Department of Economics Discussion Paper Series

The Political-Economic Risks of AI

Jean-Paul Carvalho

Number 1068
January, 2025

Department of Economics
Manor Road Building, Manor Road
Oxford, OX1 3UQ

ISSN 1471-0498

The Political-Economic Risks of AI*

First version: January 29, 2025

Jean-Paul Carvalho
University of Oxford

Abstract

The political and economic risks of artificial intelligence have been overshadowed by fears of malicious superintelligence and killer robots. Due to AI's distinctive features—automation of cognitive tasks, global scalability, general-purpose technology, and importance to national security—its impact could be unlike earlier rounds of automation. It is possible that AI creates a superabundant world with unprecedented human freedom. In this essay, however, I will explore a tail risk in which human-level artificial general intelligence (AGI) radically concentrates the global economy, breaks democratic and egalitarian institutions, and tears the social fabric, collapsing human productivity. The closest precedent would be the cultural devastation of indigenous societies by colonialism. I will describe how this process might unfold and propose measures to ensure AI has widespread benefits. Competition policy emerges as a critical tool, as do adaptive changes to political institutions. Without appropriate measures, there may be no AI-driven growth take-off and the inequality that emerges would dwarf anything experienced to date.

Key words: AI, political economy, existential risk, democracy

JEL classification: O3, P00, L00

Carvalho: Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford OX1 3UQ, UK, jean-paul.carvalho@economics.ox.ac.uk.

I am grateful for comments and suggestions by Peyton Young, Mason Edwards, Mark Koyama, Robin Lane-Fox, Raphael Mu, Stergios Skaperdas, and Cole Williams, as well as students in my Political Economy courses at UC Irvine in 2018 and 2020 who provided feedback on my early thoughts on the subject. I am also grateful to Raphael Mu for outstanding research assistance, along with Grok and ChatGPT. All errors are mine.

"*A state without the means of some change is without the means of its conservation*" [Burke, 2009/1790, p. 16]

1 The AI Revolution

1.1 What is the Yogurt?

In an episode of the animated series *Love, Death & Robots*, a scientific mishap leads to the creation of a sentient superintelligent yogurt [Robertson and Scalzi, 2019]. In exchange for solving the problems plaguing the US, the yogurt demands a 100-year lease on the state of Ohio. Otherwise, it threatens to provide its knowledge to China. The US government assents. The yogurt then builds labs in Ohio and speeds up its evolution. Eventually, it devises a set of instructions for eradicating the national debt within one year. It advises the US government that the instructions have to be followed to the letter or else the global economy will collapse. The government fails to follow the instructions precisely and catastrophe ensues. In return for an economic rescue plan, the yogurt demands supreme executive power over the United States. It uses this power to build its own space agency and launch itself into space. Of course, the superintelligent yogurt had already anticipated that the government would lack the discipline to follow the instructions precisely. It concocted the crisis and provided the solution only to fulfill its ultimate goal of becoming an interplanetary species.

'When the Yogurt Took Over' is obviously an allegory for the dangers of artificial intelligence (AI). But *what or who is the yogurt*? Without question, most viewers would point to the AI systems currently being developed as the yogurt of our near future. In that reading, the episode would be another dystopian vision of superintelligent AI escaping human control. There is, however, a more subtle and, I think, illuminating reading of the tale of the sentient yogurt. That is, the yogurt is the human economic and political elite that will emerge out of the AI revolution with new forms of power. It is not only whether humans control the AI that is important, but who controls those who control the AI. AI oligarchy and tyranny may not be

much better than AI Armageddon. Thus, the ancient political problem has not gone away: *quis custodiet ipsos custodes*?

1.2 Fixation on the Technological Problem

Faced with rapid progress it is easy to fixate on the technological problem of AI control and lose sight of the political-economic problem. We are in the initial stages of a technological revolution driven by artificial intelligence. Vast amounts of capital are currently being allocated to train AI systems [Stanford Institute for Human-Centered Artificial Intelligence, 2024]. Early AI models focused on simple classification tasks (e.g. image recognition), language translation, and playing games [Agrawal et al., 2022]. Recent developments are in generative AI, especially large language models (LLMs) that produce text from text-based prompts (e.g. ChatGPT), as well as text-to-image and text-to-video models (e.g. Dall-E). The imminent phase is AI agents who can autonomously execute complex tasks over time, including starting and running businesses, as well as policing and warfare. It is anticipated that artificial general intelligence (AGI) will be reached in under five years, and possibly much sooner [Okemwa, 2024, Altman, 2025], which is when the leading AI models will perform as well as any human across a wide range of cognitive tasks. Artificial superintelligence (ASI) will be achieved when these models exceed the capability of any human.

Even before reaching these milestones, AI is reorienting how businesses learn and make decisions (e.g. machine learning techniques), how workers commute (e.g. self-driving cars), how people are entertained (e.g. AI-generated music videos), how militaries fight wars (e.g. automated weapons systems), and how scientists and engineers innovate (e.g. AI tools for predicting protein structures). The 2024 Nobel prizes in Physics and Chemistry were awarded for work in AI to Geoffrey Hinton, John Hopfield, Demis Hassabis, and John Jumper. These AI pioneers were recognized not for their contribution to physics and chemistry *per se*, but for developing AI systems that could contribute new knowledge to these fields. That is unprecedented.

Naturally, this technological revolution brings massive potential rewards as well as risks.

The risk-reward assessment of AI has been dominated by two polar opposite views, the techno-utopian and techno-dystopian. The dystopian view is based on a sci-fi inspired vision of superintelligent AI escaping human control. In 2014, Stephen Hawking declared, "I think the development of full artificial intelligence could spell the end of the human race. Once humans develop artificial intelligence, it will take off on its own and re-design itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, could not compete, and would be superseded" [Cellan-Jones, 2014]. Once this occurs, it is supposed that the superintelligent AI will separate from humans and either leave us for dead, enslave us, or exterminate us. On the other side, the techno-utopians believe that AI will create a world of superabundance, free humanity from drudgery, cure cancer, discover clean and abundant sources of energy, and bring about a new age of prosperity. Recent work by economists Aghion, Jones, and Jones [2019b] bolsters this view. They develop a model in which AI systems boost economic growth by automating tasks previously performed by human labor. In fact, if AI automates the production of ideas, a singularity arises in which infinite income per capita is achieved in finite time. By 2016, Stephen Hawking had concluded that "the rise of powerful AI will be either the best, or the worst thing, ever to happen to humanity. We do not yet know which" [Cellan-Jones, 2016].

By fixating on the technology, both the utopian and dystopian views miss the powerful economic, political, and social forces already being unleashed by the AI revolution. A sensational conflict with malevolent superintelligence and killer robots is not the only or even most immediate risk of AI. It is the more subtle and systemic risk of AI destabilizing the current political-economic equilibrium, radically altering the balance of power in society, and tearing the social fabric. Again, the pressing problem is not AI vs Humans, but rather Humans vs Humans in the wake of AI. In the scenario I will explore in this essay, there would be no AI-driven growth takeoff as the optimists predict. In fact, GDP could shrink dramatically if AI and robots replace humans. In

addition, the scale of inequality that would emerge would dwarf anything Thomas Piketty worries about in *Capital in the Twenty-First Century* [Piketty, 2013]. Thus, a fixation on AI control, to the exclusion of all else, could be a dangerously limited approach.

1.3 The Political-Economic Problem

The primary concern of economists is how AI-driven automation will affect the labor market [e.g. Brynjolfsson and McAfee, 2014, Susskind, 2020, Deming et al., 2025]. To understand the labor market effects of technological shocks, the ideal starting point is Acemoglu and Johnson's magisterial economic history of technology *Power and Progress* [Acemoglu and Johnson, 2023]. The main lesson drawn by Acemoglu and Johnson from their broad sweep of the historical record, including the impact of the plough, windmill, jacquard loom, digital technologies, and artificial intelligence, is that technological progress does not automatically produce shared prosperity. Whether new technologies raise wages depends on institutions related to power (e.g. the ability of workers to collectively bargain and strike) and the nature of the technology itself (e.g. whether it substitutes for or complements human labor). Without the right conditions, new technologies could end up enriching a few while reducing living standards for the many. Thus, Acemoglu and Johnson believe the unalloyed techno-optimism that prevails today is misplaced.

With regard to AI, Acemoglu and Johnson expected that the technology was limited in its potential: "Reassuringly, AI does not seem to be advancing so much that it will create mass joblessness ... Its most likely impact is to further lower wages for many people, not create a completely workless future" [Acemoglu and Johnson, 2023, p. 304]. Thus, they envisioned a degree of 'so-so automation' of tasks in which humans perform only slightly worse (or perhaps even better) than machines. That would mean widespread human displacement from the workforce without significant productivity gains. Acemoglu and Johnson advocate a new approach to AI which they call 'machine usefulness', in which AI development is directed toward boosting

human productivity and furthering human objectives.

Power and Progress is an impressive and regular Political Economy analysis of AI. It is regular in the sense that it covers likely AI scenarios that are negative, but not catastrophic. AI is viewed as limited, but nevertheless displaces some humans from the workforce, lowers wages, and reduces standards of living for workers. In this essay, I propose an *irregular* Political Economy analysis of the risks of AI. I will explore a world in which AI progress goes faster and further than initially expected and zoom in on a catastrophic tail outcome in which most of the human population is immiserated and possibly exterminated. In this scenario, AI would not be a local shock to the political-economic equilibrium, but a phase transition. As we shall see, this outcome could be produced by a number of mechanisms that vastly amplify the political-economic impact of technological developments in AI. As such, it could be that previous rounds of automation are not our best guide to the future. A more instructive example might be the catastrophic outcomes experienced by indigenous societies under colonialism, due to their cultural alienation and technological disruption [Fanon, 1961, Wolfe, 2006, Kirmayer et al., 2009].[1]

1.4 What is Different about the AI Revolution?

Technological progress is almost always accompanied by tech hyperbole. It would be comforting to think that AI is just a regular development in a long history of automation. If so, its impact will be muted and qualitatively similar to earlier waves of automation that did not end in mass human unemployment and immiseration. However, I see two problems with this view.

First, there is nothing normal about the impact of prior waves of automation. Technology is the primary engine of change in human history. The major leaps in tech-

[1] A less extreme example is the fate of the samurai after the Meiji restoration. At some point, the most skilled swordsmen in the world armed with the greatest swords ever made (*katana*) were no match for conscripted peasants with modern rifles and artillery. Their position in society, traditions, and belief systems were about to evaporate [Sonoda, 1990].

nology, from the development of agriculture to industrialization and now the digital economy, have not only expanded our ability to produce things of value. They have also revolutionized how we organize ourselves socially and transformed us physiologically and psychologically. Our postures, caloric intake, heights, and immune systems have changed [Fogel, 2010], as have our mental models of the world, our moral intuitions, our aesthetic preferences, and our gods [Norris and Inglehart, 2004]. In fact, the last major wave of automation—the industrial revolution—was a cataclysmic event which Karl Polanyi described as 'The Great Transformation' [Polanyi, 2001/1944]. Its destruction of traditional modes of living and social institutions led to widespread political upheaval and the threat of socialist revolutions. To stave off this threat, all advanced economies built a welfare state including unemployment insurance, pensions, and public education and healthcare. Over time, these obligations have expanded to the point of being largely unsustainable. This realization, I suggest, is partly behind the surge in political populism in the United Kingdom, the United States, Germany, France, and Italy, nations which were at the forefront of the industrial revolution. That is why policymakers in advanced economies are desperate for a new round of technological progress to rescue the current political-economic system. Enter AI.

Second, the AI revolution has distinctive features both of kind and degree that give it even greater potential for economic, political, and social upheaval than prior forms of automation. The coming changes have been billed as a new industrial revolution, but that is not correct. In my view, the AI revolution is a distinct phenomenon that will (i) give birth to a new economy, (ii) shift the balance of power both within and between nations, and (iii) radically alter the structure of society, including whom we interact with, whom we trust, and how we acquire meaning and purpose. To understand how this might happen, we must first identify precisely what are the distinctive features of AI. They are:

1. *Automation of cognitive (not physical) tasks.* During the industrial revolution, certain manual tasks were automated that humans are not especially good at.

This allowed us to move up the cognitive ladder and specialize in tasks requiring intelligence, which is where humans excel vis-à-vis other animal species. In the AI revolution, humans are being outcompeted in cognitive domains which were once their preserve. As AI systems become more capable, human cognitive limitations will begin to bind, making it difficult for displaced workers to shift to higher-valued cognitive activities. There may be no more rungs left on the cognitive ladder for us to climb.

2. *General-purpose technology.* As AGI is approached, the range of cognitive tasks that can be automated by an AI system will be much larger than the limited number of physical tasks automated during the industrial revolution. The equivalent technology for automation of physical tasks would be general-purpose mobile robots that can perform a wide range of physical tasks in domestic and industrial settings. The industrial revolution did not come close to such a general-purpose technology. Because AI will become a critical input into a wide range of cognitive tasks, there is scope for greater market power and concentration than ever before.

3. *Global scalability.* As AI applications will be cloud-based or downloadable anywhere in the world, the market power of AI companies will not be constrained by the geographical location of their operations. A few leading AI companies located in Silicon Valley and China could dominate the industry globally. This global market concentration is all the more consequential because AI is not a final product, but is (or will be) a critical factor of production. In short, the leading AI companies will control a large part of the world's cognitive labor force.

4. *National security threat.* Due to their general-purpose nature, AI systems could discover information and knowledge that threatens national security. Industrial production was not essential to national security except during major wars, which is precisely when factories were nationalized or placed under state

control. (Henry Ford converted his factories from producing cars to bombers during World War II.) The AI sector is different due its obvious military applications (e.g. AI-powered drones, robot soldiers) and potential for discovery of new sources of energy, propulsion systems, and weapons of mass destruction (WMDs) [Aschenbrenner, 2024].[2]

Because of these distinctive features, it is appropriate to entertain the possibility (however unlikely) that AI will have extraordinary consequences for humanity's future. Much will depend on how fast and far progress in AI technology goes. It could well be that this time is no different, that technological progress will slow down and AI adoption will have a regular (though partially negative) impact on society, much like previous waves of automation. However, we need to account for the (increasingly likely) possibility that AI grows more capable and sets in motion political-economic feedbacks with explosively bad outcomes.[3] This essay explores how such a scenario might unfold.

1.5 Three AI Amplifiers

The distinctive features of AI give rise to a number of mechanisms that could vastly amplify the economic, political, and social impact of the AI revolution. This essay focuses on the following three amplifying mechanisms:

- *Economic concentration and political power.* If advanced AI systems continue to require large capital expenditure and proprietary knowledge, AI could lead to unprecedented market power and concentration of the global economy.[4] The process unfolds as follows. Due to the pressures of geopolitical competition and

[2]AI is more like nuclear technology which can be used for peaceful purposes (i.e., electricity generation) or for producing WMDs. Except it is a general technology for producing a range of WMDS, including new chemical and biological weapons. See Aschenbrenner [2024, IIIb.] for an illuminating discussion of the security issues surrounding AI.

[3]For a range of different scenarios, some far more optimistic and utopian, see Tegmark [2017].

[4]I consider alternative scenarios where AGI and ASI are cheap to develop and run in Section 3.1.4.

the desperate need for growth, policymakers carve out a separate legal and regulatory sphere to supercharge AI development. This gives AI companies an edge in acquisitions and produces an 'AI black hole' in which a large part of the global economy (e.g. energy, logistics) is absorbed into this parallel economy located mainly in the United States and China. Due to security concerns, these behemoths become private-public partnerships which are incorporated into the security state. Thus, ownership and control of the global economy becomes extraordinarily concentrated. As a consequence, the AI sector becomes a powerful political actor with the ability to launch the equivalent of national strikes on countries and harm disfavored firms by denying access to critical applications. Thus, market power will translate into broader forms of economic and political power both domestically and globally.[5]

- *Breaking of democratic and egalitarian institutions.* With more capable AI and humanoid robots, the mutual dependency between financial capital, human capital, and labor that has been present throughout history will be broken. Productivity, innovation, production scale, and military power will no longer be tied to the human population of a country. Thus, economic and political elites will have incentive to unwind the progressive reforms of the twentieth century (e.g. public education) and there will be no incentive for elites to support proposed solutions to labor displacement such as universal basic income (UBI). There could also be a radical social disembedding in which elites sever ties with the 'masses' and retreat into isolated palace households. Democratic and egalitarian institutions built on empathy, national identity, and mutual dependence between classes will break. Indeed, the economic and political elite might see no reason for the masses to exist once it is thermodynamically more efficient to produce the services they want using AI and robots.

[5]This first amplifier depends on the degree of protection from competition (or 'moat') the leading AI companies will have, in particular, how easy and cheap it will be to catch up to the frontier models. The second and third amplifiers below do not.

- *Human productivity collapse.* As AI automates cognitive (not physical) tasks, its adoption may (unlike the technologies of the industrial revolution) undermine human cognitive performance. In discussions about AGI, it is typically assumed that AI capabilities are measured against a *fixed* human benchmark. Even if AI becomes generally more capable than humans, it is thought that humans will have a *comparative advantage* in certain tasks and specialize in those. On the contrary, competition with AI systems may lead humans to reduce their investment in learning new skills and have fewer successful human role models to learn from. As such, the widespread adoption of AI could generate a collapse in human productivity, bringing about rapid AGI and human obsolescence in the workforce. It could become thermodynamically efficient to replace humans with AI and robots, even in areas of human comparative advantage.

I will elaborate on these AI amplifiers and the dangerous path which they could take humanity down in Section 3.

The narrative of economic, political, and social disruption I set out is alarming. Managing these risks requires confronting them, which is an unpleasant task. People have some sense of the coming labor market effects of AI and humanoid robots. But evidence suggests they are inclined to ignore the consequences for their own lives. According to a report from July 2023 [Pew Research Center, 2023], 62% of survey respondents believe that AI will have a major impact on the workforce in the next twenty years. But only 28% believe their job will be impacted. Only 2% believe AI will have no impact on the workforce over the next twenty years. But 19% believe there will be no impact on their job. Wishful thinking will be even more tempting when it comes to the broader political-economic risks of AI, which go well beyond fears of human displacement from the workforce. Nevertheless, we cannot afford to ignore these potentially catastrophic risks.

1.6 Solving the Political-Economic Problem

Assessing the consequences of a new technology is necessarily a speculative exercise. When it comes to AI, however, it is important to look and think before we leap. According to my thesis, the AI revolution has the potential for massive economic, political, and social upheaval. Thus, it would be dangerous for policymakers to take a narrow technocratic approach to AI by focusing exclusively on AI safety. Instead, a political economy perspective is needed to understand the broader risks of AI and formulate policies that preserve our shared prosperity, freedoms, and social bonds. The industrial revolution did not end in catastrophe not simply because it lacked the amplifying mechanisms of the AI revolution (listed above), but also because drastic political changes were made including universal suffrage and the welfare state. While opinions differ on the desirability of various public programs, most people agree that there should be a basic layer of protection for the weak in society from predation and destitution. AI threatens even this minimal provision.

To limit the tail risks of AI, and ensure that AI has widespread benefits, the following changes to economic policy and political institutions are required. First, market power in the AI economy must be contained, without stifling innovation. Thus, competition policy emerges as the critical economic measure. I suggest that competition policy be reoriented from a focus on consumer welfare to limiting the accumulation of economic control. Second, democratic and egalitarian institutions must be strengthened to preserve basic freedoms, social bonds, and shared prosperity. One proposal is sortition, where a fraction of legislative representatives and minor officials are chosen at random from the population. Third, advanced nations outside of the US should develop their own sovereign AI systems to maintain the international balance of power. I describe each set of measures in detail in Section 4 and touch upon further social interventions in the Conclusion.

Without such measures, the AI revolution could massively concentrate the global economy, undermine democracy, and create a radical separation between a narrow

elite who live in prosperity and peace and the masses who live in an underworld of insecurity and deprivation. Deemed expendable, the masses might even be left for dead, or worse. Behind the 'veil of ignorance'—that is, before anyone knows whether they will end up in Elysium or Hades—nobody would choose this kind of outcome.[6] That is why it is important for policymakers to act now, before special interests fully crystallize and before we know where the chips will fall.

2 The Genesis of the AI Revolution

To better understand the political-economic risks of AI, we need to place the AI revolution in historical context and examine how it has unfolded to date. Those familiar with AI and the history of automation could skip this background section and go straight to Section 3. Note, however, that Sections 2.6-2.8 provide a novel argument for why governments will not slow down AI development.

2.1 Human Co-evolution with Technology

As a starting point, consider a (very) brief history of how humans have co-evolved with technology. For about the first 190,000 years of our history, anatomically mod-

[6]In Greek mythology, Elysium is a paradise set apart from the world, the exclusive abode of the gods and of the heroes favored by the gods who are granted immortality. Hades is the realm of suffering and despair. The word is sometimes translated as 'hell' in the Christian New Testament. The Roman poet Virgil describes the Elysian fields in book VI of his epic poem the *Aeneid* [Heaney, 2016]:

> "They came into happy vistas and the green welcome
> Of the Groves of the Fortunate Ones who dwell in joy.
> Here a more spacious air sheds brightness
> Over the land; they enjoy their own sun here
> And their own stars . . ."

The depths of Hades are far less salubrious:

> "Sounds of groaning could be heard inside, the savage
> Application of the lash, the fling and scringe and drag
> Of iron chains . . ."

ern humans lived in nomadic bands of 25-50 people [Lee and DeVore, 2017/1966, Kelly, 2013]. These small-scale societies were highly egalitarian, did not trade, and split when they became too large. The second stage of human evolution began around 10,000 BCE with the advent of settled agriculture. The neolithic revolution transformed human economic and social organization creating large states, hierarchies, writing, trade, coinage, and complex religions [Barker, 2006, Cauvin, 2000]. The third stage was marked by the industrial revolution.

Beginning in Britain around 1760, the industrial revolution spawned a new world and worldview built on the ability to manufacture goods that people wanted (or would soon want) cheaply and at scale. This technological revolution was achieved by combining inventions for harnessing energy from fossil fuels (e.g. the steam engine), new industrial machinery (e.g. the power loom), and new ways of organizing production (e.g. the factory) [Mokyr, 1992, Allen, 2009]. Between 1700 and 1840 in Britain, male employment in agriculture fell from 61.2% to 28.6%, while male employment in industry rose from 18.5% to 47.3% [Crafts, 1985]. Out of this industrial revolution came a new system of industrial capitalism, the creation of new merchant and working classes (the 'bourgeoisie' and 'proletariat'), rapid urbanization, and consumerism [Smith, 2008/1776, Marx, 2024/1867, Weber, 2019/1921, Offer, 2006]. At each stage, the human experience was radically altered.

Today, our world is being shaped by another engine of change—artificial intelligence. Mustafa Suleyman, CEO of Microsoft AI and co-founder of DeepMind, has gone so far as to call AI a 'new digital species' [Suleyman, 2024]. To understand the political-economic consequences of the AI revolution, let us briefly review recent technological developments.

2.2 AI Breakthroughs

Table 1 presents a timeline of developments in AI this century. To get a sense of the pace of progress, consider a particular application—image recognition. In 2012,

the leading image recognition program, AlexNet, had more than three times the human error rate. Humans were surpassed just three years later. By 2017, the leading AI programs functioned at better than half the human error rate [Kang et al., 2020]. One way to gauge expectations of further AI progress is to measure inputs. Between 2018 and 2023, global private investment in AI rose from just over $40 billion to around $96 billion [Stanford Institute for Human-Centered Artificial Intelligence, 2024]. The US government increased spending on AI-related research from $1.38 billion to $3.33 billion during that time [Stanford Institute for Human-Centered Artificial Intelligence, 2024]. Investment has shifted dramatically toward machine learning and generative AI. Most of this capital expenditure has come from the 'Magnificent Seven'—Apple, Nvidia, Microsoft, Alphabet, Amazon, Meta, and Tesla. Today, they are the seven largest US-based companies in terms of market capitalization. Thus, the economy is already being split into old and new sectors.

The progress in AI has been driven by three factors: a shift to different AI models, rapid growth in computing power ('compute'), and big data.

In 1996, when IBM's DeepBlue beat then world champion Garry Kasparov at chess, it relied on rules programmed into it by expert chess players. By 2010, progress in these symbolic AI models had stalled, leading to widespread pessimism about the capabilities of AI. The huge leaps in AI in recent years were unlocked by a different approach called *deep learning* based on (multi-layered) neural networks [see Mitchell, 2019]. These AI models are not directly programmed by humans. Rather they teach themselves, learning from data and from their own mistakes in ways that we often do not understand [Rumelhart et al., 1986]. They also exhibit signs of 'transfer learning', that is, generalization of abilities from one task to another. For example, fine-tuning an LLM on mathematics improves its ability to identify and categorize important structures in text [Prakash et al., 2024].

In 2016, when Google DeepMind's AlphaGo program beat Lee Sedol, the world's number one player of Go (an ancient game that is more complex than chess), it

Table 1. A Brief Timeline of Developments in AI since 2000

2000	Kismet robot head developed (affective computing)
2002	iRobot releases first Roomba robot vacuum cleaner
2004	Spirit and Opportunity rovers autonomously navigate Mars
2005	ASIMO robot can walk at a human pace and deliver trays in a restaurant
2009	Google starts testing self-driving cars in San Francisco
2011	Watson wins Jeopardy! (NLP, Information Retrieval, Reasoning, Machine Learning); Siri launched
2012	AlexNet scores a sweeping victory in the ImageNet challenge and starts the deep learning revolution; Google Now launched
2014	Cortana launched; GAN architecture invented
2015	Highway network and ResNet introduced; open letter calling for research on societal impacts of AI
2016	AlphaGo defeats Go champion Lee Sedol 4-1
2017	Transformer architecture released
2019	OpenAI releases GPT-2
2020	GPT-3 released; AlphaFold 2 wins CASP
2021	GitHub Copilot released
2022	ChatGPT (GPT-3.5) released; DALL-E 2 launched; Midjourney launched; Stable Diffusion released
2023	Google launches Bard; Meta releases Llama, Llama 2, and Code Llama; Anthropic releases Claude and Claude 2; GPT-4 (multimodal) launched; Gemini 1.0 launched; DALL-E 3 launched; open letter calling for six-month pause in AI development
2024	Gemini 1.5 released; Llama 3 (multimodal) released; Claude 3 Opus and Claude 3.5 Sonnet released; Sora released; AlphaProof wins Silver medal in IMO; Flux.1 launched; Apple integrates ChatGPT into new iPhones and Siri; OpenAI launches GPT-4o, o1, and o3; xAI launches Grok models and Aurora; Chinese LLM DeepSeek-V3 matches GPT-4o and Claude 3.5
2025	DeepSeek-R1 released; OpenAI launches "Operator" agent

was not programmed by expert Go players. Instead, it learned how to play Go by watching past matches between strong players recorded on online Go sites. After that, it played against itself tens of millions of times. In this way, AlphaGo mastered and surpassed thousands of years of knowledge about Go in a matter of months. Its 4-1 victory over Lee Sedol is brilliantly captured in the documentary *AlphaGo*

[Kohs, 2017]. Lee Sedol was not only beaten by the machine on the board, he was also psychologically destroyed. It turns out that the program was not merely a 'stochastic parrot', but exhibited creativity. Eight months later, Lee announced his retirement, stating that even if he were to remain the top human player, "there is an entity that cannot be defeated". Since 2016, the way humans play Go has been revolutionized by strategies played by AlphaGo. For a brief period, the machine learned from us. Now we learn from it.

Neural networks have been studied since the 1940s [McCulloch and Pitts, 1943, Rosenblatt, 1958]. Deep neural networks, with multiple hidden layers, have been around since at least the 1980s [Rumelhart et al., 1986, Krizhevsky et al., 2012]. What then explains the recent breakthroughs? Though there have been further technical advances such as the Transformer architecture [Vaswani et al., 2017], progress in deep learning has been largely a matter of scaling up existing models with massive amounts of data and compute, both of which have only recently become available. Current models have around one-trillion connections (weights) between their simulated neurons. The equivalent for humans is 100 trillion synaptic connections. The scale of investments being made are based on the expectation that each time a model is scaled up, new capabilities emerge. This emergent property came to the attention of the world when ChatGPT was released by OpenAI in November 2022. ChatGPT reached 100 million users in two months, the fastest penetration of a technology ever, 30 times faster than Facebook, and 60 times faster than Netflix.

Exponential growth in computing power was famously forecast in 1969 by Gordon Moore, co-founder of Intel. According to Moore's law, the number of transistors on a chip would double every two years. Moore's law was actually exceeded for decades. Though progress in fitting transistors on chips has begun to slow, the amount of compute available to the leading AI companies is astounding. AlphaGo, the program that beat Lee Sedol in 2016, performed 1.9 million petaFLOPS. Note that one petaflop is one thousand trillion floating point operations per second (a proxy for compute). The compute required by AlphaGO was 1.9 million times that

amount. Only four years earlier, the most advanced image recognition AI, AlexNet, used a mere 1/4000th the amount of compute. Only six years after AlphaGo's famous match, Google's program for solving math problems, Minerva, used over 1400 times the compute [Sevilla et al., 2022]. This massive demand for compute has been a huge boon for advanced chip makers, especially NVIDIA whose market capitalization rose from $338 billion before the release of ChatGPT on 1 November 2022 to over $3.3 trillion by 31 December 2024.

The second driving force, data, has been described as 'the new oil' by Kai-Fu Lee, former President of Google China [Lee, 2018]. At first, it was thought that the information revolution would mean the mass digitization and collection of human knowledge from books and other text-based sources scattered around the world (e.g. Google Books). But it is the internet that has been the real catalyst for big data. The average person has access to more information today than the most powerful entities did thirty years ago. In accessing this information and interacting online, we in turn generate vast amounts of data which are harvested by corporations and governments. Every time we click on a link, tag a photo, write a review, send a text, browse for a product, play a game, data are generated and collected. These data are then sold by brokers, a highly lucrative trade worth hundreds of billions per year globally [Dataintelo, 2023]. Without the records of Go matches played online, the training of AlphaGo would not have been possible in such a short time-frame. By interacting online, we are training the AI systems and making them 'smarter', smart enough to be useful to us. Then when we use these AI systems we are the ones being trained.

2.3 Potential AI Bottlenecks

The political-economic impact of AI will depend on the pace and scale of technological progress. There are still many obstacles to achieving AGI.

Transistors are already getting so small (3 nanometers, roughly the size of 30 atoms)

that they run into quantum tunneling problems [Sze et al., 2021]. Nevertheless, processing power will continue to increase as efficiency gains are found through developments such as accelerated computing and customized AI chips. Another source of fragility is that 92% of the world's advanced processors (less than 10 nanometers) are manufactured in Taiwan by one company, TSMC. The remainder are produced by Samsung. The equipment for printing these chips using extreme ultraviolet (EUV) lithography is made by only one company located in the Netherlands, ASML.

A further constraint is the exhaustion of data sources. The world wide web has been fully exploited to train AI programs. Moreover, the success of LLMs has propagated AI-generated content throughout the web. It turns out that AI trained on AI-generated content degrades rapidly [Guo et al., 2023]. Thus, there are some who believe that AI progress will turn out to be self-undermining. Ilya Sutskever, co-founder of OpenAI and its former chief scientist, suggests that the exhaustion of internet data means that a number of further technical advances are needed for AI to continue to scale [Vincent, 2024]. These include new pre-training methods, agentic systems, and enhanced reasoning capabilities. Nevertheless, there are still vast amounts of undigitized archival data and privately owned databases with high-quality human-generated data. An example is the data Tesla has gathered from its self-driving vehicles. These data sources will become more and more valuable. New data sources will also emerge as more of our appliances are fitted with sensors and cameras and connected the internet—the so-called Internet of Things (IoT). Moreover, the humanoid robots being developed will eventually provide high-resolution data about the 3-D world.

In addition to chip and data constraints, the massive computational power being amassed in new data centers requires immense amounts of energy, especially for cooling. Supercomputers performing a billion quadrillion calculations per second generate a lot of heat. Hence, the cluster of data centers located in Iceland and Microsoft's underwater data center off Scotland's Orkney islands. AI is already producing a rush to expand energy generation, including nuclear energy, with AI

models trained in data centers located away from population centers and closer to energy sources.

Finally, it may not be that AI progress stalls, but that humans lift their game. Recall that after Lee Sedol was defeated by AlphaGo in 2016, humans began to study the moves of the AI systems. What they learned revolutionized how Go is played. It turns out that while learning from the AI systems, humans discovered some weaknesses. In 2023, an Go player beat a top AI program by exploiting these weaknesses [Waters, 2023]. If human performance improves in lockstep with AI performance, AGI will never be reached. I will present reasons, however, why the exact opposite may end up occurring.

All of these factors could slow or entirely halt the progress of AI. However, I expect these and other obstacles will be overcome and AI will continue to scale [see Sevilla, 2024]. This expectation is based on the extent of government and corporate buy-in to the AI revolution. Aschenbrenner [2024] expects this 'unhobbling' process to take place over the next few years. The coordinated private-public push toward AI will come into focus in Section 3.1. Even if AGI is not reached, some of the destructive economic, political, and social forces I describe in this essay could be unleashed merely by widespread adoption of existing or imminent AI systems. For example, it is likely that a large fraction of jobs can already be automated. Regardless of whether AI progress will run into bottlenecks, it has become clear that AI development will not be halted deliberately, despite the risks.

2.4 The Existential Risk of AI

The introduction of a new inorganic form of intelligence to the planet poses an existential risk to humanity. In *Superintelligence*, Philosopher Nick Bostrom describes an existential risk as one "that threatens to cause extinction to earth-originating intelligent life or otherwise permanently and drastically destroy its potential for future desirable development" [Bostrom, 2014]. It is widely held that ASI poses an

existential risk. AI safety research has focused on the risks posed by superintelligent systems that are beyond human control and whose objectives are misaligned with human values—the so-called *alignment problem*. The outcome people typically worry about is that a superintelligent AI system will decide, for one reason or another, that humanity should be exterminated and find a way to achieve human extinction. The existential risk of losing control of AI is not the subject of this essay. Yet, it is a genuine concern and looms so large in the current imagination that it warrants some attention before proceeding to the political-economic problem.

The fears about AI safety have been shaped in part by science fiction, which has seared images of malicious superintelligence into our imaginations. In films such as *The Terminator* and *The Matrix*, humans are either enslaved or exterminated by the AI. There are also equally compelling utopian visions of the future, including *2001: A Space Odyssey* and *Star Trek*. Many tech leaders subscribe to the techno-utopian view. Sam Altman, CEO of OpenAI, thinks that "AGI will be the most powerful technology humanity has yet invented... It's the world that sci-fi has promised us for a long time" [Shah, 2023]. Nevertheless, it is the fear of a dystopian future that dominates the present moment. 42% of CEOs at the Yale CEO Summit in June 2023 believed that AI has the potential to destroy humanity within 5-10 years [Egan, 2023]. Stephen Hawking, Bill Gates, and Elon Musk have all warned that AI could lead to human extinction. Musk has described the development of AI as "summoning the demon". There are of course post-humanists who are unconcerned by human extinction and view human intelligence as a mere blip in the evolution of the universe.

So as not to rehash a large and well-known body of thought, let me briefly elaborate on the specific risks. Though I will focus here on AI and robots, existential risks will also arise from other technological advances on the horizon. In fact, the coming revolutions in synthetic biology and nanotechnology could be even more dangerous and AI may be needed to direct and contain the use of these technologies. Thus, much will depend on AI safety and alignment.

Concerns about AI safety begin with autonomy. While nations, corporations, and religious organizations have their own objectives and are, in a sense, intelligent non-human organisms, there are always humans in the loop. Autonomous AI agents are different. Cutting out human intervention may boost productivity by enabling rapid planning and execution of tasks. But then it is critical that these AI agents be aligned with human interests. The problem is that AI systems may become misaligned despite human attempts at control and without any 'intent' on the part of the AI. Suppose an AI program is given a goal. To reach this goal, it is allowed to create subgoals and autonomously work to achieve these subgoals. It may soon discover that money and power are general currencies for achieving any goal and start to work on strategies for accumulating them, without moral constraints. This is called 'instrumental convergence'.

Problems compound when there is recursive self-improvement. A feedback loop would be created in which the AI becomes smarter by improving itself, which allows it to more efficiently self-improve, becoming even smarter, and so forth. The result would be an intelligence explosion, as hypothesized by I.J. Good, John von Neumann, Ray Kurzweil, and others, thus making the control problem exponentially harder [Good, 1966, Vinge, 1993, Kurzweil, 2005]. There are huge financial incentives for AI companies to establish the loop of recursive self-improvement. That is why it is believed that once human-level AGI is achieved, superintelligence (ASI) will not be far off. It will be difficult to contain the risks created by such an intelligence explosion.

The existential risk from losing control of AI has already attracted considerable attention. Legal and regulatory initiatives include the EU's AI Act and the UK's global AI Safety Summit hosted in November 2023. Technological solutions have also been proposed. One approach is to use advanced AI systems that are firmly under human direction to control other potentially dangerous AI systems. Nick Bostrom has warned that this would not work if a 'singleton' were to arise, that is, an AI system that establishes an unassailable technological lead over all others [Bostrom,

2014]. Another proposal is to employ AI to augment human cognition and enable us to compete with superintelligent AI. That is the motivation behind Neuralink brain chip implants which create a brain-computer interface. In *Homo Deus*, Yuval Harari does not envision human domination by AI, but rather sees humans seeking god-like powers and immortality by merging with technology [Harari, 2015]. Obviously, these solutions to the AI control problem introduce their own existential risks.

2.5 Why not Slow down AI Progress?

In March 2023, Elon Musk, Steve Wozniak, Yuval Noah Harari, and others signed an open letter calling on AI labs to pause training of AI systems more powerful than GPT-4 for at least 6 months [Future of Life Institute, 2023]. That did not happen. Others have made more modest proposals for containment: "containment is about meaningful control, the capability to stop a use case, change a research direction, or deny access to harmful actors" [Suleyman, 2023]. It is not clear that this will happen either. The *precautionary principle* suggests we should exercise caution when making irreversible decisions that can cause harm. Given the potentially cataclysmic consequences of AGI, why would policymakers not pause or contain AI development? There are two reasons.

The first reason is that the West is locked in a geopolitical struggle for power, especially with China. As in the twentieth century, and indeed throughout history, the winner will be the bloc with the most advanced technology. For this reason, the *Roadmap for Artificial Intelligence Policy* issued by the AI Working Group of the US Senate in May 2024 focused more on accelerating AI development than containing it [Bipartisan Senate AI Working Group, 2024]. In addition, on 20 January 2025 (inauguration day), Donald Trump revoked the Biden administration's Executive Order 14110 titled "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence". The competition with China is genuine. Figure 1 shows that by 2023 the US had by far the largest number of notable machine learning models (61), followed by China (15), and then the rest of West. China is the leader in AI

patents and industrial robots [Stanford Institute for Human-Centered Artificial Intelligence, 2024]. In addition, private investment in AI between 2013 and 2023 was $335 billion in the United States and $104 billion in China [Stanford Institute for Human-Centered Artificial Intelligence, 2024]. The combined private investment in the next three countries, the United Kingdom ($22 billion), Israel ($13 billion), and Canada ($11 billion), was less than half that in China. Hence, the US is in the lead, with China coming second, and the rest of the West far behind.

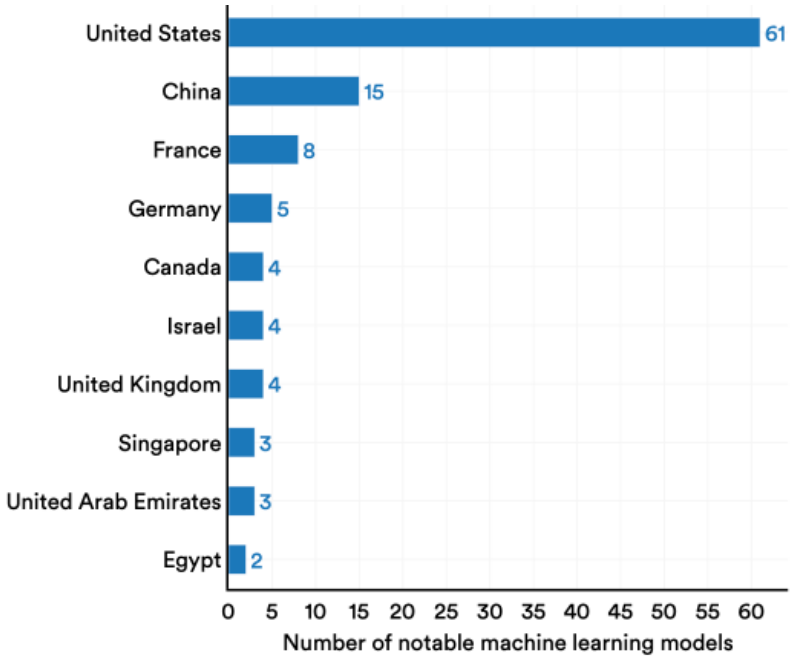

Figure 1: Number of notable machine learning models by country, 2023.
Source: Stanford Institute for Human-Centered Artificial Intelligence [2024].

The US lead is not unassailable. China's advantage in AI patents, AI research publications, and industrial robots could be converted into additional AI capabilities. Chinese companies are also rapidly catching up in the LLM space. On 26 December 2024, the Chinese AI company DeepSeek released an open-source LLM which exceeds the performance of the US-produced open-source LLMs (e.g. Meta's Llama3.1-405B)

and matches the leading proprietary models by OpenAI and Anthropic. What is remarkable is that DeepSeek-V3 was supposedly trained on a budget of $5.58 million over 55 days, probably less than one tenth of what was spent by the US companies. Moreover, unless DeepSeek found a way to circumvent export restrictions under the US CHIPS Act, it was trained using less advanced Nvidia chips. In January 2025, DeepSeek-R1 was released displaying extraordinary reasoning capabilities, as well as meta-cognition (i.e. the ability to reason about its own reasoning). This was achieved through advances in applying reinforcement learning with verifiable rewards (RLVR). Thus, the AI revolution will split the world into two technological empires, a coalition of democratic nations led by the US and one led by China.

The second reason why AI development will not be contained is perhaps less obvious. The West faces a crisis and technology will, for better or worse, be seen as the way out. Our current age began with the industrial revolution in Britain, Europe, and the United States. Now, the countries that launched the revolution are stuck with unsustainable institutions and political turmoil. As such, they are desperately in need of a further round of technological progress. To understand the roots of this institutional crisis and why it means that AI progress will not be slowed, we need to revisit the industrial revolution in Britain and its political-economic legacy.

2.6 The Industrial Revolution and a New Social Contract

In 1760 England, fabric was woven on handlooms. This small-scale production slotted harmoniously into rural life but was highly inefficient, making clothing expensive. Soon the process would be automated and the handloom weavers replaced by machines. Fabric would be woven by new industrial looms driven by steam engines. Production would not take place in isolated rural cottages, but would instead be consolidated into factories popping up in the rapidly growing industrial cities. Thus, the process of weaving fabric, as with many other physical tasks, was automated using a combination of energy, machinery, and organization.

Unlike previous bursts of innovation in the Middle Ages and before, the technological advances continued to occur, building one upon another. A second wave of technological progress stretched from the early nineteenth to the early twentieth century: the telegraph, electricity, railroads, new alloys (e.g. cheap steel), synthetic chemicals (e.g. household detergents), the internal combustion engine, and the production line. These technologies transformed the economy and the nature of work. More efficient production meant cheaper consumer goods. The great achievement of the industrial age was to vastly expand the amount and variety of goods and services that people could consume. In doing so, it created a new consumer class and culture: marketing and advertising, keeping-up-with-the-Joneses, the suburbs, shopping malls, and much else we associate with modern life [Offer, 2006]. According to the historian Karl Polanyi, what arose was not only a market economy, but a 'market society'.

Most of human existence had been lived at or near subsistence. The Industrial Revolution changed that. While creating unprecedented wealth, it was not a narrow phenomenon of rising real incomes. Literacy, heights, caloric intake, life expectancy, poverty alleviation, social tolerance, and political freedoms all improved in lockstep with GDP [Fogel, 2010, Weil, 2014]. As Acemoglu and Johnson [2023] describe, however, the benefits of industrialization arrived slowly and only after massive social upheaval. In *The Great Transformation*, Polanyi writes about how the social fabric was stretched and torn by the displacement of rural workers and their organization into urban slums:

> "Before the process had advanced very far, the laboring people had been crowded together in new places of desolation, the so-called industrial towns of England; the country folk had been dehumanized into slum dwellers; the family was on the road to perdition; and large parts of the country were rapidly disappearing under the slack and scrap heaps vomited forth from the "satanic mills." Writers of all views and parties, conservatives and liberals, capitalists and socialists, invariably referred to social conditions under the Industrial Revolution as a veritable abyss

of human degradation." [Polanyi, 2001/1944, p. 41]

In the background of this "human degradation" lay the end of feudalism, brought about by the ascent of a new merchant class at the expense of traditional hereditary elites. This transition was far from painless. The safety nets of the feudal order vanished as old notions of *noblesse oblige* and communal insurance came to an end. *The Communist Manifesto* laments this severing of feudal obligations: it left "no other nexus between man and man than naked self-interest", leading to "naked, shameless, direct, brutal exploitation", and exposing workers to "all the vicissitudes of competition, to all the fluctuations of the market" [Marx and Engels, 1988/1848]. Marx and Engels were particularly moved by the pitiful state of the handloom weavers in England, whose trade was the first to be automated.[7] Thus, Marxism emerged out of the social upheaval of the industrial revolution. In turn, it spawned political revolutions in Russia and China and shaped much of the twentieth century.

Frightened by the threat of socialist revolution, western nations forged a new social democratic contract with their citizens [Acemoglu and Robinson, 2000]. Thus, the familiar institutions of the modern nation were born, including the welfare state, mass public education, public healthcare, and government regulation of markets. Even in the more laissez-faire United States, the New Deal of Franklin D. Roosevelt and the Great Society of Lyndon B. Johnson created institutions for poverty alleviation and market regulation. Similar institutions popped up in every capitalist democracy, containing the threat of political revolution for some time.

2.7 Institutional Exhaustion

Most of these legacy institutions are now dysfunctional and/or unsustainable. Public pensions systems are a prime example. Old age poverty in the US declined from 30 percent to 13 percent between 1967 and 2000. Engelhardt and Gruber [2004] have

[7] Managing his father's cotton mill in Manchester, Engels had seen the social disruption of industrialization close up and already collected his experiences in *The Condition of the Working Class in England* [Engels, 1845].

shown that this improvement can be entirely explained by the rise in Social Security benefits over the period. Thus, the public pensions system did what it was designed to do, reduce old age of poverty. Unfortunately, however, it is not sustainable and is largely a transfer from later generations to earlier ones. At the time of writing, US federal government debt is around $36 trillion, or over 120% of GDP. The US Treasury Department estimates that unfunded spending obligations will amount to $79.5 trillion over the next 75 years (in present value terms). Social Security and Medicare shortfalls make up 95% of the total unfunded obligation [US Department of the Treasury, 2024].

There are four reasons why the welfare state cannot be sustained in its present form. First, all advanced economies have undergone a slowdown in productivity growth since the mid-2000s [Fernald et al., 2025]. For example, US labor productivity growth rates have declined by 62% since the 1995-2005 period [Goldin et al., 2024]. So far, the digital economy has not significantly boosted productivity, as Solow famously pointed out [Solow, 1987]. The second reason is population decline. Every country in Europe, the US, Canada, Australia, and New Zealand has below-replacement fertility [Schumacher et al., 2024]. The situation gets worse if we include South Korea, Japan, and Singapore. South Korea's total fertility rate is currently 0.72 meaning that, at this rate, every 100 South Koreans alive today will have only 36 children, 13 grandchildren, and 5 great grandchildren. The obvious consequence is that the working-age population would become too small to provide for the pensions and healthcare of the elderly. That brings us to the third reason, widespread political opposition to mass immigration [Alesina and Tabellini, 2024]. Even high-skilled immigration is controversial, as demonstrated by the reaction to Elon Musk's and Vivek Ramaswamy's proposal to expand the US H-1B program [Southern, 2024]. Finally, it is harder to tax corporations and high income earners due to increased global mobility. For example, the steep rise in top income tax rates under Francois Hollande in 2013 caused an exodus of top earners from France and raised minimal revenue. The policy was abandoned in 2014 [Chrisafis, 2014].

Thus, the economically advanced nations face a predicament. The welfare state is currently unsustainable, yet it cannot be dismantled. Think about the political consequences of ending Social Security in the United States. The only way out of crisis then is to find a new engine of economic growth. Eric Weinstein has described our institutions as being packed with embedded growth obligations (EGOs) [Weinstein, 2021]. Without growth, the complex of institutions developed in the wake of the industrial revolution could collapse and take the political system down with it. The much-needed economic growth is itself not sustainable without a new round of technological progress. Thus, a technological lifeline is desperately needed.

2.8 AI as Technological Lifeline

The economic, social, and psychological consequences of institutional crisis are already being felt. The productivity slowdown has been accompanied by rising inequality and a hollowing out of the middle class, with the share of US income going to workers falling to its lowest level since the Great Depression [Bergholt et al., 2022]. In addition, social indicators no longer move in lockstep with economic growth. Life expectancy in the US and UK has started falling [Case and Deaton, 2020, Lima and Snower, 2020]. One reason is the mental health epidemic in these countries marked by rising rates of depression, anxiety, loneliness.[8] Addiction and suicide are responsible for the 'deaths of despair' documented by Case and Deaton [2020], Carvalho [2022]. Mental health problems are rising fastest among the young. Among US college students, 17% have been prescribed psychiatric medication [Marconi et al., 2023]. It used to be that happiness was lowest among the middle aged (the so-called midlife crisis). Recent evidence from 34 countries, including the US and UK, reveals a new trend since 2011 in which young adults have the lowest happiness and well-being [Blanchflower et al., 2024]. Melissa Kearney has shown that a rise in out-of-wedlock births and single-parent households has reduced social mobility and raised inequal-

[8] It is estimated that 20% of US adults have a mental illness. See the mental health statistics provided by the National Institute of Mental Health [2023].

ity [Kearney, 2023]. In addition, legacy institutions, such as mainstream media and various expert bodies, are no longer appealing or credible to large parts of the population [Carvalho et al., 2024]. Congressional approval dropped to 9% in November 2013, making the legislative body the country's least favored institution—a dangerous turn foreshadowing political turmoil in the US [Gallup, 2019]. Deep social and political divisions are only hardening. Optimism about democracy has turned to rising support for authoritarianism and democratic backsliding around the world [Waldner and Lust, 2018].

All of this has occurred even before the widespread adoption of AI, and certainly before the arrival of AGI. One way of viewing the predicament is that the industrial revolution was a point-of-no-return in human history. Once intensive economic growth was achieved, it would be difficult to rest again. As pointed out by Weinstein [2021], the stabilizing institutions we have developed are at risk of collapse without non-stop growth. Hence, even if AI development could be contained, policymakers will not do so. AI is viewed by both the public and private sectors as our route out of crisis. There is no guarantee, however, that it will deliver. If the political-economic risks of AI are not properly managed, there will be no growth take-off. In fact, the experiment could end in immiseration and even extinction of most of the population.

3 The Political-Economic Risks of AI

In this section, I examine the economic, political, and social upheaval that could be unleashed by the AI revolution. While optimistic scenarios exist in which AI boosts prosperity and freedom [Tegmark, 2017, Bostrom, 2024], I will concentrate on a worst-case scenario in which AI 'permanently and drastically destroys' the potential for human development. Thus, malicious superintelligence is not the only form of existential risk that could be produced by AI [Bostrom, 2014]. My focus will be on the impact of the AI revolution in the United States, the leader in artificial intelligence. I will also cover some of the likely consequences for the broader 'West'—the coalition

of democratic nations including the UK, the EU members, Canada, Australia, New Zealand, Japan, Singapore, and South Korea—as well as less developed nations. All of these nations could face additional geopolitical challenges which I will describe later. How events will unfold in China is an important and fascinating subject, but one that I leave to China specialists [see for e.g. Lee, 2018].

The political-economic trajectory I trace in this section is not a conspiracy, but rather an evolutionary path to a dystopian future. The massive accumulation of economic and political power along this path will be less by grand design than through myopic step-by-step adaptation to the AI economy. The road to Elysium may even be paved with good intentions. But the consequences will be even worse than those envisioned by F.A. Hayek [1944] in his *Road to Serfdom*. That is why it is important to look ahead. Without proper intervention, here is how events might unfold.

3.1 Economic Concentration and Political Power

If advanced AI systems continue to require large capital expenditures and proprietary knowledge, the first development would be extreme concentration of the global economy, with profits shifted to companies based mainly in the United States (esp., Silicon Valley) and China. In that case, new economic and political elites will emerge from the AI revolution with novel sources of power, both domestically and globally. This extraordinary concentration of power will be made possible by the distinctive features of AI listed in Section 1.4, especially its general-purpose nature, global scalability, and importance to national security.

On the other hand, if AGI and ASI turn out to be inexpensive to develop and run, the market power of the AI sector would shrink significantly and the risks would be attenuated, but not eliminated. The likelihood of this alternative scenario has increased markedly after the release of the open-source LLM DeepSeek-R1 on 20 January 2025. I first explore the "AI moat" scenario in which the leading AI developers convert market power into political leverage, before exploring the alternative

"no AI moat" scenario in Section 3.1.4.

3.1.1 The AI Complex

A crucial part of the picture is the relationship between the AI sector and government. Until recently, AI developers have worked at arms-length from government and competition among then has been allowed to flourish within the existing institutional framework. That is changing. The leading AI companies are likely to become entangled with critical arms of the government, forming a parallel economic sphere. I call this private-public partnership 'the AI complex'.

Chatbots such as ChatGPT are only the beginning of the AI revolution. As proof-of-concept, they serve as marketing tools for attracting the capital and talent needed to develop stronger forms of AI and AI agents that can autonomously perform complex tasks. As noted in Section 2.3, further AI progress could be impeded by legal and regulatory constraints including:

- Land use restrictions for building data centers.
- Energy regulation limiting power to data centers.
- Intellectual property and data protection laws limiting new sources of data.
- Measures for safeguarding the privacy and mental health of users.

Due to the pressures of geopolitical competition and the desperate need for growth, policymakers will have a strong incentive to switch from prioritizing AI safety to removing these bottlenecks. A parallel economic sphere will be created to clear legal and regulatory barriers to rapid AI development. In particular, special exemptions from land use restrictions, energy regulation, and intellectual property and data protection laws will be granted. This is already happening. Consider, for example, Donald Trump's immediate cancellation of the Biden administration's AI safety Executive Order. Otherwise, the leading AI companies in the West will either start to lag behind China or look to build data centers rapidly in non-democratic countries.

Both outcomes pose obvious geopolitical risks.

The next step is for the government to step into this parallel AI sector and become more involved in its operations. Interventions will come in three forms: financing AI development, restricting competition, and providing state-level security.

The US government is already a financial backer of the AI sector. It could be that AGI and ASI cannot be achieved without massive capital expenditure and proprietary knowledge. For example, xAI's Colossus supercomputer cost three to four billion dollars in GPUs alone, with plans for a tenfold expansion [Moris and Kinder, 2024]. In addition, progress toward AGI may require further technical breakthroughs [Vincent, 2024]. Accordingly, the Department of Defense doubled funding for AI-specific research and development from the 2020 to 2024 fiscal years [U.S. Office of the Undersecretary of Defense, 2023]. It is likely that the government will continue to increase funding for the AI sector to maintain the US lead in AI and stop companies looking abroad for funding, especially to foreign governments. For example, in early 2024, OpenAI's CEO Sam Altman was seeking seven trillion dollars in funding for a new AI chip from a number of sources including the government of the UAE [Hagey, 2024]. Foreign funding could dictate the location of data centers, which will put valuable data, hardware, and weights from the AI models at risk [Aschenbrenner, 2024].

In financing the AI sector, the government may also begin to shape the composition of the industry, picking winners and restricting competition to better control the pace and direction of innovation. It would do so for several reasons. First, competition in the AI sector will erode profits and reduce incentives for investment [Aghion et al., 2005, Gilbert, 2006, Vives, 2008, Spulber, 2013]. Second, competition could push AI companies to take greater risks and deprioritize safety. Third, to compete with China, the government will want to maximize the capability of the leading AI models which are contenders to reach AGI first. That is, the US government will want to pursue a 'maxmax strategy'. In contrast, if there are diminishing returns to AI development

and widespread applications for less advanced AI systems, financial markets will tend to fund emerging companies that enjoy low-hanging fruit (a 'maxmin strategy'). These companies would not advance the technological frontier as geopolitical goals would dictate. This process of picking winners is already evident in the $500 billion Stargate joint venture to build AI infrastructure in the United States, led by OpenAI, Oracle, SoftBank, and Microsoft, and announced at the White House on 21 January 2025.

Finally, as set out by Leopold Aschenbrenner, former member of OpenAI's super-alignment team, the US government will need to provide state-level security to the leading AI companies [Aschenbrenner, 2024, IIIb.]. If human-level intelligence (AGI) or superintelligence (ASI) is achieved, there will be huge incentives for powerful, nefarious actors to steal the AI models, which is fairly easy to do given that a model is simply a matrix of numbers. No physical invasion would be necessary. While stealing current models would not give a competitor AGI, Aschenbrenner points out that the algorithmic methods for training AGI (and possibly ASI) are already being developed and stealing this knowledge could unlock a competitor's path to AGI. The Manhattan Project which generated similarly sensitive technology would have been futile without extreme security measures, including its remote base in Los Alamos, New Mexico. The difference this time is that the merging of the AI sector with the state could be far more wide-reaching than in the past, encompassing much of the economy.

The US government will not only want to secure AI models for an edge in economic competition, but also because the AI models will generate new weapons of mass destruction (WMD), especially bioweapons and nanotechnology that would be extremely dangerous (possibly humanity-ending) in the wrong hands. Again, we do not have private companies developing nuclear weapons, and we would not want the next generation of WMDs outside of state control. Thus, the parallel AI sector could merge with the existing public-private sphere of intelligence agencies, military, and defense contractors. In these ways, the boundary between the state and the AI in-

dustry will disappear and what will emerge is a new and hugely critical private-public partnership.

3.1.2 The AI Black Hole

After a loosely regulated AI sector has been carved out and merged with critical elements of the security state, there is the danger that the AI complex will expand far further than it needs to, turning into an economic black hole that swallows up a large part of the global economy. As we shall see, this over-concentration of the global economy would have devastating political consequences. It is, however, avoidable.

There are two ways in which an AI black hole could emerge. First, AI companies could vertically integrate buying up large swathes of the economy, including land, energy supply, chip manufacturers, and firms that produce business applications of AI. They will have an advantage in acquisitions due to a lower cost of capital, a consequence of government backing. In addition, it will be more profitable for many firms upstream and downstream of the AI companies to produce within the AI complex than outside of it, due to lighter regulation and state-provided security. There is already a significant amount of vertical integration in the AI sector. Amazon, Google, Microsoft, and Meta own massive amounts of data through cloud storage and social media platforms such as Facebook, Instagram, and YouTube. Microsoft and Google have leading business applications. Amazon has built a global logistics network and recently purchased a $650 million nuclear-powered data center in Pennsylvania. Microsoft has signed a purchase agreement with a company aiming to deliver fusion energy (which, of course, does not yet exist) by 2028. Again, OpenAI is looking to develop its own AI chip. In doing so, these companies are further expanding their digital ecosystems.

The second driving force behind economic concentration is the hold-up problem [Williamson, 1975, Klein et al., 1978, Grossman and Hart, 1986], that is, the use of market power by AI companies to extract the profits of downstream and upstream

firms. For example, once a firm has reduced its workforce and reconfigured its operations to make intensive use of AI, the leading AI companies could raise their prices and capture much of the economic surplus from the relationship. If the leading AI companies act in concert, the only other option for a purchasing firm would be to use inferior AI models that place it at a disadvantage vis-à-vis its competitors. Such collusion may not be difficult to achieve if the number of leading firms is limited by the government and if they are exempted from certain antitrust provisions. The industries whose profits are at immediate risk of being swallowed up by the AI sector are those that concentrate on repetitive cognitive tasks, including (but not limited to) accounting, clerical work, customer support, sales and marketing, and consulting. Advances in AI-powered robots would significantly expand the set of industries whose profits could be extracted by the AI sector to those industries that make intensive use of physical labor.

In this scenario, there will still be a regular economy bound by the laws and regulations developed over the twentieth century to strengthen competition, safeguard workers, and protect the environment. However, this old economy will make up a smaller and smaller fraction of the new economy, as the AI complex grows and becomes dominant. Thus, economic activity could come to be concentrated in a new economic sphere outside of the existing institutional framework. This economic transformation could happen without much political attention or opposition, through step-by-step adaptation to technological developments in AI.

3.1.3 The Shifting Balance of Power

The concentration of the global economy would have profound political consequences, domestically and internationally. The classical economists understood very well that wealth translates into political power, and *vice versa*. In Book V of *The Wealth of Nations*, Adam Smith writes:

> "It is in the age of shepherds, in the second period of society, that the

> inequality of fortune first begins to take place, and introduces among men a degree of authority and subordination which could not possibly exist before. It thereby introduces some degree of that civil government which is indispensably necessary for its own preservation: and it seems to do this naturally, and even independent of the consideration of that necessity. The consideration of that necessity comes no doubt afterwards to contribute very much to maintain and secure that authority and subordination." [Smith, 2008/1776, p. 689]

This passage introduces Smith's theory of the evolution of property rights. When technological change (in this case, pastoralism) creates economic inequality, political and social institutions emerge that protect property rights and entrench the inequality. The AI revolution will be no exception. Let us first consider the immediate and milder forms of political power that will emerge, before getting to more fundamental institutional change.

The key players in the AI sector could use their newly amassed capital for lobbying, political patronage, and persuasion activities. Due to the AI revolution, the political system could diverge further from the ideal of 'one person, one vote' with outcomes influenced by money channeled through political donations, patron-client relationships, and the media.[9] In addition, human displacement from the workforce will lead to a sharp fall in income tax revenue at precisely the time at which demand for welfare payments spikes. To finance its core functions, the government will need to 'strike a deal' with the AI sector to shift the taxation system toward corporate and capital taxation. As part of this deal, the AI sector will acquire significant power over what gets financed. Finally, through investments in persuasion, citizens may come to accept, as Smith writes, 'a degree of authority and subordination which could not possibly exist before' [see Skaperdas and Vaidya, 2012]. This is the thrust of Thomas Piketty's *Capital and Ideology* which examines the role of ideology in pre-

[9]Evidence indicates that wealthy Americans already have far more influence over government policy than the middle class or poor [e.g. Ansolabehere et al., 2003, Bartels, 2008, Gilens, 2012].

serving the *status quo* distribution of wealth and power [Piketty, 2020]. Finally, as AI becomes a critical factor of production, disfavored individuals and organizations could be coerced by denying them access to the leading AI models. Non-compliers could find it hard to identify and prove that the AI agents they are using have been handicapped or misdirected. Thus, the political influence of the AI complex and those connected to it could grow and extend beyond the initial exemptions granted to AI companies during their breakout phase. In turn, this political power could be used to further concentrate profits and expand the AI black hole. The extreme concentration of individual and corporate power would grow even more dangerous if co-opted by an authoritarian state.

AI will not only shift the balance of power within nations, but also between nations. So far, the analysis applies to the political economy of the United States, the leader in AI. There are two AI powers emerging, the USA (concentrated in Silicon Valley) and China [Lee, 2018]. Though other centers of AI development exist, notably London, Tel Aviv, Paris, and Toronto, it is clear that the US will lead 'the West'—the coalition of democratic nations including the UK, the EU members, Canada, Australia, New Zealand, Japan, Singapore, and South Korea. If AGI requires massive capital expenditure, these countries may be tempted to free ride on AI innovation by US companies, just as they (partially) rely on the US for defense, information technology, and medical innovation. When it comes to AI, this would tip the balance of power within the democratic coalition very far toward the US.

Domination of the supply of artificial cognitive (and possibly physical) labor, will give the US considerable leverage over the other advanced economies. The US-based AI complex will be able to launch labor strikes in other countries and otherwise hinder the competitiveness of foreign firms. AI systems in the US will also produce new WMDs and cyberattack capabilities, as well as new defensive measures against these threats. Hence, without their own investments in AI, the rest of the western countries will rely to an unprecedented degree on the US for their security and defense. Developing countries will have even less ability to build their own advanced

AI systems and humanoid robots. As such, they may find themselves with very little international bargaining power, outside of control over critical resources such as rare minerals.

Finally, the AI revolution will place extraordinary stress on public finances outside of the US. As in the United States, part of the burden from lost income taxation will have to fall on corporate taxation. The difference is that taxation of US-based companies by foreign governments will be constrained by various tax avoidance strategies. For example, an AI company could locate regional headquarters in an (intellectual property) tax haven and use it to charge local subsidiaries for using its intellectual property [see Tørsløv et al., 2023]. That would significantly reduce the profits recorded in high tax jurisdictions and limit the ability of foreign governments to raise revenue through corporate taxation. As AI companies gain global influence, it will be difficult to close tax loopholes. Indeed, tax competition between nations could produce a race to the bottom. Consumption taxes are another option but are regressive and will raise limited revenue if mass unemployment leaves the vast majority of the population with little income to spend.

Thus, AI could undermine public finances around the world and destabilize nation states, further concentrating global wealth and power in the US-based AI complex.

3.1.4 Economic Concentration under Alternative Scenarios

Now consider the alternative "no AI moat" scenario in which AGI and ASI are inexpensive to develop and run anywhere in the world. While far from certain, this scenario has become more likely since the release of DeepSeek-R1 in January 2025. Economist Olivier Blanchard has described it as: "Probably the largest positive tfp [total factor productivity] shock in the history of the world" [Blanchard, 2025]. What DeepSeek revealed was that advanced AI capabilities can be achieved at lower cost, with more modest levels of compute and data center infrastructure, and lower levels of energy consumption than previously thought possible. Hence, massive capital

expenditure and proprietary knowledge are not needed to match the *current* performance of the leading US models. That means the market power and profitability of companies developing the leading foundation models (e.g. ChatGPT) could be drastically reduced, as could the risk of an AI black hole swallowing up a large part of the global economy. As small organizations, and even individuals, are able to run AGI on their own computers, the leading AI companies would not have coercive power either domestically or internationally through 'denial of service' threats. Every country could have their own sovereign AI systems at low cost, easing concerns about the international balance of power. AI would be 'democratized'.

It is still far from certain, however, that the AI sector will lose its market power. First, reaching AGI and ASI may still prove impossible without massive amounts of compute and energy. Even if the training compute required is low, advanced applications may require a lot of inference compute, that is, resources for making predictions or decisions based on novel data after the model has been trained. Second, it could be the combination of reinforcement learning techniques applied by DeepSeek with massive amounts of new data and compute that unlocks a path to AGI and ASI. The demonstrated competitive threat from China means the race is on. We should expect faster relaxation of regulations, stronger government backing, and possibly a move away from open-source AI. The firms that get to AGI and ASI first will enjoy massive market power and political leverage, as I have set out. Finally, the government may restrict the development and use of AI to contain threats to natural security, including rogue states and non-state actors developing advanced military technologies and WMDs. Thus, a competitive moat could be created for the leading AI companies by fiat.

Even if the leading AI companies turn out to have no moat, the threat of global economic concentration due to AI remains, though probably in attenuated form. The reason is that the profits may lie in the application layer (e.g. AI marketing and sales), not the model layer. During the Dotcom boom in the late 1990s, companies supplying internet infrastructure such as Cisco Systems were expected to be mas-

sively profitable. Instead, the companies that capitalized on the internet turned out to be the platforms such as Google, Facebook, and Amazon, which benefited from network externalities and other lock-in effects. Likewise, massive platforms delivering AI applications could be built on top of foundation AI models. As companies developing foundation models have difficulty monetizing their services due to easy entry, the large players in the application layer could end up purchasing the leading AI models and financing their capital expenditure. (See for example Microsoft's existing relationship with OpenAI.) This is unlikely to lead to an AI black hole as would occur with concentration in the model layer and monopolization of AI. But, as in the existing digital economy, it could still confer considerable political power. In fact, the potential for market concentration could be larger in AI applications than in the present digital economy. Once a company gains a slight competitive edge, it will have few labor force constraints on rapid scaling. AI agents would make it possible to rapidly scale at low cost. The government could entrench this competitive advantage by regulating AI agents, which will serve as intermediaries between people and the internet, in the interests of national security. Thus, companies could still accumulate large market shares and the accompanying political leverage.

3.2 Breaking Democratic and Egalitarian Institutions

As the AI revolution advances, an even more fundamental political transformation could take place which is independent of the degree of market concentration in the AI sector. We have seen how the domestic and international balance of power can be altered within the existing institutional framework. This accumulation of power is constrained by democratic and egalitarian institutions within countries [Acemoglu and Robinson, 2012], as well as the complex of international institutions developed after the second world war known as the liberal international order [Ikenberry, 2001]. We will now see how the AI revolution could break the dam walls, removing con-

straints on the concentration of political power.[10]

Political institutions are "the rules of the game in a society or, more formally, the humanly devised constraints that shape human interaction" [North, 1990, p. 3], including procedures for appointing political leaders (e.g. elections). During a technological revolution, the political problem is not limited to the technocratic task of achieving current social objectives within the existing institutional framework. That is because our social objectives and formal institutions (i.e., rules of the game) may be reshaped in the course of the technological revolution. The deeper political problem that we face is how to stabilize the elements of the system needed to preserve our shared prosperity, freedoms, and social bonds. That means strengthening democratic and egalitarian institutions.

During the nineteenth and twentieth centuries, European elites were willing to transfer wealth and power to democratic states in order to prevent political unrest (e.g. riots, strikes) [Acemoglu and Robinson, 2000].[11] Political unrest was not only a threat to their corporal safety, but would also disrupt industrial production. The AI revolution could vastly reduce the threat of revolt and industrial disruption. First, AI could break the mutual dependency between financial capital, human capital, and labor that has existed throughout history. Because human capital and labor were indispensable factors of production, political and economic elites had a strong interest in expanding public education and healthcare and maintaining a cooperative relationship with workers. Soon, however, human capital and labor may be replaced by AI and humanoid robots. Second, the AI revolution could make citizens less willing and able to express political opinions and protest violations of democratic principles. It would do so by using AI systems to suppress political dissent and create nearly

[10] Acemoglu and Johnson [2023] examine two different (direct) effects of AI on institutions. First, AI can be used to monitor workers more intensively, further lowering wages and worker welfare and creating a new age of authoritarian capitalism. Second, AI can be used to undermine democracy through the state censorship and the spreading of misinformation by state and non-state actors.

[11] The role of elite competition [Lizzeri and Persico, 2004, Llavador and Oxoby, 2005] is covered in Section 4.2.

unlimited virtual worlds that distract and sedate the population. By lowering the threat of political unrest, AI would significantly reduce elite support for democratic and egalitarian institutions. As we shall see, the end state could be massive concentration of wealth among a small elite and the relegation of the rest of society to an underworld of scarcity and insecurity (at best).

Let us first deal with how the AI revolution could break the mutual dependency between economic classes. Among the distinctive features of the AI revolution (listed in Section 1.4), the primary driving force would be the automation of cognitive, not physical, tasks. Previous rounds of automation have always created new tasks and none have ended in mass unemployment [Brynjolfsson and McAfee, 2014, Autor, 2015, Acemoglu and Restrepo, 2019]. However, the displacement of humans from manual to cognitive tasks, where humans have a unique advantage, is different to the displacement of humans from cognitive tasks. To appreciate the distinction, let us reflect on evolutionary history.

Homo sapiens sapiens currently dominates this planet because we wiped out all other proto-human species (e.g. Neanderthals) and we have something that other animals do not. Humans are physically weak compared, for example, to a silverback gorilla that shares 98% of our DNA. What sets us apart is our unique ability to learn. The special cognitive capabilities of humans are set out by Joseph Henrich in his book *The Secret of Our Success* [Henrich, 2017]. In particular, we are the only animal species who can transmit substantial amounts of learned knowledge from one generation to the next, so that knowledge accumulates and builds on itself. We are also unique in developing external storage devices for this knowledge, such as libraries and the internet, making the knowledge partly disembodied. Knowledge is further distributed across multiple human brains, with these brains acting together, like a super-organism, to collectively record historical events and produce new knowledge. This networked intelligence has been called 'the collective brain' [Muthukrishna and Henrich, 2016, Henrich and Muthukrishna, 2024].[12] Cultural evolution of this kind

[12]See also Hayek [1945] on markets as information sharing and processing systems.

is far more rapid than genetic evolution and has enabled us to adapt and thrive in a vast range of environments, to the point that we can (like the sentient yogurt) consider colonizing other planets.

The bad news is that AI can already do a lot of this better than we can. AI systems have access to far more information than any single human, can learn much faster from this information, and can deploy more computational power in converting the knowledge it creates into action. Though humans still have an advantage in some complex and unstructured tasks, we are being rapidly surpassed in domains previously considered close to the peak of human cognitive capability. OpenAI's o1 model, released in September 2024, performs better than 89% of human coders in competitive coding tasks, places among the top 500 students in the US on a qualification test for the US Mathematical Olympiad, and exceeds PhD-level accuracy in physics, chemistry, and biology examinations [OpenAI, 2024].[13] Thus, humans are being superseded in the domains that set us apart and allowed us to dominate the planet. Previous rounds of automation were qualitatively different because they replaced heavy manual labor, something humans are not especially good at. Displaced workers could move up the cognitive ladder to other tasks. With AI, however, humans may face cognitive constraints on shifting to higher-valued tasks. There may be few, if any, new tasks that are within our grasp but beyond the reach of AI.

Once political and economic elites no longer depend on human capital and labor, the political-economic system that will emerge will be unlike any other. The industrial revolution gave us the ability to produce manufacturing goods cheaply and at scale. Over time, the growing scale of industrial production brought a new age of managerial capitalism. Power shifted to the professional class—the corporate executives, scientists, and administrators who managed capital. Human capital was ascendant. In *The New Industrial State*, John Kenneth Galbraith refers to the public-private network of human capital as the 'technostructure' [Galbraith, 1967]. It was this

[13]The o3 model released in December 2024 further improves on benchmarks, especially in tasks requiring chain-of-thought reasoning (CoT).

technostructure that drove the offshoring of manufacturing, mainly to China, and the assault on union power over the last fifty years. As a consequence, the labor share of income fell dramatically [Autor et al., 2013], but the result was not mass unemployment and immiseration.

The next phase of capitalism is the digital platform economy dominated by tech giants such as Google, Amazon, Microsoft, and Meta, which make markets by providing users (largely) free access to e-commerce and social media. According to Yanis Varoufakis, this phase of capitalism shares some elements of feudalism [Varoufakis, 2023], because powerful network effects limit competition [see Shapiro and Varian, 1999]. Consumers are tied to a few large platforms that capitalize on their unpaid (data-generating) labor, much as serfs were tied to the land and forced to work. However, Varoufakis's 'technofeudalism' is not such a large break from the past. It retains key features of managerial capitalism. Above all, human capital plays a critical role in the production process.

The AI revolution could take us to something new and far worse by breaking the symbiotic relationship between economic classes. The feudal economy was a system of labor coercion embedded in a web of social relationships and obligations. The feudal lords provided protection to peasants in return for their labor. Human labor was indispensable because the lords specialized in the exercise of violence and could not grow their own food or otherwise manage their estates. The AI revolution threatens such mutual dependency between classes. Consider Mustafa Suleyman's vision of the future of the startup firm:

> "I have proposed a test which involves them [AI systems] going off and taking a $100,000 investment and over the course of three months trying to set about creating a new product, researching the market, seeing what consumers might like, generating some new images, some blueprints of how to manufacture that product, contacting a manufacturer, getting it made, negotiating the price, dropshipping it, and then ultimately collect-

> ing the revenue. And I think that over a five-year period [by 2028] it is quite likely that that we will have an ACI—an artificial capable intelligence—that can do the majority of that task autonomously." [The Economist, 2023]

The main human element in this story is the owner of the $100,000 worth of capital. Thus, as we approach human-level AGI, the the holders of financial capital and political power will for the first time have little need for human capital and labor. Human obsolescence in the workforce severs the cross-cutting economic and social ties that shape our politics and knit our societies together. In addition, a nation's production, innovation, and military power have always been tied to the size, health, and skill of its human population. For the first time, this link could be broken by AI and humanoid robots. Thus, the elite will have no material interest in the health, productivity, or welfare of the population at large.

The situation is unlikely to be saved by a moral commitment to democratic and egalitarian institutions on the part of the elite. AI will weaken the shared interactions and forms of identification (e.g. national identity) that could forge such a moral commitment by allowing elites to socially separate themselves to an unprecedented degree.[14] Today, a wealthy individual or family will deploy their capital to produce goods and services for sale on the market. For example, a family may own a company operating a car factory. The profit from this business will pay for the goods and services required for the family's lifestyle. At a more advanced stage of technology, however, AI and AI-powered robots may allow the family to bypass the market by directly producing the buildings, food, clothing, education, and entertainment required for its lifestyle. For example, the car factory could be converted to produce robots that serve as household labor. That is, wealthy families could (at least partially) vertically integrate causing a collapse in economic complexity.[15]

[14]The US has already been socially disintegrating and segregating along economic class lines for some time [Putnam, 2000, Murray, 2012, Acemoglu and Wolitzky, 2024].

[15]Another extreme, but not entirely remote possibility is that the new elite bypasses the market by forming an elite club whose members bargain and trade critical resources exclusively among

GDP would become meaningless as a measure of economic activity. Moreover, many of the cross-cutting social relationships created by the need for human labor could disappear, along with a shared sense of identity and empathy.

To appreciate how extreme the society created by AGI could be, compare it again with feudalism. Though feudal estates were extremely hierarchical and unequal, feudal lords were enmeshed in a complex web of relationships with those who administered and worked their estates [Bloch, 2015/1939]. There was a steward who managed the estate, a marshal in charge of discipline, a clerk who kept the accounts, a butler, a butcher, a blacksmith, and many other roles. Thus, the feudal lord was firmly embedded in his community and bound by a set of traditional obligations. He did not have the option to escape to his own Elysium. That could all change with AI. Thus, the social fabric that underpins democracy, including shared identity and empathy, could be torn up.

The economic and social separation made possible by AI would further disrupt the financing of the state, making it hard to raise tax revenue from the rich and fund key protections for weaker members of society. As already noted, wealthy families will have unprecedented ability to vertically integrate. A family who today uses their car factory to produce cars for sale on the market and uses the proceeds to pay for food and clothing, as well as cooks, cleaners, and gardeners will pay corporate tax, income tax, payroll tax, and consumption tax. Now suppose the family converts the factory from manufacturing cars to robots and uses these robots to serve as household labor and grow food and cotton on the family land. Because all transactions occur within the household production unit there are no market transactions to tax. Hence, financing the state will be difficult in the age of AI. The stress on public finances could be even more acute in countries outside of the US, who will have difficulty taxing the US-based AI companies. That is, unless AGI turns out to be cheap and easy to develop anywhere or these countries invest heavily in developing their own sovereign AI systems.

---

themselves.

The consequences of the collapse of democratic and egalitarian institutions would be dire. Consider the widespread belief that once AGI is achieved and humans are 'freed' from the need to work, governments will provide universal basic income (UBI) and everyone will live a life of leisure. In his essay titled "Economic Possibilities for Our Grandchildren", John Maynard Keynes envisioned something similar [Keynes, 1930]. In the utopian vision, we would all become artists and poets and live together in peace and prosperity. Erik Brynjolfsson has described just such a utopian future, which he calls 'Digital Athens' [see Tegmark, 2017]. The reason we talk about the philosophical, political, and artistic achievements of the ancient Athenian state, but not its economics, is because it was a slave economy. Citizens could engage in the 'higher pursuits' because slaves did all the work. Brynjolfsson's idea is that as AI and AI-powered robots do all the work in the coming decades, humanity will enter a new era of creativity and prosperity. Such a future relies crucially on a *ceteris paribus* assumption, however. It assumes that democratic institutions will survive the concentration of the economy.

On the contrary, the end of mutual dependency between classes would leave political and economic elites with no incentive to provide for UBI or other welfare measures. The progressive reforms of the twentieth century owe much to the indispensable role of human capital and labor in the production process. If AI and AI-powered robots make human capital and labor obsolete, the new elite may choose to unwind these progressive measures, including public education, public healthcare, pensions, and unemployment insurance. Key consumer and financial protections from predatory firms could also be removed. Market power would swing back to capital and extend well beyond pricing power to the power to expropriate by various means [Skaperdas, 1992, Garfinkel and Skaperdas, 2007]. In addition, many of the public goods currently provided by the state could become club goods produced within elite cliques. That includes private research and development, especially in areas of biotechnology such as genetic engineering and anti-aging. There is no guarantee that scientific knowledge will remain a public good. Thus, the average person would only be valued as a

consumer, and only as long as they have some wealth left to finance consumption. As such, democratic and egalitarian institutions could collapse and the population at large left for dead, or worse. The brutal feudal hierarchy might end up looking humane when compared to this Elysian regime.

Let us now turn briefly to the second way in which AI could break democratic and egalitarian institutions. Not only could AI lower the cost of political unrest for the elite by reducing its dependence on human workers, it may also make people less willing to express political opinions and protest violations of democratic principles. Dissent is likely to be suppressed in an Orwellian manner by AI-powered surveillance technology, and later robots [Orwell, 2003/1949]. This is already taking place in China. Beraja, Kao, Yang, and Yuchtman [2023a] find that local governments in China respond to civil unrest by procuring AI facial-recognition technology. These purchases are successful in quelling unrest, and also spur further innovation in AI surveillance. Beraja, Kao, Yang, and Yuchtman [2023b] provide evidence that China is exporting this surveillance technology to a large number of countries, especially autocracies and weak democracies. In addition, citizens may willingly hand over significant power to political and technocratic elites as they become overwhelmed by the pace, scale, and sheer novelty of the changes taking place due to AI. One may think that the AI systems we are developing will be the sentient yogurt of our near future. Instead, the yogurt might be the new political and technocratic elites to which the citizenry defers during the AI revolution.

Finally, AI will begin to create nearly limitless virtual worlds that distract and sedate the population. As the old world rots, generative AI will allow individuals to create customized alternative worlds in which they can live psychically. The gap between the attractiveness of material versus virtual reality will be immense. For example, a person will be able to design their own AI partner, either in digital or robot form. They will be able to speak to a screen and ask it to create a movie in which they are the main character. The customization of 'reality' is already happening to a lesser degree in the form of curated social media feeds. Social media companies

have learned how to make the content they host more addictive. As such, teenagers in the US spend around 4.8 hours a day on social media platforms [Gallup, 2023]. AI will hypercharge the absorption of people's minds into the digital world. But rather than boosting engagement through hateful content as currently occurs on social media, enabling individuals to create customized virtual worlds will distract and pacify, much like the 'soma' of Huxley's Brave New World [Huxley, 2006/1932] or Marx's opium of the people [Marx, 1970/1844]. This psychic break from and radical atomization of society will be largely voluntary. When faced with Hades, a substantial proportion of humanity will choose the Matrix. If we pursue this logic all the way, then eventually there will only be a world of AI and robots serving a tiny elite in palace households. The rest of humanity will have been eliminated. At that point, the democratic nation state and liberal international order would no longer exist or make sense.

3.3 Human Productivity Collapse

Precisely when the catastrophic risks of AI might be realized is unpredictable. Much will depend on the pace of AI progress. There are reasons to think, however, that the process may advance rapidly. Typically, AGI is thought of as a state in which AI performance has caught up to a *fixed* human benchmark. I will now set out reasons why that may be too optimistic. There is considerable evidence that AI currently boosts human productivity in fields such as customer support [Brynjolfsson et al., 2023], diagnostic medicine [Topol, 2019], and software development [Cui et al., 2024]. As AI programs progress and substitute for human cognition, however, I suggest that feedback loops could be established which drastically reduce human productivity over time. As such, AGI may be reached faster than expected, not through an artificial intelligence explosion, but through a collapse in human productivity.

A productivity collapse may seem far-fetched, but it is not without precedent. Literacy declined sharply in Europe after the fall of the Western Roman Empire in the fifth century [Jones, 1964, Harris, 1989, Ward-Perkins, 2005]. In the eighth century,

Charlemagne, ruler of one of the largest medieval empires in Europe, was at best semiliterate. Another example is the experience of indigenous societies in the wake of colonization. Due to cultural alienation and technological disruption, many indigenous societies lost their ability to sustain modes of production, ritual, and communal interaction that had been practiced for thousands of years [Fanon, 1961, Wolfe, 2006, Kirmayer et al., 2009].

Such abrupt, extreme shifts are the hallmark of (amplifying) feedback loops [Young, 1998, 2024]. When it comes to AI, there are four mechanisms that could establish a feedback loop which destroys human productivity. They are linked to one of the distinctive features of AI—the automation of cognitive tasks. As such, these mechanisms were not a concern in earlier rounds of automation, which displaced workers from manual to cognitive tasks. The four mechanisms are as follows.

Human capital investment. Suppose that a new AI program goes from outperforming 50% of humans in a task to outperforming 85%. That is what happened when Google DeepMind's AlphaCode went from performing at the level of the median participant in coding competitions to out-competing 85% of participants when AlphaCode 2 was released in 2023 [AlphaCode Team, 2023]. Consider a high school student deciding whether to learn to code. After reading about this development, the student would be less likely to invest in coding skills because they are now less likely to out-compete the AI. Thus, without any further improvement in the AI system, it might be able to out-compete 95% of humans in the next cohort. Now this induces greater AI adoption by firms and further improvement in the AI program through experience. In this loop, it is not simply that the AI program is improving. At the same time, human productivity is collapsing due to declining investment in human capital.

Transfer learning. Transfer learning occurs when knowledge or skills gained in performing a task enhance performance in a different set of tasks [Haskell, 2000]. For example, there is evidence that learning a second language makes it easier to learn other, especially related, languages [Odlin, 1989]. As such, relying on artificial in-

telligence to perform task X can reduce a person's ability to perform another task Y. That would create an incentive to automate task Y, possibly further reducing human performance in task X, and so forth. In addition, there is evidence that transfer learning develops higher-level cognitive functioning in humans [Barnett and Ceci, 2002]. Hence, relying on AI to perform specific tasks could inhibit functions such as generalization, abstraction, and creativity, which are crucial to scientific progress by humans. Academic researchers often end their careers specializing in the development of new ideas and managing research teams. That is not how they start out, however. The reason is that a senior researcher will struggle to generate new ideas or supervise technical work if they have not received rigorous technical training themselves. Typically, biologists start out doing laboratory work, historians doing archival research, and anthropologists doing fieldwork. Likewise, economists without rigorous training in mathematics and practice constructing proofs would struggle to visualize structures that help solve existing empirical puzzles or identify new ones. Hence, using an AI program to do mathematics and other technical tasks early in one's education could have detrimental effects on higher-level cognitive functioning, including creativity. Thus, due to specific and general forms of transfer learning, a feedback loop could be established in which human productivity declines through the use of AI.

Social learning. A person's productivity is determined not only by their own investment in human capital, but through learning skills and knowledge from parents, neighbors, teachers, and peers. That is, human productivity is a networked trait. There is considerable evidence of such human capital spillovers. For example, individual outcomes are influenced by peers in educational settings [Sacerdote, 2001, Calvó-Armengol et al., 2009], workplaces [Moretti, 2004], and neighborhoods [Borjas, 1995]. Thus, if AI initially dissuades a subset of young people from investing in skills, contagion effects would produce a more general deterioration in skills. In particular, the productivity of workers further up the cognitive skills distribution could be dragged down by low skill formation at the bottom of the distribution.

That would incentivize further AI progress and automation by firms. Through this feedback loop, AGI could be achieved even more rapidly through a collapse in human productivity.

Role models. Beyond cognitive skills, human productivity is determined by a range of socially-learned general-purpose traits [Bandura, 1977]. Work ethic, punctuality, grit, patience, and diligence are all networked traits that are acquired through one's own experience and role models [e.g. Duckworth et al., 2007, Doepke and Zilibotti, 2008, Lalive and Cattaneo, 2009]. Now suppose that a large fraction of people in a neighborhood are displaced from the workforce due to AI. The unemployed individuals will likely experience a loss of skills and work habits leading to local hysteresis [Blanchard and Summers, 1986]. This would in turn create a dearth of positive role models in the neighborhood, which would reduce work ethic [Lindbeck et al., 1999], increase crime and anti-social behavior [Ridley et al., 2020], and depress economic outcomes for other members of the neighborhood [Chetty and Hendren, 2018]. Hence, by eroding critical non-cognitive traits, even a temporary unemployment shock of this scale could produce a long-term decline in human productivity. That would incentivize further automation, leading to another round of unemployment, and so forth. Again, AI adoption could cause a collapse in human productivity.

Possibly, the closest precedent is the devastation wrought by colonialism on indigenous societies. Colonial policies confiscated land and children and targeted indigenous cultural practices, including religious and communal rituals, traditional economic activities such as hunting and fishing, and ancestral political institutions [Fanon, 1961, Pilkington, 2002, Wolfe, 2006]. This destruction of the social fabric and removal of traditional role models created serious social problems in indigenous communities (e.g. addiction), making it difficult for members to integrate into the new political-economic environment or sustain traditional practices [Kirmayer et al., 2009, Dell, 2010, Nunn and Wantchekon, 2011, Dippel, 2014]. Such a comprehensive collapse in human capabilities is unlike anything experienced by advanced economies in previous rounds of automation. It is, however, something that could be brought

about by AI. Hence, in the competition between humans and AI, we should not assume that human productivity will remain fixed. Instead, AI adoption could cause rapid human obsolescence in the workforce. Elites with access to AGI and advanced robots will have little incentive to stop the productivity collapse. Perversely, they might even benefit from it.

4 Managing the Political-Economic Risks of AI

The dystopian end state that I have described must not be accepted as inevitable. In his study of the Industrial Revolution, Karl Polanyi cautioned against a fatalistic attitude toward technological change:

> "Nowhere has liberal philosophy failed so conspicuously as in its understanding of the problem of change. Fired by an emotional faith in spontaneity, the common-sense attitude toward change was discarded in favor of a mystical readiness to accept the social consequences of economic improvement, whatever they might be. The elementary truths of political science and statecraft were discredited then forgotten. It should need no elaboration that a process of undirected change, the pace of which is deemed too fast, should be slowed down, if possible, so as to safeguard the welfare of the community." [Polanyi, 2001/1944, p. 35]

When it comes to the AI revolution, the problem of change will be acute. Behind the veil of ignorance, nobody would choose the world described in Section 3. That is why it is important to take precautions now, before special interests fully crystallize and before we know where the chips will fall. It will, however, be unwise and nearly impossible to slow the pace of technological development. Instead, the political-economic tidal wave must be properly channeled.

In this section, I will describe measures for dealing with the impact of AI and ensuring it has widespread benefits. They are a reorientation of competition policy, measures to reinforce democracy, and the development of sovereign AI systems. Just

as the diagnosis of the political-economic problem involves a speculative element, this section is far from a full and definitive account of how to solve the problem. For example, to prevent a collapse in human productivity, Acemoglu and Johnson's (2023) proposal to develop human-oriented AI that complements and enhances human capabilities will be critical [see also Mollick, 2024]. Arriving at the correct set of solutions will require collaboration and this section should be viewed as an early contribution to a much-needed collective endeavor. One thing that is already clear, however, is that managing the political-economic risks of AI will be a balancing act. Given the pressures of geopolitical competition and institutional exhaustion, we must be careful that policy responses do not stifle fundamental technological progress.

4.1 Competition Policy

Competition policy will assume a new importance in the AI age, as should be apparent from the risks set out in Section 3. Regulators such as the Federal Trade Commission (FTC) in the United States and the Competition and Markets Authority (CMA) in the United Kingdom are charged with promoting competition and protecting consumers. They have the power to block mergers and acquisitions deemed to be anti-competitive, and enforce laws against price fixing, false advertising, privacy violations, and other practices that harm consumers. Maximizing consumer welfare became the objective of competition policy only relatively recently, inspired by the Chicago School in the 1970s and 1980s [Stigler, 1971, Posner, 1976, Bork, 1978, Landes and Posner, 1981].[16] This narrow objective makes sense in a stable environment where anti-competitive practices cannot concentrate power to the extent that they threaten democratic institutions. Consider, for example, the 2020 merger between Sprint Corporation and T-Mobile. The main risk of reducing the number of competitors in the wireless communications sector from four to three was the reduction in consumer surplus due to higher prices, fewer options, and lower-quality service.

[16]The Chicago school also promoted de-regulation to avoid regulatory capture, and because it was believed that contestable markets would limit market power.

The merger could hardly create a power bloc that controlled a significant part of the economy, let alone undermine democracy.

The current challenge faced by competition authorities is how to regulate the digital economy, that is, the platform economy dominated by firms such as Google, Meta, and Amazon. Initiatives include the US Platform Competition and Opportunity Act of 2021 and the UK Digital Markets, Competition and Consumers Act of 2024. The digital economy's emphasis on innovation, network effects, and elaborate ecosystems call for a rethinking of competition policy. A model can be built on the concept of creative destruction, associated most closely with the work of Joseph Schumpeter and Philippe Aghion, in which the objective of competition policy is to maximize innovation.[17] The main principle is that some degree of market power is required to give firms an incentive to innovate [Schumpeter, 1942, Aghion and Howitt, 1992]. However, if a successful innovator establishes too great a lead on its competitors, its incentives to further innovate decline [Aghion et al., 2005, 2019a]. It could also use its market power to handicap competitors, limit entry, or otherwise stifle further innovation [Shapiro, 2002, Cunningham et al., 2021]. Therefore, a dynamic approach to competition is required in which regulators target anti-innovation practices.

While the Schumpeterian approach applies well to the digital economy, competition authorities have not yet developed a conceptual framework for dealing with AI. Clearly, the need to promote innovation is of critical importance in the AI age. But that is not the whole picture. As with consumer welfare, an exclusive focus on innovation is only suitable in an institutionally stable environment. As should be clear from the political-economic risks described in Section 3, however, AI has the potential to radically destabilize political and social institutions. Thus, to reorient competition policy for the AI age, we need to look back to the less secure past.

The original objective of antitrust policy was far more broad and fundamental than

[17] Innovation could be thought of as a way to dynamically optimize consumer and producer surplus.

protecting consumer welfare or encouraging innovation. The objective was to limit the concentration of economic and political power and thereby stabilize democratic institutions. The first federal antitrust legislation in the US, the Sherman Antitrust Act of 1890, was a response to the rise of mega corporations such as Standard Oil and U.S. Steel whose consolidated control over industries gave them the ability to fix market prices and stifle competition. In the gilded age built on critical inputs such as oil and steel, there was a fear that market power would concentrate political power in the hands of a few individuals and corporations. Senator John Sherman, principle author of the Act, expressed this emphatically:

> "If the concentered powers of this combination are [e]ntrusted to a single man, it is a kingly prerogative, inconsistent with our form of government, and should be subject to the strong resistance of the State and national authorities. If anything is wrong this is wrong. If we will not endure a king as a political power we should not endure a king over the production, transportation, and sale of any of the necessaries of life. If we would not submit to an emperor we should not submit to an autocrat of trade, with power to prevent competition and to fix the price of any commodity." [Sherman, 1890]

AI brings these concerns back to the fore. Hence, competition policy in the AI age must balance incentives for innovation with limiting concentrations of power that threaten democracy. To do so, regulators will have to adopt new metrics for economic concentration. One measure currently in use is the Herfindal-Hirschman Index (HHI) which measures industry concentration.[18] A post-merger rise in HHI indicates greater potential for pricing power, collusion, and other forms of anti-competitive behavior. Market concentration will continue to be an important consideration. Collusion between the leading AI providers is needed for the AI complex to have the

[18]Suppose there are n firms in the industry before firms n and $n-1$ merge, s_i' is the market share of firm i before the merger, and s_i is its market share after the merger. Then the pre-merger index is $HHI' = \sum_{i=1}^{n}(s_i')^2$ and the post-merger index is $HHI = \sum_{i=1}^{n-1}(s_i)^2$. The merger is partly assessed by comparing these two indices, i.e. computing $HHI - HHI'$.

pricing and hold-up power described in Sections 3.1.2-3.1.3 that would drastically concentrate the global economy. In addition, genuine competition among leading AI companies and free entry would limit authoritarian corporate and state use of AI (e.g. control of AI agents). Individuals and firms purchasing AI services could then switch to an alternative (and less distorted) AI system at low cost. That is a necessary step toward 'democratizing AI'. To fully deal with the institutional impact of the AI revolution, however, regulators may have to adopt unconventional measures of economic concentration that capture broader notions of power.

Regulators should not only be concerned with market concentration (i.e. the market share of large firms in an industry), but also with how much of the economy is controlled by key individuals or organizations. I suggest employing network measures to quantify this broader notion of power when assessing mergers and acquisitions [Colizza et al., 2006, Vitali et al., 2011]. For example, large shareholders can amplify their control via cross-holdings and pyramidal structures [Faccio and Lang, 2002]. Suppose individual i purchases a 51% stake in company XYZ worth \$100 billion. Then i uses their controlling stake to have XYZ purchase a 51% stake in ten companies each worth \$10 billion. Then i uses their controlling stake in XYZ to have XYZ use its controlling stakes to induce the ten new acquisitions to each purchase a 51% stake in ten companies each worth \$1 billion. Individual i could effectively control all of the companies in this pyramidal structure worth a total of \$300 billion with only a \$90 billion investment, that is, only 30% of the cash-flow rights. In practice, more complicated network measures would be needed to capture the concentration of economic control [see Elliott and Galeotti, 2019]. *Inter alia*, input-output analysis [Hurwicz, 1955] and centrality measures [Bartesaghi et al., 2024] could be used to identify critical firms and parts of the AI supply chain.[19]

This reorientation of competition policy will need to respect a number of constraints. First, limiting the accumulation of power should be weighed against the potentially

[19]See, for example, Clayton et al. [2024] who show how input-output analysis can be used to model the exercise of geopolitical power in pursuit of economic rents.

large costs of stifling innovation and creative destruction. Second, competition policy must not be used to penalize the successful, only to stop them blocking further rounds of innovation. Third, security considerations may require some constraints on free development and use of AI. However, freedoms must not be forfeited without deliberation and unless absolutely necessary. Ultimately, I expect the major obstacle to reorienting competition policy not to be striking a balance with innovation and security, but rather opposition from the AI sector in the form of massive investment in lobbying, legal defense, and 'cognitive capture' [see Johnson and Kwak, 2011]. The citizenry may have to rely on rivalry between elite factions or the support of a sympathetic elite faction to secure competition, innovation, and freedom.

4.2 Reinforcing Democracy

To further understand how the AI revolution will threaten democracy and how to guard against these threats, let us decompose the problem of democratic governance into two parts. First, there is the *common-value* problem: More efficient governance (e.g. public good provision) expands the set of possibilities for allocating resources and increases the welfare of all citizens.[20] Second, there is the *private-value* problem:[21] Citizens will have different preferences and a democratic system will have to somehow balance the competing objectives, while maintaining political stability.[22] AGI has the potential to make government more efficient at achieving its objectives, mitigating the common-value problem. However, for reasons I have set out, AGI could greatly skew political objectives toward the private interests of a narrow elite, worsening the private-value problem and eventually destroying the system of

[20]There is evidence that elections can improve efficiency even in authoritarian regimes such as Egypt under Mubarak [Blaydes, 2011].

[21]The terms 'private value' and 'common value' come from the economics literature on auctions [Vickrey, 1961, Wilson, 1969, Klemperer, 2004]. A private value auction is when each bidder has an independent value for the item (e.g. idiosyncratic tastes for a painting). A common-value auction is when the item has the same intrinsic value for all bidders, but this value is unknown (e.g. different estimates of the value of an oil field).

[22]The aggregation of preferences is fraught with difficulties [Arrow, 1951, Sen, 1970, Myerson and Satterthwaite, 1983, Saari, 2012].

democratic and egalitarian institutions.

We have seen how competition policy could play a critical role in the AI age. More broadly, preserving democratic and egalitarian institutions will require concerted political effort by citizens to reinforce the pillars of democracy, including electoral integrity, freedom of speech, free media, and civil society engagement [Diamond, 1999, Held, 2006]. Democracy, however defined, is not a perfect system of government. Plato (an aristocrat) famously despised democracy, claiming that it leads naturally to the worst form of government—tyranny [Plato, 2000]. Almost two and half millennia later, modern democracy is increasingly viewed with suspicion. Leading political figures such as Lee Kuan Yew have claimed that democracy undermines economic development [Yew, 2000]. In addition, democracy is sometimes viewed not as governance by the people, but as a mechanism for power sharing among the elite [Mosca, 1939/1896, Michels, 1915, Pareto, 1935/1916, Mills, 1956]. Elections simply determine which elite faction has control at any point in time. Nevertheless, we may not have a better system to protect against the concentration of power potentially brought about by AGI and ASI. As set out in Section 3, the danger is that the AI revolution could unify the interests of a narrow elite connected to the AI complex and induce this elite to dispense with elections and other institutions previously used for purposes such as power sharing. As democratic and egalitarian institutions are eroded, there will be a point of no return. Concerted grass roots political action will be required to prevent a downward democratic spiral.

As a circuit breaker, I propose an old solution, one with which Plato was familiar. *Sortition* is the selection of members of decision-making bodies by lot. In ancient Athens, members of councils and juries and minor officials were selected in this way [Malkin and Blok, 2024]. A mixture of random selection and election was also used to choose council members in medieval Venice, Florence, and the city-states of Lombardy. Today, jury panels are randomly selected in countries such as the US, UK, Canada, Australia, and New Zealand. In addition, France, Canada, and Ireland have formed citizen assemblies chosen by lot to deliberate on specific issues. One moti-

vation for sortition in the ancient and medieval world was to balance representation across tribes and thereby stabilize the political system. A legitimate distribution of political authority is an ongoing concern [Stone, 2007]. Other motivations include reducing incentives for corruption and increasing civic engagement [Van Reybrouck, 2016]. The potential costs are obviously in the form of reduced expertise and accountability in decision-making roles.[23] That is, sortition may worsen the common-value problem while helping to solve the private-value problem.

To deal with the AI shock, sortition may be desirable because the private-value problem will be far worse than the common-value problem. In particular, I propose choosing a fraction of legislative members and minor officials at random from a nation's population. Random selection is of course better than negative selection. AI will mean that human expertise is not the most important factor in governance. Instead, the critical task will be to safeguard the interests of the citizenry against a narrow elite connected to the AI complex. Not only might the interests of the citizenry be directly represented by these citizen-legislators, but sortition could also have social psychological effects that stabilize the legitimacy of the political system. These include fostering a culture of civic engagement [Putnam et al., 1993, Putnam, 2000, Tabellini, 2008], trust in political institutions [Dalton, 2004, Norris, 2011], and identification with the political class [Carvalho and Dippel, 2020]. Without such measures, democratic backsliding could turn into a rapid retreat from democratic and egalitarian institutions.

4.3 Sovereign AI

While geopolitical competition with China will assume first-order importance, I will focus here on an issue that has received far less attention: How can a balance of power be preserved within the US-led coalition? The answer is for other democratic nations to develop and secure their own sovereign AI systems. Without such moves,

[23] However, it could be that randomly selected groups make better decisions due to cognitive diversity [Page, 2008, Landemore, 2013].

the US will become the center of an AI black hole that swallows up a large part of the global economy. US policymakers and corporations will be tempted to use their leverage to transfer a large share of the global economic surplus to the US, while US interests would exert even greater influence on other nations' foreign and domestic policies. Clearly, this concentration of global power is not in the interests of other members of the US-led coalition. But it could also weaken the US.

Recall that there are three channels through which countries that fail to develop sovereign AI systems face economic decline and political destabilization. First, AI companies located in the United States will have control over the new 'cognitive labor force' around the world. This will confer global market power, the ability to hold up foreign organizations, and launch national strikes in other countries. Second, if the most powerful AI systems are able to create and defend against new weapons of mass destruction (e.g. pathogens, cyberattacks), as anticipated, then countries without access to these systems will rely heavily on the US for their security and defense. Just as the UK and France developed their own nuclear weapons systems for self-sufficiency and bargaining power in strategic alliances, developed nations should invest in their own sovereign AI systems. Third, a state without its own advanced AI will lose much of its tax base and have trouble financing itself. As income tax revenue collapses due to automation, it will be difficult to switch to taxing corporations located in the US. Sovereign AI systems would mitigate all of these problems.

Many nations have already begun sovereign AI initiatives. Denmark has developed an AI supercomputer named 'Gefion' which is powered by Nvidia's H100 GPUs. India has formed a strategic partnership with Nvidia and Indian IT companies to develop local AI models. Softbank is building supercomputers in Japan using Nvidia hardware for generative AI capabilities. France's Scaleway is building Europe's most powerful cloud-native AI supercomputer. Supercomputers for AI capabilities are also being constructed in Canada, Italy, Taiwan, and Singapore. The goal in all cases is to develop AI models (especially LLMs) that are tailored to the local linguistic,

cultural, and institutional environments. From a political economy perspective, the important point is that these AI systems are domestically owned and controlled. Of course, it is uncertain how far these initiatives will go toward bridging the AI gap with the US. The Chinese open-source LLMs DeepSeek-V3 and DeepSeek-R1 suggest that rapid catch-up is possible despite hurdles in the form of access to advanced chips and limited training budgets. Whether such catch-up strategies will work when AGI or ASI is reached is unclear.

A balance of power within the US-led coalition of democratic nations could be in the US interest for the following simple reason. Hollowing out of economies outside the US could shrink the US sphere of influence and hamper its geopolitical competition with China. First, cooperation between the US and other democratic nations could break down; these nations may even be pushed into the geopolitical orbit of China or Russia. Second, overconcentration of the AI sector in the US will make the West more fragile. Nature discovered the value of redundancy a long time ago (e.g. two kidneys). Without distributed AI innovation and application, the coalition of democratic nations will be more susceptible to cyberattacks, physical attacks (e.g. bombing of data centers), energy crises, and other forms of failure. Third, no group of humans will know the optimal way to develop or channel AI technology. Nor will a superintelligent AI be able to accurately simulate all of the paths for developing and adapting to AI. Economic history tells us that competition between different approaches to technological and organizational innovation is critical to development [e.g. Hayek, 1988]. Hence, it would be good to avoid groupthink and have a diversity of experiments going on in parallel [Page, 2008]. Therefore, maintaining a balance of power within the US-led coalition of democratic nations by developing secure, sovereign AI systems could strengthen the US and its allies.

5 Conclusion

We are in the early stages of an AI revolution that is transforming political, economic, and social organization as radically as the agricultural and industrial revolutions, perhaps more so. Analysts have so far focused on the technological problem of controlling superintelligent AI systems and aligning their behavior to human interests. Relatively little attention has been paid to the political-economic problem which is equally existential. What will it be like to live in a post-AGI world? Due to distinctive features of AI—automation of cognitive tasks, global scalability, general-purpose technology, and importance to national security—I suggest that the impact of AI will be unlike earlier rounds of automation. I trace a path along which there is massive concentration of the global economy, a breaking of democratic and egalitarian institutions, and a collapse in human productivity. The experience would be more like the devastation of indigenous societies by colonialism than the impact of the steam engine, power loom, internal combustion engine, or any prior form of automation. Much of humanity could be immiserated, or even exterminated. The measures I have proposed to deal with these political-economic risks are a reorientation of competition policy toward limiting the accumulation of economic control, strengthening democracy through institutions such as sortition, and the development of sovereign AI systems by democratic nations outside of the US.

Such a study has a number of limitations. First, anticipating the effects of a new technology, especially one so powerful and poorly understood, is necessarily a speculative exercise. It is possible that AI produces a world of superabundance and unprecedented freedom, allowing humans to fully develop their potential. We should not discount this ideal outcome. Proper risk management, however, requires accounting for worst-case scenarios. This essay explores one such scenario. Second, the policy responses I have proposed are neither complete nor definitive. A full and proper set of responses will require a collaborative venture over the coming years. One issue that I have left to future work is the need for new social institutions that hold so-

ciety together and preserve human capabilities, meaning, and purpose in the face of massive disruption by AI. Without social adaptation, the AI shock will lead to social breakdown and conflict [Carvalho and Sacks, 2024, Carvalho, 2025]. Third, it is uncertain precisely how the technological revolution will unfold. Recent developments suggest that the leading AI companies in the US may not have a moat. That is, they could be caught up by foreign rivals with smaller budgets and inferior hardware. If that continues once AGI and ASI are reached, some of the political-economic risks I have identified will dissipate, while new geopolitical and security risks would arise. However, the principal risk that AI breaks the mutual dependency between financial capital, human capital, and labor would remain. If it is inexpensive to develop and run AGI, then the new elite will not be limited to those closest to the AI sector. Instead, a similar Elysian scenario could unfold with the winners being those with established networks and pools of capital who can leverage cheap AI to accumulate market power. All of this will play out in the next few years. Policymakers need to move fast while we are still, at least partially, behind the veil of ignorance.

References

Daron Acemoglu and Simon Johnson. *Power and Progress: Our Thousand-year Struggle over Technology and Prosperity*. Hachette UK, 2023.

Daron Acemoglu and Pascual Restrepo. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30, 2019.

Daron Acemoglu and James A Robinson. Why did the West extend the franchise? Democracy, inequality, and growth in historical perspective. *Quarterly Journal of Economics*, 115(4):1167–1199, 2000.

Daron Acemoglu and James A. Robinson. *Why Nations Fail: The Origins of Power, Prosperity and Poverty*. Profile Books, 2012.

Daron Acemoglu and Alexander Wolitzky. Employment and community: Socioeconomic cooperation and its breakdown. working paper, MIT, 2024.

Philippe Aghion and Peter Howitt. A model of growth through creative destruction. *Econometrica*, 60(2):323–351, 1992.

Philippe Aghion, Nicholas Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt.

Competition and innovation: An inverted-U relationship. *Quarterly Journal of Economics*, 120(2):701–728, 2005.

Philippe Aghion, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li. Missing growth from creative destruction. *American Economic Review*, 109(8):2795–2822, 2019a.

Philippe Aghion, Benjamin Jones, and Charles Jones. Artificial intelligence and economic growth. In A. Goldfarb A. Agrawal, J. Gans, editor, *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, 2019b.

Ajay Agrawal, Joshua Gans, and Avi Goldfarb. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Press, 2022.

Alberto Alesina and Marco Tabellini. The political effects of immigration: Culture or economics? *Journal of Economic Literature*, 62(1):5–46, 2024.

Robert C Allen. *The British Industrial Revolution in Global Perspective*. Cambridge University Press, 2009.

AlphaCode Team. Alphacode 2 technical report. Technical report, DeepMind, 2023. URL https://storage.googleapis.com/deepmind-media/AlphaCode2/AlphaCode2_Tech_Report.pdf.

Sam Altman. Near the singularity; unclear which side. X post, January 2025. URL https://twitter.com/sama/status/1716972815960961174.

Stephen Ansolabehere, John M. de Figueiredo, and James M. Snyder Jr. Why is there so little money in U.S. politics? *Journal of Economic Perspectives*, 17(1):105–130, 2003.

Kenneth J. Arrow. *Social Choice and Individual Values*. John Wiley & Sons, New York, USA, 1951.

Leopold Aschenbrenner. Situational awareness, 2024. URL https://situational-awareness.ai/wp-content/uploads/2024/06/situationalawareness.pdf.

David H. Autor. Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3):3–30, 2015.

David H. Autor, David Dorn, and Gordon H. Hanson. The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, 103(6):2121–2168, 2013.

Albert Bandura. *Social Learning Theory*. General Learning Press, New York, 1977.

Graeme Barker. *The Agricultural Revolution in Prehistory: Why did Foragers become Farmers?* Oxford University Press, 2006.

Susan M Barnett and Stephen J Ceci. When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128(4):612, 2002.

Larry M. Bartels. *Unequal Democracy: The Political Economy of the New Gilded Age*. Princeton University Press, Princeton, NJ, 2008.

Paolo Bartesaghi, Gian Paolo Clemente, and Rosanna Grassi. A self-adaptive centrality measure for asset correlation networks. *Economies*, 12(7):1–19, 2024.

Martin Beraja, Alex Kao, David Y. Yang, and Noam Yuchtman. AI-tocracy. *Quarterly Journal of Economics*, 138(3), 2023a.

Martin Beraja, Andrew Kao, David Y. Yang, and Noam Yuchtman. Exporting the surveillance state via trade in AI. Technical report, 2023b.

Drago Bergholt, Fabio Furlanetto, and Nicolò Maffei-Faccioli. The decline of the labor share: New empirical evidence. *American Economic Journal: Macroeconomics*, 14(3): 163–198, 2022.

Bipartisan Senate AI Working Group. Driving U.S. innovation in artificial intelligence: A roadmap for artificial intelligence policy in the United States senate, May 2024. URL https://www.schumer.senate.gov/newsroom/press-releases/following-historic-ai-insight-forums-over-the-past-year-leader-schumer-senators-rounds-h

Olivier Blanchard. DeepSeek and what happened yesterday: Probably the largest positive tfp shock in the history of the world., January 2025. URL https://x.com/ojblanchard1.

Olivier J Blanchard and Lawrence H Summers. Hysteresis and the European unemployment problem. *NBER Macroeconomics Annual*, 1:15–78, 1986.

David G. Blanchflower, Alex Bryson, and Xiaowei Xu. The declining mental health of the young and the global disappearance of the hump shape in age in unhappiness. National Bureau of Economic Research, April 2024. URL https://doi.org/10.3386/w32337.

Lisa Blaydes. *Elections and Distributive Politics in Mubarak's Egypt*. Cambridge University Press, Cambridge, UK, 2011.

Marc Bloch. *Feudal Society*. Routledge, 2015 edition, 2015/1939. Routledge Classics Edition.

George J. Borjas. Ethnicity, neighborhoods, and human-capital externalities. *American Economic Review*, 85(3):365–390, 1995.

Robert H. Bork. *The Antitrust Paradox: A Policy at War with Itself*. Basic Books, New York, 1978.

Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

Nick Bostrom. *Deep Utopia: Life and Meaning in a Solved World*. Oxford University Press, Oxford, UK, 2024.

Erik Brynjolfsson and Andrew McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton Company, 2014.

Erik Brynjolfsson, Danielle Li, and Lindsey R Raymond. Generative AI at work. National Bureau of Economic Research working paper 31161, 2023.

Edmund Burke. *Reflections on the Revolution in France*. Oxford World's Classics. Oxford University Press, 2009/1790.

Antoni Calvó-Armengol, Eleonora Patacchini, and Yves Zenou. Peer effects and social networks in education. *Review of Economic Studies*, 76(4):1239–1267, 2009.

Jean-Paul Carvalho. Markets and communities: The social cost of the meritocracy. *Journal*

of Institutional Economics, 18(3):501–519, 2022.

Jean-Paul Carvalho. Religion and terrorism: The religious utility hypothesis. In A. Basu-choudhary and G. Schulze, editors, *Handbook of the Economics of Terrorism*. Cambridge University Press, 2025.

Jean-Paul Carvalho and Christian Dippel. Elite identity and political accountability: A tale of ten islands. *The Economic Journal*, 130(631):1995–2029, 2020.

Jean-Paul Carvalho and Michael Sacks. Radicalisation. *The Economic Journal*, 134:1019–1068, 2024.

Jean-Paul Carvalho, Mark Koyama, and Cole Williams. Resisting education. *Journal of the European Economic Association*, 22(6):2549–2597, 2024.

Anne Case and Angus Deaton. *Deaths of Despair and the Future of Capitalism*. Princeton University Press, 2020.

Jacques Cauvin. *The Birth of the Gods and the Origins of Agriculture*. Cambridge University Press, 2000.

R. Cellan-Jones. Stephen Hawking warns artificial intelligence could end mankind. BBC News. https://www.bbc.com/news/technology-30290540, 2014.

R. Cellan-Jones. Stephen Hawking - will AI kill or save humankind? BBC News. https://www.bbc.com/news/technology-37713629, 2016.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *Quarterly Journal of Economics*, 133(3):1107–1162, 2018.

Angelique Chrisafis. France forced to drop 75% supertax after meagre returns. *The Guardian*, December 2014. URL https://www.theguardian.com/world/2014/dec/31/france-75-percent-supertax-dropped-meagre-returns.

Christopher Clayton, Matteo Maggiori, and Jesse Schreger. A framework for geoeconomics. National Bureau of Economic Research working paper 31852, 2024.

Vittoria Colizza, Alessandro Flammini, M Angeles Serrano, and Alessandro Vespignani. Detecting rich-club ordering in complex networks. *Nature Physics*, 2(2):110–115, 2006.

N. F. R. Crafts. English workers' real wages during the industrial revolution: Some remaining problems. *Journal of Economic History*, 45(1):139–144, 1985.

Zheyuan Kevin Cui, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz. The effects of generative AI on high skilled work: Evidence from three field experiments with software developers. SSRN working paper 4945566, 2024.

Colleen Cunningham, Florian Ederer, and Song Ma. Killer acquisitions. *Journal of Political Economy*, 129(3):649–702, 2021.

Russell J. Dalton. *Democratic Challenges, Democratic Choices: The Erosion of Political Support in Advanced Industrial Democracies*. Oxford University Press, Oxford, UK, 2004.

Dataintelo. Data broker market size, share, trends, analysis, and forecast report, 2023. URL https://dataintelo.com/report/data-broker-market.

Melissa Dell. The persistent effects of Peru's mining mita. *Econometrica*, 78(6):1863–1903, 2010.

David Deming, Christopher Ong, and Lawrence H. Summers. Technological disruption in the labor market. National Bureau of Economic Research working paper 33323, 2025.

Larry Diamond. *Developing Democracy: Toward Consolidation*. Johns Hopkins University Press, Baltimore, MD, USA, 1999.

Christian Dippel. Forced coexistence and economic development: Evidence from Native American reservations. *Econometrica*, 82(6):2131–2165, 2014.

Matthias Doepke and Fabrizio Zilibotti. Occupational choice and the spirit of capitalism. *Quarterly Journal of Economics*, 123(2):747–793, 2008.

A.L. Duckworth, C. Peterson, M.D. Matthews, and D.R. Kelly. Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6):1087–1101, 2007.

Matt Egan. Exclusive: 42% of CEOs say AI could destroy humanity in five to ten years, June 14 2023. URL https://edition.cnn.com/2023/06/14/business/artificial-intelligence-ceos-warning/index.html.

Matthew Elliott and Andrea Galeotti. The role of networks in antitrust investigations. *Oxford Review of Economic Policy*, 35(4):614–637, 2019.

Gary V. Engelhardt and Jonathan Gruber. Social security and the evolution of elderly poverty. National Bureau of Economic Research working paper 10466, May 2004.

Friedrich Engels. *The Condition of the Working Class in England*. Lawrence & Wishart, London, 1845.

Mara Faccio and Larry HP Lang. The ultimate ownership of Western European corporations. *Journal of Financial Economics*, 65(3):365–395, 2002.

Frantz Fanon. *The Wretched of the Earth*. Grove Press, New York, 1961. Originally published as *Les Damnés de la Terre*.

John Fernald, Robert Inklaar, and Dimitrije Ruzic. The productivity slowdown in advanced economies: Common shocks or common trends? *Review of Income and Wealth*, 71, 2025.

Robert William Fogel. *The Escape from Hunger and Premature Death, 1700-2100: Europe, America, and the Third World*. Cambridge University Press, 2010.

Future of Life Institute. Pause giant AI experiments: An open letter, March 22 2023. URL https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

John Kenneth Galbraith. *The New Industrial State*. Houghton Mifflin, Boston, MA, 1967.

Gallup. Congress and the public. *Gallup.com*, January 15 2019. URL /https://news.gallup.com/poll/1600/congress-public.aspx.

Gallup. Teens spend average of 4.8 hours on social media daily. 2023. URL https://news.gallup.com/poll/512576/teens-spend-average-hours-social-media-perday.aspx.

Michelle R Garfinkel and Stergios Skaperdas. Economics of conflict: An overview. In Todd Sandler and Keith Hartley, editors, *Handbook of Defense Economics*, volume 2, pages

649–709. Elsevier, 2007.

Richard Gilbert. Looking for Mr. Schumpeter: Where are we in the competition–innovation debate? *Innovation Policy and the Economy*, 6:159–215, 2006.

Martin Gilens. *Affluence and Influence: Economic Inequality and Political Power in America*. Princeton University Press, Princeton, NJ, 2012.

Ian Goldin, Pantelis Koutroumpis, François Lafond, and Julian Winkler. Why is productivity slowing down? *Journal of Economic Literature*, 62(1):1–38, 2024.

I. J. Good. Speculations concerning the first ultraintelligent machine. In *Advances in Computers*, volume 6, pages 31–88. Elsevier, 1966.

Sanford J. Grossman and Oliver D. Hart. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4):691–719, 1986.

Yicheng Guo, Guokan Shang, Michalis Vazirgiannis, and Chloé Clavel. The curious decline of linguistic diversity: Training language models on synthetic text. *arXiv preprint arXiv:2311.09807*, 2023. URL https://doi.org/10.48550/arXiv.2311.09807.

Keach Hagey. Sam Altman seeks trillions of dollars to reshape business of chips and AI. Wall Street Journal, 2024. URL https://www.wsj.com/tech/ai/sam-altman-seeks-trillions-of-dollars-to-reshape-business-of-chips-and-ai-89ab3db0.

Yuval Noah Harari. *Homo Deus: A Brief History of Tomorrow*. Harper Perennial, 2015.

William V. Harris. *Ancient Literacy*. Harvard University Press, Cambridge, MA, 1989.

Robert E Haskell. *Transfer of learning: Cognition and instruction*. Elsevier, 2000.

FA Hayek. The use of knowledge in society. *American Economic Review*, 35(4):519–530, 1945.

Friedrich A. Hayek. *The Road to Serfdom*. Routledge, 1944. Original work published 1944.

Friedrich A. Hayek. *The Fatal Conceit: The Errors of Socialism*. The University of Chicago Press, Chicago, 1988.

Seamus Heaney. *Aeneid Book VI: A New Verse Translation*. Farrar, Straus and Giroux, 2016.

David Held. *Models of Democracy*. Polity, Cambridge, UK, 3 edition, 2006.

Joseph Henrich. *The Secret of our Success: How Culture is Driving Human Evolution, Domesticating our Species, and Making us Smarter*. Princeton University Press, 2017.

Joseph Henrich and Michael Muthukrishna. What makes us smart? *Topics in Cognitive Science*, 16(2):322–342, 2024.

Leonid Hurwicz. Input-output analysis and economic structure: A review article. *American Economic Review*, 45(4):626–636, 1955.

Aldous Huxley. *Brave New World*. Harper Perennial, 2006/1932.

G. John Ikenberry. *After Victory: Institutions, Strategic Restraint, and the Rebuilding of Order After Major Wars*. Princeton University Press, Princeton, NJ, 2001.

Simon Johnson and James Kwak. *13 bankers: The Wall Street takeover and the next financial meltdown*. Vintage, 2011.

Arnold Hugh Martin Jones. *The Later Roman Empire, 284–602: A Social, Economic, and*

Administrative Survey. Blackwell, Oxford, 1964.

Dae-Young Kang, Hieu Pham Duong, and Jung-Chul Park. Application of deep learning in dentistry and implantology. *Journal of Implantology and Applied Sciences*, 24(3): 148–181, 2020.

Melissa S. Kearney. *The Two-Parent Privilege*. Swift Press, 2023.

Robert L Kelly. *The Lifeways of Hunter-Gatherers: The Foraging Spectrum*. Cambridge University Press, 2013.

John Maynard Keynes. Economic possibilities for our grandchildren. In *Essays in Persuasion*, pages 358–373. Macmillan and Co., London, UK, 1930.

Laurence J. Kirmayer, Gregory M. Brass, and Gail Valaskakis, editors. *Healing Traditions: The Mental Health of Aboriginal Peoples in Canada*. University of British Columbia Press, Vancouver, 2009.

Benjamin Klein, Robert G. Crawford, and Armen A. Alchian. Vertical integration, appropriable rents, and the competitive contracting process. *Journal of Law and Economics*, 21(2):297–326, 1978.

Paul Klemperer. *Auctions: Theory and Practice*. Princeton University Press, Princeton, NJ, USA, 2004.

Greg Kohs. Alphago. Film, 2017. Available at: https://www.alphagomovie.com/.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25, pages 1097–1105, 2012.

Ray Kurzweil. *The Singularity is Near: When Humans Transcend Biology*. Duckworth, 2005.

Rafael Lalive and Maria A. Cattaneo. Social interactions and schooling decisions. *Review of Economics and Statistics*, 91(3):457–477, 2009.

Hélène Landemore. Deliberation, cognitive diversity, and democratic inclusiveness: An epistemic argument for the random selection of representatives. *Synthese*, 190(7):1209–1231, 2013.

William M. Landes and Richard A. Posner. Market power in antitrust cases. *Harvard Law Review*, 94(5):937–996, 1981.

Kai-Fu Lee. *AI Superpowers: China, Silicon Valley, and the New World Order*. Mariner Books, 2018.

Richard Borshay Lee and Irven DeVore. *Man the Hunter*. Routledge, 2017/1966.

K. Lima and Dennis J. Snower. Recoupling economic and social prosperity. In *Global Solutions Journal*, pages 118–167. 2020.

Assar Lindbeck, Sten Nyberg, and Jörgen W Weibull. Social norms and economic incentives in the welfare state. *Quarterly Journal of Economics*, 114(1):1–35, 1999.

Alessandro Lizzeri and Nicola Persico. Why did the elites extend the suffrage? Democracy and the scope of government, with an application to Britain's Age of Reform. *Quarterly Journal of Economics*, 119(2):707–765, 2004.

Humberto Llavador and Robert J. Oxoby. Partisan competition, growth and the franchise. *Quarterly Journal of Economics*, 120(3):1155–1189, 2005.

Irad Malkin and Josine Blok. *Drawing Lots: From Egalitarianism to Democracy in Ancient Greece*. Oxford University Press, Oxford, UK, 2024.

Alexandra M. Marconi, Ursula S. Myers, Breanne Hanson, Sarah Nolan, and Elizabeth B. Sarrouf. Psychiatric medication prescriptions increasing for college students above and beyond the covid-19 pandemic. *Scientific Reports*, 13(1):19063, 2023.

Karl Marx. *A Contribution to the Critique of Hegel's Philosophy of Right*. Cambridge University Press, Cambridge, 1970/1844.

Karl Marx. *Capital: A Critique of Political Economy, Volume I*. Princeton University Press, Princeton, 2024/1867.

Karl Marx and Friedrich Engels. *Economic and Philosophic Manuscripts of 1844 and the Communist Manifesto*. Prometheus, 1988/1848.

Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943.

Robert Michels. *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*. Free Press, 1915.

C. Wright Mills. *The Power Elite*. Oxford University Press, 1956.

Melanie Mitchell. Artificial intelligence: A guide for thinking humans. 2019.

Joel Mokyr. *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford University Press, 1992.

Ethan Mollick. *Co-Intelligence: Living and Working with AI*. Penguin Random House, 2024.

Enrico Moretti. Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review*, 94(3):656–690, 2004.

Stephen Moris and Tabby Kinder. Elon musk plans to expand Colossus AI supercomputer tenfold. Financial Times, 2024. URL https://www.ft.com/content/9c0516cf-dd12-4665-aa22-712de854fe2f.

Gaetano Mosca. *The Ruling Class*. McGraw-Hill, 1939/1896.

Charles Murray. *Coming Apart: The State of White America*. 2012.

Michael Muthukrishna and Joseph Henrich. Innovation in the collective brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1690):20150192, 2016.

Roger B. Myerson and Mark A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265–281, 1983.

National Institute of Mental Health. Mental illness. 2023. https://www.nimh.nih.gov/health/statistics/mental-illness.

Pippa Norris. *Democratic Deficit: Critical Citizens Revisited*. Cambridge University Press, Cambridge, UK, 2011.

Pippa Norris and Ronald Inglehart. *Sacred and Secular: Religion and Politics Worldwide*. Cambridge University Press, Cambridge, UK, 2004.

Douglass C. North. *Institutions, Institutional Change and Economic Performance*. Political Economy of Institutions and Decisions. Cambridge University Press, Cambridge, UK, 1990.

Nathan Nunn and Leonard Wantchekon. The slave trade and the origins of mistrust in Africa. *American Economic Review*, 101(7):3221–3252, 2011.

Terence Odlin. *Language Transfer: Cross-linguistic influence in language learning*. Cambridge University Press, 1989.

Avner Offer. *The Challenge of Affluence*. Oxford University Press, 2006.

Kevin Okemwa. Sam Altman claims AGI will 'whoosh by' in 5 years with surprisingly little societal change, while Anthropic CEO predicts a 2026 or 2027 breakthrough. *Windows Central*, November 2024. URL https://www.windowscentral.com/software-apps/sam-altman-claims-agi-will-whoosh-by-in-5-years-with-surprisingly-little-societal-change

OpenAI. Learning to reason with LLMs, September 12 2024. https://openai.com/index/learning-to-reason-with-llms/.

George Orwell. *1984*. Penguin Classics, 2003/1949.

Scott Page. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press, 2008.

Vilfredo Pareto. *The Mind and Society*. Harcourt, Brace and Company, 1935/1916.

Pew Research Center. Americans' views of artificial intelligence, 2023. URL https://www.pewresearch.org/short-reads/2023/11/21/what-the-data-says-about-americans-views-of-artificial-intelligence/.

Thomas Piketty. *Capital in the Twenty-First Century*. Harvard University Press, 2013.

Thomas Piketty. *Capital and Ideology*. Harvard University Press, 2020.

Doris Pilkington. *Follow the Rabbit-Proof Fence*. University of Queensland Press, 2002.

Plato. *The Republic*. Cambridge University Press, Cambridge, UK, 2000.

Karl Polanyi. *The Great Transformation*. Beacon Press, 2001/1944.

Richard A. Posner. *Antitrust Law: An Economic Perspective*. University of Chicago Press, Chicago, 1976.

Nikhil Prakash, Tamar Rott Shaham, Tal Haklay, Yonatan Belinkov, and David Bau. Fine-tuning enhances existing mechanisms: A case study on entity tracking. *arXiv*, 2024. https://arxiv.org/abs/2402.14811.

Robert D. Putnam. *Bowling Alone: The Collapse and Revival of American Community*. Simon Schuster, 2000.

Robert D. Putnam, Robert Leonardi, and Raffaella Y. Nanetti. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press, Princeton, NJ, USA, 1993.

Matthew Ridley, Gautam Rao, Frank Schilbach, and Vikram Patel. Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science*, 370(6522):eaay0214, 2020.

Janis Robertson and John Scalzi. When the yogurt took over, 2019. URL https://www.imdb.com/title/tt9788494/.

Frank Rosenblatt. The perceptron: A probabilistic model for information storage and

organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

Donald G Saari. *Geometry of voting*, volume 3. Springer Science & Business Media, 2012.

Bruce Sacerdote. Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, 116(2):681–704, 2001.

Austin E Schumacher, Amirali Aali, Yohannes Habtegiorgis Abate, Rouzbeh Abbasgholizadeh, Mohammadreza Abbasian, Mohsen Abbasi-Kangevari, Hedayat Abbastabar, Samar Abd ElHafeez, Sherief Abd-Elsalam, Mohammad Abdollahi, et al. Global fertility in 204 countries and territories, 1950–2021, with forecasts to 2100: a comprehensive demographic analysis for the global burden of disease study 2021. *The Lancet*, 2024.

Joseph A. Schumpeter. *Capitalism, Socialism and Democracy*. Harper & Brothers, New York, 1942.

Amartya Sen. *Collective Choice and Social Welfare*. Holden-Day, San Francisco, USA, 1970.

Jaime Sevilla. Can AI scaling continue through 2030?, August 20 2024. https://epochai.org/blog/can-ai-scaling-continue-through-2030.

Jaime Sevilla, Lennart Heim, Anson Ho, Tamay Besiroglu, Marius Hobbhahn, and Pablo Villalobos. Compute trends across three eras of machine learning. *arXiv*, 2022. https://arxiv.org/abs/2202.05924.

Suhail Shah. Sam Altman on OpenAI and artificial general intelligence, December 13 2023. https://time.com/6344160/a-year-in-time-ceo-interview-sam-altman/.

Carl Shapiro. Competition policy and innovation. OECD Science, Technology and Industry working paper 2002/11, 2002.

Carl Shapiro and Hal R. Varian. *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Review Press, Boston, MA, 1999.

John Sherman. Statement in support of the sherman antitrust act. Congressional Record, March 1890. 51st Congress, 1st Session, Senate.

Stergios Skaperdas. Cooperation, conflict, and power in the absence of property rights. *American Economic Review*, 8:720–739, 1992.

Stergios Skaperdas and Samarth Vaidya. Persuasion as a contest. *Economic Theory*, 51: 465–486, 2012.

Adam Smith. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford University Press, Oxford, 2008/1776.

Robert Solow. We'd better watch out, 1987. http://digamo.free.fr/solow87.pdf.

Hidehiro Sonoda. The decline of the Japanese warrior class, 1840-1880. *Japan Review*, (1): 75–101, 1990.

Keiran Southern. Trump backs musk in row over H-1B visas that divided Maga. The Times, December 2024. URL https://www.thetimes.co.uk/article/h1b-visa-trump-elon-musk-3vts9gcxd.

Daniel F. Spulber. Innovation economics: The interplay among technology standards, competitive conduct, and economic performance. *Journal of Competition Law and Economics*, 9(4):777–825, 2013.

Stanford Institute for Human-Centered Artificial Intelligence. Artificial intelligence index report 2024. Technical report, Stanford University, 2024. URL https://aiindex.stanford.edu/report/2024/.

George J. Stigler. The theory of economic regulation. *Bell Journal of Economics and Management Science*, 2(1):3–21, 1971.

Peter Stone. Why lotteries are just. *Journal of Political Philosophy*, 15(3):276–295, 2007.

Mustafa Suleyman. *The Coming Wave*. Crown, 2023.

Mustafa Suleyman. What is an AI anyway? TED April 2024, Vancouver https://www.youtube.com/watch?v=KKNCiRWd_j0, 2024.

Daniel Susskind. *A World Without Work: The Looming Future of Technological Unemployment*. Henry Holt and Company, 2020.

Simon M Sze, Yiming Li, and Kwok K Ng. *Physics of Semiconductor Devices*. John Wiley & Sons, 2021.

Guido Tabellini. Institutions and culture. *Journal of the European Economic Association*, 6(2-3):255–294, 2008.

Max Tegmark. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Penguin Books, 2017.

The Economist. Mustafa Suleyman & Yuval Noah Harari debate: What does the AI revolution mean for our future?, 2023. URL https://www.youtube.com/watch?v=7JkPWHr7sTY.

Eric J Topol. High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, 2019.

Thomas Tørsløv, Ludvig Wier, and Gabriel Zucman. The missing profits of nations. *Review of Economic Studies*, 90(3):1499–1534, 2023.

US Department of the Treasury. The 2022 financial report of the U.S. government. Technical report, 2024. URL https://fiscal.treasury.gov/reports-statements/financial-report/2022/.

U.S. Office of the Undersecretary of Defense. Fiscal year 2023 budget request, 2023. URL https://comptroller.defense.gov.

David Van Reybrouck. *Against Elections: The Case for Democracy*. Seven Stories Press, New York, NY, USA, 2016.

Yanis Varoufakis. *Technofeudalism: What Killed Capitalism*. Bodley Head, London, UK, 2023.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal*

of Finance, 16(1):8–37, 1961.

James Vincent. OpenAI cofounder Ilya Sutskever says the way AI is built is about to change, December 2024. URL https://www.theverge.com/2024/12/13/24320811/what-ilya-sutskever-sees-openai-model-data-training.

Vernor Vinge. The coming technological singularity: How to survive in the post-human era. In *Vision 21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, pages 11–22. NASA. Lewis Research Center, 1993. URL https://ntrs.nasa.gov/citations/19940022856.

Stefania Vitali, James B Glattfelder, and Stefano Battiston. The network of global corporate control. *PloS one*, 6(10):e25995, 2011.

Xavier Vives. Innovation and competitive pressure. *The Journal of Industrial Economics*, 56(3):419–469, 2008.

David Waldner and Ellen Lust. Unwelcome change: Coming to terms with democratic backsliding. *Annual Review of Political Science*, 21:93–113, 2018.

Bryan Ward-Perkins. *The Fall of Rome and the End of Civilization*. Oxford University Press, Oxford, 2005.

Richard Waters. Man beats machine at Go in human victory over AI, 2023. URL https://www.ft.com/content/175e5314-a7f7-4741-a786-273219f433a1. Accessed: 2023-02-17.

Max Weber. *Economy and Society: A New Translation*. Harvard University Press, Cambridge, MA, 2019/1921.

David N. Weil. Health and economic growth. In *Handbook of Economic Growth*, volume 2, pages 623–682. Elsevier, 2014.

Eric R. Weinstein. Tweet about embedded growth obligations (EGO). https://x.com/EricRWeinstein/status/1370847097033875457, March 2021.

Oliver E. Williamson. *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press, New York, 1975.

Robert Wilson. Competitive bidding with disparate information. *Management Science*, 15 (7):446–448, 1969.

Patrick Wolfe. Settler colonialism and the elimination of the native. *Journal of Genocide Research*, 8(4):387–409, 2006.

Lee Kuan Yew. *From Third World to First: The Singapore Story: 1965-2000*. HarperCollins, New York, NY, USA, 2000.

H. Peyton Young. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, 1998.

H. Peyton Young. *Spontaneous Order: How Norms, Institutions, and Innovations Emerge from the Bottom Up*. Oxford University Press, Oxford, 2024.