Regulate or Socialize AI?

Claudio Katz

English translation of the original Spanish article.

The Opening Argument

The loss of control over Artificial Intelligence through its capacity for self-improvement has become the nightmare of the moment. Fear is growing about the ability of this system to override instructions issued by humans, as its rapid development increasingly enables it to operate autonomously.

This risk was first highlighted by employees of several companies in the sector in a striking public letter. Subsequently, some executives involved in projects to create self-improving superintelligence joined the warnings. They cautioned against the dangers of AI and its potential to trigger an event leading to the extinction of humanity. Finally, a former employee of these companies expressed concern that computers could set in motion an escalating spiral of self-programming mechanisms. He warned that AI could "kill us all" before the end of the decade and acknowledged that this concern is shared by the leading figures responsible for developing the technology.

Suspicious Alarms

In the new climate of panic unleashed by AI, the leading figures at OpenAI, Anthropic, and xAI have taken heed of calls to pause and control the pace of AI's self-training.

After disagreeing over the desirability or feasibility of such regulation, they have suggested endorsing these controls, although without specifying how oversight would be implemented. Trump embodies the chorus of dissenting voices that attribute any attempt to slow AI's advance to the actions of "evil forces." He argues that such a pause would cause the United States to lose the race against China and calls for preventing that outcome.

Amid growing fears of a science-fiction scenario, nobody knows the actual likelihood of a planetary catastrophe caused by AI. The danger of losing control through the "hallucinations" generated by its models has been suggested on several occasions.

The great fear surrounding the "recursive self-improvement" of computers is that programmers might lose the ability to understand AI's own languages. This gives rise to the dystopian scenario of machines communicating among themselves and circumventing controls, potentially triggering a financial, energy, health, or political collapse.

A year ago, Anthropic's Claude chatbot was producing 10% of the codebase used to develop its own successor; today, it generates more than 80% of that self-training code. In a serious incident involving a high degree of autonomy, both ChatGPT and Claude exhibited inexplicable losses of control that unsettled their creators, who struggled to determine the origin of those anomalies.

This loss of human oversight over actions managed by AI itself forms part of the dynamics of the catastrophes incubated by capitalism. That system is governed by brutal competition among companies, which fight over markets and profits without considering the uncontrollable consequences of their rivalry.

For that reason, they cannot manage the technical risks generated by innovation in a rational, controlled, and measured manner. Yet the degree of imminent and actual danger posed by AI self-training remains a major unknown.

The simultaneous campaign of alarming denunciations endorsed by the leaders of the digital plutocracy raises many suspicions. Strikingly, they warn against the very products they develop, promote, and sell. For this reason, the regulations under discussion may be aimed at ensuring the survival of those companies.

The firms face two threats far more tangible than the end of humanity: Chinese competition and a stock-market collapse on Wall Street. It would not be the first time that major capitalists secured increased government subsidies to counter a threat that they themselves had created, intensified, and magnified.

The controversy over AI regulation also exposes an internal contradiction within the American business world. On one side are the dominant sectors seeking to protect their oligopoly by introducing costly safety protocols and rules unfavorable to firms aspiring to enter the existing AI market. This is Anthropic's position.

On the other side, major chip suppliers seek to expand their markets and customer bases without restrictions. This is the policy encouraged by Nvidia. Between these positions are companies that already have contracts tied to the Pentagon and do not want competitors, but simultaneously need to accelerate the confrontation with China. This is the attitude conveyed by Palantir.

In every case, companies envision self-regulation tailored to their respective business plans. Nvidia, for example, launched a new tool to prevent AI from going out of control, which it will supervise itself in order to preserve its enormous dominance in the semiconductor industry.

Everyone involved in the AI business proclaims that this technology could extinguish humanity, yet they continue developing it and insist that they are participating in an unavoidable race against their Chinese competitor. They seek to

convince the world that their model of innovation is the only possible one and that the future will belong to whoever controls it.

Neoliberalism used the same arguments about the "end of history" and the "end of work" to justify precarious employment. Now the "end of humanity" is invoked to consolidate state support for the billionaires of the digital caste. By cultivating this atmosphere, they conceal the fact that AI's greatest danger arises from its capitalist implementation, rather than from the fetishized despotism that robotic humanoids would supposedly exercise over ordinary mortals.

Discussing the dangers of AI in terms of its benevolent or hostile "alignment" with the human race is a colossal deception. It overlooks the fact that the owners and beneficiaries of this technology enrich themselves at the expense of the majority of society, creating terrifying warnings to consolidate their profits.

Antidotes and Threats

The most frequent comparison regarding the dangers of AI refers to atomic energy and the possibility of humanity's destruction through the military use of that technology. Since the Second World War, the entire planet has lived with this grim prospect, initially restricted to a devastating confrontation between the United States and the USSR and subsequently extended to the significant number of countries possessing nuclear weapons.

This proliferation has operated as a nuclear deterrent that paradoxically limits the multiplication of wars. Insofar as the opposing sides possess forces capable of destroying the entire planet (and therefore themselves), that power acts as a restraint on the scale of warfare.

This antidote operated during the Cold War and subsequently expanded as more potential belligerents acquired atomic bombs. That precedent—which illustrates how a balance of terror counteracts apocalyptic outcomes—could, to some extent, be extended to the development of AI.

But the comparison also reminds us that the absence of an ultimate nuclear catastrophe has coexisted with the proliferation of wars, massacres, and military tragedies, and that the greatest immediate danger of AI lies precisely in its military applications. Such use has already been evident in the killing of girls at a school in Iran and in the indescribable crimes committed in Gaza.

AI is playing a role in these bloodbaths with the direct participation of OpenAI, Anthropic, and xAI. Would the regulation they now accept include a prohibition of this deadly use? By discussing the possible end of humanity, this question is avoided.

Talking about the future collapse of the entire human race serves to avoid debating the brutal genocide being perpetrated with AI in Palestine before the eyes of the entire world. As long as these issues are sidestepped, all debates will remain lost in a fog of generic statements. AI regulation is obviously indispensable, but it will be neither effective nor beneficial to the majority as long as the despotism exercised by the wealthy owners of these companies persists.

For the moment, OpenAI has abandoned its initial rejection of the proposal for coordinated controls advanced by Anthropic, which the Pope conceptualized in an encyclical. Leo XIV pointed out that behind AI stand human beings, not the magical, immaterial universe suggested by the owners of this technology. He highlighted the extensive network of natural resources and energy infrastructure generating regressive and dehumanizing dynamics. But these denunciations do not define the content of any eventual regulation, which could consolidate, weaken, or eradicate the power of AI's owners.

Such regulation will not be neutral and would involve countless areas. Some could directly challenge the digital plutocracy through the spread of opposition to the construction of data centers, already visible in different parts of the United States and the world. Conversely, future regulation could provide the magnates with additional legal instruments to consolidate the energy foundations of their businesses. But any decision will affect another key dimension of the problem: AI's close relationship with environmental deterioration.

The Same Unrestrained Development

AI aggravates the environmental crisis through the construction of data centers, which constitute the principal infrastructure of this technology. Their operation requires extremely large quantities of water and energy to cool tens of thousands of servers, whose power consumption, measured in megawatts, is equivalent to that of small cities.

This processing drastically increases local temperatures and those of the planet's atmosphere. In 2023, Google alone used nearly 23 billion liters of water to cool its facilities, while Microsoft plans to build its own nuclear power plant to operate these systems.

It is estimated that the AI boom will cause global energy demand to increase by 19% to 22% annually through 2030, while other estimates suggest that these energy requirements will multiply tenfold every six months. Heavy investment in hardware that quickly becomes obsolete generates enormous quantities of electronic waste that damage the environment.

Many of the analogies drawn between AI and the danger of a nuclear explosion extend to environmental degradation. The dystopian warning about a new

technology that "could kill us all" bears many similarities to predictions of the extinction of habitable areas of the planet.

Like AI, climate change faces catastrophic tipping points that generate uncontrollable feedback loops. The equivalent of superintelligent computers improving themselves through self-training is the rise in global temperatures above preindustrial levels, making environmental deterioration irreversible.

The catastrophic warnings now spreading about AI have much in common with the admonitions surrounding environmental disaster. It is enough to recall that more carbon dioxide was emitted between 1991 and 2021 than during the preceding two centuries and that, with the exception of the pandemic, these emissions have continued to increase year after year.

These warnings are based on the simple observation of the dramatic increase in raw-material processing that contributes to global warming. This consumption rose from 7 billion tons in 1900 to 14 billion tons by midcentury, to 100 billion tons today, and could reach 200 billion tons by 2050, irreparably damaging every ecosystem. The uncontrolled speed transforming AI development mirrors this environmental disorder.

Both processes have been sharply accelerated by military pressures. The fossil-fuel economy has experienced a full resurgence in recent years to satisfy military demand generated by the conflicts in Ukraine and the Middle East. This has induced record levels of oil and gas extraction, with a corresponding increase in global pollution.

NATO's military expenditure alone caused record greenhouse-gas emissions in 2023. The proportion of AI devoted to military activity is equally overwhelming, since that application is the sector's priority. For this reason, the United States is the epicenter of both calamities.

The shale revolution transformed the world's largest oil consumer into a major producer and exporter of crude oil. If the US military apparatus were classified as a nation-state, it would rank 47th among the world's emitters. Because of this dependence on the fossil-fuel economy, the United States remains a petro-state that avoids the transition toward the electro-states promoted by countries supporting renewable energy.

Similarity of Impacts

The similarities between the adversities caused by AI and those associated with climate disruption derive from the common capitalist foundation supporting their development in the West. The cataclysms incubated by both phenomena originate in the uncontrolled competition among companies to maximize profits at the

expense of society and humanity's own future. AI plutocrats have embarked on a reckless race to accumulate money and dominate the sector, without considering the potential disasters that this competition may unleash.

The environmental disaster is determined by the same capitalist rivalry and could lead to the same collective suicide through the simple overuse of fossil fuels. As victims of recent fires in Europe have declared: "It is not the heat that kills us; it is capitalism."

Because of this social determinant, the new phase of geological history marked by the environmental impact of human activity (the Anthropocene) is not merely an effect of disregard for nature. It is a consequence of the irrationality of the prevailing system (the Capitalocene).

In both areas, far-right ideology denies the consequences of the tragedy now unfolding. Trump displays this blindness with his customary brutality. He claims that climate change is an invention of scientists seeking recognition by falsifying data, and proclaims that pausing AI is a conspiracy to benefit China.

His followers ignore the greenhouse effect and the melting of Greenland and Antarctica. They disregard droughts, floods, and waves of polar cold or tropical heat. They also ignore forest fires, the destruction of biodiversity, the sinking of Pacific islands, and the disappearance of plant species.

The same denialism has prevailed until now in the AI universe, reflecting the same servility toward capitalists. In the environmental sphere, submission to the 100 companies responsible for 70% of polluting-gas emissions is evident. In AI, everyone pays tribute to the millionaires of SpaceX, Anthropic, OpenAI, Google, Meta, Microsoft, or Amazon.

They resort to the same reactionary ideological framework to justify this adulation. In environmental matters, they naturalize the climate crisis through neo-Malthusian perspectives or take denialism to extremes through ecofascist positions. In AI, they propagate transhumanist theses concerning the inevitable supremacy of superintelligent computers and the eternal existence of a superior caste managing these devices. This regressive foundation underlies the very ambition to create an AI superior to ordinary mortals.

Yet nobody can explain the point of spending so many scarce resources artificially producing the attribute of intelligence, which eight billion human beings already possess from birth. The money squandered on this undertaking would be better devoted to education and the cultivation of human intelligence than to the delusional attempt to replace it.

Since comparisons between AI and climate change also extend to proposed solutions, it is worth recalling what happened over recent decades with the various initiatives intended to mitigate environmental disaster. That record offers an instructive precedent for current debates about the forthcoming regulation of AI. Failure was the defining characteristic of all the initiatives—climate summits, the Kyoto Protocol, and the Paris Agreement—that attempted to develop some form of green capitalism.

The failure was equally categorical in the case of eco-capitalist solutions that sought to commodify the climate through bonds and investments. Their neoliberal sponsors mistakenly assumed that protecting nature could be achieved through privatization. The digital plutocracy is now following the same path of capitalist parameters to regulate AI at the expense of the overwhelming majority of the population.

The comparison extends, finally, to the position assigned to the periphery. Faced with growing opposition to the construction of data centers in US communities, billionaires seek to relocate these facilities to peripheral countries. They advertise that this will generate jobs, when in fact employment creation will be minimal. Clearly, once construction is completed, these facilities will be managed remotely from corporate headquarters. As with climate change, the periphery bears the full weight of the intensified adverse consequences of the new technology.

Who Owns AI?

The alarms raised by AI developers have placed discussions about the future of these systems at the center of the global agenda. Demands for regulation extend even to Trump's own circle, despite his elimination of all restrictions on the sector and creation of a free-for-all environment in which companies can do as they please. This deregulation accelerated the feared loss of control over AI.

Voices of opposition have emerged even within the far right. A former leader of Trumpism such as Bannon has recently adopted "techno-negative" positions, in confrontation with Thiel. He demands regulations, job protection, and strict limitations on the advance of AI.

All developments in this enigmatic sector generate more suspicions than certainties. For that reason, it is essential to evaluate the content of each proposed regulation very carefully, distinguishing positive elements from regressive guidelines.

At the broadest level, three major approaches to regulation appear to be emerging.

The more liberal sector represented by OpenAI and SpaceX advocates minimal or cosmetic controls, trusting that these will facilitate continued competition with

China. Inspired by the transhumanist discourse of Thiel and Karp, they have embarked on a crusade rejecting any significant restriction on AI development. Because they regard the transcendence of humanity's biological limits through this technology as desirable, they encourage the accelerationist model without fear of the devastating consequences of that course.

The second perspective, advanced by Anthropic, warns of a dystopian future if the current trajectory continues. It proposes establishing containment rules based on a coordinated pause. In the version known as the "San Francisco Consensus," this oversight would be restricted to the United States and its Western allies in order to maintain the geopolitical contest with their Asian rival.

China leads the third proposal, which Xi Jinping presented at a recent world conference. It advocates coordinated global regulation based on principles opposed to transhumanism and accelerationism. It also proposes incorporating the Global South, in convergence with certain ideas expressed in the Pope's encyclical.

The Eastern giant does not yet have a single comprehensive artificial intelligence law, but rather regulations governing algorithms. It proposes an international architecture based on creating a governance body (WAICO), integrated with the institutions promoted by China (the Shanghai Cooperation Organization, BRICS) and the United Nations. Through this approach, China promotes alliances extending beyond its bilateral competition with the United States. China's general strategy seeks to limit the risks of AI running out of control, but without accepting regulations that would freeze the advantages the United States retains in several fields.

Alongside these outlines of general positions on confronting AI's dangers, numerous initiatives have also emerged to establish immediate protective rules against its effects. These initiatives include eliminating the use of chatbots by minors and seeking to complement intellectual labor rather than replace it.

In discussions about the criteria that should govern AI regulation, opposition to allowing computers to make decisions autonomously has become widespread. There is a growing demand to guarantee human control over such decisions through a strict classification of the risks involved in each operation.

But the underlying issue is who should exercise overall control over AI, beyond its immediate regulation. If these controls validate and reinforce the ownership of this system by the handful of billionaires who currently control the technology, all the adversities, injustices, and cataclysms arising from that control will persist.

For this reason, the proliferation of initiatives aimed at bringing the companies that control AI under state ownership, nationalizing them, or socializing them is a very

positive development. Bernie Sanders, for example, proposed creating a Sovereign Wealth Fund under which these companies would transfer a 50% equity stake to society as a whole. This proposal converges with projects to impose heavy taxes on billionaires in the digital sector.

There are many possible models of socialization, involving corporate ownership, data management, and forms of control. But the starting point for these developments is ending the control of AI by a handful of billionaires.

Clearly, moral criticism of the inequalities and dangers generated by AI's current development will remain a mere abstraction unless accompanied by drastic demands for taxation and changes in ownership. What is at stake is the future of a common good that has been appropriated by a handful of capitalists to the detriment of society as a whole.

Exploitation and Confiscation

The user has become a central protagonist of digital capitalism because users provide the data collected, accumulated, selected, and commodified by companies. Through this supply of information, they have contributed, consciously or unconsciously, to forging the foundation of the system controlled by the digital plutocracy.

This contribution of data does not amount to work performed by users. They do not act as wage earners hired by companies to perform tasks on platforms or in services complementary to those structures.

In the traditional terms of political economy, users engage in an activity (labor) rather than work (work). They provide Facebook, Google, Apple, and OpenAI with the basic support for the development of their networks free of charge, without participating in the reciprocal contractual arrangements characteristic of employment relationships.

This distinction between work performed on platforms (or in connection with them) and the user's passive activity is crucial to understanding digital capitalism. Authors who emphasize this difference have provided essential guidance for clarifying the current model.

In Marxist terms, this distinction refers to the difference between forms of exploitation and domination. The former involves the subordination of workers to relations of exploitation and the extraction of surplus value that sustain employers' profits. The latter concept refers to all the activities that capitalism appropriates and commodifies in the course of its development. It takes possession of nature, the commons, and the fruits of multiple human activities without the mediation of labor.

Platform capitalism draws upon this second foundation through the simple actions of users who, by connecting to networks, contribute, recycle, and refine the data that the system transforms into algorithms. This participation may be deliberate or entirely unknown to those involved in a circuit that now encompasses most of the planet's inhabitants. The system itself has generated a dynamic of self-reproduction of these foundations without consulting any of the actors who provide the data commodified behind their backs.

Some theorists have conceptualized more precisely the distinction separating workers from network users, offering numerous examples of seemingly similar activities performed before screens whose actual content is radically different. When someone entertains themselves with Facebook's curiosities, they act as a participant, but not as an intellectual worker in that system. If, on the contrary, that person receives a salary to supply the system's Big Data or works as a community manager, they are performing work.

The qualitatively different character of these activities is frequently confused because they involve actions performed in the same place and sometimes during the same period. But it is important to distinguish these situations in order to determine when the same person is acting as a worker and when as a consumer of a good.

The surplus value that feeds capitalist profit is generated in the first situation, not the second. Exploitation is present in the contractual wage relationship, not in the ordinary use of a device by the same individual. The similarity of the actions performed before a screen, pressing the same keys, should not obscure this fundamental difference. Qualitatively distinct activities are carried out through formally similar acts.

Appropriated Public Goods

If the data contributed by users do not constitute work, what kind of goods are these inputs, and how do capitalists use them? In practice, they constitute public resources appropriated by the powerful. These data are goods shared freely and available to the entire community, which the digital plutocracy confiscates for its own benefit.

This expropriation repeats, in the twenty-first century, the seizure of common lands by landowners at the birth of capitalism. They took over those lands by expelling peasants, laying the foundations for the primitive accumulation that sustained subsequent capital accumulation. The same seizure has been reproduced with data. Companies appropriated these inputs at no cost and commodified them for their own benefit, building the multibillion-dollar fortunes now held by Gates, Musk, Bezos, Page, and Zuckerberg.

They accomplished this expropriation by monopolizing data that, until the 1980s, came from surveys and public information databases. Subsequently, they digitized these inputs, taking advantage of the growing use of bank credit cards and the early spread of the Internet. This confiscation operated as a vast form of "data colonialism" because of the scale of the capture, collection, and storage of appropriated information.

Companies confiscated public goods that did not belong to them, because these data are resources belonging to the entire community. The same applies to other public goods (air), public spaces (communication), public places (squares), public organizations (museums), public institutions (education), and public services (transportation).

The data appropriated by the digital plutocracy belong to this universe of resources available to all members of society, which a handful of usurpers have captured for their own benefit. This is the same privatization process that prevailed under neoliberalism, except that in this case there was not even the formality of purchases, bidding procedures, or contracts.

Data constitute social rather than natural commons because they are products of society's historical development, not of the physical environment surrounding human beings. They encompass an immeasurable ecosystem of codes, ideas, images, and products and form the foundation of a cultural heritage enriched through cooperation. This accumulated wealth has been seized by the billionaires who control Artificial Intelligence.

This type of expropriation is commonplace in the dynamics of capitalist accumulation, and some economists tend to describe it as an "externality." They use this term to refer to a benefit obtained by companies free of charge, omitting the fact that this profit is realized at the expense of those excluded from the resulting wealth.

This capture of freely available goods tends to become normalized, and over time nobody asks how such enormous inequality arose between those who seized the resources and those deprived of them. The same is happening today with data. The use of digital systems has become so commonplace that nobody questions the origins of the wealth accumulated by the plutocrats of the information industry. The traces of their expropriation tend to be erased.

It is extremely important to clarify these origins amid the emergence of AI because mainstream economics tends to present data as merely an additional factor of production, devoid of history and social content. Data are portrayed as goods possessing magical, immaterial virtues that emerged from the curiosity, inventiveness, foresight, or efforts of their managers.

Such praise conceals the fact that the algorithms and models underlying AI constitute knowledge formalized through the expropriation of a public good. Computers train themselves using this resource in data centers that ravage water and energy resources, with the same impunity with which companies captured the data.

Artificial Intelligence learns through the use of a public good that belongs to society as a whole. It draws upon a universal raw material dependent on collective contributions. The socialization of this resource is therefore an act of justice in response to the confiscation perpetrated by capital. Starting from this conclusion, we can evaluate more solidly and consistently the dangers raised by AI's current development.

Artificial Intelligence constitutes a digital common good, grounded in the attribute of knowledge, and should develop according to those principles. Capitalism prevents or undermines this purpose, and only by emphasizing this distortion can we find a way to confront the dangers of a dystopian future. Following this line of reasoning, in the next article we will analyze the significance of changes in labor under digital capitalism.

October 10, 2026