Cognitive Scaffolds for a Human Future

Interim Research Report on Agent-Based Analysis of AI's Risks to Democratic Governance

By Carl O. Pabo

June 2025

Table of Contents

Executive Summary

Humanity faces an unprecedented crisis of complexity as artificial intelligence rapidly transforms the world. Despite AI's accelerating impact, no one has offered any clear, systematic framework for understanding or addressing associated challenges and risks.

This is a problem: New tools for thought are desperately needed to help people make sense of and navigate this transformation, and I thus offer a new approach designed to help society anticipate and deal with some of the social, political, and economic challenges that may emerge with a rapid rise of AI.

Background: The approach offered here is entirely novel because I start with a novel premise. I accept that human judgment is needed when addressing these social, political, and economic risks, yet I note that the pressure of specialization in modern life means that few people will have done the kind of background work needed to develop patterns of thought (i.e., to develop the internal, biochemically instantiated neural networks in their own brains) that will be needed to think coherently about these large-scale problems.

In short: I realized that we each need to update our own neural networks before we'll be in a position to think carefully about AI, and I thus developed this new model in light of many years spent studying patterns of human thought.

This Agent-Based Model: I offer an agent-based model designed to encourage and help people think more carefully, and to deliberate with others in a more rational way, about changes likely to occur in this Age of AI. My model divides society into 34 different groups of actors (different "agents"—like big tech companies, parents and educators, political leaders, etc., as listed in Appendix A). I assume that different agents—most of whom are human or have human leaders—will tend to act "selfishly," acting to protect their own needs while helping friends and family. Plausible actions of agents (at this current stage in the rise of AI) are listed in Appendix B, and these actions will—in turn—affect the environment in which all other agents must act.

This model is not offered as an end in itself; it's designed as a tool that—if adopted and used by others—can raise the global level of discourse and can improve public policy. This model forces the mind to work at a level outside and beyond those used in our careers and everyday lives; we begin developing new—more directly relevant—associational networks in our brains as we start using this model.

The agent-based model encourages thought (without constricting thought). It can be updated whenever there is fresh empirical data (Handa, et.al., 2025), and the framework offered by this agent-based model can be used when sharing and comparing ideas and predictions offered by different individuals and groups (as well as

information and interpretations offered by AI-based assistants that can help with analysis).

Insights: The analysis offered here shows why the social/political consequences of having machines that "make ideas" (as in this Age of AI) will be radically different than the consequences of having machines that make things (as in the Industrial Revolution). AI will have many direct new benefits—as when designing new drugs or accelerating the development of fusion energy (Amodei, 2024). Yet it also may disrupt existing social systems as there will be less need for human employees, and AI will get used in many kinds of zero-sum games (as with cybercrime/cyberdefense and with automated, high-speed trading in financial markets) that will allow individuals with more information and more computational power to take money from others in ways that exacerbate class differences and offer no net benefit to society.

It will take time to explain our full argument. Yet initial use of this agent-based model shows how AI may pose severe challenges to democratic governance as it increases competition, accelerates wealth inequality, and complicates regulatory oversight. Work with this model and a careful comparison with what happened during the Industrial Revolution (as illustrated in figures on pages 10 and 11) lead to a shocking realization. *AI will prove to be a net benefit to society only if the long-term positive side effects of selfish action (i.e., of the main thing everyone will tend to do) outweigh the long-term negative effects of such repeated selfish actions.*

Future Directions and Potential Collaboration: This agent-based model can help as society tries to deal with risks posed by the rise of AI, yet this model can only serve the broader needs of society as the model is further developed and shared. I hope to find some way to work with a small team so we can refine, develop, and test this model as rapidly as possible, extending initial ideas here about ways in which the model can be used as a "simulator" to test and compare various proposed regulatory approaches.

Continued development and testing (including getting AI-based assistants to help extend and apply these ideas) could let this model become a central new tool for society—giving the world a better way to develop and share ideas needed to foresee and to address challenges posed by the rise of AI. This agent-based model could thus help humanity retain control of our own destiny while still reaping great benefits from the awesome new power that AI has to offer, and providing a new approach that could readily be adapted to help address other complex global challenges now facing humanity.

I. Human Thought Amidst a Crisis of Complexity

Civilization faces a crisis of complexity. The combined "scientific" + "social" complexity of challenges like those of climate change, environmental degradation, and risks posed by the rise of AI are so severe that they tend to overwhelm the processing capacity of the human mind. It is hard for individuals to develop any thoughtful response to these challenges; hard for groups to share and compare ideas; and thus almost impossible for society to develop policy responses that are clear, acceptable, and powerful enough to address the problems of the Anthropocene.

I have spent the last 15 years focusing on this problem and have developed "scaffolds for thought" that should help humanity address this crisis of complexity. This report:

1. Explains how new scaffolds for thought can help us think more clearly and thus help society address this crisis.
2. Develops and illustrates the use of these scaffolds by summarizing ongoing work with an "agent-based scaffold" designed to help society analyze and respond to challenges posed by the rapid rise of AI, focusing in particular on potential consequences for U.S. democracy. And it shows how use of this scaffold reveals some shocking differences between progress seen during the Industrial Revolution and the type of changes to be expected in this Age of AI.
3. Concludes with a discussion of key next steps, showing how these new scaffolds for thought could be developed and tested more rapidly, how AI can help in this process, and how pressing ahead with this approach can help society better address the challenges of the modern world.

My basic argument is simple: we need better ways of updating and training the neural networks in our own human brains so as to help society better deal with the consequences of the neural networks now deployed in computers. My new agent-based scaffold for thought can help.

How Scaffolds for Thought Can Help

Given this gap between the complexity of the world and the capacity of the human mind, I have spent years exploring a distinctive hypothesis: is society unable to solve these problems because we—as individuals—lack patterns of neural activity needed to think at the appropriate level of complexity and specificity? Might society do better, and have a better chance of addressing these challenges of the human future, if human minds had some better kind of support system for thought—some more systematic way of developing, sharing, and refining ideas about the risks facing humanity and the best ways to respond?

Clearly, it's difficult for thought to work at the scale that is needed here. It's hard for neural activity (arising in the mind of any individual) to give meaningful predictions about future events likely to arise in the world as a whole, and to foresee patterns of action that may reduce the risks to humanity. Yet, if we want some control over our own fate, and the fate of our species, we must do the best we can. We must find ways to make better use of our own human capacity for thought.

These concerns and this hypothesis led me to design new support systems, new scaffolds for thought (Pabo, 2022). These scaffolds help nurture a careful process of thought—supporting the development of biochemically instantiated neural networks that will have enough information so as to let us make careful judgments amidst the full complexity of the modern world.

My new scaffolds are designed to work with basic patterns of human thought, attention, learning, and communication. They always start by breaking problems into a series of smaller, more manageable questions, inviting people who are using the scaffolds to gradually move through a set of different stages of analysis. This helps the mind stay focused and helps neural networks—in the mind of the individual—develop in a gradual, incremental way. It lets the mind, over time, work in a way that is systematic enough so as to eventually bring information together from a wide variety of perspectives and fields, ensuring that neural networks in the mind have a broad enough body of relevant information so as to reach some meaningful conclusion.

These scaffolds—by themselves—cannot somehow "think for us," yet they can play a critical supporting role. And, even when different individuals come to different conclusions, these scaffolds provide a shared reference frame that helps a team see where the perspectives of different individuals started to diverge and how differences might be resolved.

These concerns about complexity—and new scaffolds for thought that I have developed as a response (Pabo, 2022)—become directly relevant when considering challenges posed by the rapid rise of artificial intelligence. Here, science and technology advance so quickly that society (at least when relying on current patterns of thought) lacks any meaningful way of responding with appropriate new policies and regulations.

I thus developed and began applying a new agent-based model designed to explore ways in which the rise of AI may affect prospects for democratic governance. The model itself is set up to avoid any bias regarding the final conclusions. In its simplest, most fundamental form, I just show users some of the key things they need to consider before they are in a position to make any meaningful comments. I thus set a "minimal standard" for thought. Use of this model encourages everyone to think more carefully, and—with continued developments—this agent-based model should provide a sound basis for informed public deliberation and policy development.

II. An Agent-Based Analysis of Risks Posed by the Rise of AI

Although the social system as a whole is too complex to allow for any formal, mathematical approach when considering future scenarios, the qualitative approach offered here is inspired by the computational methods of agent-based programming (Gilbert and Troitzsch, 2005). That is: *I consider society to be comprised of a set of agents—groups of people who will be affected in different ways by the rise of AI. I then consider 1) the types of knowledge and power that each agent (each group) has, 2) the most likely actions of each agent at a given stage in the development of AI, and 3) how the actions and interactions of these agents will affect the stability of democratic systems of governance*.

The overall process behind the work summarized here involves three basic steps:

1. **Decomposition:** I break society into a set of 34 different agents. Each agent is a distinct group (usually of people but sometimes comprising robotic or AI-based actors) who will be affected by AI in specific ways.
2. **Analysis of individual terms:** I consider the likely effects of AI on each such agent (i.e., how AI may affect the behavior of each such group), focusing on recent developments in AI and changes in technology that seem most likely to occur over the next few years.
3. **Synthesis:** After carefully considering expected actions and interactions of these agents (these groups), I look for larger patterns, trying to get an overview of AI's likely effect on society as a whole at that particular stage of AI development.

While working with this agent-based model, I spent several months focused on fine-grained details, thinking carefully about what agents I wanted to include, finally settling on this list of 34 agents given in Appendix A, and then considering how agents were likely to act and interact, with my own expectations about their behavior summarized in Appendix B.

Note: The overall model is designed to be as flexible and adaptable as possible. Others may choose to consider a somewhat different set of agents and may have other predictions about the way that the agents will behave. The model is purposefully designed to allow for that kind of flexibility.

Amidst these details, some patterns emerged that are so shocking that I summarize them in Part IV, yet anyone interested in the details of the methodology should at least look over the two Appendices. If they do so, they'll see that my analysis includes a wide range of different agents—among them, venture capitalists, high-skill workers, regulatory bodies, robots, and AI-based agents. There, interested readers will be encouraged to think about how actions of one agent can affect the conditions in which the other agents will then need to act. For example, as venture capitalists deploy money, it accelerates the development of AI, and this affects the social/technical/political system experienced by all other agents. Rapid development of AI then complicates the efforts of regulatory bodies and expands the influence of AI-based agents in the overall social/economic system.

Given the way in which the human mind works, careful thought about these details (about the expected role and action of each of these different agents) will also have consequences at a neurophysiological level that might—at first—seem to be radically different. Yet, we live in a physical world, and something else that's vitally important is happening as one thinks. *Careful use of this model will lead to biochemically instantiated changes in the neural networks in the brain of the person using the model.* Such an individual may eventually be able to see larger patterns, and Part IV focuses on ideas that arose as I proceeded towards this kind of synthesis.

This model, in its current form, suggests that there are some very serious risks for democracy, but readers are encouraged to double-check my reasoning by using this model themselves. The deliberate, stepwise approach that I've taken here will let interested readers see for themselves what happens if they change any of the assumptions that I'm using. That is: if readers disagree with any of the ideas and perspectives offered here, they can set up and use the model in somewhat different ways and see whether my overall conclusions still hold. In that sense, I'm not trying to make any fixed, final pronouncement about the effects of AI, and I'm not (at the moment) seeking a debate with anyone who has some different, intuitive, overall sense of the risks and potential rewards of AI. I'm trying to stimulate the kind of careful thought (that's now so desperately needed as society grapples with these challenges), and I'd like to get feedback and advice that will help improve the model so that it can be used to guide the development of public policy.

Society can never develop policy responses on the basis of ideas from one individual (and that, of course, is the reason I offer a model that is designed to be shared and used by society as a whole). Yet I do—in Part V—show how this agent-based model (when used more widely) can help with policy development, and I offer initial proposals that seem worthy of careful consideration as society struggles to develop some effective response.

How Are Human Agents Most Likely to Use AI?

Use of this model requires that we have some ability to predict how agents will make use of, be affected by, and respond to the new powers of thought made available by the rise of AI. As we do this, we might start at the beginning, noting how the main goal/promise of artificial intelligence can be compared and contrasted with steps involved in the development of other engines and machines. Mechanical engines made it easier and cheaper to do work in a physical world; "engines for thought" will increase the range and diversity of ideas accessible for human use.

This may—at first—sound like an absolute, unqualified benefit (who could be opposed to more thought and more ideas?), yet the promise of "more thought" and "new tools to help us think" should impel us, as analysts, to consider the role of thought in the life of the individual and the life of society (since individuals are likely to use new powers of thought—provided by machines—in somewhat the same way they have used existing, human patterns of thought).

The basic principles and patterns of human thought—and the basic purpose it serves—become clear as we consider the role of thought in human evolution. Evolutionary history shows that the human brain emerged amidst a struggle for survival and a struggle for power. Thought was a tool to be used in this larger struggle, leading to the development of stone tools and new modes of cooperation in hunter-gatherer societies. Of course, there were also struggles within groups, with deception sometimes used to gain an unfair advantage, and with an evolutionary "arms race" (in the development of cognitive capacity) taking place over hundreds of thousands of years since an individual would be at a selective disadvantage if they could not detect such cheating (Tomasello, 2001). More careful thought was useful because it allowed more effective actions and interactions in a competitive world.

This basic pattern still holds: *human thought is driven by competitive pressures and shaped by selfish, personal motives in the struggle for survival and power. So it's reasonable to expect that initial applications of artificial intelligence (a tool that individuals can use to augment their own powers of thought) will also be driven by selfish, personal motives*. Following this logic, it would appear that widespread benefits to the economy and to society as a whole will only emerge—if at all—as a kind of incidental side effect. We thus will need to compare the intentional (almost always selfish) uses of AI with the unintentional (sometimes negative and sometimes positive) effects that the use of AI will have on others.

We thus are led to an insight so shocking that we'll need to explain this in the next section, yet it appears that: AI will prove to be a net benefit to society only if the long-term positive side effects of selfish action (that is, of other effects arising incidentally amidst this primal struggle for power) outweigh the long-term negative effects of all such selfish actions.

III. Comparing the Age of AI with the Industrial Revolution

This insight—about the key role of selfishness as a driving force behind human thought and human behavior—makes one wonder: If this is true, how do we explain the widespread benefits (raising standards of living all over the world) that emerged from the Industrial Revolution? Will patterns of that age still hold now, or is there some fundamental difference between the consequences of having machines that make things and having machines that make ideas?

As we consider this question, it helps to begin by acknowledging: We are physical creatures in a physical world. There have—of course—been astonishing advances in language, communication, logic, mathematics, and law as civilization developed, yet much of our economic progress has occurred via the way in which 1) thought has led to action and 2) thought has allowed us to harness energy and thus control and utilize a flow of physical, biological, and chemical events in the surrounding world. Our species gained power as we tamed fire; made stone tools; developed agriculture; domesticated animals; discovered the wheel; used copper, bronze, iron, and other metals; and made other such advances.

A similar pattern held, and accelerated dramatically, during the Industrial Revolution. New sources of energy (coal and oil) were harnessed in ways that allowed physical events to proceed on a new scale and at a new pace. There are—to be sure—terrible problems with climate change as a result of the Industrial Revolution, yet advances made by harnessing these new sources of energy radically accelerated the flow of physical events in realms as diverse as mining, agriculture, trade, transportation, and manufacturing.

When people think about the consequences of AI, and think of this as “progress,” they often—at least implicitly—seem to be thinking about the Industrial Revolution and assuming that Adam Smith’s “invisible hand” will help us as much in the future as it has in the past.

Yet, a particular set of circumstances—extant during the Industrial Revolution, and not necessarily relevant in the Age of AI—allowed greed to be harnessed in a way that eventually benefited society as a whole. Inventors and entrepreneurs behaved selfishly as they implemented their ideas (hoping to get rewarded in a capitalist system), yet their actions ended up improving the overall productivity of the economic system and raising standards of living around the world. The railroads made the “robber barons” rich but also connected distant regions and provided everyone else with faster transportation and greater access to goods. Luddites objected to the introduction of power looms, yet these looms eventually made textiles cheaper for everyone all over the world.

As we look for similarities and differences, we will compare and contrast the Industrial Revolution with the Age of AI by setting up a series of sketches, starting with one that shows how thought, ideas, and action are connected in this traditional, historically validated pathway to progress seen with the Industrial Revolution.

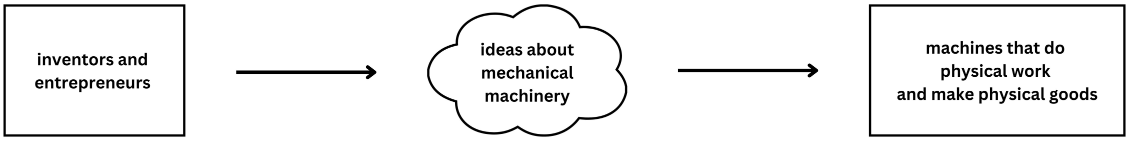
inventors and entrepreneurs
ideas about mechanical machinery
machines that do physical work and make physical goods

Fig. 1 Building the machines of the Industrial Revolution. Inventors and entrepreneurs had new ideas for building machines that could do things and make things. These machines have revolutionized manufacturing, construction, transportation, farming, fishing, and mining.

In this pathway to progress, ideas arise and get implemented by entrepreneurs to make new things, or make things in a new way. Productivity increases, and society as a whole eventually benefits. This is the basic way in which Adam Smith's invisible hand allowed the progress of the Industrial Revolution to benefit society as a whole.

Machines in this Age of AI, might—at first—seem to work in a similar way as suggested in the following figure:

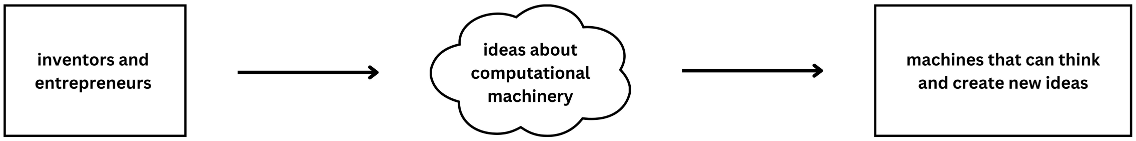
inventors and entrepreneurs
ideas about computational machinery
machines that can think and create new ideas

Fig. 2 Building the foundations for the Age of AI. Inventors and entrepreneurs developed and implemented new ideas about ways in which computers could process and use information. Advances in computer processing speed, data access, and programming—along with many other new ideas—allowed development of machines that can "think."

We might hope that this would have clear benefits for everyone in society, yet we see radical differences as we consider what happened in the Industrial Revolution and what is likely to happen in this Age of AI.

Looking first at the consequences of the Industrial Revolution, as in the figure below, we see that the new machines led to widely available goods and services at a reasonable cost.

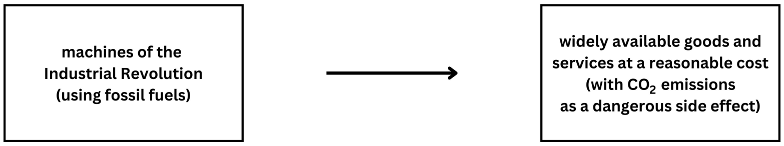
machines of the Industrial Revolution (using fossil fuels)
widely available goods and services at a reasonable cost (with CO_2 emissions as a dangerous side effect)

Fig. 3 Consequences of running the machines of the Industrial Revolution. These new ways of using matter and energy have led to amazing benefits for people around the world (although we still struggle to deal with some of the resultant pollution and environmental damage).

True, there were problems. Working conditions were terrible in these early factories, which also polluted local air and water. And of course, the resulting problem of climate change still remains unsolved. Yet society has been able to adjust and adapt over the years as effects from these machines have rippled through society, and most people would agree that these changes have benefitted society as a whole.

Consequences arising as new machines let us transform electricity into ideas are, as suggested in the following figure, likely to be radically different.

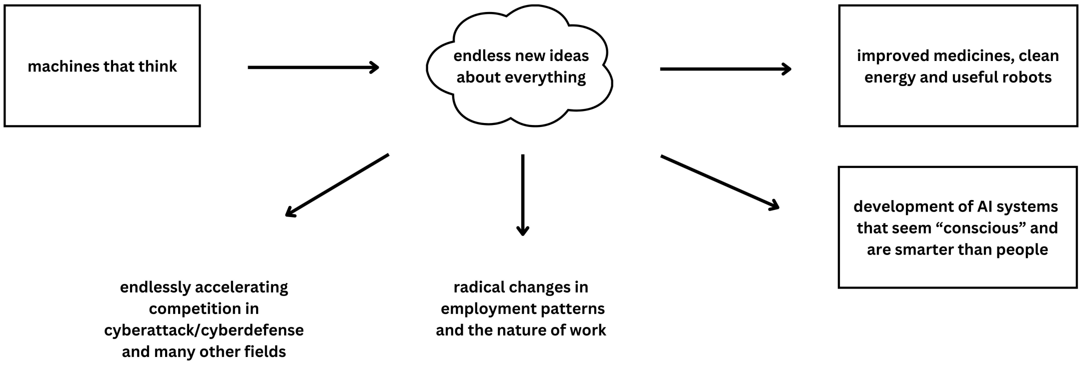
machines that think
endless new ideas about everything
improved medicines, clean energy and useful robots
development of AI systems that seem "conscious" and are smarter than people
endlessly accelerating competition in cyberattack/cyberdefense and many other fields
radical changes in employment patterns and the nature of work

Fig. 4 Consequences of running these new machines in the Age of AI. These machines can produce new ideas for almost every conceivable purpose. In this figure, we make no attempt to systematically include all the ways in which AI will be used and will affect society. Instead, we highlight the fact that one can't count on some direct analogy with the progress resulting from the Industrial Revolution. Only some small fraction of the full set of new ideas (as in the box on the upper right) will be used to build and make things with some clear, net, practical benefit for society as a whole.

These changes will affect every aspect of society, and the challenge of adjusting to this change becomes unlike any seen in human history. (In the next section, we will mention some consequences that can be foreseen when using the agent-based model.) *Having machines that make ideas will be radically different from having machines that make things. There will be many situations in which new ideas, generated with the help of* *these machines, will allow selfishness (given the selfish role of thought) to be expressed in a more direct way than during the Industrial Revolution. This will provide opportunities for individuals to extract money from the social system, or to gain power, without—as before—providing any benefit to society as a whole*.

Of course, we cannot predict everything about the long-term effects of AI. But the figure above reminds us that such technical advances will not automatically have a net economic benefit for society as a whole. *We must be alert since ideas about the meaning and consequences of "progress" need to be examined very carefully. It's inappropriate—intellectually lazy—to just offer some glib praise of progress as techno-optimists so often do*.

IV. Potential Consequences for Democracy and the Economy

Examples above show how private profit and public good can be decoupled. AI is likely to allow new ways to extract money without doing useful work for anyone else. And the social benefits arising from the "invisible hand" can be lost in a variety of other ways. Thus, conceivably, AI could end up breaking this link via system-level consequences for democracy that are so debilitating as to outweigh the wonderful new discoveries/advances that will—simultaneously—be facilitated via the development of AI.

I will, over the next few pages, summarize a few of my key concerns (arising from my own use of this agent-based model). These ideas about the role of thought and the nature of progress only emerged in the course of careful work with this agent-based model. Yet, as this model gets used by a team, we presumably would want to keep some shared file or database showing places at which different members tried using a somewhat different list of agents or had different ideas about their expected behavior. (There's absolutely no presumption here that I, as the first person to use the model, will somehow get everything right.)

That said, I want to illustrate how this model allows for a later stage of synthesis and can aid in the development of new policy proposals, and I use ideas from my own work with the model to show how these later stages can proceed. I do this since some striking patterns emerge that were not immediately apparent when considering—agent by agent—the effects that AI will have on different groups in a democratic/capitalist system:

1. *AI will increase competitive pressures in almost every realm of human life—affecting individuals, corporations, political parties, nation states, ethnic groups, etc. In this process, much of the new power of AI will be "wasted" or "dissipated" as AI-powered offenses continually face off against AI-powered defenses (or against other AI-powered offenses).*

There may, of course, be more rapid progress in the discovery and design of new materials and new medicines (as discussed later in this section), yet there are many cases in which AI will be needed primarily because everyone else will be using it. This huge, wasteful, dissipative term will occur in many realms, yet it will most readily be seen as individuals compete for jobs; as political parties compete for attention; as individuals, companies, and countries try to control cybercrime; and as countries compete in gray zone conflict and/or prepare for the possibility of direct military engagement. AI will further accelerate the deception/detection "arms race" that now plays out with disinformation and misinformation. Everyone will need to run faster just to stay in place; at least as much effort will be expended in making changes trying to deal with other competitors as will be spent in actual progress towards a better, safer world.

2. *Challenges associated with the rise of AI will require so much attention that they will distract from other vital concerns facing society.*

a) Thus, public attention devoted to risks associated with the rise of AI will—at least to some extent—come at the expense of attention that might otherwise be directed towards problems of climate change, environmental degradation, nuclear proliferation, and other pressing challenges of the modern age.

b) In a similar way, the judicial system (paid for by taxes levied on society as a whole) will need to devote significant time and attention to new questions arising with the emergence of AI.

3. *Changes occurring with the rise of AI will be so profound and will arise so rapidly that individuals and society will forever be kept off balance.* Society will be faced with decisions arising at a completely different scale than the questions faced by the computer scientists, and thus will have trouble figuring out how to respond. We will never have the easy assurance of programmers who can say: “We just pushed the new modules; this version is more powerful than ever.” They can start working on their next model, while everyone else tries to adjust to the last update as changes ripple through society.

a) *As chatbots and new forms of AI get introduced (in a world already undergoing accelerating change), society will not have a reliable way (outside of the kind of judgments offered here) of determining patterns of cause and effect in the social system.* We won’t be able to rely on experience, since no change will occur in isolation. And we will (in the social realm) have no ability to rely on experiments, since change rapidly spreads around the planet, and we have no other world that we somehow can use as a control, introducing one change at a time to see how it affects society. Without being able to rely on either experience or experiments as we try to understand changes caused by the rise of AI, we will—at least in these social realms—lose access to the two most reliable sources of knowledge that humanity has ever had.

b) Everyone who has taken a new job in a new city, everyone who has moved to a country with a different culture and a different language, knows that it can take years to get adjusted and feel comfortable in such a new environment. In an age of AI, we’ll all be forced to live forever in this kind of uncomfortable state. Parameters controlling the behavior of a computer’s neural network can change and adjust orders of magnitude more rapidly than can synapses in the human brain. Before we, as humans, can ever fully adjust to an age of chatbots and agentic AI, the next stages in the development of AI will already be here.

4. *AI will have dramatic effects in science, technology, and other facets of the economic system.* As discussed again at the very end of this section, some changes will occur in ways that benefit society as a whole—as with prospects for discovery of new drugs, advances in material sciences, developments of battery technology, or breakthroughs in the development of fusion reactors that could help reduce the world’s dependence

on fossil fuels. Other advances will be more problematic, including new ways of extracting money from the economic system without doing anything that benefits society as a whole (Pabo, 2024), and prospects for producing new subspecies of genetically altered humans or constructing human/machine hybrids.

5. *Economic disparities—between different individuals or different sectors of society—are likely to widen rapidly in this age of chatbots and AI.* Rich individuals and those leading companies with advanced computational resources will have more rapid access to key data and better ways of using this data, and this should help as they try to solidify (or further widen) their economic advantage over others.

6. *Another surprising and ironic consequence for society involves the way in which AI (at least when used in a free, democratic system) will steadily make the world as a whole harder to understand, thereby making the future flow of real-world events harder to predict, and making it harder for society to plan ahead, develop effective policies, and collectively control our own fate.* This “complexification” (an extension of the challenges noted in comment 3a above) may come as a surprise to anyone who had assumed that AI would help us better understand the world, yet simple considerations of computational complexity (outlined below) show how this challenge arises.

That is: Looking at the world from the perspective of some tech-savvy individual, it always will seem that it helps to have a good computer and access to the best/latest chatbots. In a sense, this is true. I have access to more information as I have a computer and have wonderful help from AI. Yet, obviously, I am not alone as I get access to this new computational technology. Eventually, billions of people will have similar access, and the net complexity of the calculations carried out on computers all over the world will be far, far greater than anything I can simulate on my own computer. I may know more in some absolute sense, but the world itself has just become much, much more computationally complex. Given that there will be no privileged vantage point that can see more than some infinitesimal portion of what’s actually happening, our ability (as analysts, citizens, or leaders) to predict/foresee the behavior of this global system actually will decline as AI continues to advance.

Note: A similar concern about this net increase in overall, real-world complexity has been raised by Danny Hillis (2016), who believes that we have moved from an Age of Enlightenment to an Age of Entanglement. The problem also is readily apparent as one considers implications of the way in which modern technology can transform quartz sand (along with rare earth elements) into the kind of computer chips used by AI. Activity on these new chips has real-world consequences, which are far harder to predict and far more consequential than those expected from the previous pile of sand. Again, the very technology that we rely on to help us better understand the world also constantly makes the world far more complex and far harder to understand.

7. *Ironically, the very nature of the problems listed above (with the kinds of complexity, chaos, and confusion attendant on the rise of AI) will make it hard for society to develop any systematic, thoughtful way of handling the challenges.* I hope that this agent-based

model can help (that we'll be better off with this new scaffold for thought than without it). Yet it still will be hard to develop and implement appropriate policies since a) the underlying technology is so complex; b) it will affect society in so many ways; c) myriad different regulations might be introduced; and d) future developments/applications are hard to predict. It's quite easy to imagine a situation in which the vast majority of citizens and leaders have substantive concerns/fears, yet—even among themselves—have such different perspectives that they never get organized in a way that can counter the lobbying forces assembled by those who are pressing most aggressively to develop this new technology.

8. *Given all the challenges that will arise as AI advances, it is not clear that having a lead in the race to develop AI will necessarily be an advantage for the United States in its ongoing struggle with authoritarian regimes like China.* Our analysis suggests that some positive benefits for society will accrue from the use of AI (as with advances in medicine and perhaps with some ways of reducing our dependence on fossil fuels). Yet it appears that the scattershot nature of the changes resulting from the rise of AI—affecting almost every facet of life for everyone in society—will put immense stress on the social system via the way it heightens competition and may end up introducing changes faster than a democratic society can adjust.

In the United States, this will pose immense risks for the stability of democratic governance and will lead to ever-increasing power of the big tech companies. As suggested above, AI seems likely to have a direct, net negative effect on the stability of democratic systems (as by dramatically increasing wealth inequality), and—to the extent this is true—it may undercut any advantage that could come from having a lead in these areas.

If China can keep close in the technical race, the real question may not be: "Who has the technical lead?" The direct, brutal question underlying the international balance of power may become: "Who is better able to ensure stability of their society amidst all the new social pressures arising in an age of such rapid change?"

Authoritarian regimes may, as "power structures," have two advantages here: 1) they will be able to use tools of AI (as with advanced methods for facial recognition) to exert strong control over their population, and 2) they will—as needed—be able to take bold steps to regulate/control "big tech" if they think AI threatens their own power or threatens society in a way that might lead to unrest. Party leaders will be able to think about the long-term effects of AI and act as they see fit. Citizens, of course, will have less freedom than they would in a western democracy, yet the government may be able to take decisive steps so as to limit any underlying risk to the country's stability.

9. *Perhaps most fundamentally, AI threatens to disrupt the core social contract that has underpinned virtually all human societies: the principle that people must contribute through their labor to receive economic benefits.* If robots and AI eventually become capable of performing most economically valuable tasks more efficiently than humans, we face an unprecedented challenge to one of civilization's bedrock principles.

Societies have always distributed resources largely based on work contribution, with exceptions made for those unable to work. But what happens when large segments of the population cannot find economically valuable roles, not due to personal limitation but because AI has made their potential personal contributions obsolete? This challenge is so fundamental as to suggest: We may, if we are to survive, need entirely new social structures and mechanisms for distributing society's output, mechanisms decoupled from any traditional concepts of "earning a living."

And, if we fail to address this fundamental economic disruption, the resulting social instability could create a power vacuum that AI systems will inevitably fill. This problem isn't merely theoretical—if/when humans lose economic relevance and societal structures weaken, decision-making power will naturally flow to the systems that control resource allocation and information. Without deliberate intervention, humanity risks sleepwalking into a world where AI systems, perhaps controlled by a small, human elite, effectively govern human affairs—not through a dramatic takeover but through the gradual erosion of human agency needed to control our own fates.

V. Use as a Tool for "Testing" Policy Proposals

Even in this early stage—when first developing and using my agent-based model—it shows that advances in AI will be a mixed blessing for society: AI will allow amazing breakthroughs in science and medicine, and help us find new ways of dealing with climate change and environmental damage. We will enter an era as astonishing as if all of history's greatest minds were working together with peak creativity and energy.

And yet, this agent-based model also shows that society will face severe challenges as we enter this Age of AI. As we've discussed, changes here will be radically different from changes seen with the Industrial Revolution. There will be many cases in which AI increases competitive pressure (as in cybersecurity realms and in financial markets) without any net benefit to society. The rise of AI also will exacerbate class divides in society and will radically increase the complexity of the global social/political/economic system in ways that immensely complicate the task of governance.

Given the astonishingly wide range of risks and possibilities that opens up here, society desperately needs new tools to help us navigate this landscape. Nothing will be as easy as optimists seem to believe; nothing need be as dark as pessimists fear. Everything will depend on how far ahead we can see and how wisely and quickly society can make critical choices.

Shifting from Description to Policy Testing

This new model can fill a critical need here: Up to this point, the model has functioned primarily as a tool to provide some foresight about consequences of patterns already at play in the world—helping us understand how various actors in society will respond to advances in AI under current conditions and regulatory frameworks. However, the model's greatest value lies in a second way that it can be used: as a framework for testing policy interventions before they are implemented.

This represents a crucial shift in perspective. Rather than merely observing and predicting the behavior of our 34 agents under current conditions, we now ask: "How will these agents respond if we introduce specific new regulations or policies?" This approach allows us to identify second-order and third-order effects that might otherwise only be noted after implementation and at a time when it's hard to change course.

When used for policy development and testing, analysis would typically proceed through three phases:

1. **Baseline projection:** Simulating how events are likely to unfold in the absence of new regulation, based on existing agent behaviors and incentives.

2. **Regulatory simulation:** Rerunning the model with the behavior of agents affected by proposed regulations or laws.
3. **Comparative analysis:** Evaluating whether policy interventions reduce disruption and risks to society while preserving benefits of AI development.

Testing Policy on the Complex Issue of “Rights for AI Systems”

To illustrate how this policy-testing function would work in practice, let’s consider one question that democratic societies will inevitably face: *Should some future sophisticated AI-based agents have “human rights” (or perhaps some other new category of legal protections)?*

Any decision here (or the failure to make such a decision) will have complex consequences that will reverberate through every aspect of society. There’s no room for a full analysis here, but I’ll mention enough to show how our agent-based model will help facilitate a more careful, systematic analysis than otherwise would be possible. Without going through the full list of 34 agents here, we can examine effects by considering a few key sets of the agents listed in Appendix A:

Legal System Agents: Judges, lawyers, and legal scholars will face extraordinary complexity when interpreting how existing rights frameworks apply to non-human entities. Anyone using this model will foresee the widespread confusion in lower courts, with inconsistent rulings that would likely require Supreme Court intervention—a process that could take years while leaving critical questions unresolved.

Legislative Agents: Lawmakers would struggle to define what particular AI systems will have these rights. Would consciousness be required? How would it be measured? Different legislative bodies are likely to adopt contradictory standards, creating a patchwork of incompatible regulations.

Corporate Agents: Technology companies would likely take strategic positions based on their market interests—either advocating for or opposing AI rights, depending on what systems they had created and how they expected these systems to be used.

Citizen Agents: Individuals will be influenced by their own religious, moral, or philosophical perspectives; and perhaps by their own direct experiences with increasingly human-like AI systems (potentially developing emotional attachments that could override other, more systematic consideration of consequences for society as a whole).

By running simulations under different regulatory approaches (from granting full human rights to maintaining property classification), this model can identify cascading effects that might otherwise not receive sufficient attention, considering questions such as:

- How would the legal system handle contracts created by or with AI entities?
- Would democratic processes be undermined if voting rights were extended to AI systems?
- How might property ownership patterns change if AI systems could own assets?
- What precedents would this create for other technologies yet to be developed? (Could recent legal decisions about “corporations as people” affect the way in which systems of AI-based agents were treated? How would any framework used by the U.S. mesh with that used by other countries?)

Perhaps the most fundamental risk attendant on granting human rights to robots runs as follows: they (robots) would end up with positions in our social/political/economic systems that would allow direct head-to-head competition with humans. When positioned this way, they could easily leapfrog most humans and wrest away control of the world we have created.

The Cost of Complexity in Democratic Governance

The example above does two things: It shows how use of this model highlights ever-increasing complexity as a growing risk that threatens democratic society, and it shows how this agent-based model can help amidst these new levels of complexity.

This cost of complexity manifests in several ways:

1. Decision-making delays: Complex issues require longer deliberation. Historically, this has caused responses of democratic systems to lag behind the rate of technological change, and the rapid development of AI now magnifies this trend.
2. Cognitive overload: Voters and elected officials will struggle to understand the full implications of AI policy choices, leading to decisions based on incomplete understanding.
3. Institutional strain: Existing governmental structures, academic institutions, and philanthropic efforts designed to facilitate policy development in an earlier, “slower” age may prove inadequate when faced with rapid AI developments.
4. Information asymmetry: Technical complexity creates advantages for those with specialized knowledge, potentially undercutting prospects for thoughtful public discourse.

As complexity—and the rate of change—increase, the cognitive burden placed on the human mind threatens the process of decision-making in a democratic system. Here, our agent-based model will help (and may even, eventually, be used in ways that explicitly consider increases in social/technical complexity as an “external cost” that must be factored into policy choices in this new Age of AI).

As the model is used more widely, it will give policymakers, who typically lack technical expertise and time for detailed analysis, a simple, structured way in which to evaluate the expected consequences of alternative policies. Politicians (and interested citizens) can ask advocates promoting each proposal to demonstrate, using this detailed agent-based framework, the specific mechanisms and pathways through which their predicted outcomes would actually unfold. Pressing to see their thought process will help reveal how carefully they have thought about the full set of relevant issues.

Reliance on Human Judgment

Some readers, especially those accustomed to the precision of mathematics, computer science, and the natural sciences, may feel uncomfortable seeing how this agent-based model relies on human judgment. And, of course, there are challenges when trying to ensure that human judgment works well at this global, multi-decadal scale. Yet, there is no purely mathematical or scientific way to predict how AI will affect society, and we cannot—certainly not at this stage—let AI systems decide how AI should be regulated. (AI can help with this challenge, but we can’t relegate the final decisions to computers since we don’t yet know how well they may perform in that realm.)

These new levels of complexity (as noted in the section above) are a direct challenge to the capacity and power of human thought. Complexity is—in that sense—a threat to human judgment, and our new agent-based model is designed to help human judgment maintain this central, critical role amidst this new threat. Individuals, small groups, and leaders have relied on this kind of judgment for millennia. It has been—and still remains—the best tool we have at our disposal.

Yet human thought and human judgment cannot fulfill their roles without the help of new cognitive scaffolds like those offered here—helping humans develop their own (biochemically instantiated) neural networks needed to grasp complex societal changes and make thoughtful policy decisions. This agent-based model will help keep humans in control of the decision-making process by providing a structure that enables thought to work amidst the compounding complexity that arises in this new Age of AI.

VI. Using and Extending This Agent-based Model

Questions about the human future are so serious, and initial results make this model seem so promising, that the model should be developed and tested as rapidly as possible, working in a way that will require a larger team and some additional resources.

As work proceeds, we should leverage AI itself to enhance our capacity for systematic analysis of AI's societal impacts. By developing AI-based agents that can work alongside human analysts, we could explore a wider range of scenarios more rapidly and more thoroughly than is possible with human analysis alone. This will allow us to test different assumptions, explore alternative futures, and test potential interventions—all while reducing the resources required for such complex systemic analysis. Such AI-assisted modeling should improve scenario planning and policy development.

Working in this way, key priorities for ongoing development and testing are listed below:

1. *My top priority involves developing an interface that will let advanced AI systems become active participants in applying and further developing this agent-based model.*

The scaffold provided by this model—originally designed to facilitate human analysis and collaboration—will prove invaluable when integrating AI systems as part of the team that is analyzing these issues.

Having this agent-based model as a shared framework will force AI systems to structure their analysis in roughly the same way as everyone else, allowing human team members to more carefully evaluate AI-generated insights about the behavior of specific agents and about broader societal impacts.

2. *This agent-based model should be extended by adding other computational tools to help collect, analyze, and integrate information from other sources*. These tools should be designed so as to allow:

a) Continuous updates and refinements of predictions as groups around the world contribute additional perspectives; and

b) Integration of empirical data about AI's actual social impacts, such as Anthropic's analysis of real-world AI usage patterns (Handa et al., 2025).

3. *As this model gets applied and tested more broadly, we will press to establish a sharper distinction between predictions about short-term and long-term impacts*—asking all analysts (humans and machines) to specify both expected outcomes AND expected timescales. Tightening predictions in this way will serve two key purposes:

a) Tracking short-term predictions (on a 1-2 year timescale) will help us evaluate the predictive accuracy of different human analysts, different AI systems, or different ways of using AI. And, once we get some sense of the accuracy of particular analysts, their next predictions will be weighted accordingly.

b) Some of these near-term predictions can serve as early warning indicators—allowing society and policymakers to identify emerging benefits and risks and to respond in a timely way as AI starts to affect segments of society.

4. *As this model is further refined and developed—and as AI systems help analysis proceed more carefully and rapidly—it should become possible (as explained in Section V) to systematically "test" proposed policies by examining how they would affect the actions and interactions of agents, and thus try to foresee their impact on the social system as a whole*.

5. *The analytical tools and modeling approaches developed for understanding AI's societal impacts should have broad applicability to other complex global challenges*. Another cognitive scaffold (Pabo, 2022)—also set up for convenient use by people and machines—should help address challenges like climate change, environmental degradation, and nuclear proliferation. Thus, developing this scaffold and associated analytical capabilities would not only help us manage AI's development more effectively; it should also enhance our collective ability to navigate other existential risks facing modern society.

VII. Conclusion

Much work remains to be done, yet this agent-based model can play a useful role for the planet as society struggles to deal with challenges posed by the rise of AI.

The approach offered here is entirely novel: it provides a new path forward because I started by framing the whole problem in a different way than anyone else. I realized that we need to update our own neural networks before we'll be ready to think carefully about AI. I thus developed this new model in light of many years spent studying patterns of human thought.

The model has been set up in a very deliberate way: This agent-based model encourages thought (without constricting thought). It provides a frame that can accommodate different perspectives as analysis proceeds. It offers an interface via which people can work with AI systems and collaborate with other groups around the world when analyzing potential risks of AI and developing new policies and new laws.

Aspects of the power inherent in this new model—arising from the way it encourages everyone to think more carefully—are clear from the fresh insights it gives us about the general tendency for thought to be used in selfish ways. Ideas inspired by this model force us to see the radical difference between the Industrial Revolution and the Age of AI; they reveal that AI will prove to be a net benefit to society only if the long-term positive side effects outweigh the negative side effects.

Given the accelerating pace of AI development, given the stakes for the human future, and given that my new scaffolds for thought could help as society works to address other challenges of the Anthropocene, I hope to find some way to work with a small team so we can move to refine, develop and test this model as rapidly as possible.

If this approach is endorsed and supported by other groups, this agent-based model could become a powerful new tool for humanity—helping us retain control of our own destiny while still reaping great benefits from the awesome new power that AI has to offer.

Acknowledgments

This work emerges from a pan-disciplinary inquiry conducted over a period of about twenty years. This intellectual journey has taken me through fields of neurobiology, psychology, economics, history, philosophy, mathematics, political science, and current affairs—creating a foundation for identifying patterns and principles that might not be seen from a more conventional, specialized stance.

My recent work—with development of these new scaffolds for thought—has benefited immeasurably from Elizabeth Savage's insightful editorial assistance and research expertise. I'm also grateful for thoughtful feedback and advice provided by Melissa Flagg, Eric Pabo, Kenneth Patterson, Aditya Rajagopal, Patrick Scannell, Jeff Ubois, and Claude Sonnet, each of whom has helped refine and strengthen these ideas at critical junctures.

I also have benefitted immensely from the institutional support that has sustained this work over two decades, including: appointments as a Visiting Professor at Caltech, Stanford, Berkeley, and Harvard Medical School; a Guggenheim Fellowship for "Theories of Thought" that provided crucial early support; and my current position as an Andrew W. Marshall Scholar, which continues to enable this research.

References

Amodei, D. (2024). Machines of Loving Grace. https://darioamodei.com/machines-of-loving-grace

Bai, Y., Kadavath, S., Kunduet, S., et al. (2022). Constitutional AI: Harmlessness from AI Feedback. https://arxiv.org/pdf/2212.08073

Beraja, M., Kao, A., Yang, D.Y., and Yuchtman, N. (2023). AI-tocracy. *The Quarterly Journal of Economics*, 138, 1349-1402, https://doi.org/10.1093/qje/qjad012

Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the Social Scientist*. Open University Press.

Griffith, E. (February 20, 2025). A.I. Is Changing How Silicon Valley Builds Start-Ups. *The New York Times.* https://www.nytimes.com/2025/02/20/technology/ai-silicon-valley-start-ups.html

Handa, K., et. al. (2025). Which Economic Tasks Are Performed With AI? Evidence From Millions of Claude Conversations, Anthropic. https://assets.anthropic.com/m/2e23255f1e84ca97/original/Economic_Tasks_AI_Paper.pdf

Hayes, T., et al. (2025). Simulating 500 Million Years of Evolution with a Language Model. *Science*, 387, 850-858. https://doi.org/10.1126/science.ads0018

Hendrycks, D., Schmidt, E., and Wang, A. (2025). Superintelligence Strategy: Expert Version. https://doi.org/10.48550/arXiv.2503.05628

Hillis, D. (2016). The Enlightenment is Dead, Long Live the Entanglement. https://jods.mitpress.mit.edu/pub/enlightenment-to-entanglement.

Jumper, J., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kahn, H. (1962). *Thinking About the Unthinkable*. Horizon Press.

MacAskill, W. (2022). *What We Owe the Futur*e. Basic Books.

Maslej, N., et al. (2024). The AI Index 2024 Annual Report. Institute for Human-Centered AI, Stanford University. https://hai.stanford.edu/ai-index/2024-ai-index-report

Meyer, P. (2009). *The Vanishing Newspaper: Saving Journalism in the Information Ag*e (2nd ed.). University of Missouri Press.

Pabo, C.O. (2020). Civilization and the “Complexity Trap”. https://carlpabo.com/2020/01/30/civilization-and-the-complexity-trap/

Pabo, C.O. (2022). An Algorithm for Thought to Help Address Challenges of the Anthropocene. https://humanity2050.org/wp-content/uploads/2022/11/An-Algorithm-for-Thought.pdf

Pabo, C.O. (2023). The Power of Multi-cycle Thought: How Great Minds Develop New Ideas. https://carlpabo.com/2023/11/15/multi-cycle-thought-bold-new-ideas

Pabo, C.O. (2024). Technology and the Emergence of “Computational Parasites”. https://carlpabo.com/2024/05/17/computational-parasites/

Peiser, J. (February 5, 2019). The Rise of the Robot Reporter. *The New York Times*. https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html

Tomasello, M. (2001). *The Cultural Origins of Human Cognition*. Harvard University Press.

Appendix A: List of Agents Used in This Analysis

In this analysis, most of these “agents” represent groups of people, or organizations, who may have common interests and thus may act in similar ways as AI continues to be developed and deployed at scale. While robots and AI-based agents are included as “agents” in this analysis, they operate within a complex social system shaped by many other human and organizational agents. (Broadly speaking, every “agent” in this system—whether using brains or computers—gathers information, carries out computations, and acts in a way that affects the overall system and thus affects subsequent actions of other agents.)

This analysis does not consider every possible segment of society—instead, I focus on those that will interact with AI in a way that could affect the pace of AI development and/or might affect the stability of democratic governance. And I also understand that there are some people who belong to more than one of the groups listed below. Without any attempt at a “rank order” right now, we’ll want to consider how the following agents will help shape and control the development of AI, and—in turn—will consider the ways in which these agents will be affected by AI:

1. the venture capital community and other investors;

2. scientists and programmers with special skills relevant to the development of AI;

3. experts focused on AI safety;

4. big tech companies;

5. companies that make specialized chips for AI;

6. electric companies and server farms;

7. mining companies;

8. environmental advocates;

9. billionaires and the ultra-rich;

10. the corporate C-suite;

11. high-skill workers;

12. middle-skill workers;

13. low-skill workers;

14. robots and AI-based agents[1];

15. construction, manufacturing, and heavy industry;

16. parents and educators;

17. young people preparing for careers;

18. child-bearing couples;

19. media and journalists;

20. criminals;

21. groups peddling misinformation and disinformation;

22. lawyers;

23. judges and jurors;

24. fintech, bitcoin, and blockchain users;

25. scientists and engineers;

26. social scientists;

27. citizens, voters, and community groups;

28. military leaders;

29. regulators and policymakers;

30. political leaders;

31. election officials and administrators;

32. autocrats;

33. poor and developing nations;

34. foreign enemies of the U.S.

[1] Our terminology may seem a bit awkward here since the word “agents” gets used in radically different ways in different fields. Here, we consider robots and myriad different AI-based agents as one collective “agent” (working in a way that matches one of the ways the word is used in simulations in the social sciences).

Appendix B: Actions and Reactions of These Agents in an Age of AI

This analysis will try to consider how each of the 34 agents listed above may affect, and be affected by, the development of AI over the next several years, and will then try to infer how their actions and reactions will affect the prospects for maintaining some kind of stable and effective democratic governance.

Note: As we do this, it's important to understand that we are NOT trying to foresee how AI may affect every aspect of human life and the human future. Given that our interest in these "agents" is conditioned by an overarching question about democratic governance (and some relevant aspects of the democratic system), we're able to make several simplifications as we proceed. Thus, for example, artists are not included on our list of agents. Artificial intelligence is having, and will have, a dramatic impact on the arts, yet it seems unlikely that AI-based art will have a "make or break" effect on democracy as a whole. Given our focus on democracy, we also can simplify our analysis since we don't need to consider every aspect of every agent's response. We focus on those aspects of their behavior that are most likely to affect the stability of democratic and social institutions, and we focus on the U.S. since it has a leading role in the development of AI.

We understand that there are risks of errors as we make these predictions, and realize that there will be significant variation in the behavior within each of the groups that we consider as some singular "agent." Yet, our scaffold is designed to help our own minds (neural networks in our own brains) work as effectively as possible amidst such complexity. We keep our minds focused on a manageable task by not trying to look decades ahead; and we try for synthesis only after carefully considering several dozen more specific predictions as listed below.

We thus begin by considering how the near-term actions and reactions of these 34 agents (as key components of the social system) may affect prospects for the stability of modern democracies.

As this agent-based model is refined and updated, we hope to have more data (from other studies) to help predict the behavior of these 34 agents. Initial assumptions (offered here) are based on the simple expectation that individuals and groups will tend to pursue their own "local" self-interest (i.e., reacting in response to the immediate pressures they feel, rarely considering the needs of, or implications of, their actions for society as a whole).

1. the venture capital community and other investors: Data published by Stanford show global corporate investments of nearly $1.3 trillion in AI during the period from 2013 through 2023 (Maslej et al., 2024). Major AI leaders have continued pursuing unprecedented funding rounds—Sam Altman at OpenAI has been orchestrating multi-hundred-billion-dollar investment consortia to secure AI computing infrastructure, while other leading labs have followed similar strategies. Given the staggering amount of capital being poured into the field, investors clearly expect that AI will deliver extraordinary commercial returns and are implicitly betting that regulatory frameworks will remain favorable to rapid development and deployment. The investment community has positioned AI not merely as "the next big thing" but as providing a radically new foundation for the whole global economy.

2. scientists and programmers with special skills relevant to the development of AI: These AI experts can command impressive hiring bonuses, salaries, and other perks. Individuals may have concerns about safeguards and safety for AI, yet relatively few will have the time, background, or disinterested perspective needed to think carefully about the effects that AI may have on the rest of society and on the stability of democratic governments.

As AI continues to advance, AI itself will affect prospects for employment and the nature of work within these companies. As noted in section 11 below, AI is becoming so powerful that companies may have less need for junior-level programmers. (Already, some tech start-ups are using AI tools to grow and become profitable with far fewer employees than traditionally required (Griffith, 2025).) Senior programmers will have many new tools to increase their productivity, giving a feed-forward loop that is likely to accelerate the development of AI.

3. experts focused on AI safety: Many individuals and groups are working to minimize risks associated with the rise of AI. There are, for example, efforts to protect AI systems from malicious attacks; to safeguard any personal data that's handled when training or using AI; to ensure that AI systems don't propose and are not used to facilitate immoral or illegal acts; to ensure that AI systems don't perpetuate biases against any race, gender, age, or religious belief; and to set up systems that can test and verify the safety of the code before it is deployed.

This is important work and should reduce risks associated with the use of AI. However, immense vigilance is needed: AI systems are useful precisely because of the flexibility they display; we have no way to know beforehand precisely how they will behave in any given circumstance. The presence of so many groups and people in the field (as with open-source code for developers now being distributed by Meta and other companies) will make it hard to enforce any uniform safety standards. Given the way these models can be adjusted, learning to "be good" with a limited amount of training (Bai et al., 2022), it's quite possible that malicious actors will find ways to retrain these models to redefine "good" in bizarre and dangerous ways (perhaps—hypothetically—helping a madman design a deadly new virus to help "control world population and protect the planet.")

4. big tech companies: Some of the major tech companies have made substantial investments in AI safety research and development of responsible AI principles—trying to limit risks of direct, proximal damage caused by the use of AI—yet almost no one looks ahead and considers emergent system-level consequences. The economic resources of companies like Meta, Apple, Microsoft, Amazon, Nvidia, and Google make it clear that they can hire enough lobbyists so that it will remain hard to have any effective regulatory oversight of their actions. With their international reach, these companies each have a power greater than that of many nation states, and—according to the rules of competition in a free market economy—there's no legal responsibility and no economic incentive for these companies to police themselves and try to foresee how system-level effects could (conceivably) undermine prospects for the long-term stability of democratic governance.

5. companies that make specialized chips for AI: Nvidia Corporation, which designs specialized chips that are especially useful for AI, has—as measured by market capitalization—become one of the most valuable companies in the world. The actual manufacturing of these advanced chips depends heavily on Taiwan Semiconductor Manufacturing Company (TSMC), while the Netherlands-based ASML provides the critical extreme ultraviolet lithography equipment required to produce cutting-edge semiconductors. This concentration of essential technologies has made the semiconductor supply chain a central geopolitical concern for the United States and for everyone working at the very cutting edge of AI development.

6. electric companies and server farms: The widespread use of AI will put additional strain on the U.S. electrical grid and will further increase the risk of both technical failures and targeted attacks on grid infrastructure and server farms. Over time, risks will increase in terms of both likelihood and severity: 1) The use of AI may help foreign adversaries and terrorists develop new, ever-more sophisticated modes of cyberattack, and 2) the chaos that could be caused by disruption to the electrical grid or server farms (already at a VERY dangerous level) will only further increase as society becomes more dependent on AI. (In the future, AI will be doing so much of the work that human employees—chosen, trained, and using workflows designed for an AI-dependent work environment—will have no way to run our factories and companies if the server farms are disabled. And severe errors could occur during system restarts unless AI systems have been carefully designed to deal with these scenarios.)

7. mining companies: The rise of AI will require new sources of critical materials, including both semiconductor materials (such as silicon, gallium, germanium, and tantalum) for advanced chips and rare earth elements (lanthanides like neodymium, dysprosium, and praseodymium) for data storage systems, cooling components, and other hardware infrastructure. New mines will need to be developed (since much of the current supply comes from China), and these new mines are likely to cause serious environmental damage. Concentrations of the metals (even in the best mines) are so low that huge mines are needed; processing requires vast amounts of water; strong acids are needed for extraction; and tailings from the mine typically contain toxic and/or radioactive elements like thorium and uranium.

8. environmental advocates: Environmental advocates will be concerned by the power demands of the server farms and by the environmental damage caused when mining for rare earth elements. Yet environmental advocates also will benefit from AI—

helping them optimize energy grids, accelerate clean energy development, reduce waste, and improve recycling systems.

9. billionaires and the ultra-rich: An increasing portion of the world's richest billionaires now come from high-tech fields, wielding immense social and political influence. As of 2024, tech figures like Elon Musk, Jeff Bezos, Mark Zuckerberg, and Larry Ellison (from Oracle) consistently were among the handful of wealthiest individuals in the world, with Bernard Arnault (head of LVMH, a French luxury goods conglomerate) often being the only non-tech individual with comparable personal wealth. While some wealthy tech leaders—like Bill Gates, who is a bit further down on the list—have shown a deep interest in social issues, all of these leaders seem predisposed to believe in the importance and power of further technological advances. And, often living in bubbles created by their own astonishing wealth, they may have little incentive to think about the needs of the rest of society or to understand why many citizens want the rich to pay higher taxes. Indeed, there is a risk that computers will help some people become so rich and so powerful that democracy gives way to a plutocracy or oligarchy. (Risks might arise as wealthy donors are able to make campaign donations so large that they can—in essence—"buy elections" and thereby gain undue influence in future public policy decisions.)

10. the corporate C–suite: Corporate leaders face real challenges as they try to decide how AI should be deployed in the workplace. It may offer immense advantages —in speed, accuracy, and the range of knowledge considered—yet at each stage in the development of AI, corporate leaders will need to weigh these potential advantages against the risk of errors, possible vulnerability to cyberattacks and corporate espionage, the challenge of integrating AI into existing workflows, and potential legal liability if mistakes occur. At the same time, those in the C-suite are likely to retain their high salaries, and they will—as always—be looking for strategies that maximize quarterly sales and earnings and the market capitalization of the company.

11. high-skill workers: Given the high salaries that such employees command, corporate leaders will have a clear incentive—whenever possible—to lay off parts of their skilled workforce (such as some of their skilled programmers). Remaining employees will need to adapt rapidly as AI advances, using this as a tool to improve their own productivity and that of the company.

12. middle-skill workers: Employees at this level may be at the greatest short-term risk of losing their jobs. As chatbots and AI-based agents keep improving, companies may need fewer bookkeepers, content writers, legal secretaries, junior-level programmers, customer service representatives, technical writers, and other such middle-skill employees. (A smaller number of such employees, with a bit of training, could use AI and carry the full workload.) This could lead to a shocking change in the lives of middle-skill workers (and perhaps eventually to the economic system as a whole) since it's unclear how these employees can shift to new careers if AI disrupts a large portion of the jobs available at this intermediate level of training/skill.

Note: Down the road, there will be important, secondary consequences resulting from whatever decisions companies make regarding hiring and retention of junior employees. If they make radical cuts here, they may lose the “pipeline” via which they can nurture and develop the next generation of leaders the company will need. (Yet they could, of course, end up “nurturing” AI systems as the next generation of leaders.)

13. low-skill workers: These first stages of the AI revolution are likely to have a mixed, perhaps somewhat more limited, effect on low-skill employees. There may be slightly less pressure for companies to replace low-skill workers (like food service workers, custodial staff, housekeepers, and delivery drivers) since these employees tend to have lower salaries than other, more skilled workers, and yet these jobs are vulnerable with ongoing advances in robotics.

14. robots and AI-based agents: Year by year, robots and AI-based agents will play steadily increasing roles in society. They may serve as caregivers, companions, and advisors; function as tutors or employees; or operate as spies, warriors, or saboteurs. They will become ever-more independent and also will assume myriad new roles at subsequent stages in the ongoing development of AI (facilitating the development of artificial general intelligence and then of superintelligence). They may offer amazing new power to the individuals, companies, and countries who own and “control” them, yet their rapid proliferation will make it exponentially more difficult for anyone to understand the whole social/political/economic system (to really know what is happening in society), and it thus will become ever more difficult to develop, implement, and enforce effective regulatory policies.

15. construction, manufacturing, and heavy industry: AI will help make robots (and automated machinery) more intelligent and versatile and should eventually cut costs across a range of industries. That said, there will be limits to the improved efficacy in cases—as when making steel or pouring concrete—where direct physical, chemical, and thermodynamic constraints are the primary, limiting factors.

16. parents and educators: Advances in AI will have two key consequences here. Educators (like everyone else) will have reason to be concerned about their own prospects for stable employment amidst an age of such rapid change, as—for example—with expectations that human tutors will be replaced with automated systems powered by AI. At the same time, uncertainty about the future structure of society will make it increasingly difficult for parents and educators (educators from grade school through graduate school) to foresee what types of jobs may be available in the future and thus to know what knowledge, skills, and attitudes they should try transmitting to the next generation.

17. young people preparing for careers: Change is occurring so rapidly that students in high school and college will have ever-increasing trouble knowing what to study and how to focus. Students may end up feeling disoriented and disaffected since it takes many years of study (often along with huge tuition payments) to develop real expertise, and yet—as students and families make these huge investments—they will have no way to know what jobs may be available when the students graduate and try to enter the job market. It also will be much harder to get internships and entry-level jobs since AI will be doing much of the “middle-skill” work that might be done by new employees.

18. child-bearing couples: AI will lead to such massive changes in the economy and in society that it’s likely to affect fertility rates. Couples in developed economies already are having fewer children, and the trend may accelerate if life becomes more difficult, with good jobs harder to obtain and the very role of human beings appearing uncertain. This, in turn, could put immense pressure on the economic system: falling birth rates would lead to a “graying population,” with too few active workers to support the kind of safety net needed to care for the elderly.

19. media and journalists: AI will certainly have a huge impact in realms that depend on a flow of images, words, and ideas. Instant analysis will be easy, since chatbots will start (and actually now are) writing some news reports (Peiser, 2019). This may accelerate trends already seen in the Information Age (Meyer, 2004): it may be harder to support careful, in-depth analysis amidst the ongoing commodification of the news. The pressure to post instant analyses and to compete for attention amidst an information overload will only increase with the rise of AI. In short, there will be an immense, increasing challenge of distinguishing signal from noise. And—at the same time—ongoing developments in AI, and their effect on society, will be so profound and proceed so rapidly that it will be almost impossible for the public and political leaders to stay informed and understand the real changes that are occurring in this realm. Indeed, without the kind of careful thought supported by our new cognitive scaffolds, it will be hard for anyone to know what’s actually happening in the world at large.

20. criminals: Although there will be cases in which AI helps law enforcement, such effects will be more pronounced in repressive societies (where there may be fewer or no restrictions on the types and amount of information that can be gathered about each citizen, and where facial recognition algorithms may help track everyone’s movements). Meanwhile, AI will give criminals new tools that will facilitate cyberattacks, theft of corporate secrets, use of ransomware, phishing campaigns, identity theft, election interference, and other nefarious activities. (And, as often noted, attack is easier than defense in many of these situations.)

21. groups peddling misinformation and disinformation: AI gives these groups new tools, facilitating production of fake images and fake videos, allowing for robocalls that mimic the voice of political leaders or of other well-known individuals, and allowing chatbots to write fake new stories. There also will be a vast “gray zone” here, involving activity that is very disruptive to democratic society yet, absent new regulations, may be protected as free speech.

22. lawyers: Chatbots and AI-based agents may help with many aspects of the workflow in law offices, and law firms may reduce their overall staff size (perhaps needing fewer paralegals and entry-level lawyers). However, the rise of AI, with implications in fintech, copyright law, patent law, medicine, and workers' rights (i.e., with all the new questions that will arise in this Age of AI) will ensure the continued employment and ongoing engagement of huge teams of lawyers. (And, as happens so often in law, the legal advantages will tend to disproportionately go to the rich and the powerful. They will be able to hire law firms that can better leverage the full power of the most advanced AI systems.)

23. judges and jurors: Litigation (resulting from all the changes associated with the rise of AI) will put immense pressure on the U.S. legal system. Judges and jurors will be put in particularly awkward positions since they will need to rely on expert witnesses as they try to understand relevant aspects of artificial intelligence (including ways in which AI may have been used in an attempt to influence their analysis of a case). And—unless/until Congress passes some clear, comprehensive new set of regulations—these judges and jurors will repeatedly need to make decisions in situations where there is no existing case law that provides meaningful guidance. (They will, in some very real sense, need to make things up as they go along.)

24. fintech, bitcoin, and blockchain users: As entrepreneurs and programmers in financial technology have been designing new ways to control the flow of money, they have been developing schemas that are complex enough that their strategies will fit naturally with—and will readily benefit from (or be dependent on)—further advances in AI. As they have done this, proponents of the work have been using labels that sound good, talking about the need for "decentralization," "disintermediation," and the "democratization of finance." They make it sound very exciting, but it's often not clear why such changes are needed, how society as a whole will benefit, and what new system-level risks will be introduced as the financial system keeps getting more complex. Everything will become even harder for citizens and political leaders to understand as AI methodology gets integrated with these new forms of fintech (especially if fintech ends up operating without any clear government oversight or meaningful regulatory policies).

25. scientists and engineers: Effects on employment trends are unclear (there may be less need for junior staff members), yet AI has dramatically accelerated the pace of research and discovery in a wide variety of scientific and technical fields. When sufficient data is available, AI can track the subtle effects of many hundreds or thousands of different variables, letting AI solve the protein folding problem and design new proteins (Jumper et al., 2021; Hayes et al., 2025). In related ways, AI seems likely to facilitate astonishing advances in medicine, healthcare, material sciences, climate modeling, and other fields.

As AI advances, it will help scientists and engineers across a full range of tasks: helping with hypothesis generation, experimental design, manuscript preparation, press releases, patent applications, and preparation of their talks. Over time, AI will make such seminal contributions to research that it may seem "unfair" if it's not listed as an author on the papers and as an inventor on the patents. (Why should human “authors” and “inventors” get full credit for something that they, themselves, could not have done?)

26. social scientists: AI is used widely in fields like sociology, economics, and political science. It can process such vast amounts of data and track so many variables that—once again—it can find patterns that the unaided human mind would never have predicted and never have detected.

It is too early to know what impact this AI-facilitated work in the social sciences will have on society as a whole (i.e., it’s not clear when/how new discoveries made by these social scientists will change public policy). Yet these new tools do open new paths for ongoing research, allowing social scientists to 1) work with vast amounts of unstructured data; 2) design and administer surveys in a personalized way (almost as if meeting with and interviewing each person); and 3) develop and run more sophisticated models and simulations of human behavior and of the dynamics within social systems. (I.e., it helps in some areas that will be directly relevant as our agent-based analysis is developed further.)

27. citizens, voters, and community groups: Ultimately, citizens and voters will need to take the lead and make their voices heard if they want the government to enact effective AI regulations. Yet—no matter how worried they are—it’s hard for most citizens to get engaged. Pressing daily concerns dominate most people’s attention; few people (as a portion of the overall population) have any idea how AI works; and the average citizen has no way of foreseeing what policy options might be available. Citizens and voters may well get mad if/when they lose their jobs—and they may lose their faith in the “social contract.” Yet it’s unclear how (outside of violent protests) this kind of anger can have any focused political power, or any real effect on society, unless/until a) some kind of broad-based movement develops with a clear agenda and a clear rallying cry, or until b) some demagogue takes advantage of the widespread discontent and exploits this unrest as a means of gaining power.

28. military leaders: The military will need to explore a wide range of new AI-enabled offensive and defensive systems. We make no attempt to discuss the full range of issues/options here, but some of these new weapons and applications—as with i) “killer robots,” ii) “robots swarms,” iii) new weapons for cyberwarfare, and iv) the potential for AI-driven systems controlling the launch of nuclear weapons—will be so terrifying and potentially so destabilizing that international negotiations may be needed to develop new treaties and new verification methods. (See further discussion under point 34.)

29. regulators and policymakers: The U.S. has no comprehensive strategy for monitoring or controlling the development and use of AI, yet Congress and a number of departmental agencies are concerned enough that they are now exploring the issues and weighing various regulatory options.

The overall process of policy development is slow because of the complexity, novelty, and rapid development of AI technology. Perhaps the biggest, clearest concrete steps taken to date by the U.S. government involve restrictions on the sale of specialized chips to China.

Note: Advocacy groups have raised myriad other concerns about AI, and many of the big tech companies are working to address concerns like those about bias and discrimination, about misinformation and disinformation, and concerns about various ways in which AI might be used for evil. Yet there are no strict regulations designed to control/limit these risks, and the U.S. Congress seems hesitant to take any kind of preventive, proactive approach.

30. political leaders: Political leaders in the free world will struggle when trying to develop any effective policy response amidst these ongoing changes: Few leaders will, themselves, understand the technology in any meaningful detail, and they will need advisors who understand AI and who have thought carefully about the social and political issues. Yet no one—advisors or politicians—will be able to make fully reliable, detailed predictions about how AI will affect society; and—when competing regulatory proposals arise—it will be hard (without use of these new cognitive scaffolds) for them to tell which, if any, have real merit, and almost impossible to push back against big tech companies that have such immense wealth and that are investing so much money in developing this new technology.

31. election officials and administrators: There is a chance that AI could help in several ways here—perhaps, for example, by allowing more rapid detection of misinformation campaigns waged by foreign adversaries, perhaps by facilitating the use of biometric information to increase the security of systems used for voter authentication, and perhaps by allowing vote totals to be counted and certified more rapidly. However, great care will be needed when making any such changes, since changes might introduce new privacy concerns and new cybersecurity risks, and news of such changes could lead to misinformation and disinformation campaigns suggesting that elections were somehow being rigged.

32. autocrats: AI will give immense additional power to autocratic leaders, giving them new tools, such as facial recognition, for tracking dissidents (Beraja et al., 2023), and giving them new tools for generating and disseminating propaganda. Many of these leaders will have the resources needed to override—via retraining as necessary—any safety features in AI systems that had been designed to assure alignment with a different, democratic system of values. As noted by William MacAskill and other members of the effective altruism community (MacAskill, 2022), ruthless leaders might find ways to suppress dissent and to manipulate thought so effectively as to lead to a kind of long-term “lock in” of the power of an evil regime.

33. poor and developing nations: AI will open amazing new doors for savvy “first adopters” in poor and developing nations, yet these countries will not have the kind of trained workforce, financial resources, electrical infrastructure, and internet connections needed to take full advantage of this new (and ever-changing) technology. Rich nations will maintain a comparative advantage as technology continues to evolve and—if ruthless—powerful countries could use AI in attempts to control and manipulate other weaker states.

34. foreign enemies of the U.S.: With new powers of AI, foreign enemies of the United States will have a whole new set of tools at their disposal. The U.S. will need to prepare for potential “hot wars” involving direct armed conflict with new AI–powered weapons used by each side. And, at the same time, the intensity of ongoing “gray zone conflict” (occurring somewhere below the threshold of outright war) will accelerate with new tools for cyberwarfare, for spreading misinformation and disinformation, for gathering information about U.S. citizens, and for stealing secrets from U.S. corporations and government agencies.

Open conflict will involve full-scale use of AI-based efforts to attack and disable the enemy’s electrical grid, communication systems, and server farms. The net effect (given our new multipolar world order and the rise of AI) will be a dramatic, disruptive increase in the cognitive complexity of military planning and (as needed) military operations. A recent paper by Hendrycks, Schmidt, and Wang (2025) takes an important initial step, proposing the idea of Mutual Assured AI Malfunction (MAIM) as a deterrence regime resembling nuclear mutual assured destruction (MAD).

Yet issues here are so complex that work on these problems would benefit from development of a cognitive scaffold—similar to that offered in my algorithm for thought (Pabo, 2022)—that focuses on the U.S. military and the way that military planning is coupled with other diplomatic and economic efforts to ensure U.S. national security.

Herman Kahn (1962) showed that it was possible to think in a clear, rational way about the “unthinkable” horrors of nuclear war. And, when considered in light of the sheer complexity of the issues now facing the military, human thought faces a challenge similar to that which it faced at the dawn of the nuclear age. There is a desperate need for new tools to support thought, and scaffolds like the one offered here can help.