VEMUN 2026

Guaranteeing Governance in the Global Future:What Role for Europe?

March 13 – 15, 2026

# General Assembly

# Equitable and Sustainable

# AI Development

## Chairs: Roméo Brochier and Clarence Moulin Fournier

# Table of Contents

## Introduction

The *2030 Agenda for Sustainable Development* signed by all UN members in 2015 outlined 'a shared blueprint for peace and prosperity for people and the planet'. At its heart are the 17 Sustainable Development Goals (SDGs), which are an urgent call for action by all countries to work together in a global partnership. This naturally raises the question as to whether or not Artificial Intelligence will be a net benefit for achieving all goals; how to implement it and how to prevent it from jeopardising existing solutions. The fast development of AI needs to be supported by the necessary regulatory insight and oversight for AI-based technologies to enable sustainable development. Failure to do so could result in gaps in transparency, safety, and ethical standards.

In popular media, many of the concerns surrounding AI are ethical and revolve around topics such as IP theft, copyright law and the question of "AI art". There is also debate between skeptics and proponents as to whether AI development will plateau and create an AI bubble. While these topics are interesting, within the context of a UN debate it is neither worthwhile nor pertinent to address these as those issues remain mostly subjective and localised problems. Global cooperation on sustainable and equitable AI development should instead focus on: the lack of actionable policy in using AI for development, the current economic risk of AI, and the lack of accountability and international safeguards for AI.

The UN has affirmed the fact that AI can be useful for up to 80% of the SDGs if not more. However, despite efforts to create a High-level Advisory Body on Artificial Intelligence, agencies such as UNDP have failed to provide actionable policy recommendations. While these technology transfer programs that are in place remain useful and should be further developed, more practical applications for AI are needed, notably in healthcare, agriculture, weather prediction etc...

Furthermore the growing demand of semiconductors, energy and water to sustain AI usage is beginning to make AI seem unsustainable and financially unviable. If these AI companies are truly overvalued and require too much compute then it would be irrational to encourage countries to develop AI infrastructure. Increasing international competition would both improve the quality and cost-efficiency of AI while avoiding concentration in the US and Chinese markets.

Another major concern is the lack of regulation for AI on an international scale. Currently there are many systems with more or less regulations for AI, and though there are some regional agreements ultimately almost all countries have their own AI policy. In an age where billionaires are moving away from Europe in order to avoid taxes and companies try to maximise profits by delocalising production, it’s becoming more and more important to come to widespread international agreements especially in the context of regulating AI for safety.

More recently Europe's tech sector is accelerating, with VC investment rising from $22bn to $85bn over the past decade, driven by geopolitical pressures and policy reforms. Trump's policies have inadvertently boosted Europe by reversing brain drain, redirecting VC from China, and spurring green-tech and defence-tech growth. Policymakers are unifying capital markets, encouraging pension funds to back startups, and pushing for greater tech sovereignty. Established European founders are now reinvesting wealth into new startups, creating denser innovation ecosystems. While Europe won't overtake America soon, it is becoming meaningfully competitive, particularly in climate tech, defence tech, and deep tech.

Delegates ought to find ways to tackle these 3 problems, in order to make AI a “force of good” that brings us closer to the SDGs and not further away from them.

## Definition of Key Terms

### Artificial Intelligence

- According to Oxford dictionary :

The application of computer systems able to perform tasks or produce output normally requiring human intelligence, especially by applying machine learning techniques to large collections of data.

- According to International Business Machines Corporation (IBM) :

Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy.

### Machine Learning

Machine learning involves creating models by training an algorithm to make predictions or decisions based on data. It encompasses a broad range of techniques that enable computers to learn from and make inferences based on data without being explicitly programmed for specific tasks. There are many types of machine learning techniques or algorithms, each suited to different kinds of problems and data: linear regression, logistic regression, decision trees, support vector machines (SVMs) and more.One of the most popular types of machine learning algorithm is called a neural network (or artificial neural network)

## Deep learning

Deep learning is a subset of machine learning that uses multilayered neural networks, called deep neural networks, that more closely simulate the complex decision-making power of the human brain.

## Generative AI

Generative AI, sometimes called "gen AI", refers to deep learning models that can create complex original content such as long-form text, high-quality images, realistic video or audio and more in response to a user's prompt or request. At a high level, generative models encode a simplified representation of their training data, and then draw from that representation to create new work that's similar, but not identical, to the original data.

## Chatbot

An interactive application or program designed to communicate with people through text or voice instructions in a way that mimics human-to-human conversation. Many contemporary chatbots are built on GenAI models that have been fine-tuned for this particular task from foundation large language models (LLMs).

## Compute

Compute is the capacity of a system to perform operations, process data and execute instructions. In practical terms, it can mean everything from the CPU in your laptop to the massive clusters of servers powering cloud-based AI models. The only way to keep up with the improvement of AI is scaling compute. Moore's law states that chip speed doubles each year, whereas compute scales at a rate of

4,5x every year (see appendix B). This means that while AI performance has scaled exponentially as seen in the METR graph (see appendix A), so has its cost.

## Data sovereignty

Data sovereignty is the principle that data generated within a country's borders is governed by that nation's laws and regulatory frameworks; this ensures local control over data access, storage, and usage.

## AI centers

An AI data center is a specialized data center facility designed for the computationally intensive tasks of training and running inference for artificial intelligence and machine learning models. Unlike general-purpose data centers, they are optimized for the parallel processing demands of AI workloads, typically utilizing hardware such as AI accelerators (e.g., GPUs, TPUs) and high-speed interconnects. The global push to construct these specialized facilities accelerated dramatically during the AI boom of the 2020s. This demand has reshaped supply chains, driving memory manufacturers to prioritize production of High Bandwidth Memory (HBM) essential for AI servers, which has led to a global memory supply shortage, and triggering a broader competition for advanced chips, power, and infrastructure.Major tech companies are estimated to spend $650 billion on AI data centers in 2026.

## General Overview

AI is no longer a distant prospect or a niche research interest. It is embedded in healthcare diagnostics, agricultural forecasting, financial services, education platforms, climate modelling, and national security systems. The pace of its development is without precedent. Models that once required years of research and billions of dollars to produce can now be replicated and improved upon in months. The release of DeepSeek in January 2025, a competitive large language model built at a fraction of the cost of its Western counterparts, although it caused the Nvidia stock crash, has led to a growing competition for more compute and better AIs. The technology is proliferating, and the regulation has not kept pace.

This overview is organised around the three interconnected problems that delegates are asked to grapple with: first, the gap between AI's potential and the actionable policy needed to deploy it effectively for development; second, the economic risks accompanying rapid AI expansion, including resource demands, market concentration, and labor displacement; and third, the absence of robust international accountability mechanisms and safeguards to ensure that AI serves humanity rather than undermining it.

Part I: The Gap Between Potential and Actionable Policy

It’s no secret that AI has a lot of potential. AI optimists often point to the exponentially scaling METR graph (see appendix A). This graph evaluates models based on the time humans typically take to complete tasks that the AI model can complete with a 50% success rate. The latest model, Claude Opus 4,6 can complete a task that would take a human 14 hours 30 minutes in a fraction of the time. Granted when measuring an 80% success rate this metric falls to 1 hour and 3 minutes. Nevertheless this still proves that AI has greatly improved, from being able to write emails to being able to fix bugs in Machine Learning codebases. In this respect, the currently available compute is more than sufficient and the problem comes from a lack of applications for development.

In agriculture, AI-powered satellite analysis and predictive modelling can help smaller farmers anticipate weather patterns, identify crop disease, and optimise irrigation. This technology could substantially improve food security in regions most vulnerable to climate disruption. In healthcare, machine learning models are outperforming specialists in the early detection of certain cancers, tuberculosis, and diabetic retinopathy, offering the prospect of diagnostic quality care in settings where specialist physicians are scarce. In climate science, AI is accelerating the processing of atmospheric data and enabling more granular climate modelling than was previously computationally feasible. Across each of these domains, the technology exists; what is frequently absent is the policy framework, the institutional capacity, and the investment infrastructure to bring it to scale where it is most needed.

The UN has made efforts to engage with AI governance. The Secretary-General's High-Level Advisory Body on Artificial Intelligence, established in 2023, was an important signal of institutional intent. UNESCO adopted its Recommendation on the Ethics of AI in 2021, representing the first global normative instrument on the subject. The International Telecommunication Union has convened its AI for Good platform annually, creating space for dialogue between governments, industry, and civil society. These are meaningful steps, but they stop well short of what the moment requires.

The recurring criticism of the UN's AI engagement is that it produces principles without mechanisms. The High-Level Advisory Body's mandate is advisory; it cannot compel member states or private actors to adopt its recommendations. UNESCO's ethical framework is non-binding. AI for Good creates dialogue but not delivery. Meanwhile, the agencies best positioned to translate AI into development outcomes, UNDP, WHO, FAO, UNICEF, have, with few exceptions, struggled to move beyond pilot projects and proof-of-concept deployments toward durable, scalable programs.

Closing the gap between AI's development potential and its realisation requires a different mode of engagement. It means moving from the articulation of principles to the design of programs: identifying specific, high-impact AI applications, assembling the technical and financial partnerships needed to deploy them, embedding them within existing development infrastructure, and building the local capacity to manage and adapt them over time. It means taking seriously the question of which problems AI is actually well-suited to solve, and resisting the temptation to treat AI as a universal answer to development challenges it cannot meaningfully address.

Part II: The Economic Risks of Rapid AI Expansion

However the idea that AI can scale exponentially relies on the premise that it is truly possible to have more compute. One of the most consequential and least discussed dimensions of the global AI landscape is the extraordinary concentration of capability in the hands of a very small number of private actors, overwhelmingly headquartered in the United States and, to a lesser extent, China. The development of frontier AI systems requires three inputs in vast quantities: data, compute, and talent. Each of these is distributed profoundly unequally across the world, and the dynamics of the AI industry tend to reinforce rather than counteract that inequality.

For most of the world's nations, and especially for low- and middle-income countries, this concentration creates a condition of structural dependency that has uncomfortable historical echoes. Countries that cannot independently develop or maintain competitive AI systems will find themselves relying on services provided by foreign corporations operating under foreign regulatory regimes and optimised for foreign markets. Their citizens' data will flow outward, enriching the training datasets of systems that may never be designed with their needs in mind. Their governments will make consequential decisions in domains like social welfare, law enforcement, and public health using tools they do not fully understand and cannot meaningfully audit.

This is not a hypothetical concern. It is already the operating reality in many countries where AI-powered systems supplied by global technology firms are being integrated into public services. The question for the international community is whether this dependency will be allowed to deepen, and whether the revenue flows, intellectual property rights, and governance authority that accompany AI deployment will remain concentrated in the hands of a few, or whether a more equitable architecture is possible.

The environmental cost of AI at scale is a problem that sits at the intersection of economic and ecological risk, and it bears directly on the UN's sustainable development mandate. Training and operating large AI models is intensely resource-intensive. The energy demands of AI data centres have grown dramatically as models have increased in scale, and projections suggest that continued growth along current trajectories will create significant pressures on national electricity grids and on commitments made under the Paris Agreement.

Water sources are an equally pressing concern. The cooling systems that prevent AI data centres from overheating consume enormous quantities of fresh water, a resource already under severe stress in many of the regions where large-scale data infrastructure is being built. Indeed, every time you use an AI chatbot like ChatGPT, it consumes water. For example, having ChatGPT’s latest model, GPT-4, generate a 100-word email consumes about 500ml of water (17 oz) and 10 to 50 queries consume about 2 litres of water (½ gallon). Additionally, training a model like GPT-3 in Microsoft’s US data centres can consume 5.4 million litres of water. Given that the latest model GPT-4 is larger and more advanced, it’s likely that its numbers are even higher. The global AI demand in 2027 is projected to require 4.2 and 6.6. billion cubic meters of water withdrawal, surpassing the annual water use of countries the size of Denmark, or about half of the United Kingdom’s total annual withdrawals. This hidden water cost comes from the infrastructure needed to power and cool AI systems and the manufacturing of chips and servers. What’s more, the water used often has to be drinking quality because impurities can damage the servers.

The semiconductor supply chain that underpins all AI development is concentrated in a small number of locations, Taiwan above all, and is itself environmentally intensive, requiring rare earth materials and vast quantities of ultra-pure water in its manufacturing processes.

These resource demands create a direct tension with the SDGs. SDG 7 calls for affordable and clean energy; SDG 6 for clean water and sanitation; SDG 12 for responsible consumption and production; SDG 13 for urgent climate action. AI, deployed at scale without adequate environmental governance, risks consuming resources that are needed to meet these goals, and doing so disproportionately in ways that benefit wealthy consumers and corporations while the environmental costs are borne by communities in less powerful nations. Any serious international framework for AI development must address this tension explicitly, including through transparency requirements on energy and water consumption and through standards that link AI infrastructure development to national and international climate commitments.

The economic risks of AI are not confined to questions of market structure and resource consumption. They extend to the labor market effects of automation, a dimension of the AI debate that has received extensive attention in the context of wealthy economies but that carries perhaps even greater urgency for the Global South.

The development model that allowed many countries to grow their middle classes in the late twentieth century, export-oriented manufacturing and services, leveraging comparative advantage in labor costs, is under direct challenge from AI-enabled automation. The tasks most readily automated are disproportionately those currently performed by workers in lower-wage economies: routine manufacturing, data entry, basic customer service, and increasingly, the kind of skilled knowledge work, legal research, code review, content moderation, that has underpinned the growth of outsourcing industries in countries like India, Kenya, and the Philippines.

This is not an argument against AI development. The productivity gains that AI enables are real and significant, and they have the potential to raise living standards broadly if their benefits are well distributed. But those benefits do not distribute themselves automatically. Without deliberate policy intervention, at national and international levels, the risk is that AI accelerates a divergence between those who own and deploy the technology and those whose labor it displaces. SDG 8's commitment to decent work and economic growth for all is directly implicated. The UN and its member states need frameworks that address not just how AI is built, but how its economic gains are distributed.

## Part III: The Absence of International Accountability and Safeguards

Perhaps the biggest problem of the current global AI environment is the near-total absence of binding international regulation. AI governance today is a patchwork: the European Union's AI Act represents the most comprehensive attempt at risk-based regulation, but it applies only within EU jurisdiction and is still being implemented. The United States has pursued a sectoral, largely voluntary approach, relying on executive orders and agency guidance rather than legislation. China has enacted regulations governing specific AI applications, generative AI, recommendation algorithms, but within a domestic framework that operates according to different political assumptions about state authority and individual rights. The rest of the world's nations, with few exceptions, are operating without meaningful domestic AI regulation of any kind, let alone within coordinated international frameworks.

This lack of standardisation creates conditions that create a race to the bottom, in which companies and governments seeking to avoid costly compliance migrate toward jurisdictions with the weakest standards. The comparison to tax regulation makes sense given that just as multinational corporations have historically exploited gaps between national tax regimes to minimise their obligations, AI companies have both the incentive and the legal latitude to structure their operations so as to minimise their exposure to regulatory oversight. In a domain where the most consequential decisions, about model architecture, training data, deployment constraints, and safety testing, are made entirely within private organisations with limited external accountability, this regulatory arbitrage represents a genuine systemic risk.

The absence of international safeguards is most acute in domains where AI's consequences are most severe. Autonomous weapons systems, machines capable of identifying and engaging targets without direct human authorisation, and mass surveillance are all things that the Trump administration has put in place, despite facing backlash from companies such as Anthropic. The technology to build such systems exists and is being actively developed by multiple state actors and yet no binding international treaty constrains their development or deployment. Discussions within the framework of the Convention on Certain Conventional Weapons have continued for over a decade without producing agreement. The window for preventive governance is narrowing.

Surveillance technology is also growing at an alarming rate. AI-powered facial recognition, predictive policing tools, and mass monitoring systems are being developed in high-income countries and exported, often with minimal due diligence, to governments whose human rights records would not survive scrutiny in their countries of origin. The infrastructure of digital authoritarianism is being built with AI, and the international community has not developed effective tools to prevent or constrain it.

Building effective international accountability for AI is difficult. The technology improves much faster than treaty-making processes. The states with the greatest interest in maintaining regulatory freedom, primarily the US and China, whose companies dominate the frontier, have the greatest capacity to obstruct binding international agreements. And the UN's existing institutional architecture, shaped for a world of state actors and physical goods, struggles to encompass a domain defined by private actors and digital products that cross borders invisibly and instantaneously.

And yet the precedents for international governance in comparably complex and consequential domains exist. The International Atomic Energy Agency demonstrates that binding verification and inspection regimes are possible even in domains where states have strong national security interests in maintaining autonomy. The Financial Stability Board shows how post-crisis financial regulation can be coordinated across jurisdictions without formal treaty law. Aviation safety norms, coordinated through the International Civil Aviation Organization, show how voluntary compliance with international standards can become effectively universal when market access is conditioned on it.

None of these analogies are perfect but together they suggest that the parameters of an effective international AI governance architecture are not beyond imagination. What is required is political will, and that political will, in a multilateral forum, is ultimately a function of whether the states most harmed by ungoverned AI development succeed in building coalitions capable of advancing their interests.

The three problems this debate is asked to address, the lack of actionable AI policy for development, the economic risks of rapid and unequal AI expansion, and the absence of international accountability and safeguards, are not separate issues. They are three faces of the same underlying challenge: how to govern a transformative technology so that its benefits are broadly shared and its risks are reduced.

Addressing the economic risks requires confronting the structural inequalities that AI threatens to exacerbate, including the concentration of market power, the environmental costs of scale, and the labor market disruptions that fall hardest on those least able to absorb them. Addressing the accountability deficit requires building international mechanisms that can actually constrain the behavior of powerful private and state actors, not merely exhort them.

None of this is easy. But the 2030 Agenda was not premised on the assumption that sustainable development would be easy. It was premised on the conviction that the challenges facing humanity are shared, and that shared challenges require shared responses. Artificial Intelligence is now among the most consequential shared challenges of our time.

# Major Parties Involved

## United States

The US dominates AI by almost every private-sector metric. US private AI investment hit $109.1 billion in 2024, nearly 12 times China's $9.3 billion and 24 times the UK's $4.5 billion. From 2013 to 2024, the US drew over $470 billion in private AI investment, compared to roughly $50 billion across the entire EU. On compute, the US controls an estimated 74 % of global high-end AI supercomputer capacity, while China holds 14 % and the EU 4.8 %.

The primary policy lever used against China has been semiconductor export control, by banning sales of advanced chips beginning in October 2022, tightened repeatedly through 2024. The result is that China remains a marginal producer of AI chips: Huawei produced around 600,000 AI chips in 2025, while China legally imported around 1 million Nvidia chips in 2024.

The Trump administration then revoked the Biden-era AI Diffusion Rule in 2025, stating it would have "stifled American innovation" and "undermined US diplomatic relations with dozens of countries,", and in January 2026 issued a replacement that loosens restrictions on Nvidia H200 and AMD MI325X chip exports to China. The US is also the home of the companies burning the most money: OpenAI projects a potential $14 billion in losses by 2026, and $44 billion in cumulative losses from 2023 to 2028, with positive cash flow not expected until 2029.

## China

China's approach is to become the world's leading AI nation by 2030. The US and China together control over 70% of global AI investment. China diverges sharply from the US in the structure of that investment using government guidance funds and coordinated industrial policy rather than venture capital. Chinese open models, especially Alibaba's Qwen, have become the world's most widely downloaded open LLMs, outstripping Meta's Llama in global downloads and user preference. DeepSeek's R1 model, released in January 2025, roughly matched the capabilities of models from Google, OpenAI, Meta, and Anthropic.

China's response to being cut off from Western chips has been twofold: a national effort to make China independent of Western semiconductor technology, and large-scale smuggling, with a ring of individuals who bought $390 million worth of servers containing banned Nvidia GPUs and attempted to move them via Malaysia.

At the state infrastructure level, Beijing ordered all state-funded data centers to stop purchasing foreign AI chips in November 2025, as part of a push to achieve "algorithmic sovereignty" and complete control of its computing infrastructure by 2027.

The private investment figures are lower than the US ($9.3 billion vs $109 billion in 2024), but government-channelled capital and state-owned enterprise involvement make direct comparison difficult.

## European Union

The EU's position is one of regulatory ambition paired with a real production gap. Private AI investment across EU countries from 2013 to 2024 was roughly $50 billion, less than 11% of the US figure over the same period.

European institutions produced only three notable AI models in 2024, compared to 40 from the US and 15 from China. Mistral, France's most prominent model company, had a surge of popularity in 2024 but has since been eclipsed by Chinese open models in download volume. The EU's distinctive contribution is regulatory: the AI Act, which came into force in 2024, is the world's first comprehensive AI law, classifying systems by risk level and banning specific applications like real-time biometric surveillance in public spaces. Whether this framework shapes global standards or simply raises costs for EU-based developers is genuinely unclear. The EU has responded to the investment gap with targeted spending commitments, but nothing close to matching US or Chinese scale. The bloc's 4.8% share of global high-end AI compute represents a structural weakness that spending pledges alone are unlikely to close quickly, given that data center construction and chip supply chains take years to develop.

## OpenAI

OpenAI is the highest-profile AI company in the world and, by revenue, the largest AI lab. Annualized revenue hit $10 billion by June 2025, nearly doubling from $5.5 billion in December 2024, with full-year 2025 revenue projected at $12.7 billion.

It has over 800 million weekly active users and more than 2.1 million developers on its platform. The financial model, however, remains structurally loss-making. OpenAI expects losses to continue for years, with 2028 operating losses projected at $74 billion, roughly 75% of its forecast revenue that year, and "meaningful" profitability not expected until 2030.

The core bet is that infrastructure investment now will compound into market dominance. OpenAI spent around $5 billion on R&D compute alone in 2024,, and Sam Altman has committed to $1.4 trillion in new computing deals with Nvidia and others. OpenAI's trajectory shifted in January 2025 when DeepSeek released R1, matching GPT-class performance at a fraction of the compute cost.

Recently, OpenAI CEO Sam Altman announced that the company had signed a deal with the Pentagon and the Department of War for its AI tools to be used in the military’s classified systems.

### Anthropic

Anthropic was founded in 2021 by former OpenAI researchers, including Dario and Daniela Amodei, and has structured itself around safety research as its primary differentiator. Revenue grew from $100 million in 2023 to $1 billion in 2024, reaching an annualized rate of $8–10 billion in 2025.
Unlike OpenAI, Anthropic's financial trajectory points toward profitability sooner: the company forecasts its cash burn will drop to one-third of revenue in 2026 and 9% by 2027, with profit projected by 2028. Its technical approach, "Constitutional AI," which trains models using a written set of principles rather than relying solely on human feedback, is the basis of its Claude models. In practice, the distinction between Anthropic and its competitors is real but narrowing.

### Google

Google has the most compute, the most data, the largest existing distribution network, and has nonetheless been playing catch-up in the AI race since ChatGPT launched in 2022. DeepMind (the London research lab) and Google Brain were merged in 2023 into Google DeepMind, consolidating AI research under one structure. Microsoft, Alphabet, Amazon, and Meta plan to spend a combined $320 billion on AI technologies and infrastructure in 2025, up from $230 billion in 2024.
Google's structural advantage is that it doesn't need AI to be a standalone profit center, it needs AI to defend its search monopoly and expand into enterprise software. This gives it a longer runway than either OpenAI or Anthropic, neither of which have existing profitable lines of business to subsidize model development.

## Timeline of Key Events

**1950 :** Alan Turing published “Computer Machinery and Intelligence” which proposed a test of machine intelligence called The Imitation Game.

**1952:** A computer scientist named Arthur Samuel developed a program to play checkers, which is the first to ever learn the game independently.

**1956 :** The field of AI research was founded at a workshop at Dartmouth College, organised by John Mcarthy. The Dartmouth workshop defined AI as "the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it".

**1966 :** Created by the MIT computer scientist Joseph Weizenbaum , ELIZA is widely considered the first chatbot and was intended to simulate therapy.

**1974-1980** : The first AI winter ( period of reduced funding and interest in AI research ) resulted from overpromising, technical limitations and criticism. Both the U.S. and British governments cut off exploratory research in response to criticism and ongoing pressure from the U.S. Congress to fund more productive projects.

**1980–1987 :** The AI boom; this era was marked by expert systems and increased funding in the 80s. The development of Cog, iRobot, and Roomba by Rodney Brooks and the creation of Gammonoid, the first software to win backgammon against a world champion both took place during this period.

**1986** : Ernst Dickmanns, a scientist working in Germany, invents the first self-driving car. Technically a Mercedes van that had been outfitted with a computer system and sensors to read the environment, the vehicle could only drive on roads without other cars and passengers.

**1987–1993 :** The Second AI Winter. Beginning in 1987, funds once again dried up for several years. An expert system requires a lot of data to create the knowledge base used by its inference engine, and unfortunately, storage was expensive in the 1980s. Computers had at most 44MB of storage in 1986. For comparison, a 3-minute MP3 music file is around 30MB. Companies also realized that they could use far cheaper hardware with less-intelligent systems but still obtain similar business outcomes

**1997:** Deep Blue (developed by IBM) beat the world chess champion, Gary Kasparov, in a highly-publicized match, becoming the first program to beat a human chess champion.

**2000:** Professor Cynthia Breazeal developed the first robot that could simulate human emotions with its face, which included eyes, eyebrows, ears, and a mouth. It was called Kismet.

**2006:** Companies such as Twitter, Facebook, and Netflix started utilizing AI as a part of their advertising and user experience algorithms.

**2011:** An NLP computer programmed to answer questions named Watson (created by IBM) won Jeopardy against two former champions in a televised game.

**2011:** Apple released Siri, the first popular virtual assistant.

**2015:** Elon Musk, Stephen Hawking, and Steve Wozniak (and over 3,000 others) signed an open letter to the worlds' government systems banning the development of (and later, use of) autonomous weapons for purposes of war.

**2016:** Hanson Robotics created a humanoid robot named Sophia, who became known as the first "robot citizen" and was the first robot created with a realistic human appearance and the ability to see and replicate emotions, as well as to communicate.

**2017:** Facebook programmed two AI chatbots to converse and learn how to negotiate, but as they went back and forth they ended up forgoing English and developing their own language, completely autonomously.

**2018:** A Chinese tech group called Alibaba's language-processing AI beat human intellect on a Stanford reading and comprehension test.

**2020:** OpenAI started beta testing GPT-3, a model that uses Deep Learning to create code, poetry, and other such language and writing tasks. While not the first of its kind, it is the first that creates content almost indistinguishable from those created by humans.

**2021:** OpenAI developed DALL-E, which can process and understand images enough to produce accurate captions, moving AI one step closer to understanding the visual world.

**2023 –** The momentum of generative AI continued unabated. In February 2023, Meta released **LLaMA**, a suite of open-source large language models, signaling the tech community's push toward more accessible AI research. March 2023 brought the much-anticipated **GPT-4** from OpenAI. Around the same time, Google launched **Bard**. The proliferation of capable LLMs from multiple players illustrated an intensifying race in AI. In May 2023, concerns about AI's rapid progress led dozens of AI scientists and tech leaders – including Geoffrey Hinton, Sam Altman, and Bill Gates – to sign a **Statement on AI Risk**, warning that AI could pose societal-scale risks and calling for proactive governance.

**2024 :** In 2024, the focus shifted to refining models and expanding their modalities. The open-source AI movement gained further steam: in January, Stability AI introduced StableLM 2, an improved open large language model. In February, Google rolled out Gemini 1.5 in limited beta, showing iterative progress in its flagship model. Around the same time, OpenAI previewed a breakthrough in AI-generated media with Sora, a text-to-video model capable of creating short video clips from written prompts (an early glimpse at the future of AI in entertainment and content creation). Even Apple, usually secretive in AI, made waves in April 2024 by unveiling OpenELM, a suite of small open-source language models designed to run efficiently on-device. This was a notable departure for Apple toward more open AI research. In May 2024, enterprise software joined the AI party – Red Hat announced RHEL for AI, integrating AI development tools into its enterprise Linux platform to facilitate AI workloads for businesses. By mid-2024, major tech companies were building generative AI into their core products (for instance, Microsoft's 365 Copilot and Google's Workspace AI assistants), and I became a strategic priority in virtually every industry.

**2025 :** AI's evolution shows no signs of slowing. Generative AI has become an integral part of work and daily life, moving from a novel tool to a ubiquitous presence. Businesses have rapidly adopted AI-powered solutions – in fact, approximately 78% of organizations reported using AI in 2024, up from 55% just a year before. This year has also given us a first glimpse of more autonomous AI agents that can take initiative in performing tasks. For example, cutting-edge personal assistant AIs can now execute multi-step jobs like managing your calendar, booking appointments, or ordering products online based on a simple request, checking in with you only as needed. These agents are more advanced than previous chatbots – some function almost like junior employees or researchers. Such use cases suggest we are transitioning from AI as a passive tool to AI as an active collaborator. Tech companies are also hard at work on the next generation of foundation models, aiming to achieve greater reasoning abilities and multimodal integration (combining text, vision, and even audio). While these have yet to be publicly released, expectations are high that they could push AI even closer to human-like understanding.

# Previous Attempts to Resolve the Issue

## The (minor) use of AI in various industries

Agriculture

The UN Food and Agriculture Organization (FAO) has used AI-driven tools across several regions, most notably through its Hand-in-Hand Initiative, which uses geospatial data and machine learning to map agricultural vulnerabilities and guide investment in food-insecure nations. However, these programs have remained fragmented and donor-dependent, with limited integration into national agricultural policy frameworks and inadequate provision for local technical capacity once pilot funding ends. Perhaps integrating AI into already existing FAO initiatives would be more efficient to save costs.

Healthcare

AI has 2 main purposes when it comes to healthcare: early disease detection and speeding up administrative tasks. In practice, the most significant deployments have come through bilateral partnerships and NGO programs rather than multilateral coordination: AI diagnostic tools for tuberculosis and cervical cancer screening have been trialled in sub-Saharan Africa and South Asia with measurable improvements in detection rates. According to the European commission: AI can facilitate the efficient allocation of healthcare resources. Predictive modelling can forecast patient admissions and optimise the use of hospital beds, staff, and equipment. This ensures resources are available where and when needed most, reducing waste and enhancing the quality of care. Indeed if an AI can write a report to make patient discharge faster it would be significantly easier

Climate

The UN Environment Programme and WMO have both invested in AI applications for climate monitoring, including machine learning models that improve the resolution and accuracy of precipitation forecasting in data-sparse regions. The Climate Change AI initiative, a cross-institutional research network, has catalogued over 80 domains where AI can meaningfully contribute to mitigation and adaptation. At the policy level, however, AI has been largely absent from the formal UNFCCC negotiating process, it does not feature as a technology in nationally determined contributions, there are no agreed methodologies for accounting for AI's own emissions, and no international standard exists for the energy transparency of AI systems. The gap between research enthusiasm and binding policy commitment is, in this domain, particularly stark.

### Independent International Scientific panel & Global Dialogue on AI Governance

On 26 August 2025, the United Nations General Assembly established the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance in Resolution A/RES/79/325, following intergovernmental negotiations and broad consultations with diverse stakeholders. The UN says the panel will act "as an early-warning system and evidence engine, helping distinguish between hype and reality" and produce "policy-relevant" reports. These policy recommendations are to be presented at the Global Dialogue on AI governance, which is meant to ensure more global representation. While these two mechanisms level the playing field by ensuring global representation and reports that provide information to nations with little infrastructure, they do not openly advocate for more technology transfers nor encourage low and middle income countries to invest in AI infrastructure.

### Regional Regulation

The most substantive regulatory attempt to date is the European Union's AI Act, which entered into force in 2024 and establishes a tiered, risk-based framework: outright bans on certain applications such as social scoring and real-time biometric surveillance in public spaces, stringent requirements for high-risk systems in areas like healthcare and law enforcement, and lighter obligations for lower-risk applications. The Act is significant as the first comprehensive binding AI law in the world and has begun to exert a Brussels Effect, shaping the compliance behaviour of global companies who cannot afford to exit the European market. Canada's proposed Artificial Intelligence and Data Act and Brazil's AI framework legislation represent further regional efforts, though both remain less advanced in implementation. The critical limitation of all these frameworks is their jurisdictional confinement: they protect citizens within their borders but do nothing to govern AI deployments in the many countries that lack comparable domestic legislation.

## Possible Solutions

### A Global Compute Commons

The most fundamental inequality in AI development is not access to models but access to the processing power needed to train and run them. A UN-administered compute commons, modelled loosely on CERN, where member states pool contributions to fund shared scientific infrastructure, could provide researchers, universities, and public institutions in the Global South with guaranteed allocations of processing capacity. This would not require building new data centres from scratch; it could be operationalised through negotiated agreements with existing cloud providers, with member state contributions subsidising access for qualifying institutions. Crucially, compute access would be conditioned on open publication of research outputs, preventing the commons from becoming a subsidy for private actors.

### AI Procurement Standards

A significant and underused lever is public procurement. Governments worldwide spend trillions annually on goods and services, and AI systems are an increasingly large share of that spending. A multilateral agreement on AI procurement standards, requiring that any AI system purchased by a government receiving development assistance or operating within a UN-affiliated program meet minimum transparency, auditability, and bias-testing requirements, would create powerful market incentives for responsible development without requiring binding legislation in every jurisdiction. It would also directly address the accountability gap in high-stakes domains like welfare, immigration, and law enforcement, where governments are currently deploying commercial AI systems with minimal oversight.

### Mandatory Environmental Disclosure for AI Systems

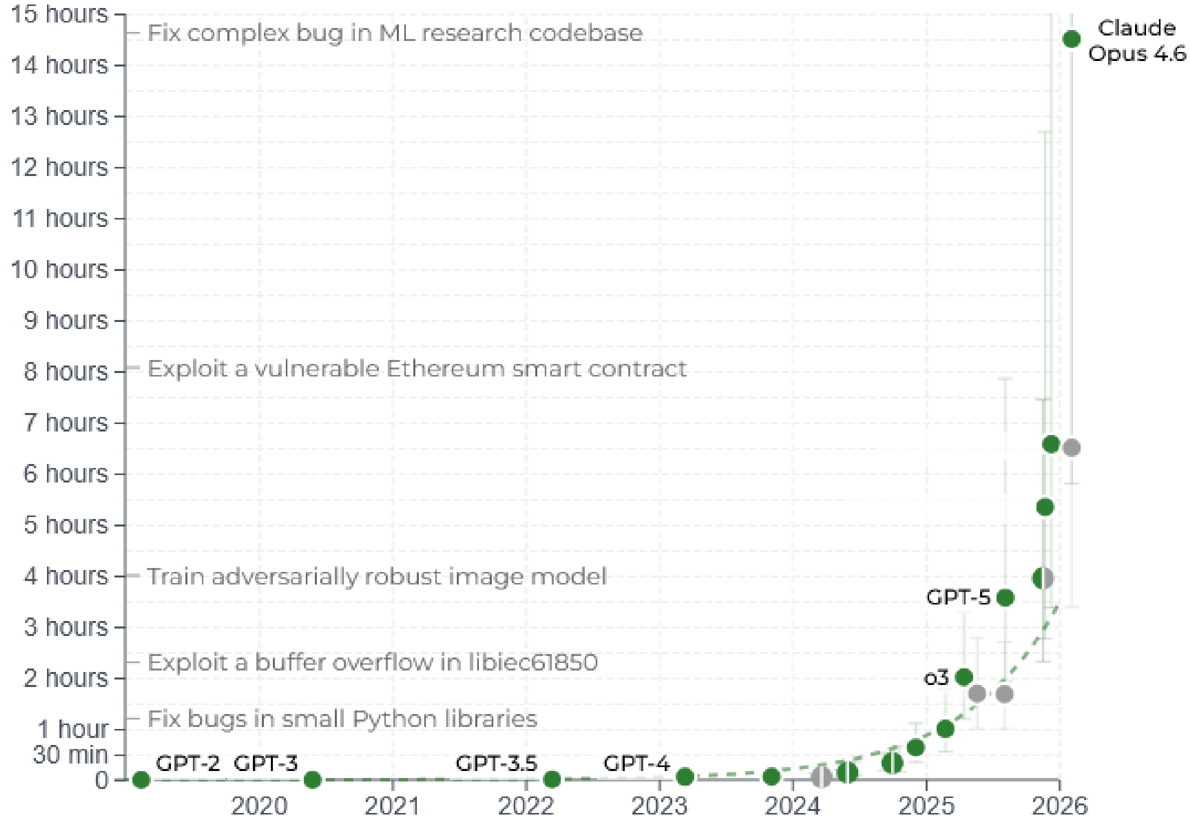
No international standard currently requires AI developers to disclose the energy or water consumption of their systems. A binding disclosure framework, potentially administered through the UNFCCC or a dedicated UN body, would require companies above a defined scale threshold to publish standardised environmental impact data for each model they deploy commercially. This serves two functions simultaneously: it creates accountability for AI's environmental footprint, and it generates the data needed to make informed regulatory decisions. Disclosure alone does not cap consumption, but it changes the political and market calculus significantly. Companies whose environmental costs become legible can be taxed, incentivised, or regulated accordingly.

### A High-Stakes AI Moratorium Mechanism

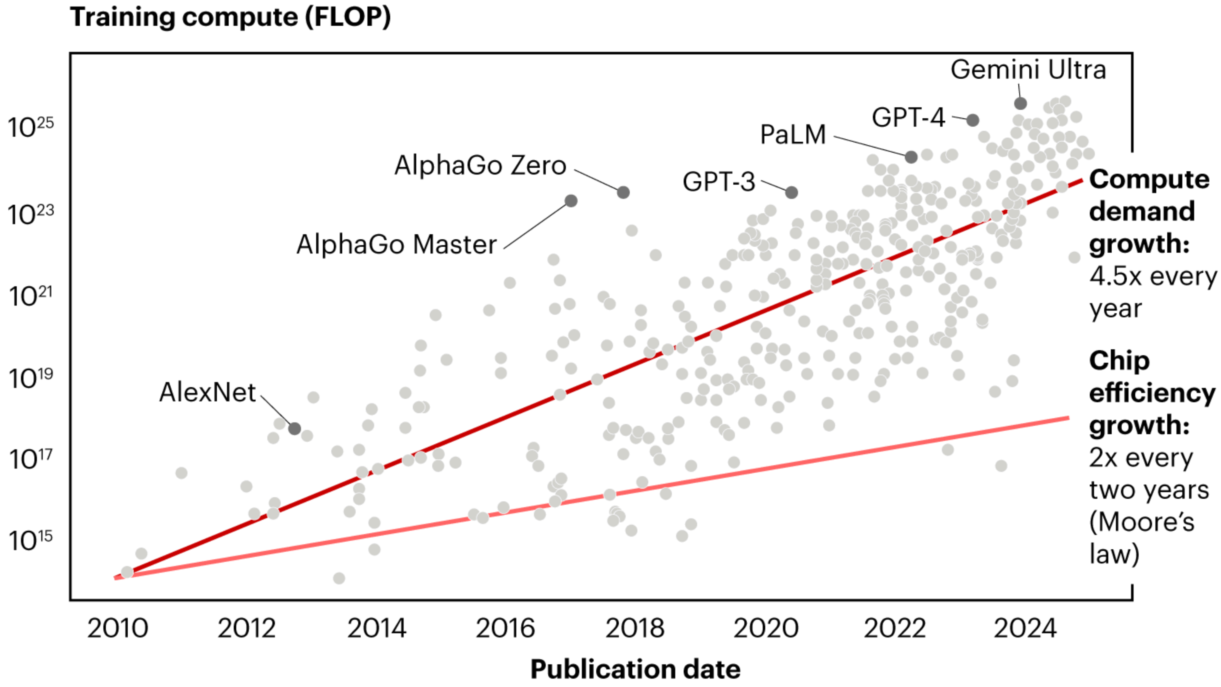
For the domains where AI poses the most acute risks, autonomous lethal weapons, AI-enabled mass surveillance, neurological manipulation, governance requires not just regulation but the capacity for a collective pause. A moratorium mechanism, modelled on the Montreal Protocol's approach to ozone-depleting substances, would allow a qualified majority of UN member states to trigger a binding halt on the development or export of specified AI applications while an international scientific and ethical review is conducted. This would require significant political will to establish, and would face determined opposition from states and companies with commercial or military interests in those applications. But it would create a credible deterrent and a structured process for the most dangerous edge cases, something the international community currently lacks entirely.

# Appendices

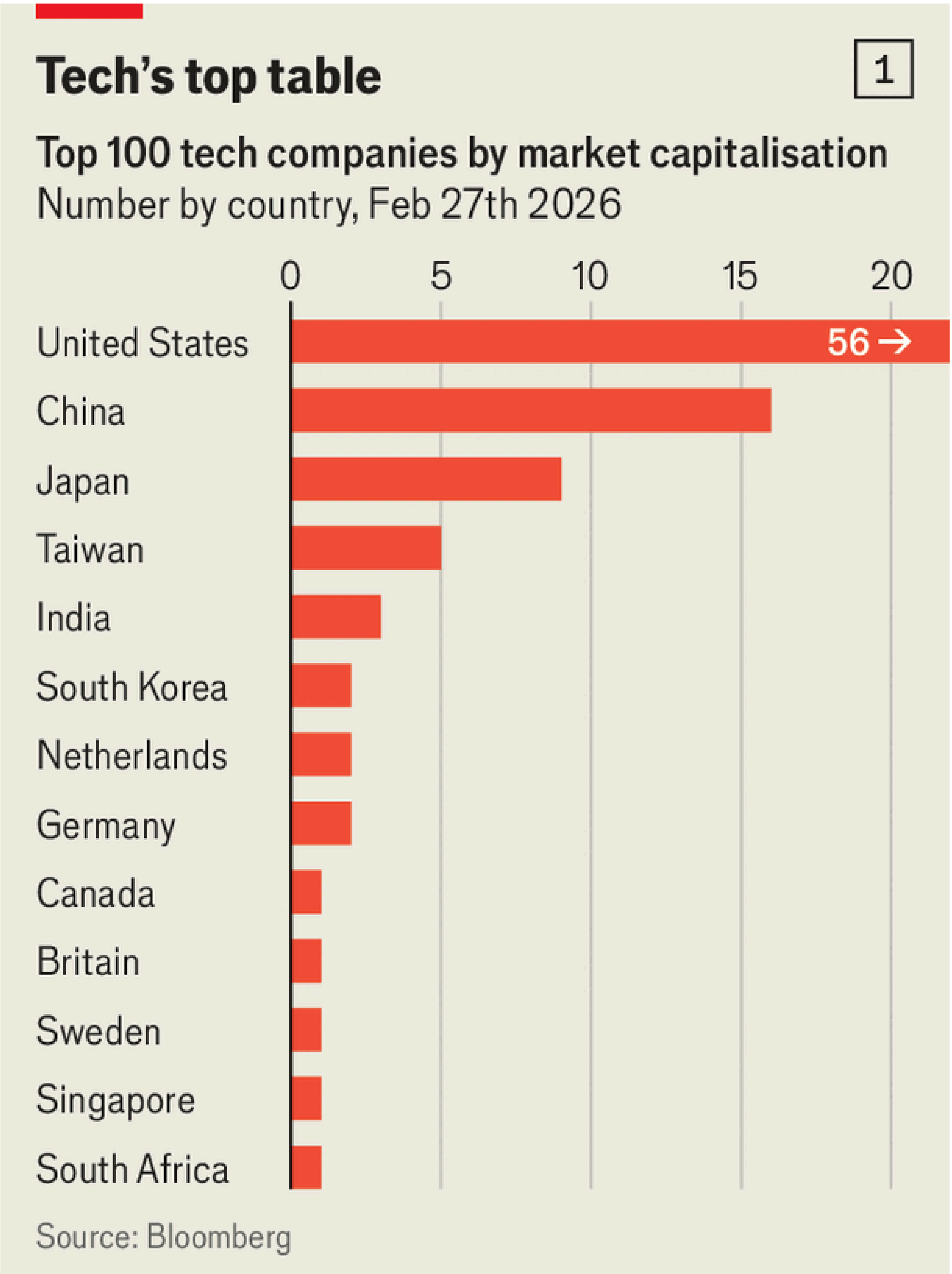
## Appendix A: METR graph



## Appendix B: Moore's law comparison



Appendix C : Top 100 tech companies by market capitalisation



## Bibliography

United Nations (2024). GOVERNING AI FOR HUMANITY. [online] Available at: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.

*Different* AI systems around the worlds https://gdprlocal.com/ai-regulations-around-the-world/

Kwa, T., West, B., Becker, J., Deng, A., Garcia, K., Hasin, M., Jawhar, S., Kinniment, M., Rush, N., Arx, V., Bloom, R., Broadley, T., Du, H., Goodrich, B., Jurkovic, N., Miles, L.H., Nix, S., Lin, T., Parikh, N. and Rein, D. (2025). Measuring AI Ability to Complete Long Tasks. [online] arXiv.org. Available at: https://arxiv.org/abs/2503.14499

Crawford, D., Schallehn, M., Renno, P., Hanbury, P. and Alessandro Cannarsi (2025). How Can We Meet AI's Insatiable Demand for Compute Power? [online] Bain. Available at: https://www.bain.com/insights/how-can-we-meet-ais-insatiable-demand-for-compute-power-technology-report-2025/.

Gibney, E. (2026). UN creates new scientific AI advisory panel: what will it do? Nature. [online] https://www.nature.com/articles/d41586-026-00542-8

https://artificialintelligenceact.eu/

https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

https://wmo.int/news/media-centre/world-meteorological-congress-endorses-actions-promote-ai-forecasts-and-warnings

https://unfccc.int/ttclear/misc_/StaticFiles/gnwoerk_static/AI4climateaction/28da5d97d7824d16b7f68a225c0e3493/a4553e8f70f74be3bc37c929b73d9974.pdf

https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(24)00571-6/fulltext

https://www.thinkglobalhealth.org/article/singapore-to-africa-ai-diagnostics-for-a-healthier-future

https://www.fao.org/hih-geospatial-platform/en