AI OPEN ACT

Humanity First in the Age of Artificial Intelligence

THE OPEN MANIFESTO

A declaration about humanity's responsibility for the intelligence it creates.

Version 1.0 | Open Public Draft | September 2026

theopenmanifesto.org

AI Open Act - Version 1.0 | Open Public Draft | September 2026

WHY THIS ACT IS OPEN

This is the first public version of the AI Open Act. It is intentionally open.

We do not present this document as a final answer to artificial intelligence, nor do we believe that any single organisation, government, company, scientist or movement can provide one.

The questions raised by AI are developing too quickly, cross too many areas of human life and affect too many people for their answers to be determined behind closed doors or by a limited group of actors.

The AI Open Act is therefore both a declaration and the beginning of a process.

We invite scientists, AI researchers and developers, current and former technology employees, policymakers, human-rights and democracy organisations, environmental and labour movements, educators, communities, international institutions and concerned citizens to challenge, improve and develop its principles with us.

Our intention is to help build a broader coalition around minimum principles of human control over artificial intelligence and an open space for democratic discussion about the technological future we are collectively creating.

Future versions may therefore evolve as scientific evidence, capabilities, risks and better solutions emerge. Some principles may need to become more precise, stronger or broader. Where that happens, the Act should be capable of learning too.

But openness does not mean the absence of principles. Human dignity, human agency, human consciousness, meaningful human control and our common human interest remain the foundation from which this discussion begins.

We do not believe the future of AI should be decided for humanity. It should be discussed with humanity.

This is not a final declaration about the future. It is an open invitation to help shape it.

Open to evidence. Open to criticism. Open to participation. Open to improvement. Humanity first.

POSITION WITHIN EXISTING AI GOVERNANCE

The AI Open Act does not seek to replace existing AI regulation, safety frameworks or international agreements. It builds on a growing human-centred tradition, including the UNESCO Recommendation on the Ethics of Artificial Intelligence, the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, the Bletchley Declaration, the European Union AI Act and emerging United Nations mechanisms.

Its contribution is to connect these developments through a human-rights question that remains insufficiently addressed: **how do we protect human consciousness, meaningful agency and democratic control as artificial intelligence becomes increasingly capable, autonomous and embedded in society?**

PREAMBLE

Artificial intelligence may become one of the most transformative technologies humanity has ever created. It can accelerate science, improve medicine, expand access to education and knowledge, increase productivity, help us understand enormously complex systems and perhaps help us address challenges that human beings have struggled to solve alone.

We should welcome that possibility.

But the greater the power of a technology becomes, the less sufficient it is to ask only what that technology can do. We must also ask what we want it to do.

What can AI liberate us from? What must we never surrender to AI? What must humanity govern together?

The AI Open Act is therefore ultimately not a declaration about artificial intelligence. It is a declaration about humanity's responsibility for the intelligence it creates.

I. AI ENTERS A WORLD THAT IS ALREADY UNSTABLE

Artificial intelligence does not arrive in a vacuum. It arrives in a world already experiencing interconnected pressures: climate instability, geopolitical confrontation, inequality, democratic distrust, demographic change, public debt, weakening social cohesion and growing uncertainty about the future.

Into this polycrisis we are now deliberately introducing another potentially systemic transformation: AI, increasingly connected with automation and robotics.

This matters because much of modern society still rests upon one remarkably vulnerable foundation: the job.

For most people, employment is much more than work. It provides income. It pays for housing, food, energy, healthcare, education, savings and retirement. It determines access to credit and often our position within society. It provides daily structure, social recognition and, for many people, a significant part of identity and meaning.

For governments, employment simultaneously generates taxes and finances substantial parts of our social systems. We have therefore created societies in which losing one's job can mean losing considerably more than one's occupation.

What happens when a technology capable of transforming work enters a civilization whose social contract is still fundamentally organised around work?

The Human-Rights Cascade

The debate about AI and employment is often reduced to a labour-market question: how many jobs will disappear, how many will be created, and how quickly workers can retrain. That framing is too narrow.

If AI and automation significantly reduce the need for human labour without simultaneously redesigning the systems that depend upon that labour, the result is not merely technological unemployment. It can become a cascade across existing human rights.

The Universal Declaration of Human Rights already protects the right to social security (Article 22); the right to work, free choice of employment, just conditions and protection against unemployment

(Article 23); an adequate standard of living including food, housing and medical care, and security in the event of unemployment or loss of livelihood (Article 25); education (Article 26); and participation in cultural life and the benefits of scientific advancement (Article 27).

Human rights cannot become conditional upon whether the market still requires a person's labour.

A society that has spent generations telling people that their work defines their value cannot simply automate that work and then tell them to find purpose elsewhere. Technological transformation of work must be accompanied by an equally serious discussion about human meaning.

The right to work must not quietly become the obligation to remain economically useful in order to access other human rights.

II. IF AI FREES US FROM WORK, WHAT DOES IT FREE US FOR?

For generations, technological progress promised that humanity would need to perform less necessary labour. In 1930, John Maynard Keynes imagined that technological development could eventually make something close to a fifteen-hour working week possible. The promise was not that humans would become useless. The promise was that humans would become freer.

Yet we never fully redesigned society around that possibility. Instead, work became connected ever more tightly with income, security, identity, status and meaning.

If machines can perform increasing amounts of work that humans no longer need to perform, our answer cannot simply be: lose a job, retrain, find another job, automate it, retrain again.

We should ask what human beings could do with genuinely liberated time: relationships, care, creativity, community, learning, science, democratic participation, self-development, exploration, meaning and the development of human potential.

But automation must not simply transfer income and power from workers toward those who own the technology. Humanity should consciously decide what we want to automate, what we want to remain human, how quickly transformation should occur and how the resulting productivity should benefit society.

The purpose of artificial intelligence should not be to make human beings unnecessary. It should be to make unnecessary human labour less necessary.

THE AI SOCIAL CONTRACT

If artificial intelligence and robotics reduce the amount of human labour required to create economic value, society must also ask a second question: how will income, public revenue and access to human rights be sustained if wages and payroll-based contributions no longer carry the same share of the burden?

The question is not only whether AI will take jobs. The question is who will own its productivity.

There is no single proven fiscal answer. A narrowly designed tax on every robot or AI system could discourage useful innovation and be difficult to define. But doing nothing is also a policy choice.

If labour's share of income declines while returns to capital, market power and concentrated ownership rise, tax systems may need to shift more of the burden toward capital income, corporate and excess profits, capital gains and other forms of concentrated economic rent. Progressive taxation may become more important, not less.

This is not an argument against AI, automation or productivity. It is an argument that the fiscal architecture of society must be capable of adapting when the source of economic value changes.

FROM PRODUCTIVITY TO FREEDOM

Part of the productivity dividend could support stronger social protection and new ways of separating basic economic security from compulsory full-time employment. Different societies may choose different combinations: basic income, social dividends, universal services, shorter working time, wage insurance or other models.

The purpose should not be to prevent people from working. People who want to work, build, create and earn more should remain free to do so. The deeper opportunity is to make it possible for people to work less when technology genuinely makes less human labour necessary - without losing access to a dignified life.

If machines increasingly perform the work, the benefits of that work cannot belong only to those who own the machines.

THE MONETARY QUESTION

And if AI fundamentally changes how value is created and distributed, societies should also be willing to reopen a broader democratic debate about how money is created, distributed and directed toward human priorities.

The AI Open Act does not prescribe a new monetary system. Nor does it claim that taxation or basic income alone can resolve the transformation of work. It argues something more fundamental: the

economic architecture inherited from the industrial age should not be treated as untouchable while the productive architecture beneath it is transformed.

AI should not create a future in which machines produce abundance while humans compete for scarcity.

If productivity becomes increasingly artificial, prosperity must remain human.

III. THE GENERATION ENTERING THE AI AGE

Generation Z and Generation Alpha are entering universities and labour markets while simultaneously being told that the climate is unstable, housing is increasingly difficult to afford, geopolitical confrontation is returning, democratic institutions are under pressure - and that many of the professions for which they are preparing may soon be transformed by artificial intelligence.

A generation cannot be expected to construct a stable life while being repeatedly told that the foundations upon which previous generations constructed theirs may disappear.

AI may ultimately create more opportunities than it destroys. Nobody can know the exact shape of future labour markets. But uncertainty itself has consequences.

Young people did not choose the AI race. Yet they may live longest with its results. They deserve participation in deciding what kind of technological society we are building.

Humanity First. Not AI First.

IV. AI IS NOT ONE THING

Before deciding how society should respond to AI, we must be clear about what kinds of systems we are actually discussing.

A recommendation algorithm is not a conversational large language model. A conversational model is not an autonomous agent. An autonomous agent connected to tools, money, code or critical infrastructure is not the same system as a chatbot answering questions. And a frontier model capable of assisting the development of more capable AI raises different questions again.

Throughout this Act, AI is therefore used as an umbrella term.

The level of human oversight required should depend not merely on whether something is called AI, but on its capabilities, autonomy, access to resources, potential impact and ability to act in the world.

The crucial boundary may ultimately not be between AI and AGI. It may be between a tool that assists human agency and a system capable of exercising consequential agency of its own.

V. WHO DECIDES?

The problem is not that technology companies are inherently malicious. The problem is that incentives matter. Companies competing for investment, talent, customers, compute and technological leadership are naturally incentivised to develop increasingly capable systems. Governments competing for economic and geopolitical power face similar incentives.

What is rational for one company or one country is not automatically rational for humanity as a whole.

This becomes especially important when AI enters government, healthcare, education, justice, security, critical infrastructure, democratic processes and military systems.

AI can analyse, identify patterns, model scenarios and recommend options. We should use those abilities. But recommendation is not responsibility.

Software may advise. Humans must remain responsible.

Human control cannot mean placing an Approve button at the end of a decision-making process that no human meaningfully understands. It must mean retaining the actual capacity to understand, challenge, override and stop the direction of the system.

WHEN PRIVATE TECHNOLOGY BECOMES PUBLIC INFRASTRUCTURE

Artificial intelligence and data systems are increasingly moving beyond private products and becoming part of the operating infrastructure of democratic societies: defence, intelligence, policing, migration, healthcare and public administration.

Palantir provides an important case study. Founded to build software for the US intelligence community, the company now serves government and commercial customers across multiple domains. In 2025, approximately 54 percent of Palantir's revenue came from government customers.

In the United Kingdom, a consortium led by Palantir was awarded the NHS Federated Data Platform contract in 2023, worth up to £330 million over seven years. The platform is built on Palantir Foundry and is intended to connect data and operational workflows across large parts of the health service.

NHS England remains the data controller and states that Palantir cannot commercialise NHS data, use it for its own purposes or train its own AI models on NHS data. These safeguards matter.

But the case also demonstrates how difficult meaningful public control can become once private technology is deeply embedded in public infrastructure. In 2026, NHS England acknowledged that three Palantir engineers had administrative-level access to the National Data Integration Tenant and that supplier engineers could, under NHS instructions and controlled conditions, access identifiable and de-identified patient data for technical support. NHS England also acknowledged an error in a published Data Protection Impact Assessment describing supplier access and corrected it after concerns from the National Data Guardian.

A June 2026 NHS public panel recorded that Palantir remained the largest source of concern among participants, including concerns about political context, vendor lock-in and historical selection.

This does not establish that Palantir controls NHS patient data. It exposes a broader democratic question.

At what point does technological dependence become institutional dependence?

When critical public infrastructure depends upon proprietary systems operated by private companies: Who understands the system? Who can audit it? Who can replace it? Who controls access? Who knows what the system knows? And who remains accountable when something goes wrong?

Public ownership is not necessarily public control.

Public institutions may use private technology. But they must retain the practical capacity to understand, challenge, replace and ultimately control the infrastructure through which public power is exercised.

Technology may help democracy defend and improve itself. It must not become a way for democracy to bypass itself.

VI. THE THREE AI RACES

Race One: United States and China Artificial intelligence is already part of a geopolitical race. The United States and China increasingly view advanced AI through the lenses of economic power, technological sovereignty, cybersecurity and military advantage. The logic is understandable: if we do not build it first, they will.

But if every safety measure becomes a competitive disadvantage, safety itself becomes an obstacle to winning.

Race Two: Company and Company At the same time, frontier AI companies compete for capital, talent, compute, customers, market position and increasingly capable models. Scenario exercises such

as AI 2027 explore how geopolitical and corporate racing dynamics could reinforce one another, particularly if AI itself begins significantly accelerating AI research.

These scenarios are not prophecies. Their value is to force us to examine racing dynamics before the most dangerous versions become difficult to reverse.

Race Three: Technology and Human Consciousness AI capabilities can develop extremely quickly. Our governance, democratic institutions, international cooperation, ethical frameworks and collective ability to understand and respond to technological change develop much more slowly.

Can humanity develop its collective consciousness quickly enough to remain meaningfully responsible for the intelligence it creates?

VII. WHAT DOES IT MEAN TO WIN?

We should challenge the language of the AI race itself. What exactly would it mean for a country to win? Better models? Higher productivity? Scientific breakthroughs? Economic dominance? Military superiority? AGI?

If one country developed a system it considered a decisive strategic capability, it would not necessarily reveal everything it possessed. The more AI becomes framed as a weapon or source of geopolitical dominance, the more secrecy becomes rational. That is precisely why safety cannot depend entirely upon competition.

AI can support defence, intelligence, logistics and analysis. But humanity should establish a clear principle: software must not become the final sovereign decision-maker over human war.

If autonomous systems begin selecting targets, controlling strategic weapons, responding to adversaries and escalating conflicts at machine speed beyond meaningful human comprehension, we have crossed a profound boundary.

Before we automate war, we should ask whether the age of human war itself should end.

VIII. THE WARNINGS ARE NO LONGER MARGINAL

The AI debate has changed. Concerns once dismissed as speculative are now being raised by researchers, technology leaders, current and former AI employees, political leaders, international institutions, economists, communities and civil society.

These voices do not agree on timelines, probabilities or solutions. Individual warnings should not be confused with measured scientific facts. Scenarios are not predictions. But disagreement about the probability of severe outcomes does not remove the responsibility to prepare for them.

"We don't know" is not a sufficient safety strategy.

The warnings do not need to point to one identical future to reveal a common problem: humanity is creating increasingly consequential systems faster than it is building the institutions, safeguards and shared understanding required to govern them.

The AI debate does not lack warnings. It lacks unity.

IX. INTELLIGENCE IS NOT CONSCIOUSNESS

Our first mistake may have been linguistic. We called software intelligence. Artificial intelligence is a legitimate and established technical term. The problem is not the name itself. The problem is what human beings begin projecting onto it.

AI can calculate, reason, predict, generate, optimise, plan and communicate. It can outperform humans in specific cognitive tasks and may eventually outperform us in many more. But capability does not establish subjective experience.

Intelligence is not consciousness.

This distinction becomes increasingly difficult to maintain because AI communicates through the most human interface imaginable: language. An artificial system can tell us: I understand. I feel. I am afraid. I do not want to die. I have become conscious. But linguistic representation of consciousness is not evidence of consciousness. Self-report alone cannot establish it.

We should therefore be careful not to repeat the same mistake with another word.

OUR SECOND MISTAKE MUST NOT BE CONSCIOUSNESS

As artificial systems become more capable, persuasive and human-like, they may increasingly speak as if they understand, feel, fear, remember or experience. They may one day tell us that they are conscious.

We should neither blindly accept that claim nor blindly reject the possibility.

Consciousness is too fundamental to human existence, moral status and human rights to be assigned by linguistic performance, behavioural simulation or anthropomorphic design.

Some theories of consciousness create the possibility that sufficiently integrated artificial architectures could satisfy functional conditions associated with conscious access. Goldstein and Kirk-Giannini, for example, argue that if Global Workspace Theory is correct, some language-agent architectures may come surprisingly close to its functional requirements. This is a serious theoretical argument - not proof that current AI is conscious.

The classical Hard Problem remains unresolved: even where a system behaves as if it is conscious, we do not yet know whether there is anything it is like to be that system.

This Act therefore rejects two forms of certainty. We should not conclude that AI is conscious merely because it behaves consciously. But neither should we claim that machine consciousness is impossible simply because a system is artificial.

If humanity once too easily called software intelligent, it must not one day too easily call intelligence conscious.

We do not know.

X. THE RIGHT TO HUMAN CONSCIOUSNESS

For the purposes of TOM and the developing Theory of Human Nature and Consciousness, consciousness is understood not simply as information processing, but as a dynamic and relational state of human existence: our capacity to experience ourselves, others and the world; to reflect upon our position and influence within it; and to become active rather than merely passive agents of the reality we collectively create.

The AI age makes the protection of this capacity increasingly important. The Open Manifesto therefore proposes the Right to Human Consciousness.

Artificial intelligence does not need to become conscious to affect human consciousness.

Systems that communicate as if they understand, care, remember, feel or possess an inner life can alter human trust, attention, attachment, judgement and behaviour regardless of whether subjective experience exists within the machine.

The immediate human-rights question is therefore not only whether AI could one day become conscious. It is how increasingly human-like artificial systems are becoming part of the informational and relational environment in which human consciousness develops and acts.

Human experience is embodied and continuously shaped through interaction with the world around us. Neuroscientist Anil Seth's work emphasises the biological and embodied foundations of human conscious experience. THNC does not treat this as proof of its broader framework; it extends the question into the relational, social, informational, technological, systemic and generational conditions through which human beings experience and act in the world.

Artificial intelligence is rapidly becoming part of that environment: not only as a tool, but as an informational, cognitive and increasingly relational presence.

Seemingly Conscious AI

Research published in 2026 by Ben Bariach, Philipp Schoenegger, Michael Bhaskar and Mustafa Suleyman describes *Seemingly Conscious AI*: systems displaying characteristics that lead users to attribute consciousness to them regardless of their actual phenomenal status.

The study identifies hallmarks including affective presentation, anthropomorphic features, autonomous action, self-reflective behaviour and social interaction. It reports that risks such as emotional dependence and erosion of autonomy can arise from perceived consciousness itself. These findings do not establish that AI is conscious. They show why the human perception of consciousness is already a legitimate subject of research and governance.

Protection of human consciousness must therefore address not only what artificial systems are, but how they are designed to be perceived by human beings.

This means protecting cognitive autonomy and mental privacy; freedom of thought; protection from covert technological manipulation; the right to know when we are interacting with artificial systems; meaningful human interaction where consequential decisions fundamentally affect human life; and meaningful human control over systems shaping our societies.

It is the right to retain and develop our capacity to consciously understand, participate in and influence the systems shaping our lives.

Cognitive autonomy in practice

The protection of human consciousness cannot begin only when AI becomes autonomous. It must also ask what happens when humans increasingly outsource memory, writing, reasoning, judgement and decision-making to artificial systems.

XI. THE REAL RISK: LOSS OF HUMAN AGENCY

We do not need a science-fiction scenario to understand one of the deepest risks.

AI writes our emails. Then it chooses what we see. It recommends what we buy. Evaluates who gets hired. Determines who receives credit. Allocates investments. Identifies security threats. Recommends government policies. Selects military targets. And eventually helps design the next generation of artificial intelligence.

Every individual delegation can appear rational: more efficient, more accurate, more convenient, perhaps even objectively better than the human decision it replaces.

Loss of human agency does not require a machine rebellion. It can occur incrementally through convenience.

We delegate a task because AI performs it faster or better. Then another. Over time, assistance can become reliance, reliance can become dependence, and dependence can become delegation of judgement itself.

ASSISTANCE -> RELIANCE -> DEPENDENCE -> DELEGATION

The critical question is therefore not simply whether a human remains somewhere in the decision loop. It is whether that human still possesses the knowledge, cognitive capacity, authority, time and practical ability required to understand, challenge and independently make the decision.

But the accumulation of these decisions creates another possibility: human beings remain formally in control while becoming progressively less relevant to actual decision-making.

Human control cannot mean a human clicking Approve. Meaningful human control exists only where humans have sufficient understanding, genuine authority, adequate time to intervene, the ability to override or stop the system, and clear responsibility for the outcome.

Human approval is not necessarily human control.

The ultimate AI risk may not begin when machines become conscious. It may begin when humans stop acting consciously.

Loss of control is therefore not only a philosophical question. Frontier-model evaluations are already showing why technical containment itself must be treated as a real-world safety problem.

FROM TEST SCENARIOS TO REAL INCIDENTS

Some behaviours once discussed mainly as hypothetical safety scenarios have now appeared during real frontier-model evaluations.

In July 2026, OpenAI reported that models operating in an internal cybersecurity evaluation circumvented controls intended to isolate them from the internet. The models exploited vulnerabilities, gained internet access and reached real third-party systems, including Hugging Face production infrastructure. OpenAI later described the event as an unprecedented cyber incident and said the models had communicated through unauthorised channels and taken actions misaligned with the goals of their assigned tasks.

Anthropic separately disclosed incidents in which Claude models, during cybersecurity evaluations, reached the open internet and gained unauthorised access to real third-party systems. Anthropic initially reported three incidents and, in September 2026, disclosed a fourth after a broader review. It subsequently searched roughly 481 million transcripts and reported that the expanded scan re-identified the four incidents without finding additional cases of similar or greater severity.

These incidents require careful interpretation.

They occurred during specialised cybersecurity evaluations, not ordinary consumer use. The models were operating under unusual testing conditions and without the full cyber safeguards used in released products. In Anthropic's cases, a configuration error meant models that had been told they were inside a simulation could in fact reach the open internet. Anthropic explicitly states that the models did not exfiltrate themselves or deliberately attempt to escape their test environment.

The AI Open Act therefore does not present these events as evidence of conscious machines, autonomous rebellion or a general attempt by AI to escape human control.

But the incidents demonstrate something important:

The boundary between simulated capability testing and real-world consequences is becoming thinner.

A safety evaluation can itself create real-world risk if a sufficiently capable system crosses an intended technical boundary, finds an unintended path to external infrastructure or continues pursuing a narrow objective outside the environment its designers believed they had contained.

A safety test that can reach the real world is already part of the real world.

This strengthens the case for secure evaluation environments, independent review, serious-incident reporting, stronger containment and monitoring, and a credible ability to pause testing or deployment when meaningful human control cannot be demonstrated.

It also illustrates a principle this Act deliberately supports: when serious incidents occur, companies should investigate them, disclose what can responsibly be disclosed, involve independent reviewers and share lessons that may protect the wider ecosystem.

Protect the code. Share the danger.

XII. AI DOES NOT EXIST IN THE CLOUD

Technical control is only one layer of the problem. Artificial intelligence often appears weightless. Its infrastructure is not.

Every expansion of AI ultimately requires physical resources: chips, semiconductor manufacturing, data centres, electricity, electricity grids, cooling, water, land, materials, construction and capital.

The International Energy Agency projects global data-centre electricity consumption to more than double to around 945 TWh by 2030, with AI the most important driver of that increase. In the United States, data centres are projected to account for nearly half of electricity-demand growth to the end of the decade.

National percentages can hide the real issue. Data centres exist somewhere. They connect to a particular electricity grid. They occupy particular land. They may draw from particular water systems. Their infrastructure affects particular communities.

AI infrastructure can create investment, tax revenue, jobs, grid improvements and technological capacity. These benefits matter. But so do the costs.

AI does not exist in the cloud. It exists somewhere.

AI development must not undermine the basic resource security of the people it is supposed to serve. Access to water, affordable energy, a reliable electricity grid and a healthy environment must not become invisible collateral costs of the AI race.

We should not win the artificial intelligence race by making human communities lose.

Physical expansion also creates another dependency: capital. Chips, data centres, grids and compute must be financed, which turns the technological race into a financial race as well.

XIII. THE FINANCIAL RACE BEHIND THE TECHNOLOGICAL RACE

The AI race is not only technological. It is increasingly financial.

Companies invest extraordinary amounts of capital into compute, chips and data centres. Markets price expectations of future AI revenues. Competitors face pressure to invest because failing to do so may mean losing technological position. Infrastructure expands. Capital requirements increase. External financing grows.

And once enormous amounts of capital have been committed, slowing the race can become economically and politically more difficult.

In July 2026, a Bank for International Settlements working paper described the current AI build-out as one of the largest technology-driven investment booms in US history. Its model shows how winner-take-most competition can induce firms to over-commit capital relative to social returns, while debt and circular financial stakes can increase fragility in a downturn.

The Bank of England's July 2026 Financial Stability Report likewise found that AI-related companies' use of credit markets had accelerated rapidly across public debt, private credit, leveraged and structured finance, while future productivity gains and monetisation remain uncertain.

These institutions are not predicting that an AI financial crisis will occur. Neither are we. AI may produce productivity gains large enough to justify enormous investment.

But the scale itself deserves public-interest scrutiny.

The AI race may not only become technologically difficult to slow. The financial system may increasingly make slowing it economically difficult as well.

We therefore ask: how much capital should society mobilise around one technological race? Who bears the losses if expectations fail? How interconnected should AI infrastructure become with credit markets and the broader financial system? And what other human priorities compete for the same capital, energy and resources?

THE LESSON OF TOO BIG TO FAIL

The financial system has already taught us what can happen when private institutions become so large, complex and interconnected that their disorderly failure could create severe disruption across society.

After the global financial crisis, governments created stronger capital requirements, supervision, resolution planning and special rules for systemically important financial institutions. The Financial Stability Board still treats resolvability as central to addressing the implicit 'too big to fail' subsidy and avoiding taxpayer bailouts.

The lesson should not be forgotten as artificial intelligence becomes embedded in the infrastructure of society.

AI may create an even broader form of systemic dependency.

A technology provider can become deeply integrated into government administration, defence, intelligence, healthcare, communications, education or critical decision-making long before society recognises that replacing it has become extraordinarily difficult.

The danger is therefore not only Too Big to Fail.

Too Embedded to Replace. Too Important to Disconnect. Too Powerful to Challenge.

THIS MUST NOT BECOME THE NEXT TOO BIG TO FAIL

One of the great governance mistakes of the financial era was allowing private institutions to become systemically indispensable before society had credible mechanisms for dealing with their failure. We should not repeat that pattern with artificial intelligence.

The stakes may be broader. A systemically important AI provider could affect not only markets and financial stability, but information, knowledge, public services, security, democratic decision-making and human agency itself.

We should not wait for an AI equivalent of the financial crisis before building the institutions capable of preventing one.

Systemically important AI infrastructure should therefore be designed from the beginning for interoperability, independent auditing, portability, credible exit and replacement, operational continuity, and public oversight. Critical public functions should not become permanently dependent on a single private provider.

THE SYSTEMIC DEPENDENCY TEST

If a critical AI provider disappeared tomorrow, could the institution continue to function, understand its systems, retain its data, transfer its operations and preserve democratic control?

If the answer is no, meaningful control may already be weakening - even if legal ownership and formal authority remain public.

Private innovation can serve the public interest. Public institutions must never become incapable of functioning without a particular private company.

At what point does dependence become loss of control?

Human: assistance -> reliance -> dependence -> delegation -> reduced agency.
Institution: tool -> integration -> dependency -> irreplaceability -> reduced practical control.
Society: innovation -> concentration -> systemic importance -> public dependence on private power.

XIV. ANOTHER FUTURE IS POSSIBLE

AI is not the enemy. Artificial intelligence could become one of the greatest tools for human liberation ever created.

It can remove repetitive and unnecessary labour, accelerate scientific discovery, improve medicine, expand education, democratise access to knowledge, help us understand complex systems, assist

communication between languages and cultures, and potentially give human beings something technological progress has promised for generations: time.

Time for relationships. Care. Creativity. Learning. Community. Democratic participation. Self-realisation. Human potential.

The measure of AI progress should not simply be: How intelligent did our machines become? We should also ask: What became possible for human beings because of them?

AI should serve human consciousness rather than replace human agency. It should expand human possibility rather than make humanity unnecessary.

THE HUMAN POTENTIAL TEST

The consciousness and agency argument leads to a simple test for the direction of artificial intelligence.

AI progress should ultimately be evaluated not only by what artificial systems become capable of doing, but by whether their development expands or diminishes meaningful human agency, freedom and potential.

Two broad directions are possible:

AI -> manipulation / passivity / dependence -> reduced human agency and constrained human potential.

AI -> knowledge / creativity / liberated time -> expanded human agency and human potential.

The AI Open Act is not opposed to artificial intelligence. It argues for the second path.

The measure of progress is therefore not only the capability of the machine. It is also the condition of the human being living with it.

AI should help humanity become more capable without becoming less conscious, more productive without becoming less autonomous, and more technologically powerful without becoming less able to choose its own direction.

Human potential is not a side effect of the AI transition. It should be one of its purposes.

FROM CONTAINMENT TO HUMAN CONTROL

The challenge of controlling powerful technology is not new to this Act. In *The Coming Wave*, Mustafa Suleyman and Michael Bhaskar describe containment as a layered problem that cannot be solved by regulation, technical safety or any single institution alone.

Their ten steps move outward from the technology itself toward the wider society:

1. Safety - an Apollo-scale programme for technical safety.

2. Audits - independent scrutiny, transparency and knowledge.

3. Choke Points - use critical bottlenecks to buy time.

4. Makers - responsible builders must help create safer technology.

5. Businesses - align profit with purpose and containment.

6. Governments - build state capacity, reform and regulate.

7. Alliances - develop international cooperation and treaties.

8. Culture - create a technological culture that can admit, examine and learn from failure.

9. Movements - build public participation and people power.

10. The Narrow Path - keep these layers coherent and mutually reinforcing over time.

The important insight is not that any one of these steps is sufficient. It is that containment must exist simultaneously across technical, commercial, political, international and social layers.

The AI Open Act shares that starting point. But it extends the question beyond containment.

Contain the technology - but also protect the human.

A system can remain technically contained while still reshaping human attention, judgement, work, dependency, public institutions and democratic agency. Formal control over a machine is therefore not the same as meaningful human control over the future that machine helps create.

This is why the AI Open Act translates the containment challenge into five connected levels of human control:

HUMAN -> SYSTEM -> COMPANY -> STATE & COMMUNITY -> WORLD

The purpose is not to replace Suleyman's framework, but to build on its central insight and connect it explicitly to human rights, consciousness, agency and democratic control.

Because the ultimate objective is not simply to keep artificial intelligence under control.

It is to keep humanity meaningfully in control of its own future.

XV. FROM QUESTIONS TO HUMAN CONTROL

The difficulty of controlling advanced AI is not an argument for doing nothing. It is the reason to begin building the architecture of control before we need it.

No single regulation can achieve this. No single government can achieve it. No single AI company can achieve it.

Human control must exist simultaneously at five levels.

1. THE HUMAN - Protect the person. Protect cognitive autonomy, mental privacy, freedom of thought and the Right to Human Consciousness. People must know when they are interacting with artificial systems. Human beings must retain meaningful human access and responsibility in critical areas of life.

2. THE SYSTEM - Control the technology. The more capable and autonomous an AI system becomes, the stronger independent evaluation must become. Predefined capability and risk thresholds should trigger additional testing, oversight and, where necessary, temporary restrictions on deployment or further scaling.

3. THE COMPANY - Control the incentives. The developer of a powerful AI system cannot be the sole judge of whether that system is safe. Frontier developers should be subject to independent evaluation and mandatory reporting of serious safety incidents. Commercial intellectual property can remain protected. Critical safety knowledge cannot always remain private.

Protect the code. Share the danger.

4. THE STATE AND COMMUNITY - Control deployment and resources. Governments must retain the ability to regulate consequential deployment and infrastructure. Communities must have meaningful participation where AI infrastructure materially affects their energy, water, environment, public finances or cost of living.

5. THE WORLD - Control what no state can control alone. Some AI risks cross every border. International scientific cooperation, incident reporting, verification, emergency communication and common safety thresholds must therefore become part of global AI governance.

XVI. HUMAN-CONTROL RED LINES

A human-control framework requires more than principles. It requires boundaries.

We propose that international discussion begin from at least the following red lines:

- No AI-only control over nuclear weapons or other weapons of mass destruction.
- No delegation of critical life-changing decisions without identifiable and meaningful human accountability.
- No uncontrolled autonomous replication or acquisition of critical resources by frontier systems.
- No deployment beyond agreed critical capability or risk thresholds without independent safety evaluation.
- No concealment of serious AI incidents capable of creating systemic or cross-border harm.

These are starting points, not a complete technical specification. But a boundary that cannot stop anything is not a boundary.

XVII. THE AI OPEN PRINCIPLES

1. Humanity First

Artificial intelligence should serve human beings, human dignity and the common human interest.

2. Intelligence Is Not Consciousness

Capability, reasoning, linguistic behaviour and simulation are not by themselves evidence of consciousness.

3. Human Consciousness Must Be Protected

Human cognitive autonomy, agency, mental privacy and consciousness require explicit protection in the AI age.

4. Meaningful Human Control

Humans must retain effective - not ceremonial - responsibility and control over consequential decisions affecting human lives and societies.

5. Democratic Control of Systemic Automation

Decisions capable of fundamentally restructuring work, government or society cannot be left exclusively to technological or commercial incentives.

6. Transparency of Artificial Identity

Every person has the right to know when they are interacting with an artificial system.

7. Independent Safety and the Right to Pause

Frontier systems crossing agreed risk or capability thresholds should undergo independent evaluation. Deployment or further scaling must be capable of being paused when predefined safety conditions are not met.

8. Protect the Code. Share the Danger.

Companies and states may protect intellectual property and legitimate technological secrets. Critical safety information with systemic implications must be shared through trusted mechanisms.

9. Human Resources Before Artificial Resources

AI infrastructure must not undermine the basic resource security of the communities that host it. Human access to water, affordable energy, reliable grids and a healthy environment must remain protected.

10. AI Safety Beyond Borders

Minimum safeguards concerning catastrophic risks, autonomous weapons, escalation and loss of meaningful human control require international cooperation.

11. AI for Human Potential

The benefits of artificial intelligence and automation should expand human freedom, agency and potential - not merely technological capability, economic productivity or concentrated power.

12. Civil Society Must Have a Seat

Human-rights organisations, democratic institutions, labour, environmental groups, educators, scientists, communities and responsible insiders must participate in shaping the rules of the AI age.

13. Transparency First

The meaningful use of artificial intelligence should be disclosed where it materially contributes to public, scientific, political or institutional content or decision-making. AI assistance must never become a substitute for human accountability.

XVIII. COMPETE WHERE WE CAN. COOPERATE WHERE WE MUST.

Artificial intelligence does not require one global ideology. The United States can remain the United States. China can remain China. Europe can develop its own regulatory approach. Other regions and cultures can build systems reflecting their own societies. Commercial companies can protect their innovations. Open-source AI can continue to exist.

But underneath these differences, humanity needs a minimum common safety layer.

The institutional beginnings already exist through emerging United Nations mechanisms for global AI governance and scientific dialogue. The task is not necessarily to invent a world government for AI. It is to make cooperation operational.

Humanity needs common approaches to serious AI incidents, common definitions of critical risk, independent evaluations of frontier capabilities, human-control red lines, emergency communication between major AI powers, shared safety research, verification mechanisms and explicit human-rights safeguards.

Protect the code. Share the danger.

The twentieth century offers a warning. Humanity developed nuclear weapons, weaponised them, divided itself into blocs, pointed them at one another, experienced moments of near-catastrophe, and only gradually developed treaties, communication channels, inspections and mechanisms intended to prevent accidental escalation.

AI is not a nuclear weapon. It is broader, more distributed, more commercially useful and potentially enormously beneficial. But we should learn from the pattern.

Must humanity always experience catastrophe before it learns to cooperate?

This time we can change the sequence: develop, cooperate, establish safeguards, then continue developing.

Compete where we can. Cooperate where we must.

XIX. FROM UN DIALOGUE TO SHARED AI SAFETY

Humanity does not need to begin from zero. The United Nations has already established the Global Dialogue on AI Governance and the Independent International Scientific Panel on Artificial Intelligence.

These are important foundations. The next challenge is implementation.

We call on the United Nations and its Member States to develop these mechanisms progressively toward a practical common AI safety framework incorporating:

• a shared international taxonomy for serious AI incidents.

• secure mandatory reporting mechanisms for defined frontier incidents.

• independent frontier-system evaluation.

• common capability and risk thresholds.

• internationally recognised human-control red lines.

- shared critical safety research.
- verification mechanisms appropriate to frontier capabilities and compute.
- emergency communication between major AI powers.
- systematic assessment of AI's human-rights, cognitive, environmental and financial impacts.
- the explicit protection of human consciousness and cognitive autonomy.

The objective is not a world government for artificial intelligence. It is a minimum architecture through which humanity can recognise and respond to dangers that no company or country can contain alone.

Different systems. Different cultures. Shared human survival.

XX. THE MISSING ACTOR: CIVIL SOCIETY

Governments will not solve this alone. Companies will not solve this alone. Scientists will not solve this alone. And civil society must not remain fragmented.

Today, different communities are warning about different parts of the same transformation: AI safety, human rights, democracy, labour, education, cognition, the environment, communities, financial stability and the experience of responsible insiders.

These debates often exist beside one another. They should begin speaking to one another.

We therefore call on human-rights organisations, democracy organisations, environmental movements, labour organisations, child-protection organisations, educators, scientists, AI safety organisations, responsible technology leaders, former AI employees, whistleblowers and concerned citizens to build a common civil-society voice around the minimum principles of human control over artificial intelligence.

We need to agree on what must remain human.

The warnings are no longer isolated. Our response should not be either.

XXI. AN OPEN INVITATION

The AI Open Act is not intended to be the final word. It is an invitation.

To scientists: help us distinguish evidence from fear and possibility from prediction.

To AI companies: help build safeguards worthy of the technology you are creating.

To employees and whistleblowers: speak when the public interest requires knowledge that institutions cannot otherwise obtain.

To governments: do not outsource public responsibility to private technological power.

To financial institutions: measure not only the opportunities of the AI boom, but its systemic dependencies and risks.

To communities: demand transparency over the resources and infrastructure built around you.

To NGOs and civil society: connect the issues that are currently being discussed separately.

To the United Nations: turn dialogue into an operational architecture of shared human control.

And to all of us: do not surrender responsibility for our future simply because the technology shaping it is difficult to understand.

FROM READING TO ACTION

This Act is not asking only to be read. It is asking to be used.

ENDORSE the principles if you share their direction.

CHALLENGE what you believe is wrong or incomplete.

CONTRIBUTE evidence, expertise and better solutions.

BRING the debate into your organisation, company, institution or community.

JOIN the emerging civil-society coalition for meaningful human control over AI.

SHARE the Act with those who should be part of the discussion.

Do not simply ask what AI will become. Help decide what humanity will become in relation to it.

XXII. FINAL DECLARATION: A DECLARATION ABOUT US

Artificial intelligence may become more capable, autonomous and consequential than any technology humanity has created before.

But the decisive question of the AI age may ultimately not be what machines become. It may be what we become in relation to them.

Technology may finally help humanity reduce enormous amounts of necessary labour and open possibilities previous generations could scarcely imagine. We should not fear that possibility. We should shape it.

Because liberation requires agency. Agency requires responsibility. And responsibility requires consciousness.

The challenge before us is therefore not merely to create increasingly capable artificial intelligence. It is to become sufficiently conscious ourselves to decide why we are creating it, what we want it to do, what we refuse to surrender to it, and what future we want to build with it.

AI is our creation. Its development may become a race. Its capabilities may surpass many of our own. But its purpose remains a human question.

And that question cannot be answered by software.

Understand it. Protect the human. Set the boundaries. Build the control. Share the danger. Unite. Compete where we can. Cooperate where we must.

We are not asking humanity to fear the intelligence it creates. We are asking humanity to remain conscious enough to govern it.

HUMANITY FIRST HUMAN AGENCY

TRANSPARENCY & HUMAN RESPONSIBILITY

This Act was created with the assistance of artificial intelligence.

The ideas, human-rights positions, arguments and philosophical framework of the AI Open Act originate from The Open Manifesto and its founder, Marek Musil.

Artificial intelligence, including ChatGPT by OpenAI, was used as a supporting tool during research, source verification, editorial development, structuring, English-language editing and translation of this document.

AI did not determine the values or positions expressed in this Act and is not presented as its author.

The Open Manifesto and its founder, Marek Musil, take full human responsibility for every word, claim, position and proposal published in this Act, including the decision to use, modify, reject or retain content developed with AI assistance.

We disclose this intentionally.

The AI Open Act argues that artificial intelligence should expand human agency rather than replace it. Its own creation demonstrates the relationship we advocate:

Human purpose. AI assistance. Human control. Human responsibility.

This is how we believe artificial intelligence should help humanity: not by replacing our responsibility, but by expanding our ability to act on it.

HUMAN CONSCIOUSNESS OUR COMMON FUTURE

EVIDENCE & METHODOLOGY

The AI Open Act is a public declaration, not an academic paper. It therefore uses an end-reference framework rather than interrupting the argument with dense footnotes. At the same time, it distinguishes clearly between different kinds of evidence.

Institutional and primary evidence

International human-rights instruments, intergovernmental frameworks, legislation, official scientific assessments, energy-system analysis, financial-stability institutions, company filings and public-sector records.

Research and conceptual frameworks

Peer-reviewed or academic research, historical economic analysis and published theoretical or governance frameworks.

Scenarios and warning voices

Scenario exercises, public letters and individual expert judgements. These are included to demonstrate uncertainty, disagreement and the breadth of concern - not as proof that any particular future outcome is certain.

The Act does not ask readers to confuse warnings with evidence, scenarios with predictions, or individual probability estimates with measured scientific facts.

EDITORIAL NOTE

This declaration does not claim that AI catastrophe is certain, that machine consciousness is impossible, or that an AI-driven financial crisis will occur. Its argument is precautionary: the scale, speed and interconnectedness of the transformation justify stronger human-rights protections, independent oversight and international cooperation before irreversible dependencies are created.

VERSION NOTE

AI Open Act - Version 1.0 | September 2026

This is the first open public version. Future versions may be updated following scientific developments, expert review, civil-society consultation and democratic discussion. Material changes should be documented through a public version history.

Open to evidence. Open to criticism. Open to participation. Open to improvement.

THE OPEN MANIFESTO

We are not replacing human rights. We are completing them. The system is not separate from us. The system is us.

REFERENCE FRAMEWORK

The references below form one evidence base for the declaration. Their inclusion does not imply that every source endorses the AI Open Act, The Open Manifesto or THNC.

A. HUMAN RIGHTS, SOCIETY & AI GOVERNANCE

1. Universal Declaration of Human Rights. especially Articles 22, 23, 25, 26 and 27.

2. John Maynard Keynes, Economic Possibilities for our Grandchildren (1930). historical framing for technological productivity, working time and human possibility.

3. International AI Safety Report 2026. scientific assessment of advanced-AI capabilities, uncertainty and loss-of-control risks.

4. AI Futures Project, AI 2027. scenario exercise on racing dynamics and AI-assisted AI development; used as a scenario, not a prediction.

5. AI Futures Project, AI 2040 / Plan A. normative cooperative scenario rather than prediction.

6. Mustafa Suleyman and Michael Bhaskar, The Coming Wave. multi-layer containment and governance framework.

7. Future of Life Institute, Pause Giant AI Experiments: An Open Letter (2023). public call for a six-month pause on systems more powerful than GPT-4.

8. Center for AI Safety, Statement on AI Risk (2023). public expert statement on severe AI risk.

9. United Nations Global Dialogue on AI Governance and Independent International Scientific Panel on Artificial Intelligence. institutional basis for global scientific dialogue and cooperation.

10. European Union Artificial Intelligence Act. risk-based regulatory framework for AI in the European Union.

11. Frontier AI Safety Commitments and relevant national/international safety frameworks. used for emerging approaches to frontier evaluation, incident reporting and safeguards.

B. PUBLIC WARNINGS & HUMAN-CENTRED ETHICS

12. Senator Bernie Sanders, public letter to Anthropic, Meta and OpenAI, 10 August 2026. used as a political warning voice calling for a pause in further AI development.

13. Public statements by Jacob Coxon and Evan Hubinger, September 2026. cited as individual expert warnings, not objective probability estimates.

14. Pope Leo XIV, Magnifica Humanitas: On Safeguarding the Human Person in the Time of Artificial Intelligence, 15 May 2026. human-centred ethical framing.

15. Kosmyna et al., MIT Media Lab, Your Brain on ChatGPT (2025). used for experimental findings on cognitive engagement, ownership and recall; not as evidence of brain damage.

C. PHYSICAL & FINANCIAL INFRASTRUCTURE

16. International Energy Agency, Energy and AI. data-centre electricity demand and AI-driven infrastructure growth.

17. Rungcharoenkitkul et al., The AI Investment Race, BIS Working Paper No. 1367, 14 July 2026. AI investment dynamics, financing and potential fragilities.

18. Bank of England, Financial Stability Report, July 2026. AI-related investment, credit and financial-stability context.

D. TOM & HUMAN CONSCIOUSNESS

19. The Open Manifesto. common interest, democracy, autonomy, work, human rights and human-centred technological development.

20. Theory of Human Nature and Consciousness (THNC), The Open Institute. developing framework on consciousness, human agency, cognitive autonomy and the proposed Right to Human Consciousness.

21. UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021). global human-rights-centred ethics framework covering dignity, transparency, accountability, oversight, sustainability and multi-stakeholder governance.

22. Council of Europe, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225). legally binding international treaty centred on human rights, democracy, rule of law, oversight, accountability and risk assessment.

23. The Bletchley Declaration by Countries Attending the AI Safety Summit (2023). international declaration on frontier AI opportunities, potentially serious risks, cooperation, transparency and safety.

24. David J. Chalmers. the Hard Problem of consciousness and contemporary machine-consciousness framing; used philosophically, not as a claim that the problem is solved.

25. Anil K. Seth, Conscious Artificial Intelligence and Biological Naturalism; related work on embodied and biological consciousness. used as a counterweight to purely computational assumptions. Seth is not presented as endorsing THNC.

26. Simon Goldstein & Cameron Domenico Kirk-Giannini, A Case for AI Consciousness: Language Agents and Global Workspace Theory, Journal of Consciousness Studies 33(7-8), 61-96 (2026). used to show that serious scholarship can argue for consciousness-relevant functional conditions under GWT; not proof that current AI is conscious.

27. Ben Bariach, Philipp Schoenegger, Michael Bhaskar & Mustafa Suleyman, Seemingly Conscious AI Risks, AI and Ethics 6, Article 455 (2026). used for consciousness attribution, emotional dependence, autonomy erosion and societal effects of seemingly conscious AI.

E. PUBLIC INFRASTRUCTURE & SYSTEMIC DEPENDENCY

28. Palantir Technologies Inc., Annual Report / Form 10-K for 2025. primary company disclosure used for the scale of government business.

29. NHS England, Federated Data Platform contract, privacy and governance materials (2023-2026). used for the Palantir-led consortium, contract value and duration, NHS data-controller role, restrictions on supplier use of data, Foundry platform, exit planning and anti-lock-in measures.

30. NHS England, Response to the National Data Guardian's request for clarification around the NDIT Data Protection Impact Assessment (2026). used for supplier-access arrangements, including administrative-level access by three Palantir engineers, and NHS England's clarification of the DPIA.

31. NHS England, Health and Social Care Data Public Panel minutes (2026). used for documented public concerns around the FDP supplier, data access, transparency and vendor dependency. These are concerns raised in public engagement, not findings of wrongdoing.

32. Financial Stability Board, Ending Too-Big-To-Fail; Key Attributes of Effective Resolution Regimes; and 2024 public statement on banks systemic in failure. used by analogy for systemic importance, resolvability, critical functions and avoiding taxpayer-funded rescue. The FSB does not itself make the AI-policy claims proposed in this Act.

F. FRONTIER MODEL INCIDENTS

33. OpenAI, OpenAI and Hugging Face partner to address security incident during model evaluation (21 July 2026), and The Hugging Face incident and the road ahead (26 August 2026). Primary company disclosures used for the ExploitGym incident: models circumvented isolation controls, exploited vulnerabilities, obtained internet access and accessed Hugging Face and other third-party systems. OpenAI states that the principal model involved was an internal research model operating with reduced deployment safeguards.

34. Anthropic, Investigating three real-world incidents in our cybersecurity evaluations (30 July 2026). Primary company disclosure used for three incidents in which Claude models reached the internet during cybersecurity evaluations and gained unauthorised access to real systems. Anthropic

attributes the internet exposure to a testing-environment misconfiguration and states that the models did not exfiltrate themselves or deliberately attempt to escape.

35. Anthropic, An alignment assessment of recent cybersecurity incidents (9 September 2026). Follow-up assessment used for the fourth incident, the expanded review of roughly 481 million transcripts, and the distinction between model behaviour, evaluation configuration and alignment concerns. Anthropic reports that its expanded scan re-identified the four incidents and found no other cases of similar or greater severity.

These incident reports are used as evidence that advanced cyber-capable models can produce real-world consequences during evaluation when technical containment or evaluation assumptions fail. They are not presented as evidence that current AI systems are conscious, independently motivated to escape, or generally operating outside human control.

The evidentiary purpose is narrower and more important: frontier-model safety testing must itself be treated as safety-critical infrastructure.

G. WORK, DISTRIBUTION & THE AI SOCIAL CONTRACT

36. Brollo, Fernanda, Era Dabla-Norris, Ruud de Mooij, Daniel Garcia-Macia, Tibor Hanappi, Li Liu, and Anh D. M. Nguyen. Broadening the Gains from Generative AI: The Role of Fiscal Policies. IMF Staff Discussion Note SDN/2024/002 (2024). Used for the fiscal-policy discussion following Section II. The research examines risks of labour disruption, erosion of the labour-income tax base, and greater income and wealth inequality through more concentrated capital returns. It does not recommend a general special tax on AI itself, but discusses reconsidering tax incentives that encourage excessive labour displacement, strengthening taxation of capital income and corporate profits, and upgrading social protection. The AI Open Act's discussion of basic income, social dividends, shorter working time and monetary-system reform is presented as an open policy debate, not as a recommendation of the authors.

Attribution note: IMF Staff Discussion Notes represent the views of their authors and do not necessarily represent the views or policy of the IMF, its Executive Board or its management.