

THE PHILOSOPHY OF AI: LEARNING FROM HISTORY, SHAPING OUR FUTURE

HEARING

BEFORE THE

COMMITTEE ON
HOMELAND SECURITY AND
GOVERNMENTAL AFFAIRS
UNITED STATES SENATE
ONE HUNDRED EIGHTEENTH CONGRESS

FIRST SESSION

NOVEMBER 8, 2023

Available via the World Wide Web: <http://www.govinfo.gov>

Printed for the use of the
Committee on Homeland Security and Governmental Affairs



U.S. GOVERNMENT PUBLISHING OFFICE

53–996 PDF

WASHINGTON : 2024

COMMITTEE ON HOMELAND SECURITY AND GOVERNMENTAL AFFAIRS

GARY C. PETERS, Michigan, *Chairman*

THOMAS R. CARPER, Delaware	RAND PAUL, Kentucky
MAGGIE HASSAN, New Hampshire	RON JOHNSON, Wisconsin
KYRSTEN SINEMA, Arizona	JAMES LANKFORD, Oklahoma
JACKY ROSEN, Nevada	MITT ROMNEY, Utah
ALEX PADILLA, California	RICK SCOTT, Florida
JON OSSOFF, Georgia	JOSH HAWLEY, Missouri
RICHARD BLUMENTHAL, Connecticut	ROGER MARSHALL, Kansas

DAVID M. WEINBERG, *Staff Director*

LENA C. CHANG, *Director of Governmental Affairs*

MICHELLE M. BENECKE, *Senior Counsel*

EVAN E. FREEMAN, *Counsel*

AVERY M. BLANK, *Counsel*

WILLIAM E. HENDERSON III, *Minority Staff Director*

CHRISTINA N. SALAZAR, *Minority Chief Counsel*

KENDAL B. TIGNER, *Minority Professional Staff Member*

LAURA W. KILBRIDE, *Chief Clerk*

ASHLEY A. GONZALEZ, *Hearing Clerk*

CONTENTS

Opening statements:	Page
Senator Peters	1
Senator Johnson	8
Senator Hassan	14
Senator Butler	16
Senator Hawley	18
Senator Blumenthal	21
Senator Ossoff	24
Senator Rosen	31
Prepared statements:	
Senator Peters	35

WITNESSES

WEDNESDAY, NOVEMBER 8, 2023

Daron Acemoglu, Ph.D. Institute Professor, Department of Economics, Massachusetts Institute of Technology	3
Margaret Hu, Taylor Reveley Research Professor, Professor of Law, Director, Digital Democracy Lab, William & Mary Law School	5
Shannon Vallor, Ph.D. Baillie Gifford Chair in the Ethics of Data and Artificial Intelligence, Director, Centre for Technomoral Futures, Edinburgh Futures Institute, The University of Edinburgh	7

ALPHABETICAL LIST OF WITNESSES

Acemoglu, Daron, Ph.D.:	
Testimony	3
Prepared statement	37
Hu, Margaret:	
Testimony	5
Prepared statement	55
Vallor, Shannon:	
Testimony	7
Prepared statement	67

APPENDIX

Senator Hawley Chart	84
Senator Hawley Letter to Inspector General	85

THE PHILOSOPHY OF AI: LEARNING FROM HISTORY, SHAPING OUR FUTURE

WEDNESDAY, NOVEMBER 8, 2023

U.S. SENATE,
COMMITTEE ON HOMELAND SECURITY
AND GOVERNMENTAL AFFAIRS,
Washington, DC.

The Committee met, pursuant to notice, at 10:01 a.m., in room SD-562, Dirksen Senate Office Building, Hon. Gary Peters, Chair of the Committee, presiding.

Present: Senators Peters [presiding], Hassan, Rosen, Blumenthal, Ossoff, Butler, Johnson, and Hawley.

OPENING STATEMENT OF SENATOR PETERS¹

Chairman PETERS. The Committee will come to order. We are living through perhaps one of the most exciting times in human history as artificial intelligence (AI) becomes more advanced each and every day.

AI tools have the capacity to revolutionize medicine, expand the frontiers of scientific research, ease the burdens of physical work, and create new instruments of art and culture. AI has the potential to transform our world for a better place, but these technologies also bring new risks to our democracy, to civil liberties, and even our human agency.

As we shape and regulate AI, we cannot be blinded by its potential for good. We must also understand how it will shape us and be prepared for the challenges that these tools will also bring. Some of that work will be accomplished with innovative policy, and I am proud to have passed numerous bills to improve the government's use of AI through increased transparency, responsible procurement, workforce training and more.

I have convened hearings that explore AI safety risk, procurement of these tools, and how to prepare our Federal workforce to properly utilize them. But as policymakers, we also have to explore the broader context surrounding this technology. We have to examine the historical, the ethical, and philosophical questions that it raises. Today's hearing and our panel of witnesses give us the opportunity to do just that. This is not the first time that humans have developed staggering new innovations. Such moments in history have not just made our technologies more advanced, they have affected our politics, influenced our culture, and changed the fabric of our society.

¹ The prepared statement of Senator Peters appears in the Appendix on page 35.

The Industrial Revolution is one useful example of that phenomena. During that era, humans invented new tools that drastically changed our capacity to make things. The means of mass production spread around the world and allowed us to usher in modern manufacturing economy. But that era brought with it new challenges. It led to concerns about monopolies, worker safety, unfair wages, and child labor. It produced the weapons that were used to fight two world wars. In short, it was not just about technology that could be used for good.

I am grateful our first witness, Daron Acemoglu, has studied these phenomena. He has not only examined the history of technological change, but also the democratic institutions that are needed in response. In the 20th century, we had trade unions to protect workers' rights, and effective government regulation to keep those industries in check. What tools do we need to meet this moment and what else should we learn from the history?

Artificial intelligence also brings unique challenges. The history of technological change has largely centered on human strength and how we can augment it through the use of new machines. AI will affect physical work, but unlike other technologies, it is more directly tied to our intellectual and cultural capacities. It has already introduced new ways to ask and answer questions, synthesize information, conduct research, and even make art.

These qualities, the ability to understand ideas and create culture, are the very foundation of our humanity. We must work to preserve them as they become influenced by artificial tools. Perhaps most importantly, AI's influence on these capacities is not neutral. These tools, like the humans who make them, are biased. We must define what values lie at the core of our human experience and create technological tools that support them.

Our second witness Shannon Vallor, will be a helpful resource in understanding these ethical questions. She studies the way that new technologies reshape our habits, our practices and moral character. With her help, we can understand the values embedded in these technologies and the effect that it will have on our human character.

Finally, we will explore AI through a constitutional law framework. AI poses risks to our civil liberties. New surveillance tools can be used to target vulnerable communities. Biometric systems like facial recognition can endanger a citizen's right to due process. Advanced technology brings renewed questions about our privacy and our personal information.

If we do not understand how AI can be used in ways that erode our constitutional rights, it can pose a grave danger to our democracy and civic institutions. Our third witness, Margaret Hu, will help us understand these intersections. She researches the risk that AI possesses to constitutional rights, due process, and civil liberties.

Artificial intelligence has already begun to shape the fabric of our society. Our response cannot come through piecemeal policy alone or isolated technological fixes. It must include a deeper examination of our history, our democracy, and our values, and how we want this technology to shape our future.

We must look to the past and learn the lessons of previous technological revolutions. We must answer the ethical questions that AI poses, and use these new technologies to build a world where all humans can thrive. We must protect our civil liberties and democratic institutions against risk that these tools can pose.

This hearing provides an excellent opportunity to focus on this work, and I would like to thank our witnesses for joining us today. We certainly look forward to your testimony. It is the practice of this Homeland Security and Governmental Affairs Committee (HSGAC) to swear in witnesses. If each of you would please rise and raise your right hands. Do you swear to the testimony you give before this Committee will be the truth, the whole truth, and nothing but the truth, so help you, God?

Dr. ACEMOGLU. I do.

Dr. VALLOR. I do.

Ms. HU. I do.

Chairman PETERS. Thank you. You may be seated.

Our first witness is Daron Acemoglu. Professor Acemoglu is an economist at the Massachusetts Institute of Technology (MIT). His work focuses on the intersection of technological change with economic growth, prosperity, and inequity.

Professor, welcome to the Committee. You are recognized for your opening comments.

**TESTIMONY OF DARON ACEMOGLU,¹ INSTITUTE PROFESSOR,
DEPARTMENT OF ECONOMICS, MASSACHUSETTS INSTITUTE
OF TECHNOLOGY**

Dr. ACEMOGLU. Thank you for inviting me to testify on this important topic. I will argue that there is a pro-human meaning pro-worker and pro-citizen direction of artificial intelligence that would be much better for democracy and shared prosperity.

Unfortunately, we are currently on a very different and worrying trajectory. Digital technologies have already transformed our lives, and AI has further amplified these trends, but all has not been good. U.S. inequality has served since 1980, many workers, especially men without a high school degree or just a high school degree, have seen very significant declines in their real earnings, and inequality has multiplied in other dimensions as well.

My research indicates that the most important cause of these trends is automation, meaning the substitution of machines and algorithm for tasks previously performed by workers. Automation accounts for more than half of the increase in U.S. inequality. Other trends such as offshoring and imports from China have played a somewhat smaller role.

Technological change is a force for good, but we need to use it the right way. During the mechanization of agriculture and the three decades following World War II, automation was rapid, but the U.S. economy created millions of good jobs and built shared prosperity. The main difference from the digital age was that the new technologies not only automated some tasks, but also created new ones for workers.

¹The prepared statement of Dr. Acemoglu appears in the Appendix on page 37.

Henry Ford's factories used new electrical machinery that automated some work, but at the same time, they also introduced many new technical tasks for blue collar workers. Simultaneously, manufacturing became much more intensive in information activities and created a lot of jobs through these channels as well, such as in design, planning, inspection, quality control, and accounting. Overall, new tasks were critical for employment and wage growth during these eras.

Unfortunately, my research showed that emerging AI technologies are today predominantly targeting automation and surveillance. The emphasis on surveillance is, of course, much more intense in China. We are already seeing the social and democratic implications of the rising inequality. In the United States, areas that have been most hit by Chinese competition or the introduction of robots show much greater degrees of polarization. Inequality undermines support for democracy, and this lack of support makes democracies more unstable and less capable of dealing with challenges.

This path is not inevitable. To improve human performance. We need to think beyond creating AI systems that to that seek to achieve artificial general intelligence or human parity. The emphasis on general intelligence is not just a Kymira, but distracts from the more beneficial users of digital technologies to expand human capabilities.

Making machines useful to humans is not a new aspiration. Many people were working on this agenda as early as the 1940s, and many technologies that have been foundational to our lives today, including the computer mouse, hyperlinks, menu-driven computer systems came out of this vision.

Machine usefulness is more promising today than in the past. The irony of our current age is that information is abundant, but useful information is scarce. AI can help humans become better problem solvers and decisionmakers by presenting useful information. For example, an electrician can diagnose rare problems and accomplish more complex tasks when presented useful information by AI systems.

The analog to the pro-worker agenda in communication is a pro-citizen perspective to provide better information to individuals and enable them to participate in deliberations without manipulation or undue bias. The opposite approach is one that focuses on surveillance, manipulation, manufacturing false conformity. The evolution of social media illustrates this manipulative path with algorithms used for creating echo chambers and extremism.

The survival of any political regime depends on how information is controlled and presented. Authoritarian rulers have understood this for ages. The rulers in China, 2,200 years ago, reputedly burned books and executed people who could rewrite them to control information. The anti-democratic use of computers is clearly visible in Russia, Iran, and China.

Whoever controls information matters no less for democratic regimes. Digital platforms' monopoly over information today is completely unprecedented. Their business model is based on monetizing data via digital ads and much work in social psychology documents, and in fact, unfortunately teaches platforms how to in-

crease engagement by manipulating user perceptions and presenting them with varying stimuli and emotional triggers.

AI is a new technology, but as you pointed out, history offers important clues about how to best to manage it. The British Industrial Revolution is today remembered as the origin of our prosperity. This is true, but only part of the story. The first 100 years of the British Industrial Revolution were simply awful for the working people. Real income stagnated, working hours increased, working conditions deteriorated, and health and life expectancy became much worse in the face of uncontrolled epidemics and intensifying pollution.

The more positive developments after 1850 were due to a major direction of technology away from just automation and toward pro-human goals. This was embedded in fundamental political and social changes, including democratization and new laws to protect worker voice and worker rights.

Just like during the industrial revolution, we have widely different paths ahead of us. A pro-human direction of AI would be much better for prosperity, democracy, and national security. Yet that is not where we are heading. My five minutes is up, but I will be happy to discuss later policy proposals for redirecting AI toward a more beneficial trajectory. Thank you.

Chairman PETERS. Thank you, Professor.

Professor Hu is a Professor of Law and Director of the Digital Democracy Lab at William & Mary Law School. She is a constitutional law expert, and her research focus focuses on the intersection of technology, civil rights, and national security. Professor Hu previously served in the Civil Rights Division at the U.S. Department of Justice (DOJ).

Professor Hu, welcome to the Committee. You are recognized for your opening comments.

TESTIMONY OF MARGARET HU,¹ TAYLOR REVELEY RESEARCH PROFESSOR, PROFESSOR OF LAW, DIRECTOR, DIGITAL DEMOCRACY LAB, WILLIAM & MARY LAW SCHOOL

Ms. HU. Good morning. It is an honor to be a part of this critically important dialog on the philosophical and historical dimensions of the future of AI governance. The reason we must consider the philosophy of AI is because we are at a critical juncture in history where we are faced with a decision; either the law governs AI or AI governs the law.

Today, I would like to place AI side by side with constitutional law. Doing so allows us to visualize how both function on a philosophical level. It also provides us with a window into how they are philosophically in conversation with one another and gives us a method on how we must best respond when we see that they are philosophically in conflict with one another.

Constitutional law is more than just the text of the Constitution and cases. Similarly, AI is more than its technological components. AI can be understood as more of a philosophy than a technology. Like constitutional law. They are both highly philosophical in nature. Specifically, AI is animated by multiple sciences and philoso-

¹The prepared statement of Ms. Hu appears in the Appendix on page 55.

phies, including epistemology, a science in philosophy concerning the structure of knowledge and ontology; the philosophy of existence.

AI and the law is highly complex and requires grappling with these interdisciplinary consequences, just as constitutional law is highly nuanced and contextualized. In the past year, we have entered a new phase of large commercially driven AI investments. This new phase brings into sharp relief the need for a dialog on rights-based AI governance.

The creators of generative AI have shared that their ambition is to advance artificial general intelligence, which aims to surpass human capacities. Generative AI and Artificial General Intelligence (AGI) ambitions force us to confront these epistemological and ontological questions head on and with some urgency in a constitutional democracy.

AI is already being deployed as a governing tool in multiple contexts. AI, particularly due to its combined ontological and epistemological powers, as well as its combined economic, political, and social power, has the potential to evolve into a governance philosophy as well as potentially a governance ideology.

AI is constitutive of not only a knowledge structure, but also a market structure in an information society and a governing structure in a digital political economy. The incentives of AI privatization and the exponential growth of datafication can operate as an invisible governing superstructure under an invisible and potentially unaccountable hand. Additionally, AI can execute both private and public ordering functions, sometimes without authorization, rapidly shifting power toward centralized and privatized and often automated and semi-automated methods of governing.

The Constitution is inspired by a philosophy of how to guarantee rights and how to constrain power. Constitutional law is animated by commitment to a governing philosophy surrounding self-governance through a republican form of government. In theory and philosophy, it separates and decentralizes power, installs checks and balances to prevent or mitigate power abuses and supports a government that is representative of the people by the people for the people.

An important question at this critical juncture is how to ensure that AI, as it potentially evolves into a governing philosophy, will not compete with and rival constitutional law as a governing philosophy in a way that sacrifices our philosophical commitments, to fundamental rights, and to constraints on power, including separations of power.

The Constitution is more than a text; it is a philosophy. AI is more than a technology; it is a philosophy. I would like to return to my opening question, which is; will AI govern the law or will law govern AI. In order to preserve our democracy and reinforce it, there can only be one answer. The law must govern AI.

Thank you.

Chairman PETERS. Thank you, Professor.

Finally, Shannon Vallor is a Professor at the University of Edinburgh. She is appointed to the university's department of philosophy, as well as Edinburgh Futures Institute. Her research centers

on the ethical challenges of AI and how these new technologies reshape human character, habits, and practices.

Professor Vallor, wonderful to have you here all the way from Edinburgh. You may proceed with your opening comments.

**TESTIMONY OF SHANNON VALLOR,¹ BAILLIE GIFFORD CHAIR
IN THE ETHICS OF DATA AND ARTIFICIAL INTELLIGENCE,
DIRECTOR, CENTRE FOR TECHNOMORAL FUTURES, EDIN-
BURGH FUTURES INSTITUTE, THE UNIVERSITY OF EDIN-
BURGH**

Dr. VALLOR. Thank you, Chairman Peters, and distinguished Members of the Committee for this opportunity to testify today. It is a profound honor to address you on a matter of such vital importance to the Nation and the human family.

I direct the Center for Technomoral Futures at the University of Edinburgh, which integrates technical and moral knowledge in new models of responsible innovation and technology governance. My research is focused on the ethical and political implications of AI for over a decade. It is deeply informed by our philosophical and historical perspectives on AI's role in shaping human character and capabilities.

The most vital of these capabilities is self-governance. This capability to reason, think, and judge for oneself how best to live underpins the civil and political liberties guaranteed by the U.S. Constitution and by international law. It also underpins democratic life. My written testimony explores the deep tension between AI and our capacity for democratic self-governance, and some important and powerful lessons from history for resolving it.

The power of AI is one we must govern. In modern democracies, free peoples may not justifiably be subjected to social and political powers which determine their basic liberties and opportunities, but over which they have no say, which they cannot see and freely endorse, and which powers are in no way constrained by or answerable to them.

Many of the greatest risks of AI technologies have arrived before the promised social benefits, which prove harder to deliver at scale. Yet, the gap between AI's social power and our democratic will to govern it remains vast. As a result, public attitudes toward AI are souring. This is a grave warning for those of us who want AI technologies to mature and succeed for human benefit. GMOs and nuclear power also suffered public backlash in ways that greatly limited their beneficial use and advancement. AI may become a similar target.

Yet, we do know how to govern AI technologies, and responsible AI researchers have given us plenty of tools to get started. The United States has a long and proud history of regulatory ambition in making powerful and risky technologies safer, more trustworthy, and more effective, all while fueling innovation and enabling wider adoption. It was done first in the 19th century with steamboat regulation, then automobiles, aviation, pharmaceuticals, and medical devices to name just a few.

¹ The prepared statement of Dr. Vallor appears in the Appendix on page 67.

This required the courage to give manufacturers, operators, and users irresistible incentives to cooperate. It required the capacity to keep learning, and innovating, and adjusting our regulatory systems to accommodate technological change. It also required persistence of shared governance aims in the public interest across changes in political administration.

This was all within our democratic capacity and still is, but the political will to use that capacity is now damaged for many reasons. The mischaracterization and misuse of AI technologies makes this problem worse by undermining our confidence and our own capabilities to reason and govern ourselves. This was predicted by early AI pioneers.

In 1976, Joseph Weizenbaum lamented that intelligent automation was emerging just when humans have, “ceased to believe in, let alone to trust, our own autonomy.” Norbert Wiener, who developed the first theories of machine learning and intelligent automation, warned in 1954 that for humans to surrender moral and political decisionmaking to machines, “is to cast our responsibility to the winds and to find it coming back seated on the whirlwind.”

Yet, many of today’s powerful AI scientists and business leaders claim that the truly important decisions will soon be out of our hands. As just one example, OpenAI Sam Altman has suggested that we are merely the biological bootloader for a form of machine intelligence that will dwarf ours not just in computing power, but in wisdom and fairness.

These careless and unscientific AI narratives are pressing on democratic cultures already riddled with stress fractures. If we do not assert and wisely exercise our shared capacity for democratic governance of AI, it might be the last chance at democratic governance we get.

Had AI arrived in a period of democratic health, none of its risks would be unmanageable. But we are in a weakened political condition and dangerously susceptible to manipulation by AI evangelists who now routinely ask, “What if the future is about humans writing down the questions and machines coming up with the answers?” That future is an authoritarian’s paradise.

The question upon which the future of democracy hangs, and with it our fundamental liberties and capacity to live together is not, “What will AI become and where is it taking us?” That question is asked by someone who wants you to believe that you are already out of the driver’s seat. The real question is, “What kind of future, with AI, will democracies choose to preserve and sustain with the power we still hold?” One where human judgment and decisions matter, or one where they don’t.

Thank you to the Committee.

Chairman PETERS. Thank you, Professor.

You can see there is a lot going on today. We have people coming and going. But Senator Johnson has to leave shortly. Senator Johnson, if you have a moment for a question or two, you are recognized.

OPENING STATEMENT OF SENATOR JOHNSON

Senator JOHNSON. First of all, I appreciate this hearing. I really do. I thank the witnesses. I like the hearing title Philosophy of AI

because I think it is crucial. I have been interested in science fiction all my life, and now we have been holding these seminars or hearings here in the Senate trying to understand what this is. But I have also been reading some pretty interesting science fiction books.

Science fiction writers are unbelievably prescient. These things are researched pretty well. These things go off in different directions, and some pretty troubling directions, which is, I think, what the Chairman was talking about in his opening remarks as well as what you are talking about.

Professor Vallor, you are talking about our capacity to regulate this and our ability to do so. President Eisenhower in his farewell address not only talked about the military industrial complex, he also warned us about government funding science and research. That would lead to scientists more concerned about obtaining a government grant than really addressing pure scientific knowledge. It could lead to a scientific and technological elite that drove public policy.

I think that was the concern with AI. He was concerned about human beings that were technologically and scientifically elite. Now we have computer capability that's going to vastly outpace our ability in terms of volumes and speed of calculations. It is highly concerning. I would argue just with the latest pandemic. Scientific research was certainly looking at how to take a virus with gain of function, make it more dangerous, and then come up with a countermeasure anticipating biological warfare.

I would argue that obviously got out of hand. We do not know the exact origin, but I guess I am less convinced that we are going to really be able to control this and that we have the governing capacity to do so. It is hard to really put a question in on this, but this is an incredibly important issue in question, and I really do not know whether this dysfunctional place is going to come up with the right answers.

I think of the Hippocratic Oath; first, do no harm. Again, I am not a computer scientist. I cannot even begin to grapple with how they create these algorithms. We have a few smart people that know this, that are warning us about certainly a possibility of AI destroying this country, destroying humanity. I guess just speak to that.

Dr. VALLOR. Happy to. Thank you very much. I think the important point in what you are saying is that AI today is an accelerator of many of the dynamics that are currently present in our society, in our economy. Among those, for example, is rising economic inequality, and declining social and economic mobility, which has been an issue now in this country for decades.

One of the greatest worries about AI is that it will accelerate those trends unless we actively govern AI in ways that ensure that its benefits are realized by everyone who has a right to have the infrastructure that AI will build, serve them.

I will just say that, I think, the fact that we have done this in the past with other technologies that were at the time equally unprecedented, equally powerful, equally challenging to regulate, actually if we have the political will, leaves us in a better place than we have ever been to govern a complex technology like AI because

we have 200 years of experience in regulatory innovation, in adjusting the incentives for powerful actors.

It has been effective before. That is why airplanes now are safer than driving. That would not have happened if we had stepped back and let airlines operate without any kind of regulatory oversight, accountability, or constraint.

Senator JOHNSON. But regulating transportation devices is completely different. I would even say nuclear power versus nuclear weapons is completely different than this that we cannot even begin to grapple with once this is unleashed and it is starting to learn and maybe even becoming self-aware. What is that going to actually result in?

Professor Acemoglu, you talked about, we have all talked about growing inequality. I would argue that we have put ourselves in a huge pickle over the last couple of generations as we have told all of our children you have to get a 4-year degree. To me, the greatest threat AI represents to loss of jobs are those college-educated kids that machines can learn a lot quicker. You are seeing what is happening with ChatGPT.

Certainly, we are seeing a real shortage of workers in manufacturing, in the trades. We are always going to need those folks. Unfortunately, certainly in Wisconsin, those employers try to hire people, our kids are not doing it because we send them all to college and they think that kind of work is beneath them.

They are screaming for legal immigration reform, which I am all for, but, here is an instance where we did not really regulate, but society en masse told all of our kids, you have to get college educated thereby implying that being in construction or being a trades person was somehow a lesser occupation, you are a second class citizen. I think all work has value. Why do you not speak to that in my remaining time?

Dr. ACEMOGLU. That's a very important point. I have also come to believe exactly like you said that we have undervalued a lot of important work. But we have not just undervalued it philosophically, we have also failed to create the right training environment and the right technologies for these workers.

It is a tragedy in this country that we have a tremendous shortage of skilled craftspeople, electricians, but they are not even paid that well.

Senator JOHNSON. They are starting to get paid well.

Dr. ACEMOGLU. They are going to get paid somewhat better because of scarcity, but they can be paid even more if we create the right environment for them to acquire more expertise and provide the tools for them.

The promise of AI is really, if we strip all of the hype, is really in doing that. Because what is generative AI good at? It is taking a vast body of knowledge and some specific context, and finding what is the relevant bit of that vast body for that specific context. That is a great technology for training. That is a great technology for boosting expertise. That is the way to actually use AI in a way that is not inequality inducing.

Now, you have raised another very important point, which many economists also make, which is; well, these ChatGPT-like technologies are going to go after college-educated jobs. I am not actu-

ally sure. This is not the first technology that has been promised to automate white collar workers or white collar work and therefore reducing equality that way.

My work finds that many of these technologies end up actually going after the lower skilled jobs. Like, you are not going to automate managerial jobs or those people who have power, but you are going to do to the sort of information technology (IT) security-type jobs, which are not very well paid anyway.

Moreover, that is not a very effective way of reducing inequality because what happens to people who let us say used to do IT security or, advertisement writing, et cetera, they go and compete for other white collar jobs that were lower paid, and the burden, again, falls on lower educated workers.

You are 100 percent right. Four-year college for everybody is not the solution, but skills for everybody, building expertise for everybody is the right solution.

Senator JOHNSON. Sorry, I cannot stay around. This really is an important subject. Thank you.

Chairman PETERS. Thank you, Senator Johnson.

I want to kind of have a little bit of a dialog perhaps, kind of ask a broad question and then we will go through. A little different than hearings where there is questions and answers. If you want to chat among each other too, that would be very much appreciated because you bring some different perspectives here.

I am going to ask a very broad question. First, Professor Acemoglu, I hope you can answer based on your historical research, your understanding of economics and framing that. All of you of have some specific examples that would be helpful for us to have in the record. Professor Hu, I would like to obviously hear your perspective based on your understanding of constitutional law. Professor Vallor, I hope you can do it based on your study of a future worth wanting. What should humans want and how do we achieve that?

The first open question is that there is a popular line of thought out there touted by many influential people that unfettered technological innovation will solve all of our problems, and it is going to lead to increased well-being for all. Just let it go and we should all be very happy about the end result. Do each of you agree with that line of reasoning? If not, why and what should we be thinking about? We will start with you.

Dr. ACEMOGLU. I completely disagree. First of all, we as humans decide how to use technology. Technology is never a solution to problems. It could be a helper or it could be a distractor exactly like you said in your opening remarks. Moreover, unfettered competition is not always the vehicle for finding the right direction of technology. There are so many different things we can do with technology, and profit incentives sometimes align with the social good and sometimes do not.

I am certainly not arguing that government bureaucrats or central planning could be a rival to the market processes for creating entrepreneurial energy or innovative energy. I do not think there is any evidence that anything better than the market process for innovation has been invented by humans. But that does not mean that the market process is going to get the direction of technology

right, and it does not mean that without regulation, we are going to use these technologies the right way.

That is why exactly like Professors Hu, and Vallor also pointed out, we need the right regulatory framework, and the right regulatory framework has to be broadly construed. It is not like we create the technologies and then we put some regulations on how they can be used. I think we need to create the right ecosystem where social input, democratic input, and government expertise are part of setting the agenda for innovation.

In the past, U.S. Government played a very important leadership role in many of the transformative technologies from antibiotics, computers, sensors, aerospace, nanotechnology, and of course, the Internet. I think the right priorities for redirecting technological change in a socially beneficial direction is very important, and that is the way to make use of these technological innovations.

But if I could add one other thing, which is a reaction to the question that Senator Johnson raised, which is, is it possible to regulate AI? I certainly believe it is possible to regulate AI, but I agree with Senator Johnson that it is much harder than the previous technologies. But the reason for that is not just the nature of technology, it is because we have become completely mesmerized with AI in the wrong way.

Both Professors Hu and Vallor emphasize AI is a philosophy. You could say AI is also an ideology. We have been chosen one specific way of perceiving that AI ideology as this general intelligence that is going to solve our problems, it is going to take away human agency, and it is not only dangerous, it is also really making it much harder for us both to find the right technologies to solve social problems and to regulate it.

I think we need a general change in perspective to help with the regulation of AI. Thank you.

Chairman PETERS. Thank you.

Professor Hu. We can go beyond this clock. You are all professors. I know you usually like to expand on your answers, and you are free to do that.

Ms. HU. Thank you for that very important question. I think this type of techno utopianism is something that we really need to look at with an eye of skepticism and especially in a constitutional democracy. We need to ask the question whether or not we have the proper means in order to achieve those ends. With that type of invitation to see technology as something that can solve all problems and needs to be unfettered, I think that it poses the problem that the ends may be not justifying the means.

In a constitutional democracy, we must always consider the means. I think that especially when we are faced with very compelling ends that are being presented before us; that AI can resolve pressing issues in national security or in health, for example, then it seems even more compelling. But I think that this really also opens the door to the conversation of whether or not when we are thinking about AI regulation, we really need to think about an ex-ante approach and not just an ex-post approach.

In the law, oftentimes, it is highly reactionary. We look at the harms, and then we try to find some type of structure to deal with those harms. But with AI, I think that this is now a moment for

us to ask what type of laws and regulations and rules do we need in order to anticipate the harms before they occur and address them.

Chairman PETERS. Thank you, Professor.

Professor Vallor.

Dr. VALLOR. Thank you for this important question. I will echo some things that my fellow witnesses have said. First of all, technology is a tool. We solve problems often with the help of technology, but technology does not solve problems. When we start expecting technology to solve our problems for us without human wisdom and responsibility guiding it, our problems actually tend to get worse.

That is what a lot of people are seeing with AI, that we have something that is not being used wisely and responsibly as a tool to solve problems, but something that we are increasingly expecting to solve our problems for us. To Professor Acemoglu's point, that is really undermining some of the confidence and ambition that we need in order to govern AI responsibly and wisely.

In response to Senator Johnson's earlier question, we certainly cannot cut and paste from aviation or any other sector, a regulatory model that is going to work for AI, but we do not need to. We can do what we have done every time before, which is innovate in the sphere of governance and adjust different incentives and powers within the AI ecosystem until we have the results that we want.

But this brings me to my second point. You talked about the idea of unfettered technological innovation, and this ideology that that kind of unfettered innovation leads us to human well-being for all. But notice that we always hear this promise now made with the word "innovation", almost never the word "progress". There is a reason for that.

Technology, not just the machines we build, but the techniques and systems that we create for all of history has been an essential driver of human progress. But that means meaningful, measurable improvements in things like life expectancy, infant mortality, sanitation, literacy, political equity, justice and participation, economic opportunity and mobility, and protections of fundamental rights and liberties.

Today, there is more advanced technology in the United States than anywhere else, but we have actually started seeing measurable declines in many of those metrics that I just mentioned. What does that tell us about the connection between technology and progress? It suggests that it is breaking down, because we have substituted the concept of innovation where we do not need to prove that a new technology actually meets a human need only that we can invent a market for it often by changing our social infrastructure so that we cannot opt out of it.

We need to go back to the heart of technology, which is the ambition to improve the human condition. You asked about this work that I have done on building a future worth wanting. The Spanish philosopher Jose Ortega Y Gasset said in 1939, that technology is, strictly speaking, not the beginning of things. He says it will within certain limits, of course, succeed in realizing the human project,

but it does not draw up that project. The final aims it has to pursue come from elsewhere.

Those aims come from us. We have to identify what we want and need technology to help us achieve. AI can be a powerful tool in helping us do that, but not if we treat innovation as an end in itself.

Chairman PETERS. Thank you, Professor.

Senator Hassan, you are recognized for your questions.

OPENING STATEMENT OF SENATOR HASSAN

Senator HASSAN. Thank you very much, Chair Peters. I want to thank you and the Ranking Member for holding this hearing, and a thank you to the witnesses for being here. We really appreciate it.

Professor Vallor, I want to start with a question to you. Your testimony discusses the possibility that bad actors can use AI in ways that threaten national security, such as in the bioengineering and nuclear fields. What would you recommend Congress do to minimize national security risks posed by the misuse of AI?

Dr. VALLOR. I think one point is that we have to be realistic and recognize that AI is not a technology that we will always be able and everywhere to keep out of the hands of bad actors.

One of the important things to recognize is that this is part of the risk profile of AI that needs to be managed in the same way that we manage many other inherently risky technologies that can also be abused and exploited.

One of the most important things is to identify what are the powerful incentives that bad actors have to abuse this technology, and where can we remove those incentives, or increase the costs for bad actors abusing AI in harmful ways. For example, the use of AI to produce disinformation is a worry for a lot of researchers. But actually there are cheaper ways to produce disinformation that a lot of bad actors have been relying on. It is not clear, for example, that AI will be the most attractive path for people who want to do harm through that, through that pathway.

I think from a national security perspective, obviously need to have close monitoring of AI developments. This is something that we need in the commercial ecosystem as well—forms of early warning systems, where we see incidents being reported back to us that we can then chase back. Many platform companies can be incentivized to do that kind of incident reporting, so that if we see signs of bad actors exploiting their tools, we have some advanced warning and ability to act.

Senator HASSAN. Thank you very much.

Professor Acemoglu, as we discuss AI and how it can magnify threats to democracy, I am particularly concerned about Chinese AI tools that are used for surveillance and censorship, and how these tools may undermine democracy and freedom around the world. Based on your research, how are Chinese surveillance and censorship tools spreading throughout the world, and what is the effect of these tools on democracy and free speech?

Dr. ACEMOGLU. Thank you, Senator Hassan, and you are absolutely right to be concerned. China is not at the frontier of most AI areas, but facial recognition, censorship and other control tools

have received the most investment in China. That is one area in which China is on a par with the United States and other nations that are leading the AI knowledge.

Those tools are not being developed intensively in China, but they are also being used very much both at the local level and the national level for controlling the population. There is evidence suggesting that they are not completely ineffective. In fact, one of the things that is quite surprising in China is that the middle class has multiplied and there are a lot of aspirations, but those aspirations are not reflecting themselves in the political domain. A lot of that is because of this very intense use of data collection control.

You are also absolutely right that those technologies are not just staying in China, China is actively exporting them to many other nations. Only the Huawei company has exported surveillance technologies to more than 60 other countries, most of them non-democratic, and those countries are also using them for surveillance.

This is part of the reason why AI leadership coming from the United States is so important because the United States has the resources, scientific resources and corporate resources to set the direction of research and it can choose a very different one from China. If the United States makes those choices, other countries will follow because the advances in the United States are going to provide profit opportunities for companies.

This is part of the reason why setting the right priorities with government support, but with also shifting priorities in the corporate and the tech world is so important. Thank you.

Senator HASSAN. Thank you. Another question for you. In today's political climate, extremism can sometimes boil over into acts of violence. Just last week, the Committee heard from the Federal Bureau of Investigation (FBI) Director Christopher Wray, that the most persistent terrorist threats to the United States are small groups or lone actors who sometimes commit acts of violence after watching or reading online content that glorifies extreme or hateful views.

Professor, what lessons can we learn from history about how major technology advancements can contribute to a climate of extremism, and what recommendations do you have for Congress to mitigate how AI systems may contribute to extremism in the United States?

Dr. ACEMOGLU. Thank you for this important question as well. I think it is inevitable that digital tools are going to be used for spreading misinformation and disinformation. It cannot be stopped. But then again, the printing press was used for the same thing, the radio was used for the same thing, and lots of other vehicles were available to actors for fermenting extremism.

The issue is that AI and digital platforms in general increase the capabilities for bad actors to use these tools, and this is an obvious area for regulation. But more importantly, I think we have to ask questions about how the business models of the leading tech companies is playing out in this domain. Part of the reason for the phenomena that you are pointing out is that many of these digital platforms are actually not just displaying misinformation, but they are actively promoting it. I think displaying misinformation is very difficult to solve, but promoting is a choice. It is a choice that they

make because of their business model which is based on monetizing information via digital ads.

This is something that provides a lot of alternative directions for us. It is possible to use AI technologies in a way that is much more reliable in a way that does not create the most pernicious echo chambers in a way that does not promote misinformation and disinformation.

I think three types of policies are particularly important. Government regulation of where extremism is taking place and going after it is very important. The government has to invest more in tracking where this is happening, and I think your Committee is at the forefront of this.

Second, I think digital ad-based business models are creating a lot of negative social effects. My proposal has been for a while that we should consider digital ad taxes, meaning, taxes on advertisement that uses personalized ads collected from digital platforms. I think when we do that, we are both going to discourage, to some extent, the most pernicious users of these digital ads, but second, open up the market for alternative models.

The marriage of data collection, and venture capital, and other kinds of funding has created a business environment in the tech world where the most successful companies are those that try to collect as much data as possible, and try to get as much market share as possible for sometimes a decade or more.

They do not even make money, but they can get funding because this is the way of the future as viewed by venture capitalists. But that also means that alternative business models cannot enter because the market is being captured by these things. A meaningful digital ad tax would actually be a pro-competitive tool.

Then the final policy is data markets. Right now, a lot of this is also completely entangled with digital platforms being able to take data as they wish. I think we need to have better regulation about who has rights to data, and also perhaps start building legislation to create data markets in which, for example, creative data artists or writers have collective data ownership rights.

This way there will be a fairer division of the gains from digital technologies and AI, but also it could encourage better use of data and better ways of developing new monetization models in the digital age. Thank you.

Senator HASSAN. Thank you very much.

Thank you, Mr. Chair, for your indulgence.

Chairman PETERS. Thank you, Senator Hassan.

Senator Butler, you're recognized for your questions.

OPENING STATEMENT OF SENATOR BUTLER

Senator BUTLER. Thank you so much, Chair, and colleagues for helping us have more deep discussion about AI. This has been a topic that all of us have been talking about, it feels like in my short time, for a good deal of time. I appreciate all of you for your work and your leadership on the topic.

Dr. Hu, I think I will start with you, if that's OK. You have been doing an incredible amount of academic examination in this area. I understand that AI could become a critical asset to stakeholders in the criminal justice system.

However, we have already begun to see cutting edge artificial intelligence-based technology like facial recognition systems drive wrongful arrests of innocent people. The reality is that this technology is already widening pre-existing inequities by empowering systems that have long histories of racist and anti-activist surveillance.

Here's my question. I am curious to hear your thoughts, really, on how we can best use this sort of tension because without action, we know that that communities of color will disproportionately continue to disproportionately face the harmful consequences of this technology. I would love to hear your thoughts on how we can best respond, acknowledging that it is a technology that is going to exist, and we have these sort of built in inequities in our current system.

Ms. HU. Yes, thank you so much for that important question. I think that this is one of the critical inquiries that we are faced with when we are talking about AI. That it can, in the way in which it absorbs vast oceans of data, also absorb very historically problematic information and then translate that and interpret that in ways that are not consistent with our constitutional values or principles or civil rights.

This is particularly troubling in the field and in the way in which the technologies are being enrolled in criminal justice and criminal procedure because of our deep commitments to fairness in criminal justice and the ways in which we have those protections embedded in the Fourth, Fifth, Sixth Amendments, for example, of the Bill of Rights. How are we now faced with these types of evolutions in AI and these technologies and algorithmic decisionmaking in particular, in ways that we are having a hard time trying to preserve those fundamental constitutional rights.

I think that this is an opportunity for us to try to think through exactly what types of new jurisprudential methods and interpretations do we need in order to, for example, expand our interpretation of the Fourth Amendment in a way that encompasses these types of challenges so that we can stay true to our first principles of protections.

Senator BUTLER. Thank you so much for that.

Ms. Vallor, if I could turn to you quickly. I think the three practical recommendations in your written testimony to the Committee are very compelling. I was struck by the idea that just like the Dutch childcare benefits scandal, it is inevitable that we will get some of this stuff wrong, and even despite our best efforts.

Can you talk a little bit about why you think it is so important to create new systems of liability, contestability, and redress for impacted groups, which adjacent to my first question often includes the most vulnerable communities?

Dr. VALLOR. Absolutely. Thank you for that important question. I think we have seen plenty of evidence that if we do nothing, the use of AI technologies will continue to disproportionately harm the most vulnerable and marginalized communities here in this country and also in other countries. As you noted, it has been seen in multiple places in the world where this dynamic occurs.

A researcher in our field, Abeba Birhane, has described these technologies as conservative, not in the political sense, but in the

sense that they take patterns of the past and they literally conserve them and push them into the present and the future. They make it harder for us to overcome some of the patterns of the past that we are rightly committed to addressing.

We have to direct AI then as a tool for identifying harmful patterns, harmful biases, and mitigating those. AI can be used as a tool for that as well and has been in many cases. It comes down to who absorbs the risks that new technologies inevitably introduce. New technology can be completely safe or risk-free, but it's about who absorbs those risks, and who reaps the benefits.

When you allow large companies and wealthy investors to reap the benefits of innovation in ways that push all the risk and cost of that process onto society, and particularly onto the most vulnerable members of society, as we are seeing today, you produce an engine of accelerating inequality and accelerating injustice.

What Congress needs to do is to ensure that those who stand to profit the most from innovation are asked to take on most of those costs and risks. We have done this before in areas like environmental regulation with the Polluter Pays Principle, right? When it's implemented correctly, it actually incentivizes for profit companies to build safety and responsibility into their operations so that instead of spending money to have to clean up pollution, they can spend money to make their operations cleaner and safer in the first place.

I would love to see that dynamic be pursued in AI regulation as well, where we think about how can we incentivize companies to build more responsibly in the first place. I think we can, obviously, begin where we already have some power, and that is something that has come out of bills in this Committee to address the uses of AI in the public sector, to address uses by Federal agencies.

You see also in the Executive Order (EO) recently released many moves to empower and instruct Federal agencies to begin to take action so that we can in a way start by making sure that government uses of AI are appropriately governed, audited, monitored, and that the powers that government has to use AI are used to increase the opportunity, and equity, and justice in society rather than decrease it. Which can happen even when we do not intend it if we are not actually implementing many of the measures that I and other panelists here have described in the regulatory environment.

Senator BUTLER. Thank you.

Thank you, Mr. Chair.

Chairman PETERS. Thank you, Senator Butler.

Senator Hawley, you are recognized for your questions.

OPENING STATEMENT OF SENATOR HAWLEY

Senator HAWLEY. Thank you very much, Mr. Chair. Thanks to the witnesses for being here.

I want to start my time with a piece of oversight business, if I could. Last week when the Secretary of Homeland Security was here, Secretary Mayorkas, I asked him about a whistleblower claim. A whistleblower who had come forward to my office and alleged that as many as 600 security special agents from Homeland Security Investigations (HSI), 600, had been removed from felony

investigations, including particularly child exploitation investigations and sent to the Southern Border to do things like make sandwiches for illegal immigrants.

That's a quote from the whistleblower, not from me. Here is what she said. "We are being told to shut down investigations to go hand out sandwiches, and escort migrants to the shower, and sit with them while they are in the hospital, and those types of tasks.¹ Now, Secretary Mayorkas did not deny this. He did say that, well, they are working on fentanyl, or they may be working on fentanyl claims while they are at the border.

After that testimony, multiple additional whistleblowers came forward to my office from across the country, different whistleblowers unrelated to each other from different offices across the country, and directly contradicted Secretary Mayorkas' testimony. One whistleblower said Secretary Mayorkas was, and I am going to quote him now, "Absolutely lying," and that agents were not in fact being reassigned to investigate fentanyl cases.

Another whistleblower claimed that he was reassigned to the border to, in his words, "Babysit illegal immigrants." A fourth whistleblower confirmed that special agents had been pulled off child exploitation investigations, and all of these whistleblowers provided documentation about being asked to drop felony investigations, move to the Southern Border to conduct, essentially, ministerial tasks along the lines that the first whistleblower alleged.

Mr. Chair, of course, I do not personally know whether this is accurate or not. I know now we have multiple whistleblowers who are all alleging the same thing. They are also pointed out to me these whistleblowers that there may be violations of the law. In fact, the whistleblowers allege that these practices violate 31 USC 1301, that it violates the Office of Management and Budget (OMB) Circular A76, that it violates in internal U.S. Immigration and Customs Enforcement (ICE) travel policies.

What I have done, Mr. Chair, as per my normal practice and the practice that I think all of us follow on this Committee, I have collected this information. I have written a letter to the Inspector General (IG) of the Department of Homeland Security (DHS) asking his office to investigate these claims, which I am sharing with the Committee today, and I have asked him to report back to me and to the Committee so that we can see what he says. I would like to submit this for the record,² if I could, Mr. Chairman.

I want to thank you for your work always with whistleblowers, and for those who come forward to my office and other offices before. I am putting this on the record. We will see what he says. I hope that he will look into this, and he will get back to us and we can evaluate these claims. Thank you.

Chairman PETERS. Without objection.

Senator HAWLEY. Now, turning if I could, to you, Professor Acemoglu. Let me ask you a little bit about AI in the recruiting and hiring context. My understanding is that, increasingly, companies are using AI recruiting tools to what they would say enhance

¹ The quote referenced by Senator Hawley appears in the Appendix on page 84.

² The information referenced by Senator Hawley appears in the Appendix on page 85.

efficiency in the hiring process. This is especially true among large established companies.

My concern is this, is that AI's application to recruitment is often controversial because hiring is an inherently subjective process. We were just discussing, in fact, some of the issues when you use AI to make what we might call "people decisions" and to some of the biases that AI tends to scoop up and replicate. One example of this is Amazon in 2018, where it's reported that AI software used in the hiring process systematically discriminated against women.

My question to you is this, where would you draw the line on AI decisionmaking in hiring practices? What should we be aware of or concerned about there?

Dr. ACEMOGLU. Excellent question. Thank you very much, Senator Hawley. I think that's a very difficult question. I am very concerned about all users of AI that takes away human judgment, especially when human judgment is very important. This becomes particularly concerning when AI practices then legitimize things that were not previously completely accepted.

Let me give you an example. For instance, imagine that we have an AI system that puts a lot of weight on somebody having completed a 4-year college for essentially a semi-manual task. It is quite easy how that might come about. Four-year college workers are doing much better in the labor market. But in for many semi-manual tasks, those college skills are not that important. But if the AI system starts putting that emphasis, it is going to start turning a lot of good candidates down.

The more it does that the more it becomes accepted that you should really have a 4-year college to become an electrician. Then our social norms and our expectations completely shift, even though the original decision to turn down people who had just a high school degree was not right.

This is not a hypothetical situation, because we are having a lot of similar cases happen when AI systems are engaged in decision-making, especially when people do not know how to evaluate them. There's a lot of evidence, for example, that doctors who get AI recommendations do not know how to evaluate it, and they sometimes get confused where the AI recommendation comes from. They may put overweight on things that they should not really be overweight because they do not understand the blackbox nature of the system.

I think human judgment, and the expert opinion of people who have accumulated expertise is going to be very important. This is particularly true when we start using AI, not just for recruitment, but lots of other human resource tasks. For example, promotion, or deciding who's doing well, or how to assign workers to different shifts.

We are going to do much better, which if we do something broadly consistent with what I try to emphasize in my written testimony; choose a pro-human direction of AI. Which means that we try to choose the AI technologies trajectory in a way that empowers humans. Then we train the decisionmakers so that they have the right expertise to work with AI, and that includes valuing their own judgment not becoming slaves of the AI system. Thank you for that question.

Senator HAWLEY. Oh, very good, and your answer touches on something that I think is so important that we cannot lose sight of who has control of the AI, and who the AI is benefiting. I have said over and over, I am sure that these giant corporations who are developing AI, I am sure that they will make lots of money on it. I have no doubt about that. Will it be good for the people that they employ and in particular, will it be good for working people in this country? I am less certain about that.

I see my friend Senator Blumenthal across the dais, in a hearing that we had recently, I still remember the testimony of a large corporate executive who just remarked offhand that it was wonderful that AI was doing things like replacing people who work at fast food restaurants. I think he just expected everyone to agree because, of course, those are not creative tasks. It is good we can do without them.

I thought wait a minute, it is easy for you to say as you sit in your position in the C-suite, maybe not so much for the person for whom that is the first job that is getting a foothold in the labor market from which she can advance to something else.

I think who controls the AI and what the biases are in it in a way that you point out is very important. Thank you, Mr. Chairman.

Thank you, Senator Hawley.

Senator Blumenthal, you are recognized for your questions.

OPENING STATEMENT OF SENATOR BLUMENTHAL

Senator BLUMENTHAL. Thank you. I will just expand on the line of questioning that Senator Hawley was asking because he and I actually had been having hearings on the Judiciary Privacy, Technology and the Law Subcommittee. The labor aspects have been perhaps less important for us than preventing the deep fakes and impersonations. We have developed a framework for legislation, including a licensing regime with an oversight entity and testing before products are released so as to prevent the kind of deep fakes and impersonations that so scare people. At the same time, preserve the promise of AI that I think all of you and we too agree is very important.

But the impact on the labor market in terms of inequality, aggravating inequality, eliminating tasks without creating new tasks, I think is a very important point that you made, Professor. You say, and I know that you cite in footnote 5, a lot of studies that have been done on electrician and plumbers, could you make it real for us how can AI enhance the work done by electricians and plumbers? Then also, what new tasks can you give us an example of how AI could create new tasks so that it can be pro-worker, as you say, pro-citizen, pro-human?

Dr. ACEMOGLU. Thank you very much, Senator Blumenthal. Let me actually start giving a different example than the electrician, I will come to the electrician, educators, teachers. A lot of the emphasis today is to use AI in classrooms for automated grading, automated teaching, and also large language models that take the place of experts in forming students.

But actually, one of the problems in many U.S. schools is that a lot of students are falling behind. There is quite a bit of evidence

in the education science literature showing that personalized teaching is tremendously useful for the students. If we had the resources to have a teacher work with one or two students identifying their weaknesses, and how the material could be presented to them so that they could understand, they could have a chance to catch up. But they do not have those resources, they do not have those opportunities, so those students fall behind. That is part of our educational crisis right now.

One quite feasible direction of AI, it is actually well within the technological frontier, it does not even require any advances, is use of existing AI tools in real time to identify which students are having trouble with which part of the curriculum. You can do that actually as the class progresses, and then provide suggestions to teachers.

You would need more and better trained teachers to do that. But you provide suggestions to these teachers to say, let us take these two or three students and present the material differently, spend a little bit more time give some remedial help. That's the kind of system recommendation that AI can easily do. You can see here that the tasks that the teachers will start performing will become new tasks. The current educators, they teach to the class 30 people or something, they do not have this aspect of identifying and working one on one in a systematic way. Those would be the examples of new tasks.

Having given that example to the educators, I come to the electricians. It is exactly the same issue. Electricians are going to have more and more problems as, for example, electrification of the grid, or new electrical machinery. There are going to be more and more rare problems, more and more troubleshooting problems. Right now, even for the very regular issues that I have in my house, an electrician will come as a semi-skilled electrician, is not the very best. They will look at the problem and they cannot solve it, and they have to go and they have to do some research, and then another expert comes and they try to deal with these issues.

One way that you would make them much better, and this will help in general a lot of semi-skilled craftspeople is that real time AI tools would draw from the expertise of a lot more electricians with similar problems and would make them recommendations. They can do on the spot troubleshooting, problem solving, and deal with the more complex and new tasks that are going to emerge with the changing environment.

The benefit of that is not just it is going to help with our shortage of electricians, it is going to increase the earning capacity of electricians and the economy, but it is actually going to be an equalizing tool. Because who is going to benefit most from this? It is not going to be the very best electrician because he or she would have been able to solve these problems, it is going to be those with middle expertise who are good enough to do certain tasks, but they need help, they need training, they need additional recommendations to deal with the more complex problems.

That is where the promise lies. Thank you for your question.

Senator BLUMENTHAL. Thank you. That is a really helpful answer. Will that in turn, address the phenomenon of growing inequality in our system?

Dr. ACEMOGLU. I think it has a real chance of being a very contributing factor. It is not going to be sufficient by itself, but one of the major reasons for why we have so much inequality is that we have not helped low education workers. We have replaced their jobs exactly like Senator Hawley pointed out, and we have not given them new opportunities, and we have not given them new tools.

Those workers can become much more productive if we give them better technologies and better training opportunities. Again, AI has that capacity, especially generative AI. Forget the hype, I really think the hype is misleading. But there are some very impressive aspects of it. The most impressive one is that you can load on a tremendous amount of information, and then give some clues about a context, and it finds from that vast amount of information which bits are relevant for that context.

If we use that, we can really deploy it for making more helpful technologies for low education workers, for skilled craftsmen, for semi-skilled craftspeople, for service workers, for healthcare workers, for educators.

Senator BLUMENTHAL. That is a very exciting prospect. At the same time, you know, I guess there is good AI and less effective AI. I read an article recently about hallucination that said that there is a variation of three percent hallucination to 27 percent hallucination, depending on the system. I hope the plumber or electrician gets the more accurate version rather than 27 percent because they will be fired.

Dr. ACEMOGLU. That is actually a very important point. Right now, you could not use ChatGPT or similar models to do that exactly because they are not providing reliable information. This goes back to Senator Hawley's comment. These technologies are developed in a way that is good for the bottom line of the large companies, but not good for the workers or for the people.

That is actually very easy to deal with. If instead of training these models on the vast amount of unreliable information on Reddit and also lots of other places, if you give them reliable information so the training set of these models is much more reliable, then the information they will give us much more reliable.

Why are we training these models on the entire Internet and the speech patterns that you see on Twitter, Facebook, Reddit, and so on? Because the agenda of many of these companies was to create the hype that these are general intelligence-like technologies, and to do that they wanted to mimic the way that humans talk. The amount of information was not important. It was just important to get the human-like speech out of this.

So different agendas. One is good for the corporations, the other one is going to be good for the workers. I think this is where government leadership is going to be important.

Senator BLUMENTHAL. Thank you very much. Fascinating topic, and my time is expired. But there is a lot more to discuss, and appreciate all your work. Thanks, Mr. Chairman.

Chairman PETERS. Thank you, Senator Blumenthal.

Senator Ossoff, you are recognized for your questions.

OPENING STATEMENT OF SENATOR OSSOFF

Senator OSSOFF. Thank you, Mr. Chairman. thank you to our panelists for your testimony, and your expertise, and your work.

Obviously, there are and will be intersections between privacy law and privacy policy, and any regulatory regime established that touches on or manages the development and deployment of artificial intelligence.

Dr. Acemoglu, you mentioned in your statement suggestions about a property rights model for data Professor Hu you cited some of the work of Jack Balkin in your opening statement, who as I understand it, suggested a fiduciary model for data whereby custodians and recipients of data from persons would have inherent duties of care and confidentiality to the individuals whose data they have collected, and which they are storing or using.

When I think about the failure of Congress to make effective privacy law, one of the things we see is an effort to imitate the European Union's (EU's) regime. Another thing we see is a sort of Whac-A-Mole regulatory approach that looks at current problems faced by consumers and individuals and tries to isolate and target them with certain specific regulatory prohibitions, but doesn't seem to propose any kind of more basic law upon which fundamental obligations could be established that judges then over time could evolve into a more comprehensive regime protecting the privacy of individuals.

Professor Hu, if you could just opine for a moment on your thoughts on the notion of a fiduciary model as a means for establishing some fundamental obligations for software companies, Internet platforms, and others across the private sector and the public sector who will receive data from private individuals?

Ms. HU. Yes. Thank you so much for that excellent question. I do think that we are seeing a renegotiation of the social contract. This is where Jack Balkin's theory of the fiduciary model for privacy and the information fiduciaries in which you have, especially under First Amendment rights, an acknowledgement that you have a triangle of negotiation of constitutional First Amendment rights between the tech companies, the citizen, and the government. That new renegotiation of rights and obligations is the representation of our modern digital economy.

But I want to go back to your question about do we need something more fundamental. I think that this is where we are opening the dialog to potentially needing a constitutional amendment that enshrines privacy as a fundamental right. If we look at that as a launching pad in which to through at some type of constitutional amendment, empower Congress to enact legislation in order to try to ensure that fundamental privacy rights are extended to all citizens, then I think we do not need to see it as much of a negotiation in a triangle with the companies.

As we have heard from the other witnesses, we have to always ask the question of who is benefiting and how our data, for example, is being monetized in a way that is adverse to the best interests of the citizenry.

Senator OSSOFF. Thank you, Professor Hu. You know it is an intriguing proposition, of course, procedurally, in terms of the difficulty of the process. Such an amendment would require a tremen-

dous amount of effort. It is not to say it may not be worth the effort, but a statute. Although the record of Congress thus far enacting any kind of meaningful privacy statute is a failure, I think that there is an interest on both sides of the aisle in privacy law.

Dr. Acemoglu, could you comment, please, on your reaction to this proposal of a fiduciary model for the protection of data and how it contrasts with other sort of property rights regime, which you have suggested in your opening remarks?

Dr. ACEMOGLU. I think we just do not know which one of these different models is going to be most appropriate for the emerging data age. I think the fiduciary model has a lot of positives. That European Union's General Data Protection Regulation (GDPR) regulation, it was motivated by the right philosophy, but at the end, we are seeing that it has backfired. It is not very effective, and it may have actually advantaged some of the large companies because they are better able to bear the costs of complying with the regulation.

I think the general point that we should bear in mind is data, and who controls data is going to become more and more important. It has become one of the major reasons why the tech sector has become more oligopolized, because a few companies have a big advantage in controlling data.

So privacy issues are very important as Professor Hu also mentioned, which is privacy is a right. But I think right now they are completely intersecting about who controls data, and that is the reason why I think I am tempted to favor models in which we try to systematize data markets.

At the end of the day, if data is going to become the lifeblood of the new AI economy, it is not going to be OK to treat data as an afterthought to solve privacy issues. We really need to institute the right sort of regulations and legislation about what rights people have to different types of data that they have created, and whether those rights are going to be exercised individually or collectively.

That is actually a very tricky, but new issue. The most natural thing for economists, and I think for policymakers, is to say, "OK. We are going to create property rights on data. So you own your data." That would not become a very workable model, both because it will be very expensive for individuals to track whether their data is being used, but there are also lots of market-driven reasons for why individual data rights may not work. After all you may find, my data is about identifying cats, other people can do that as well as me. That creates a lot of race to the bottom. So you may need some sort of collective ownership of data.

Senator OSSOFF. With my remaining time, Dr. Acemoglu, you have also talked a lot about centralization. The development of these frontier models is very energy intensive, technology intensive. This is IP produced at great cost, and there are few entities in the world with access to the processing power to do it.

Just comment, if you could, on the risks of centralization, of ownership of such models, and what kind of policy remedies might be available to Congress, if they are necessary at all, in order to prevent some of the negative consequences of such centralization and cut market concentration.

Dr. ACEMOGLU. Again, this is an area we just do not know enough about because there are some people who think open source is going to be a sufficiently powerful check on the power of the largest tech companies. On the other hand, there is a lot of doubt about whether open source is going to work. I think the most important issue is exactly like you have pointed out; there are two resources that are very centralized at the moment which is compute power, which is becoming more and more expensive because there is a shortage of the compute power at the moment, and second is data.

Bth of these are going to create potentially a much more monopolized system which is not good for innovation. It is not good for the direction of innovation because then it is going to be just a few companies that set the agenda.

I think antitrust tools are a very effective one. I am talking as much about stopping mergers and acquisitions. If you look at over the last 10 years, the largest tech companies have acquired dozens of rivals. Often, they actually sideline the technologies of those rivals because they just do not want the competition.

The second thing is to create the data infrastructure that I was talking about that is going to be a channel to create more competition. Then the final one that I think we should think about is whether there are reasons for the government to get involved in the allocation of compute power. If it becomes more and more scarce and it is a critical resource, especially if a critical resource from a national security point of view, I think the government may need to worry about where that compute power is going. Thank you.

Chairman PETERS. Thank you, Senator Ossoff.

Last question for all three of you. We will start with Professor Vallor, and we will work this way just to change things up a little bit here. Another really broad question. We have talked about a variety of issues here today at the Committee. There is a huge conversation going on across the world right now, but what do you think is missing from the AI conversation right now?

I want to be specific to governments and lawmakers. To those of us sitting up here who are thinking about how we deal with these issues. Is there something missing from the conversation that you really think we should be thinking about?

Professor Vallor, you are going to be the first shot at that, and then we will work down the dais.

Dr. VALLOR. Thanks for that question. I think one of the things that is not entirely missing, but I think it is underemphasized in the current regulatory conversation is the ability to see these systems as governable, as opposed to things that are being thrust upon us. As if they are natural forces that have arrived.

Every AI technology has been shaped by human decisions to serve particular ends, and driven by particular incentives that our systems have set. I do not think we talk enough about the incentives that we have created for some of the harms that we are seeing perpetuating and accelerating across the AI ecosystem, both profit incentives and power incentives, and where those can be changed.

I also think we are not talking still enough about ensuring that when we use AI, that we are using it to amplify human intelligence rather than becoming a cheap replacement for it. AI tools, as have been mentioned, are overhyped, not because they are not powerful, but because their power is of a different sort than the people who market them want us to believe.

These tools are not intelligent in the way that we are. They do not have the capacity for good judgment and common sense. They are very powerful pattern amplifiers, and that is a tool that we can use in our decisions. But many people are still talking about AI as if it is going to be able to make the hard decisions for us. These systems do not make decisions. They make calculations. Even if we automate a process by letting a computer determine the outcome, that is a human decision.

I do not think we are going to be served well if we forget how to make decisions well or even lose sight of the fact that we are still the ones making the decisions no matter how much we are relying on the technology. Because in that case, we are making decisions in the dark, and that is a terrible strategy for human progress.

Chairman PETERS. Thank you, Professor. Professor Hu.

Ms. HU. Thank you so much, Mr. Chair for that question. I think part of what is missing from the discussion is whether or not a fundamental assumption is being challenged by AI. That assumption is that the rule of law can precede all other forms of power, and that the law can govern effectively, especially if you have these tech companies and AI seeing themselves as co-equals and being able to speak with the law as an equal. Therefore, you might have a difficulty of having then the AI, apparently, in some instances, being presented as something that can now precede the law.

If we are going to, I think, really address how the law will govern AI, I think we need to understand that that is the fundamental question. Under our constitutional democracy under Article 1 of the Constitution, it gave Congress the power to legislate. But how is AI trying to challenge that power?

Chairman PETERS. Thank you. Professor.

Dr. ACEMOGLU. I think I have been emphasizing this for a while, but it is still, I believe, underappreciated that there are different directions in which we can develop AI tools, and that we have to make the right choices understanding what these different directions are.

I am an economist. As I responded to Senator Hawley, of course, the profit motive and the benefits to corporations matter. Those are very important. But I think we are underestimating how much the founding ideology or vision of AI has influenced the way that the industry has developed.

That founding vision, as I have argued in a number of different contexts, is that we have a desire or a social benefit from creating autonomous machine intelligence. Meaning, machines that are as intelligent as humans, and they are autonomous. If once we do that, a number of conclusions follow from that.

One, is a lot of automation, because if machines are really intelligent and autonomous, then they should do a lot of the tasks that we do because they can perform them as well as we can do. Second,

much less need for human judgment because they are autonomous and intelligent. Third, much less emphasis on humans actually controlling them.

But this vision has become completely foundational to the tech industry. A lot of the emphasis on general intelligence follows from that. I think it is very difficult to change the direction of the tech industry, with only regulation, unless we also cultivate different types of priorities among tech leaders, and the leading engineers, and computer scientists.

That is why I have emphasized, not just my work, but many important people. Professor Vallor also emphasized for example, Norbert Wiener, and many other inspiring scientists, as early as 1949 and 1950s, came up with different visions, but they have been completely overshadowed by the artificial general intelligence or autonomous machine intelligence vision.

Putting that on the table, encouraging a broader perspective on AI and encouraging or articulating the idea that having pro-human AI is both feasible and desirable is both missing, and I think quite important for the future of this industry. Thank you.

Chairman PETERS. Thank you. Another question for all three of you. This time, we will start with Professor Hu. We will mix it up here, then Professor Vallor, and Daron. This is a tough question, but given the fact that how complex this issue is and all of the issues that we have talked about, but as lawmakers we have to distill things down to concrete actions that we need to take.

If the United States government can do just one thing, one thing to increase the chances that AI is going to increase the well-being for everyone, not just a few, but for everyone, what would that one thing be given your areas of expertise? Professor Hu.

Ms. HU. I think the one thing that I would prioritize is an amendment to the Civil Rights Act of 1964, so that we incorporate and then try to anticipate and address the types of AI-driven civil rights concerns that we have seen over the last decade.

I think that we can see that across the spectrum of the ways in which the AI and automated systems and algorithmic decision-making can cut across discrimination in the criminal justice context, housing, mortgage financing, in employment. That would be the one thing that I would emphasize.

Chairman PETERS. Thank you. Professor Vallor.

Dr. VALLOR. I think I would emphasize examining the misaligned incentives that we have permitted in the AI ecosystem, particularly with the largest and most powerful players, and learn the lessons from the past where we have had success realigning the incentives of innovation with the public interest so that we can create clear and compelling penalties for companies who innovate irresponsibly, for companies that get it wrong because they have not put in the work to get it right.

While at the same time, perhaps capping the liabilities or reducing the risk for innovators who do invest in innovating safely and responsibly, and then want to find new ways of using those tools to benefit humans. Because we often see some of the good actors are hearing about the risks of AI systems, the ways that they might fabricate falsehoods, or the way that they may amplify bias,

and that can actually reduce innovation and narrow it to only those powerful actors who can afford to get it wrong.

I think if we adjust those incentives, so that the best and most innovative actors in the ecosystem are rewarded for innovating responsibly, and the most powerful ones have to be held liable for producing harms at scale, then I think we can see a way forward that looks much more positive for AI.

Chairman PETERS. Thank you. Professor Acemoglu.

Dr. ACEMOGLU. Thank you for this question. There is no silver bullet. But I think one of the first steps is to redress the fact that the vision of AI is pushing us more and more toward automation, surveillance, and monitoring.

This is really an ecosystem. Senator Johnson pointed out the Eisenhower quote that government support for university scientists could have negative consequences because it makes scientists cater to the government needs.

Right now, it is actually much worse when it comes to AI. All leading computer scientists and AI scientists in leading universities are funded and get generous support from AI companies, and the leading digital platforms.

It really creates an ecosystem in academia, as well as in the industry, where incentives are very much aligned toward pushing more and more for bigger and bigger models, and more and more of this machine intelligence vision and trying to automate a lot of work.

I think if we want to have a fighting chance for an alternative, the government may need to invest in a new Federal agency which is tasked with doing the same things that the U.S. Government used to do, for example, with DARPA or with other agencies of playing a leadership for new technologies. In this instance, that would be more pro-worker, pro-citizen agenda.

I think something along the lines of, for example, the National Institutes of Health (NIH), which has both expertise and funding for new research, could be very necessary for the field of AI with an explicit aim of investing into things that are falling by the wayside to more pro-human pro-worker pro-citizen directions. Thank you.

Chairman PETERS. Thank you.

Professor Vallor, you have written quite a bit about the connection between technology and human values. Would you share some concrete examples of this connection, and in particular, talk about but based on your research, how you see AI changing our basic societal values?

Dr. VALLOR. Sure, thank you. First of all, I think it is important to recognize, and you mentioned this in your opening remarks, that AI is not neutral. In fact, no technology is neutral. All technologies are mirrors of human values. Every technology that human beings have ever created has been a reflection of what humans at particular times and places that was worth doing, or enabling, or building, or trying.

But technologies also change what we value. If we think of these kinds of AI systems that we are building today trained on human-generated data that reflects the patterns of our own behaviors and past judgments, we are using AI much like a mirror. We are look-

ing at AI increasingly to tell us what we value to reflect the patterns and preferences, and to instruct us on the patterns and preferences that we and others hold.

This, to me creates a very perverse relationship between technology and values. Because instead of our most fundamental human values, the things that are connected most deeply to our need for shared human flourishing, instead of those driving the tools that we need—and this is to Professor Acemoglu's point of there is so much untapped opportunity to direct AI to address unmet needs and areas from health, to infrastructure, to the environment. But that is only if the values that are connected for us to shared human flourishing are what are driving those decisions.

Instead, what is happening, and I mentioned this earlier in my testimony, is that we are looking at a mirror of ourselves with these systems that actually reflects very old patterns of historical valuation, very old prejudices, very old biases and priorities, that do not in fact, reflect the values of the human family as a whole.

I think it is partly about being able to recognize what our values are without having to find them in the AI mirror. In that way, we can ensure that the technology continues to be shaped by the values that we hold most deeply.

Chairman PETERS. Thank you, Professor.

Professor Acemoglu, a question, do you believe or would you argue that AI is either causing or going to cause increased dysfunction in government? How would we manage that? I think you have written on some of these areas.

Dr. ACEMOGLU. I do not think right now AI is causing increased dysfunction in the government yet, except that I think we are falling behind the necessary regulation and building of the necessary expertise in AI in the government. It is wonderful to see this and several other Senate committees deal with the issues of AI because I think the lawmakers need to be at the forefront of it.

But as we move forward, and AI systems become more widely used, exactly like my fellow witnesses have pointed out, we need to introduce the rights safeguards for making sure that individual rights, including privacy rights, but more importantly, human and civil rights are correctly recognized and protected.

I do not see huge issues there in the United States at the moment, but there are a few local law enforcement agencies that started using systems that are not very reliable for law enforcement. That needs to be brought under control.

But you can see from China and other countries how the emphasis on surveillance and monitoring is already having a tremendous effect. It is particularly important for democratic countries to set the right legislation to ensure that both companies and government agencies are not tempted to follow China in the use of AI in the next decade.

Dr. VALLOR. Can I just briefly add to that?

Chairman PETERS. Please, go.

Dr. VALLOR. Just pointing out that that dynamic is an excellent example of how AI can in fact warp our human values because it can cause us to become increasingly resigned to control and efficiency as values that become more accepted and important than

particular liberties and considerations of justice that are inscribed in our Constitution and in international law.

I think it is also important to remain anchored in those value commitments that are written in those documents for a reason and ensure that we are not letting the direction of the technology that is currently ungoverned, undermine those very commitments.

Chairman PETERS. Thank you.

Professor Hu, you have talked about changing the social contract, do you want to talk about that in relation to the—

Ms. HU. Yes, absolutely. I think that what we are seeing is really a quadrilateral situation with our social contract where the rights are being mediated and negotiated across the spectrum of not just the government and citizens as it used to be when we first established the social contract. But now, it is negotiated across the government, citizens, civil society, the tech companies, and then the AI as the fourth vertices. I think that we need to think through whether or not that type of negotiation and mediation is consistent even with our constitutional democracy at all in the first instance.

Chairman PETERS. Thank you.

Senator Rosen, you recognized for your questions.

OPENING STATEMENT OF SENATOR ROSEN

Senator ROSEN. Thank you, Chair Peters. This really is an important hearing, and thank you all for being here today.

Lo and behold, you set me up perfectly for my question. We are going to follow-up about prioritizing values in AI because I am a former computer programmer, and systems analysts, and so I understand how evolving technology can revolutionize how Americans work. But as you have already been talking about, in all technology there are traces of human values, whether how it is used, or the math behind it, many large language models (LLM).

So human bias and human values, they are baked into the system in some form or fashion. Some LLMs we know perform better under pressure. For example, when a user tells a model that their job is at risk, or people will be hurt if a certain decision is made, that is one thing. But those same human values can make large language models more fallible and easier to be manipulated.

I am going to go to Dr. Vallor, because in one recent study, we found it was easier to evade an AI's system safety mechanisms when the system thought the stakes were higher. You have talked about this with Senator Peters' question, what should we consider when balancing values like efficiency versus accuracy? In what context? Should more accuracy be required from the model than efficiency, and vice versa?

Dr. VALLOR. That is a great question. The answer is one that, I think, highlights the need to invest more in the kind of interdisciplinary expertise around these systems that is needed to make these kinds of decisions wisely. Because whether, for example, efficiency, or accuracy matters more depends entirely on the sociotechnical context that you are designing and deploying the system in.

If you are using an AI system to make high stakes, irreversible decisions, right, where it is life or death and there is no undoing it if you get it wrong, then very clearly accuracy becomes a far

more vital priority than efficiency. In a lower stakes environment where what you are trying to do is simply automate a process in a way that actually uses resources in the most efficient way so that you do not have a lot of waste, which is something, obviously, that from an environmental standpoint is of great urgency, right, then accuracy in that case, perhaps matters less than the efficiency with which the system can drive the operation.

But one of the things that we have not talked about although I think it is in the background, is that AI is not just large language models. AI is not just even machine learning, right, AI is a host of many different kinds of technologies that are suited for different purposes and that work well in some environments and not in others.

What I think what we really need to see more investment in is the kind of combined technical, and social, and political, and legal expertise such that people understand the capabilities of the systems and their effects at the same time.

Right now what we have, is we have a lot of very smart technologists who understand very well the emerging capabilities of the systems they are building, but have a very limited view of the contexts in which they are being deployed and the likely consequences. On the other hand, you have a lot of social scientists, and humanities researchers, and law scholars who understand those contexts deeply, but are often not given the opportunity to/or have not invested enough time themselves in understanding the technical capabilities of the systems that they are proposing regulating or governing.

But at the University of Edinburgh and lots of other programs around the world, we are seeing more of an investment in that kind of interdisciplinary expertise that will be needed to govern these systems well, and perhaps this is something that Congress can accelerate as well.

Senator ROSEN. It is not enough to have a brain, maybe you have to have heart as well. That is when you marry both together in the easiest of ways. It is much more complicated than that, for sure. I have a few minutes left out, I want to move from this. We could talk about this all day because I really think this is the fear and the future of AI. This is the fear that we have; will it have a heart, right? There will be really plenty smart, but we need more.

But I want to move on to international cooperation because that is also really important. Last week, multiple countries, including China, India, Japan, they signed their first international agreement regarding AI, which is committed to deploy and develop AI in a safe and responsible way.

But this agreement, of course, is very broad and only in the first steps. Dr. Acemoglu, in the United States, we hold ideals like equality, democracy, and freedom, values that are not always shared by our foreign adversaries. That is why “safe and responsible” is in the eye of the beholder, perhaps.

How do we ensure that international standards hold these key values? Again, we are talking about values, the same as the last question, prioritize them when we set standards of allowable AI uses.

Dr. ACEMOGLU. That is a very important issue, and I do not think we can control what China will do directly. But we can form alliances with our allies and neutral countries for developing the right ethical standards. But both for this type of relationship with friendly countries, but even for where the direction of innovation will go in China and Russia, I think, U.S. scientific leadership is really important.

The same sort of concerns that people are voicing about AI right now, were also voiced when it came to mitigating climate change. The key was we can try to fight climate change in Europe and the United States, but China and India won't. But what we have seen over the last decade is that when the world led by Western countries, but some others as well, invests in renewables, that renewables become attractive in China and India as well.

I think it is the same thing. If we let China set the leadership, AI is going to go more and more in direction that takes control and agency from humans; it does not value equality, and it emphasizes surveillance, monitoring, and censorship. But if we can take that leadership and push it in a direction that's much more pro-human, supporting of democracy, supporting of equality, I think it will even have beneficial effects on China.

Senator ROSEN. Mr. Chair, can I ask one last thing, a follow-up on this. Because we talked about the human and the values, and there is a workforce transition that we are grappling with, and Dr. you have been talking about pro-worker use of AI, could you explain to us here in Congress what you mean by that because all of this does not happen on its own, and we have to prepare our workforce all across the spectrum. Could you just as a finish up just explain that to us?

Dr. ACEMOGLU. Absolutely.

I think it is very important. If I may, I will go back to the values discussion that Professor Vallor and Senator Peters had, and you contributed as well. The values that we are going to have for AI really depends on the direction in which we use AI and who controls AI.

If AI is in the hands of a few elite entrepreneurs and tech barons, that's going to be a very different direction of AI. We want to put the resources to help low skilled workers. The more we do not put the resources to train them to provide them with the technologies and the knowledge, the more they will appear to us that they are useless. That will centralize AI, that will centralize all of the resources.

Of the pro-worker agenda is using these AI tools to help the workers and provide them with better information. I think the capabilities that we have with the even existing knowledge in AI research is we can provide much better information so that educators become better decisionmakers. Skilled craftspeople like electricians, plumbers become decisionmakers. Healthcare workers, nurses become better decisionmakers. Blue collar workers can be much better. We have seen in a few companies how the right type of augmented reality in AI tools using to train them in precision work can increase their productivity. There is a lot we can do for the workers, and there is a lot that we can do for democracy.

Senator ROSEN. Thank you. That is a great way to finish. I appreciate it.

Chairman PETERS. Thank you, Senator Rosen.

I would like to take this opportunity to thank our witnesses. Thank you for being here. We are certainly all grateful for the work you do and for your contribution to this very important conversation. Today's hearing offered a new perspective on artificial intelligence. Our witnesses helped us step back from some of the exciting developments and the hype of AI, and consider historical, ethical, and philosophical questions that this technology possesses.

In order to ensure that AI truly works on behalf of the American people, we cannot stop here. We must continue this deeper examination of our values our agency and the future that we want from this technology as we build it in the years ahead.

The record for this hearing will remain open for 15 days until 5 p.m. on November 23, 2023, for the submission of statements and questions for the record.

This hearing is now adjourned.

[Whereupon, at 11:21 a.m., the hearing was adjourned.]

A P P E N D I X

Chairman Peters Opening Statement As Prepared for Delivery Full Committee Hearing: A.I. Philosophy November 8, 2023

The Committee will come to order.

We are living through one of the most exciting times in human history – a moment of unprecedented technological change. Artificial intelligence becomes more advanced every day. These tools have the capacity to revolutionize medicine, expand the frontiers of scientific research, ease the burdens of physical work, and create new instruments of art and culture. A.I. has the potential to transform our world for the better.

But these technologies also bring new risks to our democracy, civil liberties, and even our human agency. As we shape and regulate A.I., we cannot be blinded by its potential for good. We must also understand how it will shape us – and prepare for the challenges these tools could bring.

Some of that work will be accomplished with innovative policy. I am proud to have convened prior hearings that explore the safety risks of A.I., federal procurement of these tools, and how to prepare our workforce to properly utilize them.

But as policymakers, we also have to explore the broader ideas surrounding this technology. We have to examine the historical, ethical, and philosophical questions that it raises. Today's hearing – and our panel of witnesses – gives us an opportunity to do just that.

This is not the first time that humans have developed staggering new innovations. Such moments in history have not just made our technologies more advanced. They affected our politics, influenced our culture, and changed the fabric of our society.

The Industrial Revolution provides is one key example that phenomenon. During that era, humans invented new tools that drastically changed our capacity to make things. The means of mass production spread around the world, and allowed us to usher in the modern manufacturing economy.

But that era brought with it new challenges. It led to concerns about monopolies worker safety, unfair wages, and child labor. It produced the weapons that were used to fight two World Wars. In short, it wasn't just about the benefits of new technology.

I'm grateful that our first witnesses – Daron Acemoglu – has studied this phenomenon. He has not only examined the history of technological change – but also the democratic institutions that are needed in response.

In the 20th century, we had trade unions to protect worker's rights and effective government regulation to keep these new industries in check. What tools do we need to meet our moment? What else should we learn from this history?

Artificial intelligence also brings unique challenges. The history of technological change has largely centered on human strength and how we can augment it through the use of new machines.

A.I. will affect physical work. But it is more directly tied to our intellectual and cultural capacities. It has already introduced new ways to ask and answer questions, synthesize information, conduct research, and make art.

Those qualities – the ability to understand ideas and create culture – are the foundation of our humanity. We must work to preserve them as they become influenced by artificial tools.

Perhaps most importantly, A.I.’s influence on those capacities is not neutral. These tools – like the humans who make them – are biased. We must define what values lie at the core of our human experience – and create technological tools that support them.

Our second witness – Shannon Vallor – will be a helpful resource in understanding these ethical questions. She studies the way that new technologies reshape our habits, practices, and moral character.

With her help, we can understand the values embedded in these technologies, and the effect they will have on our human character.

Finally, we will explore A.I. through a constitutional law framework. A.I. poses a risk to our civil liberties. New surveillance tools can be used to target vulnerable communities. Biometric systems like facial recognition can endanger a citizen’s right to due process. Advanced technology brings renewed questions about our privacy and personal information. A.I. tools can also pose a grave danger to our democracy and civic institutions.

Our third witness – Margaret Hu – will help us understand these intersections. She researches the risks that A.I. poses to constitutional rights, due process, and civil liberties.

Artificial intelligence has already begun to reshape the fabric of our society. Our response cannot come through piecemeal policy alone, or isolated technological “fixes.” It must include a deeper examination of our history, our democracy, and our values – and how we want this technology to shape our future.

We must look to the past and learn the lessons of previous technological revolutions. We must answer the ethical questions that A.I. poses and use these new technologies to build a world where humans can thrive. And we must protect our civil liberties and democratic institutions against the risks that these tools can pose.

This hearing provides an excellent opportunity to focus on this work. I’d like to thank our witnesses for joining us today, and I look forward to hearing your testimony.

Written Testimony

Hearing on “The Philosophy of AI: Learning from History, Shaping Our Future”

Senate Committee on Homeland Security and Governmental Affairs

Daron Acemoglu

Institute Professor, MIT

November 8, 2023

Chairman Peters, Ranking Member Paul, and Members of the Committee,

Thank you for inviting me to testify today on this important subject. I will start with a summary of my main points and then expand upon them in the rest of the testimony.

Executive Summary: Digital technologies have already ushered in a multifaceted economic, social and political transformation. Artificial intelligence (AI) promises to amplify these epochal changes, for good and bad. Although these tools have tremendous potential to expand our production, communication, and informational capabilities, they also pose major risks to economic prosperity, social cohesion, democracy, and national security — as did many other transformative technologies in the past.

These risks are rooted in three related social changes: (1) economic shifts, especially greater inequality, brought about by new technologies can create social and political tensions; (2) digital tools, including AI, alter who controls information and how that information can be used and manipulated, with direct implications for political behavior and democracy; (3) these technologies also unleash myriad social changes, affecting aspirations and norms, with potentially far-reaching effects. All of these risks apply to both democracy and national security. It is critical to understand them, learn from history about when humanity has and has not managed to develop institutions and norms to deal with similar risks, and chart a clear-eyed regulatory course to guard against the worst eventualities.

My overarching argument is that there is a pro-human (meaning pro-worker and pro-citizen) direction for AI tools that would be much better both for shared prosperity and for democracy, and therefore for national security. We need to take AI risks seriously because—although a pro-human direction for AI could strengthen prosperity, democracy, and security—we are currently on a very different and worrying trajectory.

The AI Revolution: The Possibility of a Pro-Human Direction

By the 2010s, digital technologies had already fundamentally changed our economy and communication, and AI has further amplified these trends. Computers expand our capacity for computation and information processing. When their capabilities are applied to physical processes, they can enable both automation of tasks previously performed by humans (for example, numerically controlled machinery for manufacturing and software systems for office work) and the creation of new tasks for workers (for example, computer-assisted design). They also change how we collect and process information.

Basic computers are pre-programmed with step-by-step instructions to perform basic operations. This keeps their applications limited to routine tasks (that can be performed in a repetitive manner within a stable, predictable environment). AI expands these capabilities by providing greater flexibility, both because not everything needs to be pre-programmed (the learning aspect of AI) and because non-routine tasks can also be tackled by these tools. This flexibility not only amplifies the possibilities for automation, but could also enable better human decision-making and the capacity to perform new and more complex tasks.¹

To improve human performance, we need to think beyond creating AI systems that aim to acquire “higher-level capabilities” or “artificial general intelligence” (AGI) whereby machines can achieve parity with, and even surpass, humans in all mental tasks. Leaving aside whether AGI is even feasible within the next several decades (which, I believe, it is not), my argument is that the emphasis on general intelligence distracts us from the more beneficial uses of digital

¹ Russell, Stuart and Peter Norvig (2009) *Artificial Intelligence: A Modern Approach*, Pearson; Susskind, Daniel (2021) “Technological Unemployment” from *The Oxford Handbook of AI Governance*. Oxford University Press (forthcoming): <https://www.danielsusskind.com/s/Susskind-Handbook-Updated-21-September-2021.pdf>.

technologies to expand human capabilities by providing better information for human problem-solving and decision-making.

Making machines useful to humans and complementary to workers is not a new aspiration. Related ideas were articulated as early as 1949 by MIT mathematician and engineer Norbert Wiener, and put into practice by several leading computer scientists and engineers, such as Douglas Engelbart, inventor of technologies like the computer mouse and hypertext.² Such human-complementary tools, as well as new human tasks and appropriate worker training in the manufacturing and service sectors, were foundational to rapid wage growth and shared prosperity throughout the 1950s, 1960s and early 1970s. Nevertheless, this “machine usefulness” agenda has often been overshadowed by other priorities in today’s tech industry, including a strong focus on automation and monetization models based on data collection and digital ads.

Machine usefulness may be even more promising today than in the past. Many of the routine tasks that workers previously performed have already been automated, so a large fraction of current jobs centrally depends on nonroutine problem-solving and decision-making tasks. Empowering workers to perform these tasks more effectively, and to accomplish even more sophisticated decision-making tasks, will require providing workers with better information and decision-support tools. Recent advances in AI, especially in generative AI tools, are particularly well-suited to this type of information provision.³ An irony of our current digital era is that information is abundant, but useful information is scarce. Generative AI can help by recognizing the relevant context and presenting information that is useful for problem-solving, human decision-making, and performance in new, more complex tasks. For example, an electrician can much more effectively diagnose rare problems and accomplish complex tasks when empowered with AI tools that present information and recommendations based on knowledge accumulated from similar cases. In essence, AI holds great potential for training and provision of useful

² Acemoglu, Daron and Simon Johnson (2023a) *Power and Progress: Our Thousand-Year Struggle over Technology and Prosperity*, Hachette Public Affairs [hereafter *Acemoglu and Johnson (2023a)*], Chapter 9.

³ Acemoglu, Daron, David Autor and Simon Johnson (2023) “Can We Have Pro-Worker AI? Choosing a Path of Machines in Service of Minds” MIT Shaping the Future of Work policy memo: <https://shapingwork.mit.edu/wp-content/uploads/2023/09/Pro-Worker-AI-Policy-Memo.pdf>.

information to support expert workers, such as educators, medical personnel, software developers, and other skilled craft workers such as electricians and plumbers.⁴

These potential pro-worker gains from AI are not hypothetical. Several studies already show that, if used to provide better information or successfully speed up the more-routine parts of tasks for workers, generative AI can provide a significant boost to productivity.⁵

There is an analogue to the pro-worker perspective when it comes to how AI tools are used for communication. What we might call a “pro-citizen” perspective would leverage these tools to provide better information to individuals about the political and social world, and enable them to participate in deliberations without manipulation or undue bias. The opposite approach would be to deploy these tools for surveillance, manipulation, manufacturing false conformity, and fomenting emotional outrage and extremism.

The evolution of social media in the 2010s illustrates the manipulative path, where much evidence shows that algorithmic tools were used to boost user engagement by intensifying echo chambers, spreading extremist or misleading content, and exploiting emotional triggers.⁶ Although there is much we do not know about the systemic political and social effects of social media, existing evidence suggests that they have contributed both to mental health problems and polarization in politics.

The alternative, pro-citizen direction was articulated by a number of optimistic takes on the Internet and social media in the early 2000s, which viewed these tools as expanding democracy and increasing the accountability of authoritarian rulers.⁷ While these hopes were

⁴ Acemoglu, Daron (2021) *Redesigning AI: Work, Democracy, and Justice in the Age of Automation*, Boston Review Forum [hereinafter Acemoglu 2021].

⁵ Brynjolfsson, Erik, Danielle Li and Lindsey Raymond (2023) “Generative AI at Work.” NBER Working Paper no. 31161; Noy, Shakked and Whitney Zhang (2023) “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence.” *Science*, 381(6654): 187–192; Peng, Sida, Eirini Kalliamvakou, Peter Cihon and Mert Demirel (2023) “The Impact of AI on Developer Productivity: Evidence from GitHub Copilot.” arXiv Working Paper no. 2302.06590.

⁶ Brady, William, Julian Wills, John Jost, Joshua Tucker and Jay Van Bavel (2017) “Emotion Shapes the Diffusion of Moralized Content in Social Networks” *PNAS* 114(28): 7313–7318; Braghieri, Luca, Ro’ee Levy and Alexey Makarin (2022), *American Economic Review* 112(11): 3660–3693; Wu, Tim (2016) *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*, PRH Knopf; Acemoglu, Daron, Asuman Ozdaglar and James Siderius (2023) “A Model of Online Misinformation,” *Review of Economic Studies* (forthcoming).

⁷ Acemoglu and Johnson (2023a), Chapter 10.

soon dashed, prototypes of more pro-citizen AI tools have been developed successfully, including at significant scale in Taiwan, where they are used to increase the transparency of government processes and services, encourage democratic participation, and enable citizen input on government policies.⁸

With the widely different possibilities that AI tools present, it becomes particularly important to understand which direction we are heading in today, and how different paths might impact our economy and democracy.

Computers, AI and the Economy

It is well known that the American economy has not generated shared prosperity over the last four decades. U.S. labor market inequality has surged since 1980. Many workers—especially men with only a high-school education or less—have experienced significant declines in their real earnings. Even men with a college degree have seen only limited gains. In the meantime, racial inequities have widened, and regional disparities have multiplied. These sweeping changes have many causes, ranging from globalization to the transformation of U.S. labor market institutions.⁹

These trends cannot be separated from the effects of computers. My research indicates that the most important cause of increasing inequality has been automation—the substitution of machines and algorithms for tasks previously performed by workers. Automation is not a recent phenomenon: for example, it was pervasive during the British Industrial Revolution and in the United States during the 1800s.

Two types of evidence illustrate the inequality implications of today's automation technologies, such as robotics and office software. First, in regions where industrial robots have been adopted faster, employment and wages have fallen more and inequality has increased.¹⁰

⁸ Tang, Audrey (2019) "A Strong Democracy Is a Digital Democracy," *New York Times*: <https://www.nytimes.com/2019/10/15/opinion/taiwan-digital-democracy.html>; Lanier, Jaron and E. Glen Weyl (2020) "How Civic Technology Can Help Stop a Pandemic," *Foreign Affairs*: <https://www.foreignaffairs.com/articles/asia/2020-03-20/how-civic-technology-can-help-stop-pandemic>.

⁹ Acemoglu, Daron and David Autor (2011) "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of Labor Economics*, Chapter 12, 4(B): 1043–1171; Autor, David (2019) "Work of the Past, Work of the Future," *AEA Papers and Proceedings*, 109: 1–32.

¹⁰ Acemoglu, Daron and Pascual Restrepo (2020) "Robots and Jobs: Evidence from U.S. Labor Markets," *Journal of Political Economy*, 128(6): 2188–2244.

Second, demographic groups employed to perform routine tasks in industries undergoing rapid automation have almost uniformly suffered large declines in their real wages. My work finds that automation accounts for more than half, and perhaps as much as three quarters, of the surge in U.S. wage inequality.¹¹ Other trends, such as offshoring and competition from cheaper Chinese imports, have also played a role, but they have been less important than automation.¹²

Automation has contributed to specific dimensions of economic disparities as well. It accounts for 80% of the sizable increase in the college wage premium.¹³ It has also widened the gap between Black and White Americans. Automation, together with globalization, has been a major driver of regional disparities.

Concurrently with the increases in labor market inequality, the gap between capital owners and labor has also widened in the United States. The share of national income distributed as capital (e.g., interest, profits, dividends, capital gains) has risen sharply since the 1990s, as the share to labor has fallen. Among capital owners and managers, the very top have done particularly well, so that the shares of the richest 1% and 0.1% of Americans in national income have skyrocketed.¹⁴ Automation appears to have been an important factor in these trends as well.

What about the benefits of automation? During the mechanization of agriculture and in the three decades following World War II, automation was rapid, and the U.S. economy created millions of good jobs and achieved broadly shared prosperity. The main difference is that new technologies throughout these eras not only automated existing tasks, but also created new tasks for workers. Henry Ford's breakthroughs in automobile manufacturing were due to the application of new electrical machinery that routinized and automated some manual low-skill and artisanal tasks. But they also introduced a range of new technical tasks and critical functions for blue-collar workers. Simultaneously, manufacturing became much more intensive in

¹¹ Acemoglu, Daron and Pascual Restrepo (2022) "Automation, Tasks and the Rise in U.S. Wage Inequality," *Econometrica*, 90(5): 1973–2016 [hereinafter *Acemoglu and Restrepo (2022)*].

¹² Autor, David, David Dorn, and Gordon Hanson (2013) "The China Syndrome: Local Labor Market Effects of Import Competition in the United States" *American Economic Review* 103(6): 2121–2168; Acemoglu, Daron, David Autor, David Dorn, Gordon Hanson and Brendan Price (2016) "Import Competition and the Great U.S. Employment Sag of the 2000s," *Journal of Labor Economics* 34(S1.2): S141–S198 [hereinafter *Acemoglu et al. (2016)*]; Acemoglu and Restrepo (2022).

¹³ Acemoglu and Restrepo (2022).

¹⁴ Piketty, Thomas and Emmanuel Saez (2003) "Income Inequality in the United States, 1913–1998," *Quarterly Journal of Economics* 118(1): 1–41; Moll, Benjamin, Lukasz Rachel and Pascual Restrepo (2022) *Econometrica* 90(6): 2645–2683.

information activities, including design, planning, inspection, quality control and a variety of back-office tasks. These new tasks were critical for the expansion of employment in the automobile industry and throughout the U.S. economy during this era.¹⁵

The same forces were foundational during the decades that followed World War II, which witnessed both robust productivity growth and rapid wage growth that was broadly shared across demographic groups. While automation of certain manufacturing and white-collar tasks was rapid during this growth episode, so was the introduction of a range of new worker tasks, which were supported by novel technologies and training. Put simply, the 1950s, 1960s and early 1970s ushered in rapid growth and shared prosperity because automation went hand-in-hand with new tasks that increased workers' contributions to the production process.¹⁶ There has been a breakdown of shared prosperity since the 1980s because digital technologies have been used predominantly for automation and insufficiently for creating new tasks.

It is not certain whether the automation focus of early computers and other digital technologies was because their capabilities made automation easier (but did not enable enough new task creation) or if, instead, there was an excessive focus on automation that neglected alternative, potential pro-worker uses of these technologies. My research suggests that the latter explanation is more likely. Two pieces of evidence support this interpretation. First, when incentives and priorities were aligned, computers were used in more pro-worker ways, and relatedly, cross-country differences indicate that when institutions and policies encourage more pro-worker outcomes, technologies respond as well.¹⁷ Second, despite the great promise of digital tools, productivity gains from computers and other digital technologies have been disappointing, and there is some evidence that this has been because of an overemphasis on automation.¹⁸

¹⁵ Acemoglu and Johnson (2023a), Chapter 7; Hounshell, David (1985) *From the American System to Mass Production, 1800–1932: The Development of Manufacturing Technology in the United States*, Johns Hopkins University Press [hereinafter Hounshell (1985)].

¹⁶ Acemoglu, Daron and Pascual Restrepo (2019) "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives*, 33(2): 3–30; Autor, David, Caroline Chin, Anna Salomons and Bryan Seegmiller (2022) "New Frontiers: The Origins and Content of New Work, 1940–2018" NBER Working Paper no. 30389.

¹⁷ Acemoglu and Johnson (2023a), Chapter 7; Hounshell (1985).

¹⁸ Acemoglu et al. (2016); Gordon, Robert (2016) *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*, Princeton University Press; Acemoglu, Daron, David Autor and Christina Patterson (2023) "Bottlenecks: Sectoral Imbalances and the U.S. Productivity Slowdown," NBER Macroeconomics Annual 2023(38).

The evidence I have presented so far pertains to pre-AI digital technologies. AI can be used in even more versatile ways than early computers and opens up new pathways for pro-worker technologies. Nevertheless, recent research also suggests that emerging AI applications are targeting automation. For instance, establishments with jobs and tasks that can be automated by AI have been at the forefront of AI adoption in the late 2010s and the same establishments have concurrently cut back on their hiring.¹⁹ There is also evidence suggesting that AI tools are often used for surveillance and intensive monitoring, which are costly for workers and do not generate much productivity growth.²⁰ The emphasis on surveillance is much more intense in China, triggered by the Communist Party's preoccupation with controlling the population's political beliefs and activities. The available evidence suggests that surveillance and censorship tools produced in China are spreading throughout the world.²¹

The evolution of the market structure for AI providers will matter for how these tools are incorporated into business applications and who captures the gains. Two very different paths appear possible. In one, open-source models and the entry of many small generative AI providers would create a fairly competitive environment.²² This is not a panacea against many of the problems I have pointed out, but would at least prevent monopolization of these tools, as well as monopoly control of the massive amounts of data that the tech industry is collecting. The alternative is a much more oligopolistic structure, where a few companies that have a lead in data collection and in building large, all-purpose generative AI models, such as GPT-4, will become dominant. Along this path, these leading companies would capture a disproportionate share of the profits from new AI tools. This oligopolistic model would intensify inequality (as a few companies earn most of the profits), and could increase the influence of these players on the direction of technology and how centralized control of information plays out (an issue I return to below).²³

¹⁹ Acemoglu, Daron, David Autor, Jonathon Hazell and Pascual Restrepo (2022) "Artificial Intelligence and Jobs: Evidence from Online Vacancies," *Journal of Labor Economics* 40(S1): S293-S340.

²⁰ Acemoglu and Johnson (2023a), Chapter 9; Hounshell (1985).

²¹ Acemoglu and Johnson (2023a), Chapters 9 and 10; Beraja, Martin, Andrew Kao, David Yang and Noam Yuchtman (2023) "Exporting the Surveillance State via Trade in AI," NBER Working Paper no. 31676.

²² Masiello, Betsy and Derek Slater (2023) "Will Open Source AI Shift Power from 'Big Tech'? It Depends," *Tech Policy Press*: <https://techpolicy.press/will-open-source-ai-shift-power-from-big-tech-it-depends/>.

²³ Acemoglu, Daron and Simon Johnson (2023b) "Big Tech Is Bad, Big AI Will Be Worse," *New York Times*: <https://www.nytimes.com/2023/06/09/opinion/ai-big-tech-microsoft-google-duopoly.html>.

In summary, digital technologies and AI present a range of different possibilities for our economic future. A pro-worker direction would leverage these tools to make human workers more productive and provide them with better information for problem-solving and critical decision-making. The alternative, using them for automation and surveillance, would not bring the same benefits for workers and would tend to boost inequality. The evidence we currently have suggests that digital automation has been at the root of the spectacular increase in U.S. inequality since 1980, and early AI technologies are going in the same direction of automation and surveillance.

Social and Democratic Implications of Inequality

A thesis dating back at least to Alexis de Tocqueville maintains that democracy depends on limiting social and economic inequality.²⁴ People need to participate in democratic decision-making as citizens, and this becomes more likely when they are active socially, when they believe that they have a chance to succeed economically, and when they are getting their fair share of the economic benefits. Economic crises often lead to political instability, and recent evidence from the United States and from around the world supports these concerns.

In the United States, areas that were most severely disrupted by the rise in Chinese competition, which led to business closures and greater joblessness, show a sharp increase in political polarization and support for more extremist political candidates.²⁵ These trends are visible through both changes in general attitudes and in voting behavior. There is a similar increase in polarization and support for more extremist candidates in places where the introduction of robots led to declines in employment and wages.²⁶

Additionally, my research across a large number of countries shows that high inequality undermines support for democracy among the population, and this lack of support makes democracies more unstable and less capable of dealing with adverse shocks.²⁷

²⁴ de Tocqueville, Alexis (1835) *Democracy in America*, Eds. Harvey Mansfield and Delba Winthrop (2000), University of Chicago Press.

²⁵ Autor, David, David Dorn, Gordon Hanson and Kaveh Majlesi (2020) "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure," *American Economic Review* 110(10): 3139–3183.

²⁶ Frey, Carl, Thor Berger and Chingih Chin (2018) "Political Machinery: Did Robots Swing the 2016 U.S. Presidential Election?" *Oxford Review of Economic Policy* 34(3): 418–442.

²⁷ Acemoglu, Daron, Nicolás Ajzenman, Cevat Aksoy, Martin Fiszbein and Carlos Molina (2023) "(Successful) Democracies Breed Their Own Support," *Review of Economic Studies* (forthcoming).

This body of work thus suggests that the inequality effects of digital automation have likely been a major factor in the rise of anti-democratic views in parts of the U.S. electorate and the more general political polarization of the country. By fueling extremism, these changes are raising new challenges for national security.

If AI deepens inequality, the dangers for U.S. democracy and national security could become intensified as well. The risks may be even more pronounced because several decades of non-shared prosperity could potentially have much bigger effects on grievances and discontent than shorter-term trends. This perspective suggests that it is particularly important to find a path for new AI tools that would help build shared prosperity, especially after the huge increases in inequality in the country over the last four decades.

Control over Information and the Future of Democracy

The survival of any political regime depends on the beliefs and values of its people and on how information is controlled and presented. Authoritarian rulers have understood this for ages. The top-down, highly repressive rule that the Qin Dynasty imposed in China after the Warring States period in third century B.C. was accompanied by much more intense control over information, including book burnings. The Qin rulers were sufficiently worried about information spreading in the population that they are reported to have executed the people who wrote the books, as well as those who could rewrite them.²⁸

We also see the importance of control of information, and its breakdown, in the history of Europe. The printing press, which broke the monopoly of political and religious elites on information, was not just critical for the Protestant Reformation, but was a direct contributor to the momentous political changes that followed. Pamphlets, newspapers and books that could be quickly printed and disseminated were important in the English Civil War and the run-up to the French Revolution. Many rulers, including Thomas Cromwell in his efforts to promote Henry

²⁸ Chan, Lois Mai (1972) "The Burning of the Books in China, 213 B.C.", *Journal of Library History*, 7(2): 101-108.

VIII's centralization project and Napoleon Bonaparte in his efforts to inculcate the French identity, also used the printing press and control over information that it enabled.²⁹

As information tools, computers have opened myriad possibilities for citizens to access information and communicate. They have also created new pathways for surveillance and centralized control of information. These trends are clearly visible in Iran, Russia and China. For example, Internet censorship and digital surveillance have become the bulwark of the Chinese Communist Party's rule over the last two decades. Part of the reason why early hopes about the democratizing effects of social media and the Internet have been dashed is precisely because governments have been able to mobilize significant resources to establish their control over online information.³⁰

While the parallels between the current and past authoritarian efforts to control information are most evident, the issues are relevant to democracies as well, for two related reasons. First, it is not only the survival of nondemocratic regime that depends on who controls information. Threats from extremist groups against many democratic governments in the interwar years were intensified when information manipulation through print media and radio became widespread—for example, by Father Coughlin in the United States and the insurgent Nazi Party during the Weimar Republic in Germany.³¹ Second, the political and social consequences of concentration (or even monopolization) of information in the hands of a few private corporations, which has been a defining feature of the digital era, are poorly understood.

Powerful media empires have been a mainstay of the 20th century. But the extent to which digital platforms today have access to a vast amount of public and private information is unprecedented. Google can track the metadata of hundreds of millions of users from their search, email, driving and shopping activity. Amazon has even more detailed data on purchasing behavior and evolving preferences for its massive user base, and social media platforms have unparalleled data about the emotional, political and social states of billions of people. There is

²⁹ Einstein, Elizabeth (1993) *The Printing Revolution in Early Modern Europe*, Cambridge University Press; Darnton, Robert (1995) *The Forbidden Best-Sellers of Pre-Revolutionary France*, W.W. Norton & Co; Raymond, Joad (2005) *The Invention of the Newspaper: English Newsbooks 1641–1649*, Oxford University Press.

³⁰ Acemoglu and Johnson (2023a), Chapters 9 and 10; Beraja, Martin, David Yang and Noam Yuchtman (2023) "Data-Intensive Innovation and the State: Evidence from AI Firms in China," *Review of Economic Studies* 90(4): 1701–1723.

³¹ Acemoglu and Johnson (2023a), Chapter 10.

mounting evidence that these platforms have sometimes manipulated this information for their own private benefits, as recent lawsuits against Meta, Amazon and Alphabet attest.

Reining in anti-competitive behavior is important for a fair and functioning market economy. But the monopolized control of information may have even more sweeping implications for democracy and national security.

The business model of most digital platforms in the industrialized world today depends on collection of personal data, which is then monetized via individualized digital ads. This business model encourages a particular way in which data is collected, processed and used. Most importantly, digital ads are more profitable when users are more frequently on the platform and are more intensively engaged with the offered content. This has prompted the development of algorithmic methods for boosting user engagement, often by presenting more provocative, sensational and emotionally charged material. A growing body of work in social psychology documents—and, in fact, teaches businesses—how platforms can increase engagement by manipulating user perceptions and presenting them with varying stimuli, often in the form of emotional cues.³² These cues take different forms for different subpopulations—related to envy and insecurity for teenagers and based on extremist and sensationalist material for politically engaged populations. Regardless of its exact form, this business model appears to have contributed to polarization and political disengagement, as well as to mental health problems within the population.

Algorithms have played an important role in the emergence of these new online business models. More powerful AI capabilities are significantly expanding what platforms can do, as they are able to sift through more massive amounts of data and identify new person-specific vulnerabilities.³³

Although the evidence is not conclusive, a reasonable reading of prior studies is that generative AI tools will amplify the manipulative capabilities of digital platforms, and depending on how the industry evolves, the degree to which control of information becomes more

³² Eyal, Nir (2014) *Hooked: How to Build Habit-Forming Products*, PRH Portfolio.

³³ Haidt, Jonathan and Eric Schmidt (2023) “AI Is About To Make Social Media (Much) More Toxic: We Must Prepare Now,” *The Atlantic*: <https://www.theatlantic.com/technology/archive/2023/05/generative-ai-social-media-integration-dangers-disinformation-addiction/673940/>.

monopolized could increase as well. This perspective suggests that risks to democracy and national security from how AI develops are first order.

As in the realm of production technologies, there are critical choices about the path of AI in communication. The pro-citizen path would empower individuals, provide safeguards against misinformation and offer new tools for community-building and pro-democracy activities; the manipulation path would be controlled by a few corporate players and government entities that can mislead users as they choose. Which one we choose will have major implications for democracy and national security.

Broader Social Implications

New technologies sometimes inculcate new values and practices. When they do so, they can deeply impact our political and social arrangements. The printing press is a case in point.

In the case of AI, an important set of issues will center on how people's social priorities and perceptions of their communal duties evolve.³⁴ If we proceed along the path that sidelines workers and disempowers citizens, we may also expect a greater level of disengagement and passivity within the population.³⁵ Although there is little evidence on these issues, the rise in mental health problems and withdrawal of younger generations into online environments—away from real-world social networks, community-level activities and traditional political participation—may auger more systemic social changes, with potential effects on democracy. Historical evidence suggests that economic and social means of encouraging active citizenry is critical for democratic institutions.³⁶

The direction of AI technologies—between pro-worker vs. automation-focused and between pro-citizen vs. top-down manipulative—will almost certainly impact norms and aspirations. If new technologies sideline workers, discourage political participation and promote

³⁴ Sandel, Michael (1998) *Democracy's Discontent: America in Search of a Public Philosophy*, Harvard University Press Belknap.

³⁵ Wike, Richard and Alexandra Castillo (2018) "Many Around the World Are Disengaged From Politics" *Pew Research Center Report*; Gale, Steven and Mat Burrows (2022) "Disillusioned Youth: A Danger to Democracy" *Wilson Center*: <https://www.newsecuritybeat.org/2022/06/disillusioned-youth-danger-democracy/>.

³⁶ Acemoglu, Daron and James A. Robinson (2012) *Why Nations Fail: The Origins of Power, Prosperity and Poverty*, PRH Crown Business.

manipulation on online platforms, they may raise new risks to democracy and national security. We would be wise to study and track how political and social values, means of information consumption and political participation will change in the age of AI.

Learning from History

AI is a new technology. But there have been other transformative technologies, and history offers important clues about how to manage them. The most critical lesson is that democratic governance, both in the context of how to best develop and use new technologies and how to confront the social changes that they are unleashing, is critical.

The Industrial Revolution, which started in Britain sometime in the middle of the 18th century, is today remembered as the beginning of our modern age, ushering in greater prosperity, better health and unparalleled comfort to current generations. This is true, but is only part of the story. The first 100 years of the Industrial Revolution were simply awful for the working classes. Real incomes stagnated, working hours increased, working conditions worsened in the modern factories that were highly disciplinarian and inhospitable, and life expectancy and health conditions deteriorated in the face of intensifying pollution and uncontrolled epidemics. These pitiful conditions cannot be understood without recognizing that a lot of early industrial technologies focused on automation and worker control, and there were essentially no institutions protecting workers economically or socially.³⁷

The more positive developments after 1850 stand in stark contrast to this earlier era. But there was nothing automatic about this turnaround in events. It took a major redirection of technology away from a singular focus on automation and toward the goals of raising worker productivity and increasing their roles in production, and this was embedded in fundamental political and social changes. In Britain, only the very privileged could participate in politics at the beginning of the 19th century, until the adoption of universal male suffrage and then universal adult suffrage between 1867 and 1928.³⁸ Trade unions, which had been banned and heavily

³⁷ Acemoglu and Johnson (2023a), Chapter 6.

³⁸ Acemoglu, Daron and James A. Robinson (2006) *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.

prosecuted, became legal, and workers began to campaign for higher wages and better working conditions. Draconian “master and servant” laws that made it difficult for workers to quit their jobs were abolished. Investment in public infrastructure and public health started cleaning up the cities and brought diseases under control.³⁹ Mass education and basic social safety net provisions improved the skill level and welfare of the working classes.

The general lesson is that a pro-human direction for technology, democracy and shared prosperity were symbiotic during the Industrial Revolution and similar periods of history. This lesson may have particular importance today, as it reminds us that risks against shared prosperity and democracy are tightly entangled.

Choosing the Right Path for AI

I have presented my take on the state of current knowledge about digital technologies and their economic, social and political effects. We have widely different paths ahead of us on how AI technologies will develop and their implications for productivity, inequality, social norms and democracy.

Given the fast-changing nature of AI technologies, it is impossible to be certain on any of these issues. Nevertheless, there is considerable evidence that a pro-human direction of AI—pro-worker in production, pro-citizen in communication—would be a better foundation on which to build renewed shared prosperity and would likely create a far better environment for democracy.

Is our current trajectory likely to become pro-human? I believe the answer is no. There is ample evidence that the automation focus of AI is continuing and that the most pernicious uses of centralized information by private and government actors are intensifying. It is important to understand the reasons for these global trends and develop regulatory, institutional and policy responses.

Let me first elaborate on why we are currently heading in the wrong direction. Many U.S. corporations are focused on cost-cutting due to competitive pressures or to target short-run

³⁹ Steinfeld, Robert (1991) *The Invention of Free Labor: The Employment Relation in English and American Law and Culture, 1350–1870*, University of North Carolina Press.

performance metrics. This often implies that increasing the contribution of employees to long-run performance does not receive as much attention as it deserves. Additionally, the tech industry has been dominated by the business models of the largest corporations, which prioritize the development of automation tools or monetization based on digital ads, as I have explained. Finally, the emphasis on artificial general intelligence and more broadly on reaching “human parity” have also become key metrics in the tech industry, sidelining the objective of maximizing machine usefulness.⁴⁰

If this diagnosis is correct, then government policy may be necessary to redirect AI in a more pro-human direction. There is no silver bullet for doing so, and just recognizing that a pro-human direction is both feasible and desirable may be an already important step. But there are also some policy ideas worth considering.

First, the government can set goals that are more pro-worker, and this should start with additional support for the research and development of human-complementary AI technologies. Though it is hard to target the direction of AI in the abstract, there are many areas in which opportunities for human-complementary work abound. These include education and training, healthcare, and technologies for better decision-making in skilled craft work.⁴¹ Just as DARPA orchestrated investments and competitions to foster the development of self-driving cars and dexterous robotics, sufficient investment from the federal government can foster breakthroughs in pro-worker AI.⁴² A new federal agency (perhaps “National Institutes of Computing” modeled on the National Institutes of Health) may be necessary to coordinate research effort, funding and priorities in this area.

Second, the U.S. tax code currently places a heavier burden on firms that hire labor than on those that invest in algorithms to automate work. Firms and workers jointly face a tax burden

⁴⁰ Acemoglu, Daron, Danielle Allen, Kate Crawford, James Evans, Michael Jordan, Divya Siddarth and E. Glen Weyl (2021) “How AI Fails Us,” *Edmond J. Safra Center for Ethics, Harvard: Justice, Health, and Democracy Impact Initiative*: https://ethics.harvard.edu/files/center-for-ethics/files/howai_fails_us_2.pdf; Brynjolfsson, Erik (2022) “The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence,” *Stanford Digital Economy Lab: Human-Centered Artificial Intelligence (HAI) Lab*: <https://digitaleconomy.stanford.edu/news/the-turing-trap-the-promise-peril-of-human-like-artificial-intelligence/>.

⁴¹ Acemoglu, Daron (2021) *Redesigning AI: Work, Democracy, and Justice in the Age of Automation*, Boston Review Forum.

⁴² Gruber, Jonathan and Simon Johnson (2019) *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream*, Hachette PublicAffairs.

of 25–30% for labor, while firms and capital owners pay only about 5% net when investing in equipment and computers.⁴³ A more symmetric tax structure, where marginal taxes for hiring and training labor and for investing in equipment and software are equated, may provide a better foundation upon which to build pro-worker technologies. A fairer tax system can be achieved by reducing or eliminating payroll taxes, lessening capital income tax deductions, or increasing corporate income taxes.

Third, well-functioning data markets may be an important step for pro-human AI.⁴⁴ Large-scale data collection is at the root of the manipulative use of information in online platforms. Moreover, the current path of generative AI tools is partly shaped by the possibility of using freely available data with few constraints, which also discourages investment in higher-quality data. Legislation can support a functioning data market by establishing property rights to workers and citizens for their data, which could then enable companies to invest in high-quality data. Such an environment could be a better basis for a more pro-worker trajectory of AI and may also discourage the dominance of business models based solely on maximizing user attention and engagement.

Fourth, digital ad taxes may be needed to discourage most manipulative online practices.⁴⁵ Such taxes can also have the pro-competitive effect of opening up online markets to alternative business models, such as those based on subscription (like Netflix and several other streaming platforms) or voluntary contributions (like Wikipedia), which are often crowded out because of the ability of ad-based models to grow rapidly.

Fifth, it is important to redirect technology away from an excessive focus on monitoring and surveillance, which is already receiving a big boost from large-scale AI investments in China. U.S. government leadership and funding can again play a central role here, but other regulations are also worth considering. OSHA could be tasked with regulation of surveillance in

⁴³ Acemoglu, Daron, Andrea Manera and Pascual Restrepo (2020) “Does the U.S. Tax Code Favor Automation?” *Brookings Papers on Economic Activity*.

⁴⁴ Lanier, Jaron and E. Glen Weyl (2018) “A Blueprint for a Better Digital Society” *Harvard Business Review*: <https://hbr.org/2018/09/a-blueprint-for-a-better-digital-society>.

⁴⁵ Acemoglu and Johnson (2023a), Chapter 11; Romer, Paul (2021) “Taxing Digital Advertising”: <https://adtax.paulromer.net/>.

workplaces to reduce negative impacts on workers from excessive monitoring, and new laws and regulations can clarify the limits to personal data collection and surveillance on digital platforms.

Finally, with any new technology, new ideas, entrepreneurs, companies and researchers are critical to chart different directions. Creating a more pro-competitive environment, by reducing the dominance of a few firms by means of antitrust and other regulatory measures, may be important as well.

TESTIMONY AND STATEMENT FOR THE RECORD

Margaret Hu
Taylor Reveley Research Professor and Professor of Law
Director, Digital Democracy Lab
William & Mary Law School

“The Philosophy of AI: Learning from History, Shaping Our Future”

Committee on Homeland Security and Government Affairs
U.S. Senate

November 8, 2023

Good Morning, Chairman Peters and Committee Members:

I am Margaret Hu, Taylor Reveley Research Professor and Professor of Law, and Director of the Digital Democracy Lab, at William & Mary Law School in Williamsburg, Virginia. It is an honor to be a part of this critically important dialogue on the philosophical and historical dimensions of the future of Artificial Intelligence (AI) governance.

The reason we must consider the philosophy of AI is because we are at a crossroads: either the Law governs AI, or AI governs the Law.

The first decade of my law career was dedicated to the Civil Rights Division of the U.S. Department of Justice, and for the past decade I have served as a Constitutional Law researcher and professor. I would like to approach this topic from the perspective of AI and the Law generally, and specifically AI and Constitutional Law.

Placing AI side-by-side with Constitutional Law allows us to visualize how both function on a philosophical level. It also provides us with a window into how they are philosophically in conversation with one another and how best to respond when one may philosophically conflict with the other.

I. The Philosophy of AI

A. AI and the Law: Understanding AI within Philosophical Frameworks

AI and the Law is distinctive from other areas of law, just as Constitutional Law is distinctive from other areas of law.

AI systems are more than their technological components and products. AI systems are not born independently, and AI does not operate in the abstract. Put another way, experts contend that emerging technologies can exhibit or reflect a philosophical or ideological standpoint, even inadvertently.¹ Even in the absence of instructions to do so, AI and automated systems can digest collective narratives, translate theories, and reflect hierarchies.² The way AI is designed and launched, and then interpreted and applied, is embedded within an intricate system of preexisting historical, philosophical, political, and socioeconomic structures.

AI therefore should be understood as more of a philosophy than a technology. Like Constitutional Law, AI is highly philosophical in nature.³ Specifically, AI is animated by multiple sciences and philosophies, including epistemology,⁴ a philosophy concerning knowledge structure and creation;

¹ See, e.g., MEREDITH BROUSSARD, MORE THAN A GLITCH: CONFRONTING RACE, GENDER, AND ABILITY BIAS IN TECH (2023); SAFIYA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE OPPRESSION (2018); RUHA BENJAMIN, RACE AFTER TECHNOLOGY (2019); YARDEN KATZ, ARTIFICIAL WHITENESS: POLITICS AND IDEOLOGY IN ARTIFICIAL INTELLIGENCE (2020); SIMONE BROWNE, DARK MATTERS: ON THE SURVEILLANCE OF BLACKNESS (2015); CATHY O'NEILL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019); JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019).

² Id. See also, e.g., Jessica Eaglin, *Racializing Algorithms*, 111 CALIF. L. REV. 755 (2023); Brandon L. Garrett and Cynthia Rudin, *The Right to a Glass Box: Rethinking the Use of Artificial Intelligence in Criminal Justice*, CORNELL L. REV. (forthcoming); Oren Bar-Gill, Cass R. Sunstein, and Inbal Talgam-Cohen, *Algorithmic Harm in Consumer Markets*, HARV. PUB. L. WORKING PAPER NO. 23-05 (2023).

³ See generally Daniel Susser, *Artificial intelligence and the body: Dreyfus, Bickhard, and the future of AI* in Müller, V. (ed.) *Philosophy and Theory of Artificial Intelligence*, 5 STUDIES IN APPLIED PHILOSOPHY, EPISTEMOLOGY AND RATIONAL ETHICS 277–287 (2013) (citing *inter alia* HUBERT L. DREYFUS, WHAT COMPUTERS CAN'T DO: THE LIMITS OF ARTIFICIAL INTELLIGENCE (1972); HUBERT L. DREYFUS, WHAT COMPUTERS STILL CAN'T DO: A CRITIQUE OF ARTIFICIAL REASON (1993); MARK H. BICKHARD AND LOREN TERVEEN, FOUNDATIONAL ISSUES IN ARTIFICIAL INTELLIGENCE AND COGNITIVE SCIENCE: IMPASSE AND SOLUTION (1995)); MARVIN L. MINKSY, THE EMOTION MACHINE: COMMONSENSE THINKING, ARTIFICIAL INTELLIGENCE, AND THE FUTURE OF THE HUMAN MIND (2006); MARVIN L. MINKSY, THE SOCIETY OF THE MIND (1986)).

⁴ See, e.g., Bo-chiuan Su and Batnasan Luvaanjalba, *The Effect of Hubert Dreyfus's Epistemological Assumption on the Philosophy of Artificial Intelligence*, Int'l Conf. on Human-Computer Interaction, HCI 2021: HCI IN BUS., GOV'T AND ORGS. 630-44 (July 3, 2021).

semantics,⁵ the scientific and philosophical study of natural and artificial languages; and ontology,⁶ a philosophy concerning existence. AI orders vast oceans of data. For the past few decades, various forms of AI technologies have processed oceans of data into translatable information, algorithmic systems, predictions and correlations, and other results. With natural language processing and large language models (LLMs), generative AI has introduced unprecedented accessibility to AI. With ease and speed, generative AI converts unthinkable vast galaxies of information into content which can include text, depictions, sound, and code; and can imitate a wide range of reasoning processes.⁷

AI and the Law, therefore, is a highly complex field that requires grappling with its interdisciplinary consequences, just as Constitutional Law is a highly nuanced and contextualized area of law that also demands an interdisciplinary lens.

B. Understanding AI as a Governing Philosophy

In the past year, we have entered a new phase of large commercially driven AI investments. This new phase brings into sharp relief the need for a dialogue on rights-based AI governance. The creators of generative AI have shared that their ambition is to advance Artificial General Intelligence (AGI), which aims to surpass human capacities.⁸ Generative AI and AGI ambitions force us to confront these epistemological and ontological questions head on and with some urgency in a constitutional democracy.

AI is already deployed as a governing tool in multiple contexts.⁹ AI, particularly due to its combined

⁵ See, e.g., Yoav Shoham, *Temporal logics in AI: Semantical and ontological considerations*, 33 ARTIFICIAL INTELLIGENCE 89-104 (1987).

⁶ *Id.* See also Roman Krzanowski and Pawel Polak, *The meta-ontology of AI systems with human-level intelligence*, 73 PHILOSOPHICAL PROBLEMS IN SCIENCE (ZFN) 197-230 (2022).

⁷ See, e.g., Yejin Bang, et al., *A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity*, [arXiv:2302.04023](https://arxiv.org/abs/2302.04023) COMPUTER SCIENCE, COMPUTATION AND LANG. (Feb. 2023).

⁸ See Evgeny Morozov, *The True Threat of Artificial Intelligence*, N.Y. TIMES (JUNE 30, 2023) (citing Sam Altman, CEO of OpenAI, *Planning for AGI and Beyond* (Feb. 24, 2023), <https://openai.com/blog/planning-for-agi-and-beyond>) (“Our mission is to ensure that artificial general intelligence—AI systems that are generally smarter than humans—benefits all of humanity.”).

⁹ See generally VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (2018); Danielle K. Citron & Ryan Calo, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L. J. 797 (2021); Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 15 (2019); Jonathan Zittrain, *The Hidden Costs of Automated Thinking*, THE NEW YORKER (July 23, 2019). See also *supra* notes 1-2 and *infra* note 10.

epistemological and ontological power, and its economic, political, and social power, has the potential to evolve into a governance philosophy and governance ideology. AI is constitutive of not only a knowledge structure but also a market structure in an information society and a governing structure in a digital political economy. The incentives of AI privatization and the exponential growth of datafication can operate as an invisible governing superstructure under an invisible and unaccountable hand. Additionally, AI can execute both private and public ordering functions,¹⁰ sometimes without authorization, rapidly shifting power towards centralized and privatized, and automated or semi-automated, methods of governing.¹¹

C. Understanding Constitutional Law as a Governing Philosophy

The Constitution is inspired by a philosophy of how to guarantee rights and constrain power. Constitutional Law is animated by a commitment to constitutional democracy, a governing philosophy surrounding self-governance through a republican form of government. In theory and philosophy, it separates and decentralizes power; installs checks and balances to prevent or mitigate power abuses; and supports a government that is representative “of the people, by the people, for the people.”¹²

An important question at this critical juncture is how to ensure AI as a governing philosophy will not compete with and rival Constitutional Law as a governing philosophy in a way that sacrifices our philosophical commitment to fundamental rights and the separation of powers. The Constitution is more than its text. It is a philosophy. AI is more than technology. It is a philosophy.

II. Three Opaque Boxes: Design, Translation, and Application

For the purposes of better understanding the historical and philosophical dimensions of the future of AI governance, I would like to discuss the idea of serial opaque and black boxes in a comparative way, placing AI’s opaque boxes side-by-side with opaque boxes in Constitutional Law.

Some refer to AI as a black box as it can be difficult, and in some circumstances is impossible, to

¹⁰ See, e.g., Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54 (2019); Joshua A. Kroll *et al.*, *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017); Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

¹¹ See *supra* notes 1-2, 8-10, and *infra* notes 13-14.

¹² President Abraham Lincoln, The Gettysburg Address (Nov. 19, 1863). See also GARRY WILLIS, LINCOLN AT GETTYSBURG: THE WORDS THAT REMADE AMERICA (2006).

understand its workings, as AI is never entirely transparent.¹³ In this vein, we can say that AI encompasses several opaque boxes in its lifecycle of design, translation, and application. This can also be said about Constitutional Law. Constitutional Law can be difficult to understand in its design, translation or judicial interpretation, and application.

A. Three Opaque Boxes of AI

The black box of AI often refers to the multiple layers of automation that make AI and emerging technologies—the algorithmic decisionmaking or generative AI results, for example—inscrutable. When looking at how AI is more of a philosophy than a technology, it is useful to think of a series of three opaque boxes: how philosophy may play a role in the opaque box of the technology itself; how philosophy may be immersed in the opaque box of designer, marketer, and user translation; and how philosophy may be integrated into an opaque box in the application of AI systems and social systems, and then how AI may evolve into forms of governance and governance philosophies or ideologies.

1. AI: Opaque Box in Design

Black box AI often refers to the technical infrastructure of the AI technology itself. Frank Pasquale states that “‘black box AI’ refers to any natural language processing, machine learning, textual analysis, or similar software which uses data which are not accessible to the data subject, or which deploys algorithms which are either similarly inaccessible, or so complex that they cannot be reduced to a series of rules and rule applications comprehensible to the data subject.”¹⁴ In this way, the first opaque box of AI involves the black box of the tool itself: the inscrutability of technological method; the lack of transparency and explainability; the opacity of the algorithm or incomprehensibility of the generative AI protocol, for instance, that reached its result.¹⁵ Multiple AI ethics principles and AI risk frameworks, for example, seek

¹³ See FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* (2015); Scott J. Shackelford, Anjanette Raymond, *et al.*, *Should We Trust a Black Box to Safeguard Human Rights? A Comparative Analysis of AI Governance*, 26 UCLA J. OF INT’L L. AND FOREIGN AFFAIRS 35 (2022).

¹⁴ Frank Pasquale, *Normative Dimensions of Consensual Application of Black Box Artificial Intelligence in Administrative Adjudication of Benefits Claims*, 84 LAW AND CONTEMPORARY PROBLEMS 35, 36 (2021).

¹⁵ See generally Janelle Shane, *YOU LOOK LIKE A THING AND I LOVE YOU: HOW ARTIFICIAL INTELLIGENCE WORKS AND WHY IT’S MAKING THE WORLD A WEIRDER PLACE* (2019).

to elevate attention on how biases can enter into the data and design of AI and automated systems.

2. AI: Opaque Box in Translation

Relatedly, the second opaque box of AI involves opacity in the human-system interaction with the AI. Beyond the black box of the technology, what conscious or unconscious systems drive, shape, and translate the outcomes desired by the designers and the builders of the AI? There is often a lack of comprehensibility of how the AI systems are built. The way in which a human interacts with AI in the lifecycle of AI design, testing, and decisionmaking can reflect, translate, and then incorporate multiple layers of history, philosophy, and ideology. Those who build, market, and deploy AI can infuse AI systems with these philosophical and ideological commitments. For those using such technologies, these commitments are not subject to scrutiny. Indeed, the philosophical and ideological commitments underlying AI may be invisible to its developers and marketers.

3. AI: Opaque Box in Application

Philosophically, some experts have noted that AI is a meta-technology with meta-ontological implications.¹⁶ Consequently, understanding the third opaque box of AI requires meta-theorization, similar to meta-theories applied in Constitutional Law:¹⁷ an interrogation of the way in which AI can enter the bloodstream of society and other preexisting complex systems, for instance, through its application and justification. Here, the opaque box encompasses the way in which AI may be shaping the world around us, affecting us directly and indirectly, without any awareness on our part.

B. Three Opaque Boxes of Constitutional Law

Constitutional Law is also more of a philosophy and a series of decisionmaking processes than simply constitutional provisions and court opinions. If seen as an analogy to AI as a philosophy, Constitutional Law can also be viewed as a series of opaque boxes. Constitutional Law has been referred to as a black box

¹⁶ See Krzanowski & Polak, *The meta-ontology of AI systems*, *supra* note 6.

¹⁷ See, e.g., Sanford Levinson, *The Audience for Constitutional Meta-Theory*, 63 U. COLO. L. REV. 389 (1992). See also *infra* notes 25-28.

in some contexts.¹⁸ To emphasize the benefit of a comparative method, some scholars have examined and compared what they have referred to as the black box of judicial decisionmaking against the black box of algorithmic decisionmaking.¹⁹

1. Constitutional Law: Opaque Box in Design

The first opaque box of Constitutional Law is the text of the Constitution. The philosophy of the Constitution was not stated explicitly in the text of the document beyond what was implied in the Preamble: “We the People of the United States, in Order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common defense, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity, do ordain and establish this Constitution of the United States.”²⁰ The *Federalist Papers* provided the philosophical backstory of the Constitution,²¹ as did the notes on the Constitutional Convention debates,²² speeches, news reports, pamphlets, letters, and diaries of the Founding Fathers and their contemporaries.²³ Other records form the archives of the Constitution, and also include the archival records of the formation of the Bill of Rights and the Reconstruction Amendments.²⁴

2. Constitutional Law: Opaque Box in Translation

The second opaque box of Constitutional Law involves judicial decisionmaking and the variability of judicial interpretation and translation of the text. Theories of constitutional interpretation are abundant. Jeff

¹⁸ See, e.g., Martha Minow, *The Constitution as Black Box During National Emergencies: Comment on Bruce Ackerman's Before the Next Attack: Preserving Civil Liberties in an Age of Terrorism*, 75 *FORDHAM L. REV.* 593 (2006).

¹⁹ See, e.g., Wim De Mulder et al, *Are Judges More Transparent than Black Boxes? A Scheme to Improve Judicial Decision-Making by Establishing a Relationship with Mathematical Function Maximization*, 84 *LAW AND CONTEMPORARY PROBLEMS* 47 (2021). See also Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 *COLUM. L. REV.* 1829 (2019).

²⁰ U.S. CONST. (1787).

²¹ Alexander Hamilton, James Madison & John Jay, *FEDERALISTS PAPERS: A COLLECTION OF ESSAYS, WRITTEN IN FAVOUR OF THE NEW CONSTITUTION, AS AGREED UPON BY THE FEDERAL CONVENTION, SEPTEMBER 17, 1787*, IN TWO VOLUMES (first ed. 1788).

²² JAMES MADISON, *NOTES OF DEBATES IN THE FEDERAL CONVENTION OF 1787*; EDWARD J. LARSON AND MICHAEL P. WINSHIP, *THE CONSTITUTIONAL CONVENTION: A NARRATIVE HISTORY FROM THE NOTES OF JAMES MADISON* (2005).

²³ See generally GORDON S. WOOD, *THE CREATION OF THE AMERICAN REPUBLIC, 1776-1787* (1969).

²⁴ See generally MICHAEL J. KLARMAN, *THE FRAMERS' COUP: THE MAKING OF THE UNITED STATES CONSTITUTION* (2016); AKHIL AMAR, *THE BILL OF RIGHTS: CREATION AND RECONSTRUCTION* (1998); MARK A. GRABER, *PUNISH TREASON, REWARD LOYALTY: THE FORGOTTEN GOALS OF CONSTITUTIONAL REFORM AFTER THE CIVIL WAR* (2023).

Powell, for instance, explores why any attempt to distill the original intent of our founding fathers under originalism is a contested interpretive strategy.²⁵ Philip Bobbitt sets forth several modalities of constitutional interpretation to make more transparent the weighted influences that feed into interpretive methodologies applied in judicial translation of the text, including: historical, textual, doctrinal, prudential, structural, and ethical tools of interpretation.²⁶

3. Constitutional Law: Opaque Box in Application

Just as the third opaque box of AI application requires meta-theorization, the third opaque box of Constitutional Law involves meta-theories of Constitutional Law's operationalization and influences.²⁷ The third opaque box of Constitutional Law is a systems-wide approach to questioning the way that it functions to allocate power and rights. This can present itself through history, social movements and its symbolism, theorization, and its interplay with other philosophical or ideological commitments.²⁸

III. Comparing AI with Constitutional Law as a Governance Philosophy

Throughout history, governance philosophies have competed with one another. AI rules are already being referred to by one AI company as a constitution.²⁹ Over time, AI may be seen as a philosophy of governance that competes with Constitutional Law and the rule of law.³⁰ AI is progressing in its

²⁵ H. Jefferson Powell, *The Original Understanding of Original Intent*, 98 HARV. L. REV. 885, 888 (1985) (The "original 'original intent' was determined not by historical inquiry into the expectations of the individuals involved in the framing and ratifying the Constitution, but by consideration of what rights and power sovereign polities could delegate to a common agent without destroying their own essential autonomy. Thus, the original intentionalism was in fact a form of structural interpretation") (internal citations omitted). See also JACK M. BALKIN, *LIVING ORIGINALISM* (2011); DAVID A. STRAUSS, *THE LIVING CONSTITUTION* (2010).

²⁶ PHILIP BOBBITT, *CONSTITUTIONAL FATE: THEORY OF THE CONSTITUTION* 7 (1982).

²⁷ See *supra* Levinson, *Constitutional Meta-Theory*, *supra* note 17.

²⁸ See generally ERWIN CHEMERINSKY, *NO DEMOCRACY LASTS FOREVER: HOW THE CONSTITUTION THREATENS THE UNITED STATES* (2023); HEATHER COX RICHARDSON, *HOW THE SOUTH WON THE CIVIL WAR* (2020); JACK M. BALKIN, *CONSTITUTIONAL REDEMPTION: POLITICAL FAITH IN AN UNJUST WORLD* (2011); SANFORD LEVISON, *CONSTITUTIONAL FAITH* (1988); MARK TUSHNET, *RED, WHITE, AND BLUE: A CRITICAL ANALYSIS OF CONSTITUTIONAL LAW* (1988); KENNETH W. MACK, *REPRESENTING THE RACE: THE CREATION OF THE CIVIL RIGHTS LAWYER* (2012).

²⁹ Kevin Roose, *What if We Could All Control A.I.? Researchers at Anthropic asked roughly 1,000 Americans to write rules for their A.I. chatbot. The results could be a model for future kinds of A.I. governance*, N.Y. TIMES (Oct. 17, 2023) ("What if an A.I. company let a group of ordinary citizens write some rules, and trained a chatbot to follow them? The experiment, known as 'Collective Constitutional A.I.,' builds on Anthropic's earlier work on Constitutional A.I.").

³⁰ For discussions on similar digital crossroads, see, e.g., LAWRENCE LESSIG, *CODE AND OTHER LAWS OF CYBERSPACE* (1999); LAWRENCE LESSIG, *CODE: AND OTHER LAWS OF CYBERSPACE, VERSION 2.0* (2006); BRUCE SCHNEIER, *DATA AND GOLIATH: THE HIDDEN BATTLES TO COLLECT YOUR DATA AND CONTROL YOUR WORLD*;

integration into multiple forms of governance, public and private, and is centralizing its importance in national security realms, such as information warfare and cyber conflict. Because AI serves as a shortcut to governance, it potentially opens the door for accelerated undemocratic power reallocation. This makes it an obvious target for adversaries of democracy to target and control.

In information warfare, it is significant to note that foreign and domestic adversaries are actively and concertedly working to shift governance philosophies and ideologies through psychological operations. The risks of disinformation and misinformation campaigns, exacerbated by the vulnerabilities in generative AI, guarantee that cognitive manipulation will be a cornerstone of a new chapter of the Cold War. Constitutional democracy may be modified in ways that are difficult to see and assess, as AI governing systems may lack transparency, AI deployed may be opaque and almost impossible to challenge, and the philosophical dimensions of the AI may swim in the subterranean levels of social and individual consciousness.

A. Learning from History

A philosophy can speak to another philosophy. In other words, analyzing one philosophy against another can bring clarity to both. A philosophy like a constitutional democracy needs to speak to another philosophy, like the epistemology and potential ideology that is infused in AI, to consider the way in which one may or may not be consistent with the other. When a philosophy like a constitutional democracy tries to speak to AI as a technology, it may struggle, and a crucial moment to converse about risks to civil and human rights, and fundamental constitutional rights, might be lost. There may be lapses in dialogue due to an incompatibility in the framing of discourse, a lack of shared vocabulary, and a failure to comprehend the consequences of the conversation.

Constitutional Law is driven by a set of philosophical commitments, such as equality, due process under the law, limited government of enumerated and defined powers, and other principles and values. AI, by contrast, presents itself as a technology whose philosophical commitments are embedded, hidden,

BENJAMIN WITTES AND GABRIELLA BLUM, *THE FUTURE OF VIOLENCE: ROBOTS AND GERMS, HACKERS AND DRONES: CONFRONTING A NEW AGE OF THREAT* (2015). *See also supra* notes 1-4, 8-10, and 13.

and even denied. In the name of achieving technological efficiencies, it can reshape and even undermine core constitutional philosophical commitments.

In a constitutional democracy, the rule of law precedes power. Our Founders proclaimed that equality and freedom were endowed rights and a form of inalienable self-sovereignty. The signing of the Declaration of Independence, the Constitutional Convention, the ratification of the Bill of Rights were products of a deep historical and philosophical struggle. Going forward, the point of decision is: will AI be applied in a way that is consistent with our constitutional philosophy, or will it alter it, erode it, or mediate it?

B. Shaping Our Future

We have close to 250 years of experience in trying to make sense of how to preserve the democratic experiment through a close examination of the meaning and consequences of the U.S. Constitution. In a constitutional democracy, consent of the people and government by the people are paramount. The social contract implies a direct relationship between the people and their government. Similarly, rights are intended to be directly accessed and the failure to grant rights can be directly challenged.

In AI and our digital economy, it is said that technology has transformed people into the product. This and other emerging technologies, and other policy developments that have intersected with technology's evolution, have transformed the relationship between government and its citizenry.³¹ The social contract has been radically altered.³² Jack Balkin describes the need to fix the "grand bargain" of the modern political economy that spins on the axis of digital information.³³ Under his theorization, certain constitutional rights, such as the First Amendment, can be grasped and better protected if seen as now operating along three coordinates of a triangle: the government, the speaker, and the technology

³¹ See, e.g., WITTES & BLUM, *supra* note 30.

³² *Id.* (contending that social contract could be undermined by multiple trajectories, including technological, national security, and privatization and decentralization developments).

³³ See, e.g., Jack M. Balkin, *Fixing Social Media's Grand Bargain*, HOOVER WORKING GP. ON NAT'L SEC., TECH., AND LAW, AEGIS SERIES PAPER NO. 1814 (Oct. 16, 2018).

companies.³⁴ Building upon Balkin's triangle, a rights-based AI governance framework inquiry can be framed as a constitutional quadrilateral, whereby constitutional rights operate along four vertices: government, citizen and civil society, technology companies, and now AI systems. We are at a critical juncture where we must grapple with whether constitutional rights were meant to be mediated in this way: whether we can still access equality and due process; and the fundamental freedoms of autonomy, and expressive and associational rights; and other privileges, when knowledge and other interests are digitally mediated, and when it may be nearly impossible to challenge the methods and outcomes of these negotiations.³⁵

As the capacities of AI evolve, several risks will grow exponentially and more rapidly than we can anticipate, including reexamining definitions of personhood and citizenship; the erosion of privacy and autonomy through data-driven AI; drawing separation between real and virtual; incentivizing data breaches, cyber conflict, and AI weaponry; the risks of predictive policing and autonomous weaponry; information warfare and dis/misinformation campaigns; inscrutability and opacity of AI decisionmaking, and black box or opaque box accountability challenges; AI displacement of human knowledge and judgment, and labor; exacerbation of bias and discrimination; retrenchment of historical impulses that bend toward marginalization and classification; and social credit systems, datafication, scoring, and profiling, and exploitation of digital identities; and other data tracking, databasing, and cybersurveillance harms.

AI releases extraordinary potential, like many prior industrial, scientific, and technological revolutions. It can expand commerce and productivity, mapping insights through its predictive potential like automation tools have done in the past. The paramount inquiry of AI and the Law as a field of study

³⁴ See, e.g., Jack M. Balkin, *Free Speech is a Triangle*, 118 COLUM. L. REV. 2011 (2018) ("On one corner are nation-states and the European Union. On the second corner are privately owned internet-infrastructure companies, including social media companies, search engines, broadband providers, and electronic payment systems. On the third corner are many different kinds of speakers, legacy media, civil-society organizations, hackers, and trolls.").

³⁵ My prior research has attempted to explore these questions in various legal and technological contexts. See, e.g., Margaret Hu, *Big Data Blacklisting*, 67 FLA. L. REV. 1735 (2015); Margaret Hu, *Algorithmic Jim Crow*, 86 FORDHAM L. REV. 633 (2017); Margaret Hu, *Biometrics and an AI Bill of Rights*, 60 DUQUESNE L. REV. 283 (2022).

and practice is how to harness the potential of AI and regulate its application in a way that is consistent with, and not irreversibly harmful to, our constitutional democracy.

IV. Conclusion

The future of AI governance and automation will likely reflect similar historical and philosophical tensions, as the conflicts and struggles will likely rhyme with the past.³⁶ In a constitutional democracy, it is important to start with the philosophy and history of democratic governance. To ground how AI must be interrogated for its governance impact, the humanities and philosophies that underscore an “analogue democracy” must serve as a guide in a “digital democracy.”³⁷

If we look at AI too literally as only a technology, we run the risk of not fully grasping its impact as a challenge to philosophical foundations. We know from history that those who design, litigate, and interpret Constitutional Law can permeate it with historical biases and antidemocratic ideologies, sometimes in ways that are undertheorized and misunderstood, when reaching the intersection of two roads.

The serial opaque boxes of AI pose similar risks and vulnerabilities. As we are at the intersection of two roads, oversight of AI and other emerging technologies forces the threshold question at the crossroads: will AI govern the law or will law govern AI? To preserve a constitutional democracy, there is only one answer: it must be the latter.

³⁶ See generally KEVIN ROOSE, *FUTUREPROOF: 9 RULES FOR HUMANS IN THE AGE OF AUTOMATION* (2021); DARRELL M. WEST AND JOHN R. ALLEN, *TURNING POINT: POLICYMAKING IN THE ERA* (2020); Bruce Schneier and Davi Ottenheimer *Robots Are Already Killing People*, *THE ATLANTIC* (Sept. 6, 2023).

³⁷ See, e.g., Jamie Bartlett, *The war between technology & democracy*, *MEDIUM* (Sept. 18, 2018); NANJALA NYABOLA, *DIGITAL DEMOCRACY, ANALOGUE POLITICS: HOW THE INTERNET ERA IS TRANSFORMING POLITICS IN KENYA* (2015).



THE UNIVERSITY of EDINBURGH
Edinburgh Futures Institute

Centre for
Technomoral Futures



Testimony of Prof. Shannon Vallor

Baillie Gifford Chair in the Ethics of Data and Artificial Intelligence
Director, Centre for Technomoral Futures at the Edinburgh Futures Institute
Co-Director, BRAID (Bridging Responsible AI Divides)
The University of Edinburgh

For the U.S. Senate Committee on Homeland Security and Governmental Affairs

Hearing Date: November 8, 2023

Thank you Chairman Peters, Ranking Member Paul, and distinguished Members of the Committee for this opportunity to submit my testimony today. It is a profound honor to address you on a matter of such vital importance to the nation and the human family.

For over a decade, my research as a philosopher of technology has focused on the ethical and political implications of artificial intelligence, robotics and algorithmic automation. I currently direct the Centre for Technomoral Futures at the University of Edinburgh, which integrates technical and moral expertise in new models of responsible innovation and technology governance. I have worked as an AI ethicist in the United States and the United Kingdom, in both academia and the tech industry, and as an independent ethics advisor to UK and Scottish government bodies on public sector uses of AI and data.

In the UK, I lead multi-disciplinary teams of computing, social science and humanities researchers as part of two Responsible AI initiatives funded by the national science funding agency UK Research and Innovation. In the *Trustworthy Autonomous Systems* program, our team is working to strengthen human responsibility for autonomous systems. Our BRAID program, *Bridging Responsible AI Divides*, builds new partnerships between government, voluntary organizations, civil society and industry to drive the growth, embedding and adoption of responsible AI knowledge and practices, while leveraging the arts and humanities to strengthen and sustain the UK's AI ecosystem.

1. Introduction

My research is interdisciplinary, but deeply informed by philosophical and historical perspectives on technology's role in shaping human values, character and capabilities. Among the most important of those capabilities is *self-governance*. This capability, to

reason, think and judge for oneself how best to live, underpins the individual civil and political liberties guaranteed by the U.S. Constitution and by international law. It also underpins democratic life. In a democratic society, decisions about how best to live must be made in political cooperation with those whose fates are intermingled with ours.

The link between AI and our capacity for self-governance has two dimensions that interweave in complex ways, not unlike a Mobius strip. One side of the strip is our need, in any democratic society, to jointly exercise responsible self-governance of the new social, political and economic powers that AI technologies inject into our institutions. The second side of the strip is the way that AI technologies can undermine our confidence in, and will to exercise, those same human capabilities of self-governance.

My testimony will expand on these two interrelated challenges for AI and democracy; I will conclude with some brief, but I hope, encouraging reflections on lessons from history for the responsible democratic governance of AI technologies.

2. AI as a Subject of Democratic Governance

As a new source of immense socioeconomic and political power, AI is *something we must govern*. Why? Because ungoverned, or ungovernable, social and political power is deeply incompatible with democratic principles. A core principle of modern democracies is that free peoples may not *justifiably* be subjected to social and political powers which determine their basic liberties and opportunities, but over which they have no say, which they cannot see and freely endorse, and which powers are in no way constrained by or answerable to them.

A point of clarification is needed. AI is not a single technology, but many; AI is not one monolithic power, but a smorgasbord of them. Many types and uses of AI do not affect our fundamental liberties and opportunities at all. So it does not make sense to talk about our need to govern AI *as a whole*. However, for the sake of clarity in this testimony, take 'AI' here to refer to the subset of AI technologies that increasingly *do* affect our basic liberties and opportunities to flourish, either by greatly amplifying the power of certain individuals and existing institutions, or by generating new powers that can be exercised upon us.

Now, in a democratic society, *might does not make right*. Power is not self-justifying. Any exercise of power that extends beyond the private conduct of the individual, in ways that significantly impact the opportunities, welfare and liberties of others, always requires social and political legitimization. We do this through the joint exercise of our political capacity. Legislation and regulation are one way that democratic peoples, through their elected representatives, jointly govern the powers that impact us. Expressing and enforcing shared moral and political norms is another way we collectively self-govern. Adopting

professional and technical standards is yet another. Market incentives do some of the work (though far less than some economists imagine), and various forms of public and organizational policy not enacted in law do much of the rest. For most significant forms of power, a combination of these tools is needed. What matters is that the right incentives are created to ensure the responsible, trustworthy and politically legitimate use of that power, under appropriate, mutually agreed upon, and reliably enforced constraints.

In the United States, as with most if not all democratic countries, AI as a new source of power has yet to be socially or politically *legitimized* by the necessary acts of effective governance. These technologies are operating in most domains as ungoverned and growing powers; either because the necessary laws, regulatory bodies, policies, norms and standards to govern them have yet to be created, or more often, because the existing modes of governance that already apply are simply not being used or reliably enforced.

Nor is the growing power that AI technologies generate being distributed equally across our society, so that all of us might choose to use this new power for our benefit. Rather, the power and benefits these technologies afford are increasingly concentrated in very few hands, especially those (corporate) hands that are *already* operating with undue influence on our democratic systems of self-governance. This concentration only amplifies the economic inequality that has been rising in this country for over 40 years, all while intergenerational socioeconomic mobility fell well behind most other developed countries – a linked phenomenon known by economists as the Great Gatsby Curve.¹

If AI technologies were an equalizing force, as well as transparently and widely beneficial, generating little risk to society or vulnerable groups, then closing the AI governance gap would not be so urgent. Unfortunately, we have a growing pile of evidence for the many harmful effects of AI systems and related technologies on humans living today. While the potential for AI technologies to benefit society remains immense, far fewer of those shared benefits have materialized than were once confidently predicted as imminent.² There are many promising applications of AI that we *can* use now, in health research, energy research, and environmental management;³ but the most socially beneficial use cases suffer from low investment, and very few are being deployed commercially at scale.

¹ Steven N. Durlauf, Andros Kourtellos, Chih Ming Tan (2022), 'The Great Gatsby Curve,' *Annual Review of Economics* 14:1, 571-605. <https://www.annualreviews.org/doi/10.1146/annurev-economics-082321-122703>

² For example, in 2015 IBM's Watson for Health AI was widely predicted to soon outperform human doctors; it crashed out into a commercial and scientific failure just a few short years later. Fully automated, safe and reliable driverless cars, and robots that can relieve exhausted caregivers, have been 'just around the corner' for over a decade.

³ Sophie Bushwick, '10 Ways AI Was Used For Good This Year,' *Scientific American*, Dec 15 2022. <https://www.scientificamerican.com/article/10-ways-ai-was-used-for-good-this-year>

The disparity between the timelines of AI benefit and AI risk is striking. The Committee has heard much about AI's risks in previous testimony, and I will not detail them here. Unlike the benefits, the risks have suffered no delay in arriving.⁴ And yet we have done vanishingly little about them. Most Americans, and billions of others around the world, remain highly vulnerable and exposed to unjust and discriminatory automated decisions, dangerous AI malfunctions, unwarranted algorithmic surveillance, false arrests and profiling, misidentification, disinformation, fraud, and unsafe or illegal content. As a result, public attitudes toward AI are souring,⁵ a serious warning sign for those of us who *want* the technology to mature and succeed. Other technologies, from GMOs to nuclear energy generation, have historically suffered public backlash in ways that greatly limited their beneficial use and advancement, and AI is very much at risk of a similar backlash.

This is not a *scientific* problem. We have studied AI's risks extensively over the past decade, and there is a mountain of evidence about their causes. They are not mysterious, but fairly well understood. And while there is a lot still to learn about how to manage these risks, we are already well on our way. That is the good news! For nearly a decade, researchers in the closely related fields of AI and data ethics, Responsible AI, machine learning fairness, and AI trust and safety have been generating a truly impressive body of powerful tools for documenting, evaluating and auditing AI systems, for anticipating and measuring harmful AI outcomes, and for mitigating their risks.

Organizations like NIST, the IEEE Standards Association, the British Standards Institute, the Alan Turing Institute, the Ada Lovelace Institute, AI Now, the World Economic Forum, and many others have released ample bodies of guidance on how to use these tools. Research programs like BRAID, which I co-direct in the UK, are creating new pathways to embed, test, and adopt Responsible AI tools in industry, government and third sector organizations that want to use AI safely and successfully.⁶ If we mandated the judicious and responsible use of these tools tomorrow, across the AI ecosystem, it would take a bit of time for companies to comply and for the effects to show up, but they would come. Sooner, I would wager, than will safe driverless vehicles.

⁴ Important advances have been made with the help of AI in areas such as protein-folding research, disease diagnosis, flood forecasting, and automated transcription and captioning tools. But the benefits have thus far been dwarfed by the harmful and unjust patterns that routinely emerge in automated algorithms used in hiring, education, policing, public benefits fraud detection, bail risk, hospital triage, and numerous other high-stakes endeavors. The harms are not theoretical: they include false arrests, lost jobs, bankruptcies, separation of children from innocent families, premature deaths and suicides. See <https://incidentdatabase.ai/>

⁵ Alec Tyson and Emma Kikuchi, 'Growing public concern about the role of AI in everyday life,' Pew Research Center, Aug 28 2023. <https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/>

⁶ <https://braiduk.org/fellowships>

However, the use of these tools in high-impact and high-risk AI systems remains sporadic, opaque, inconsistent and underincentivized. For example, in 2022 and 2023, just as today's new generative AI tools were being released, several AI companies made heavy cutbacks to their in-house AI ethics and trust and safety teams, or even removed them altogether.⁷ Recently, many of the same corporations rebranded such efforts under the label of 'AI safety', with a new focus. Instead of bringing back the experienced, expert teams they laid off, or using the tools those teams created to responsibly address the individual and social harms caused by their existing products and business models, several of these companies now seek government and philanthropic funding to invest in technical study of the unknown, longer-term risks of future 'frontier' models that could be more dangerous than those we have today.⁸

Yet many of the same commercial AI leaders warn that government interference in their efforts, whether through regulatory constraint, demands for transparency, or greater exposure to liabilities for AI harms, will only delay or diminish their capacities to get ahead of these so-called 'existential risks' for our benefit. The implication is clear: as long as we stand aside and let them work (ideally subsidized by public investment), they will keep us all safe from the AI bogeyman that they assure us threatens human eradication.

This problem is not scientific or technical. It is *political*. The history of political thought that shaped this nation tells us how we should assess promises of safety and security from those who seek release from accountability as the payment. As the philosopher John Locke said in 1689 of similar promises from unaccountable monarchs: "This is to think that men are so foolish that they take care to avoid what mischiefs may be done them by polecats or foxes, but are content, nay, think it safety, to be devoured by lions."⁹ Nor is safety enough, if it comes at the expense of our liberty and capacity for self-determination. As noted by Jean-Jacques Rousseau in 1762's *The Social Contract*, humans can also find safety and tranquility in dungeons.¹⁰

I do not blame the powerful corporations behind today's AI technologies for evading their social responsibilities to be accountable and answerable to the rest of us. I have worked with and for technology companies. I have collaborated with many researchers and

⁷ Gerrit De Vynck and Will Oremus, 'As AI booms, tech firms are laying off their ethicists,' *The Washington Post*, Mar 30 2023. <https://www.washingtonpost.com/technology/2023/03/30/tech-companies-cut-ai-ethics/>

⁸ Parmy Olson, 'There's too much money going to AI doomers,' *The Washington Post*, Aug 16 2023. https://www.washingtonpost.com/business/2023/08/16/ai-apocalypse-there-s-too-much-vc-money-going-to-doomers/ada4388e-3bed-11ee-ae6d-40c039a855ba_story.html

⁹ John Locke (1689/1884). *Two Treatises on Civil Government*. London: Routledge, p. 230–40.

¹⁰ Jean-Jacques Rousseau (1762). *The Social Contract*, trans. G.D.H. Cole, https://web.archive.org/web/20171215154541id_/https://www.ucc.ie/archive/hdsp/Rousseau_contrat-social.pdf

engineers employed by them (and still do). Virtually all *wanted* to make fairer, more transparent, and more beneficial tools, and most did their best to make that happen, with what power they had. I have spoken at conferences where well-paid machine learning engineers got *teary* telling me how much they wanted to build responsibly. Many of the most groundbreaking and effective research outputs and practical tools in the field of Responsible AI have, in fact, been created by people who were at the time employed by large tech companies, such as Timnit Gebru, Meg Mitchell, and Rumman Chowdhury.¹¹

The problem isn't the corporations. The problem is a decades-old political failure to give those companies the proper incentives to align their business activities with the wider public interest, or even to comply with minimal standards of transparency, accountability, safety and fairness. No one questions the need for this compliance in other high-stakes, safety-critical fields, like aviation or pharmaceuticals. Who would suggest (especially in the aftermath of the 737 MAX) that Boeing should have been *freer* from regulatory obligations, and trusted to innovate without oversight, transparency and accountability?

Would anyone suggest that Johnson & Johnson be allowed to release a powerful new heart drug in 'beta', with no independent safety testing and no disclosure of efficacy, risks or side effects beyond what the company elects to disclose? Yet Tesla and Cruise have both enjoyed precisely this license for testing autonomous driving technologies on public roads, at least until the California Department of Motor Vehicles recently suspended Cruise's robotaxi deployment, apparently for concealing safety issues and critical evidence in a case where a pedestrian was struck and dragged under one of the vehicles.¹² Last week, it was revealed by a *New York Times* investigation that these supposedly 'driverless' taxis actually required remote human operators to intervene every few miles of travel – something Cruise had apparently not disclosed in any of its marketing.¹³

Corporations are not natural persons or entities. They are socially constructed and legally empowered entities. They were created centuries ago by government charters and continue to exist by virtue of government and social licenses to operate. They follow the incentives given by the legal and commercial order in which they operate. In this they are no different from anything else that operates without the benefit of a moral and social conscience.

¹¹ For an excellent overview of their contributions, see *Time* magazine's 2023 list of the '100 Most Influential People in AI': <https://time.com/collection/time100-ai/>

¹² Russ Mitchell, 'Cruise sidelines entire U.S. robotaxi fleet to focus on rebuilding 'public trust,' *The Los Angeles Times*, Oct 27 2023. <https://www.latimes.com/business/story/2023-10-27/cruise-shuts-down-robot-cars-rebuild-public-trust>

¹³ Tripp Mickle, Cade Metz, and Yiwen Lu, 'G.M.'s Cruise moved fast in the driverless race. It got ugly.' *The New York Times*. Nov 3, 2023. <https://www.nytimes.com/2023/11/03/technology/cruise-general-motors-self-driving-cars.html>

If you need to change such an entity's behavior, you don't argue with it, or try to shame it. You first study its incentives and then modify them, so that it is sufficiently rewarded for pursuing a better course of action and/or penalized for not doing so.

AI companies and their executives are legally incentivized to maximize shareholder value (too often, in the short term), within whatever constraints the jurisdictions they operate in supply and adequately enforce. Multinational corporations will exploit opportunities for regulatory arbitrage where they are available. But most tech companies will comply with reasonable, well-enforced regulation that raises the floor for responsible conduct, as long as it enables continued commercial viability and innovation (which effective governance does, by greatly lowering the risks for those who invest in, adopt or use the technology).

If a dog chases a kid on a bicycle, I don't blame the dog. I blame the owner who didn't leash it. I don't blame tech companies for not making AI safe, fair and accountable. I blame the democratic political institutions that have neglected their duties to competently govern the powers that now impinge upon their citizens' vulnerabilities and liberties.

It is therefore encouraging to see moves this year in Europe and the UK to consider, albeit in very different ways, empowering regulators to act on AI harms. In the United States, the Blueprint for an AI Bill of Rights and the October 30 Executive Order are welcome steps toward ensuring that AI governance addresses the most urgent risks presented by the technology, while enabling more of its benefits to be developed and distributed to wider publics. It also promises to incentivize more competition in the AI ecosystem, which is another important goal. But the mechanisms of adequate enforcement, and the incentives for corporations to seriously adopt these measures, remain weak and underspecified.

3. AI as a Stressor Upon Democratic Agency: Philosophical and Historical Perspectives

Today, researchers looking at AI's impact on the health of our democracies tend to focus on the impact of disinformation, misinformation and algorithmic echo chambers. As we see in the social media context, these can diminish our political sense of a shared reality, enable new forms of political persuasion and manipulation, and make social cooperation, cohesion and consensus harder for us to build and sustain. These are well studied concerns¹⁴ that I won't testify to further, except to note that some researchers believe that AI 'deepfakes' and other modes of AI disinformation may be less impactful than the lower-cost, lower-effort modes of political disinformation that have been proliferating with little cultural resistance for decades, even before the social media revolution.

¹⁴ Todd C. Helmus (2022). Artificial Intelligence, Deepfakes and Disinformation: A Primer. *The Rand Corporation* https://www.rand.org/content/dam/rand/pubs/perspectives/PEA1000/PEA1043-1/RAND_PEA1043-1.pdf

What is less often discussed today is the impact of AI technologies on human confidence in our democratic capabilities to reason and govern ourselves, and the downstream effects on our willingness to hold political space for human liberty and self-determination. Yet a slightly longer historical view reveals that philosophers, sociologists, political theorists, and computer scientists have been warning us about this since the start of the digital revolution. 1960s computing pioneer Joseph Weizenbaum anticipated the corrosive effect of algorithmic automation on human liberty, lamenting in 1976's *Computer Power and Human Reason* that just when humans have "ceased to believe in—let alone to trust—[our] own autonomy," we have "begun to rely on autonomous machines."¹⁵

He was far from the first. In 1954, French sociologist and philosopher Jacques Ellul foresaw an inevitable conflict between human self-determination and machine efficiency, predicting that the efficiency of the technical order (In French, *technique*) would soon be seen as the highest social and moral value. As a consequence, Ellul predicted, human spontaneity and insistence upon our rights of self-determination would be relentlessly pathologized and forced out of the system. He wrote that "The combination of man and technique is a happy one only if man has no responsibility."¹⁶

That same year, cybernetics pioneer Norbert Wiener, who developed the first theories of machine learning and intelligent automation upon which our modern AI systems depend, warned of a future where humans imagine that they can surrender the burdens and inconvenience of moral and political decision-making to intelligent machines. He wrote that "to throw the problem of responsibility on the machine, whether it can learn or not, is to cast [our] responsibility to the winds, and to find it coming back seated on the whirlwind."¹⁷

In Isaac Asimov's 1955 short story *Franchise*, he imagined a future of 2008 in which Americans have surrendered our voting power to the supercomputer Multivac, a mechanical 'black box' that in the story's telling, cannot be fully understood even by its own designers, and yet is said to know the collective American political will with even greater precision than Americans do. Today there is no Multivac and most Americans retain the power to vote. Yet we are increasingly being told by powerful AI scientists and business leaders that AI systems will soon be smarter, wiser, and more rational than we are, if they aren't already. They're not just intelligent, we are told, they are becoming *superintelligent*. They aren't just human-like, they're *superhuman*!

¹⁵ Joseph Weizenbaum (1976). *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: W.H. Freeman & Co., p. 9.

¹⁶ Jacques Ellul (1954/1964 trans.), *The Technological Society*. New York: Vintage Books. p. 136.

¹⁷ Norbert Wiener (1954). *The Human Use of Human Beings: Cybernetics and Society*. Boston: Da Capo Press/Houghton Mifflin, p. 185.

The evidence for such claims requires moving the scientific goalposts for demonstrating ‘intelligence’ and ‘humanness’ by a mile, but in an era of surging political irrationality and division, the implication of such promises is clear: the truly important decisions can’t be entrusted to humans any more. Leave it to the machines, who will exercise the intelligent control and reasonableness that you no longer believe your neighbor has.

OpenAI’s Sam Altman has previously expressed confidence that we are merely the ‘biological bootloader’ for a new, higher form of intelligence that will dwarf ours, not just in brute computational ability (a far more reasonable claim) but in *wisdom*.¹⁸ Despite the mountain of evidence that today’s most powerful AI systems, including GPT models, tend to relentlessly amplify rather than eliminate unfair human bias, Altman has said, ‘I think we’ll find out we can make GPT systems *way less biased* than any human.’¹⁹ Of course, we could strive (as we have for centuries) to become more equitable and fair in our judgments, and GPT, which is trained on our data, could follow our lead. For Altman this is too many steps. He thinks that without our ‘emotional load,’ a mathematical tool like GPT can somehow leapfrog humanity’s stalled moral development, leaving us in the dust.

Turing Award winner Yoshua Bengio says that we may have superhuman AI systems already, since “we call an AI superhuman if it outperforms humans on a vast array of tasks.”²⁰ Notice the staggering reduction of what it means to be ‘human’ – a human is now simply an *underperforming completer of (computational and economically valuable) tasks*. Indeed, OpenAI has already redefined artificial general intelligence (AGI), which used to mean ‘a machine that can think and reason just like a human,’ as a system that can “surpass human capabilities in a majority of economically valuable tasks.”²¹

None of these AI leaders claim that these systems are sentient. Bengio and Altman do not claim that AI models have emotions, or a moral conscience, or a sense of justice, or the capacity for love, loyalty and self-sacrifice, or aspirations beyond those dictated to them, or the ability to take responsibility for their own actions. These capabilities have now been excised from the concept of general intelligence, since they don’t fit into the box of ‘economically valuable tasks.’ Nor do they apparently remain central, vital capabilities of the human person. *After all, a machine can be ‘superhuman’ without them.*

¹⁸ Sam Altman, ‘The Merge,’ Dec 7 2017. <https://blog.samaltman.com/the-merge>

¹⁹ Lex Fridman Podcast #367 (2023), ‘Sam Altman: OpenAI CEO on GPT-4, ChatGPT, and the Future of AI,’ <https://auphonic.com/media/blog/LexFridmanPodcast367-transcript.html>

²⁰ Yoshua Bengio, ‘FAQ on catastrophic AI risks,’ Jun 24 2023. <https://yoshuabengio.org/2023/06/24/faq-on-catastrophic-ai-risks/>

²¹ Mark Sullivan, ‘Why everyone seems to disagree on how to define AGI,’ *AI Decoded*, Oct 18 2023, <https://www.linkedin.com/pulse/why-everyone-seems-disagree-how-define-artificial-general-iaudc/>

It is not AI systems that pose the greatest risk to our humanity right now. It's the people telling us that our humane capabilities are *outmoded*, particularly the forms of moral and political agency that are the fragile foundations of democratic life. After all, if these are nonessential for superhuman intelligence, why carry on with them? What place do they have in social decision-making, or our choices about our shared human future? The philosopher and theologian Hans Jonas, like Ellul, Weiner, and Weizenbaum before him, saw this coming. He wrote in 1984, "we need wisdom most when we believe in it least."²²

Just as when Jonas wrote at the height of the Cold War, today the security of the human family hangs on our capacity for wise self-governance and collective political reason in service of shared values. Yet we believe in these things with far less sincerity and fervor than we did in the 18th century. As philosopher of technology Langdon Winner observed in 1986, even the concept of 'shared values' has lost its capacity to help us wisely steer technological development to desirable ends, since we no longer see 'values' as we once did, as objective features of a desirable and just state of human affairs to be pursued by democratic processes. Instead, we see values as personal, arbitrary whims of individuals, immune to political challenge and reasoned deliberation. We substitute the older, harder question of democracy that this country's founders asked – in Winner's words, 'how are we to live together, gracefully and with justice?'²³ – with the technocrat's cheap replacement: 'how can we calculate the sum of individual human preferences?'

In 1984, Jonas was clear that we must recover our confidence and belief in shared political wisdom if we hope to have any chance to avert irreversible environmental and climate destruction, as well as nuclear and biological holocaust. We can add to that list the need to safely govern AI. We have the same need for democratic renewal now as we did then, only more so, as global confidence in political reason guided by democratic norms is even lower than when Jonas and Winner wrote.

AI in its present form does not promote that renewal. If anything, AI's growing use to automate high-stakes decision-making, alongside the myth of 'superhuman AI,' threatens our already damaged collective confidence in human moral and political agency. AI hype fosters a narrative in which our unique human capacity for practical wisdom – the virtue that Aristotle called *phronēsis* – is not a political tool that democratic peoples can wield to responsibly craft better futures, but an outmoded relic waiting for the scrap heap.²⁴

²² Hans Jonas (1984), *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago: University of Chicago Press, p. 21

²³ Langdon Winner (1986/2020), *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, Second Edition. Chicago: University of Chicago Press, p. 162.

²⁴ For more on practical wisdom and AI, see Vallor (2016) *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, New York: Oxford University Press, and the forthcoming (2024) *The AI Mirror: Reclaiming Our Humanity in an Age of Machine Thinking*, New York: Oxford University Press.

AI is not an intrinsically destructive set of technologies. They aren't even the most dangerous technologies we've ever made (nuclear weapons still hold that dishonor). AI is just coming along at a very bad time, when our political and moral will to exercise democratic wisdom is flagging, and when even our basic belief and confidence in democratic institutions and the value of democratic ways of life are deeply damaged. Surveys of young generations around the world keep telling us this, and AI isn't the cause.²⁵ Instead, young people cite the failure of their elected political leaders to work together constructively, and reliably serve the common public interest, rather than their donors and most powerful lobbyists.

It may not be the cause of our democratic malaise, but AI, particularly the careless uses and narratives in which AI is painted as a superior replacement for high-stakes human decision-making, is one more force pressing on democratic cultures already riddled with stress fractures. If we don't assert and wisely exercise our shared capacity for democratic governance of AI, *it might be the last chance at democratic governance we get.*

4. Looking backward, to find a road forward

Current media and academic debates about the impact of AI on society rarely take a historical perspective. AI is treated as *sui generis*, a technology unlike anything we have ever encountered, and one which we are repeatedly told that humans are radically unprepared and ill-equipped to govern. This narrative is deeply flawed and misleading, yet it is endlessly amplified by powerful interests who stand to profit from its acceptance.

Like all compelling but incomplete narratives, this one embeds important grains of truth. Today's artificial intelligence technologies *are*, in many important ways, unprecedented. Yet so were the printing press, the steam engine, the airplane and the automobile. You may say: 'Sure, but these tools couldn't destabilize whole societies, or destroy our humanity!' Yet the printing press was once widely feared as an agent of widespread political and moral destruction. So was nuclear power. So was genetic engineering. Ironically, even *steam* power sparked 19th century fears about the evolution of intelligent machines that could replace or enslave humans.²⁶

²⁵ Open Society Foundations, 'Generational Shift: New Global Poll Reveals Large Minorities of Young People Lack Faith in Democracy to Deliver on Their Priorities,' Sep 11 2023. <https://www.opensocietyfoundations.org/newsroom/generational-shift-new-global-poll-reveals-large-minorities-of-young-people-lack-faith-in-democracy-to-deliver-on-their-priorities>; see also Foa, R.S., Klassen, A., Wenger, D., Rand, A. and M. Slade (2020). "Youth and Satisfaction with Democracy: Reversing the Democratic Disconnect?" Cambridge, United Kingdom: Centre for the Future of Democracy. https://www.cam.ac.uk/system/files/youth_and_satisfaction_with_democracy.pdf

²⁶ See the 'Book of the Machines' in Samuel Butler's 1872 novel *Erewhon* (New York: Penguin).

Not all of these fears were wrong. The printing press *did* radically remake society, although we now think for the better, as it helped to usher in a new political possibility: well-informed peoples capable of democratic self-governance.²⁷ Nuclear energy and genetic engineering share a more ambivalent destiny. Each has delivered immense benefits, yet they still retain the potential to end intelligent life on this planet.

In comparison with nuclear power and bioengineering, AI has been described as promising even more potential benefits, and equal or greater dangers, including human extinction. Along with many other AI researchers, I believe many of the ‘existential risks’ of human destruction attributed to AI are wildly exaggerated and scientifically unsound. Those who use the AI risk narrative to draw political and public attention away from the immediate and undeniable existential threats of runaway climate change and growing nuclear destabilization are, in my view, doing serious harm.

Another harmful narrative is that the U.S. must accelerate AI research without regulatory guardrails to avoid losing an AI ‘arms race’ with China. While much more research is needed to keep up with the new cybersecurity and defense risks presented by AI systems, a runaway proliferation of unsafe, unregulated AI technologies will only endanger us all. New AI techniques are not the sort to remain state secrets for long, and in any case, China has its own internal reasons to keep AI from developing without tight human controls.

Even so, AI *does* present many incontrovertibly real risks to human flourishing, some very grave indeed. For example, even if AI developments do not threaten human survival on their own, AI might be used in ways that undermine our ability to competently manage truly existential climate, nuclear, and bioengineering risks. For that reason I will not use this testimony to press the case against what its critics call AI ‘doomerism.’²⁸

For if AI is a grave existential risk to humanity, the very worst thing we could do is accept the false narrative that AI is a superhuman power beyond our control, something ungovernable. Any competent historical view tells a very different story. The lesson of the 20th century, from nuclear anti-proliferation treaties to voluntary scientific moratoria on germline engineering, to our remarkable successes in strengthening cultures of responsible, safe, and well-regulated innovation in the aviation and pharmaceutical industries, is clear. We have figured out how to govern many technologies that are new, dangerous, opaque, hard to control and contain, and challenging to predict.

²⁷ This was not possible in earlier eras of ‘democratic’ government, when ‘peoples’ could still not rule themselves, but small bodies of wealthy, elite men might vote as equals as they ruled the rest.

²⁸ Matteo Wong, ‘AI doomerism is a decoy,’ *The Atlantic*, Jun 2 2023.

<https://www.theatlantic.com/technology/archive/2023/06/ai-regulation-sam-altman-bill-gates/674278/>

We cannot make any new technology perfectly safe and risk-free, AI included. Even well-designed governance systems can fail to prevent harm, or fall victim to regulatory capture. But we absolutely *can* govern AI in ways that enable public trust. We currently lack the political will, but certainly not the capacity, because we've done it before.

Millions of people will board an airplane in this country today without a second thought, knowing that it is statistically far safer than driving. Yet in the 1960s, airplanes fell out of the sky with alarming regularity. Enhancing governance of the aviation sector, by better incentivizing safety investments, mandating global industry cooperation with regulators, and capping liability for responsible actors, *did not halt innovation*. Planes eventually stopped falling out of the sky every week, but they also got more efficient, and faster, with new features and services, while carrying more passengers than ever before.

Of course, AI presents new governance challenges. We cannot use the template of civil aviation regulation for AI; governance is not a copy and paste operation. Yet every high-stakes technology has presented unique and novel governance challenges, and until unconditional surrender to powerful corporate interests and lobbies became our political default, we didn't let that stop us.

For example, my PhD student Bhargavi Ganesh has studied the U.S. history of steamboat regulation in relation to the challenges of AI governance, and in April 2023 as part of our BRAID policy program, she presented her early findings to staffers at the UK's Office for AI and Department for Digital, Culture, Media and Sport (DCMS) as a source of lessons for how to successfully govern AI. The U.S. in the 19th century was an ambitious pioneer in successful regulation of a novel, powerful technology like the steam engine, which held huge potential economic benefits but had a perverse habit of violently blowing people up in ways that even technical experts could often not predict or fully explain.

This was not seen as a reason to delay action. Just as with AI today, public fears and distrust of steamboat travel were a serious problem for the industry. Starting in 1838, the U.S. federal government took bold, quick steps, in several stages, regulating flexibly and iteratively, adjusting various industry, operator, and regulatory incentives through successive legislative and professional bodies until the desired results were achieved. These efforts built up the marine safety code and culture that we still rely on today.

It wasn't a one-shot success. The first few attempts didn't create all the right incentives, or empower regulatory authorities to succeed. It took decades for steamboat governance to stabilize and for a marine safety culture to mature. 20th century governance of civil engineering, automobiles, aviation, nuclear power, medical devices, and pharmaceuticals all had their own growing and learning pains, yet each took valuable lessons from the

regulatory successes and failures before them. How long before we start with AI? If it might take a decade or more to see the full effect, can we afford to wait? The bills this Committee has passed are an excellent start down this road. The Biden administration's new executive order is similarly welcome, yet limited in its enforcement mechanisms. It will take a concerted and bipartisan effort of Congress to provide them.

What should our priorities be for meaningful and democratic AI governance? Here are three practical recommendations, developed with input from the Ada Lovelace Institute, our partners in the UK BRAID program:

1) Require AI developers to conduct pre-deployment assessments of potential impact on fundamental rights, with participation from affected groups and users – most urgently for public sector applications of AI for decision support. This follows recent moves by the Dutch government²⁹ in the aftermath of the childcare benefits AI scandal which forced the government to resign in 2021, after causing immense suffering for thousands of innocent families and children.³⁰ Mandating fundamental rights algorithmic impact assessments (FRAIA) would also complement the 2022 AI Workforce Training Act, by enabling those tasked with procuring AI systems for public sector use to get the documentation and evidence they need to make informed decisions. Such a regulatory requirement will require careful attention to ensuring that private companies have adequate incentives and capacity to meaningfully comply, but over time the benefits could be considerable.³¹

2) Require independent third-party audits for high-risk AI systems throughout their lifecycle. Algorithmic impact assessments are essential for responsible deployment decisions, but they cannot guarantee that an AI system's safety profile will be as expected once deployed, or stable over time. Downstream interactions or subtle changes in the deployment environment can radically alter a model's performance or outcomes. Post-deployment audits are therefore essential for AI systems that impact people's fundamental rights and opportunities. Third-party audits can be conducted by a regulator, or another entity that follows standardized practices and procedures and is accountable to a regulator

²⁹ Janneke Gerards, Mirko Tobias Schäfer, Arthur Vankan and Iris Muis, 'Impact Assessment Fundamental Rights and Algorithms,' Mar 31 2022. <https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms>

³⁰ Melissa Heikkilä, 'Dutch scandal serves as a warning for Europe over risks of using algorithms,' *Politico*, Mar 29 2022. <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/>

³¹ Andrew D. Selbst (2021), 'An institutional view of algorithmic impact assessments,' *Harvard Journal of Law and Technology* 35:1, <https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf>

(as in financial services or product safety testing).³² The U.S. could lead in incentivizing and building out the necessary infrastructure for a third-party audit ecosystem.

3) Institute stronger mechanisms for contestability, liability, and redress for avoidable and significant AI harms, to re-internalize the costs of preventable harm and developer negligence currently being imposed upon vulnerable publics. Such mechanisms have played a vital role in building and sustaining the safety cultures of civil aviation, civil engineering and pharmaceutical development. Currently, victims whose fundamental rights are violated by careless AI design and deployment have few paths, if any, to seek recourse. The paths that do exist are often unaffordable or inaccessible to the most vulnerable and impacted groups. New systems of legal liability, contestability and redress are needed to incentivize AI developers and deployers to meet a high standard of care.³³

5. Conclusion³⁴

Some AI leaders, like Meta's Yann LeCun, conclude that because today's AI tools aren't a likely road to AGI, they pose no grave threat to humanity.³⁵ Unfortunately, these techno-optimists are also mistaken. The threat is there. Many have just misunderstood its nature, because without a philosophical and historical perspective, it can be hard to see that the danger is not really from AI itself.

Had AI arrived at a different historical moment, in a period when global confidence and commitment to democratic norms and values was more robust and secure, I do not think any of the risks we see today would be unmanageable or 'existential' in nature. The extremity of the danger from AI arises from our own current weakened moral and political condition, which has compromised our collective will to legitimize and exercise appropriate control over this new form of power, or direct it to just and beneficial ends.

It is this weakened condition that enables some to see the prospect of automating human culture, judgment and meaning as a *selling point*. Marketers of AI-powered plot generating apps for both children and adults now advertise as a benefit the chance to surrender the most vital parts of storytelling: envisioning where a story might go, what a character's

³² Deborah Raji's work on this subject leads the field; see most recently Inioluwa Deborah Raji (2022), 'From algorithmic audits to actual accountability: Overcoming practical roadblocks on the path to meaningful audit interventions for AI governance.' In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIIES '22)*. Association for Computing Machinery, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3514094.3539566>

³³ AI accountability and recourse are core priorities of our BRAID research program in the UK; see also this related report from our partner The Ada Lovelace Institute: Ugo Pagallo (2022), 'The way ahead on AI liability issues,' <https://www.adalovelaceinstitute.org/blog/the-way-ahead-on-ai-liability/>

³⁴ Elements of this conclusion are adapted from Vallor (2023), published Aug 12 2023 in BBC Science Focus: <https://www.sciencefocus.com/future-technology/will-ai-make-humans-dumber>

³⁵ <https://twitter.com/ylecun/status/1637603426682150912?s=20>

backstory and motivations might be, or what unexpected futures we might open from the present. Millions now use generative AI chatbots to summarize scientific research or government meetings; no need to bother deciding for yourself what novel or important ideas are worth remembering. Your child can use a chatbot to summarize the lecture they didn't attend, and write the exam demonstrating that they understood it. Their professor can then use the very same chatbot to generate the feedback the student receives.

After all, why not liberate ourselves from the work of forming and articulating our own thoughts, telling our own stories, and making our own decisions about what matters? That prospect chilled cybernetics pioneer Norbert Wiener, who saw the power of encoding thoughts into language as the source of our unique cognitive liberty and the heart of our political capacity, "as specifically human as any interest can be. *Speech is the greatest interest and most distinctive achievement of man.*"³⁶ And yet today, large language AI models are built and sold to do the speaking *for us*, as well as the thinking and judging that our power of speech enables (whether we speak audibly or by other signs).

Those who protest handing over our humane capacities for thinking, creating and self-governing to mechanical 'Skinner boxes,'³⁷ are now often met with the cynical proposal that we ourselves are nothing more than meat Skinner boxes. What more are humans, really, than helpless stimulus-response machines, for whom self-determination is just a comforting illusion?³⁸

With AI barely on the horizon, Hans Jonas warned us in 1984 of the existential risk of a future that celebrates the "quenching of future spontaneity in a world of behavioral automata," putting "the whole human enterprise at its mercy."³⁹ He didn't say whether these automata would be machines or people. I think the ambiguity was intended.

On social media and commercial tech stages, generative-AI evangelists are now asking: what if the future is merely about humans *writing down the questions*, and letting something else come up with the answers? That future is an authoritarian's paradise. Self-governance – not just the ability, but the *desire* and *will* to jointly author our own futures and tell our own stories – is the perpetual enemy of unaccountable power.

³⁶ Wiener (1954), p. 85.

³⁷ <https://twitter.com/neilturkewitz/status/1662495973438881795?s=20>

³⁸ For a compelling philosophical challenge to this view, see John Martin Fischer, 'Some scientists say we don't have free will. As a philosopher I say, of course we do,' *The Los Angeles Times*, Oct 22 2023. <https://www.latimes.com/opinion/story/2023-10-22/humans-free-will-biology-neuroscience>

³⁹ Jonas (1984), p. 118.

Dismantling a democratic way of life is costlier and riskier than convincing people that the rare and hard-won treasure they hold is worthless. In 1911, philosopher and mathematician Alfred North Whitehead claimed that “civilization advances by extending the number of important operations which we can perform without thinking of them.”⁴⁰ But if you want a future for democratic ways of life, ask yourself: what thoughts do the civilized *keep*?⁴¹

AI is causing very real harms right now, at scale, from algorithmic discrimination and disinformation, to growing economic inequality, to the surging environmental costs of training computationally intensive models. These urgently demand robust, reasonable, and sustained regulatory and political action to incentivize responsible cultures of safe and trustworthy AI development and use. Without these measures, we will see continued growth of public fear and distrust of AI, and suppressed adoption of its beneficial uses.

But if we fear a future without humanity, or a future without democratic freedoms and meaningful human agency, AI isn’t what will steal it from us. The question upon which the future of democracy hangs, and with it our fundamental liberties and capacity to live together and thrive on this planet for very much longer, is not ‘what will AI become, and where is it taking us?’ That question is only asked by someone who wants you to believe that *you’re already out of the drivers seat*.

The real question is the one that the most prescient philosophers of technology and computing pioneers have been asking for 75 years now: what kind of future with AI will we and our elected leaders choose to preserve and sustain, with the power we still retain? One where human autonomy, judgment, and decisions *matter*? Or one where they don’t?

Much of the talk about AGI and existential risk is a dangerous distraction from what’s going on right in front of us. It’s not a violent uprising by machines. It’s a slow, quiet *human* devaluation of the political and cultural currency of our own capacity for collective wisdom and self-governance. That’s the endgame. Our humanity is the stake.

Thank you to the Chairman, the Ranking Member and the Members of the Committee for the opportunity to testify today.

⁴⁰ Alfred North Whitehead. *Introduction to Mathematics*. New York: Henry Holt, 1911, p. 46.

⁴¹ Shannon Vallor (2021), ‘The Thoughts the Civilized Keep,’ *Noema*, Feb 2 2021.
<https://www.noemamag.com/the-thoughts-the-civilized-keep/>

“—
“We are being told to shut down investigations to go hand out sandwiches and escort migrants to the shower and sit with them while they’re in the hospital and those types of tasks.”

~ DHS Whistleblower

JOSH HAWLEY
MISSOURI
115 RUSSELL SENATE OFFICE BUILDING
TELEPHONE: (202) 224-6154
FAX: (202) 228-0526
WWW.HAWLEY.SENATE.GOV

United States Senate
WASHINGTON, DC 20510-2509

COMMITTEES
JUDICIARY
HOMELAND SECURITY
AND GOVERNMENTAL AFFAIRS
ENERGY
SMALL BUSINESS
AND ENTREPRENEURSHIP

November 8, 2023

The Honorable Joseph V. Cuffari
Office of Inspector General
U.S. Department of Homeland Security
245 Murray Lane SW
Washington, DC 20528

Dear Inspector General Cuffari,

I write to request an immediate investigation into whistleblower allegations that the Department of Homeland Security (DHS) has shut down or delayed investigations into child exploitation and other crimes in order to transfer hundreds of special agents to process illegal immigrants at the southern border. No fewer than four separate DHS whistleblowers have contacted my office alleging a severe mismanagement of staffing and resources at the direction of senior officials within DHS, allegations which may demonstrate a violation of law. These allegations merit an immediate investigation.

On October 31, 2023, I questioned Secretary of Homeland Security Alejandro Mayorkas before the Homeland Security and Governmental Affairs Committee (HSGAC) regarding whistleblower allegations that he removed 600 Homeland Security Investigations (HSI) special agents from felony investigations to facilitate immigration processing functions at the southern border. According to the whistleblower, agents "are being told to shut down investigations to hand out sandwiches and escort migrants to the shower." This led to child exploitation suspects and fentanyl dealers evading indictment, including child molesters. When asked whether special agents had been pulled away from cases and reassigned to the border, Secretary Mayorkas testified, "Combatting the fight against fentanyl, yes."

Following the hearing, three more whistleblowers from across the country contacted my office, corroborating these claims. One alleged that Secretary Mayorkas was "absolutely lying" and that agents were not, in fact, being reassigned to investigate fentanyl cases. The whistleblower corroborated the claim that there have been at least 600 agents at a given time reassigned to the border. A third whistleblower claims that he was also reassigned to the border to "babysit" illegal immigrants; his duties included holding doors for private contractors who prepared the sandwiches for illegal immigrants, who were "fed better than my kids." A fourth whistleblower confirmed that special agents had been pulled off child exploitation investigations. According to some of the whistleblowers, much of their time was spent on "stand-by," as offices at the border were not expecting their arrival.

Most concerning, all the whistleblowers allege that important criminal investigations stalled and suspects were not arrested or indicted as a result of these reassignments. Agents were pulled from investigations ranging from child exploitation to drug trafficking to counterterrorism. One whistleblower was pulled from child exploitation cases "that haunt your soul when you aren't

available and able to ensure the children involved aren't being abused." Another whistleblower alleges that hands-on child molesters were not arrested and further provides specific examples of child exploitation indictments that were delayed or adversely affected due to these reassignments.

These whistleblowers have also provided documentation supporting their claims. For example, documents provided to my office show that HSI supervisors acknowledged the problem of HSI special agents sitting around without being assigned tasks. Additionally, the whistleblowers provided my office with a Memorandum of Agreement (MOA) between U.S. Customs and Border Protection (CBP) and U.S. Immigration and Customs Enforcement (ICE) to detail HSI special agents to the southern border. The MOA confirms the whistleblowers' allegations that their duties would include "[p]hysically observing [migrants] in holding areas," sitting with migrants at the hospital, and escorting migrants from place to place. Notably, the MOA was signed by the Acting Commissioner of CBP and a senior ICE official in April 2023, though the reassignments allegedly have been occurring for over a year. My office will transmit relevant documents to inform your investigation in a manner that protects the anonymity of whistleblowers.

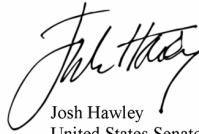
The Department's decision to relegate its highly-trained criminal investigators to processing and feeding illegal immigrants, rather than investigating child exploitation cases, is bad policy. But it also may be illegal and in violation of existing regulations. For example, whistleblowers allege that these practices violate 13 U.S.C. § 1301(a) of the Anti-Deficiency Act, Office of Management and Budget (OMB) Circular A-76, and internal ICE travel policies. Under the Inspector General Act of 1978, as amended, you have the authority to conduct investigations into the programs and operations of DHS. I therefore write to request an immediate investigation into these practices.

The American public deserves transparency and accountability, especially when it concerns the safety and security of our nation and the integrity of our law enforcement agencies. Please include in your investigation answers to the following questions:

1. Since January 2021, how many HSI special agents have been reassigned to the southern border to perform migrant processing functions?
 - a. How many of these special agents were required to pause or abandon investigations into criminal activity due to their reassignments?
 - b. How many of the investigations delayed or abandoned because of reassignments involved child exploitation, fentanyl trafficking, or counterterrorism?
 - c. How is DHS measuring the effectiveness of these reassignments in reducing the number of migrants at the southern border?
2. How many complaints identifying potential waste, fraud, and abuse have DHS officials received as a result of HSI reassignments?
 - a. Have any whistleblowers been subject to retaliatory action as a result of their complaints?

3. Are HSI special agents able to defer their reassignments in order to continue critical investigations?
4. Did the Department violate any provisions of law or internal policies by reassigning special agents to the border, including but not limited to violations of the Anti-Deficiency Act, OMB Circular A-76, or ICE travel policies?
5. Secretary Mayorkas testified to Congress that special agents had been reassigned to the border to combat the fight against fentanyl. Was this statement true?

Sincerely,



Josh Hawley
United States Senator