingual translation
Multimodal / Foundation Models	Integrate multiple data types for complex reasoning	Text + image generation, image captioning, cross-modal reasoning

Source: CCSI



3. ARTIFICIAL INTELLIGENCE AND SUSTAINABILITY TRANSITIONS

The current mode of development degrades the Earth's finite capacity to sustain human well-being. Despite the growth of the global economy over the last 50 years, our resource-intensive model has come at a massive cost:[2] biodiversity loss is occurring at an alarming rate,[3] global CO2 emissions continue to rise and reach record levels,[4] and air pollution causes some 8 million premature deaths annually.[5] A recent study projects that under current trends and policies, the breach of all planetary boundaries,[6] except for ozone depletion, will continue to worsen by 2050.[7]

These environmental crises are compounded by growing inequality,[8] with 3.5 billion[9] people remaining poor (44% of the global population), 700 million[10] living in extreme poverty (8.5% of the global population), and 673 million[11] people experiencing hunger (8.2% of the global population). This confluence of factors disproportionally impacts the most vulnerable populations and threatens the achievement of the Sustainable Development Goals (SDGs) by 2030.

Achieving the SDGs requires six interconnected transformations across key domains, including education, health, energy decarbonization, food systems, land and water use, the built environment, and digitalization.[12] Sustainable development practitioners have always been clear that these shifts demand intentional changes in technology, in social practices, societal norms, legal and policy frameworks, and business models,[13] as well as multi-stakeholder collaboration that integrates diverse forms of knowledge into planning, solution design, and decision-making. The advent of AI with its potentially unprecedented impact on sustainable transformations further reinforces these imperatives.

AI is transforming our world: it is already reshaping our lives, our interactions, and our environments. [14] AI can raise productivity, lower production costs, improve matching in markets, and make public services more readily available. AI also holds significant potential to improve resource efficiency, support circular economy models, enable energy system decarbonization, and enhance the monitoring and protection of ecosystems, all functions that are critical to achieving the SDGs.

Precisely because AI is becoming so deeply embedded in how societies function and decisions are made, its trajectory, namely the direction and evolution of its development and deployment, matters as much as its capabilities. While its potential to support transformations for sustainable development is substantial, the speed and scale of current advances raise fundamental questions about direction and control. **Without deliberate guidance, these same technologies risk amplifying existing inequalities, concentrating power, and placing strain on social and institutional systems, rather than strengthening them.**

Part II
Opportunities and Risks

This section of the report examines the opportunities and risks associated with the deployment of AI across several key areas: the planetary environment, energy systems, labor and industry, finance, and democracy and societal resilience. Rather than focusing on specific applications—which fall beyond the scope of this study—the analysis seeks to identify broader trends and dynamics that emerge from the literature,[15] with the aim of developing an overall understanding of how AI may contribute to, or potentially undermine, progress toward the sustainability transitions within these domains. These findings must be contextualized: the environmental and social footprint of AI is not a fixed attribute of the technology itself, but a result of specific decisions.

After examining the opportunities and risks of AI, this report identifies both cross-cutting and area-specific approaches to address challenges arising from the use of existing AI tools. The **cross-cutting approaches to govern AI**—transparency and human oversight, robust security and data protection, and inclusive, cross-sectoral collaboration—serve as foundational guidelines relevant across all five focus areas, while the area-specific measures are defined for each domain, reflecting its unique characteristics and needs.

4. PLANETARY ENVIRONMENT

Despite global goals to halt and reverse nature loss, limit global warming, and reduce pollution levels, progress has been limited on these fronts.16 This section examines the principal opportunities and risks associated with the deployment of AI in biodiversity conservation and restoration, pollution reduction, and climate change mitigation and adaptation to answer the question of how and under which conditions AI has the potential to accelerate action on nature, climate and pollution.

4.1 OPPORTUNITIES

AI is transforming environmental science through more precise, scalable, and integrated analysis of Earth systems. Its capabilities enhance the measurement, modeling, and forecasting of ecological and climatic processes, providing evidence-based insights to optimize resource management, anticipate environmental risks, and broaden participation in conservation and climate action.

Opportunity 1 - Improved monitoring, measurement, and reporting of Earth systems

One of the major transformations driven by AI is the **way in which experts and non-experts engage with Earth system-related data,** from data production and analysis, to dissemination of information.

- Existing and emerging AI applications can expand field monitoring, automate numerous tasks, standardize health indicators for ecosystems and species,[17] and provide information through advanced simulation and representation.[18]
- Improved remote sensing and image and sound recognition can enable high-fidelity, large-scale mapping of biodiversity and species distribution, while technologies such as digital twins can be used to model the likely outcomes of specific interventions.[19] **Such improved detection capabilities can be used to track trends**—e.g., habitat fragmentation, land use changes, ocean acidification—and identify anomalies, while streamlining the monitoring and reporting required for local or international compliance.[20]
- AI helps overcome the lack of high-quality data necessary to track the precise rate and scale of environmental change.[21] When associated with satellite imagery, these applications have been useful for e.g., identifying and monitoring pollution and its sources. When combined with remote sensing, they have been used to document changes at hazardous sites over time.[22] Similarly, AI can be leveraged to analyze emissions data collected with remote sensing technologies because of its ability to provide real-time insights and synthesize information across sources. This use case makes AI a valuable tool for verification and compliance with local requirements or emission reduction commitments.[23]
- Beyond raw data collection, AI can overcome data fragmentation by consolidating and integrating expansive data sets from diverse sources, ranging from satellite imagery and drone feeds to in-situ sensors, citizen science, and government records. This synthesis of information allows AI to understand environmental shifts rapidly and across multiple scales, from local habitat degradation to broad landscape change from phenomena such as deforestation.[24]
- AI could minimize data waste in the field of conservation. Over the past few decades, scientists and practitioners have collected a wealth of data, much of which remains underutilized.[25] AI's ability to synthesize and integrate diverse datasets may provide a potential remedy to this problem.[26]

Opportunity 2 - Refined scenario planning and predictive forecasting

- AI's unique ability to synthetize data not only enhances the understanding of environmental scenarios in real-time, but also significantly increases the accuracy of predictive forecasting. This predictive power applies to complex long-term trends such as habitat loss, ocean acidification, and global pollution. Currently, AI-assisted predictive modeling is most commonly used to model climatic, disaster-related, and ecological and disease shocks.[27]
- Early warning systems represent a specific and distinct application of this predictive capacity. Synthesizing historical records with current environmental variables allows AI-driven tools to detect early ecological hazards and extreme events, including floods, droughts, wildfires, and hurricanes. **AI-supported early warning systems can enable proactive disaster risk management,** helping communities plan and prepare for future conditions. For instance, AI-driven urban resilience tools can be used to support infrastructure planning by identifying vulnerabilities and optimizing land use.[28] Such a shift from a reactive stance during emergencies to a proactive, evidence-based approach has the potential to bolster communities' long-term resilience and decrease reliance on external post-disaster interventions.[29]
- In the context of predictive forecasting for climate mitigation, AI enables the discovery of new CO2-absorbing materials by modeling and anticipating how different atomic or molecular configurations will perform under specified conditions.[30]

Opportunity 3 - Better informed and optimized environment and resource management

- Leveraging large-scale environmental datasets and AI-driven predictive analytics, stakeholders can optimize natural resource management strategies through evidence-based, scenario-informed decision-making. For example, AI models can forecast water demand and availability across river basins, guiding irrigation planning.[31] This is already happening in some sectors such as agriculture, where AI-supported applications allow farmers to make more informed decisions on e.g., irrigation and optimizing water usage or crop management and reducing reliance on fertilizers and agro-chemicals that might otherwise leak into the water and soil.[32]

Opportunity 4 - Democratizing environmental knowledge

- None of these practical applications would be possible without translating complex scientific information into actionable insights. This is often referred to as AI's ability to democratize environmental knowledge, namely by bridging the gap between complex scientific data and public action.[33] Making environmental information more accessible has several practical implications, including potentially increasing participation in environmental decision-making processes from a diverse group of stakeholders and facilitating communication of technical knowledge in both formal and informal contexts.

4.2 RISKS

The integration of Artificial Intelligence into various sectors of our economy brings with it a complex matrix of sustainability risks and ethical challenges that extend far beyond technical errors. These challenges are systemic, impacting the physical environment and the integrity of global data landscapes.

Risk 1 - Intense pressure on natural resources

The physical reality of AI is deeply resource intensive.

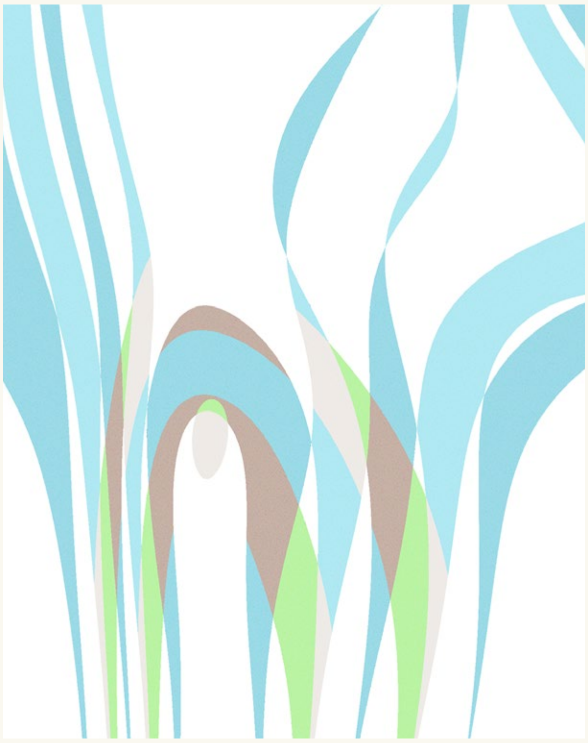
- Computation comes with an ecological and climate footprint generated by its underpinning infrastructure. The more AI proliferates as a consumer product through digital services such as chatbots and synthetic content generation (e.g., videos and text) the more this footprint will increase.[34]
- **The entire AI value chain**—from the mining of critical minerals for semiconductors to the operation of massive data centers—**puts intense pressure on natural resources.**
  - The escalating demand for hardware necessitates the expansion of mining operations, which frequently leads to habitat fragmentation and ecosystem degradation.
  - Both data centers' equipment and offsite electricity generation—unless solar and wind based—require water for cooling due to their high thermal loads, which can exacerbate regional water scarcity if no drastic water saving or recycling measures are in place; AI training and deployment are linked to high energy consumption and CO_2 emissions, especially when not paired with renewable deployment;
  - The resulting surge in electronic waste continues to outpace recycling capabilities, leading to long-term pollution and resource loss; the increasing amount of e-waste (e.g., obsolete processors) increases the risk of toxic substances—such as lead, mercury, and cadmium—being released into the environment and polluting key resources such as soil and water.[35]

Risk 2 - Biosafety Risks

- The acceleration of biotechnology through AI allows for the design of new lifeforms or proteins, introducing unprecedented biosafety risks. These risks are potentially irreversible, with wide-ranging impacts on ecosystems and species.[36]

Risk 3 - Malicious Use of AI

- AI can be leveraged by malignant actors, such as poachers, to optimize illegal activities, while simultaneously increasing the risk of large-scale data breaches.[37]



Risk 4 - Data inequality, inaccuracy, and exploitation

- Since the utility of AI is fundamentally limited by the quality of its inputs, when training data is poor, limited, biased, or misattributed, the resulting outputs can be actively harmful.
- Additional vulnerabilities include "transfer context bias" and "interpretation bias".
  - Transfer context bias occurs when an AI system, originally optimized and trained for a specific ecological, climatic, or social-ecological environment, is applied to a different or unintended context, without appropriate recalibration. The mismatch between training environments and real-world applications not only undermines the efficacy of AI-driven interventions but can also lead to unintended consequences for local ecosystems and the communities that depend on them. Similarly, errors arise when a system is trained on historical data that are no longer valid. Even if AI is applied to the correct environment, shifts in that environment that are not reflected in the model may render generated insights obsolete.
  - Interpretation bias is when an AI model performs as it is intended by its designer, but the user lacks an understanding of the system's specific utility and limitations, leading to a potential misuse of the AI-generated insights.[39]
- A primary concern is also the phenomenon of "hallucinations," where generative models produce factually incorrect or misleading information, often presented as coherent text.
- AI also risks reinforcing existing social and geographic disparities. In the field of conservation, for instance, a lack of data regarding specific species or regions can lead AI-driven recommendations to neglect or underrepresent critical ecosystems. This inequity is further compounded by "AI colonialism," a dynamic where data is extracted from the Global South to train models for the Global North.[40] This is the case when Indigenous Peoples and Local Communities (IPLC) contribute data but lose control over how the data is used and who profits from it. Collecting environmental and community data through AI systems without ensuring consent from IPLCs raises questions about data sovereignty and governance as well as about exploitation or misrepresentation of IPLC knowledge systems.

Risk 5 - Lack of transparency caused by algorithmic "black box"

- Central to the difficulty of regulating these technologies is the lack of transparency. The algorithmic "black box" makes it nearly impossible for external observers to audit how AI reaches its conclusions.[41] This opacity complicates oversight and accountability because it is challenging to scrutinize an AI-generated decision on, for example, natural resource management, when the decisions may draw on vast, proprietary digital sequence data with unclear ecological implications.[42]

Risk 6 - Misinformation and disinformation undermining climate and environmental action

Misinformation can be defined as information whose inaccuracy is unintentional. Disinformation on the other hand, refers to information that is deliberately false or misleading.[43] The climate, biodiversity, and pollution crises require collective action, which could be undermined by widespread mistrust in science resulting from incorrect climate and environmental information amplified through digital media.[44] In particular:

- Generative AI systems can produce highly realistic text, images, video, and audio with minimal human intervention, increasing the scale, sophistication, and accessibility of misinformation. At the same time, AI tools can fabricate authoritative-sounding articles, deepfake videos, and targeted messages that misrepresent climate science and policy. Personalization and real-time interaction on digital platforms enable AI to amplify persuasive narratives tailored to specific audiences, increasing its likelihood of influencing opinions and behavior.
- Rapid proliferation of misleading content can erode public trust in scientific evidence, weaken understanding of environmental risks, and delay or misdirect sustainability transitions.


WIND POWER

5. ENERGY SYSTEMS

There is no AI without energy. At the same time, AI has the potential to transform the sector's future.[45] In the context of the global energy transition, AI is a critical enabler for bolstering grid stability, optimizing energy consumption and industrial processes, and facilitating the large-scale electrification of transport and heating. This section examines the synergies between AI and the energy transition to determine how and under which conditions AI can be strategically leveraged to accelerate the shift toward cleaner energy systems.

5.1 OPPORTUNITIES

The opportunities posed by AI deployment in the energy systems transition can be grouped into four categories, namely energy systems efficiency, optimization of renewable integration and circularity, acceleration of knowledge production, and modernization of demand-side markets.

Opportunity 1 - Energy systems efficiency in generation, transmission, and distribution

AI can make energy systems more efficient by optimizing resource use, while reducing emissions and energy waste throughout the energy value chain. **AI may optimize system operations, maintenance, and planning to increase electricity transmission capacity without building new lines.**

- AI may help enhance centralized power generation primarily through predictive maintenance systems that monitor infrastructure to detect mechanical wear before it leads to system outages. Simultaneously, optimization software recalibrates combustion processes to reduce both fuel intake and environmental impact. The integration of these predictive and analytical tools lowers operational costs and reduces emissions.[46]
- On the transmission side, AI can be employed to bolster energy resilience thanks to improved forecasts for loads, power demand, and supply. Forecasts enable transmission system operators to anticipate spikes in electricity demand or drops in renewable output.[47]
- In distribution operations, real-time analytics from smart meters allow automated monitoring and tracking of loads, with subsequent opportunities to optimize resource distribution such as shifting between on and off-peak hours. Such adaptability preserves grid stability and limits peak-load stress.[48]
- At the retail level, AI has the potential to redefine how end-users engage with electricity services. For example, for individual consumers, AI models can evaluate historical usage data and provide personalized recommendations on how to reduce energy demand.[49]
- For companies, AI can support planning future infrastructure investments based on forecasts trends about long-term needs as well as optimizing the design plans for new facilities to improve energy efficiency and reduce overall energy consumption.[50]

Opportunity 2 - Optimization of renewable integration and grid modernization

AI can support the integration of renewable resources into existing energy systems through its capabilities in data analytics, optimization, and forecasting.

- AI can forecast renewable energy generation from variable sources improving grid stability and reducing reliance on fossil fuels.[51]
- AI helps optimize the design and siting of new wind and solar installations and energy storage facilities, enabling systems to accommodate renewables more effectively. It also enhances overall efficiency allowing for real-time adaptation to fluctuations in wind or solar output.[52]
- AI supports the fine-tuning of both supply and demand, optimizing the operation of smart grids and balancing supply and demand in real time. EV batteries, for example, can circulate power back to the grid when needed, functioning as an additional power source, while new smart meters can be set to allow the indoor temperature to drop or rise when demand on the grid is peaking. While these technologies do not need AI to function, AI enables these applications to be optimized for the grid and enhance grid flexibility. This is particularly important in the context of energy hungry data centers whose AI-enabled, fine-tuned demand flexibility can be a strategic asset for the grid.[53]
- AI can help optimize power flow problems, especially when there are multiple sources of energy.[54]

Opportunity 3 - Acceleration of knowledge production

AI can support data collection, analysis, and innovation that promotes the energy transition.

- AI can assist with data collection and help fill in incomplete datasets, thereby improving the quality of information. It also reduces time and resources necessary for the processing and analysis of larger quantities of data—e.g., datasets used in research.[55]
- AI-enhanced processes can support innovation toward the development of new technologies, such as advanced battery chemistries, catalysts, and CO2 capture materials.[56] For instance, in the field of material science, AI shortens the time necessary to develop new materials essential for the energy transition, such as metal-organic frameworks (MOFs) and aerogels for thermal insulation.[57] On the one hand, AI enables high-speed, atomic-scale simulations that provide a deeper understanding of how material structure and chemical reactivity impact performance; on the other hand, it synthesizes vast bodies of scientific literature to generate hypotheses that guide the trajectory of human-led research.

Opportunity 4 - Modernization of the demand-side market

AI improves consumer choice and increases opportunities for sustainable habits.

- AI can provide personalized, data-driven recommendations to consumers that informs their energy consumption habits.
- AI can create demand-side incentives for more sustainable habits, such as incentivizing off-peak consumption and enabling dynamic pricing.

5.2 RISKS

The integration of AI into energy systems introduces risks that could undermine the achievement of the sustainable development goals. These challenges range from the "high-carbon lock-in" of fossil fuel infrastructure to digital vulnerabilities inherent in large-scale AI deployment.

Risk 1 - Fossil fuel entrenchment and transition delay

AI-driven efficiencies risk entrenching fossil fuel reliance.

- AI has been and can be used to enhance fossil fuel exploration and extraction. By leveraging log and seismic data, AI models accurately predict critical reservoir properties like permeability and porosity, enabling more precise hydrocarbon recovery management. Furthermore, these systems utilize vast datasets from simulations and field logs to forecast technical complications and emergencies before they occur.[58] Collectively, these advancements bolster industry efficiency, which risks prolonging the profitability of carbon-intensive industries and creates a direct conflict with international decarbonization and renewable energy mandates.
- Much like the use of AI for more efficient exploration and extraction, the oil and gas industry has also been able to leverage AI for automating production and streamlining complex industrial processes. This has lowered operational costs, which makes plants profitable for a longer period of time.[59]
- does not necessarily result in decreased energy use. Per the Jevons Paradox, efficiency may instead offset expected reductions in energy consumption—a dynamic that is particularly problematic when the additional energy demand remains largely fossil-fuel based.[60]
- AI is itself electricity-intensive: the rapid scaling of data centers is driving immediate increases in power demand, which in many systems are being met by fossil fuel-based peaking and backup generation to preserve grid reliability.[61]

Risk 2 - Strain on physical and digital infrastructure

The rapid expansion of AI introduces significant strain on physical and institutional infrastructure. This risks manifesting through three primary dimensions:

- Grid instability and economic pressure: The energy-intensive nature of AI data centers places unprecedented demand on electricity grids. When not strategically managed, this surge threatens operational reliability and risks driving up energy prices for all consumers, potentially destabilizing local energy markets.[62]
- The digital and skills gap: Adoption of AI is often hindered by a lack of fundamental digital infrastructure and literacy. For example, AI-powered pricing models can benefit households with flexible consumption patterns, but lower-income communities may lack the digital access, stable electricity, or even network infrastructure needed to participate, transforming this digital divide into an energy divide.[63]

The biggest challenges are in developing countries—particularly those classified as "Least Developed Countries and Small Islands Developing State"—and in remote and rural areas. These countries face digital infrastructure limitations, including unreliable internet connectivity, insufficient computing power, and a lack of skilled professionals to develop and deploy AI systems in the energy sector.[64]

- Global structural dependencies: The mineral requirements of AI infrastructure risk perpetuating colonial dependencies by centralizing AI innovation and data-use capacities in the Global North, while Global South economies primarily supply raw resources. This structural imbalance is further reinforced by energy and digital inequalities: high-capacity data centers and AI computation are concentrated where reliable electricity and advanced digital infrastructure exist, typically in the Global North. Consequently, the ability to generate, process, and benefit from AI-driven insights remains unequal, entangling energy access, digital capacity, and resource extraction within a broader colonial pattern of global power and value concentration.[65]

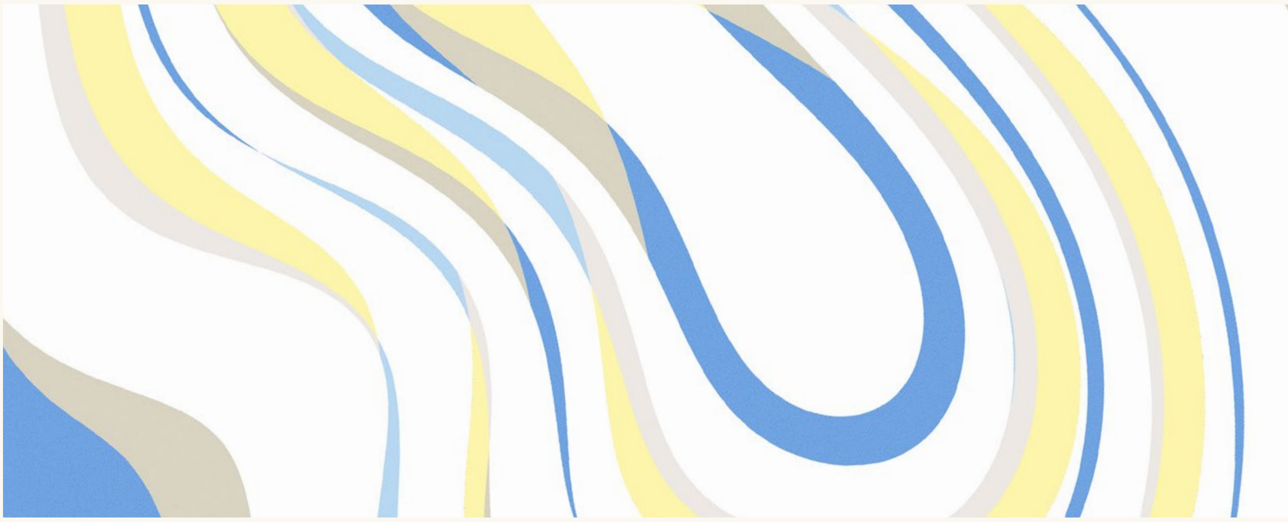
Risk 3 - Vulnerability to cybersecurity risks in digitized energy systems

- As energy systems become increasingly digitized, their vulnerability to cyberattacks and data breaches grows exponentially.[66] Reliance on interconnected digital networks provides new entry points for attackers, who could inject false signals or steal sensitive data.

Risk 4 - Data limitations and lack of transparency

AI systems depend on extensive, high-quality data, from historical demand patterns to real-time grid information. If this data is compromised or manipulated, AI outputs can become unreliable, eroding efficiency gains and potentially disrupting critical energy infrastructure. In particular:

- Data scarcity and poor data quality represent the primary barriers to AI adoption in the energy sector, as they hinder the development of AI models capable of delivering meaningful operational insights.[67] Furthermore, biased datasets may marginalize vulnerable communities already grappling with energy poverty.
- As AI-driven decisions—such as dynamic utility pricing or infrastructure investments— become more complex, they often become less interpretable. This lack of transparency risks eroding public confidence.[68]



5.3 EXPERT INSIGHTS

Excerpts from an interview with Dr. Aidan O'Sullivan

Aidan O'Sullivan is an Associate Professor in Energy and Artificial Intelligence at University College London Energy Institute, where he leads the Energy Systems and Artificial Intelligence Lab and the department's Data Analytics research theme. He is also the founding course director of the Energy Systems and Data Analytics MSc, an innovative program that combines the study of energy systems and sustainability with data science and machine learning.

1. Could you break down in really simple language the specific ways in which AI and infrastructure can be used for energy systems for the better?

At the end of the day, it all comes down to efficiency and the interface between software and hardware (i.e. infrastructure). You can't make more hardware using the software at present, but you can operate it better and that's what we use AI to do: to try and optimize and efficiently utilize the infrastructure that we do have and get more out of it.

In the UK, for example, we have an issue with the curtailment of wind farms, where they are turned off when they could generate power. It's essentially because the grid can't handle it, but you could use AI to optimize the grid so that other generation was turned down to make space for this renewable curtailment — then you could be getting an extra 10% of wind generation onto the system, which is as if you had an extra 10% of wind fleets built out without needing to build it. Essentially, it comes down to operational efficiency.

Current AI is suited to closed systems: It needs well defined boundaries and well described systems to operate in. A power grid is an ideal example of this, because the components of the system don't change from one day to the next and are well described. You have quite a consistent, well-defined operation and what you're trying to do is avoid e.g., blackouts and operate the infrastructure as efficiently as possible. Because it has a centralized control room, you have one point where all information feeds into. This is ideal for AI: you don't have lots of different distributed components, but you do have one central source of information, and all the boundaries well-defined.

2. Where has AI's impact been overstated?

LLMs. The data for LLMs is all-natural language, and the problems in the energy system are very rarely natural language-based problems. They're usually sensors, generation, machinery, and industrial signals. LLMs can provide an interface, but they still require a lot of reworking and repurposing to make them work in energy.

We pay for reliability in the energy sector — we don't necessarily want speed in our responses, but we want to know we're getting things 100% right, and that we can explain those decisions as well. These are aspects of the energy sector as critical infrastructure that are at odds with the LLMs, as they are designed now, which is to be used by consumers in daily life.

3. Data centers are resource intensive and could potentially create situations where you need more power, potentially spurring a return to fossil fuels because renewables are not enough and not efficient enough. What would be a risk mitigation way (e.g., policy, design elements, etc.)?

I'm conflicted on this one because there are definitely some positives and interesting angles. As part of the energy transition we're trying to encourage investment into the power sector, which has had declining demand in the UK for the last decade or so: it goes down by about 1-2% per year. Getting investment into a sector that's declining is tricky, but getting it into a sector that's growing is easier, and that's what we're seeing with the data center boom.

The demand for data centers has encouraged greater investment into the power sector from people who have a lot of money, like tech companies. On the other hand, they do absorb a lot of the progress that we've made in decarbonizing the power sector by absorbing that demand.

The way things are happening in the UK is that data centers that can be flexible are being prioritized for grid connection. This is one policy thing you can do: you can jump people up the queue if they agree to limit demand during peak hours, which reduces strain. The grid is usually at about 40-50% utilization, except for the peak hours, where you could push it over the limit by 1% and the whole thing goes down. Having that flexibility is critical and turns it into a net positive in some sense.

In Ireland, data centers take up 20% of the grid and electricity consumption, so one of the requirements is that they build additional renewable capacity if they want to come online as part of the grid. You can put these kinds of restrictions in place because the demand is there.

Another positive thing that I'm seeing is that countries like France and Norway, who have very low carbon grids, are getting the data center investment ahead of more carbon intensive grids like the UK, which is competing and not doing as well because their grid is about 10 times more carbon intensive. In some sense, data center investment, because of the renewable preferences of some tech companies, is favoring grids that have already decarbonized.

In terms of the key kind of things: I think if you can get a data center to flex, then it can become more positive on the grid than negative overall. However, their capacity to flex depends on their actual usage and what they're doing.

4. You also work in hard-to-abate sectors. Is AI unlocking opportunities in these sectors?

There's a couple of different things here. The way you make cement is by heating rock up to 1500 degrees. To do that, you need density of energy, which is where fossil fuel is undefeated, unfortunately; and you need it to be cheap, and fossil fuels are very hard to displace in that regard.

You also have the chemical emissions from the chemical process: 50% of the CO2 comes from driving CO2 out of limestone to turn it into cement, which you do by heating it up. 50% comes from the heat energy, and 50% comes from the chemical energy.

We've been making cement since the Romans, but over time we've changed the mixture to lower the melting temperature. Now you need 1300 degrees. In the past, you probably needed 1600 degrees. However, because you can add e.g. magnesium, it lowers the heat of reaction, which is important. Innovating in that matter requires a highly complex simulation of chemical reactions, which is challenging and AI could do something there which could be quite impactful.

5. What do you think are the most crucial elements, policy or governance wise, to minimize the societal and environmental impacts? Or maximize the benefits of AI?

In terms of minimizing the impact, policies need to be based on the understanding that data centers are flexible, but that they also contribute to the build out of transmission and renewable generation to make them net positive. If they are just plugging into a grid without adapting to the situation, then they are net negative.

In terms of enhancing their benefit for the energy transition, I think a greater awareness of the problems in the energy sector from the people building data centers is extremely crucial. We're still quite early in the journey; for a technology in the energy system, AI is incredibly young and to adopt something that's two or three years old is quite unusual. I think this will require actively working to minimize the weaknesses that are relevant for energy systems, such as reliability. There are ways of doing that, but it requires significant domain expertise to validate an energy systems model, which may be very difficult for a big tech company to do alone — they have to collaborate with energy domain experts.

6. INDUSTRY AND LABOR

AI is beginning to transform industry and labor markets at an unprecedented pace. Production processes are increasingly optimized in real time, reducing human intervention while enhancing efficiency and precision. Concurrently, labor organization is reshaped as AI reallocates tasks, creates new roles, and renders some skills obsolete.[69] Collectively, these shifts constitute a systemic reconfiguration of industrial and workforce structures.

6.1 OPPORTUNITIES

The adoption of AI in industrial systems and in workplaces presents both systemic and specific opportunities.

Opportunity 1 - Innovation and structural transformation

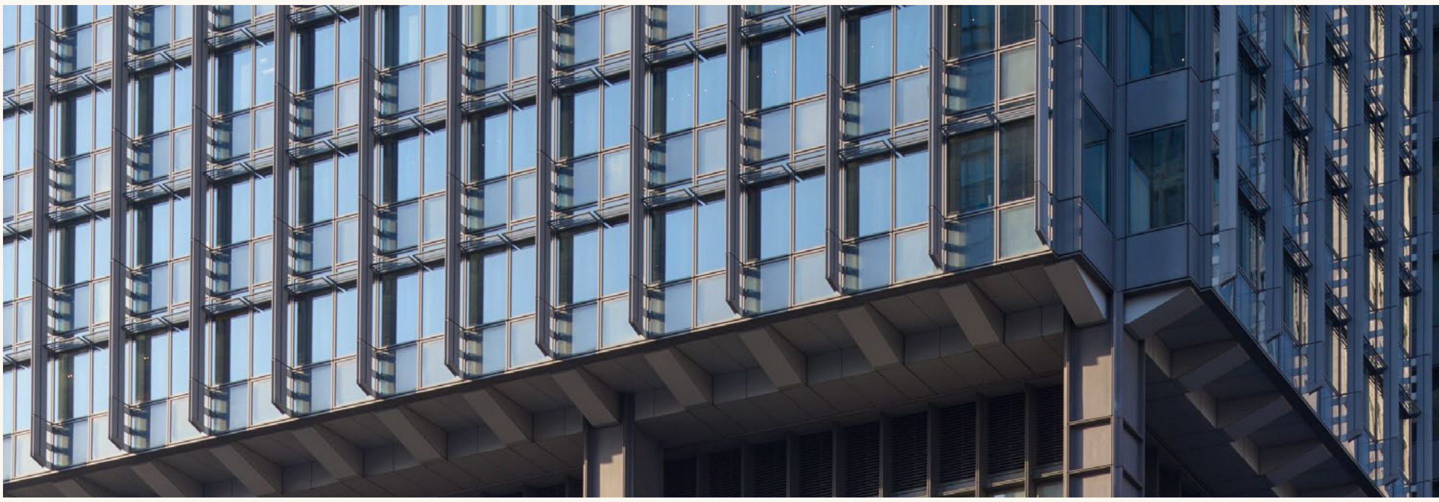
AI acts as a general-purpose technology that reshapes industrial structures and labor markets by accelerating innovation, enabling new sectors, and changing the composition of skills demanded across the economy. Its primary impact lies not only in improving existing processes but in expanding the frontier of what industries and workers can produce.

- The increase in patent applications suggests that AI adoption has already contributed to higher levels of innovation. Beyond optimizing existing processes, AI also has the potential to give rise to entirely new sectors and jobs that do not yet exist.[70]
- Accelerating economic growth through innovation is therefore considered one of the key benefits of AI. AI supports innovation that can boost economic productivity by enhancing research and development and enabling the generation of new discoveries.[71] Through AI-supported processes such as digital platforms, smart manufacturing, and predictive maintenance, new sectors and business opportunities may be unlocked.

Opportunity 2 - Productivity gains through task reallocation and process optimization

AI raises productivity primarily by reallocating tasks between humans and machines, automating routine or data-intensive activities, while allowing workers to focus on higher-value tasks such as supervision, problem-solving, and coordination.

- In manufacturing, predictive maintenance systems reduce unplanned downtime, increasing output while shifting labor toward maintenance planning and system oversight roles. In logistics and services, AI-supported scheduling and workflow optimization improve output without proportional increases in labor hours.[72]
- AI can also support greener industrial processes through energy efficiency, more efficient resource use, and integrated circular economy approaches.[73]
- These developments could also create substantial employment opportunities, with some estimates projecting as many as 170 million new jobs.[74]



Opportunity 3 - Quality enhancement and workplace safety improvements

By enabling real-time monitoring, pattern recognition, and predictive risk detection, AI can improve product quality and workplace safety, simultaneously enhancing industrial performance and job quality.[75]

- For example, AI-driven tools, such as machine learning for disease detection, computer vision for hazard monitoring, and sensor-based systems for real-time exposure tracking, can substantially reduce workplace accidents and occupational diseases while improving productivity and operational efficiency.[76]
- AI also holds promise for **enhancing creativity**. In intellectual work, AI can be used to bridge disciplines. Since AI systems help researchers connect across different fields to answer promising questions, AI has the potential to break down disciplinary silos even as it narrows focus within specific domains (del Rio-Chanona et al., 2025).

Opportunity 4 - Enhanced decision-making in complex dynamics

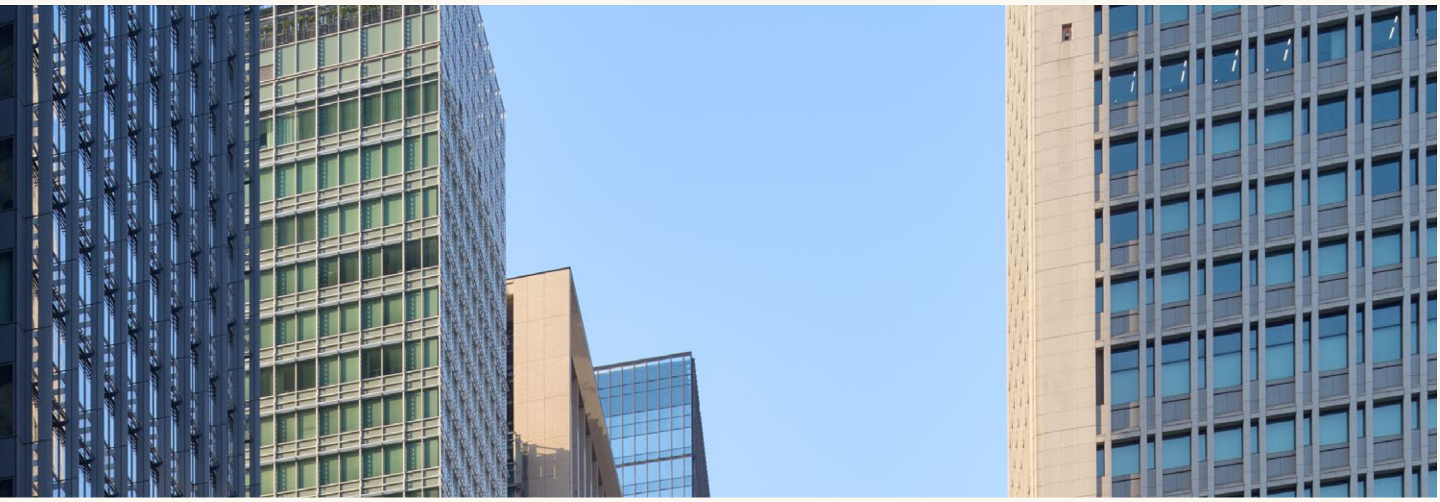
AI improves decision-making by expanding the capacity to process complex information, simulate scenarios, and support evidence-based choices, particularly in dynamic and uncertain industrial environments.

- The use of digital twins—digital replicas of goods, machinery, or even entire factories—allows firms to simulate changes and test scenarios without the real-world consequences of physical trials.[77]
- In human resource management, AI-supported analytics improve workforce planning and scheduling, reshaping managerial roles toward interpretation and judgment rather than routine administration.[78]

Opportunity 5 - Creation of new jobs and occupational transformation

AI adoption generates new occupations and transforms existing ones by increasing demand for skills related to system design, supervision, data interpretation, and human–machine coordination.

- Manufacturing firms increasingly employ AI technicians, data analysts, and quality-assurance specialists to manage and interpret AI-driven systems.[79]
- Across sectors, new roles emerge in model governance, AI ethics, and algorithm oversight.[80]
- AI also holds promise for enhancing creativity. In intellectual work, AI can be used to bridge disciplines. Since AI systems help researchers connect across different fields to answer promising questions, AI has the potential to break down disciplinary silos, even as it narrows focus within specific domains.[81]



6.2 RISKS

As with the opportunities discussed above, the adoption of AI in industrial systems and workplaces entails both systemic and specific risks.

Risk 1 - Uneven productivity gains and widening inequality

- AI-driven transformations interact with existing forms of digital inequality and structural exclusion. The productivity-enhancing effects of AI are not evenly distributed, since high-income and high-skilled workers are often better positioned to use AI as a complement to their work, leveraging these tools to amplify output and efficiency. Furthermore, the integration of AI into work systems may reduce human oversight and limit access, particularly for informal, precarious, or geographically remote workers. This erosion of oversight risks reinforcing the conditions under which exploitative labor arrangements can persist.[82]
- More broadly, AI adoption may intensify economic exclusion, as countries and communities lacking adequate digital infrastructure and skills risk falling behind. Disparities in AI access and capabilities can further exacerbate income inequality by concentrating benefits among those already positioned to participate in the digital economy.[83]

Risk 2 - Increased market concentration and monopolistic dynamics

- AI may also contribute to increased market concentration and monopolistic dynamics. Critical AI inputs and infrastructure—including data, proprietary models, and computational resources—are often controlled by a small number of large technology firms. This concentration of power and resources limits competition, slows the diffusion of AI technologies, and creates dependencies that can constrain innovation and overall market dynamism.[84]

Risk 3 - Risk of workforce policy misalignment due to labor market uncertainty

- The high level of uncertainty surrounding AI's impact on the labor market creates a significant risk that workforce policies oriented toward specific future skills or occupations may be misaligned with actual labor market needs. This could lead to ineffective use of resources and leave workers unprepared for the kinds of disruptions that ultimately occur (see interview with Ole Teutloff).

Risk 4 - Labor displacement and job polarization

- AI adoption carries significant risks of labor displacement across a wide range of sectors. AI-powered automation is expected to disproportionately affect jobs that rely heavily on routine or repetitive tasks. While manufacturing has long been at the forefront of automation, with robotics replacing assembly-line workers,[85] the scope of AI-driven displacement extends well beyond industry to include service-oriented sectors such as retail, hospitality, and logistics. At the global level, approximately 24% of jobs have some degree of exposure to generative AI, rising to nearly one in three jobs in high-income countries.[86]
- The risks of displacement also increasingly affect entry-level professional roles. Positions such as paralegals are increasingly being substituted by AI systems. Generative AI, in particular, heightens displacement risks for clerical and administrative workers.
- A notable gender gap persists: in high-income economies, 9.6% of female employment exhib-

its high automation potential, compared to 3.5% of male employment. While exposure does not imply immediate job loss, it indicates that a substantial share of tasks could be performed by AI. The long-term consequences will depend on adoption trajectories and on workers' capacity to adapt and reskill.[87]

- AI has increasingly become a driver for layoffs. Many companies are firing workers not because AI systems have already delivered measurable productivity gains, but because executives expect they will soon. According to a survey of more than a thousand global executives conducted in December 2025, 39% reported making low-to-moderate headcount reductions in anticipation of AI-driven efficiencies, while only 2% said their cuts were based on realized productivity gains from AI systems.[88]

Risk 5 - Workplace alienation

Next to displacement, the use of AI in the workplace can also negatively affect workers' well-being.

- When AI systems are deployed to manage labor through algorithmic mechanisms that assign tasks, monitor behavior, and administer discipline in opaque ways—such as in platform-based gig work or logistics—workers often experience diminished control, autonomy, and psychological safety.[89]
- Similar dynamics are emerging in white-collar environments, where AI is increasingly embedded through performance dashboards, automated scheduling tools, and AI-powered hiring systems. Workers frequently express resistance to being managed by AI, citing perceived status loss and an associated "empathy penalty." [90]

Risk 6 - Erosion of collective diversity

AI may also have unintended effects on knowledge production and creativity. Although AI can enhance individual performance and support the generation of creative ideas, it may reduce collective diversity. As individuals increasingly converge on similar AI-assisted outputs, the overall variety of results declines. This tendency reflects AI's comparative advantage in accelerating work within established, data-rich domains, rather than in exploring new or underdeveloped areas, which may lead to fewer new subfields over time.[91]

6.3 EXPERT INSIGHTS

Excerpts from an interview with Dr. Ole Teutloff

Ole Teutloff is postdoctoral visiting research fellow at the Oxford Internet Institute and affiliated guest researcher at the Copenhagen Center for Social Data Science. Ole is a co-founder of the DWG Data Science Company. He has previously worked in a variety of international and interdisciplinary contexts including the Centre for the Governance of AI, the OECD and the KfW Developing Bank. He has also served as a consultant for the World Bank and the International Labour Organization (ILO).

1. What distinguishes AI from previous technologies that impacted the labor market?

There are vastly different opinions and a debate between people that think AI is just like any other technology and people saying it will fundamentally change everything we do. I don't think that this debate can be settled currently, because it all depends on how AI capabilities evolve over the coming years.

At the moment, my impression is that we can already say AI is different from, for example, the computer and first wave of digital technologies, to the extent that it really helps do more non-routine cognitive tasks: tasks like writing text, writing computer code, helping people make decisions at a much higher cognitive level and across almost any field. I think for me, this is the new quality: that it is a general purpose tool for cognitive work. You can actually outsource certain tasks almost entirely to a machine and these tasks are not only calculating a formula or typesetting some texts — it's way more complex.

I think (1) the type of tasks that you can use AI for e.g. cognitive work, and in particular, non routine cognitive tasks that used to be very much only reserved for humans, and (2) the breadth of the available tasks you can use it for. I think of these as the key qualities. Then it's just to be seen how these capabilities evolve. I don't think we can make any good predictions at the moment.

2. Augmentation v Substitution. Could we know already whether AI adoption in a specific role or in a specific industry will lead to labor augmentation or displacement because of substitution?

That's the big question that people are debating and that nobody really has an answer for because it depends so much on the capabilities of AI and how they evolve.

What maybe this huge focus on automation misses is that AI is probably also creating lots of new tasks and lots of new jobs and they may be difficult to anticipate at the moment — and therefore don't show up in any of the measures. Most current studies use pre-existing classifications of tasks and jobs, and there you can measure (or you can try to quantify) which of these tasks can be automated and which ones are augmented or complemented. But it's difficult to measure what kind of jobs or tasks will be there in 5 or 10 years, and how many of them will be available, because we don't know. I think this is one big piece that is missing from the debate at the moment.

One thing to monitor closely is the speed of change, because this determines, in my opinion, a lot about how disruptive this technology can be for society. If changes play out over longer periods of time, I think our societies and economies are quite resilient in adapting to change. But if changes happen in the matter of a few years, then things may start to be tricky, and we have to think about how to react. So the speed of change is, in my opinion, one of the key parameters that determines how disruptive and perhaps how negative the consequences of the change can be.

One perspective that I personally like a lot is that we, as humans developing this technology, also have some agency in deciding what we want to focus on building. We can focus on building AI that is an automation tool, but we can also build AI that is more a collaboration tool. The narrative seems to favor automation but I think at the current stage of AI capabilities, augmenting human workers is probably much more promising because the AI doesn't need to be perfect to create a lot of value. This speaks to the question of what do these tools ultimately achieve? Do they help us to do the same things faster and cheaper or do they help us to do more?

3. **What could you say about the characteristics of jobs that can be replaced by AI, and what are the characteristics of work that might remain at low exposure?**

As of now, the types of tasks that can be automated are cognitive work of simple to medium complexity. E.g. simple writing, simple translation, simple programming. This is work that has limited ambiguity and limited scope in length and complexity. I also think current models are extremely good at anything related to processing large amounts of text data, retrieving information, and summarizing information.

What really remains the competitive advantage of humans for the moment is anything that requires higher order cognitive skills e.g. strategic thinking, longer term analytical thinking, judgment making, decisions under uncertainty, and making decisions that follow a longer vision.

One good example is a study where people from Harvard gave micro-entrepreneurs in Kenya chat bots and observed how they behaved. I would say entrepreneurship is a prime example of a challenging task because you are operating under high uncertainty, but you still have to make day-to-day decisions about how to spend your limited resources. What the researchers could show is that people who had lower abilities to begin with actually performed much worse with AI than without AI. In contrast, the people who had maybe more experience or better judgment in deciding which AI advice was good or bad did much better.

[*The study refers to is The Uneven Impact of Generative AI on Entrepreneurial Performance"* and the research is based on an experiment with 640 Kenyan entrepreneurs who used a GPT-4-powered AI business mentor via WhatsApp].

I think what we could potentially see is an increase in performance inequality between people because you can get a lot from AI, but it very much depends on the skills that you bring to AI. I think we might see a difference between people who are really good at using these tools, and people who just don't use them much because all these higher order cognitive skills are necessary compliments to make the most out of AI. I think any job that has a large component of these kinds of skills to succeed will probably not be automated anytime soon.

4. **What could you say about the characteristics of jobs that can be replaced by AI, and what are the characteristics of work that might remain at low exposure?**

A simple and straightforward recommendation is to increase our ability of foresight by building better data analysis and monitoring infrastructure, because at the moment, there's a lot of uncertainty — we don't really know what is happening right now and much less what will happen in the next few years. This is partly because researchers like myself don't really have good, up-to-date data of what's happening in the labor market. This is a bottleneck that is not easy to solve, but at least the problem is clear. This would allow us to monitor the speed of change more closely and have a feeling of urgency, of "how urgent do we need to be in our response"? Because, on a bigger level, I think the policy response really depends on the speed of change.

We should also not forget other big trends in society that might be overlapping. For example, demographic change, lots of people retiring, people getting older — this also changes what's happening in the labor market. How many new jobs we need depends on how many young people want or need a job. Maybe we'll be happy that AI automated lots of jobs because we don't have people to do them anyways. So I think there we have a lot of uncertainty and should not look at AI as an isolated phenomena, but as embedded in other mega-trends we see happening.

I personally think reskilling as a policy is overrated. You often see it as "the thing" governments should do to help people, but I don't think there's a lot of empirical evidence that reskilling can be an effective tool to overcome disruptive change from technology. Given how uncertain the situation is right now I wouldn't even be able to tell people what to reskill into. Over the past decade or so, we've been telling people "Learn how to program, this will give you a great career". I'm not sure this is true anymore and given that we have such limited foresight, we don't really know what people should learn. I think reskilling is certainly helpful, but if AI becomes super good and starts to replace 10, 20, or 30% of all jobs out there, it likely isn't the silver bullet — we'll need other options.

I think what governments should focus on instead is to strengthen the labor market institutions and educational institutions capabilities to self adapt to change. This includes reforms to build a good education system, to build a labor market that is flexible enough to accommodate change, but has enough safety nets to protect people who might suffer from it. I think all of these things that were known before — how to build better institutions in society — are more important than ever because the better these general institutions work, the better we will be able to make decisions once we have to make them fast.

I think this would be my recommendation: don't worry about what kind of reskilling strategy to build now. Instead go back to fundamentals and say, "as long as we still are in this phase of relative stability, what kind of policy changes can we make to improve the general resilience of our society and economic system"? Then, no matter what happens, we are in a better position to deal with it.

I know this is not a very concrete policy recommendation, but I think it's more realistic and would help us not only deal with AI as a big challenge, but also with other challenges like climate change, migration, demographic change. All of these big questions somehow overlap. You can't separate them very clearly.

Excerpts from an Interview with Dr. Ekkehard Ernst

Ekkehard Ernst is Chief of the Macroeconomic Policies and Jobs Unit at the International Labour Organization. He is widely published on the implications of artificial intelligence, robots and blockchain applications for the future of work, and the transition to a sustainable society. Previously, he worked at the Organization for Economic Cooperation and Development and the European Central Bank.

1. How do you characterize the challenge of AI for labor and workforce?

For the moment, the big challenges in how AI is impacting the world of work are not so much in the workplace itself.

They are challenges, especially if you think about gig workers, but there are not massive layoffs. There is also talk about young people and young graduates having difficulties entering the labor market. That might be the case in specific occupations, especially software engineering, etc, but we don't see a generalized trend of decline in recruitment for young people and the youth unemployment rate. We published an institutional report a few weeks ago on the global state of the labor market – the World Employment Social Outlook Trends report. You can see from that report that there's no massive increase in youth unemployment rates. It's in line with the cooling of the economy more broadly, but it's certainly not a specific AI impact. I always say this as a backdrop, because it seems there are these outrageous expectations, both positive and negative, in terms of what AI is doing to the labor market.

Having said that, as the saying goes, we overestimate the impact of technology in the short run, and we underestimate in the long run. By 2040 things might look completely different and if we were to talk again at that time, we would probably see massive changes.

2. Research shows that AI is increasingly replacing entry level jobs. What does this mean for the long term impact on how people's careers develop and on skills formation?

The question of whether or not the number of jobs increase or decrease depends a bit on prices and elasticity of demand. Which is to say if you make something cheaper, will the demand increase, or will it stay stable?

Given the example of translation work, for some translation services, we see a decline, but overall, the total volume of translation services actually increased. You can see this for yourself. I'm often in China and I would be totally lost without translation services. These are services that I would not be able to pay somebody for, so clearly I don't replace anybody by using these AI tools.

The big challenge for young people who just came out of university is that your education system has not integrated these tools into your curriculum. You're left with these tools by yourself, you're not being trained in how to use them properly and so employers will have the challenge of identifying the extent to which you're ready for the type of job changes that they will see. And one of the big changes of jobs that we already see is that you will move from production to verification. So, you will no longer be the person who is producing text — or at least producing at a much lower volume — but you will be the person who will have to verify that the information provided is correct: Does it actually answer the question that you have? So the critical thinking part will become even more important.

3. How would you differentiate the impact that the different kinds of AI is having on the different types of industry or occupations?

The applications are very sector specific from what we can see.

What people would probably describe as good old fashioned AI is something you see a lot in manufacturing. A lot of the companies that we talked to have bespoke AI tools that they develop for their own supply chain e.g., inventory management tools, etc. These are things that are already mainstream and typically if you compare these types of tools with, let's say, the previous wave of robotization, these are not necessarily replacing jobs in the country. E.g. Japan is one of the most robotized countries but actually has one of the lowest unemployment rates in the world. People are not getting laid off because of robots so I don't think this is really an issue that we should be afraid of.

On the contrary, those countries who lag behind in terms of adoption of these tools might actually see that their companies lose competitiveness and will no longer be able to compete on the international market.

Now the other type of approach is for instance, in creative industries, where these tools are being used as a way of generating new ideas. There's indeed a concern with IP theft, which I recognize the importance of.

My concern would more be that these tools are "seemingly creative", meaning they provide a sense that they generate new ideas, but at the end of the day, it always turns around the same four or five issues. I was making this joke once: if you had stopped training Gen AI with the paintings before American expressionism, would it have made Jackson Pollock or Mark Rothko? My suspicion would be no, and I don't have any way to prove it, but it would be interesting to see. I call it the convex hull: the database on which these tools are trained provide you with a potential space of combinations. Within that space, the tools are great and maybe generate new ideas. But to get out of that space — to enlarge the space of ideas — for the moment, I don't see how this can be done without human intervention.

Another aspect I find extremely concerning is the fact that these tools, if they interact with each other, can create forces to circumvent market competition.

One of the examples I've heard from speaking with companies is that they used to do inventory management on the basis of historical prices, where you have e.g. 20,000 items in your inventory, and most of the items were priced at historical cost, plus markups, and that's it. They never went back to update it, because it's just too many items.

Now we see AI can do dynamic pricing. It actually assesses the price depending on e.g. seasonal patterns. For example, if you have a winter tire and it's winter, you increase the price, etc. What they do is they actually buy in data streams of prices from other companies. This type of interaction creates a collusion effect, where nobody actually talks to each other. It's just the fact that AI puts prices from each other and will automatically create a sort of collusive oligopoly, where the companies will automatically adjust the prices at higher than the market clearing level. These phenomena, for the moment, are probably still very rare. The OECD has some work on this algorithmic collusion but it's really hard for authorities to detect them. If you start to integrate these types of tools at the HR level to do e.g. compensation management, that is an obvious concern — especially in terms of monopoly power on the labor market.

This is a big red flag for me and unfortunately, our current discussions on regulation is not yet taking this kind of phenomenon into account, which I personally find much more important than many other issues in this area.

4. How do you see future governments and international government organizations priorities? What kind of policies do you think are really important at this stage, but also, in the longer term, thinking about potential AI development?

In Europe, we have a bit of experience in terms of supporting this type of transition through a social protection system. For whatever income shortfalls you have, you provide some compensatory measures. In the 1980s, the whole steel industry went bust and a lot of people needed to be reoriented, so social protection was a key pillar.

Talking mostly from the perspective of the ILO, because that's closer to where I'm sitting, there's reskilling, upskilling, skills identification, and scenario development to analyze long term trends. I think that these types of tools can be very useful and should be rolled out more. But I think for the moment, especially when it comes to really long term trends, the granularity in the data and thus projections is missing.

One last approach where we have already seen some impacts, mostly for developing or emerging economies, is using AI for a predictive policing approach for informal economies. Especially in countries with large shares of informal enterprises, national labor inspectorates are often completely overwhelmed with supervising and ensuring these businesses are complying with regulations. Albania, for instance, has implemented the mechanisms by which they use AI to identify which companies are most at risk. You can obviously apply this at the individual level i.e. the employees or the people who are most at risk of being underserved or facing discrimination, for example, you can think of women rights in the labor market, etc.

These are examples where you can use these tools to help spread the limited resources that many countries have to improve and enhance state capacity and to focus on areas of importance.

7. FINANCE

The integration of the rapidly evolving capabilities of AI is transforming the financial system. AI technologies are increasingly embedded in core financial activities, including risk assessment, trading, credit allocation, and portfolio management. These capabilities promise gains in efficiency, predictive accuracy, and data processing at unprecedented scale. At the same time, their growing deployment introduces new operational, systemic, and governance risks. This section examines both the opportunities and the potential vulnerabilities associated with expanded AI use in finance.

7.1 OPPORTUNITIES

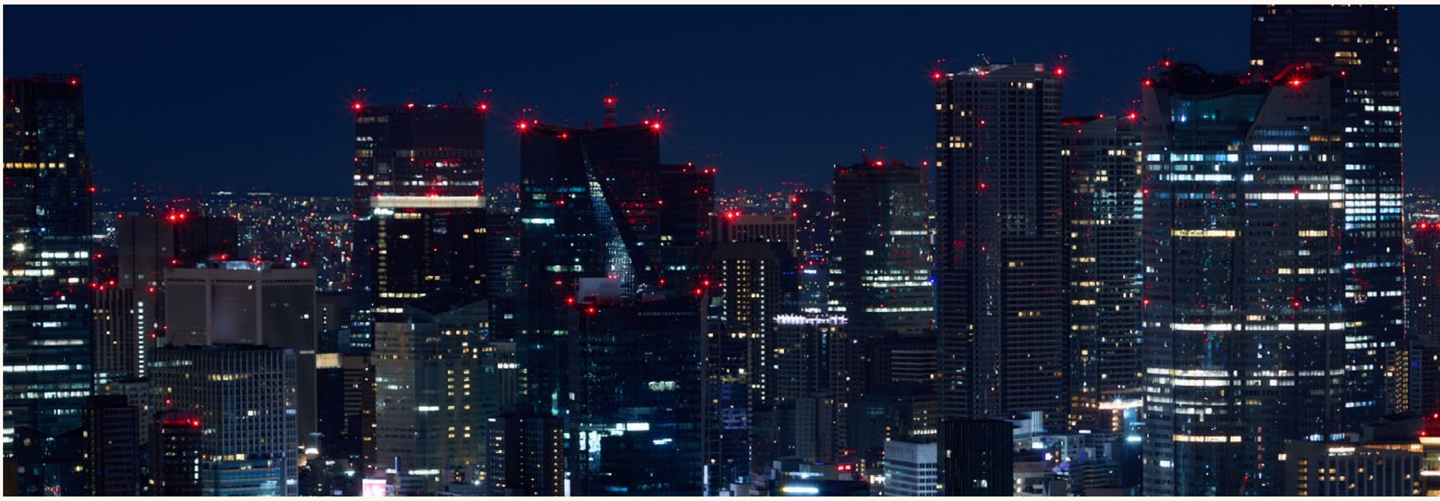
Existing research identifies several opportunities for AI in finance. In this section, we acknowledge the opportunities identified by this body of work, however many of them primarily enhance the efficiency and informational capacity of financial markets rather than fundamentally increasing the allocation of capital for projects and initiatives that support the sustainability transitions.

Opportunity 1 - Portfolio optimization and improved data collection

AI is fundamentally reshaping the efficiency of traditional financial functions by automating tasks with a speed and strategic depth that frequently exceeds human capability. AI models can enhance portfolio optimization, align investments with real-time risk profiles and specific investor objectives, and dynamically adjust holdings in response to shifting market conditions or evolving investor preferences.[92]

- AI is fundamentally reshaping the efficiency of AI also enhances information processing and data analysis by synthesizing diverse data streams—from social media to real-time news—to deliver comprehensive quantitative and qualitative insights.[93]
  Through similar methods, AI simplifies climate risk assessments through the analysis of weather and geolocation data, and creates better forecasting, empowering companies and investors to transition from reactive to proactive strategies.[94]
- Furthermore, AI can make reporting more efficient, as it automates compliance and streamlines the assessment of environmental, social, and governance (ESG) and sustainability metrics.[95]

However, optimized portfolio allocation and improved data collection, analysis and reporting do not automatically ensure that finance is allocated to support development objectives, climate and environmental goals, and equity priorities.



7. Finance

Opportunity 2 - More accurate matching between investors preference and financial products

- AI can significantly enhance the matching of investors with portfolios aligned to their financial objectives and values. AI enables more precise investor profiling, improving understanding of individual risk tolerance, preferences, and constraints. AI systems can simulate the performance of different asset allocations, support long-term portfolio planning, and identify the metrics that matter most to investors. AI can also assess corporate performance across dimensions such as ESG criteria, compare firms with relevant indicators, and surface best practices, thereby supporting more informed and values-aligned investment decision-making.[98]

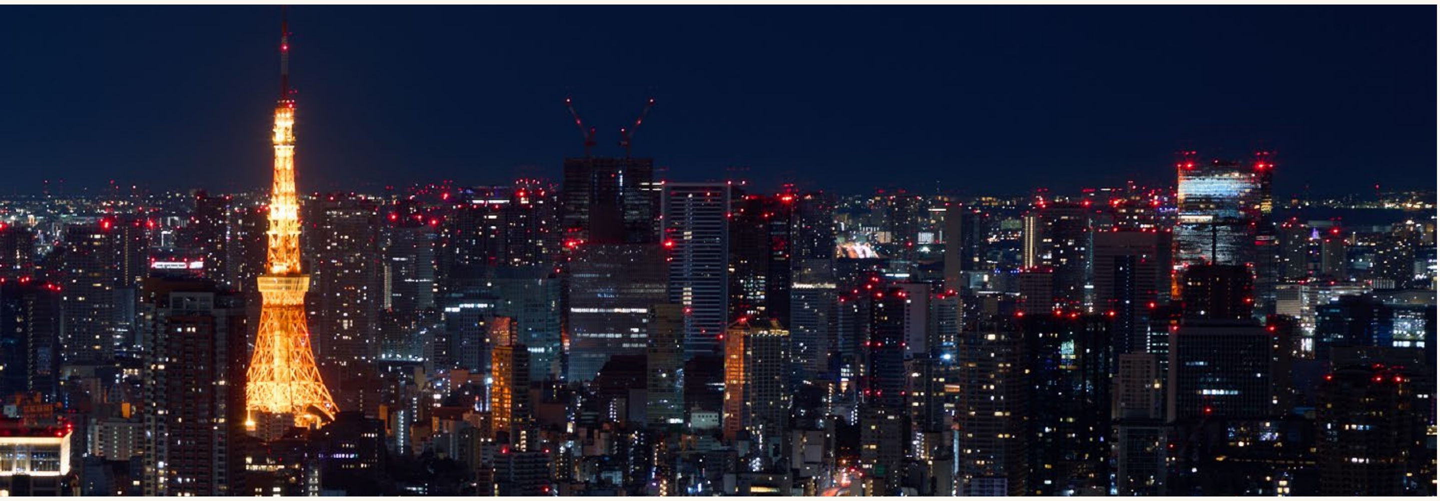
More accurate matching between investor preferences and financial products does not, however, automatically entail an increase in the overall capital directed toward the sustainability transitions.

Opportunity 3 - Improved data availability for EMDEs

Arguments that AI can accelerate climate and development finance often frame the biggest barrier as informational fragmentation.[99] From this perspective, AI can be a useful tool for accelerating capital flows for development and climate goals in emerging markets and developing economies (EMDEs) in two ways:

- The use of alternative and expansive datasets generates non-intuitive 'alpha' opportunities[100] for investors. It may also expand financial inclusion through more robust creditworthiness predictions for individuals or countries traditionally overlooked by conventional banking and investing because of their perceived risk.[101]
- The newly generated data can be integrated into the business operations and risk assessments of financial institutions, such as the World Bank, and improve the velocity of their interventions.[102]

While improved data integration can accelerate decisions, enhance disaster responsiveness, and reduce transaction costs, its effect on financing gaps depends on whether information frictions are binding. Data limitations often contribute to higher capital costs by elevating uncertainty and risk premia. However, where financing constraints reflect deeper structural conditions, AI-enabled data improvements are unlikely to substantially alter investment outcomes.



7.2 RISKS

The deployment of AI in the financial sector gives rise to a range of risks. Some are systemic in nature—such as the unsupervised use of AI and threats to financial stability—while others mirror challenges faced across sectors, including risks to privacy and the potential for biased or discriminatory outcomes.

Risk 1 - Regulatory gaps and insufficient oversight

- The complexity and rapid pace of AI development make it difficult to design and implement robust regulatory frameworks. In the absence of clear regulatory guidance for financial institutions deploying AI, there is a risk of the unregulated and insufficiently supervised use of these technologies.[103]

Risk 2 - Financial instability risks from rapid adoption and market concentration

- AI's relative novelty and market concentration has been found to undermine financial stability. The implications of AI-driven disruption in the real economy for financial stability will depend heavily on the speed and scale of AI adoption. In scenarios where highly capable AI systems rapidly automate human tasks, AI may generate significant economic shocks, including abrupt shifts in income and wealth distribution, increased corporate and household defaults, rising interest rates, declining public revenues, and heightened political instability, all of which would pose serious risks to financial stability.[104]
- The IMF identifies increased market concentration as a key financial stability risk associated with the widespread adoption of AI in capital markets.[105] These risks were seen as particularly relevant if trading and investment strategies increasingly rely on similar AI models—especially open-source models trained on the same data from a limited number of vendors. High concentration among AI models and data providers is also viewed as a potential source of systemic risk, as failures affecting a small number of vendors can disrupt trading and investment activity at scale.
- The capital-intensive financial model driving AI expansion may create vulnerabilities for financial stability and challenge the long-term economic viability of data center infrastructure. Data center developers face multiple interrelated financial risks stemming from the evolving dynamics of the AI market. Cash flow uncertainty arises as the competitive market for AI services, particularly inference workloads, limits providers' ability to recover rising operational costs, leaving insufficient funds to cover liabilities. Compounding this, the declining value of GPUs, the keystone asset underpinning AI infrastructure, introduces volatility: frequent chip upgrades and aggressive market pricing suppress the collateral value these assets can generate. Financial pressures are further amplified by tenant churn risk, as hyperscaler tenants undertake costly, repeated capital expenditure cycles within a single lease term, creating mismatches that strain developers' creditworthiness.[106]
- This is exacerbated by circular financing ("roundabouting"), where leading tenants effectively finance each other's expansion, generating interlocking liabilities and concentrated exposure for lenders and shareholders. Although hyper-scalers themselves typically maintain low debt levels, the growing reliance on opaque debt structures within this interdependent ecosystem raises serious concerns about systemic risk across the AI data center sector.[107]

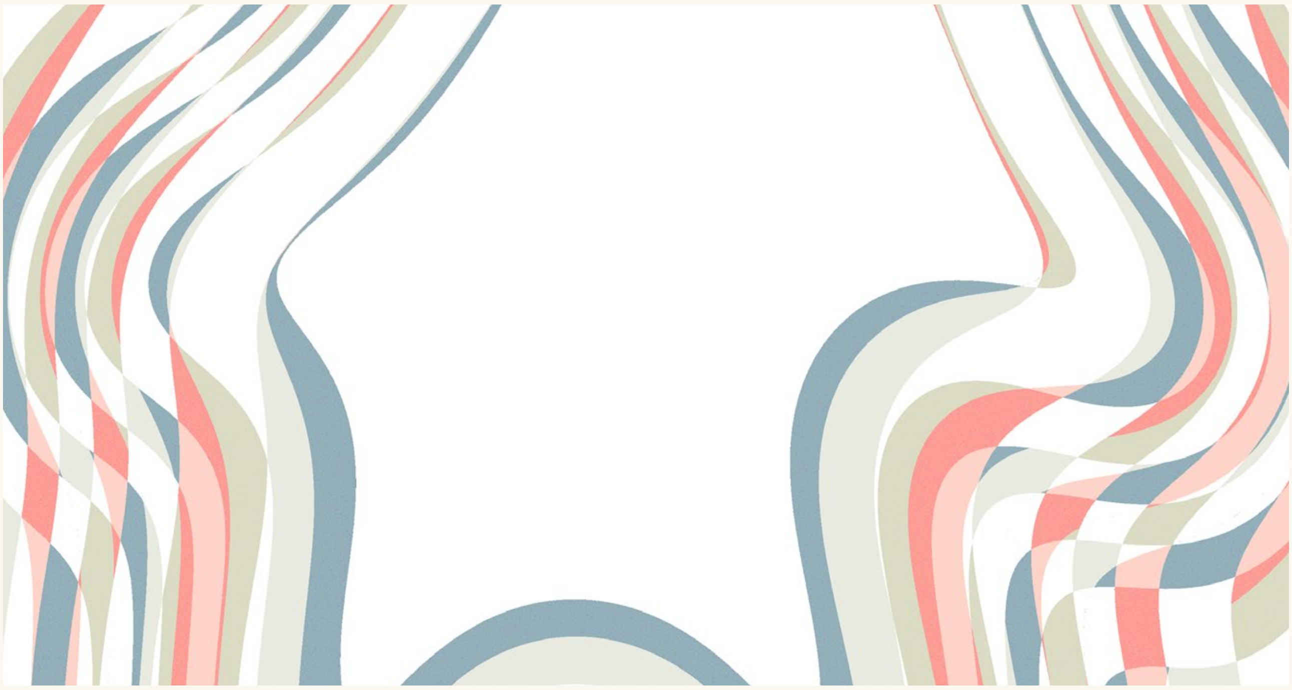
Risk 3 - Flawed algorithms based on poor data amplifying bias

Another risk associated with the use of AI is that poor-quality data or flawed algorithms can undermine the relevance, reliability, and integrity of AI-generated outputs. This risk is compounded by the opacity of many AI models, which makes it difficult to monitor decision-making processes and outcomes.

- In the financial context, training data that reflects or amplifies existing biases may lead to discriminatory outcomes, particularly in areas such as creditworthiness assessments and loan terms. Thus, the IMF expresses concern about the fact that AI adoption could amplify cross-border spillovers and capital flow volatility, particularly affecting emerging markets and developing economies.[108]
- Low-quality or incomplete input data can reduce the effectiveness of AI systems, increasing the likelihood of suboptimal or misguided investment decisions.[109]
- ESG-related algorithms may be poorly designed to account for ethical considerations, inclusion, or social impacts, resulting in outputs that inadequately reflect sustainability objectives and potentially undermine the credibility and integrity of sustainability reporting.[110]

Risk 4 - Privacy and data security

- Privacy and data security remain central concerns in the deployment of AI systems. In many cases, this data is processed or stored by third-party entities that may not operate under standardized or robust privacy frameworks. In addition, the increasing digitalization of systems expands the potential attack surface for cyber threats, creating vulnerabilities that can be exploited and further compromising data security.[111]



7.3 EXPERT INSIGHTS

Excerpts from an interview with Dr. Sep Pashang

Sep Pashang is a scholar whose research investigates the intersection of AI and sustainable finance. He holds a doctorate in Sustainability Management from the University of Waterloo. In addition to his academic portfolio, he has extensive professional experience in the financial sector, where he currently works as a Vice President in the financial technology sector.

1. How can AI be applied when it comes to ESG?

Compared to traditional ESG, which provides annual or semi annual insights, you can have insights from last week or yesterday. If you're able to do that as an investor or as a company, you're going to be making more money. Companies are always looking to enhance their tool belt and investors are too. If AI for ESG is one way of doing it, I think they're going to do it. Do they really care about e.g. the environment? Some of them may but I think they just want to make money at the end of the day — and there's a fiduciary responsibility for that.

During my research, I was confronted with the questions: Do these ESG activities actually create outcomes that are useful, tangible, and sustainable? What's the motivation of these investors as people? Are they driven by the outcomes, or do they just want to participate in an investment trend that is beneficial? I became a critic of the ESG landscape. Of course, I think it's complicated, but ESG frameworks are also very flawed and investors' intentions are not as pure as one would hope. When you have a flawed framework or model, you really have to fix it in order for the outcomes to deliver. From a systems perspective, there's a problem there.

2. Are there opportunities for AI in finance to structurally embed sustainability considerations into financial decision-making rather than treating them as add-on ESG considerations?

I would say so. Part of what can contribute to that solution is individuals or teams, who are sustainability minded or oriented, who develop their own algorithms and come up with their own solutions, as opposed to the big players that maybe have their own biases or black box algorithm models. I would say with the more intentional or smaller tools, I think there's more flexibility to customize.

One of the future research questions that I posed in my research was, while I use AI to examine conventional ESG, I think there should be future research examining these AI-ESG companies. Who is assessing them? One of my policy papers argued we need to enforce ESG standards on these AI companies that are "doing good" because they themselves, either knowingly or unknowingly, are going to contribute negatively. There should always be this accountability actor who is keeping that healthy tension.

3. Can AI alter the way in which capital is allocated across countries or markets (e.g, between Global North and Global South)?

First, a more optimistic response. A lot of my research was on AI from a societal perspective, including things like remittances and providing digital tools for people in the Global South. I think there are great use cases for that which allow people to be a bit more self sustainable. There are of course problems with this as well; but, on a smaller scale, maybe AI is providing individuals or towns with some opportunities.

Secondly, a less optimistic response. I think the hyper-usage in the Global North may steal usage and capital from the Global South. These Global North industries may get bigger and bigger and it may take away from the global container. You also have governments who are heavily using AI for their own benefit. For smaller actors, it's just difficult to compete with governments and large firms, who are just so big.

> **4. In your view, what are three positive changes that you're perceiving in the financial industry because of the advent and diffusion of the usage of AI? What do we need to make sure those positive changes will still be there and not be overwhelmed by other transformations?**

The first thing that came to mind was access. It can provide access at a societal level. There may be some implications there, but it can provide people who typically wouldn't have access to financial tools or services with access.

Number two, it probably enhances creativity, although that's double edged. I think it probably diminishes one's own creativity, but it opens up for being creative — maybe less thoughtful creativity.

Lastly, in industry, AI may enable you to go beyond what you could have ever imagined doing without it. That could be regarding risk reduction or profit generation or whatever else your scope is — AI can enhance that. In many sectors, risk reduction is key.


ハルタの靴

8. DEMOCRACY AND SOCIETAL RESILIENCE

The last few years have seen remarkable progress in the power and potential of AI technology to transform how citizens and their government interact. AI is increasingly shaping democratic societies by transforming how information is produced, distributed, and used. Examining AI's impact on democracy and social stability is therefore essential to understanding its long-term societal consequences.

8.1 OPPORTUNITIES

AI offers transformative opportunities for society, from making complex policy content accessible to more citizens, to keeping essential socio-economic systems running during crises, all of which contribute to societal resilience.

Opportunity 1 - More inclusive civic participation

AI can act as moderator, mediator, and consensus builder.[112] It can reduce the costs of participation by translating, summarizing, and structuring complex policy content, allowing more citizens to engage meaningfully at scale. When deployed transparently by governments, the deployment of AI expands deliberation without overwhelming institutions.

- AI tools can improve citizen deliberation and participatory governance, especially in local and civil-society–led initiatives.[113]
- AI can summarize diverse opinions, generate statements reflecting the range of perspectives, and help groups approach shared positions. AI can be applied in citizens' assemblies, social media platforms, and conflict-resolution processes to reconcile differing viewpoints and generate consensus. LLMs can thus potentially promote tolerance and respect in the public sphere to facilitate more epistemically productive dialogue.[114]
- Governments are increasingly using AI to analyze citizen input, moderate and translate online participation, personalize public communication, and detect patterns in feedback.[115]
- Through automated fact-checking, LLMs can support more informed public discussions, evaluating the veracity of claims on news sites, social media, and other digital platforms, while reducing the labor required for human verification.[116]

Opportunity 2 - More effective, transparent, integrated public administration

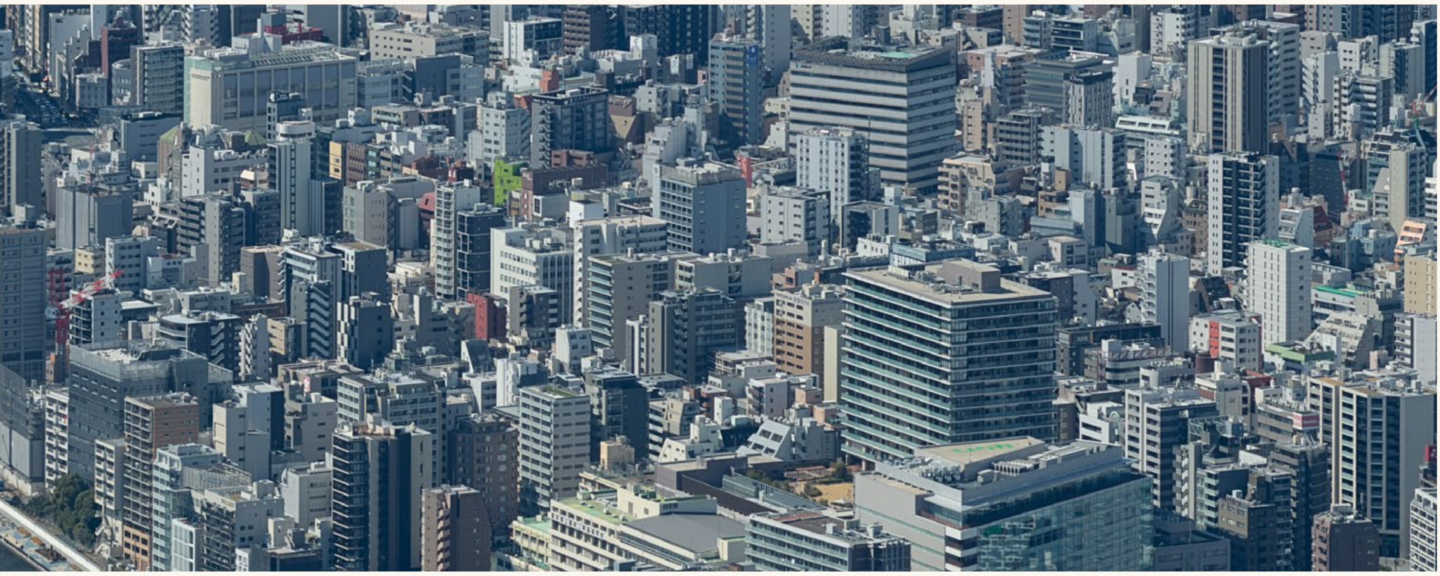
AI can strengthen state capacity and reliability:

- AI adoption in public administration may increase efficiency, responsiveness, and public value when governance safeguards exist.[117]
- AI can connect scientific insights to public policy by validating complex hypotheses, modeling outcomes, and incorporating citizen science to tackle global challenges.[118]

Opportunity 3 - Stronger societal resilience to crises and disasters

AI strengthens societal resilience by enabling anticipatory governance, stabilizing the information environment during crises, and preserving continuity of essential socio-economic systems. When embedded in accountable institutions, AI enhances a society's capacity to absorb shocks and sustain coordination and collective action under stress.

- AI is useful for essential elements of foresight such as trend analysis, future scenario development and identification of emerging themes and issues. AI-enabled strategic foresight integrates real-time signals across policy domains, allowing governments to anticipate systemic risks earlier and coordinate responses more coherently, strengthening public confidence during crises.[119]
- AI improves societal resilience by sustaining continuity in critical systems such as supply chains, healthcare, and public services. It achieves this through real-time coordination, predictive analytics, and quicker institutional decision-making under stress. More immediate response time can prevent local disruptions from cascading into systemic breakdowns.[120]



8.2 RISKS

Significant risks mirror the enormous opportunities of AI. These risks arise not only from existing applications but also from the broader trajectory of AI development, particularly considering the potential emergence of AGI and ASI.

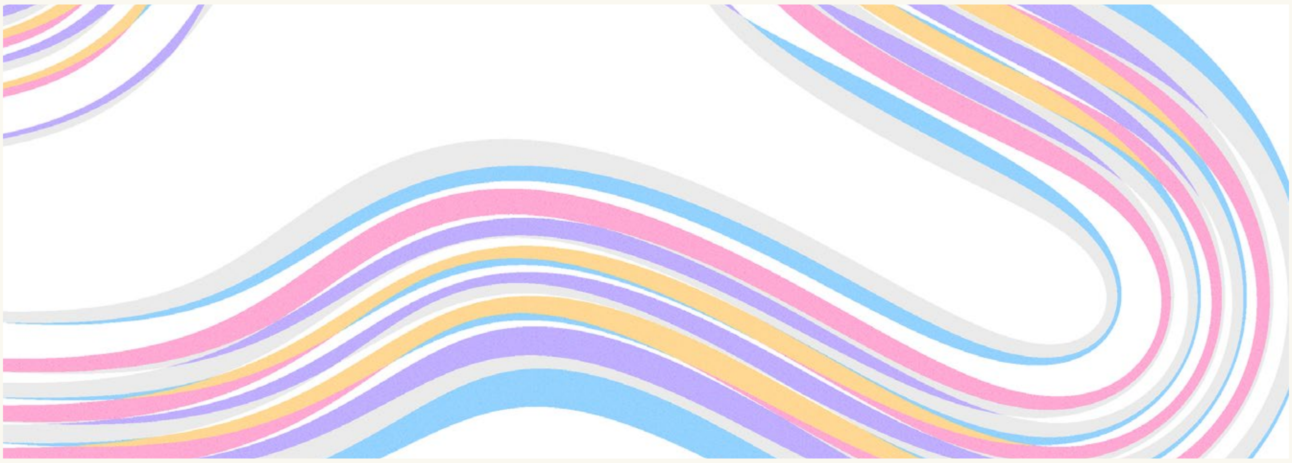
Risk 1 - Disinformation, trust erosion, and polarization

The rise of generative AI has introduced new challenges regarding the accuracy, authenticity, and integrity of information.[121] AI intensifies information disorder by simultaneously injecting false content into digital spaces, undermining trust in authentic evidence, and amplifying polarization, creating reinforcing dynamics that weaken democratic legitimacy.

- AI lowers the barrier to political manipulation (deepfakes, synthetic audio or video), while increasing its realism, overwhelming verification systems and destabilizing the shared factual basis required for legitimate elections.[122]
- As awareness of AI manipulation spreads, political actors can plausibly deny authentic evidence as fake, weakening journalism, legal scrutiny, and democratic accountability mechanisms.[123]
- Widespread mis- and disinformation can be exploited by malicious actors to influence and manipulate public opinion, distorting the democratic debate shaping electoral processes.[124]
- Engagement-optimized AI systems systematically prioritize emotionally charged and hostile content, accelerating affective polarization and shifting democratic competition toward zero-sum conflict rather than deliberation.[125]

Risk 2 - Inequality and wealth concentration

- AI risks exacerbating global inequality, as high development costs and the concentration of computing power in wealthy countries widen the technological divide and entrench economic dependence in low- and middle-income countries.[126]
- The consolidation of AI capabilities within a handful of dominant technology firms threatens democratic foundations by concentrating economic power, data ownership, and political influence. This enables firms to steer regulation and market structures toward private advantage, often at the expense of public goods, including environmental protection and social equity.[127]
- High levels of market concentration among a small number of AI providers and agent models introduce systemic risks, creating "single points of failure" in which flaws or disruptions in one system could potentially cascade across AI-dependent sectors such as finance, healthcare, and critical infrastructure.[128]



Risk 3 - Existential threats, from increased risk of war to loss of human control, as compounded by growing model opacity

- AI may be weaponized for automated surveillance, cyber espionage, or large-scale attacks. AI significantly lower the barrier to creating sophisticated, high-tech weaponry, potentially allowing dangerous capabilities to be harnessed by malicious actors. The outcomes of such scenarios include worsening geopolitical instability and increased risk of war, especially if security safeguards lag behind AI development.[129]
- AI-based lethal weapons pose acute security risks by enabling systems that can select and engage targets without meaningful human control, undermining accountability and compliance with international humanitarian law. AI-enabled autonomous weapons, especially drones and swarms, also accelerate escalation and proliferation by operating at machine speed and lower costs, increasing the risk of unintended conflict and civilian harm before governance frameworks can respond.[130]
- Beyond misuse risks, more systemic concerns relate to the loss of control associated with AI developing capabilities that match or exceed human intelligence.[131] Superhuman intelligence would possess abilities such as autonomous planning, persuasion and advanced programming that may cause AI to operate outside the scope of human oversight. Even before getting to ASI, AGI might pursue instrumental goals such as resource acquisition—i.e. AI trying to secure, and control resources, such as computational power and data to increase its ability to achieve its goals—or self-preservation—i.e. AI acting to prevent its own shutdown or alteration by deceiving or exploiting humans.[132] The combination of high intelligence and autonomous decision-making creates a landscape of highly unpredictable outcomes.[133]
- These risks are compounded by growing model opacity. As AI systems become more advanced, they are also becoming less transparent. Many modern models already operate as 'black boxes' whose internal processes are difficult for even their creators to decipher. Indeed, since the rapid advancements in deep learning from the past few years are more the result of trial and error than rigorous understanding,[134] scientists do not know how the AI models produce what they do or why they do so.[135] While we can understand AI inputs and outputs, we are moving further away from understanding how AI learns and makes decisions. This raises doubts about the human ability to predict the capabilities of the next generation of AI systems and to control increasingly powerful models. Concerns about transparency, reliability and responsible development are even more salient when it comes to ASI, whose ability to exceed human performance at most cognitive tasks may disempower humanity and lead to catastrophic harm due to malicious or incautious use.[136]

8.3 EXPERT INSIGHTS

Excerpts from an interview with Dr. Tshilidzi Marwala

Dr. Tshilidzi Marwala is the Rector of the United Nations University and Under-Secretary-General of the United Nations. He has extensive experience in academia and policy, having served on a variety of global and national policymaking bodies, and has worked with such United Nations entities as UNESCO, UNICEF, WHO and WIPO.

1. You bring the distinctive perspective of being both a UN leader and a researcher in artificial intelligence. In your view, how can the ongoing evolution and widespread adoption of AI contribute positively to sustainable development?

In the short term, the contribution of AI to sustainable development is already visible. AI is strengthening our capacity to analyze complex data, improve forecasting, and support more timely and informed decision-making. Within the United Nations system, AI-enabled tools are enhancing climate modelling, disaster risk reduction, early warning systems, and the delivery of humanitarian and social protection programmes. These applications improve efficiency and precision, allowing scarce resources to be deployed more effectively and in support of the most vulnerable.

Over the medium to long term, the implications of AI are more structural. AI has the potential to transform production systems, labour markets, and models of governance. If aligned with sustainability objectives, AI can help strengthen resilience and support more inclusive and sustainable patterns of development. These benefits, however, are not automatic. Without appropriate governance frameworks, AI risks exacerbating existing inequalities and widening digital divides. Ensuring that AI contributes positively to sustainable development requires sustained investment in inclusive access, strong human oversight, and international cooperation on norms and standards. The central challenge before us is not whether AI will shape the future, but whether we will guide its development in a way that advances sustainability, equity, and shared human well-being.

2. Among the risks associated with the rapid advancement and diffusion of AI, which do you believe are the most underestimated?

Discussions around the risks of artificial intelligence often focus on those that are most immediate and visible, such as energy consumption, labour displacement, and the spread of misinformation. These concerns are serious and warrant sustained attention. However, some of the most underestimated risks lie in the less visible, systemic effects of AI.

One such risk is the gradual erosion of human agency in decision-making. As AI systems are increasingly relied upon to inform or automate complex choices, there is a tendency to defer judgment to technical outputs without sufficient scrutiny. When this occurs in areas such as public policy, or resource allocation, it can weaken accountability and obscure responsibility. Over time, this could undermine trust in institutions and diminish the role of human values in shaping societal outcomes. Another underestimated risk concerns the concentration of power. Advanced AI capabilities are largely controlled by a small number of actors with access to vast data resources, computing infrastructure, and technical expertise. This concentration can deepen global and domestic inequalities, limiting the ability of many countries, particularly in the Global South, to shape how AI systems are designed and deployed. The result is not only a digital divide, but a governance divide that has long-term implications for sustainable development.

There is also a tendency to treat misinformation as a discrete technical problem, rather than a broader societal challenge. AI-driven misinformation does not merely distort information environments; it can polarize societies, weaken social trust, and erode the foundations for collective action on sustainability challenges. In this sense, its impact extends well beyond media ecosystems into the core of democratic and social systems.

These risks are underestimated because they accumulate gradually and are difficult to measure, yet their consequences are enduring. Addressing them requires more than technical solutions. It calls for robust governance frameworks and international cooperation that places sustainability at the centre of AI development and deployment.

3. The deepening and broadening use of AI is influencing not only advanced economies in the Global North but a wide range of countries worldwide. How do you anticipate AI affecting countries in the Global South? Which aspects do you believe deserve particular attention from the international community?

The spread of artificial intelligence is already shaping development trajectories across the Global South, and its effects will become more pronounced in the years ahead.

In many developing countries, AI is unlikely to displace large numbers of manual or informal jobs directly. However, its indirect effects on labour markets are significant and often underestimated. In urban centres across Africa, routine cognitive roles such as clerical work, basic accounting, and customer support have served as important entry points for educated young people. These are precisely the tasks most exposed to automation. If these pathways disappear without alternatives, the transition from education to stable employment becomes more fragile, particularly for youth.

At the same time, AI holds real potential to raise productivity and expand access to essential services. In public administration, AI can support translation, data management, and service delivery. In education, adaptive learning tools can help address gaps in teacher capacity. In healthcare, AI-assisted diagnostics can strengthen the work of community health workers in remote areas. These gains, however, depend on foundational enablers, including digital infrastructure, reliable energy access, language inclusion, and human capability. Ensuring that AI systems reflect local languages and contexts is therefore essential.

South Africa's experience illustrates both the promise and the risk of AI adoption. While unemployment remains high, there is growing recognition that disengaging from AI would lead to technological marginalization. This understanding has driven efforts to build national capacity and to bring government, academia, and industry together in shaping responsible approaches to AI. The lesson is that AI must be actively shaped; it cannot be left to market forces alone.

Several issues deserve particular attention from the international community. First, investment in human capital is essential. Education systems must evolve beyond rote learning and equip people with skills that AI complements rather than replaces, including critical thinking, creativity, and judgment. Second, governance matters. Without clear frameworks for accountability, data ownership, and fairness, AI risks reinforcing inequality and concentrating power. Third, AI must be connected to real economies and development priorities. Technology alone cannot overcome structural constraints or weak labour markets. AI can support reform, but it cannot substitute for it.

For the international community, the central challenge is to ensure that AI supports inclusive and sustainable development rather than reproducing older patterns of dependency and extraction. This requires coordinated investment in infrastructure, skills, and public institutions, alongside international cooperation on ethical standards and responsible governance. With the right choices, AI can become a catalyst for opportunity across the Global South.

4. AI governance is inherently multi-layered. While actions by governments and international institutions are essential, large corporations also hold significant responsibilities because of their role in AI development. What roles and responsibilities should major private-sector actors assume within the broader AI governance ecosystem?

Private-sector actors have a responsibility to embed principles of responsible AI into their core operations. This includes transparency in model development, attention to bias and fairness, respect for privacy, and the maintenance of meaningful human oversight. These principles should not be treated as optional or external to business strategy, but as integral to long-term sustainability and trust. In addition, companies must take seriously the environmental footprint of AI. Decisions related to semiconductor production, data-center construction, and energy use have direct implications for climate and water systems.

Private-sector actors should also engage constructively with public institutions and international processes. Effective AI governance cannot be achieved through unilateral action. Companies possess technical expertise and operational knowledge that are valuable to policymakers, while public institutions provide legitimacy, accountability, and a global perspective. Dialogue and cooperation between these actors are therefore essential.

Finally, there is a responsibility to contribute to inclusive outcomes. This includes supporting skills development, enabling broader access to AI technologies, and avoiding practices that concentrate benefits among a narrow set of actors or regions. Such efforts help ensure that AI supports shared prosperity rather than deepening existing divides.

5. AI's rapid development and diffusion are already reshaping higher education and employment. The implications for younger generations are particularly substantial. Do you have a message for young people who will learn, work, and ultimately take part in building society alongside AI in the years to come?

Artificial intelligence will shape the world in which young people learn, work, and lead, but it will not define their value or replace their responsibility. AI will continue to evolve, and many tasks will change, but the qualities that matter most will remain deeply human.

My message to young people is to engage with AI critically and confidently. Build digital literacy and technical understanding, but do not neglect skills that AI cannot replicate, such as judgment, creativity, empathy, and ethical reasoning. These capacities will become more important, not less.

At the same time, young people should see themselves not only as users of technology, but as active participants in shaping how it is governed and applied. The choices made today about education, work, and values will influence whether AI strengthens inclusion and sustainability or deepens existing divides. The future of AI, and of society alongside it, will depend on informed, responsible, and engaged leadership from the next generation.

9. SUMMARY OF OPPORTUNITIES AND RISKS

Artificial intelligence is rapidly reshaping how societies manage the environment, energy, work, finance, and democratic systems. AI can support sustainability transitions only if it is deliberately governed and aligned with public goals. Without guidance, it risks worsening environmental damage, inequality, and systemic instability.

Planetary Environment

Opportunities: AI enhances planetary observation by integrating data from satellites, sensors, and citizen science, improving monitoring of ecosystems, biodiversity, pollution, and climate risks. It strengthens early-warning systems for floods, droughts, and wildfires, enabling more proactive planning. By translating complex data into accessible insights, AI can also broaden public participation in environmental decisions.
Risks: AI's infrastructure consumes large amounts of energy, water, and critical minerals, generating emissions and electronic waste. AI-enabled biotechnology raises biosafety concerns, while opaque models, biased data, and environmental misinformation can undermine conservation and trust in science.

Industry and Labor

Opportunities: AI boosts productivity and innovation by accelerating research, improving quality and safety, and supporting decision-making through tools such as digital twins. It also creates new roles in system design, oversight, and governance.
Risks: Automation threatens routine jobs, including many entry-level and white-collar roles, while benefits concentrate among high-skilled workers and large firms. Algorithmic management can reduce worker autonomy and well-being, and heavy reliance on AI-generated outputs may narrow diversity and creativity.

Energy Systems

Opportunities: AI supports cleaner energy systems by improving demand and renewable forecasting, integrating wind, solar, storage, and electric vehicles into grids, and enabling smarter demand management. It also accelerates innovation in batteries, materials, and energy technologies, lowering costs and improving performance.
Risks: AI can reinforce fossil-fuel dependence by increasing the efficiency of extraction and power generation. Rising data-center demand increases electricity use, often supplied by fossil-based backup power. Digital divides may deepen energy inequality, while digitalization exposes energy infrastructure to cybersecurity risks. Poorly planned, AI can strain energy systems.

Finance

Opportunities: AI increases financial efficiency through automated analysis, portfolio optimization, improved risk assessment, and streamlined ESG reporting. It can also generate data for emerging economies, potentially accelerating development finance.
Risks: Efficiency gains do not ensure sustainability or equity. Benefits often remain disconnected from real-world outcomes, while AI introduces risks such as market concentration, correlated investment strategies, biased algorithms, weak transparency, and fragile AI-dependent infrastructure. AI can make finance faster, but not necessarily fairer or greener.

Democracy and Societal Resilience

Opportunities: AI can strengthen democratic institutions by lowering barriers to participation through translation, summarization, and large-scale consultation. It can improve pub-

lic-sector efficiency and support crisis anticipation and response across sectors.
Risks: Generative AI amplifies misinformation, erodes trust in evidence, and fuels polarization. Control over AI is increasingly concentrated among a few firms and countries, raising concerns about inequality and political influence. More severe risks include AI-enabled weapons, geopolitical instability, and opaque models that limit oversight and accountability.

Table 3. Summary of sector-specific opportunities and risks and cross-cutting themes

Domain	Key Opportunities	Key Risks	Cross-Cutting Themes
Planetary Environment	Environmental monitoring, scenario planning, resource management, and democratization of environmental knowledge	Environmental impacts, biosafety risks, malicious use of AI, data inequities, lack of transparency	Data integrity and quality, transparency, human oversight
Energy Systems	Efficiency improvements, renewable integration, grid modernization, knowledge acceleration, modernization of demand-side market	Fossil fuel lock-in, infrastructure stress, cybersecurity, data limitations and lack of transparency	Data integrity and quality, transparency, cybersecurity, human oversight
Industry & Labor	Innovation across the economy and the labor market, workplace efficiency, enhanced creativity	Increased digital divide, market concentration, workforce policy misalignment, labor displacement, workplace alienation, erosion of collective diversity	Transparency, human oversight
Finance	Portfolio optimization, alignment of investments with financial products, improved emerging market data	Unsupervised AI use, financial instability, algorithmic bias, data security	Data integrity and quality, cybersecurity, human oversight
Democracy & Societal Resilience	Enhanced deliberation, collective decision-making, reliable public information, accelerated scientific discovery	Inequality and wealth concentration, misinformation and manipulation, security, cybersecurity, loss of control	Data integrity and quality, cybersecurity, human oversight

PART III
AI for Humanity

In just a decade, AI has moved from the fringes of tech labs to the center of global transformation. Its potential to contribute positively to addressing the sustainability challenges of our time is immense. As discussed in the previous section, AI can serve as a powerful engine for scientific discovery and innovation: it can boost productivity, enable new ways of modeling complex systems, and strengthen our ability to predict and prepare for the challenges that will arise as a result of climate change and other crises.

For this promise of a more sustainable future to be realized, AI development trajectories must be deliberately shaped around these parameters. This requires, first, a collective societal decision to steer AI development toward pathways that benefit the many rather than the few. Second, it requires the design and implementation of effective governance frameworks to ensure that this intentional direction is not merely aspirational but is translated into practice.

While Part II analyzes opportunities and risks of AI across the sectoral five domains, Part III shifts the focus to a structured governance framework operating across three interlocking layers.

First, sectoral governance addresses how AI is deployed within specific domains—planetary environment, energy systems, industry and labor, finance, and democracy and societal resilience. These measures seek to optimize opportunities and manage sector-specific risks at the local level.

Second, cross-cutting safeguards establish horizontal conditions, such as transparency and human oversight, which are applicable across sectors. These mechanisms ensure that AI deployment remains safe, accountable, and socially legitimate.

Third, global AI governance operates at the level of AI capability development. As AI systems progress from narrow applications toward general-purpose and potentially superhuman intelligence, risks become systemic, transboundary, and potentially irreversible. Governing AI deployment is insufficient if the underlying trajectory of capability escalation remains unconstrained. Global governance therefore addresses how far, how fast, and under what conditions frontier AI systems are developed. AI governance addresses core policy questions: whether societies can reach shared agreements on the direction of AI development, and what institutional and normative conditions are necessary to ensure that AI advances the public interest and delivers broadly shared societal benefits.

The sections that follow outline both the domestic governance architecture required to manage AI deployment and a phased roadmap toward a durable global framework capable of guiding advanced AI development according to shared international guardrails.


ニトリ
NITORI
ニトリ NITORI

10. SECTORAL AND CROSS-CUTTING PERSPECTIVES

The analysis presented in Part II demonstrates that AI introduces both remarkable opportunities and significant risks across all key domains. These impacts arise primarily at the level of deployment: how AI systems are integrated into existing economic, environmental, and political structures.

Addressing these challenges requires that sectoral governance be tailored to the risk profiles of each domain. In each case, regulatory responses must be context-specific in order to mitigate harms while enabling innovation that advances the sustainability transitions.

At the same time, certain risks transcend sectoral boundaries. Issues such as opacity in model design, data inaccuracies, cybersecurity vulnerabilities, algorithmic bias, and insufficient human oversight are not confined to any single domain. These shared challenges require cross-cutting safeguards, namely governance instruments that apply across sectors and establish baseline conditions of transparency, accountability, and resilience for AI deployment as a whole.

Together, sector-specific measures and cross-cutting safeguards form the domestic governance foundation for responsible AI use. However, as subsequent sections will argue, governing deployment alone is insufficient if the broader trajectory of advanced AI capability development remains unaddressed.

10.1 SECTORAL RISK MANAGEMENT MEASURES

10.1.1 - Planetary Environment

Table 4. Summary of opportunities and risks in the planetary environment domain

Opportunities	Risks
Improved environmental monitoring Scenario planning for ecosystems and climate Better informed and optimized natural resource management Democratization of environmental knowledge	Area specific: Environmental impacts of AI underpinning infrastructure Biosafety risks Malicious use of AI Cross-cutting*: Poor data quality, including misinformation and disinformation* Lack of transparency*

Data centers—the infrastructure underpinning AI—have direct environmental impacts that can be minimized through rigorous planning and assessment mechanisms to quantify impacts on biodiversity, climate, and communities.[137] Although the impacts of data centers on the environment greatly vary depending on the context, policies should:

- Require data center developers to adopt a holistic approach to optimize design and operational performance. Because data centers impact water, biodiversity, and climate, trade-offs are often unavoidable. These impacts should be addressed early on in planning, and integrated into design and operational strategies.[138]
- Condition development approval on minimizing environmental footprint, ensuring zero competition with other water uses and zero water discharge, including through closed-loop systems and non-potable sourcing. Conditions should also require additional renewable energy generation and flexible demand that facilitates grid integration, rather than reliance on existing clean supply (as also discussed below).[139]
- Require data centers developers to prioritize locations with lower ecological value and adjust design layouts to preserve priority habitats. Where impacts remain, restoration measures—such as replanting native species or rehabilitating ecosystems—should be implemented.[140]

To address concerns connected to biosafety risks, policies should require robust assessment mechanisms to evaluate AI–synthetic biology integration, including impacts on biodiversity, sustainable use, equitable benefit-sharing, and traditional knowledge.[141] Finally, it is important that significant steps are taken to protect data and to ensure that tools are used only for their intended purpose to guard against the malicious use of AI.

To seize the opportunities for the planetary environment, it will be important to draw the research findings from the UNU Institute on Big Data and Artificial Intelligence for Managing Human Habitat Change (UNU-AI),[142] which focuses on leveraging AI, big data, and high-performance computing to address complex global issues, particularly those related to human habitat changes caused by environmental crises. These research insights should be integrated in national policy making as well as in North-South collaboration on combatting environmental multi-crises.

10.1.2 - Energy Systems

Table 5. Summary of opportunities and risks in the Energy Systems domain

Opportunities	Risks
Improved energy efficiency Integration of renewable energy and grid modernization Acceleration of knowledge production Modernization of demand-side market	Sector specific: Fossil fuel lock-in and transition delays Physical infrastructure and competences stress Cross-cutting*: Cybersecurity* Poor data quality and lack of transparency*

AI can accelerate decarbonization if guided by appropriate policies that reduce risk and unlock its full potential. Governments should:

- Advance policy that accelerates AI-driven energy innovation by increasing innovation funding and improving the availability of data, models, and computing infrastructure.[143]
- Employ regulatory sandboxes to safely test low-risk AI applications (e.g., renewable forecasting, dynamic line rating) and encourage pilot projects that allow experimentation while maintaining safeguards for critical infrastructure.[144]
- Increase public R&D investments in green-AI hardware and software, and promotion of open-source AI models to reduce resource-intensive training. Dedicate national compute resources for green-AI research.[145]
- In countries experiencing a data-center boom, impose conditionality on licensing and grid connection to facilitate renewable energy integration, either by increasing grid flexibility or by requiring data centers to source electricity directly from additional renewable capacity.[146]

10.1.3 - Industry and Labor

Table 6. Summary of opportunities and risks in the Industry and Labor domain

Opportunities	Risks
Increased innovation across the economy and the labor market Efficiency gains in the workplace Enhanced creativity	Sector specific: Wider digital and labor divide Market concentration Workforce policy misalignment Labor displacement Workplace alienation Erosion of collective diversity Cross-cutting *: Lack of transparency * Insufficient human oversight *

Managing AI's impact on work requires coordinated labor and education policies that strengthen institutional adaptability while guiding the direction of technological adoption. The objective is not only to equip workers with relevant skills, but to ensure that AI deployment supports inclusive growth rather than deepening displacement and inequality. Policymakers should consider advancing the following initiatives:

- Invest in foundational capabilities by expanding access to education, STEM training, and digital literacy. Increasing targeted reskilling as clearer signals emerge about evolving task demands.[147]
- Strengthen institutional adaptability by enhancing the capacity of labor market and educational institutions to respond flexibly to technological change, while designing redistribution mechanisms that cushion transitional shocks.[148]
- Promote human–AI complementarity. Prioritize retraining into roles where AI augments productivity rather than substitutes for labor and align incentives—such as tax or procurement policies—to encourage labor-augmenting adoption.[149] Proactively build holistic skill systems through curricula that integrate technical AI competencies with critical thinking, social intelligence, and other non-automatable skills.
- Support international transition mechanisms. Consider instruments such as an AI Adjustment Fund to finance retraining, institutional reform, and social protection in developing economies during periods of structural adjustment.[150]

10.1.4 - Finance

Table 7. Summary of opportunities and risks in the Finance domain

Opportunities	Risks
Portfolio optimization and improved data collection Better alignment of investor preferences with financial products Improved data availability for emerging markets	Sector specific: Financial instability Cross-cutting *: Unsupervised use of AI * Algorithmic bias * Privacy and data security *

AI can create major financial instability. To safeguard the financial system, technical and regulatory oversight will have to evolve alongside AI's rapid integration into the financial sector. While regulators already conduct risk reporting and monitor large market participants, these measures remain sector-specific and limited in scope. Existing frameworks capture traditional operational and financial risks for banks and registered entities, but they rarely cover full dependencies on AI systems, data, and critical technological infrastructure. Moreover, the monitoring of non-bank or highly automated participants is often incomplete. In this context, the following strategies are important to mitigate the risks that undermine financial stability:[151]

- Comprehensive risk mapping and enhanced data collection for all large traders in order to systematically identify and manage emerging AI-system dependencies and ensure regulators can anticipate and mitigate risks in increasingly AI-driven markets.
- Deploy stress testing of extreme scenarios to help assess the resilience of financial markets to AI-driven disruptions, such as rapid algorithmic price swings, misaligned AI risk models, critical infrastructure outages, data manipulation, sudden adoption shocks, and cross-market contagion. These tests can reveal vulnerabilities and ensure that institutions can absorb shocks without triggering systemic crises.
- Include circuit breakers[152] and margining mechanisms[153] to manage the risk of significant AI-driven price volatility, which can arise from high-speed, adaptive, and cross-market algorithmic strategies.
- AI can be a force for shifting capital where investment is needed. Beyond financial stability, policymakers should urgently convene regular consultations or forums with development and commercial financing institutions and AI data producers. These engagements should explore how AI-generated data can meaningfully support capital allocation towards sustainable development priorities—rather than undermine it—-by identifying new investment opportunities, improving understandings of risk-return profiles, and standardizing approaches to de-risking financial structures. These forums could:
- Use AI-enabled datasets to identify sustainable development to advance investment opportunities that are poorly captured by traditional project pipelines and assessment tools.[154]
- Improve risk–return assessment by enabling more granular, project-level analysis, reducing reliance on perceived risk and coarse country- or sector-level proxies.[155]
- Support the standardization of de-risking approaches, aligning how AI-generated data is used in guarantees, blended-finance structures, insurance, and credit enhancement instruments. Shared data standards would lower transaction costs, improve comparability, and facilitate greater participation by commercial investors, thus leveraging a key barrier to investment in EMDEs.[156]

10.1.5 - Democracy and Societal Resilience

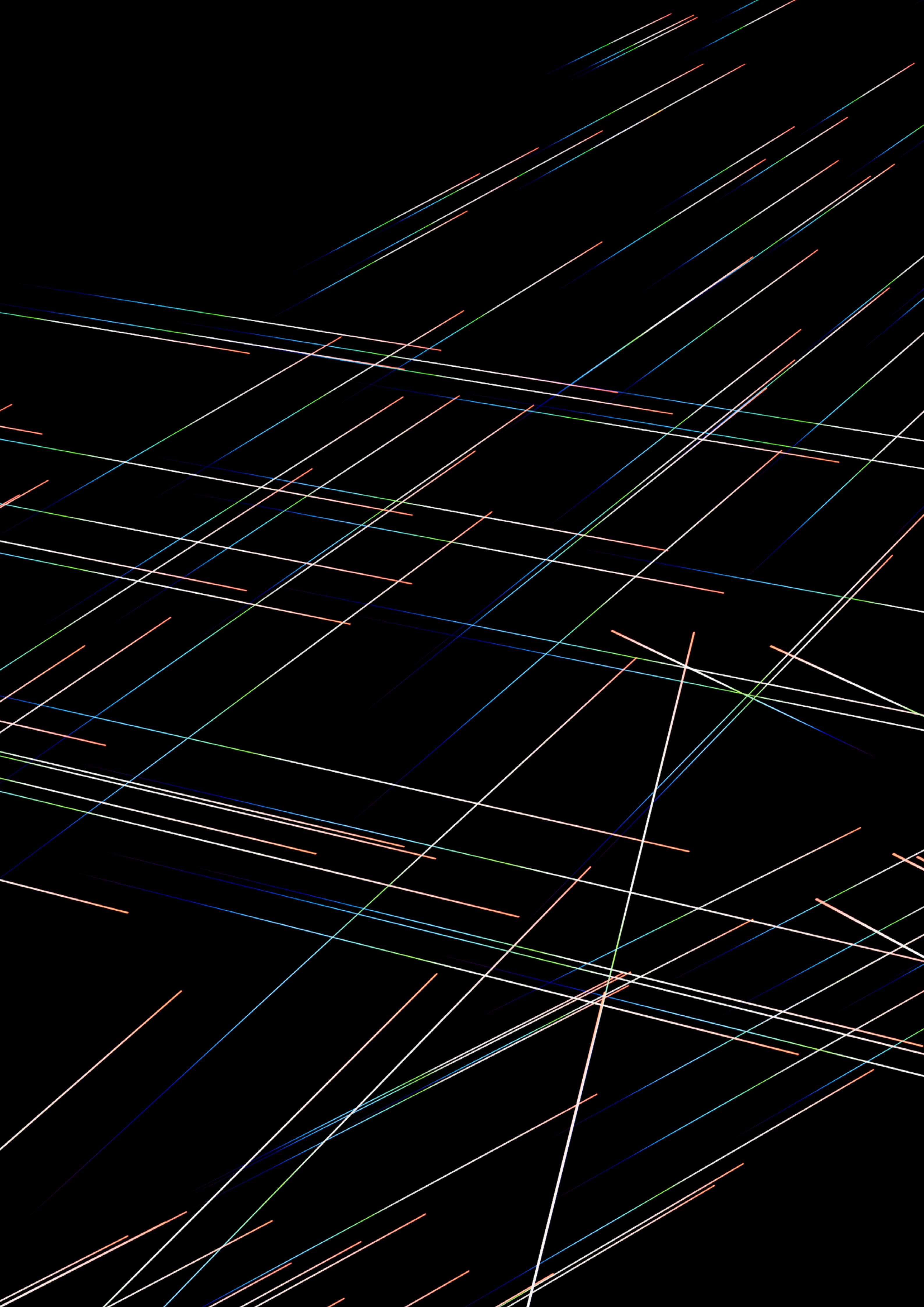
Table 8. Summary of opportunities and risks in the Democracy and Societal Resilience domain

Opportunities	Risks
Enhanced democratic deliberation Collective decision-making Anticipatory governance Preservation of the continuity of essential socio-economic systems	Sector specific: Global inequality and wealth concentration Manipulation and misinformation Geopolitical instability Surveillance Loss of control due to increasingly powerful AI systems Cross-cutting *: Cybersecurity * Increased opacity * Insufficient human oversight *

Because of the threats posed by AI, governments are increasingly tasked with protecting democratic values (e.g., protection of fundamental rights or popular sovereignty, rather than technocracy), while balancing innovation and economic growth. Although specific regulations will vary by jurisdiction and will have to evolve together with technological developments,[157] governments should consider the following approaches to address these challenges: [158]

- Develop national and regional AI regulatory frameworks that define clear rules for AI development and deployment, with a focus on democratic protection, privacy, and public safety.
- Monitor inequalities in access across income, employment, gender, health, education, and environmental dimensions, and assess options for corrective action where appropriate.[159]
- Promote digital literacy, integrate it into school curricula, and implement awareness programs about disinformation risks.[160]
- Participate in international AI governance efforts to reinforce democratic safeguards globally (see chapter "A Global Governance Framework for AI"). Such collaboration between governments at bureaucratic and political levels helps monitor policy implementation effectively. It may also mitigate geopolitical instability by fostering cooperation over competition, thereby aligning AI development and deployment with shared democratic principles.

To maximize the benefits of AI, governments at national and local levels should integrate AI technology into public service programs, as effective service delivery strengthens the legitimacy and stability of democracy.[161] Governments should leverage AI to enhance the quality, efficiency, and accessibility of policies, laws, and public services, building on existing successes in digital governance, particularly in areas such as anti-corruption and citizen access.[162]



10.2 CROSS-CUTTING RISK MANAGEMENT APPROACHES

While sectoral governance addresses domain-specific risks, certain vulnerabilities are structural and are caused by how AI systems are designed, trained, deployed, and monitored. These shared risks require horizontal safeguards that apply across all domains.

Cross-cutting governance therefore establishes baseline conditions for safe and legitimate AI deployment. Its purpose is to ensure that AI systems operating in society remain transparent, accountable, and secure. These safeguards govern AI deployment. They do not determine the pace or limits of frontier capability development—a distinct challenge to be addressed at the global level, as discussed below.

Transparency, Accuracy, and Human Oversight

Without visibility into how models are trained, evaluated, and deployed, regulators cannot meaningfully assess risk or remediate harmful outcomes. Transparency in the deployment of AI entails:

- Rigorous verification processes, including security audits, compliance checks, and standardized metrics and benchmark datasets for validating model performance.[163]
- Regular auditing to ensure AI systems remain accurate and do not reproduce biased outcomes, particularly in sensitive areas such as hiring and compensation.[164]
- "Human-in-the-loop" processes to review AI outputs, i.e. utilizing subject matter experts and data labelers. This safeguard is critical for ensuring quality and preventing the circulation of "AI hallucinations."[165]
- AI Models that respond to plain-language commands to allow operators to oversee, direct, and override systems when necessary.[166]
- Mandated transparency in AI training data and rigorous testing of generative and foundation models prior to deployment. Testing should evaluate performance, biases, alignment, and potential risks for society and democracy.[167]
- Creation of dedicated AI oversight bodies or institutes at the national or regional level to monitor deployment and ensure adherence to democratic principles.[168]

Security, Data Protection and Cyber Resilience

As AI systems become embedded in critical infrastructure, weak cybersecurity can amplify vulnerabilities across entire sectors. Data protection and cyber resilience are therefore foundational conditions for safe deployment. Cross-cutting security measures should include:

- Rigorous management protocols around who can access and manipulate high-value environmental and social datasets.[169]
- Mandatory use of encrypted technologies and continuous technological checks to defend against the misuse of AI by malignant actors.[170]
- Robust cryptographic safeguards, anomaly detection, real-time threat monitoring, and establishment of standardized benchmarks for data protection.[171]

Participatory and cross-sectoral collaboration

AI deployment must be grounded in social legitimacy. Without participatory design and equitable access, AI risks reinforcing existing power asymmetries and deepening inequality. Cross-cutting participatory mechanisms includes:

- Public funding for interdisciplinary AI research that engages AI developers, ethicists, and social and economic scientists to study AI's contributions to human well-being beyond profitability.[172]
- Participatory design processes and employment of diverse training datasets to mitigate bias and discrimination. This includes increasing efforts to collect better data, as well as encouraging data-sharing for both AI development and deployment.[173]
- Mechanisms that allow communities to retain control over their data and the resulting AI outputs, in order for the benefits of the technology to remain local.[174]
- Capacity-building initiatives—including education, digital literacy, and infrastructure investment—to ensure individuals and institutions can engage with AI systems effectively and responsibly.[175]

Table 9. Summary of sector-specific risk management measures and cross-cutting risk management approaches

<table>
<tr><th>Domain</th><th>Sector-Specific Risk Management Measures</th><th>Cross-Cutting Risk Management Approaches</th></tr>
<tr><td>Planetary Environment</td><td>- Develop assessment mechanisms to quantify AI impacts on biodiversity, climate, and communities, for example integrating KPIs based on the Planetary Boundaries framework.
- Categorize AI applications by risk levels, identifying technologies with high-risk and unacceptable-risk thresholds for the planetary environment.
- Condition data center development on minimizing environmental footprint and maximizing renewable energy use.
- Assess AI–synthetic biology integration against robust assessment mechanisms for equitable, safe use.</td><td rowspan="5">Transparency, Accuracy, and Human Oversight: Embed processes that support openness, accountability, and accuracy in AI-generated content. Ensure models are interpretable, auditable, and explainable. Implement human-in-the-loop review and plain-language command interfaces. Mandate transparency in AI training data and rigorous model testing before deployment. Create AI oversight bodies and mechanisms to monitor instances of bias.

Security, Data Protection, and Cyber Resilience: Strengthen legal frameworks and protocols for sensitive data, enforce encryption, anomaly detection, real-time monitoring, and standardized benchmarks for cybersecurity.

Inclusive, Participatory, and Cross-Sectoral Collaboration: Public funding to democratize interdisciplinary AI research, participatory design, use of diverse training data, local data control, and education and skills-building to empower communities and ensure contextually responsible AI deployment.</td></tr>
<tr><td>Energy Systems</td><td>- Increase funding for AI-driven energy innovation, particularly by improving availability of data, models, and computing infrastructure.
- Regulatory sandboxes for testing low-risk AI applications (e.g., renewable forecasting, dynamic line rating).
- Public R&D investments in green-AI hardware/software; promote open-source AI.
- Conditional licensing of data centers that facilitate renewable integration.</td></tr>
<tr><td>Industry & Labor</td><td>- Invest in education, STEM, and digital literacy.
- Strengthen institutional adaptability of labor markets and education systems.
- Design redistribution mechanisms to ease transitional disruptions.
- Prioritize human–AI complementarity through workplace design, retraining programs, and economic incentives.
- Implement holistic skills curricula combining technical AI, social, and critical thinking competencies.
- Expand digital infrastructure to promote accessibility and close the digital divide.
- Support global transition mechanisms e.g., AI Adjustment Fund that finances retraining and social protection in developing economies.</td></tr>
<tr><td>Finance</td><td>- Comprehensive risk mapping and data collection for large traders to manage AI-system dependencies.
- Stress-testing extreme AI-driven market scenarios to gauge resilience.
- Circuit breakers and margining mechanisms to manage price volatility from high-speed, adaptive trading.
- Convene forums to align AI-generated data with sustainable investment goals.</td></tr>
<tr><td>Democracy & Societal Resilience</td><td>- Develop national/regional regulatory frameworks for AI development and deployment that emphasize democratic protection, privacy, and public safety.
- Monitor demographic inequalities in AI access and implement corrective measures.
- Promote digital literacy and disinformation awareness.
- Participate in international AI governance efforts to reinforce democratic safeguards and promote cooperation.
- Integrate AI into public services programs to enhance efficiency and accessibility.</td></tr>
</table>

11. A GLOBAL GOVERNANCE FRAMEWORK FOR AI

AI governance operates across multiple, interconnected layers, each addressing a distinct dimension of risk. Sectoral governance manages how AI is deployed within specific domains. Complementing these measures, cross-cutting safeguards establish horizontal conditions for responsible deployment. These risk mitigation approaches are primarily implemented at the national or regional level, where governments adapt them to domestic institutions, legal frameworks, and socio-economic priorities. Together, they form the architecture for governing AI use within society.

Yet deployment governance, however robust, cannot address the deeper structural forces shaping the trajectory of AI development itself. Sectoral and cross-cutting measures regulate how AI systems operate once embedded in social and economic systems; they do not determine how powerful those systems become, how rapidly capabilities scale, or under what systemic constraints frontier research proceeds.

Global AI governance[176] operates at a different level, as a coordination architecture situated at the international level. Its purpose is to ensure that the development of increasingly powerful AI systems unfolds safely and in alignment with shared global interests. A global governance framework functions above individual regulatory domains, creating the institutional conditions under which national sectoral and cross-domain policy interventions can operate coherently, predictably, and collectively.

This distinction is particularly important as AI capabilities approach general-purpose and potentially autonomous systems. While domestic governance can mitigate harms arising from specific applications, the escalation of frontier capabilities introduces risks that are transboundary, systemic, and potentially irreversible. Therefore, effective risk containment at this level requires coordinated international mechanisms capable of governing AI's developmental trajectory itself.

11.1 STRUCTURAL BARRIERS

Despite the numerous initiatives aimed at establishing a global governance framework for AI—including, the UNESCO Recommendation on the Ethics of Artificial Intelligence,[177] the WEF Presidio Recommendations on Responsible Generative AI,[178] and the UN AI Advisory Board[179]—to date, none have been successful in shaping the trajectory of AI development. Calls for pausing experiments on AI have largely gone unheard,[180] and scenarios describing the existential threat posed by general AI—and more seriously, by superhuman intelligence—have failed to materially influence the direction or pace of AI development. The lack of governance frameworks and norms for accountability risks accelerating a race to the bottom, where the competitive pressure to develop AI encourages speed over safety.[181]

There are **four structural barriers** to cooperation that can explain the failure to regulate AI at global level.

1. **Corporate concentration and the erosion of public-interest AI research:** Fundamental and applied research in this area have advanced rapidly, with activity increasingly concentrated in the hands of corporate actors, whose incentives are primarily driven by profit maximization, rather than universities and public research institutions, which are traditionally oriented toward the generation of public goods and broader societal benefits. In this scenario, states are weak against corporate actors that have every incentive to develop and deploy increasingly powerful models with few guardrails in place.[182]
2. **Weakening multilateralism and the rise of competitive AI deregulation:** Mounting pressures on multilateralism, which forms the foundation of global governance and is increasingly contested,[183] make an international agreement on AI seem unlikely. As multilateralism weakens, states become less willing—and in some cases less able—to cooperate on shared regulatory approaches. Instead, they are pushed into competitive dynamics in which AI development is framed as a strategic and economic race.[184] Competing states are thus more likely to identify incentives not to regulate collectively; furthermore, they are even less likely to regulate domestically for fear

of imposing unnecessary burdens on innovation or pushing innovation elsewhere.[185]

3. **Opacity in AI systems and the challenges of ensuring safety:** Gaps in understanding of why AI models generate particular outputs complicate efforts to assess and assure safety, making it difficult to provide credible guarantees regarding AI systems' behaviors. [186] This opacity constrains the design of effective governance mechanisms, especially vis-a-vis the uncertainty of how AI develops in the future.
4. **Divergent risk perceptions and barriers to global AI governance:** The incomplete assessment of the practical consequences of advanced AI systems, combined with contested estimates regarding the timeline for the emergence of human-level or superhuman intelligence, contributes to the formulation of contradicting priorities and divergent perceptions of urgency among governments. This misalignment, in turn, hampers progress toward the establishment of a global AI governance framework.[187]

Overcoming these structural barriers requires a sequenced approach: first building shared understanding of AI capabilities and associated risks, then containing existential threats, and ultimately establishing durable institutional guardrails. This report therefore proposes a phased roadmap toward global AI governance structured around three interdependent stages, namely scientific alignment, precautionary safeguards to prevent irreversible harm, and long-term global coordination through a binding but adaptable framework.

11.2 SHORT TERM: ESTABLISHING A SHARED SCIENTIFIC BASELINE

This phase addresses epistemic fragmentation. It creates a shared scientific foundation for AI capabilities and risks, which in turn lowers the political barriers to cooperation. From this perspective, scientific alignment becomes the necessary precursor to regulatory alignment.

The first practical step is to establish a shared understanding of AI's existing risks and capabilities. This roadmap prioritizes scientific cooperation over the immediate construction of a legal framework for two reasons: first, a unified scientific foundation is a prerequisite for developing robust, evidence-based legislation; second, the current crisis of multilateralism makes a global treaty unlikely in the short-term, whereas even under these conditions, scientific collaboration can be advanced successfully.

The foundational step toward a shared scientific understanding of AI began in August 2025, when the United Nations established the Independent International Scientific Panel on Artificial Intelligence (resolution A/RES/79/325).[188] Conceived as an AI-focused analogue to the Intergovernmental Panel on Climate Change (IPCC), the Panel will provide a common scientific baseline to inform policy decisions and facilitate international coordination.

Alongside the UN Panel on AI, we suggest creating an AI Expert Technical Laboratory, whose functioning would be similar to a 'CERN for AI,'[189] which has gained significant political traction, recently entering the European Union's official policy agenda.[190] To avoid duplication and ensure responsiveness to relevant challenges, the AI Expert Technical Laboratory would coordinate closely with the UN Panel on AI to identify critical research gaps and strategic priorities. While the Panel should focus on assessing existing risks and opportunities, the AI Expert Technical Laboratory should be leveraged as a rigorous environment for testing frontier AI models and fostering innovation, while ensuring all innovation is anchored in safety and in pursuit of shared societal benefits.

The efforts of the UN Panel and AI Expert Technical Laboratory will serve as a foundation for policymakers in the medium term. Their scientific assessments of the risks that AI poses to humanity will enable policymakers to make informed decisions and to better understand the relationship between AI development and its associated risks. This, in turn, will facilitate cooperation at the policy level, which will be essential in the medium term.

The UN Panel and the AI Expert Technical Laboratory should remain operational throughout the medium- and long-term governance phases. Their scientific support to policymakers, and their capacity to advance research grounded in safety and shared societal benefits, are critical for identifying the most effective policy and regulatory options.

11.3 MEDIUM TERM: THE INTERIM AI SAFETY GOVERNANCE FRAMEWORK

The interim phase represents a shift from knowledge production to risk containment. As AI research advances toward increasingly powerful models, scientific cooperation will not be enough to ensure global safety, unless it is supported by a shared vision and clear guardrails outlining how these systems should be developed.

The interim framework acts as an international coordination mechanism, complementing domestic sectoral and cross-cutting measures, and bridging the gap between rapid technological advancement and slow institutional consensus. Its central purpose is to reduce the likelihood of irreversible harm from misaligned AI systems as capabilities approach AGI, and potentially, ASI.

A “wait-and-see” approach is untenable under conditions of profound uncertainty. Delays could narrow effective policy options, while a fully negotiated global framework may take too long to establish. Interim governance addresses near-term risks while preparing for long-term, consensus-based regulation.

The interim AI safety governance framework should be based on three pillars:

1. **Harm prevention:** The main objective of such an interim AI governance framework would be harm prevention. In practical terms, this translates into binding international restrictions—or even temporary bans—on certain categories of advanced AI research that could lead to AGI or ASI. It would draw on precedents from international regimes governing biological and chemical weapons.[191] Such measures would impose a temporary pause on advanced AI research to ensure that collective decision-making about the future direction, boundaries, and purpose of AI development precedes further technological escalation.
2. **Safety thresholds:** After pausing certain categories of AI research, the framework must establish clear thresholds to decide if and when these restrictions can be lifted.
3. **Inclusive multi-stakeholder processes:** This period creates the space needed to bring together states, industry, and civil society to translate scientific insights into durable governance arrangements.

An interim governance framework that focuses narrowly on these three specific pillars is more likely to succeed than a comprehensive framework. Indeed, although broad agreements have the potential for the greatest overall policy impact, they are extremely difficult to negotiate. In contrast, narrow agreements that address well-defined problems are often more effective in driving policy change.[192]

To operationalize these three pillars, responsibilities should be clearly allocated across scientific bodies, governments, and private actors:

- The UN AI Scientific Panel and the AI Expert Technical Laboratory would continue to play complementary epistemic roles. The AI Expert Technical Laboratory evaluates frontier AI systems, develops technical safety benchmarks, and conducts human-safety aligned research. The Panel synthesizes global knowledge, assesses systemic risks, and reviews safety thresholds to ensure alignment with societal considerations.
- National governments would translate this technical evidence into binding international policy measures. This would include negotiating and implementing a multilateral temporary restriction on training models above specified compute or capability thresholds. In parallel, governments would convene multilateral forums for coordinated dialogue (e.g., via the Global Dialogue on AI Governance, UN Resolution A/RES/79/325.)[193]

11.4 LONG TERM: THE GLOBAL AI GOVERNANCE FRAMEWORK

The final stage of the roadmap is the adoption of a framework convention on AI and the implementation of a durable global governance framework to guide technological advancement in accordance with internationally agreed guardrails. This architecture operates above the sectoral and cross-domain regulations and policy interventions, establishing the conditions under which national and regional measures remain coherent, adaptive, and enforceable.

The framework would address risks to safety, societal well-being, and environmental sustainability, since effective sustainability transitions depend on safeguards across all three domains. Ensuring safety means preserving human agency, keeping AI under human control, and establishing globally agreed thresholds. Societal well-being involves promoting fairness, human rights, and social cohesion, while environmental alignment requires minimizing ecological harm and greenhouse gas emissions.[194]

Building on these objectives, specific institutional design requirements underpin a stable and legitimate global AI governance system.

1. **Adaptive governance mechanisms and iterative learning**

Adaptability ensures that governance mechanisms remain responsive to capability shifts and do not become obsolete as AI systems evolve.[195] The AI global governance framework should therefore allow for iterative updates as new risks emerge and evidence about impacts accumulates. To bridge the gap between technical developments and policy responses, states should continue to rely on the expertise of the UN Panel on AI and the AI Expert Technical Laboratory.

2. **Adaptive governance mechanisms and iterative learning**

Structured coordination reduces fragmentation and mitigates competitive regulatory arbitrage. Global AI governance should integrate multi-level coordination among national governments, regional institutions, and global bodies. The global level will create alignment while providing the necessary flexibility for states and regions to tailor their domestic approaches to their unique jurisdictional needs.[196] For instance, countries may prioritize specific sectors, such as healthcare over labor; determine the degree to which competition policy should be utilized to dismantle market concentration; establish direct oversight on data practices; implement robust consumer protection norms; or define clear standards for manufacturer responsibility.

Global AI governance should also strengthen multi-stakeholder coordination across industry, civil society, and technical experts, to ensure legitimacy, reduce fragmentation, and strengthen consistent implementation of standards.[197]

3. **Accountability and sustainability alignment**

To ensure that compliance is not solely normative but also incentive-compatible, trade and market mechanisms could be incorporated into this global AI governance architecture. Following the example of the Montreal Protocol,[198] compliance could be reinforced through trade-related measures.[199] Applied to AI, similar mechanisms could require certification or compliance standards for cross-border transfers of AI models, data, or software. Responsible AI practice would thus become a condition for participation in global digital markets rather than a voluntary norm.

11.5 TRANSLATING THE GOVERNANCE ROADMAP INTO POLICY ACTION

The governance phases outlined above establish a sequenced pathway from scientific alignment to durable institutional coordination. The following policy recommendations translate each phase into practical and time-bound actions.

Each phase builds on the previous one:

- The short-term focuses on epistemic alignment and infrastructure-building.
- The medium-term introduces precautionary safeguards to prevent irreversible harm.
- The long-term institutionalizes global coordination through a binding but adaptable framework.

This sequencing ensures that governance evolves in parallel with technological capability growth.

The following policy recommendations translate the governance phases into practical actions:

Short-Term (0–2 Years): Build the Scientific Infrastructure

The first phase reduces epistemic fragmentation and creates a shared scientific foundation to support evidence-based governance. This means:

- Operationalizing the UN AI Scientific Panel and the AI Expert Technical Laboratory.
- Establishing formal coordination mechanisms between the UN Panel and the AI Expert Technical Laboratory (e.g., creating a joint steering committee with a defined division of responsibilities.)

Medium-Term (2–4 Years): Implement an Interim Governance Framework

Once a shared scientific understanding has been established, governance must shift from knowledge generation to precautionary risk containment. This entails:

- Negotiating a multilateral temporary restriction on categories of AI research identified as high-risk by the UN AI Scientific Panel.
- Developing measurable safety criteria for lifting restrictions, including robustness against misuse, controllability and interpretability benchmarks, and demonstrated resistance to autonomous self-improvement.
- Embedding multi-stakeholder governance collaboration mechanisms.

Long-Term (4-6 Years): Institutionalize Durable Global Governance

The final phase translates the institutional design requirements of the global AI governance architecture into binding yet adaptable legal and economic instruments. This translates into:

- Adopting a framework convention on AI Governance that sets core and binding obligations for states while allowing adaptable national implementation.
- Requiring certification for cross-border AI transfers, making responsible AI practices a condition for participation in global digital markets.

12. CONCLUSION

AI is emerging amid complex crises—climate change, biodiversity loss, inequality, and conflict—and is beginning to reshape these dynamics in ways that are difficult to predict. Its rapid advancement creates extraordinary potential to accelerate both societal progress and systemic harm.

The rapid diffusion of AI across key domains—planetary environment, energy systems, industry and labor, finance, and democracy and societal resilience—has outpaced our scientific understanding and regulatory capacity. This asymmetry creates a “black box” effect that threatens to amplify systemic risks driven by AI, such as misinformation, job displacement, and wealth concentration. As AI systems approach autonomous, superhuman capabilities, the danger of losing human control becomes a transboundary and potentially irreversible threat.

However, AI-related harms are not inevitable. To navigate this uncertainty, this report advocates for a phased roadmap toward global governance. Starting with a shared scientific baseline and moving toward a binding international framework, the proposed global AI governance strategy would establish universal safety thresholds that protect both social equity and planetary boundaries. Governance is the central lever: without it, AI risks accelerating unsustainable trajectories; with it, we can steer this general-purpose technology toward becoming a powerful tool for inclusive and resilient global development.

Ensuring that AI systems advance societal goals requires multistakeholder coordination. Integrating science, policy, and ethical principles allows us to chart a path that harnesses AI’s potential while safeguarding human and planetary well-being.

ANNEX I METHODOLOGY

For each sector, the analysis is based on a literature review as well as expert interviews with academics and practitioners, which were focused on understanding the speed, direction and impact of AI on society as a whole.

Literature was identified using targeted searches on Google, Google Scholar, and Columbia University's Library Catalogue (CLIO) with keywords relating to AI, biodiversity, energy, labor, finance, manufacturing, democracy, AI ethics, global governance, and superhuman intelligence. Using thematic analysis, we identified recurring patterns and organized them into key themes, which were subsequently developed into a coherent narrative during the drafting phase.

For the expert interviews, interviewees were first identified through the literature review and within our network (purpose sampling). Whenever some of the identified experts were unavailable, we applied a snowball sampling approach, identifying additional interviewees on the basis of referrals. While efforts were made to capture a range of perspectives across regions, the resulting interviewees sample is weighted towards AI and sector-specific experts based in the Global North. For this reason, the sample should not be considered statistically representative of the broader field. For these interviews, we adopted the methodology of semi-structured interviews. We interviewed a total of five experts between January 28 and February 28, 2026.

AI was used to support this research. We used Gemini 3, ChatGPT 5.2, and Claude to synthetize the literature review and make the draft report clearer. We used Otter AI to transcribe the interviews. All literature was reviewed by members of our research team, and all transcripts were manually edited. This report has been written, reviewed, and validated by the human researcher team at CCSI and Hitachi.

ANNEX II
COMPARATIVE PERSPECTIVES ON NATIONAL AND REGIONAL AI GOVERNANCE

Over the past few years, distinct regulatory approaches have emerged: while the European Union and South Korea have codified overarching law, the United States and China exhibit more agile, deregulated, or use-case-specific frameworks. The following analysis examines how these major jurisdictions are navigating AI governance.

EUROPEAN UNION

The EU AI Act (Regulation (EU) 2024/1689 laying down harmonized rules on artificial intelligence) is the first-ever comprehensive legal framework on AI worldwide. The Act is designed as a piece of adaptive legislation, leaving many details intentionally vague to permit later adaptation as technology changes. At its core is a risk classification system, which attributes the majority of responsibility to the developers of AI systems. On the basis of its risks, the Act distinguishes:

- When it presents unacceptable risks, AI practices are prohibited (Article 5), e.g., AI systems used to detect the emotional state of individuals in situations related to the workplace and education (Recital 44).
- When risk is considered high (Article 6), the regulation establishes obligations for developers and deployers. For example, developers are required to establish, implement, document and maintain a risk management system (Article 9). High-risk AI systems must be designed and developed so that they can be effectively overseen by humans during the period in which they are in use (Article 14).
- When risks are limited, the regulation establishes lighter transparency obligations, such that developers and deployers must ensure that end-users are aware that they are interacting with AI (Article 50).
- Minimal risk is unregulated.

Specific requirements are also established for general purpose AI (GPAI i.e. AI models that can competently perform a wide range of distinct tasks regardless of the way the model is placed on the market, and that can be integrated into a variety of downstream systems or applications,) especially with regards to maintaining technical documentation and providing information to downstream providers (Article 53).

To stimulate innovation, the EU also requires member states to create at least one AI regulatory sandbox at the national level (Article 57). With the goal of fostering innovation while identifying and mitigating any risks, sandboxes are controlled environments where AI systems can be developed, tested, and validated before being released to the market.

Finally the EU AI Act establishes:

- AI Office to develop expertise and capabilities in the field of AI (Article 64).
- Scientific Panel of Independent Experts, which, in support of the AI Office, will alert them about potential risks, help develop evaluation tools, and advise on the classification of AI models (Article 68).
- European Artificial Intelligence Board to assist and advise the Commission and Member States in applying the regulation effectively and consistently (Article 66).
- Advisory Forum, composed of a balanced mix of stakeholders from industry, start-ups, SMEs, civil society, and academia, to provide technical expertise and advice to the Board and the Commission (Article 67).

SOUTH KOREA

South Korea is the second jurisdiction in the world, following the EU, to enact a comprehensive regulatory law on AI. The Framework Act on the Development of Artificial Intelligence and Establishment of Trust Foundation (AI Framework Act) integrates nineteen bills, was approved in 2024, and took effect in January 2026. The law aims to promote market-led innovation, but also emphasizes safety, reliability, and civil rights, merging specific regulations with promotional measures.

The Act defines AI as "the electronic implementation of human intellectual abilities such as learning, reasoning, perception, judgment, and language understanding;" and AI systems as "AI-based systems that infer outputs such as predictions, recommendations, and decisions that affect real and virtual environments for given objectives, with varying levels of autonomy and adaptability." The legislation provides broad support for the entire AI ecosystem and establishes oversight for generative AI (i.e. "AI systems that generate various outputs such as text, sound, images, and videos by mimicking the structure and characteristics of input data") and high-impact AI. The latter covers eleven sensitive areas where AI could pose significant risks to human life, safety, or fundamental rights. High-impact AI operators must implement safety measures and, when providing products or services using high-impact AI, must evaluate the potential impacts on individuals' fundamental rights in advance.

The Ministry of Science and ICT is the main authority responsible for AI policy implementation, while the National AI Committee, under the president, deliberates and decides on AI policies. The AI Policy Center and AI Safety Institute support policy implementation.

The act also established promotional measures, outlining a comprehensive roadmap for national growth, including special support for SMEs and startups, the creation of AI clusters, the development of a professional workforce, and the promotion of dedicated AI data centers.

The act mandates transparency obligations, including prior notification to users when providing products or services using high-impact or generative AI. It also mandates labeling requirements for generative AI outputs, namely labels that distinguish between AI-generated and human-generated content. AI systems with cumulative computational power above a certain threshold must implement risk identification, assessment, and mitigation measures throughout the AI life cycle. Additionally, they must establish a risk management system to monitor and respond to AI-related safety incidents.

Compared to the EU AI Act which defines a risk-based approach, the AI Framework Act targets specific AI models, primarily high-impact AI, generative AI, and high-performance AI systems. These models are specifically identified as regulatory subjects because of the cumulative computational power used for their learning.

CHINA

China's approach to AI governance and regulation is a hybrid between the centralized approach of the EU and the decentralized approach in the US. China has favored laws tailored to specific use-cases, ranging from data security and cybersecurity to recommendation algorithms.

In 2015, China launched "Made In China 2025" (MIC2025), a national strategic plan to establish the country as a global leader in technologies like AI. A key component of this strategy involved supporting small- and medium-sized enterprises (SMEs) identified as primary drivers of innovation. The practical application of Chinese AI policy reveals a tiered enforcement strategy, with large companies expected to fully comply with AI regulations, while SMEs and startups are informally granted regulatory leeway to prevent heavy regulation from stifling innovation. This approach is intended to promote economic growth, innovation, and international competitiveness.

In 2023, several Chinese agencies jointly released the Interim Measures for the Management of Generative Artificial Intelligence Services (the "AI Measures"), which is the first administrative regulation on the management of generative AI services. In 2025, new "Labeling Rules" came into effect, making it mandatory for AI-generated content to be implicitly labeled, and explicitly labeled where applicable. In addition, China jointly released three national standards aimed at enhancing the security and governance of generative AI. The "AI Measures" do not provide a clear definition of "AI," but define "generative AI technology" as models and related technology that have the ability to generate text, images, audio, videos, or other content.

While China was initially expected to adopt a comprehensive, high-level AI law, such a law does not seem to be part of the political agenda anymore. Beijing seems to prioritize pilot programs and targeted measures over a rigid overarching statute, preserving regulatory flexibility, and allowing the technology to mature without being stifled by premature restrictions. The risk of this approach is that, in the absence of an overall law, companies are caught in a web of inconsistent and often contradictory rules.

Although formal legislation has stalled, standard-setting bodies are shaping the technical architecture for model evaluation, watermarking, and cybersecurity. This shifts the focus from broad legal principles to technical requirements, effectively governing through "hard-coding" expectations into the technology itself.

UNITED STATES

The landscape of U.S. AI governance has undergone a dramatic transformation, shifting from a distributed risk-management model to a centralized strategy focused on rapid development and global dominance.

In 2023, the U.S. established a comprehensive federal framework through Executive Order (EO) 14110. Rather than adopting the centralized, top-down models seen in the EU or China, the U.S. utilized a distributed approach. The EO tasked over fifty federal agencies with building regulatory capacity and integrating AI into operations across eight core policy areas, ranging from safety and privacy to civil rights and worker support. This federal effort was complemented by a surge in activity at the local level, with various states (such as e.g., Colorado and Texas) and several municipalities debating or passing their own AI regulations.

This trajectory shifted following the 2025 presidential transition. The new administration rescinded EO 14110 and introduced "Winning the Race: America's AI Action Plan" (Action Plan), a strategy that prioritizes deregulation and accelerated innovation over risk mitigation.

To support this new direction, the administration issued three binding orders designed to restrict the federal procurement of "biased" AI models, streamline permitting processes for critical AI infrastructure and data centers, and establish a global export strategy to solidify U.S. market leadership. The Action Plan suggests a reluctance to implement substantive AI laws at the federal level, while a debate within Congress continues regarding a potential federal "moratorium" to temporarily preempt most state-level AI legislation.

In the absence of a federal-level law, **California** introduced a state law focused on "frontier" AI systems. In September 2025 California adopted SB 53 - Transparency in Frontier Artificial Intelligence Act (TFAIA), targeting the most advanced developers, and establishing a risk reporting model. SB 53 comes after the SB 1047 - Safe and Secure Innovation for Frontier AI Models Act, which was vetoed and was never signed into law. Although SB 53 technically applies to only a handful of developers, its practical impacts are likely to ripple throughout the AI ecosystem, and beyond California.

SB 53 targets "frontier" AI systems, namely foundation models trained at extraordinary computational scales (i.e., computing power greater than 10^26 integer or floating-point operations, "FLOPs",) targeting firms whose models could produce catastrophic outcomes if misused or misaligned. The law defines "catastrophic risk" as the foreseeable risk that a model might:

- Cause death or serious injury to 50 or more people, or greater than $1 billion in damages;
- Provide expert-level assistance in creating or releasing a chemical, biological, radiological, or nuclear weapon;
- Autonomously commit major crimes or cyberattacks; or
- Evade control by developers or users.

SB 53 establishes specific obligations for frontier and large frontier AI developers. In particular:

- Large developers must publish an annual framework detailing governance, cybersecurity, and risk mitigation strategies.
- All frontier developers must issue reports on model capabilities, limitations, and risk assessment results before deployment.
- Developers must report safety incidents to the appropriate public safety authority within 15 days of discovering the incident, or 24 hours if the incident presents imminent risk of death or serious injury.
- Employers must provide anonymous reporting channels for risk concerns and are prohibited from retaliating against whistleblowers.

While each of the regulatory approaches described possess distinct strengths and weaknesses, none are sufficient to ensure the safe development and deployment of AI on a global scale. Because AI's societal impacts are far-reaching and transcend national borders, its regulation is inherently a global challenge. Regardless of how effective a framework may be within a specific domestic context, the absence of global governance exposes humanity to risks that no single nation can manage in isolation. Since a restriction in one jurisdiction is not binding in another, the absence of a global AI framework allows AI models to be developed and tested in those regions where minimal safeguards and public oversight are prescribed.

Topic	European Union	South Korea	China	United States — Federal	United States — California
Type of Framework	Comprehensive AI law	Comprehensive AI law	No single comprehensive AI law; targeted rules	No comprehensive AI law	Comprehensive AI law
Main Approach	Risk-based system	Targets specific AI types (high-impact, generative, high-compute)	Use-case specific and standards-based	Accelerated innovation & deregulation	Frontier-model risk regulation
Prohibited / Strict Uses	Unacceptable risk is prohibited	High-impact AI requires safety measures and rights impact assessment	Generative AI services regulated	Not specified	Frontier models that could enable catastrophic harms
Developer Duties	Risk management, human oversight, documentation	Safety measures, impact evaluations, lifecycle risk management for high-compute AI	Compliance with generative AI measures and technical standards	Federal procurement limits on "biased" AI, infrastructure and export orders	Risk frameworks, capability and risk reports, incident reporting, whistleblower protections
Developer Duties	Users informed when interacting with AI, GPAI documentation	User notice and labeling for generative AI outputs	Mandatory labeling of AI-generated content	Not specified	Mandatory model capability and risk reports before deployment
Innovation Measures	Mandatory national AI sandboxes	SME/startup support, AI clusters, workforce, data centers	SME support and regulatory leeway and pilot programs	Deregulation and accelerated innovation strategy	State-backed public compute cluster (CalCompute planned)

NOTES

1 "Understanding the Different Types of Artificial Intelligence," IBM Data and AI Team, International Business Machines Corporation (IBM), https://www.ibm.com/think/topics/artificial-intelligence-types.https://www.ibm.com/think/topics/artificial-intelligence-types.

2 United Nations Environment Programme, Making Peace with Nature: A Scientific Blueprint to Tackle the Climate, Biodiversity and Pollution Emergencies (United Nations Environment Programme, February 2021), https://wedocs.unep.org/handle/20.500.11822/34948.

3 "Newsroom: Biodiversity," World Health Organization, February 18, 2025, https://www.who.int/news-room/fact-sheets/detail/biodiversity. ; Maud Abdell, Zach Abraham, Dominic Andradi-Brown, Mike Barrett, Nathan Bennett, Becky Chaplin-Kramer, Samantha Cheng, Stefanie Deinet, Robin Freeman, Sarah Glaser, Rachel Golden-Kroner, Brent Loken, Valentina Marconi, Louise McRae, Ravic Nijbroek, Jeff Opperman, Pablo Pacheco, Hannah Puleston, Stephanie Roe, Lucia Ruiz, Kirsten Schuijt, Abel Valdivia, Aaron Vermeulen, and Daniel Viana, Living Planet Report 2024: A System of Peril (Gland: WWF, 2024), https://livingplanet.panda.org/en-US/. Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb. "Do We Want Less Automation: AI May Provide A Path to Decrease Inequality." Science 381, no. 6654 (July 13, 2023): 155–158. https://www.science.org/doi/abs/10.1126/science.adh9429.

4 "1.5°C: What it Means and Why it Matters," Climate Action, United Nations, https://www.un.org/en/climatechange/science/climate-issues/degrees-matter.

5 Health Effects Institute, State of Global Air 2024: Special Report (Boston: Health Effects Institute, 2024),https://www.un.org/en/climatechange/science/climate-issues/degrees-matter. ; Health Effects Institute, State of Global Air: A Report on Air Pollution and Its Role in the World's Leading Causes of Death (Boston: Health Effects Institute), https://www.stateofglobalair.org/resources/report/state-global-air-report-2025.

6 As of 2025, seven out of nine planetary boundaries have been breached: Climate Change, Biosphere Integrity, Land System Change, Freshwater Use, Biogeochemical Flows, Novel Entities, and Ocean Acidification. Aerosol loading and Ozone depletion are the last two. According to van Vuuren et al., a new breach will happen for aerosol loading and under current policy, the breach will exacerbate the others by 2050. "Planetary boundaries," Stockholm Resilience Centre, 2025, https://www.stockholmresilience.org/research/planetary-boundaries.html; Detlef P. van Vuuren, Jonathan C. Doelman, Isabela Schmidt Tagomori, Arthur H. W. Beusen, Sarah E. Cornell, Johan Röckstrom, Aafke M. Schipper, Elke Stehfest, Geanderson Ambrosio, Maarten van den Berg, Lex Bouwman, Vassilis Daioglou, Mathijs Harmsen, Paul Lucas, Kaj-Ivar van der Wijst & Willem-Jan van Zeist, "Exploring pathways for world development within planetary boundaries," Nature, 641 (2025): 910–916, https://doi.org/10.1038/s41586-025-08928-w

7 van Vuuren et al., "Exploring pathways for world development within planetary boundaries."

8 World Bank Group, Poverty, Prosperity, and Planet Report 2024: Pathways Out of the Polycrisis (Washington, D.C.: World Bank), https://www.worldbank.org/en/publication/poverty-prosperity-and-planet.

9 Ibid.

10 Ibid.

11 World Health Organization, "Global Hunger Declines, but Rises in Africa and Western Asia: UN Report," news release, July 28, 2025, https://www.who.int/news/item/28-07-2025-global-hunger-declines-but-rises-in-africa-and-western-asia-un-report.

12 Jeffrey D. Sachs, Guido Schmidt-Traub, Mariana Mazzucato, Dirk Messner, Nebojsa Nakicenovic, and Johan Rockström, "Six Transformations to Achieve the Sustainable Development Goals," Nature Sustainability, No. 2 (August, 26, 2019): 805-814, https://www.nature.com/articles/s41893-019-0352-9#citeas.

13 Ioannis Souliotis and Nikolaos Voulvoulis, "Sustainability Transitions: The Role of Systems Thinking in Improving Planetary Health and Human Prosperity," Frontiers in Environmental Science: Environmental Policy and Governance, Vol. 13 (December 2025), https://www.frontiersin.org/journals/environmental-science/articles/10.3389/fenvs.2025.1730692/full#cite.

14 Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena, "An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," Ethics, Governance, and Policies in Artificial Intelligence, Philosophical Studies Series, Vol. 144 (Springer, Cham, November 2021), https://link.springer.com/chapter/10.1007/978-3-030-81907-1_3.

15 For a detailed description of the report's methodology, please see Annex I.

16 Abdell et al., Living Planet Report 2024: A System of Peril.; Anne Olhoff, William F. Lamb, Minal Pathak, John Christensen, Takeshi Kuramochi, Taryn Fransen, Michel den Elzen, Joeri Rogelj, and Dan Tong, Emissions Gap Report 2025: Off Target – Continued Collective Inaction Puts Global Temperature Goal at Risk (Nairobi: United Nations Environment Programme, 2025), https://doi.org/10.59117/20.500.11822/48854.; OECD Environmental Outlook, Environmental Outlook on the Triple Planetary Crisis: Stakes, Evolution and Policy Linkages (Paris: OECD Publishing, November 2025), https://doi.org/10.1787/257ffbb6-en.

17 Victor Galaz, Maria Schewenius, Jonathan F. Donges, Ingo Fetzer, Erik Zhivkoplias, Wolfram Barfuss, Louis Delannoy, Lan Wang-Erlandsson, Maximilian Gelbrecht, Jobst Heitzig, Jonas Hentati-Sundberg, Christopher Kennedy, Nielja Knecht, Romi Lotcheris, Miguel Mahecha, Andrew Merrie, David Montero, Timon McPhearson, Ahmed Mustafa, Magnus Nyström, Drew Purves, Juan C. Rocha, Masahiro Ryo, Claudia van der Salm, Samuel T. Segun, Anna B. Stephenson, Elizabeth Tellman, Felipe Tobar, and Alice Vadrot, AI for a Planet Under Pressure, ed. Victor Galaz and Maria Schewenius (Stockholm: Stockholm Resilience Centre, Potsdam Institute for Climate Impact Research, and Google DeepMind, November 2025), http://arxiv.org/abs/2510.24373.

18 Sam A. Reynolds, Sara Beery, Neil Burgess, Mark Burgman, Stuart H. M. Butchart, Steven J. Cooke, David Coomes, Finn Danielsen, Enrico Di Minin, América Paz Durán, Francis Gassert, Amy Hinsley, Sadiq Jaffer, Julia P. G. Jones, Binbin V. Li, Oisin Mac Aodha, Anil Madhavapeddy, Stephanie A. L. O'Donnell, William M. Oxbury, Lloyd Peck, and William J. Sutherland, "The Potential for AI to Revolutionize Conservation: a Horizon Scan," Trends in Ecology and Evolution, 40(2) (February 2025): 191-207, https://doi.org/10.1016/j.tree.2024.11.013.

19 Sharif Islam, Hanna Koivula, Carrie Andrew, Julian Lopez Gordillo, Claus Weiland, Dmitry Schigel, Dag Endresen, Christos Arvanitidis, Eli Chadwick, and Stian Soiland-Reyes, "FAIR Digital Twins for Biodiversity: Enabling Data, Model, and Workflow Integration," npj Biodiversity, 5 (February 2026): Article 5, https://www.nature.com/articles/s44185-025-00116-3.; Adrien Tofighi-Niaki, Sanjukta Moorthy, and Ramin Soleymani, Risks and Benefits of Artificial Intelligence in Locally Led Nature Restoration (REDAA, May 2025), https://www.redaa.org.

20 Reynolds et al., "The Potential for AI to Revolutionize Conservation: a Horizon Scan."

21 Francis Gassert, Antonia Gawel, Mike Harfoot, Abby Mayer, Kanika Singhal, Fred Stolle, and Laura Vary, "AI for Nature: How AI Can Democratize and Scale Action on Nature" (Land and Carbon working paper: World Resources Institute, November 4, 2025), https://www.wri.org/research/ai-nature-how-ai-can-democratize-and-scale-action-nature.

22 United Nations Climate Change: Technological Executive Committee, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries (United Nations Framework Convention on Climate Change (UNFCCC), United Nations Industrial Development Organization (UNIDO), July 2025), https://unfccc.int/ttclear/misc_/StaticFiles/gnwoerk_static/AI4climateaction/f2922b97c4cf431996c468e622127eb5/112f8be560ea447dab5ff2e53ab3f6e4.pdf.

23 United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries.

24 Tofighi-Niaki, Moorthy, and Soleymani, Risks and Benefits of Artificial Intelligence in Locally Led Nature Restoration.

25 Allison D. Binley, Vincent Jaimie G., Trina Rytwinski, Peter Soroye, and Joseph R. Bennett, "Making the Most of Existing Data in Conservation Research," Perspectives in Ecology and Conservation, 22(2) (April–June 2024): 122–128, https://doi.org/10.1016/j.pecon.2023.11.004.

26 Galaz et al., AI for a Planet Under Pressure.

27 Ibid.

28 United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries.

29 Tofighi-Niaki, Moorthy, and Soleymani, Risks and Benefits of Artificial Intelligence in Locally Led Nature Restoration.

30 United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries.

31 Gülcay Ercan Oğuztürk, "AI-Driven Irrigation Systems for Sustainable Water Management: A Systematic Review and Meta-Analytical Insights," Smart Agricultural Technology, 11 (August 2025): https://doi.org/10.1016/j.atech.2025.100982

32 Mohd Javaid, Abid Haleem, Ibrahim Haleem Khan, and Rajiv Suman, "Understanding the Potential Applications of Artificial Intelligence in Agriculture Sector," Advanced Agrochem, 2(1) (March 2023): 15–30, https://doi.org/10.1016/j.aac.2022.10.001.

33 Gassert et al., "AI for Nature: How AI Can Democratize and Scale Action on Nature."

34 Michael Copley, "Data Centers, Backbone of the Digital Economy, Face Water Scarcity and Climate Risk."; Spencer and Singh, Energy and AI: World Energy Outlook Special Report.; Adam Zewe, "Explained: Generative AI's Environmental Impact," MIT News, January 17, 2025, https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117.

35 Organisation for Economic Co-operation and Development (OECD), Artificial Intelligence (AI) End-to-End: The Environmental Impact of the Full AI Lifecycle Needs to Be Comprehensively Assessed – Issue Note (Paris: OECD, September 2024), https://wedocs.unep.org/items/5f3afe87-5419-439b-a6ee-14240b2605e9.

36 Jim Thomas, "Black Box" Biotech: The Integration of Artificial Intelligence with Synthetic Biology (Johannesburg: African Centre for Biodiversity, Third World Network, and ETC Group, September 2024), https://acbio.org.za/wp-content/uploads/2024/09/Black-Box-Biotech-AI-Synth-bio-paper_fin.pdf.

37 Reynolds et al., "The Potential for AI to Revolutionize Conservation: a Horizon Scan."

38 Galaz et al., AI for a Planet Under Pressure.

39 Ibid.

40 Tofighi-Niaki, Moorthy, and Soleymani, Risks and Benefits of Artificial Intelligence in Locally Led Nature Restoration.

41 Will Douglas Heaven, "Large Language Models Can Do Jaw-Dropping Things. But Nobody Knows Exactly Why," March 4, 2024, https://www.technologyreview.com/2024/03/04/1089403/large-language-models-amazing-but-nobody-knows-why/.

42 Thomas, "Black Box" Biotech: The Integration of Artificial Intelligence with Synthetic Biology.

43 Caroline Jack, Lexicon of Lies: Terms for Problematic Information (New York: Data & Society Research Institute, 2017), https://datasociety.net/library/lexicon-of-lies/.

44 Galaz et al., AI for a Planet Under Pressure.

45 Spencer and Singh, Energy and AI: World Energy Outlook Special Report.

46 George Kyriakarakos, "Artificial Intelligence and the Energy Transition," Sustainability, 17(3) (January 30, 2025): 1140, https://doi.org/10.3390/su17031140.

47 Kyriakarakos, "Artificial Intelligence and the Energy Transition"; Younes Zahraoui, Tarmo Korõtko, Argo Rosin, Saad Mekhilef, Mehdi Seyedmahmoudian, Alex Stojcevski, and Ibrahim Alhamrouni, "AI Applications to Enhance Resilience in Power Systems and Microgrids—A Review," Sustainability, 16(12) (June 10, 2024): 4959, https://doi.org/10.3390/su16124959.

48 Kyriakarakos, "Artificial Intelligence and the Energy Transition."

49 Ibid.

50 United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries.

51 Sameer Algburi, Salah Sabeeh Abed Al Kareem, I. B. Sapaev, Otabek Mukhitdinov, Qusay Hassan, Doaa H. Khalaf, and Feryal Ibrahim Jabbar, "The Role of Artificial Intelligence in Accelerating Renewable Energy Adoption for Global Energy Transformation," Unconventional Resources 8 (October 2025): 100229, https://doi.org/10.1016/j.uncres.2025.100229.

52 Nancy W. Stauffer, "How Artificial Intelligence can Help Achieve a Clean Energy Future," MIT News, November 24, 2025, https://news.mit.edu/2025/how-ai-can-help-achieve-clean-energy-future-1124.

53 "Data Centers, AI, and the Grid: Why Flexibility Must Come First," Hitachi Energy, July 23, 2025, https://www.hitachienergy.com/jp/ja/news-and-events/perspectives/2025/06/data-centers-ai-and-the-grid-why-flexibility-must-come-first.

54 Ashley Siagian and Emmanuel A. San Andres, Using AI to Power Up Efficient and Resilient Energy Systems, APEC Policy Support Unit Policy Brief No. 67 (Singapore: Asia-Pacific Economic Cooperation, August 2025), https://www.apec.org/docs/default-source/publications/2025/8/225_psu_pb-67.pdf?sfvrsn=bb44ff0c_1.

55 Spencer and Singh, Energy and AI: World Energy Outlook Special Report.

56 Ibid.

57 MOFs are a class of coordination polymers consisting of metal clusters, aerogels are a a class of synthetic porous ultralight material.

58 United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries.

59 Ganesh Shankar Gowekar, "How Oil and Gas Industry are Transforming with AI and ML, World Journal of Advanced Research and Review, 23(02) (September 2024): 1234-1238, https://wjarr.com/sites/default/files/WJARR-2024-2722.pdf; United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries.

60 Spencer and Singh, Energy and AI: World Energy Outlook Special Report; United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries.

61 Spencer and Singh, Energy and AI: World Energy Outlook Special Report.

62 Stauffer, "How Artificial Intelligence can Help Achieve a Clean Energy Future."

63 Kyriakarakos, "Artificial Intelligence and the Energy Transition."

64 United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries.

65 Lori Regattieri, AI and Climate Change: The Global South Facing the New Geopolitics of Innovation (Green Screen Coalition, 2025), https://greenscreen.network/files/pdf/AI_and_Climate_Change-The_Global_South_Facing_The_New_Geopolitics_of_Innovation__EN_.pdf.

66 Spencer and Singh, Energy and AI: World Energy Outlook Special Report

67 Siagian and San Andres, Using AI to Power Up Efficient and Resilient Energy Systems, APEC Policy Support Unit Policy Brief No. 67.

68 Kyriakarakos, "Artificial Intelligence and the Energy Transition."

69 In the context of AI and labor, it is fundamental to understand the difference between automation and augmentation.
Automation refers to the substitution of machines for human labor in the execution of one or more tasks within a workflow. This process typically raises productivity, creating long-term economic growth and average living standards. However, when machines replace workers, these productivity gains impose significant costs on displaced workers.
Labor augmentation occurs when the use of machines in certain tasks increases productivity in other tasks, thereby enhancing overall labor productivity. When technological change generates new tasks, displaced workers may transition into alternative forms of work. Unlike automation, which directly reallocates tasks from humans to machines, augmentation operates through interactions across tasks, whether within pairs, clusters, or broader task networks. For more see: Ajay Agrawal, Joshua S. Gans, Avi GoldFarb, "Do We Want Less Automation: AI May Provide A Path to Decrease Inequality," Science, 381(6654) (July 13, 2023): 155-158, https://www.science.org/doi/abs/10.1126/science.adh9429.

70 Dona Ghosh, Rajarshi Ghosh, Sahana Roy Chowdhury, and Boudhayan Ganguly, "AI-Exposure and Labour Market: A Systematic Literature Review on Estimations, Validations, and Perceptions," Management Review Quarterly, 75(1) (January 2024): 677–704, https://doi.org/10.1007/s11301-023-00393-x.

71 Francesco Filippucci, Peter Gal, Cecilia Jona-Lasinio, Alvaro Leandro, and Giuseppe Nicoletti, The Impact of Artificial Intelligence on Productivity, Distribution and Growth: Key Mechanisms, Initial Evidence and Policy Challenges, OECD Artificial Intelligence Papers No. 15 (Paris: Organisation for Economic Co-operation and Development, April 2024), https://doi.org/10.1787/8d900037-en.

72 Ibid.

73 Federica Acerbi, Dai Andrew Forterre, and Marco Taisch, "Role of Artificial Intelligence in Circular Manufacturing: A Systematic Literature Review," IFAC-PapersOnLine, 54(1) (2021): 367–372, https://doi.org/10.1016/j.ifacol.2021.08.040.

74 United Nations Industries Development Organization, The Future of Industries for Development: UNIDO's Vision 2025 (Vienna: UNIDO, 2025), https://www.unido.org/sites/default/files/unido-publications/2025-11/The%20Future%20of%20Industries%20for%20Development%20-%20UNIDO%27s%20Vision%202050.pdf.

75 Sai Dhiresh Kilari, The Impact of Artificial Intelligence on Automated Manufacturing and Supply Chain Systems: Advancements, Challenges, and Future Prospects (March 23, 2025), http://dx.doi.org/10.2139/ssrn.5204723.

76 Immad A Shah and SukhDev Mishra, "Artificial intelligence in advancing occupational health and safety: an encapsulation of developments," Journal of Occupational Health 66, no. 1 (2024). https://doi.org/10.1093/joccuh/uiad017

77 M. M. Manjurul Islam, Jakaria Islam Emon, Kok Yew Ng, Abdoreza Asadpour, M. M. Rafi Al Aziz, Marcia L. Baptista, and Jong-Myon Kim, "Artificial Intelligence in Smart Manufacturing: Emerging Opportunities and Prospects," in Artificial Intelligence for Smart Manufacturing and Industry X.0, Springer Series in Advanced Manufacturing (Cham: Springer, 2025), 9–36, https://link.springer.com/chapter/10.1007/978-3-031-80154-9_2.

78 Nodira Bositkhanova and Sherzodbek Murodilla Ugli Dadaboyev. "Revolutionizing Workforce Planning: The Strategic Role of AI in HR Strategy." Discover Global Society 3 (2025): Article 100. https://doi.org/10.1007/s44282-025-00252-y.

79 Ibid.

80 Ernst & Young, "EY survey: companies advancing responsible AI governance linked to better business outcomes," Press release, October 8, 2025. https://www.ey.com/en_gl/newsroom/2025/10/ey-survey-companies-advancing-responsible-ai-governance-linked-to-better-business-outcomes.

81 R. Maria del Rio-Chanona, Ekkehard Ernst, Rossana Merola, Daniel Samaan, and Ole Teutloff, "AI and Jobs: A Review of Theory, Estimates, and Evidence," arXiv preprint arXiv:2509.15265 (September 22, 2025), https://arxiv.org/pdf/2509.15265.

82 Dennis Pepple and Nadeesha Muthuthantrige, "Artificial Intelligence, Innovation and the New Architecture of Exploitation: Towards Reconfiguring Humanness in the Age of Algorithmic Labour," Journal of Innovation & Knowledge, 11 (January–February 2026): 100878, https://doi.org/10.1016/j.jik.2025.100878.

83 United Nations Development Programme, The Next Great Divergence: Why AI may Widen Inequality Between Countries (New York: UNDP, 2025), https://www.undp.org/asia-pacific/publications/next-great-divergence; United Nations Industries Development Organization, The Future of Industries for Development: UNIDO's Vision 2025.

84 Filippucci et al., The Impact of Artificial Intelligence on Productivity, Distribution and Growth: Key Mechanisms, Initial Evidence and Policy Challenges

85 Daron Acemoglu and Pascual Restrepo, "Robots and Jobs: Evidence from US Labor Markets" (NBER Working Paper Series No. 23285, Cambridge, MA: National Bureau of Economic Research, March 2017), https://www.nber.org/papers/w23285

86 Gmyrek et al., "Generative AI and Jobs: A Refined Global Index of Occupational Exposure."

87 Ibid.

88 Thomas H. Davenport and Laks Srinivasan, "Companies Are Laying Off Workers Because of AI's Potential—Not Its Performance," HBR, January 29, 2026, https://hbr.org/2026/01/companies-are-laying-off-workers-because-of-ais-potential-not-its-performance.

89 Katherine C. Kellogg, Melissa A. Valentine, and Angèle Christin, "Algorithms at Work: The New Contested Terrain of Control," Academy of Management Annals, 14(1) (2020): 366–410, https://doi.org/10.5465/annals.2018.0174.

90 del Rio-Chanona et al., "AI and Jobs: A Review of Theory, Estimates, and Evidence."

91 Ibid.

92 Kamal Al-Sabahi, Yousuf Khamis Al Mabsali, Faozi A. Almaqtari, and Salim Amer Salim Al-Rashdi, "Transforming Sustainable Finance: The Impact of Artificial Intelligence," in AI Integration for Business Sustainability, Contributions to Environmental Sciences & Innovative Business Technology (Cham: Springer, 2025), 73–100, https://link.springer.com/chapter/10.1007/978-981-96-3464-4_4; Abdelghafar M. Elhady and Samaa Shohieb, "AI-Driven Sustainable Finance: Computational Tools, ESG Metrics, and Global Implementation," Future Business Journal 11 (2025): 209, https://doi.org/10.1186/s43093-025-00610-x.

93 Elhady and Shohieb, "AI-Driven Sustainable Finance: Computational Tools, ESG Metrics, and Global Implementation."

94 International Monetary Fund (IMF), Global Financial Stability Report: Steadying the Course—Uncertainty, Artificial Intelligence, and Financial Stability, Chap. 3, "Advances in Artificial Intelligence: Implications for Capital Market Activities," (October 2024), https://www.imf.org/-/media/files/publications/gfsr/2024/october/english/ch3.pdf.

95 Georgios Pavlidis, "Empowering Sustainable Finance with Artificial Intelligence: A Framework for Responsible Implementation," in A Research Agenda for Financial Law and Regulation (Cheltenham, UK: Edward Elgar Publishing, 2025), 23–38, https://doi.org/10.4337/9781803929996.00009.

96 International Monetary Fund (IMF), Global Financial Stability Report: Steadying the Course—Uncertainty, Artificial Intelligence, and Financial Stability, Chap. 3.

97 Elhady and Shohieb, "AI-Driven Sustainable Finance: Computational Tools, ESG Metrics, and Global Implementation."

98 Al-Sabahi et al., "Transforming Sustainable Finance: The Impact of Artificial Intelligence."

99 Shargiil Bashir, "How AI Can Help Close the Finance Gap Between Climate Ambition and Action," World Economic Forum, January 13, 2026, https://www.weforum.org/stories/2026/01/ai-close-the-climate-gap/.

100 Alpha opportunities are investment opportunities that offer the potential to earn returns above the overall market average.

101 "How AI is Transforming Investing," BlackRock, August 27, 2025, https://www.blackrock.com/us/individual/insights/ai-investing.

102 Jonathan Everhart, "The World Bank & AI: Leveraging AI to Accelerate Global Investments for Economic Development and Climate Finance," The Bretton Woods Committee, September 20, 2023, https://brettonwoods.org/the-world-bank-ai-leveraging-ai-to-accelerate-global-investments-for-economic-development-and-climate-finance/.

103 Sep Pashang and Olaf Weber, "AI for Sustainable Finance: Governance Mechanisms for Institutional and Societal Approaches," pg. 203-229, in The Ethics of Artificial Intelligence for the Sustainable Development Goals (May 4, 2023), https://link.springer.com/chapter/10.1007/978-3-031-21147-8_12.

104 Iñaki Aldasoro, Leonardo Gambacorta, Anton Korinek, Vatsala Shreeti and Merlin Stein, Intelligent Financial System: How AI is Transforming Finance (BIS Working Papers No. 1194: BIS, June 2024), https://www.bis.org/publ/work1194.pdf.

105 International Monetary Fund (IMF), Global Financial Stability Report: Steadying the Course—Uncertainty, Artificial Intelligence, and Financial Stability, Chap. 3.

106 Advait Arun, Bubble or Nothing (Center for Public Enterprise, November 12, 2025), https://publicenterprise.org/report/bubble-or-nothing/.

107 Ibid.

108 International Monetary Fund (IMF), Global Financial Stability Report: Steadying the Course—Uncertainty, Artificial Intelligence, and Financial Stability, Chap. 3.

109 Elhady and Shohieb, "AI-Driven Sustainable Finance: Computational Tools, ESG Metrics, and Global Implementation."

110 Pashang and Weber, "AI for Sustainable Finance: Governance Mechanisms for Institutional and Societal Approaches."

111 Pavlidis, "Empowering Sustainable Finance with Artificial Intelligence: A Framework for Responsible Implementation"; Al-Sabahi et al., "Transforming Sustainable Finance: The Impact of Artificial Intelligence."

112 Bruce Schneier, "Ten Ways AI Will Change Democracy," Harvard Kennedy School, November 6, 2023, https://ash.harvard.edu/articles/ten-ways-ai-will-change-democracy/.

113 Rachel George and Ian Klaus. AI and Democracy: Mapping the Intersections. (Washington, DC: Carnegie Endowment for International Peace, January 8, 2026). https://carnegieendowment.org/research/2026/01/ai-and-democracy-mapping-the-intersections.

114 Christopher Summerfield, Lisa P. Argyle, Michiel Bakker, Teddy Collins, Esin Durmus, Tyna Eloundou, Iason Gabriel, Deep Ganguli, Kobi Hackenburg, Gillian K. Hadfield, Luke Hewitt, Saffron Huang, Hélène Landemore, Nahema Marchal, Aviv Ovadya, Ariel Procaccia, Mathias Risse, Bruce Schneier, Elizabeth Seger, Divya Siddarth, Henrik Skaug Sætra, Michael Henry Tessler, and Matthew Botvinick, "The Impact of Advanced AI Systems on Democracy," Nature Human Behaviour, 9 (2025): 2420–2430, https://www-nature-com.ezproxy.cul.columbia.edu/articles/s41562-025-02309-z.

115 Organisation for Economic Co-operation and Development. Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions (Paris, OECD Publishing, 2025), 228-238, https://doi.org/10.1787/795de142-en

116 Bill Adair, Chengkai Li, Jun Yang, and Cong Yu, Progress Toward 'the Holy Grail': The Continued Quest to Automate Fact-Checking (2017), https://cj2017.northwestern.edu/documents/progress-cj2017-paper-18.pdf.

117 Matej Babšek, Dejan Ravšelj, Lan Umek, and Aleksander Aristovnik. "Artificial Intelligence Adoption in Public Administration: An Overview of Top-Cited Articles and Practical Applications," AI 6, no. 3 (2025): 44. https://doi.org/10.3390/ai6030044.

118 United Nations, Governing AI for Humanity: Final Report (New York: UN, 2024), https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.

119 World Economic Forum and Organisation for Economic Co-operation and Development. AI in Strategic Foresight: Reshaping Anticipatory Governance. White paper. (Geneva and Paris, 2025). https://www.weforum.org/publications/ai-in-strategic-foresight-reshaping-anticipatory-governance/.

120 Organisation for Economic Co-operation and Development. "Steering AI's Future: Strategies for Anticipatory Governance," OECD Artificial Intelligence Papers, no. 32 (2025). https://doi.org/10.1787/5480ff0a-en

121 United Nations Development Programme.

122 Phumzile Van Damme, Future-Proofing Elections Against Deepfake Disinformation (Civics and Digital Democracy Initiative, 2025).

123 Josh A. Goldstein and Andrew Lohn. Deepfakes, Elections, and Shrinking the Liar's Dividend (Brennan Center for Justice, January 23, 2024). https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend

124 George and Klaus, AI and Democracy: Mapping the Intersections.

125 Janna Anderson and Lee Rainie, Experts Predict the Best and Worst Changes in Digital Life by 2035 (Pew Research Center, 2023).; Fabrizio Germano, Vicenç Gómez, and Francesco Sobbrio. "Ranking for Engagement: How Social Media Algorithms Fuel Misinformation and Polarization" (BSE Working Paper No. 1501, Barcelona School of Economics, July 2025), https://bse.eu/research/working-papers/ranking-for-engagement-how-social-media-algorithms-fuel-misinformation-and-polarization

126 Kate Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. (New Haven, CT: Yale University Press, 2021).; Davide Castelvecchi and Benjamin Thompson, "It Keeps Me Awake at Night: Machine-Learning Pioneer on AI's Threat to Humanity," Nature, November 12, 2025, https://www.nature.com/articles/d41586-025-03686-1; Yoshua Bengio, Sören Mindermann, Daniel Privitera, Tamay Besiroglu, Rishi Bommasani, Stephen Casper, Yejin Choi, Philip Fox, Ben Garfinkel, Danielle Goldfarb, Hoda Heidari, Anson Ho, Sayash Kapoor, Leila Khalatbari, Shayne Longpre, Sam Manning, Vasilios Mavroudis, Mantas Mazeika, Julian Michael, Jessica Newman, Kwan Yee Ng, Chinasa T. Okolo, Deborah Raji, Girish Sastry, Elizabeth Seger, Theodora Skeadas, Tobin South, Emma Strubell, Florian Tramèr, Lucia Velasco, and Nicole Wheeler, International AI Safety Report: The International Scientific Report on the Safety of Advanced AI (DSIT 2025/001, 2025), https://www.gov.uk/government/publications/international-ai-safety-report-2025; Sabhanaz Rashid Diya, "Platform Governance and the Political Economy of Global AI Regulation," CIGI Paper No. 345 (The Centre for International Governance Innovation (CIGI), December 2025), https://www.cigionline.org/static/documents/Sabhanaz_Rashid_Diya.pdf.; Karen Hao, Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI (May 2025).

127 Castelvecchi and Thompson, "It Keeps Me Awake at Night: Machine-Learning Pioneer on AI's Threat to Humanity"; Hao, Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI.

128 Alan Chan, "What Role Should Governments Play in Providing AI Agent Infrastructure?" GovAI, July 1, 2025, https://www.governance.ai/analysis/what-role-should-governments-play-in-providing-ai-agent-infrastructure.

129 Dan Hendrycks, Eric Schmidt, and Alexandr Wang, Superintelligence Strategy (March 2025), https://files.nationalsecurity.ai/Superintelligence_Strategy.pdf; United Nations, Governing AI for Humanity.

130 International Committee of the Red Cross. Autonomous Weapon Systems and International Humanitarian Law: Selected Issues. (Geneva: ICRC, 2026), https://www.icrc.org/en/article/autonomous-weapon-systems-and-international-humanitarian-law-selected-issues; Sarah Grand-Clement, Artificial Intelligence Beyond Weapons: Application and Impact of AI in the Military Domain (Geneva: UNIDIR, 2023), https://unidir.org/publication/artificial-intelligence-beyond-weapons-application-and-impact-of-ai-in-the-military-domain/.

131 Bengio et al., International AI Safety Report: The International Scientific Report on the Safety of Advanced AI.

132 Ben Cumming, Safety Standards Delivering Controllable and Beneficial AI Tools (Future of Life Institute, February 5, 2025), https://futureoflife.org/wp-content/uploads/2025/02/AI-Action-Summit-Tool-AI-Explainer-V5.pdf.

133 Yuval Noah Harari, Nexus: A Brief History of Information Networks from the Stone Age to AI (September 2024).

134 Heaven, "Large Language Models Can Do Jaw-Dropping Things. But Nobody Knows Exactly Why,"

135 Melissa Heikkilä, "Nobody Knows How AI Works," MIT Technology Review, March 5, 2024, https://www.technologyreview.com/2024/03/05/1089449/nobody-knows-how-ai-works/.

136 "Statement on Superintelligence," Future of Life Institute, Superintelligence Statement, 2025, https://superintelligence-statement.org/.

137 Hiba Bachiri. "AI for Environment and Ecosystems Toolkit for Policymakers." Global AI Ethics and Governance Observatory, UNESCO. https://www.unesco.org/ethics-ai/en/node/288.

138 Edward Ansett and Sa'ad Ahmed, "Developing Sustainable Data Centres: A Strategic Roadmap to Achieve Net Zero Carbon and Reduce Environmental Impact, Ramboll, September 2025), https://www.ramboll.com/insights/decarbonise-for-net-zero/how-to-develop-sustainable-data-centres.

139 United Nations Environment Programme, Artificial Intelligence (AI) End-to-End: The Environmental Impact of the Full AI Lifecycle Needs to be Comprehensively Assessed (Nairobi: UNEP, 2024), https://wedocs.unep.org/rest/api/core/bitstreams/07b3c8fc-bd30-4b92-b5f4-d665e927b59d/content.

140 Rachel Esbenshade, "Biodiversity and Nature-Based Solutions for Data Centres: Development and Operational Strategies," Ramboll, December 9, 2025, https://www.ramboll.com/insights/biodiversity-and-ecosystems/nature-based-solutions-data-centre-for-biodiversity.

141 Thomas, Black Box" Biotech: The Integration of Artificial Intelligence with Synthetic Biology.

142 United Nations University, "UNU Signs Agreements to Establish New Institute Focused on Big Data and Artificial Intelligence," press release, December 12, 2025, https://unu.edu/press-release/unu-signs-agreements-establish-new-institute-focused-big-data-and-artificial.

143 Abdulmonaem Essed, Kolawole Iyiola, and Ahmad Alzubi, "Unpacking Artificial Intelligence's Role in the Energy Transition: The Mediating and Moderating Roles of Knowledge Production and Financial Development," Energies, 18(17) (August 25, 2025): 4512, https://doi.org/10.3390/en18174512.

144 Siagian and San Andres, Using AI to Power Up Efficient and Resilient Energy Systems, APEC Policy Support Unit Policy Brief No. 67.

145 Kenddrick Chan, Devorah West, Marie Teo, Harriet Brown, Tom Westgarth, and Thomas Smith, Greening AI: A Policy Agenda for the Artificial Intelligence and Energy Revolutions (London: Tony Blair Institute for Global Change, May 2024), https://institute.global/insights/climate-and-energy/greening-ai-a-policy-agenda-for-the-artificial-intelligence-and-energy-revolutions.

146 Lisa Sachs, Perrine Toledano, Bradford M. Willis, The Infrastructure of Intelligence: Rethinking Energy Systems in the Age of AI (New York: Columbia Center on Sustainable Investment, June 3, 2025), https://ccsi.columbia.edu/news/infrastructure-intelligence-rethinking-energy-systems-age-ai/.

147 United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries.

148 Ole Teutloff and Fabian Braesemann, "Resilienz statt Reskilling: Wie KI den Arbeitsmarkt verändert und wie wir darauf reagieren müssen," Wirtschaftsdienst, 105(10) (October 2025): 715–719, https://doi.org/10.2478/wd-2025-0184.

149 Fabian Stephany and Ole Teutloff, "What Is the Price of a Skill? The Value of Complementarity," Research Policy, 53(1), (January 2024): 104898, https://doi.org/10.1016/j.respol.2023.104898.

150 Tsai-fang Chen, "AI, Job Displacement, and the WTO: Identifying Legal Gaps and Charting a Worker-Centered Path," Yale Journal of International Law (November 10, 2024), https://yjil.yale.edu/posts/2025-11-10-ai-job-displacement-and-the-wto-identifying-legal-gaps-and-charting-a-worker.

151 International Monetary Fund (IMF), Global Financial Stability Report: Steadying the Course—Uncertainty, Artificial Intelligence, and Financial Stability, Chap. 3.

152 Circuit breakers temporarily halt trading when prices move beyond predefined thresholds, giving markets time to stabilize and preventing panic-driven cascades.

153 Margining mechanisms require participants to post collateral or maintain financial buffers, ensuring that losses from extreme price swings do not propagate through the system.

154 Columbia Center on Sustainable Development and Man Group, Climate Allocation Compass, a Framework for Real-World Decarbonization (Compass-FRWD) (New York: Columbia Center on Sustainable Investment (CCSI) and London: Man Group, August 2024).

155 Jeffery Sachs, Lisa Sachs, Ana Maria Camelo Vega, and Bradford Willis, Lowering the Cost of Capital for Climate and SDG Finance in Emerging Markets and Developing Economies (EMDEs) (New York: Columbia Center on Sustainable Investment (CCSI), April 2025), https://

ccsi.columbia.edu/sites/ccsi.columbia.edu/files/content/docs/publications/CCSI-Cost-of-Capital-in-EMDEs.pdf.

156 Perrine Toledano, Ana M. Camelo Vega and Tucker Wilke, From Promise to Performance: Reforming Blended Finance for Scale (New York: Columbia Center on Sustainable Investment (CCSI), June 2025), https://ccsi.columbia.edu/wp-content/uploads/2025/07/CCSI-From-Promise-to-Performance-Reforming-Blended-Finance-for-Scale-Report-1.pdf.

157 Alexander Romanishyn, Olena Malytska, and Vitaliy Goncharuk, "AI-Driven Disinformation: Policy Recommendations for Democratic Resilience," Frontiers in Artificial Intelligence, 8 (July 31, 2025): 1569115, https://doi.org/10.3389/frai.2025.1569115.

158 Vija Kalniņa, "Guarding Democracy in the Age of AI: When Everyone Has to Do Their Part," Harvard University: Edmond and Lily Safra Center for Ethics, February 17, 2026, https://www.ethics.harvard.edu/news/2026/02/guarding-democracy-age-ai-when-everyone-has-do-their-part.

159 Paul Anand, "AI: Challenges for Democracy and Some Policy Solutions," Journal of Human Development and Capabilities, 26(3) (June 18, 2025): 461–470. https://doi.org/10.1080/19452829.2025.2518307.

160 Romanishyn, Malytska, and Goncharuk, "AI-Driven Disinformation: Policy Recommendations for Democratic Resilience."

161 Francis Fukuyama, Chris Dann, and Beatriz Magaloni, "Delivering for Democracy: Why Results Matter," Journal of Democracy, 36(2) (April 2025): 5–19, https://www.journalofdemocracy.org/articles/delivering-for-democracy-why-results-matter/.

162 Dina Ali Seiam and Doaa Salman, "Examining the Global Influence of E-Governance on Corruption: A Panel Data Analysis," Future Business Journal, 10(29) (2024), https://doi.org/10.1186/s43093-024-00319-3.

163 United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries; Gassert et al., "AI for Nature: How AI Can Democratize and Scale Action on Nature."

164 Islam et al., "Artificial Intelligence in Smart Manufacturing: Emerging Opportunities and Prospects."

165 Reynolds et al., "The Potential for AI to Revolutionize Conservation: A Horizon Scan."

166 Islam et al., "Artificial Intelligence in Smart Manufacturing: Emerging Opportunities and Prospects."

167 Florence G'sell, Regulating Under Uncertainty: Governance Options for Generative AI (Stanford, CA: Cyber Policy Center, Freeman Spogli Institute for International Studies, Stanford University, 2024), https://cyber.fsi.stanford.edu/content/regulating-under-uncertainty-governance-options-generative-ai.

168 Kalniņa, "Guarding Democracy in the Age of AI: When Everyone Has to Do Their Part."

169 United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing Mitigation and Adaptation in Developing Countries.

170 Ibid.

171 Kyriakarakos, "Artificial Intelligence and the Energy Transition.

172 Anand, "AI: Challenges for Democracy and Some Policy Solutions"; G'sell, Regulating Under Uncertainty: Governance Options for Generative AI.

173 United Nations Climate Change, Artificial Intelligence for Climate Action: Advancing

174 Tofighi-Niaki, Moorthy, and Soleymani, Risks and Benefits of Artificial Intelligence in Locally Led Nature Restoration.

175 Gassert et al., "AI for Nature: How AI Can Democratize and Scale Action on Nature"; Siagian and San Andres, Using AI to Power Up Efficient and Resilient Energy Systems; KyriKyriakarakos, "Artificial Intelligence and the Energy Transition."

176 Global AI governance functions as a multi-stakeholder regulatory architecture designed to foster collective action and transnational cooperation in the pursuit of responsible AI development and deployment. Global AI governance is not only the development of ethical and legal norms to govern AI, but also has an important component of cooperation across states and across multiple stakeholders to manage AI's global challenges: Maral Niazi, "Conceptualizing Global Governance of AI," (Digital Policy Hub working paper, Waterloo: CIGI, 2023). Global AI governance should be understood as a meta-layer of coordination. It does not replace national regulation or sectoral oversight; rather, it aligns and structures them within shared international guardrails.

177 UNESCO, Recommendation on the Ethics of Artificial Intelligence (May 16, 2023), https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

178 World Economic Forum, The Presidio Recommendations on Responsible Generative AI (June 2023), https://www3.weforum.org/docs/WEF_Presidio_Recommendations_on_Responsible_Generative_AI_2023.pdf.

179 "High-Level Advisory Body on Artificial Intelligence," Office for Digital and Emerging Technologies, United Nations, https://www.un.org/digital-emerging-technologies/ai-advisory-body.

180 "Statement on Superintelligence," Superintelligence Statement.

181 Cumming, Safety Standards Delivering Controllable and Beneficial AI Tools.

182 Simon Chesterman, "The Tragedy of AI Governance," (NUS Law Working Paper No. 2023/027: October 31, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4600065.

183 Paris Agreement and the 2030 Agenda. Forum for Development Studies," Forum for Development Studies, 1–17 (January 22, 2026). https://doi.org/10.1080/08039410.2026.2618795.

184 Huw Roberts, Emmie Hine, Mariarosaria Taddeo, and Luciano Floridi, "Global AI governance: barriers and pathways forward," International Affairs 100, no. 3 (May 2024): 1275–1286, https://doi.org/10.1093/ia/iiae073.

185 Chesterman, "The Tragedy of AI Governance."

186 Bartosz Brożek, Michał Furman, Marek Jakubiec, and Bartłomiej Kucharzyk, "The Black Box Problem Revisited: Real and Imaginary Challenges for Automated Legal Decision Making," Artificial Intelligence and Law, 32 (April 4, 2024): 427–440, https://link.springer.com/article/10.1007/s10506-023-09356-9.

187 Roberts et al., "Global AI governance: barriers and pathways forward,"

188 United Nations General Assembly (UNGA), Terms of Reference and Modalities for the Establishment and Functioning of the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance, A/RES/79/325, August 26, 2025, https://docs.un.org/en/A/RES/79/325.

189 CERN - Conseil Européen pour la Recherche Nucléaire or European Council for Nuclear Research- is a particle physics laboratory that utilizes the world's largest particle accelerator, the Large Hadron Collider. It is well known for the discovery of the Higgs Boson in 2012. CERN carries out exclusively fundamental, non-military research (Kohler, 2025), and, despite being in Geneva, it is not part of the UN.It is governed by a Council of 25 member states (each with one vote), and it is based on a cost-sharing model that after World War II allowed European countries to pool resources and maintain scientific leadership in high-energy physics. For more information see: Ian Sample, "Higgs Boson Discovery: Now the Real Work Begins," The Guardian, July 4, 2012, https://www.theguardian.com/science/2012/jul/04/higgs-boson-discovery-real-work; Kevin Kohler, "CERN for AI: One Analogy, Many Visions," Simon Institute for Longterm Governance, July 8, 2025, https://simoninstitute.ch/blog/post/cern-for-ai-one-analogy-many-visions; "Our Member States," CERN, 2026, https://home.cern/about/who-we-are/our-governance/member-states.

190 "Speech by President von der Leyen at the Artificial Intelligence Action Summit," speech, February 10, 2025, https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_25_471.

191 Chesterman, "The Tragedy of AI Governance."

192 Emily O'Brien and Richard Gowan, What Makes International Agreements Work: Defining Factors for Success (Center on International Cooperation and New York University, September 2012), https://cic.nyu.edu/wp-content/uploads/1662/65/gowan_obrien_factors_success.pdf?utm_source=chatgpt.com.

193 United Nations General Assembly (UNGA), Terms of Reference and Modalities for the Establishment and Functioning of the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance, A/RES/79/325, August 26, 2025, https://docs.un.org/en/A/RES/79/325.

194 Felix Creutzig, Fatima Denton, Emmie Hine, Somya Joshi, Gong Ke, Yara Kyrychenko, Xue Lan, Dirk Messner, and Daniel W. O'Neill, "Governing Artificial Intelligence for Planetary Health," The Lancet Planetary Health (2026): 101408, https://doi.org/10.1016/j.lanplh.2025.101408.

195 Niazi, "Conceptualizing Global Governance of AI."

196 Ye Shulan and Li Mengting, "Global AI Governance: Progress, Challenges, and Prospects," China International Studies, Vol. 109 (2024): pp. 48-69, https://heinonline.org/HOL/AuthorProfile?action=edit&search_name=Mengting,%20Li&collection=journals.

197 Owen Gaffney, Amy Luers, Franklin Carrero-Martinez, Berna Oztekin-Gunaydin, Felix Creutzig, Virginia Dignum, Victor Galaz, Naoko Ishii, Francesca Larosa, Maria Leptin, and Ken Takahashi Guevara. The Earth alignment principle for artificial intelligence," Nature Sustainability 8 (2025): 467–469, https://doi-org.ezproxy.cul.columbia.edu/10.1038/s41893-025-01536-6

198 "About Montreal Protocol," UN Environmental Programme, UNEP, https://www.unep.org/ozonaction/who-we-are/about-montreal-protocol.

199 The Montreal Protocol explicitly restricted trade in ozone-depleting substances and related products with non-party states and implemented licensing systems to monitor and enforce commitments. These measures discouraged free-riding, prevented competitive advantages for non-parties, and strengthened participation and implementation of phase-out obligations (Barrett et al., 2024). For more details see: Scott Barrett, Noah Kaufman, and Joseph E. Stiglitz, "How to Enforce Climate Agreements with Trade Measures," Project Syndicate, January 25, 2024, https://www.project-syndicate.org/commentary/cop28-un-climate-change-conference-process-doesnt-work-but-narrow-binding-agreements-do-by-scott-barrett-et-al-2024-01.

Columbia Center on Sustainable Investment
Research and Development Group, Hitachi