Critique of Systemic Reason: Ilyenkov on Artificial Intelligence

Isa Jacobs, Sidney Sussex College Cambridge, 4 November 2026

In his collection of sci-fi stories *The Cyberiad: Fables for the Cybernetic Age*, written in the 1960s, Stanisław Lem introduced Trurl and Klapaucius, two inventors of all sorts of fantastical machines, including an apparatus that can create anything in the world beginning with the letter "n"—"nuclei" and "noses" but, apparently, not "Nothing":

> The machine sat still. Klapaucius rubbed his hands in triumph, but Trurl said:— "Well, what did you expect? You asked it to do nothing, and it's doing nothing."— "Correction: I asked it to do Nothing, but it's doing nothing."— "Nothing is nothing!" (Lem 2014, 5)

Another of their machines insists that two plus two is seven and then threatens the engineers if they fail to repeat its faulty result. Besides building cybernetic gadgets, Lem's engineers are immersed in the study of the General Theory of Dragons, a scientific paradigm dissatisfied with the simple fact that dragons do not exist. They believe that dragons can be non-existent in several different ways—a problem that can only be solved through cybernetics. There are "anti-dragons," "minus-dragons," and "imaginary dragons" (Lem 2014, 84).

They even found an entirely new field, "statistical draconics," a probabilistic theory of virtual dragons. Using a "dracotron," the draconical cyberneticians quantize a dragon, measuring the energy of its particles and even determining the coordinates of its tail (Lem 2014, 87). In Lem's parody of the cybernetics of his time, life is reduced to statistical fluctuation.

At the same time, Ilyenkov, too, was critically engaged in the cybernetic turn in the Soviet Union (after a rocky start, see Slava Gerovich, cybernetics was hailed by technooptimists as a new tool to control economic flows of distribution and decentralise socialist market planning). I will now turn to some ideas from a recent paper I co-authored with Jiří Růžička for *Marxism & Sciences* on dialectical critiques of artificial intelligence, cybernetics and machine thinking in Czechoslovakia and the Soviet Union, which places Ilyenkov in dialogue with the Czech Marxist thinker Josef Cibulka.

Ilyenkov and other materialist dialecticians viewed machine intelligence as primarily oriented towards systemic control, thereby risking to reproduce the alienating logic of capitalism as an automaton. Picture the beast of capitalism that Marx described in *Capital* I, Chapter Fifteen, entitled "Machinery and Modern Industry": "a mechanical monster whose body fills whole factories, and whose demon power, at first veiled under the slow and measured motions of his giant limbs, at length breaks out into the fast and furious whirl of his countless working organs."

From the perspective of Ilyenkov's dialectical, embodied and socialized conception of thinking, as a collective working through contradictions, intelligence can never be reduced to such a monstrous machine; social or *enactive* intelligence is neither a computational nor a cybernetic process. It cannot be captured by Lem's "statistical draconics" or other quantifying monsters which reduce qualitative differences to abstract functions and turn the particular human being into a generalised tool.

As Ilyenkov put it in a 1966 article on cybernetics, co-written with Anatoliy Arsenyev and Vasily Davydov (all three all members of the Scientific Council on Cybernetics), capitalist production is “a grandiose machine, that has become an end-in-itself, and has turned human into a means, a raw material for the production and reproduction of its insatiable organism. This grandiose machine, consisting of millions of partial machines, has gone beyond the control of human reason and will, it has become smarter and more powerful than any separate human individual that within it is playing the unenviable role of a cog.” The only way out of this machine, in order to liberate humans from their function as cogs, Ilyenkov and his co-authors write, is “the communist transformation of the whole system of relations between human and human.”

In May 1962, the Academy of Sciences, where Ilyenkov worked, cohosted the All-Union Conference on Philosophical Questions of Higher Nervous Activity and Psychology in Moscow. More than a thousand Soviet physiologists, psychologists and philosophers discussed the latest developments in computation and cybernetics and how it affected materialist-dialectical conceptions of human thinking within a socialist society.

In the Marxism & Sciences paper, we decided to frame the AI debate that played out in the Eastern Bloc in the 1960s and 70s as a struggle between *systemic* and *dialectical* reason: AI representing a form of systemic reason, concerned with processes of problem-solving, while dialectical reason is the creative and socially shared grappling with contradiction, with the logic of the other, or another logic. For Ilyenkov, thinking cannot be simulated by machines or grasped by measuring brain waves. We think through many organs: our hands, tools, friends.

Ilyenkov’s foundational disagreement with cybernetic conceptions of intelligence lies in his conception of mind: where cyberneticians claim that “it is *the brain* that thinks,” Ilyenkov insists that “it is not the brain that thinks, but the human with the help of the brain” (Arsen’ev, Ilyenkov and Davydov 2024, 157). One example is the famous brain in a vat experiment, an idea that fascinated both Feuerbach and Luria. For Ilyenkov, separated from the human body, the brain “will be thinking just as little as any telegraph pole or mould spread on the rocks of distant planets” (Arsen’ev, Ilyenkov and Davydov 2024, 157).

If one wanted to create a truly intelligent AI, they argue, it would not be enough to build an isolated brain but one would need to recreate an artificial version of the whole of human civilization – a truly *social intelligence*. While cybernetics has proven that “it is possible to reproduce only the machine-like functions of intelligence,” he argued in the 1966 essay, it can not generate intelligence itself whose special feature “consists precisely in the constant process of ‘sublation’ of any finite scheme, in revealing logical contradictions within this scheme and resolving them by further investigation of *the concrete reality*, namely this reality itself” (Arsen’ev, Ilyenkov and Davydov 2024, 164).

Artificial intelligence, for the Soviet dialecticians, “can be created only on the basis of dialectical logic and in no case on the basis of axioms and postulates of the logic of formal-mathematical constructions” (Arsen’ev, Ilyenkov and Davydov 2024, 164).

Ilyenkov pushed his critique of machine intelligence further in his 1968 sci-fi book *On Idols and Ideals* (Ob idolakh i idealakh) where he claimed that the binary logic of artificial intelligence was incompatible with truly human, that is, dialectical thought. Staging a dispute between a cybernetician and a philosopher, the book also criticised the famous mathematician Andrey Kolmogorov, an early proponent of artificial intelligence.

Artificial intelligence, for Ilyenkov, lacks two main qualities of thinking: mediation and imagination. Where machines only operate according to their own logic, human minds are able to mediate the logic of another through *ideals* (which Alan Díaz Alva recently called a "social algorithm").

Ilyenkov considered cybernetics—which detaches thinking from its social milieu—a false science that had regressed to pre-Marxist materialism. He strongly opposed the idea that thinking machines could be more intelligent than humans. For Ilyenkov, no computer can think dialectically; it can solve problems but not sublate contradictions. Humans can think through the logic of an other, he argues, whereas machines only operate according to their own precoded logic. A truly philosophical question, only resolvable through further study, will always remain uncomputable to a machine.

Machines can be media, in this framework, but never mediators. What they lack is imagination (*voobrazhenie*), the ability to come up with new things, to create objects of fantasy. It is imagination for Ilyenkov that enables human beings to see the world through the eyes of another, and to thereby actively participate in this shared world created by collective activity, within a system of ideals called culture.

It is important to note that Ilyenkov was not a Luddite: rather than being against machines as such he insisted on the necessity to transform human-machine relationships, in order to prevent machines from transforming humans into soulless slaves blindly serving the machine. For Ilyenkov, any human-machine relationship is at its core a human-to-human relationship mediated by a machine. It is not the machine itself that alienates humans but, as Ilyenkov examined in *Idols and Ideals*, it merely mediates existing alienated social relations.

A similar point can be made about Large Language Models (LLMs) today: rather than being a source of alienation, LLMs mediate and process existing socio-economic conditions. For Ilyenkov, humans need to consciously shape the ways they use and interact with machines in order to prevent becoming merely living flesh as an energy resource for the machine, consumed entirely and transformed into a calculating and labouring cog in the machine.

If the machine thinks abstractly, to quote Hegel, who thinks concretely? For Ilyenkov, thinking is a communal activity, "not a mental act that is taking place in the skull of the individual, in the secret spaces of the brain's grey matter" (Ilyenkov 2019, 53). Spinoza's "somatic monism" (a term by Andrey Maidansky) was foundational to Ilyenkov's post-Cartesian theory of thinking as an expression of collective, self-conscious activity. Like Spinoza, Ilyenkov considered the interaction between body and mind as an objective motion in the world.

For Ilyenkov, thinking manifests itself in "the entire 'non-organic human body' that stands objectively over and against an individual human being, the body of civilization, including tools and temples, statues and offices, factories and po- litical organizations, ships and toys—all that with which we are involved from the moment we are born and enter the human family." (Ilyenkov 2019, 13).

Body and mind, the two Cartesian halves, are united in one thinking body, which thinks by molding its own form to shape other bodies. Like blood, thinking streams through society's vessels and arteries, shocking through the entire system. Thinking is shaped by ideals, similar to the form of a jar growing under the hands of a potter, as Ilyenkov

described it in *Dialectical Logic*. Thinking happens within the interactivity of hands, clay, and tools. The ideal is situated neither in the piece of clay nor the body of the potter but arises from the activity of transforming the clay into a jar. Rather than a systemic code, the ideal functions as a social algorithm.

For Ilyenkov, thinking requires more than just a brain; it needs the whole of the human body and its social activity, embedded in a larger network of socio-economic relations. The brain needs other "organs," Ilyenkov contends, both organic and inorganic ones, such as tools and sense organs: "In particular it is the organs that provide the sensory-object contact of this brain with the world outside of it, something like eyes, ears, tactile hands and other 'external receptors'" (Arsen'ev, Ilyenkov and Davydov 2024, 158).

Ilyenkov's engagement with AI is most pronounced in his sci-fi story "The Mystery of the Black Box," a post-humanist tale featuring all sorts of non-human thinking machines, including a brain on spider legs, a lazy flying saucer, a deaf ear and a brainless set of hands. In their gatherings, the machines celebrate the overcoming of the human. They worship the Black Box, an advanced thinking machine that appears capable of resolving any problem. Nobody remembers when or how the Black Box was created; it was always simply there. If the Black Box is given two mutually inconsistent prompts, it produces no output but remains silent. Like Lem's machine, the black box ultimately does nothing.

What is the Black Box's secret? At the end, it turns out, unsurprisingly, that the box is empty. It was never able to think from the start but merely produced random outputs. Only communism, Ilyenkov concludes his parable, can put an end to any machine worship, discarding old ideals and bringing idols down from heaven to earth. Communism, I cite him, "is not a fairy tale about some bright future, but a real movement of modernity" (Ilyenkov 2020, 495).