AI and the HAL Revisited:
Why Computation Was Never the Binding Constraint

Roshan Shah
ECON 89: Asian Economic Systems
Prof. Steven Rosefielde
April 26, 2026

I. Introduction

In October 1970, a Ukrainian cyberneticist named Viktor Glushkov stood before the Soviet Politburo with a proposal that should have changed the twentieth century, namely OGAS, a real-time national computer network that would have linked roughly twenty thousand terminals at factories across the USSR while routing data through two hundred regional centers and into a central computing facility in Moscow. Glushkov had been working on the system since 1962 and had even drafted a plan to replace physical money with electronic payments, and the CIA had worried about the project for nearly as long as he had been building it. The plan died politically at the hands of Finance Minister Vasily Garbuzov, whose opposition helped push it into review limbo before the full technical case could prevail, because he understood that a working OGAS would render his ministry redundant.

That vote reveals the central argument of this analysis, because the hardest problem with AI-enabled planning was always primarily political and institutional, even when computation also imposed a real constraint. The same Politburo that funded ICBMs and Sputnik rejected a working national network in favor of what Garbuzov dismissively called simple computers to blink lights in hen houses, because the officials who had to approve the project were the same officials whose jobs it would eliminate. Half a century later, the question of whether AI can

finally make central economic planning viable has returned with the framing barely changed, except that we now possess something close to the computing power Glushkov could only imagine.

Rosefielde's Chapter 19, “New Age Planning,” anchors that question well, since his framing of AI as a double-edged sword cuts in both directions at once. On one edge, AI could in principle make socialist planning superior to markets, an optimistic HAL that finally tames the calculation problem, while on the other, AI in the wrong hands becomes the perfect instrument of tyranny, a HAL that surveils and allocates with no human friction. Rosefielde concludes that neither extreme proves realistic, and that the institutional pathologies which collapsed planning in the twentieth century still bind in the twenty-first.

My thesis follows from this framing: AI has substantially solved the computational dimension of the socialist calculation problem, but it does not, by itself and absent significant institutional reform, address the two things that actually collapsed planned economies, namely the motivation problem and the freedom problem. OGAS supplies the strongest historical illustration, China's surveillance state supplies the contemporary one, and even American surveillance capitalism runs on the same underlying logic, differing in the identity of those who control the objective function rather than in the structure of the problem itself.

II. The Calculation Problem and What AI Actually Solves

Understanding what AI changes requires first understanding what it leaves untouched, and the socialist calculation debate offers the right place to start. The Austrian economist Ludwig von Mises opened the debate in 1920 with an institutional argument: rational economic calculation collapses under socialism because prices for capital goods cannot exist without private ownership and voluntary exchange. The Misesian objection assumes that the planner has

all the relevant physical information about the economy, and shows that even in that case, the planner cannot rationally allocate resources, since without genuine prices generated by buying and selling under private property, there is no common metric for comparing one production method against another. A planning board might know the physical quantities of inputs and outputs, yet have no way to determine whether using a thousand tons of steel to build tractors is better than using it to build trains. Mises took care to grant that planners could be brilliant and well-intentioned, since the critique targeted the system's logic; even what he called a board of supermen would still fail (Mises, 1920).

Friedrich Hayek, another Austrian economist, deepened the argument across the 1930s and 1940s by recasting it as epistemological. In “The Use of Knowledge in Society” (1945), he distinguished two kinds of knowledge: scientific knowledge that can be written down and centralized, and knowledge of the particular circumstances of time and place, which remains tacit, local, and constantly shifting. This second kind of knowledge exists nowhere in concentrated form, and any attempt to transmit it to a central authority distorts it in transit. The 1945 essay emphasizes prices as knowledge-conveying devices, so that when the price of tin rises, millions of actors substitute and conserve without ever needing to learn why. Hayek's later work on competition as a discovery procedure strengthens this point by arguing that the process of competing in real markets is what generates the relevant information in the first place, which means central planners cannot simply collect the data after the fact, since the data does not exist absent the process (Hayek, 1945).

Two socialist economists, Oskar Lange and Abba Lerner, proposed a workaround in the 1930s in which a Central Planning Board would set trial prices, state managers would produce where price equals marginal cost, and the board would adjust prices in response to surpluses and

shortages until equilibrium emerged. This tâtonnement approach claimed to replicate competition while removing monopoly distortions, but Hayek's rebuttal cut to the core of the proposal by pointing out that the model already assumed the data existed somewhere accessible. In real markets, cost functions, technologies, and preferences only emerge through the process itself, through entrepreneurial discovery and the discipline of profit and loss under private ownership, and the Lange model therefore required the answer in order to compute the answer.

Lange himself seemed to concede this later in his career. In 1965, he wrote a paper suggesting that electronic computers could finally solve the millions of simultaneous equations central planning requires, and although he never quite said so explicitly, the move amounted to an admission that market socialism had served as a theoretical patch all along. The faith had quietly migrated from clever institutional design to raw computing power, and that migration forms the through-line connecting Lange to today's AI revival.

The contemporary version of the Lange move comes from W. Paul Cockshott, a computer scientist at the University of Glasgow, whose 2019 paper in World Review of Political Economy Rosefielde explicitly grants in Chapter 19, since a modern supercomputer running Cockshott's harmony algorithm can plan an economy with two hundred million distinct products in roughly twenty-two minutes, and a 2021 PLOS ONE paper extended the framework with AI-based planning models at scale. What Cockshott shows is that the arithmetic feasibility constraint, which Mises emphasized when he wrote in 1920 of the millions of equations a planning board would have to solve, no longer binds. What Cockshott does not show is that the deeper Misesian objection has been answered, since his labor-value framework still does not generate prices the way private exchange does, and Hayek's epistemological challenge about dispersed knowledge survives independently. Rosefielde concedes the arithmetic feasibility point

cleanly, and so should we, while keeping the underlying institutional and informational objections in view.

But the concession does far less work than it appears to, because Hayek's epistemological objection survives completely untouched. Hayek never argued about whether a computer could do the arithmetic; he argued about whether the inputs to the arithmetic existed in centralizable form. AI excels at optimizing over data it can see, yet it cannot conjure into existence the dispersed, tacit, constantly revised knowledge that prices encode through the act of being bid for. And even the data that planners can see corrupts the moment a planning system makes falsifying it rational, which is exactly what the Soviet experience demonstrates.

III. What AI Cannot Solve: The OGAS Proof

OGAS offers the strongest historical illustration we have of why the motivation and freedom problems are structurally sticky, even though OGAS itself was not AI planning and was never deployed at scale. The reason it works as an illustration is that it isolates the variable with unusual precision: the Soviets had the political will to plan, the engineering talent to build the network, and a leader in Glushkov who understood the system at the deepest technical level, yet what they lacked was a set of institutions that would let the system actually run. The argument is not that OGAS proves AI planning cannot work, but that the failures which doomed OGAS continue to recur in every system where AI tries to coordinate economic activity at scale.

The architecture itself was striking in its modernity, with a central computer center in Moscow at the top, two hundred regional centers occupying the middle tier, and up to twenty thousand local terminals operating at the base, all collecting economic data continuously and adjusting targets in real time. Glushkov's moneyless electronic payment proposal anticipated digital finance by several decades, and by 1962 he had calculated that the Soviet planning

bureaucracy would need to grow forty-fold by 1980 if it continued on paper, making OGAS the only alternative to bureaucratic collapse. US intelligence took Soviet cybernetics seriously, and Arthur Schlesinger Jr., then Special Assistant to President Kennedy, separately warned of an "all-out Soviet commitment to cybernetics" in a memorandum to the White House.

And then it died in a Politburo meeting, for reasons that had nothing to do with technical feasibility and everything to do with institutional incentives. The first reason traces to what I will call the motivation problem, and the historian Benjamin Peters, in his archival study How Not to Network a Nation (MIT Press, 2016), produced the best formulation of the failure when he observed that "the capitalists behaved like socialists, and the socialists behaved like capitalists." ARPANET succeeded because American institutions collaborated openly across institutional boundaries, while OGAS collapsed because Soviet ministries competed for turf at every level of the system. Each ministry understood that a centralized data network would expose its data to rivals, render bureaucratic layers redundant, and shift power from officials to computer scientists, so the incentive to obstruct was a rational response to a genuine threat. When the Politburo refused to fund the full project, OGAS disintegrated into hundreds of isolated, non-interoperable local systems known as the ASUP networks, with each ministry building its own incompatible version as a kind of defensive moat, and that fragmentation reflected the equilibrium that the institutional incentives produced (Peters, 2016).

Beneath that political failure lies a deeper one that any AI planning system necessarily inherits. Soviet managers and workers systematically falsified data, gamed plan metrics, and pursued private interests over plan objectives throughout the history of the command economy, out of rational calculation, since a manager who reported accurate data about his factory's true capacity would receive a harder target the following year while a manager who understated

capacity received a softer one. An AI optimizer requires honest inputs in order to function, but the planning system itself generates the incentive to falsify those inputs, and this problem grows more severe as the optimizer becomes more sophisticated, since a better algorithm produces better gaming rather than better compliance. A simple planning algorithm that tracks gross output gives managers one margin to manipulate, while a sophisticated AI system that tracks dozens of indicators and cross-references reported data against physical input consumption creates an adversarial dynamic in which managers must become correspondingly sophisticated, learning to game the cross-referencing logic and coordinate with inspectors whose own careers depend on the appearance of compliance rather than its reality. Garbuzov's veto captures the upper limit of this logic perfectly, since he blocked OGAS precisely because it would succeed, and that success would expose his ministry as redundant, which means the motivation problem operated at the very top of the system, well above the shop floor where falsification was more visible.

The second reason OGAS failed traces to what I will call the freedom problem. Any AI planning system must specify what exactly it is trying to maximize, and that specification represents a political act rather than a technical one, since whoever controls the algorithm's goals controls the economy. In the Soviet case the objective function was nominally social utility, but in practice the Communist Party set it, which means that moving the objective function from dispersed market actors to a central algorithm concentrated political power rather than depoliticizing resource allocation. Rosefielde puts the point sharply when he notes that workers, given the choice, might have preferred markets, private ownership, and entrepreneurship over state planning and behavioral control, and that the Soviet collapse therefore reflected Gosplan's

mechanical failures and, more fundamentally, the Party's unwillingness to sacrifice its sacred cows for the people's benefit (Rosefielde, Ch. 19).

Chile's Cybersyn project reinforces the same conclusion from a very different political direction, since it represents the most democratic version of the experiment we have. Designed by the British cyberneticist Stafford Beer for Salvador Allende's government between 1971 and 1973, Cybersyn combined a telex network called Cybernet, Bayesian forecasting through a system called Cyberstride, an economic simulator called CHECO, and a control room known as the Opsroom. The project achieved one documented success when it managed supply chains during the October 1972 truckers' strike using a skeleton crew of loyal drivers, and although Pinochet's coup in September 1973 ended the project before its full deployment, the structural limits had already become visible, since the system depended on voluntary data submission by managers with every incentive to distort their numbers and lacked the democratic accountability mechanisms Beer himself had hoped to build. Even the most well-intentioned AI planning system runs into the same motivation and freedom constraints (Medina, 2011).

Rosefielde's broader point holds that AI advances from 2010 to 2026 have resolved neither problem, since he notes that AI advanced rapidly without lifting aggregate GDP growth or improving COVID-19 management, and the reason is that institutions shape outcomes while AI does not change institutions. The same fragmentation that turned OGAS into hundreds of incompatible ASUP networks has a contemporary analogue in the way global AI governance has splintered into incompatible national strategies, with each jurisdiction optimizing for its own bureaucratic and political interests rather than for collective welfare. But these two problems, motivation and freedom, undermine central planning and also distort how AI operates inside market economies, which brings us to the neoclassical framework Rosefielde asks us to apply.

IV. The Pareto Lens: Does AI Pass the Efficiency Test?

The First Welfare Theorem holds that under perfect competition, complete markets, no externalities, and perfect information, the competitive equilibrium achieves Pareto optimality, meaning that no reallocation can improve one person's position without worsening another's, while the Second Welfare Theorem adds that any Pareto efficient allocation can be supported as a competitive equilibrium given appropriate lump-sum redistribution of initial endowments. These results provide the benchmark for evaluating what AI actually does to economic efficiency, and the verdict turns out to be genuinely mixed.

The case in AI's favor proves real and worth stating fairly. AI reduces search friction across labor markets, financial markets, and consumer markets, which by itself moves outcomes toward the Pareto frontier, and it pushes transaction costs toward zero for digital interactions, enabling forms of contracting and monitoring that previously could not scale, while it also improves price discovery in financial markets through algorithmic execution. The economists Andrew Berg, Edward Buffie, and Luis-Felipe Zanna, in IMF Working Paper 18/116 (2018), find that automation tends to be good for growth even when it worsens distribution, and they identify conditions under which growth gains can compensate for distributional losses, particularly when productivity gains diffuse widely and when policy responses redistribute the surplus. The growth case is real, even if the unconditional Pareto case is not.

AI has nevertheless created or sharpened at least three new market failures the standard framework was not built to handle. The first involves algorithmic collusion, and Emilio Calvano and his coauthors, writing in the American Economic Review in 2020, demonstrated that AI pricing algorithms left to interact in a repeated environment learn to sustain supracompetitive prices with no explicit agreement and no communication between them, since two algorithms

simply discover through trial and error that cooperation pays. The result delivers a Pareto-inferior, cartel-level pricing outcome that standard antitrust law cannot detect, since none of the behaviors the law was designed to punish have actually occurred, and the Canadian Competition Bureau documented the same dynamic in real digital markets in 2026 (Calvano et al., 2020).

The second involves data monopolies. Data functions as a non-rival input, meaning that one firm's use does not deplete it for another, yet first-mover advantages in data accumulation produce a structural lock-in that compounds with network effects. The legal scholars Tejas Narechania and Ganesh Sitaraman, writing in the Yale Law and Policy Review in 2024, argue that data concentration generates market power which standard antitrust analysis cannot reach, because data advantages accrue before any anticompetitive pricing behavior surfaces, and the result manifests as the winner-take-all dynamic that has produced the current set of AI incumbents (Narechania and Sitaraman, 2024).

The third comes from the MIT economist Daron Acemoglu, whose 2024 NBER paper estimates AI's contribution to total factor productivity at no more than 0.66 percent over ten years, and less than 0.53 percent once the analysis adjusts for the difficulty of automating hard-to-learn tasks where decisions depend on context and outcomes are not easily measurable. He also flags a separate concern about AI's negative externalities in what he calls “bad tasks,” where AI is deployed to generate misinformation, manipulative advertising, addictive social media engagement, or AI-powered cyberattacks, all of which contribute to GDP without contributing to welfare. There is a related but distinct point about the dignity and social function of work, which I find compelling but should not put in Acemoglu's mouth: a labor market that pays human workers less because AI does the same task more cheaply registers as efficiency in the national accounts even when the human displacement carries real social cost (Acemoglu, 2024).

Rosefielde's productivity paradox provides the empirical capstone for all three of these failures, since AI has advanced rapidly and yet aggregate GDP growth in advanced economies has not improved. The economists Erik Brynjolfsson, Daniel Rock, and Chad Syverson respond with a J-curve argument holding that gains from general-purpose technologies show up only after firms reorganize around them, the same way electrification took decades to fully translate into output, while Acemoglu counters that the current trajectory delivers the wrong kind of AI and that no amount of patience will fix substitution. The micro-macro gap proves telling, since individual task-level studies show AI improving performance by 20 to 40 percent while macro GDP impact remains near zero, and that gap reflects what happens when a powerful technology gets deployed inside unreformed institutional structures.

On Pareto grounds, then, AI passes the efficiency test in narrow competitive market settings, but the conditions that test requires, including no data monopolies, no algorithmic collusion, and augmentation rather than substitution, are precisely the conditions that concentrated AI deployment erodes. Rosefielde's institutional framework explains why this dynamic represents a structural feature rather than a temporary problem awaiting correction. But Pareto efficiency only measures whether resources are allocated without waste. It says nothing about whether the resulting society is free, democratic, or equal, and that is the gap Rosefielde's framework insists on filling.

V. Beyond Pareto: AI and the Non-Pareto Utilities

A society can achieve Pareto optimality while remaining deeply unfree, anti-democratic, and unequal, and standard welfare economics has nothing to say about any of that. Rosefielde insists that any honest evaluation must judge an economic system on three additional

dimensions, namely freedom, democracy, and egalitarianism, and on all three current AI deployments score badly.

Take freedom first, in the substantive sense of meaningful individual autonomy. The COMPAS algorithm, used in criminal sentencing across the United States, was shown by ProPublica in 2016 to rate Black defendants as high-risk 45 to 77 percent more often than white defendants with identical criminal histories. Amazon scrapped a recruiting tool in 2018 after Reuters reported that the system, trained on a decade of male-dominated hiring data, systematically downgraded resumes from women and even penalized phrases such as “women's chess club,” while a 2024 Lehigh University study found that LLM-based mortgage underwriting required Black applicants to carry 120 additional credit score points to receive equivalent approval rates. The political scientist Virginia Eubanks, in Automating Inequality (2018), extends the pattern into welfare eligibility and child protective services, where what she calls the “digital poorhouse” automates and scales the biases of the bureaucracies it replaces. The deeper point is that algorithmic decision-making encodes the preferences of whoever sets the objective function, and in market economies that group means advertisers and capital owners (ProPublica, 2016; Eubanks, 2018).

Freedom also carries an epistemic dimension that AI specifically erodes, since recommendation algorithms optimize for engagement rather than truth and therefore produce filter bubbles and progressive radicalization pipelines. A 2024 Nature paper identifies three structural violations of human autonomy from algorithmic personalization, namely cognitive dependency, reduction of choice architecture, and relational autonomy violations through behavioral modification, and the algorithms in question actively shape preferences as much as they respond to them.

On democracy, the empirical evidence has now grown strong enough to take seriously. The economists C. Y. Cyrus Chu, Juin-Jen Chang, and Chang-Ching Lin published a study in PNAS in 2025, titled “Why does AI hinder democratization?”, covering many countries over the past decade and finding an empirical association between AI/ICT advancement and democratic narrowing in fragile democracies and authoritarian regimes. The mechanism they emphasize is technology complementarity, since states enjoy structural first-mover advantages in data collection through citizenship records, tax records, and surveillance infrastructure that AI amplifies into political power, shifting the balance away from civil society and toward incumbent governments. The Carnegie Endowment's 2026 report identifies four major intersections of AI with democratic life, including elections, deliberation, government services, and social cohesion, and finds materially negative effects on all four under current deployment patterns (Chu, Chang, and Lin, 2025).

Disinformation forms the most visible piece of this story, since Frontiers in AI reported in 2025 that deepfake content grew 550 percent between 2019 and 2024. AI-generated audio of a Slovak election candidate days before the 2023 vote and a Biden impersonation robocall in the New Hampshire primary supply the documented cases, but the deeper democratic harm shows up in what scholars now call the liar's dividend, because when deepfakes saturate the information environment, even authentic content becomes deniable, and truth as a democratic resource degrades systematically.

On egalitarianism, UNESCO's 2024 study Bias Against Women and Girls in Large Language Models found striking patterns of bias in widely used systems, including Llama 2 outputs that placed women in domestic roles roughly four times more often than men and produced negative completions about gay people in around 70 percent of cases tested. NIST's

facial recognition evaluation showed false positive rates 10 to 100 times higher for Black and Asian faces than for white faces, with direct downstream consequences in criminal justice and border security where these systems get deployed. At the macroeconomic level, Federal Reserve data show the top 1 percent share of US net worth rose from roughly 28 percent in 2006 to roughly 31 percent in 2021, and a growing economic literature now documents that AI patent intensity within sectors is associated with declining labor income shares in those sectors, suggesting that Piketty's r > g dynamic, in which the rate of return on capital exceeds the growth rate of the economy, may now be reinforced by a technology that raises the return on capital relative to the return on labor (UNESCO, 2024).

The UNDP's December 2025 report, The Next Great Divergence, warns that AI risks widening the gap between developed and developing nations by foreclosing the manufacturing-led development path that lifted East Asia, and the economist Dani Rodrik's premature deindustrialization thesis grows more urgent in this context, since labor-intensive sectors such as garment manufacturing, which industrialization theory has long treated as a stepping stone, face significant disruption from AI-driven automation in the countries that depend on them.

The verdict across all three non-Pareto dimensions therefore reads uniformly negative. Freedom suffers as algorithmic decisions in criminal justice, employment, and credit restrict meaningful autonomy and encode bias at scale, democracy suffers as AI narrows democratic corridors empirically and erodes the shared epistemic ground deliberation requires, and egalitarianism suffers as AI concentrates returns on capital and encodes racial and gender bias. A system can pass a narrow Pareto test and still fail comprehensively on all three dimensions, and that gap reflects exactly what Rosefielde's framework exists to expose. The question, then, is what happens when all of these dynamics, the motivation problem, the freedom problem, and the

failures across Pareto and non-Pareto dimensions, converge in a single contemporary case. China provides the answer.

VI. China as the HAL Scenario

China supplies the closest real-world approximation we have to Rosefielde's tyrannical HAL, and the architecture of the AI state there reaches further than anything else currently operating. The Social Credit System, in simplified terms, draws on state administrative records, commercial data, social and behavioral data, and scoring assessments produced by both government and private actors, with real-world consequences spanning access to credit, travel, education, and employment, although in practice the system remains fragmented and uneven across jurisdictions rather than fully unified. Skynet, the urban camera network, and Sharp Eyes, its rural counterpart, had expanded into a surveillance infrastructure commonly estimated at over 700 million cameras by 2023, many with real-time facial recognition, and since 2024 all large language models deployed in China must pass a government review confirming that their outputs reflect “socialist core values.” AI in China operates as a tool of political direction built to reinforce Party authority.

The Next Generation AI Development Plan, released by the State Council in 2017, lays out a three-stage roadmap to global AI leadership by 2030, with AI positioned as both an economic tool and a political control mechanism simultaneously. The political scientists Xu Xu, Genia Kostka, and Xun Cao found in their 2022 study in the Journal of Politics that Chinese authorities actively manage public information about the Social Credit System to manufacture consent, since citizens support the system more the less they know about its coercive applications (Xu, Kostka, and Cao, 2022).

The DeepSeek shock in January 2025 cut in an unexpected direction. DeepSeek's open-source model matched GPT-4 performance at a fraction of the compute cost, triggering roughly a one trillion-dollar US stock market selloff, and the lesson is that institutional pressure shapes AI development more than raw compute does. US chip export controls forced Chinese researchers to optimize for efficiency in ways that delivered strong results, and that observation itself reinforces a Rosefielde-style point about the primacy of institutional context.

But even China's system runs into the same constraints that killed OGAS. The political scientists L. Jason Anastasopoulos and Jie Lian, writing in the Journal of Democracy in 2026, describe a calibration dilemma at the heart of authoritarian AI. Any predictive surveillance system requires a decision threshold, and lowering it generates collateral repression through false positives that produce backlash, while raising it produces blind spots through false negatives that miss real threats. The result is what they call threshold whiplash, meaning cycles of tightening and loosening that produce brittleness rather than omnipotence, since local officials falsify Social Credit data, ministries resist sharing data across jurisdictional lines, and the appearance of total surveillance functions as a governance tool in its own right because the system works partly because people believe it reaches further than it actually does, which Anastasopoulos and Lian call the panopticon bluff (Anastasopoulos and Lian, 2026).

China's most important structural advantage over the Soviets proves also the most ironic, namely that a functioning market economy operates beneath the surveillance state, providing price signals and competitive incentives that keep the productive sector running even while the political sector imposes top-down control. The parts of China's economy that perform best remain the parts least centrally directed by AI, and that observation captures the OGAS lesson

precisely: the tyrannical HAL scenario remains real and dangerous, but it self-limits through the same institutional pathologies that constrain all planned systems.

VII. Personal Assessment

My own view, after working through this material, holds that Rosefielde gets the central argument right and that the contemporary AI debate significantly underappreciates it. Almost every public conversation about AI frames itself as a technology debate concerning capability, safety, and alignment, while the comparative economic systems lens reveals what that framing leaves out, namely that AI governance remains inseparable from the institutional question of for whom an economic system operates and who controls the objective function. Technology cannot transcend institutions, since OGAS proved this in 1970, China's Social Credit System demonstrates it today, and the sociologist Shoshana Zuboff's concept of surveillance capitalism shows the same dynamic operating in market economies, where platforms deploy AI to extract behavioral surplus and sell prediction products to advertisers and insurers. Whether the algorithm gets controlled by Garbuzov's ministry, the Communist Party, or a San Francisco platform company changes the specific distortion while leaving the structural source of the problem intact.

Where I want to push back slightly is on what I take to be Rosefielde's underweighting of a third path between the two HALs, since his framework correctly identifies the institutional conditions that determine AI's impact while perhaps leaving too little room for the possibility that those conditions could be changed through deliberate democratic action. Acemoglu's 2024 NBER paper argues that AI could be redirected toward augmenting human workers rather than substituting for them, and that the deployment trajectory reflects a policy choice rather than a technological inevitability, while early work on democratic AI governance also deserves serious attention, including a Nature 2026 study on DAO-based AI deliberation, One Project's

framework for public democratic claims on AI, and DemocracyNext's AI localism proposals. None of this work refutes Rosefielde but rather extends him, since his framework correctly identifies the institutional conditions that determine AI's impact while leaving open whether democratic institutions retain enough strength to direct AI toward human welfare when both market and state actors carry overwhelming incentives to direct it elsewhere.

Where I ultimately come down is this: AI functions as a general-purpose technology in roughly the same sense electricity did, and its impact depends almost entirely on the institutional context in which it gets deployed, since inside unreformed market structures it amplifies inequality and market concentration while inside authoritarian states it amplifies surveillance and control. The right question is whether democratic institutions retain enough strength to direct it toward human welfare. Rosefielde's framework supplies the right tools for asking that question, and his verdict that the motivation and freedom problems remain binding accurately describes where things stand right now, although whether that verdict holds permanently depends on choices that remain political rather than technological.

VIII. Conclusion

OGAS died because it would have worked, and the people in the room understood that working would render them redundant. That dynamic has not gone away, since AI systems capable of genuinely optimizing for democratic welfare threaten the people who currently control the objective function in every existing system, including capital owners and advertisers in market economies and party vanguards and surveillance bureaucracies in state economies. The motivation problem operates at the very top, and that is the part of the analysis that gets missed most often.

AI has solved the computational dimension of the calculation problem, since Cockshott rightly shows that 200 million products can be planned in 22 minutes, but AI has not solved, and absent significant institutional reform cannot solve, the motivation problem or the freedom problem, with OGAS supplying the historical illustration and China's threshold whiplash providing the contemporary confirmation. On Pareto grounds, AI's case reads as mixed, since it can improve efficiency in narrow competitive markets while generating new market failures that erode the conditions Pareto optimality requires, and on the non-Pareto dimensions of freedom, democracy, and egalitarianism, the verdict reads uniformly negative across the range of evidence examined.

Rosefielde's double-edged sword therefore remains double-edged, and AI has not resolved the calculation debate so much as sharpened it by removing the one objection, namely computational infeasibility, that always carried the least philosophical weight. Both the optimistic HAL and the tyrannical HAL prove more plausible in 2026 than they appeared when Kubrick filmed the original, and yet neither outcome stands as inevitable. The binding constraint was always primarily institutions, motivation, and freedom, and that observation captures Rosefielde's lasting contribution to the debate.

Works Cited

Acemoglu, Daron. “The Simple Macroeconomics of AI.” NBER Working Paper No. 32487, 2024.

Anastasopoulos, L. Jason, and Jie Lian. “The Limits of Authoritarian AI.” Journal of Democracy, 2026.

Berg, Andrew, Edward F. Buffie, and Luis-Felipe Zanna. “Should We Fear the Robot Revolution? (The Correct Answer is Yes).” IMF Working Paper No. 18/116, 2018.

Boettke, Peter J., and Rosolino A. Candela. “Hayek and the Theory of Central Planning.” Review of Austrian Economics, 2023.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. “Artificial Intelligence and the Modern Productivity Paradox.” NBER Working Paper No. 24001, 2017.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. “Artificial Intelligence, Algorithmic Pricing, and Collusion.” American Economic Review 110, no. 10 (2020): 3267–3297.

Carnegie Endowment for International Peace. AI and Democracy: Four Intersections. Washington, DC, 2026.

Chu, C. Y. Cyrus, Juin-Jen Chang, and Chang-Ching Lin. “Why does AI hinder democratization?” Proceedings of the National Academy of Sciences 122, no. 19 (2025): e2423266122.

Cockshott, W. Paul. “How Feasible are Jack Ma's Proposals for Computerized Planning?” World Review of Political Economy 10, no. 3 (2019): 302–325.

Cockshott, W. Paul, and Allin Cottrell. Towards a New Socialism. Nottingham: Spokesman, 1993.

Eubanks, Virginia. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York: St. Martin's Press, 2018.

Hayek, Friedrich A. “The Use of Knowledge in Society.” American Economic Review 35, no. 4 (1945): 519–530.

Xu, Xu, Genia Kostka, and Xun Cao. “Information Control and Public Support for Social Credit Systems in China.” Journal of Politics 84, no. 4 (2022): 2230–2245.

Medina, Eden. Cybernetic Revolutionaries: Technology and Politics in Allende's Chile. Cambridge, MA: MIT Press, 2011.

Mises, Ludwig von. “Economic Calculation in the Socialist Commonwealth.” Archiv für Sozialwissenschaften, 1920.

Narechania, Tejas N., and Ganesh Sitaraman. “An Antimonopoly Approach to Governing Artificial Intelligence.” Yale Law and Policy Review 43, no. 1 (2024): 95–168.

Peters, Benjamin. How Not to Network a Nation: The Uneasy History of the Soviet Internet. Cambridge, MA: MIT Press, 2016.

ProPublica. “Machine Bias.” By Julia Angwin et al., May 23, 2016.

Rosefielde, Steven. Socialist Economic Systems: 21st Century Pathways. London: Routledge, 2023, ch. 19.

UNDP. The Next Great Divergence: AI and Global Inequality. United Nations Development Programme, December 2025.

UNESCO. Bias Against Women and Girls in Large Language Models. Paris, 2024.