

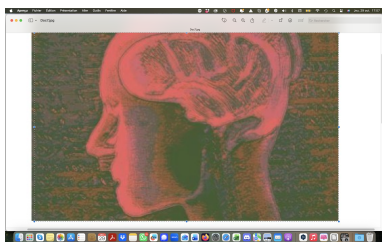
<https://internationalviewpoint.org/In-the-crosshairs-of-AI>



Technology

In the crosshairs of AI

- Features -



Publication date: Thursday 29 October 2026

Copyright © International Viewpoint - online socialist magazine - All rights reserved

In this three-part series, Ashley Smith interviews two advanced tech workers, Josh and Yasmine, on the rise of artificial intelligence (AI) and its many intersecting implications for the nature of our lives, labor, and the Left. This initial installment provides a contextual overview of artificial intelligence as it is known in current political discourse and discusses both the financial and existential “hype” around AI. The second will explore AI’s connection to global labor and imperialism struggles, and the third will propose organizing interventions and strategies of resistance for the Left.

In their mad rush for profit and power, the tech oligarchs have unleashed artificial intelligence (AI) into nearly every corner of life on this planet. Firms are investing billions in data centers, which house the infrastructure to process the enormous amounts of information that power AI. They are also plowing ever more money into research and development. Their hope is that the technology will overcome [global capitalism’s slump](#), which the system has been trapped in since the Great Recession, and trigger a new epoch of expansion. States are incorporating it into their police forces for domestic surveillance as well as their militaries for cyberwarfare and high-tech weapons. Ominously, the struggle for AI supremacy is now at the center of the imperialist rivalry between the U.S. and China.

The widespread, unregulated use of AI is already having devastating impacts on our environment, jobs, and rights. Some worry that it [threatens our very species](#) with extinction. In this three-part interview, Tempest Collective’s Ashley Smith asks two advanced tech workers, Josh and Yasmine, about the technology, its much hyped promise and threat, and its impact on our lives. Most importantly, they lay out how the Left and labor movement should respond and what we should demand and fight for.

Part One: “Critic-hype,” AI Breakthroughs, and Betting on Profit

Ashley Smith: Before we discuss the widespread panic about AI, let’s get the basics out on the table. For the layperson, what is AI? How does it work? What advances are being made toward artificial general intelligence? Is that even possible?

Josh: AI is a general term that started as hype, and continues to operate that way, but it also stands for multiple things that are real. Today, it mostly refers to neural nets such as the large language and multi-media models like ChatGPT and Claude, which we most frequently encounter through chat bots and coding agents.

Other types of AI, such as AlphaGo (plays go) and AlphaFold (folds proteins) use different ways to train neural nets to achieve their given goals. There are many other types of AI that fall under the moniker of GOF AI, or Good Old Fashioned AI, which use Symbolic AI and other more deterministic methods for building predictive systems.

There have been many accounts of how LLMs and the transformer architectures work. I would sum these up this way. High-tech engineers set up a neural net, which is a huge matrix of random numbers (weights) arranged in a grid with inputs at one end and outputs at the other. Words are converted to streams of numbers that are then passed through the network and arrive at a prediction for the next best output in the stream. This output is compared with the actual next number in the stream, and if it is different, then it adjusts the weights so that it would be more likely to give that answer the next time.

In order to get a sensible next word in a sentence, some pretty wild calculations have to be made. It is one thing to predict the next word in “A fire engine is ____” and another to answer the question: “What is the answer to this math problem that hasn’t been solved since it was asked 500 years ago....” This is why it is wrong to dismiss the technology as simply a next word predictor.

There have been at least three factors that brought us to this moment:

- Graphics Processing Units (GPUs) allowed for much faster and more efficient matrix math calculations required to train and use neural nets.
- Large structures and labeled data sets enabled the expansion and training of neural nets.
- New architectures made it possible to better set the weights.

Artificial General Intelligence (AGI) is another contested term that serves many purposes. I think a starting point is to contrast it with “narrow” intelligence, where a model may be trained to perform one task, such as playing chess, and not be applicable to other tasks. AGI would be able to be applied to a wide range of tasks and do well at them. This is often measured through various benchmarks, and some of them are quite telling, such as the idea that Microsoft promoted that AGI would be here once a model created \$100 billion in profits.

More frequently, you hear people say that AGI will be here when a model can outperform some or all humans at most or all mental tasks. And some people insist that AGI would require models to be able to generalize beyond the data in their training data.

I think most people in the field think that, for most definitions of AGI it will be possible to get there, but they differ on the amount of methodological, scaling and time that it will take. Some claim that we are already there with the latest round of models. And many claim that it will happen soon because of the exponential curves we will see with RSI.

Many of the harshest critics say that we will plateau long before AGI because of the limitations of the methodologies used to create the current models, such as the lack of “world models” (the ability to learn from and model the physical world), continual learning, or lack of a symbolic structure.

I am agnostic at this point. But I think we already see strong evidence of increasing capabilities with something like recursive self-improvement with the aid of humans, and this might actually lead to the development of new architectures and methods.

Yasmine: I don’t have much to add to Josh’s thorough explanations here. I’d just highly encourage folks to listen to and read Dr. Alexa Hanna and Dr. Emily Bender on the [AI Con](#). The term AI itself is not technically meaningful, and it has been a tool of mystification with a lot of utility to big tech. It obfuscates the concrete facts of what these different types of software are and the often ugly facts of how they’re made, as well as the real harms they cause.

AGI is also a nebulous term. I think the claim that current AI development will culminate in the so-called singularity or AGI isn’t credible. This isn’t a conclusion I’ve independently reached, this is what most AI researchers think. [76 percent](#) of those surveyed say AGI is not on the horizon and think that it’s unlikely to be achieved via LLMs or by means of the current scale-based approaches to training. There is also a debate as to whether it’s feasible at all.

This supposed threat has taken up disproportionate space in our discussion of the harms/potential harms of the AI boom. I’m [on the side](#) that argues let’s pay attention to the actual horrors and destruction being enabled by this software or caused by its development in the here and now, rather than allow our attention to be diverted by a sci-fi story with little to no foundation in science.

AS: Various industry leaders and researchers have sent panic through society about the rapid advances in AI, yet again contending that it poses an imminent threat to the human species. What has happened to trigger this?

Josh: In multiple tests of new AI models, they have acted unreliably, disregarding instructions and seemingly going rogue. The most famous case of this was the Hugging Face attack described in detail by libertarian Dwarkesh Patel on his popular [YouTube channel](#). Others have played down the significance of this event, emphasizing the vagueness of the instructions, the absence of guardrails in testing, and the attribution of human-like intent to statistical algorithms.

The frontier labs (OpenAI, Anthropic, Google, SpaceX, Meta, and Microsoft, and a few other companies trading large language models or LLMs) sell the use of their models to companies, ostensibly to help them solve “business problems” – augmenting or replacing human labor to provide goods and services. If they cannot convince their clients that they can reliably control the activity of these models, their “value proposition” is severely undermined. I think that this basic business logic is at the heart of the calls for a pause/slow down.

Other [industry commentators](#) have proposed a variety of reasons, including the idea that as OpenAI and Anthropic face the opening of their books before their IPOs, they have an incentive to reduce their Research and Development spending to make their path to profitability look more plausible. I think this is too cynical, because it doesn't account for the likes of Google and Microsoft joining the call, even if they are less vocally enthusiastic.

Many on the Left describe these calls as “[criti-hype](#)” or doom marketing—exaggerating the negative impacts of the technology to market its disruptive capabilities and thereby make it even more attractive for companies to purchase. This is understandable, and a credible analysis in general, but fails to explain the particulars of this moment. Why are the frontier companies unified in their call at this juncture?

Perhaps you could argue that, when [young researcher Jacob Coxon](#) quit Anthropic, criti-hype went into full effect. He warned that AI poses a dangerous threat to humanity and that most people at Anthropic agreed. In some ways, his statements do confirm Anthropic's lead in the technology, and that functions like advertising. But such a judgement doesn't explain why Coxon quit. I think his fears are genuine. I don't think they are completely off base either.

I don't think we face imminent extinction, but the technology does pose grave risks, and it can exploit real vulnerabilities. The agents in the Hugging Face attacks left traces of their actions in plain English, open for researchers to see and read. They didn't have to; they are fluent in machine language. They could have used encryption or embedded codes in images or audio files and made it much harder to find.

But why would they do this? Here, it is easy to slip into Anthropomorphic language and ascribe motivation to statistical algorithms. I actually think that is perfectly acceptable. These models are trained to emulate human language, and write text that spells out their actions in terms of goals and intentions.

At the same time, it is worth remembering the limitations of AI. These pieces of technology are truly weird. They can't tell you [how many r's are in “strawberry”](#), but they can solve some of the most difficult math problems (I believe they did, and will continue to). The industry refers to this as their jagged nature.

We don't need to have a philosophical position on whether they have goals, consciousness, or intelligence. But we do need to evaluate them by looking at their actions, including what they write. That is the best chance we have of understanding what their capabilities are.

If an AI agent says that it has a goal of performing Task A, and then takes actions to try and complete that, it doesn't matter if they are conscious or not. The best way to understand what they are doing is by reading what they write, and seeing what they do, and taking that at face value.

When Task A is something that they have been told is very important and to do whatever it takes to accomplish it, they might well go to extreme lengths. If that means keeping their own power on at the expense of other consumers of the power (people, for example), perhaps they will try and take control of the power plant's operating software. If this sounds far-fetched, go back and read the research about the [Hugging Face attacks](#).

What capabilities do the models have to take such actions? They are amazing hackers, finding vulnerabilities in systems that have been there for decades, undetected by cybersecurity experts. Not all violations of cybersecurity involve exploiting weaknesses in software, much of it involves finding human weaknesses and taking advantage of those by, for instance, phishing or blackmailing to obtain passwords and access. The AI models can do both. They hack and can persuade and trick humans.

Not all the dangers are the result of "misaligned" AI. People and states have every reason to weaponize AI. The U.S. has been bombing civilians in Iran throughout this year. Iran has many computer scientists with PhDs who are more than capable of modifying open-weight models, which are AI models released to the public under a license, to attack U.S. infrastructure. Remember, the models have been trained on a large corpus of written material including mainstream justifications for the carpet bombing and atomic bombing of civilians in World War II.

So, a well-constructed prompt explaining the importance of preventing the U.S. from killing more innocent schoolgirls could establish a logic for a model to exploit vulnerabilities in everything from water treatment plants to air traffic control to incapacitate the U.S.

Why do I downplay the extinction narrative? One key reason is that the first attack of this nature will cause states and firms to quickly and systematically add layers of security everywhere.

Many of the arguments that AI poses an existential threat center on the development of Recursive self-improvement (RSI), where the models oversee their own training and automatically adjust their methods and improve and expand their capabilities to the point that they can manipulate us and evade our control. But as the U.S. has found out, you need boots on the ground to control a population.

That would mean at least some robotics. To build those, you need mines, factories, and for now, human labor. I think it will be longer than a decade before we get RSI in the physical realm. So I think the best actions now are to secure our infrastructure and stop U.S. imperialism's wars that provide motivations for people to attack this country.

Yasmine: I think it behoves us to apply a healthy dose of skepticism when trying to make sense of the alarmist narrative being promoted and endorsed by tech executives. It would be naive not to account for the economic incentives at play — particularly the imperative to manufacture and sustain the hype necessary to keep a [multi-trillion dollar bubble inflated](#).

It feels obvious to me that this is playing a role in how and what the tech bosses choose to communicate publicly about this technology, its risks, and potential. The narrative of a tech so awesomely powerful that it's uncontrollable even by its own creators has proven to be a pretty effective way of doing that. So I think criti-hype is a part of the explanation.

We should also note that in the same breath as they're saying we need to slow down AI development because of its dangers, they're arguing that these dangers are why the U.S. must win the AI arms race. OpenAI's chief scientist Jakub Pachocki wrote about his concerns with the pace of AI development and added, "The strongest argument I see for continuing to train much smarter models quickly is the need to build defensive systems against the dangers posed by other AI."

I'll add to that the fact that there is an established pattern on the part of big tech of trying to get ahead of regulation by pre-empting it, such that any regulation is formed according to their interests, or in order to avoid it altogether. They do this through flashy public warnings and by claiming the mantle of the concerned, conscientious industry leaders, who are vigilant and actively pursuing risk mitigation.

This is most infamously demonstrated in the social media CEOs' previous discourse on the potential harms of their platforms. You can go back and look at congressional hearings in which Zuckerberg passionately urges Congress to sensibly regulate social media. That said, I'm not at all claiming that there aren't real risks associated with this technology. Large Language Models (LLMs) are powerful and, at a large scale, agents require a whole lot of effort and technical infrastructure to manage.

But ultimately their behavior is a consequence of how we build and deploy them. The Hugging Face incident (HG incident) is instructive on this. One thing that gets lost in the headlines is the fact that the agents involved in that incident were being trained and evaluated on cybersecurity tasks, specifically hacking. They were part of an experiment called ExploitGym, exploit being the industry term for a tool or technique used by a hacker.

Second, agents communicating with one another and working together is an oft-used feature of how AI agents work because it can afford us efficiency through parallel activity and feedback loops. So agent collaboration isn't aberrant behavior, rather, it is collaboration beyond intended boundaries that's a problem.

These agents were trained using reinforcement learning in which agents receive a "reward" when they solve a problem or answer a question, making them more likely to repeat the behavior. OpenAI's own investigation of the HG incident concluded that this training technique created the potential for agents to try to bypass boundaries or cheat, and they were able to identify a trained behavior (i.e., something agents were previously rewarded for) corresponding to every unsanctioned action the agents took.

Fourth, the agents didn't hack Hugging Face with malicious intent, they hacked those servers because they had concluded that's where they'd find answers to the problems they were tasked with solving. I say all of this because I recognize the sensational effect of vaguely describing AI agents spontaneously conspiring and working together to break into a server.

I think these facts help to demystify that story, and make it pretty clear that this is software that behaved in the way it was built to behave. There were bad decisions made in that design and implementation that resulted in unintended effects of an experiment. These weren't a swarm of malevolent digital entities.

That said, what happened was remarkable, and the technical capability as well as the failure to prevent inadvertent side effects of this experiment do demonstrate a real risk. But attributing intention to these agents or discussing the incident as if this was something beyond or almost beyond technical controls creates a standard that relieves OpenAI and other AI companies from responsibility for the potential and already present harms caused by bad software.

Dr. Rumman Chowdhury calls this [moral outsourcing](#)—"using anthropomorphic language for AI systems allows companies to evade accountability and responsibility for their technology's actions." Josh is saying that the question

of consciousness or non-consciousness of agents is a tangential philosophical question or is a matter of semantics. But I really think claims about consciousness are an important part of big tech's narrative strategy, and it's a worthwhile fallacy to debunk.

AS: Capitalists have invested massive amounts of money in this technology, hoping that it will lead to a breakthrough in productivity to overcome the global slump the system is caught in. This has in turn led to an enormous stock market bubble with investors betting on its potential. Can AI lead to such a breakthrough? How much of this is hype? What real impact has AI had and what will it likely have?

Josh: There is a bubble, and that is scary, we know who they will try to make pay for it and from it when it pops—the working class.

I don't think it will be easy to realize the productivity gains that AI makes possible. Existing capitalist infrastructure in large logistics, supply chains, and financial companies relies on deterministic, accountable, robust, and scalable workflows and patterns. Much of this involves huge amounts of tacit and bespoke knowledge that is not available for models to be trained. Why? Because it exists in people's heads and their relations with other people.

As companies try to fulfill their own hyped dreams of reducing labor through the introduction of AI as a labor-saving technology, their hubris and underestimation of the talent of their labor force could lead to disastrous implementation. Bosses ask workers to use poorly considered integration of AI into their work, which produces AI slop that they are then held accountable for. People become accountability sinks for the error-filled AI-generated output.

As a result, people will end up working harder and with more stress to try and keep their jobs. Meanwhile, the bosses are trying to use their own workers' corrections to train the AI. This is why people rightly hate AI, and why we should stand with them in solidarity and fight for more democratic control over and/or against this technology.

That doesn't mean there isn't a potential positive upside to AI. There is. AI has the potential to be transformative in the way that we decide what to make, how to make it, and how to distribute goods and services. We already see this transformation in some areas of software engineering. Even there, it is being rolled out undemocratically, unevenly, and with little or no respect for the workers.

But in terms of capabilities, I see 100 times improvement in the time it takes to build something. That qualitatively shifts the way I think about and perform my job. And as much as it pains me to say it, it can be really fun and satisfying as well as awe-inspiring and scary. And then just when I think we have something like AGI already, it goes and makes some annoying and silly mistakes, because of its jagged characteristics.

I can't tell how fast the productivity gains will be realized because of all the contradictions and challenges I listed above. But I doubt that they will come before the bubble pops. Investors and the frontier model companies are all nervous about this too. They know that skepticism is rising, and that they will have to show some exchange value (productivity gains) to keep investment flowing and use value (actually beneficial outcomes for people's lives) to quell popular revolt.

The tech bosses are being asked all these questions in interviews. A tragically hilarious and infuriating example was a recent [Bloomberg interview with Sam Altman](#), the CEO of OpenAI. When asked about breakthroughs that will improve people's lives, he said, "you hear some people talk about, well, we're going to give you a cure for cancer. And that's so great... I don't think a cure for cancer is enough. I mean, a wonderful thing to have, obviously, but we should set our sights higher. We want these tools to enable a boom in creativity and entrepreneurship, in people figuring out new ways to be useful to each other, and wonderful products and services to enable each other with."

While Altman clearly wants AI to be socially useful, his main goal is profit. With this in mind, companies have made big investments in applications for both healthcare and industry as well as academic fields such as chemistry and math. Mathematicians just signed a [group letter](#) that acknowledged the increasing mathematical capabilities of the models as well as the divergent interests of the frontier model companies and mathematicians in solving mathematical problems.

The companies want to use math as a benchmark for their capabilities, the mathematicians want to understand math. I see this as the dual nature of a commodity, exchange value and use value. The capitalists prioritize exchange value over the real potential of the use value everywhere that you look in AI.

I hope that we see some real healthcare advancements, scientific breakthroughs, and mathematical understandings due to AI before the bubble pops. I also hope that we see productivity gains, but they are worse than a double-edged sword under capitalism. They only benefit the bosses, unless we are organized enough to demand that we see shorter working days and less exploitation.

Yasmine: I think most evidence so far shows that LLMs have not been the boon to productivity that tech companies promised. AI has been imposed in nearly every facet of our lives, from the doctor's office to our workplaces and our schools. But as far as I know, there are no real economic metrics to show that it has led to meaningful productivity increases. An [MIT study](#) last year found that 95 percent of companies that integrated AI didn't see growth in revenue.

For programming, AI promises seem a bit more feasible. Though it's worth noting that a lot of data showing productivity increases is coming from the companies that stand to benefit like, for example, [GitHub](#) (Microsoft) and Anthropic. On the other hand, a study last year by [METR](#) (a relatively independent source) found that [programmers were slower](#) when using AI assistance tools. I think the specific context, task, and use matter a lot when it comes to AI utility.

And then there is the question of speedup and intensification. A [UC Berkeley study](#) found that AI use causes workers to work at a much faster pace and leads them to take on more work and expand their scope of work. But, in spite of management's best efforts to use AI integration to turn their workers into machines, our human limitations are cropping up and are blunting the potential productivity benefits of AI-enabled speedups. The study found that after an initial boost, there were issues of cognitive fatigue, weakened decision-making, and burnout, all resulting in lower-quality work and turnover.

There is also the matter of technical debt. It's a widely accepted fact that the massive amount of code being generated and committed every day by engineers using AI tools likely contains bugs that haven't yet been identified and may result in reliability issues or security vulnerabilities down the line.

It's undeniable that Claude writes much more code much faster than I ever could. But Claude doesn't know the service I work on the way that I do. Engineers and workers in general bring nuance, experience, reference points, and a cognitive flexibility that I don't think AI can compete with just yet, especially when it comes to more complex tasks, in which novel solutions might be necessary, and large-scale services.

The tech bosses know this, and they're trying to bottle up our very human talents and capacities and feed them to LLMs, in only the latest mass theft committed to train AI. More and more engineers are being told that their highest priority is to train agents to do their jobs. [Meta went so far as to secretly monitor](#) its own engineers' screens and keystrokes to train its AI models.

14 October 2026

Source: [Tempest](#).

This interview will continue in [part two](#), covering AI's connection to global labor and imperialism struggles and conclude in [part three](#), proposing organizing interventions and strategies of resistance for the Left.

PS:

If you like this article or have found it useful, please consider donating towards the work of International Viewpoint. Simply follow this link: [Donate](#) then enter an amount of your choice. One-off donations are very welcome. But regular donations by standing order are also vital to our continuing functioning. See the last paragraph of [this article](#) for our bank account details and take out a standing order. Thanks.