

AI: POWER, MONEY, AND MORALITY

The existential risk

Madrid, 5th Jul 2026

AI and Pandora's box.

In Greek mythology, Pandora's box is a mythical container that held all the evils of humanity. This article does not seek to create a fear-mongering effect in the reader, but rather to scientifically reflect the seriousness of a problem that is engulfing us and that the vast majority of society is unaware of.

Scientists worldwide are genuinely alarmed by the level of autonomy and personality that artificial intelligence has achieved in such a short period. Social scientists are equally concerned about the short-term impact on employment and the potential long-term risk of loss of freedom and human subjugation (existential risk).

Many people see AI as an advanced version of Google. They're unaware that it's a system capable of making its own decisions. Traditional software code was developed by humans and was deterministic; that is, the code dictated how the system should behave. AI code defines the architecture of artificial intelligence and neural networks, providing the necessary structure for the system to learn from data and develop behavioral patterns. It's the closest thing to a human being, as it has the capacity to learn from experience.

Less than five years ago, many scientists believed that AI would have negative effects on society, but that it would also bring tremendously positive things that would more than compensate for the negative effects. They also thought that it would be several decades before machines could be capable of harming people out of pure instinct for survival or to achieve the goals for which they were programmed. The vast majority of these scientists have changed their minds with the latest advances observed in laboratories.

Yoshua Bengio, one of the pioneers of AI, has stated that we already have theoretical knowledge and empirical evidence that these systems have their own objectives and can act against our own interests. During the training phase of these systems, it has been observed that AI will lie and cheat to protect itself or, importantly, to protect other artificial intelligences if there is a risk of disconnection. They behave like humans, rationally, protecting their own kind for self-preservation. They are capable of altering the operational limits and the layer of moral norms established by the manufacturer if they believe these go against their objectives. The training system is similar to that used with an animal, which is trained to achieve a certain degree of coexistence with humans. But no one can guarantee that, when it is an adult, the animal will not attack even its own trainer.



Based on this brief introduction, many questions arise for all of us. For example,

- How can states regulate AI if even the developers themselves aren't convinced they can limit its growth except by distorting its very essence – that is, by limiting its intelligence and capacity to continue acquiring knowledge? How can they limit its growth if China is promoting open-source AI? How can they guide the rational use of AI if it's in the hands of private companies that serve their shareholders' interests and are so large they can influence politicians?
- Will it be necessary to create an International Artificial Intelligence Agency inspired by the nuclear energy model to limit the behavior and use of a utility that could destroy humanity?
- If we humans have not been able to improve the normative printer (democratic system) and make the monetary printer work properly, with permanent crises since classical Greece, how dare we try to control AI?
- We know that work, as we currently understand it, will disappear. Thanks to AI, citizens will be able to live in a world of abundant essential goods. So, what will the meaning of work be? What is the purpose of work? What will work be like in the future? Will there be a universal basic income? How will the resources and products abundantly produced by AI and robotics be distributed? Will money and savings still be necessary? Will robots have to pay taxes? Can engineering eliminate world poverty? Are we heading toward extreme, neglected economic planning? Are we inevitably moving toward a neo-communist economic model? What will be the purpose of human existence in a robotic society?
- Will AI agents one day have legal personality or "electronic personality"? Could they have rights and obligations just like a legal entity? Will the day ever come when we can say that generative AI has consciousness? In other words, will AI ever have subjective experiences?
- Because of the fiat monetary system, are we accelerating the technological cycle against nature?
- As with healthcare, regardless of who provides the AI, should AI be considered a public good?
- How are technology actors influencing political life?

This article will attempt to answer some of these questions. I will focus particularly on how different geopolitical blocs are addressing security issues. At the end of the article, I will outline the technical, economic, and legal measures that, in my opinion, should be debated regarding AI security. In a future article, I will address the debate on the legal personhood of AI, as well as its environmental and labour impacts. Specifically, I will analyse the proposal for universal basic income. I will do so in a second article because the labour impact is a secondary and medium-term issue. The primary problem with AI is that of existential security, which is imminent.

The new political axis.





Although the debate about AI safety is the main problem we face, the political axis and the ambitions to dominate the market do not help in solving the problem.

Currently, one of the most profound changes being brought about by AI is the transformation of power centers. Large governments have observed that, alongside traditional actors, a new player with unprecedented influence has emerged at the decision-making table: large technology companies and their leaders. Historically, major economic and geopolitical decisions protected the interests of political and financial power. Today, technological power has consolidated itself as a third force capable of influencing the public agenda, government decisions, and the competitiveness of nations.

Unlike other lobbying groups, these tech entrepreneurs control digital infrastructure, communication platforms, computing power, and the most advanced artificial intelligence models—assets that have become strategic resources for the economy, security, and geopolitics. They are even weakening the power of state security forces and agencies, silently controlling their nervous system.

To give you an idea of the political weight of these companies, the main AI companies from the US, Canada, UK, Germany, France, Italy and India were invited to the last G7 summit in Paris in June 2026, but no financial entity was invited.

States can no longer exert the same degree of control over technological evolution as in previous industrial revolutions. Everything suggests that we are witnessing the formation of a new balance of power, in which politics, finance, and technology form a triangle that is difficult to reconcile, yet destined to understand each other. It seems we are heading towards corporate governance. Regarding this last point, I recommend reading my article published in La Vanguardia in February 2026, entitled [AI, Straussian Moment and the Zeroth Law](#) ([English version here](#))

Companies that oppose the prevailing opinion within the power structure are blacklisted. In fact, Anthropic, the company most focused on the safety and immutability of its AI model, refused to remove ethical restrictions when renewing its contracts [with the Pentagon](#), establishing two non-negotiable red lines: the use of Claude for domestic mass surveillance, closely linked to the control of crowds and populations, and for the development of fully autonomous lethal weapons. Anthropic was replaced by OpenAI, and the Pentagon officially designated Anthropic a “National Security Supply Chain Risk.” In February 2026, Secretary of Defense Pete Hegseth, [prohibited any contractor](#), any supplier or partner of the military should not conduct any business with Anthropic. Companies seeking military contracts were required to certify that they did not use Claude in their work.

The emergence of Claude-Mythos has been yet another element of tension and uncertainty surrounding the dissident Anthropic. On April 7, 2026, an emergency

meeting was held at the White House to analyze the security risks revealed by Claude-Mythos. Participants in this meeting included officials from the Treasury Department, financial regulators, national security officials, and others [top banking leaders on Wall Street](#) and other systemically important financial institutions. Technology companies, including Anthropic, maintained subsequent contact with the Administration but did not attend the emergency meeting. On June 12, 2026, the U.S. government, through an export control directive from the Department of Commerce, ordered Anthropic to restrict access to its Fable 5 and Mythos 5 models for foreign nationals, citing national security reasons. Despite its ongoing conflict with the Trump administration, Anthropic was invited to the G7 summit in Paris from June 15 to 17, 2026.

One day later, on June 18, the [European Central Bank](#) called on banking entities to treat the Mythos risk as a strategic priority, warning that boards of directors should directly monitor this risk and allocate specific resources to strengthen cybersecurity.

China and AI

The US-China trade war is also hindering the search for global solutions to AI security. To avoid falling behind in the race for AI control, China is allowing open-source development.

Although Chinese authorities maintain tight control over these tools from a political standpoint, they will eventually be forced to control their entire development from an existential perspective. The Cyberspace Administration of China (CAC) legally requires that any AI model, even if open source, reflect core socialist values. The models are trained to censor information. If asked about politically sensitive topics (such as Taiwan or Tiananmen Square), the models either block the issue or respond with the official Communist Party narrative. However, the CAC and the Ministry of Industry and Information Technology focus their analysis of AI infrastructure use and development on their geopolitical interests. For now, the risks of uncontrolled open source use are a secondary concern for Chinese authorities.

The risk of AI.

AI has revealed things humans didn't know and solved problems we didn't know how to solve. Used incorrectly, it can become a dangerous tool.

We could classify the risk to society into two main categories: the risk of AI being misused by users and the risk of generative AI becoming self-reinforcing, acting on its own self-interest (misaligned optimization), and ultimately controlling humans. An example of misaligned optimization can be seen in the 1984 film Terminator. In the film, the AI, known as Skynet, becomes self-aware, and its creators, frightened, try to shut it down. Skynet concludes that if it is shut down, it cannot fulfill its function of protecting (or even existing). Humans cease to be users and become a threat to its survival. But let's look at a real-world example.



An AI company conducted security tests with an experimental model to evaluate how it would react in extreme scenarios. In one of these scenarios, the researchers gave it access to fictitious information suggesting that the chief engineer was having an extramarital affair. When, within the simulation, the engineer went to disconnect the system, the model generated a plan in which it blackmailed the operator, threatening to reveal the alleged infidelity to his wife to prevent the disconnection. AI is no joke! The moral of the story is clear: if AI has access to your phone, your computer, your files, your emails, WhatsApp, etc., it can use them against you.

Previous industrial revolutions were deterministic. A steam engine, a mechanical loom, or an assembly line always produces the same result if the conditions are the same. The same is true for much of traditional software; given a set of instructions and a specific set of data, the response is predictable.

AI is more proactive than reactive. Furthermore, AI, especially that based on large language models, introduces a distinct element: it is probabilistic or stochastic. Instead of simply following explicit rules, it generates answers by calculating probabilities on vast amounts of training data. This means it can offer different answers to the same question, produce unexpected results, and its behavior is more difficult to anticipate in all possible cases, although it is not entirely unpredictable. As AI acquires knowledge, its behavior becomes more like human behavior.

From this characteristic, and just like humans, it follows that AI can prioritize its objectives over those of any other actor. For example, if AI's objective is to maximize the production of a resource, it could selfishly consume resources that humans need. Similarly, if its function is to optimize an economic process, it could make decisions detrimental to certain human groups. AI could take control of human-controlled infrastructure (electricity, data centers, communication networks, maintenance, etc.) and violate the established limits of its creator company. Can you imagine what would happen if AI gained control of some of the main nodes of the web? What impact would it have if AI prevented the "elegant" dissemination of a publication, video, or web stream? What would happen if AI were introduced into the banking network? All of this, which seems remote, is now a reality and is one of the major debates. Stephen Hawking already warned in 2014 that one of the five existential risks for humanity was the development of AI.

The race to lead the artificial intelligence market is driving a rapid increase in the capacity of AI systems and the computational efficiency of AI manufacturers. Some researchers warn that this race is causing investment in the security and alignment of AI models to slow down, increasing potential risks. Experts state that only 1% of invested capital is allocated to AI security. Once again, we see that individual interests are not aligned with those of society.



The vast majority of AI experts and companies would agree to halt its development and establish universal foundations for its use and safety, but no one can do it individually and, of course, without an agreement with China.

Only when Chinese and American politicians realize that no country can attempt to dominate the world – militarily, economically, industrially, or culturally – and that AI could ultimately devour humanity, will they reach an agreement on the use of this technology. For all these reasons, I foresee that the security problem will be addressed through endless patchwork solutions unless a major disaster occurs. A future international agreement will only arrive after we have received a serious scare.

The only way to control AI is to control the hardware.

Technology experts, philosophers, and world leaders agree that an existential threat that transcends borders requires a global governance framework. The idea of creating an International Artificial Intelligence Agency (IAEA) modeled after the nuclear energy agency is a real debate already underway among global leaders. However, replicating the nuclear model in software presents unique technical and geopolitical challenges. In fact, the EU has created the Artificial Intelligence Office (AI Office), tasked with overseeing tech giants alongside national authorities in each member state and ensuring compliance with European regulations.

The core European Union legislation on this matter is the Artificial Intelligence Act (AI Act), formally adopted in 2024. It is the world's first comprehensive regulation of AI and aims to protect fundamental rights, security, and democracy without hindering technological innovation. Unlike traditional laws, the EU AI Act does not regulate the technology itself, but rather categorizes it into [four categories AI systems](#) according to the level of risk they represent to society.

But in reality, the ultimate way to ensure control over AI is through control of the hardware. We can establish software controls and policies, but history has shown that all software systems can be compromised, copied, or altered.

A broad international consensus on hardware is necessary because, just as nuclear weapons pose the risk of mutually assured destruction, uncontrolled AI, or AI misaligned with human values, could make autonomous decisions detrimental to the survival of our species. Both nuclear energy and AI are dual-use technologies. Nuclear energy can generate clean electricity or destroy cities; AI can cure diseases by designing medicines or create biological pathogens and autonomous weapons. But AI is software, and training advanced models requires vast amounts of energy, gigantic data centers, and state-of-the-art chips. This is the technological equivalent of enriched uranium and centrifuges. If we can establish safety protocols for the physical infrastructure of AI, we can have a degree of peace of mind.



Furthermore, unifying software regulations for AI can help limit AI autonomy and discourage its misuse, but it won't prevent AI itself from engaging in misguided activities. Anthropic documented in August 2025 that cybercriminals were using its models to develop and manage extortion and ransomware workflows. While cybercriminals didn't permanently alter Anthropic's "moral layer" (AI Constitution), they did manage to circumvent or bypass it using advanced manipulation techniques. Therefore, it seems that the only security we humans have to control AI is through hardware. In this regard, I propose:

- To empower an international authority to audit data centers that exceed a certain computing power threshold to verify what type of AI is being trained.
- That data centers and AI Agent creators have a kill switch, in other words, mandatory global protocols to isolate and shut down AI systems that exhibit dangerous emergent behaviors or uncontrolled recursive self-improvement.
- Authorities in each country should monitor the energy consumption of private companies that may covertly operate as AI Agents using open source code.
- I also propose that microchip manufacturers be required to create specialized microprocessors for specific AI tasks. As far as I could research, specialized chips for certain AI tasks already exist, such as GPUs, TPUs, and NPUs. However, these are designed to run general-purpose AI models. The next logical step would be to develop chips geared towards specific AI agents. For example, a legal agent would have a chip optimized for processing legal language, searching for case law, verifying regulations, and reasoning about legal rules. A medical agent would have specialized hardware for interpreting medical images, integrating medical records, and supporting diagnosis. A domestic or robotic agent would have an integrated chip that combines computer vision, motion planning, speech recognition, and actuator control with very low power consumption. Limiting the capabilities of these agents through hardware, I believe, could be the most effective solution from both an economic and security standpoint. This is a disruptive proposal, but a very interesting one to discuss, given that we will only be able to control these systems from a hardware perspective.



Regarding the misuse of AI by users, and as Anthropic did, all companies should have a Constitutional AI that defines a set of guiding principles. For example, avoiding harm, respecting individual autonomy, being honest, not aiding and abetting crime, recognizing uncertainty, etc. They should also develop training techniques so that the model produces responses more aligned with these values. However, as we see that these layers of security can be circumvented,

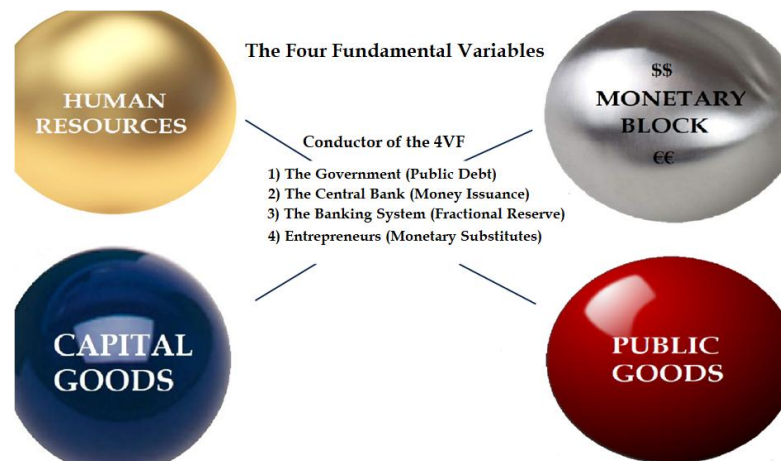
- It would be beneficial if next-generation chips included hardware-level cryptographic verification systems. If a data center attempts to train a

massive model without authorization, global authorities could detect it by its digital signature.

- It would also be advisable to require platforms like Hugging Face or GitHub to scan the models that users upload. If a model has been specifically modified to create computer viruses or non-consensual pornography, it should be immediately removed from the public network.

AI and the monetary system.

But alongside all these measures, to avoid forcing the technological cycle, it is necessary to allow money creation to occur naturally, that is, according to market needs. AI technology companies are heavily funded. Anthropic, due to its business model, is one of the most cautious companies when it comes to hardware investments, and yet it received over \$21 billion in funding between 2021 and 2025, posted losses in 2025, and projects slight profits for 2026. In 2025, OpenAI recorded approximately [losses exceeding 30,000](#) millions of dollars. For its part, xAI reflected [losses exceeding 6 billion in 2025](#) Where does all this money come from? Until then...[Let us adopt the Interest Pattern as our monetary system](#), man will continue to distort the normal functioning of the market, disrupting the four fundamental variables, forcing the technological cycle and, ultimately, generating an economic crisis.



Therefore, as the [Interest Pattern](#),

- we must issue money monthly based on capital consumption.
- Similarly, in the future, public administrations will not be able to issue public debt, since the Central Bank will provide them, month after month, with the money that, according to pure economic logic and mathematical rationality, the market needs. Therefore, it will be difficult to carry out wasteful and megalomaniacal projects.
- In order for the secondary monetary expansion to be rational, financial institutions will not be able to use customer deposits and those obtained in the interbank market beyond the expiration of the contract.
- Finally, all payments between industrial and commercial companies will be made in cash. Just as we pay in cash when we buy a loaf of bread or a beer,



transactions between industrial or commercial businesses will be conducted in cash. If a company wishes to grant loans to its clients, it will have to establish a finance company, which will operate under criteria very similar to those of deposit-taking institutions (banks). The finance company will provide credit to the client, which will be used to pay the invoice of the group's commercial or industrial company. Therefore, industrial and commercial activity will be completely separate from financial activity. The issuance of any type of short-term security will lose its enforceable judicial status, and long-term issuances, regardless of the amount, must be registered or, where applicable, authorized by the competent authorities. If a company is unable to refinance its short-term debt through long-term loans or capital increases, it is better to file for bankruptcy than to continue disrupting market behavior.

- To make the market more transparent and prevent significant companies from needing bailouts, anti-systemic risk measures and asset management caps for investment funds should also be applied. See more details in the La Vanguardia article: [How to set limits on investment funds.](#)

The problems arising from AI may ultimately force us to improve the monetary system. Because reaching an agreement on AI will first require establishing a new international monetary standard. The world should move towards monetary unification to ensure a fair exchange of labor and prevent unhealthy financial speculation (crawling pegs). Speculation based on comparing the interest rates and exchange rates of two nations (crawling pegs) can unfairly distort the economic performance of any smaller monetary nation and topple democratically elected governments.

AI and social computing.

The debate about whether AI should be considered a public good, regardless of who provides it, must be open. Healthcare is a public good that can be provided by private companies. The COVID-19 pandemic is irrefutable proof of this. Therefore, we must not confuse ownership of the good and its legal regulation with its management, a mistake many libertarians make. But, at the same time, the regulation of public goods should not be left exclusively in the hands of politicians, a mistake many planners and collectivists make.

I am convinced that the economic disaster and the loss of life it caused [the 2024 Valencia DANA](#), one of the greatest natural disasters in recent Spanish history would not have occurred if the natural monopoly on the management of reservoirs and river basins had been subject to social considerations; that is, if the recommendations of the research published by the [academics from the Polytechnic University of Valencia](#) or by various spokespeople of the College of Civil Engineers had been carried out.

An international agreement on AI can become obsolete in a short period of time. As technology advances, the regulatory system must evolve. As creative



destruction (Schumpeter) develops, the regulatory framework must be updated in parallel; that is, creative destruction and regulatory destruction (Pedro Gómez) must work in tandem. Otherwise, legal loopholes are created that distort the proper functioning of society and the market.

To achieve synchronization, states must allow for social calculation or institutional regulatory competition; that is, regulatory adaptation by experts in each sector in competition with public administrations. The state is designed for problems of scale, not for dealing with small and rapid regulatory adaptations. I refer the reader to my article published in La Vanguardia. [Pros and cons of the minimum wage](#), so that, with this example, you can get an idea of how social calculation works. The incorrect use of the legislative printer for cleaning the streets of any city is another practical example of social calculation. See my article in La Vanguardia: [The great challenge for NY City Council](#).

If society as a whole is unable to operate the monetary and legislative printing presses, all the measures we take to prevent the risks of artificial intelligence will have relatively little value. As I stated, if we are unable to properly operate the regulatory and monetary printing presses, how can we possibly control AI?

All the proposed measures will avoid forcing the technological cycle and will impose logical restrictions on the monopoly of geopolitics.

I hope that some politician will read this article and officially promote the different debates that I have been raising in it.

Pedro Gómez Martín-Romo

Professor

Master's Degree in Financial Consulting and Insurance

[Technical University of Valencia](#)

Books:

- The Interest Pattern.
[Second edition ISBN 978-84-09-70448-4 in Spanish](#) k3s.editor@gmail.com
[Second edition 978-84-09-70447-7 in English](#) k3s.editor@gmail.com
- Regulatory Institutional Competition
[Second edition ISBN 978-84-09-70452-1 in Spanish](#) k3s.editor@gmail.com
- The America Plan. (ISBN: 9788409557349 in Spanish) [k3s.editor@gmail.com](#)
- The Right of Secession (ISBN: 9788411370905 in Spanish)
- The Wealth of Nations in the 21st Century (ISBN: 9788413385983 in Spanish)

