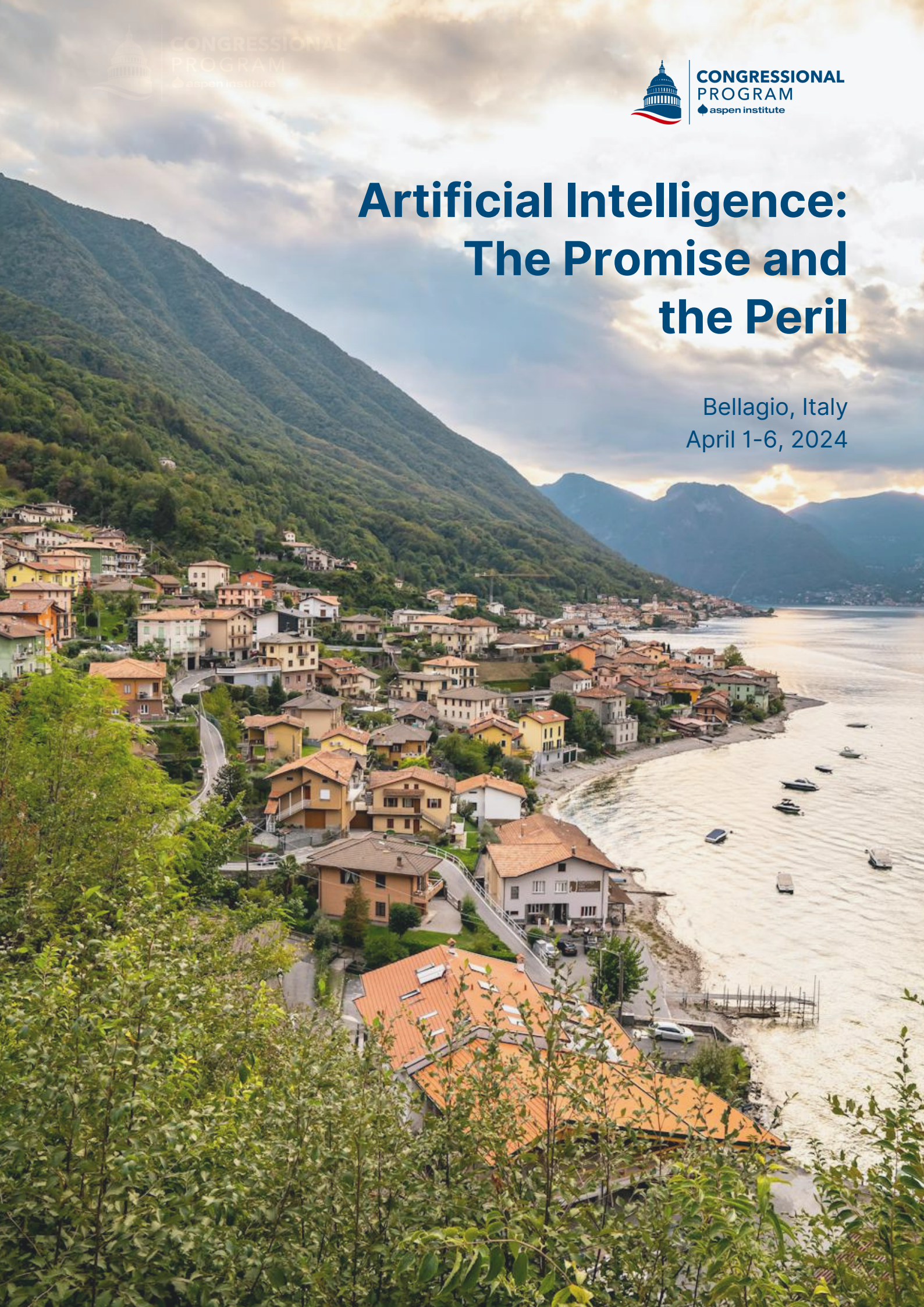
CONGRESSIONAL PROGRAM
aspen institute

Artificial Intelligence: The Promise and the Peril

Bellagio, Italy
April 1-6, 2024



Artificial Intelligence: The Promise and the Peril

April 1-6, 2024 | Bellagio, Italy

TABLE OF CONTENTS

AGENDA

MONDAY, APRIL 1

U.S. participants depart the U.S.

TUESDAY, APRIL 2:

U.S. participants arrive in Bellagio, Italy throughout the day.

7 - 9 PM: **Working Dinner**
Seating is arranged to expose participants to a diverse range of views and provide the opportunity for a meaningful exchange of ideas. Scholars and lawmakers are rotated daily.

WEDNESDAY, APRIL 3:

8 - 9 AM: Breakfast

9 - 9:15 AM: Introduction and Framework of the Conference
This conference is organized into roundtable conversations, working lunches and pre-dinner remarks. This segment will highlight how the conference will be conducted, how those with questions will be recognized, and how responses will be timed to allow for as much engagement as possible.

Speaker:
Charlie Dent, *Executive Director and Vice President, Aspen Institute Congressional Program*

9:15 - 11 AM: **Roundtable Discussion**
Artificial Intelligence 101

In this session, AI industry experts will cover the basics of Artificial Intelligence (AI), including the underlying technology, the history of AI and Machine Learning, current use cases, and what the future may look like.

Speakers:
Raffi Krikorian, *Chief Technical Officer, Emerson Collective*
Alondra Nelson, *Harold F. Linder Professor of Social Science, Institute for Advanced Study*

11 - 11:15 AM: **Break**

11:15 AM – 1 PM: Roundtable Discussion
Geopolitics of Artificial Intelligence

Perhaps more than any other technology, the global competition around AI is both a matter of national security as well as an economic and innovation race. This session will provide an overview of AI from a global perspective, highlighting security risks and international policies.

Speakers:
Klon Kitchen, *Managing Director, Beacon Global Strategies*
Eva Maydell, *Member of the European Parliament*

1 - 2 PM: Working Lunch
Discussion continues between members of Congress and experts on geopolitics of artificial intelligence with Klon Kitchen and Eva Maydell.

2 - 5 PM: Individual Discussions
Scholars will be available to meet individually with members of Congress for in-depth discussion of ideas raised in the morning sessions, including Raffi Krikorian, Alondra Nelson, Klon Kitchen, Eva Maydell, David Rhew.

5 - 6 PM: Pre-Dinner Remarks
New Frontiers: AI and Healthcare

The intersection of AI and healthcare represents a new frontier in medical innovation, promising transformative advancements across diagnosis, treatment, and patient care. Artificial intelligence applications in healthcare range from predictive analytics and personalized medicine to robotic surgeries and drug discovery. Machine learning algorithms analyze vast datasets, enabling early detection of diseases, optimizing treatment plans, and improving overall healthcare outcomes. However, this technological frontier comes with ethical and regulatory challenges, including privacy concerns, data security, algorithmic bias, and the need for transparent decision-making. Members of Congress will learn about how AI's ability to process and interpret complex medical information can accelerate research and development but also surface structural inequities.

Speaker:
David Rhew, *Chief Medical Officer, Microsoft*

7 - 9 PM: Working Dinner
Seating is arranged to expose participants to a diverse range of views and provide the opportunity for a meaningful exchange of ideas. Scholars and lawmakers are rotated daily. Discussions will focus on the role of AI in medicine.

THURSDAY, APRIL 4:

6:30 - 9 AM: Breakfast

9 - 11:15 AM: Roundtable Discussion:
Deepfakes and Democracy: the Looming Collision of AI and Elections

In 2024, almost one billion people around the globe will vote in national elections. As malicious actors continue their assault on democratic processes worldwide, AI tools can be weaponized to wrongly influence people. In this session, experts will lay out specific threats enabled by AI, including deep fakes, microtargeting of voters, and automated content distribution. The panelists will review current and potential guardrails to keep democracy on track.

Speakers:
Chris Krebs, *Chief Intelligence and Public Policy Officer, SentinelOne*
Vivian Schiller, *Vice President and Executive Director, Aspen Digital, Aspen Institute*

11:15 - 11:30 AM: Break

11:45 AM - 1 PM: Roundtable Discussion:
The Good News: AI and Innovation in Education, Healthcare, and Climate

The integration of artificial intelligence brings promising prospects for innovation in various critical sectors. **In education**, AI is revolutionizing personalized learning experiences, tailoring educational content to individual needs and optimizing teaching methodologies. It opens up new avenues for adaptive learning systems, equipping students with the skills necessary for the evolving job market. **In healthcare**, AI applications are enhancing diagnostic accuracy, streamlining administrative tasks, and accelerating drug discovery processes. **In climate change**, AI is playing a pivotal role from optimizing energy consumption to facilitating predictive modeling for environmental changes. This panel will help Members of Congress uncover how to ensure that AI-driven innovations benefit society and promote inclusive, equitable access to these technologies.

Speakers:
Vilas Dhar, *President and Trustee, Patrick J. McGovern Foundation*
Anna Makanju, *Vice President, Global Affairs, OpenAI*

1 - 2 PM: Working Lunch
Discussion continues between members of Congress and scholars on addressing the looming collision of AI and elections.

2 - 5 PM: Individual Discussions
Scholars will be available to meet individually with members of Congress for in-depth discussion of ideas raised in the morning sessions, including Chris Krebs, Vilas Dhar, Vivian Schiller, and Anna Makanju.

6 - 7 PM: Pre-Dinner Remarks
Building AI We Can Trust

New techniques have significantly improved the traditional, costly, and inefficient way to create and deploy AI models. This offers exciting new possibilities for increasing innovation, efficiency, and productivity. However, with the benefits come also additional risks besides those already considered in traditional machine learning models, and questions about the safety, development, deployment, and use of generative AI. The Members of Congress will learn about the latest techniques to build governance into the heart of the AI lifecycle (from data to models and applications), align AI models with values, identify and mitigate hallucination and other risks, address biases, and build AI guardrails. The discussion will include how AI is being used to advance scientific discovery, as well as initiatives to accelerate progress and increase participation and the broad diffusion of the benefits of AI through open innovation.

Speaker:
Darío Gil, *Senior Vice President and Director of Research, IBM*

7 - 9 PM: Working Dinner
Seating is arranged to expose participants to a diverse range of views and provide the opportunity for a meaningful exchange of ideas. Scholars and lawmakers are rotated daily. Discussions will focus on AI ‘s role in innovation and democracy.

FRIDAY, APRIL 5:

6:30 - 9 AM: Breakfast

9 - 11 AM: Roundtable Discussion
AI and the Future of Labor Market

There are mixed perceptions about AI's role in a new labor market: AI will open up new labor markets and improve employee satisfaction or lead to an epic loss of jobs. However, everyone agrees that AI will change the nature of work and workers. This panel will uncover how companies are approaching these opportunities and risks.

Speakers:
Athina Kanioura, *Executive Vice President, Chief Strategy and Transformation Officer, PepsiCo*
Mike Haley, *Senior Vice President, Research, Autodesk*

11 - 11:15 AM: Break

11:15 AM - 1 PM: Individual Discussions
Scholars will be available to meet individually with members of Congress for in-depth discussion of ideas raised in the morning sessions, including Athina Kanioura and Mike Haley.

1 - 2 PM: Working Lunch
Discussion continues between members of Congress and experts on policy takeaways from the conference.

2 - 5 PM: Policy Reflections for Members of Congress
This time is set aside for Members of Congress to reflect on what they learned during the conference and discuss their views on implications for U.S. policy.

7 - 9 PM: Working Dinner
Seating is arranged to expose participants to a diverse range of views and provide the opportunity for a meaningful exchange of ideas. Scholars and lawmakers are rotated daily. Discussions will focus on policy takeaways from the conference.

SATURDAY, APRIL 6:

6:30 - 9 AM: Breakfast

Conference participants depart to the U.S.

CONFERENCE PARTICIPANTS

MEMBERS OF CONGRESS AND THEIR SPOUSES:

Rep. Nanette Barragan

Rep. Don Beyer and Megan Beyer

Rep. Kat Cammack and Matt Harrison

Sen. Tom Carper and Martha Carper

Sen. Bill Cassidy and Laura Cassidy

Rep. Neal Dunn and Leah Dunn

Rep. Anna Eshoo

Rep. Scott Franklin and Amy Franklin

Rep. Garret Graves and Carissa Graves

Rep. Michael Guest and Haley Guest

Rep. Jim Himes and Mary Himes

Rep. Glenn Ivey and Jolene Ivey

Rep. Dave Joyce and Kelly Joyce

Rep. John Joyce

Rep. Annie Kuster and Brad Kuster

Rep. Darin LaHood and Kristen LaHood

Rep. Rick Larsen and Tiia Karlén

Rep. Ted Lieu and Betty Lieu

Rep. Greg Murphy and Wendy Murphy

Rep. Guy Reschenthaler and Jennifer Drogus

Rep. Linda Sanchez

Sen. Chris Van Hollen and Katherine Wilkens

SCHOLARS AND EXPERTS:

Vilas Dhar	*President and Trustee, Patrick J. McGovern Foundation*
Darío Gil	*Senior Vice President and Director of Research, IBM*
Mike Haley	*Senior Vice President, Research, Autodesk*
Athina Kanioura	*Executive Vice President, Chief Strategy and Transformation Officer, PepsiCo*
Klon Kitchen	*Managing Director, Beacon Global Strategies*
Chris Krebs	*Chief Intelligence and Public Policy Officer, SentinelOne*
Raffi Krikorian	*Chief Technical Officer, Emerson Collective*
Anna Makanju	*Vice President, Global Affairs, OpenAI*
Eva Maydell	*Member of the European Parliament*
Alondra Nelson	*Harold F. Linder Professor of Social Science, Institute for Advanced Study*
David Rhew	*Global Chief Medical Officer & Vice President of Healthcare, Microsoft*
Vivian Schiller	*Vice President and Executive Director, Aspen Digital, Aspen Institute*

CONFERENCE RAPPORTEURS:

Kristine Gloria	*Director of Strategic Partnerships & Innovation, Young Futures*
Matthew Rojansky	*Rapporteur and Counselor to the Aspen Institute Congressional Program*

FOUNDATION REPRESENTATIVES:

Brad Carson	*President, University of Tulsa*
John Dedrick	*Executive Vice President and Chief Operating Officer, The Charles F. Kettering Foundation*
Danielle Geanacopoulos	*Managing Director for Government Relations, The Rockefeller Foundation*
Adelaide Park Gomer	*President, The Park Foundation Board*

ASPEN INSTITUTE CONGRESSIONAL PROGRAM:

Charlie Dent and **Pamela Dent**	*Executive Director, Congressional Program and Vice President, Aspen Institute*
Tyler Denton	*Deputy Director*
Carrie Rowell	*Conference Director*
Jennifer Harthan	*Manager of Congressional Engagement*

CONFERENCE SUMMARY

Kristine Gloria
Rapporteur and Director of Strategic Partnerships and Innovation, Young Futures

Matthew Rojansky
Rapporteur and Counselor to the Aspen Institute Congressional Program

Introduction

From April 2-5, 2024, the Aspen Institute Congressional Program gathered 22 bipartisan, bicameral members of Congress in Bellagio, Italy to discuss the perils and promise of artificial intelligence. Throughout the conference, Members heard from and engaged with experts from across various domains and sectors, inclusive of civil society, national security, the tech industry, and other sectors of business.

Over a dozen experts from the U.S. and EU parliament shed light on the many ways artificially intelligent systems—from large language models to manufacturing optimization to ambient listening in patient care—show up throughout our lives, every day. Given the prevalence and scale of the technology, experts homed in on the role of the U.S. federal government in shaping AI research, development, deployment, and adoption both domestically and internationally. As one expert shared, "the headline is that there is exponential change and growth and dynamism (in AI tools), but we're still very much in the early stages, which means that there is a tremendous political opportunity to create the kind of world system that we want with these tools."

As the title of the conference suggests, AI exists in tensions. It is both uniquely transformative and potentially disruptive; simultaneously liberating and isolating; generative and destructive. The promises offered by technologies, such as artificial intelligence, are vast and currently unknowable. Already, we are witnessing just how quickly AI can change traditional workflows, industries, sectors, and our own understanding of what it means to be a human. More acutely, AI has exposed critical gaps in how the U.S. is currently managing the potential risks of the tech and its impact on citizens.

This report highlights the key conference takeaways and policy ideas that came forth amid the conference sessions.

Artificial Intelligence 101

The term **artificial intelligence** refers to a constellation of technologies designed to emulate human intelligence,[1] which now includes specific techniques such as machine learning, automated reasoning, and neural networks. Examples of AI today include

[1] Artificial Intelligence 101, Aspen Digital

ChatGPT, autonomous vehicles, voice recognition, and recommendation systems such as on music and video apps.

As experts highlighted, the journey of AI began in the mid-1950s during a summer workshop at Dartmouth College. Since then, a combination of milestones—from increased compute power to new mathematical techniques—have accelerated and redefined what we consider to be artificial intelligence. The term **generative AI** is the latest example. Generative AI, often most associated with applications like ChatGPT, is a subset of artificial intelligence technologies that creates new content, whether it be text, image, or sound. Unlike previous AI systems that are centered on decision-making outputs, generative AI is capable of both analyzing and creating content.

No matter the specific type of AI system, all rely heavily on massive amounts of data to optimally and reliably function. This data can come in essentially any form—from text, to images, to voice, to discrete numbers, to user-generated content, etc. The data is then used to "train" or teach mathematical models (aka representations) on various computational tasks such as pattern matching, automated reasoning, and predictive statistics. Experts highlighted that to have high-fidelity models, the data layer is critical. This includes having high-quality, diverse, accurate, and complete, datasets that allow for generalizability in model outputs. If the data used to train a model is incomplete or biased, the model will reproduce and/or amplify these flaws. In other words, garbage in, garbage out.

Three-Legged Stool and Governance

Experts offered several frameworks for how to formulate safeguards against harms and risks potentially posed by AI. The first included the idea of a **three-legged stool**, comprised of tech companies, civil society (e.g. activists, researchers, and philanthropy), and the government (e.g. public policymakers and regulators). Each leg of the stool is critical, yet "the tech companies dominate the conversation, and we need to figure out how to raise the prominence of the other two legs," said one expert. To do so, the expert offered a few ideas of note. First is the need to address the academic capture of research labs and scholars across university campuses. Due to a lack of public funding and resources such as the GPU processing power necessary to conduct AI research, academics turn to private industry for support. This heavy reliance on private industry results in a lack of independence in the research, calling into question the credibility and trustworthiness of results. Moreover, the absence of a robust and independent academic research community means there is no one holding industry and practitioners accountable.

In addition to increased funding of research efforts such as the National Science Foundation (NSF) National Artificial Intelligence Research Resource (NAIRR), one expert called for a national commitment to raise public education around artificial intelligence. For example, in 2018, Finland in partnership with the University of Helsinki, created, promoted, and distributed an AI online webinar that discussed pros and cons of AI. At the end of the pilot, 10% of the total Finnish population watched the

webinar, including citizens whose professional careers are not traditionally AI focused. The expert acknowledged that while this type of education opportunity could be helpful in educating the American citizenry, the practicality of cutting and pasting the Finnish approach would be difficult. Instead, for this type of campaign to be effective, the content would need to be context-relevant across various regions and demographics in the U.S.

Building off this discussion, another expert introduced the framework of **AI Governance**, which shares elements with the three-legged stool but emphasizes the need for norms and standards that move alongside with regulatory efforts. "How we think about governance needs to be how do we need to work together and in new multi-sector ways to get at the outcomes and the good stuff of AI?" posited the expert. This includes recognizing various tensions around private sector innovation, public good, social responsibilities, and the incentives that drive market value. The scholar offered the multilateral research collaboration at CERN (the European nuclear research laboratory) as an example of the good that could be unlocked with a sound commitment to innovation, established norms, and the "the need to control as necessary." Like AI, the efforts at CERN face national security implications, budget constraints, data flow and control considerations, etc. And, yet CERN is where we got the World Wide Web.

"We want to regulate towards outcomes, not the objects," said the scholar. "AI is an elusive object and not a thing you can regulate. What you can do is create the pathway that you need for the outcomes that we want." This includes utilizing existing laws when situationally appropriate to help manage harms and risks. For example, the group discussed in-depth the lever of liability law as it pertains to licensing or its current use in class action lawsuits against social media companies. Its use sends a signal to companies and organizations around what is acceptable or not. The challenge, as one participant noted, is this legal tool takes a long time and can be expensive. "It (liability) helps add friction to the system right away," countered one participant. "And I think that's important to have both, the friction and also the signaling it sends to the (American) people and the industry."

The group also debated in depth the use of thresholds within certain frameworks, like the recent Executive Order on Safe, Secure, and Trustworthy Development and Use of AI.[2] Specifically, in Section 4.2 on ensuring safe and reliable AI, the EO offers specific numerical thresholds around the use of floating-point operations (FLOPs) in data models. Scholars pushed back on this guidance suggesting that potential risks and harms are present at even much lower compute power. Instead, scholars pointed to a need for increased transparency and auditing, or risk assessments of the models used by practitioners and industry. This requires a commitment to building the necessary infrastructure that sits between private sector and civil society to conduct the audits. Additionally, another scholar suggested the need for better definitions around safe harbors for testing models.

[2] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (October 30, 223). https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

In addition to unlocking specific outcomes using AI, a governance approach can help establish trust with the American public that there is a plan, a national strategy to achieve good outcomes. According to the experts, American public discourse and sentiment around AI is largely pessimistic, especially compared with the enthusiasm for AI adoption from developing countries, such as India. Fears around work displacement and lack of public education on the technology are key drivers. Moreover, media coverage focused on the existential risks of AI[3] have contributed to the public's growing discontentment and distrust of the technology.

As one scholar summarized, "AI itself is a huge term that encompasses so many different things. And talking about it in these unilateral ways is insufficient. We need to work at multiple timeframes. Yes, we need someone to be thinking about these exponential risks that will kill us; but also, we should also be using some of these technologies. We need to have that opportunity to just walk and chew gum at the same time."

Geopolitics of Artificial Intelligence

Perhaps more than any other technology, the global competition around AI is both a matter of national security as well as an economic and innovation race. Participants traversed a variety of intersections from research and development to international partners, to the intelligence community, to data markets and the supply chain.

Europe and AI

The European Union has produced what some have called the "gold standard" for regulation with examples such as the General Data Protection Regulation (GDPR),[4] the Digital Markets Act (DMA) and the AI Act.[5] These laws are shaped by a vision that "ultimately lies in safeguarding our society and our democracies to allow for the positive use of AI," explained one scholar. Specifically, the EU AI Act sought to balance the three questions: how much does the EU need to protect privacy? How strict does regulation need to be? And, how much of it is a free market approach? In its current form, the AI Act is a risk-based, tiering system that categorizes potential harms into high-, med-, and low-risk use cases and implementations. For example, AI technologies that employ the use of social-scoring algorithms are deemed as high-risk use cases. "We [EU] are very good at setting the standard," noted one scholar. "But then we fail at engaging with our partners – U.S., Canada, UK, etc. – to bring those norms and standards into other bodies outside of the EU." Specifically, the GDPR has been criticized for dampening innovation, particularly for having the same standards for small- to mid-size companies as that of a large multinational company. "Our total focus has been on putting up the

[3] Pause Giant AI Experiments: An Open Letter (March 22, 2023). https://futureoflife.org/open-letter/pause-giant-ai-experiments/
[4] General Data Protection Regulation. https://gdpr-info.eu/
[5] The EU Artificial Intelligence Act https://artificialintelligenceact.eu/

guardrails but not how to export those guardrails or engage for further development," noted an expert.

This focus on strengthening global democracy, however, is in tension with European competitiveness and innovation in the AI field. As the expert noted, "These goals are not mutually exclusive, but they can clash. And it seems in the last couple of years we've [EU] been focusing only on the values-first approach, but how can we help an AI make us more economically competitive?" According to the World Economic Forum[6] and a study commissioned by Amazon Web Services, over one-third of EU businesses have adopted AI, a significant increase from prior years. Echoed in points made by scholars in the room, the key to unlocking AI's economic potential in the EU requires creating a more pro-innovation environment, addressing the digital skills gap, and ensuring access to businesses of all sizes to the latest technologies. "Our approach to emerging technologies has to change," said one expert. "We cannot address these new rapidly changing technologies with the same regulation. We need a more creative mindset. One that could make us compete economically."

In addition to economic competitiveness, defense will be top of mind during the next European Commission. This includes discussions around a Commissioner for Defense and a more concrete defense strategy for the EU. Like economic competitiveness, the scholar cautioned that current discussions around defense echo voices and thinking 10-15 years in the past. However, with AI and its geopolitical implications, calling for a European army is shortsighted. "What happened in 2013 in Crimea has changed the way Ukraine is looking at its military strategy, but it did not change the way most other European countries look at defense," said the expert. "The hard power has never been something that the EU thinks too much about; and the way the U.S. prioritizes it is fundamentally different. So, when it comes to regulatory interoperability or economic interoperability or national security and defense interoperability, just the mindset is very, very different."

Three Strategic Trends

In addition to the EU's perspective and approach, one expert outlined three strategic trends at the intersection of geopolitics and technology (with an emphasis on AI).

Foreign Policy and national security are a shared burden. Technology companies and private industry are now in geopolitical positions that require them to reconcile with the responsibilities inherent to holding this position. "No government, American or otherwise, is anywhere close to reasserting its unilateral activity," explained one scholar. "The US government is now a national security stakeholder not the national security stakeholder."[7] Therefore, the U.S. Congress is tasked with setting the conditions to allow for this partnership with industry to thrive. This includes simultaneously

[6] Over a third of EU companies adopt AI (February 9, 2024). https://www.weforum.org/agenda/2024/02/ai-regulation-digital-software-news-february-2024/
[7] Emphasis by speaker

holding industry accountable while continuing to enable the industry's advantage in innovation.

One participant questioned whether the traditional methods, such as export controls (e.g. video gaming chips), remain an effective tool for helping ensure America's advantage in technology innovation. For such materials, Congress has engaged in bipartisan and persistent efforts to constrain access to cutting edge hardware, like the latest chipsets. Yet, as one scholar noted, the impact of such controls is not a perfect solution as lower-end models of these chips are being repurposed with much utility. As for the algorithms that drive AI systems, the competitive edge between the U.S. and China is, as one scholar noted, "a jump-ball."

Technology is reshaping global alliances and vice versa. Today, the U.S. is leading the effort in shaping global alliances, in large part due to the innovation of private industry. These alliances are being formed around a few core categories: secure supply chains, trusted data flows, and defense alliances. "Even amongst our [U.S.] closest friends, as we're renegotiating these alliances, there are very real worldview differences that are complicating this," noted the scholar. "And nowhere is that clearer than in the European Union. Not for a lack of desire or intent. But we just fundamentally have different understandings of what safety is and what the proper orientation for that is." This difference in mindset is causing "friendly disagreement" but as noted, "if you do not have regulatory interoperability, you're not going to be able to build economic interoperability." For Congress, the call is to help orient policy towards these three categories and to help inform its allies.

Artificial Intelligence is an infrastructure challenge. AI systems require massive amounts of compute power, inclusive of the GPUs to networking cables to the server farms. All these elements will be shaped by strategic competition for access to this compute power. With funds in the U.S. becoming constrained, other actors are using investments to transition their economies into data and knowledge economies. China is particularly active in this area. Gulf states also have the capital, the natural resources (e.g. land), and the drive to invest in AI related infrastructure, raising the question of whether they will do so in alignment with the U.S. or China. "What's happening is everybody—up and down the value chain, whether it be the tech companies, private equity or venture capital—are engaging with these nation states," emphasized the expert. Additionally, as one scholar noted, these nation states are more politically agile in their ability to navigate challenges that might introduce friction into decision making processes. Participants spoke at length on the ideological differences between democratic and autocratic states, suggesting that values-based decision making introduces a level of inefficiency and de-stabilization. The counter offered is that while democratic models introduce messy and inefficient conditions, the autocratic model narrows the pathway for innovation. For example, if the Chinese government decides on a specific priority, then investments and efforts are funneled into realizing this one application; wherein, democratic models afford a diversity of solutions and approaches which spurs greater economic gains. For Congress, the task is to keep these persistent risks on the forefront while striking a balance in America's long-term technological, economical, and geopolitical leadership.

Artificial Intelligence and Elections

In 2024, almost one billion people around the globe will vote in national elections. As malicious actors continue their assault on democratic processes worldwide, AI tools can be weaponized to influence democratic politics. This phenomenon, however, is not new. As one scholar highlighted, "manipulated content and propaganda by any other name goes back millennia. False content on social media goes back more than ten years." Instead, the core question is whether the use of AI tools, like audio and video manipulation, makes a difference and or changes the game.

Already, we see various use cases that surface just how difficult it may be to assess the good versus bad in the use of AI within political campaigns. One problematic and recent example includes an incident where robocalls using an AI generated voice of President Joe Biden were used to discourage voters in the New Hampshire primaries.[8] On the other end, a similar application using AI-generated voice was employed by New York City Mayor Eric Adams to make robocalls in his own voice in several languages that he does not speak.[9] And, in January in Slovakia, a deep-fake audio of a leading candidate seemingly boasting about how he had rigged the election was leaked 48 hours prior to the national quiet period.[10] The fake audio recording then jumped across social media platforms and went viral. These examples illustrate that the technology itself–AI generated voice—is not inherently nefarious. Instead, its use by bad actors for ill-intent make it ethically, and in the case of the New Hampshire vote dissuasion campaign, criminal under federal law.

Harking back to the previous day's conversation on the three-legged stool, the expert outlined how each leg has responded thus far. For industry, U.S. tech companies, including Adobe, Amazon, Google, IBM, Microsoft, OpenAI, Snap, Tiktok and X, signed an accord in February 2024 pledging to combat the use of deceptive AI in this year's election. As part of this agreement, companies pledged to "work collaboratively on tools to detect and address online distribution" of AI-generated content that may deceive voters around the globe.[11] "It was necessary but not even close to sufficient," commented the scholar. "In reality, in an open-source world, these AI tools are out there."

---

[8] New Hampshire investigating fake Biden robocall meant to discourage voters ahead of primary. AP News. (January 22, 2024). https://apnews.com/article/new-hampshire-primary-biden-ai-deepfake-robocall-f3469ceb6dd613079092287994663db5

[9] Can Mayor Eric Adams speak Mandarin? No, but with AI he's making robocalls in different languages. https://ny1.com/nyc/all-boroughs/politics/2023/10/17/mayor-eric-adams-mandarin-ai-robocalls-different-languages-yiddish

[10] A fake recording of a candidate saying he'd rigged the election went viral. https://www.cnn.com/2024/02/01/politics/election-deepfake-threats-invs/index.html#:~:text=Michal%20%C5%A0ime%C4%8Dka%2C%20the%20leader%20of,manipulate%20votes%20at%20polling%20stations. CNN.com (February 1, 2024).

[11] Tech companies pledge to fight deceptive AI during 2024 elections. The Hill (February 16, 2024). https://thehill.com/policy/technology/4472837-tech-companies-pledge-to-fight-deceptive-ai-during-2024-elections/

Additionally, while the pledge is encouraging, U.S. tech companies have also taken steps that undermine the effectiveness of the accord. For example, some recently have cut-off API access to their content (e.g. notably Meta's CrowdTangle[12]), which inhibits the ability of researchers to conduct meaningful assessments on the information shared across these platforms. Moreover, many of these companies have shrunk and or totally disbanded their trust and safety and content moderation teams. The reasons for this vary and include considerations such as costs associated with the time, effort, reputation, and legal risks incurred by companies and civil society to moderate content. Industry also lacks any consistent policymaking around how to address information that seeks to deceive people. There is no clarity on a process or protocol or central body for reporting mis-, dis- and or false information. "There is no central place," said an expert, "so instead of trying to play whack-a-mole with all of the false content, there is now an effort to uplift the authoritative sources, like your local election official or secretary of state."

For the second leg of the stool, civil society has made efforts to increase the education and training of stakeholders across the election process to prepare them for how to handle deceptive AI generative content. This includes multi-year efforts to train secretaries of state, election officials, poll workers, and journalists across the nation. For media, specifically, there are efforts to help train them on how and what should be communicated to the public as it relates to AI and elections. As one expert noted, there are unintended consequences in causing a mass hysteria on a specific piece of content. The scholar also highlighted that "what worries me most is not the spectacular, high-profile deepfake which is going to be debunked; it is the content we cannot monitor. It is the private messaging happening on WhatsApp, Signal, Telegram and SMS." Complicating this further is what two academic researchers, Bobby Chesney and Danielle Citron highlighted in their paper, "Deepfakes, Elections, and Shrinking the Liar's Dividend."[13] "Concern about deepfakes poses a threat of its own. The theory is simple: when people learn that deepfakes are increasingly realistic, false claims that real content is AI-generated become more persuasive too." And, as one scholar summarized, this is not a cybersecurity or a technology issue. It is a people issue.

Cybersecurity, AI, and Elections

Shifting to a more cybersecurity perspective, one expert outlined four key questions to consider. The first is: *what are the risks of AI to elections?* Utilizing an information warfare and operations framework, the scholar broke this down into two core ideas: information technical and information psychological, which both include offensive and defensive strategies. From an offensive stance on the technical side, the scholar highlighted a recent collaboration between Microsoft and OpenAI to assess the use of their AI tech across user and employee bases. This effort identified several, seemingly benign cybersecurity risks, including social engineering and more sophisticated mouse

[12] Meta will sunset CrowdTangle in August. Search Enginge Land. March 15, 2024. https://searchengineland.com/meta-sunset-crowdtangle-august-438472

[13] *Deepfakes, Elections, and Shrinking the Liar's Dividend*. Brennan Center. January 23, 2024. https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend

traps or phishing. From a defensive stance on the technical side, AI tools like ChatGPT are simply improving basic functionality and automation within the entire cybersecurity field. For example, the scholar highlighted the use of AI in configuration scans in vulnerability reviews to help identify potential risks within a system and provide guidance on how to fix it at a much larger scale.

On the information psychological side, experts note the improvement in quality of AI-generated content across audio, video, and text, making it difficult to discern real versus fake. In addition, the information environment of today is starkly different from that of four-years ago. Now, the environment is increasingly fragmented into echo chambers between various social media platforms and closed-door messaging apps. "Today, we're just going to learn much slower and it's going to be a while for researchers and national security professionals to pick up on the signal," said one expert. Moreover, with more automated tools, the volume of content will increase. This, as the expert points out, can give the appearance of consensus and discussion around specific content that is entirely synthetic.

Second question: *what are the benefits of AI to elections and democratic participation?* Like the previous question, AI presents significant gains in efficiency resulting in faster creation and dissemination of content as well as defensive detection. Another advantage is in constituent engagement through precise messaging and targeting. For example, the use of chatbots to address very specific district or community related questions. AI can also be beneficial not just for candidates but for local community public services and officials to engage with the community.

Third question: *what are the outcomes that we need to figure out?* Or, in other words, why would these AI tools be used for elections? The obvious examples include the New Hampshire robocall mentioned above. These are overt, "do not vote," targeted messages. However, experts pointed to less obvious use cases that could also have outsized impacts on the election. This could look like, for example, synthetic texts and messages seemingly sent out by the secretary of state or local county recorder indicating that polls have closed and telling people not to vote or to staff the offices. Currently, ten states have issued legislation for the monitoring of use of AI generated content in campaigns. For example, Wisconsin now requires political ads to disclose when they use AI-generated audio or video content. A label must be clearly displayed at the beginning and end of all campaign-related audio and visual media distributed publicly.[14]

Fourth question: *what are the questions we are struggling with right now?* At the top is whether AI will really impact the public and overall outcome of the election. Unfortunately, there are not great metrics for evaluating the effects and rapid changes in technology are quickly reshaping where to even begin. Next is whether the cybersecurity or national security community can even really detect the various types of threats.

[14] Wisconsin Public Radio. Wisconsin political ads must now disclose if they include AI-generated content. March 21, 2024. https://www.wpr.org/news/wisconsin-political-ads-must-disclose-ai-generated-content#:~:text=Wisconsin%20political%20ads%20must%20now%20disclose%20if,according%20to%20a%20new%20law%20signed%20Thursday.

Participants once again discussed tools such as content provenance/watermarking and interoperability standards to help mitigate the negative effects of synthetic content. However, as one expert explained, there is no clear guidance on who will be managing watermarks or the authenticity of content. "What I expect to see is candidates going on the offensive continuously," mentioned a scholar, "claiming their own identities, voice, video, and providing a seal of approval etc., making it easier to manage." Lastly, time bound situations remain an open and unpredictable variable on how effectively the intelligence community and government can respond to election threats. "We don't have the luxury of time, whether it's a 40-hour restriction period or the morning of Election Day, to handle chaos," said one expert. "We're going to be really vulnerable and what worries me most is overreacting to it."

Scholars emphasized that AI is a tool. What degree of impact will the tool have on the upcoming elections remains unknown. What is clear, however, is the need to bolster the resilience of the U.S. election infrastructure. One that is currently seeing "a shrinking of government apparatuses" to ensure the safety and security of the process. This includes, as one participant highlighted, fear for the physical safety of election officials because of threats fueled by mis- or dis-information. "If I've learned anything in the last eight-years, radical transparency is the only way through this and to maintain trust with the public," said one expert.

AI Innovation: Education, Climate, and Healthcare

Amid the backdrop of the threats AI may pose to areas of American society, like elections, it is also a primary driver for economic prosperity with innovation across several verticals. Specifically, participants heard from experts on AI's transformative application in education, climate, and healthcare.

AI and Environment

"Climate is one of those great examples of an environment where it feels so non-digital," began one expert, "it's really a political, a storytelling, and a place where experience matters." This was partially due to a lack of data to help understand climate issues. Recently, however, global datasets that include information like geospatial satellite information and lived experiences, can be used to create predictive climate models. For example, Climate TRACE, is an open and accessible inventory to track greenhouse gas emissions with "unprecedented detail and speed, delivering information that is relevant to all parties working to achieve net-zero global emissions."[15] In addition to its predictive application, AI can be helpful in managing climate-fueled natural disasters, such as hurricanes and floods. Specifically, UN Global Pulse's Data Insights for Social & Humanitarian Action (Disha), combines satellite image data and AI-based analysis to empower on-the-ground groups from around the world to evaluate the impact of a disaster and efficiently facilitate targeted response efforts.[16]

[15] https://climatetrace.org/about
[16] https://disha.unglobalpulse.org/

One of the major costs of increased AI use is its high energy demand for computing. In January 2024, the International Energy Agency issued its forecast for global energy use over the next two years.[17] The report highlights that in 2022, almost 2% of total global electricity demand could be traced back to use by data centers, cryptocurrencies, and artificial intelligence. The International Energy Agency (IEA) cautions that this demand could double by 2026, which would "make it roughly equal to the amount of electricity used by the entire country of Japan."[18] With the democratization of generative AI driving user consumption, researchers are just beginning to uncover additional increases in energy consumption. Researchers from Hugging Face and Carnegie Mellon University have found that generating an image using a powerful AI model takes as much energy as fully charging a smartphone.[19] The reasons generative AI models use much more energy is due to multi-tasking as it attempts to classify, infer, and generate outputs simultaneously. According to researcher, Dr. Sasha Luccioni, "switching from nongenerative, good old-fashioned 'AI' to a generative one can use up to 30 to 40 times more energy for the exact same task."[20] In addition to energy consumption, AI and the data centers that power these systems need more and more water to cool them down. The "water footprint" of these systems can range from 500-ml of bottled water for a short ChatGPT conversation to 700,000 liters of clean freshwater used in U.S. data centers used to train the GPT model.[21] "The concentrated power that gets used to train frontier models, like ChatGPT and Claude, is a huge chunk," noted one expert. "And I think it's a technical challenge but also a social one to determine just how much (data center) capacity we want to build to let a few private companies go out and train these things." Efforts to make AI more environmentally sustainable are already being developed. This includes industry commitment towards utilizing more renewable resources to power data centers.

AI and Healthcare

The healthcare industry is poised to see substantial gains from the use and implementation of artificial intelligence across various use cases. One salient example is the use of generative AI tools in drug discovery. Moderna is one such pharmaceutical company integrating AI into all its central business functions, from R&D to

---

[17] https://iea.blob.core.windows.net/assets/6b2fd954-2017-408e-bf08-952fdd62118a/Electricity2024-Analysisandforecastto2026.pdf (p. 31).

[18] Vox.com. "AI already uses as much energy as a small country. It's only the beginning." (March 28, 2024). https://www.vox.com/climate/2024/3/28/24111721/ai-uses-a-lot-of-energy-experts-expect-it-to-double-in-just-a-few-years

[19] Hugging Face and Carnegie Mellon University. "Power Hungry Processing: Watts Driving the Cost of AI Deployment?" (Nov 2023) https://arxiv.org/pdf/2311.16863.pdf

[20] Vox.com. "AI already uses as much energy as a small country. It's only the beginning." (March 28, 2024). https://www.vox.com/climate/2024/3/28/24111721/ai-uses-a-lot-of-energy-experts-expect-it-to-double-in-just-a-few-years

[21] The Markup. "The Secret Water Footprint of AI technology." (April 2023). https://themarkup.org/hello-world/2023/04/15/the-secret-water-footprint-of-ai-technology

commercialization.[22] According to one expert, Moderna's use of OpenAI technology to streamline its analysis of clinical trial data saw a reduction of 84% in time spent processing documents and formulating dosage recommendations. In another example, one scholar shared a story of an organization with over 20-years of rural frontline healthcare experience addressing maternal health concerns in a remote area of the world with no digital interventions. After months of data analysis in collaboration with Yale University and the organization, researchers identified clusters of villages with consistent low-birth weights. The organization, medical professionals, and sociologists were then tasked to uncover why. Months later, researchers identified a cultural practice in the villages that led to consistent deficiencies on a particular set of vitamins. With this discovery, the organization then worked with the government to begin shipments of B-12 and other supplements. "This is a project that took 20-years of data from some of the most remote parts of the world and within a year essentially let us drive a population level health intervention that had massive health outcomes," shared the speaker.

In addition to research and discovery, AI has the potential to dramatically change the industry by addressing some of its biggest operational challenges, including healthcare waste, access to affordable care, and clinician burnout. One expert noted that in 2022, the U.S. spent $4.5 trillion on healthcare, of which, approximately 25% of that was spent on waste like inefficient operations and disjointed care. Moreover, according to both the American Medical Association and American Nursing Association, over 60% of physicians and nurses report burn-out, leading many to leave the practice of medicine causing downstream effects of workforce shortages. One major contributing factor to burnout is the increasing amount of administrative work, such as addressing insurance claims, documenting in electronic health records, and filling out forms. In February 2024, the New England Journal of Medicine highlighted potential "low-hanging fruit" applications in which AI tools, including generative-AI, could be most useful. This includes areas of prioritization and analysis of imaging results in radiology, differential diagnosis, enhanced doctor-patient interaction, appointment scheduling, etc.[23] Authors of the article emphasized the need to position AI as a "complementary tool rather than a replacement in health care."

Already the industry is exploring how to operationalize responsible AI principles as these new tools come on board. One example is the Coalition for Health AI (CHAI), a non-profit that works in collaboration with multiple stakeholders to define the values, purpose, and practices that ensure the safe and effective use of AI by the industry.[24] Another example is the Trustworthy and Responsible Health AI Network (TRAIN), which launched in March 2024. TRAIN is a consortium of 16 healthcare organizations and Microsoft (as the technology-enabling partner) that aims to "improve the

[22] Forbes.com. "How Moderna Is Embracing Data & AI To Transform Drug Discovery." (March 25, 2024). https://www.forbes.com/sites/randybean/2024/03/25/how-moderna-is-embracing-data--ai-to-transform-drug-discovery/?sh=34e3852afed6

[23] New England Journal of Medicine. "To Do No Harm — and the Most Good — with AI in Health Care." https://ai.nejm.org/doi/full/10.1056/AIp2400036 . February 22, 2024. (presented in David Rhew's essay)

[24] Coalition for Health AI. https://www.coalitionforhealthai.org/

trustworthiness of AI by sharing best practices" through an online registry that captures real-world outcomes among network members.[25]

AI and Education

AI intersects education in multiple ways. First is the development and use of AI tools within classrooms, from curriculum development to AI-powered individualized instruction. According to one expert, "the primary benefit of generative AI is its ability to provide students and educators personalized instruction that resource constraints would not otherwise allow." For example, the scholar highlighted a collaboration with UNICEF to develop an interactive textbook for disabled kids around the world. Using the multi-modal capabilities of ChatGPT (aka text and voice), the textbook can respond to requests such as language translation. In another example, for students who are visually impaired, text-to-speech can be utilized to help convey the information. The scholar also shared a report from the Walton Family Foundation that found 51% of teachers, including 69% of both Black and Latino educators, are already using tools such as ChatGPT.[26]

For some conference participants, the introduction of such tools in the classroom raised concerns around concepts of learning, the future of thought, and the capacity for curiosity. "In education, what we're saying to our kids is that if you don't know something, just ask AI," noted one participant. "I think this is really a degeneration literally of the brain and not questioning ourselves since we've arrived with something else (aka AI) that will do the job." Amid the concerns, scholars pushed back on whether the longer-term effects would manifest. Instead, the conversation shifted towards appropriate pedagogical entry points for the use of AI in the classroom. Instead of asking the AI for the answer, an AI should be used to test your understanding of an answer.

The second intersection is in the public education of AI. Throughout the conference, participants and experts discussed the need for increased media literacy, educational campaigns, and reskilling efforts to better prepare U.S. citizens. One major challenge for any nationally orchestrated effort to educate the public is the structure of America's public education system. This falls outside the purview of the federal government and is largely determined at the state and local levels. Therefore, creating one central curriculum around AI, information, and or media literacy becomes a distributed problem. Unfortunately, as one scholar noted, the lack of public education of AI is one contributing factor to growing pessimism and distrust in it. In the next section, we explore the downstream effects as it impacts the American workforce.

AI and Labor

[25] Healthcare IT News. "Microsoft and 16 health systems debut network for responsible AI" (March, 14, 2024). https://www.healthcareitnews.com/news/microsoft-and-16-health-systems-launch-network-responsible-ai

[26] Walton Family Foundation. "Teachers and Students Embrace ChatGPT for Education." (March 1, 2023). https://waltonfamilyfoundation.org/learning/teachers-and-students-embrace-chatgpt-for-education

In the final session of the conference, participants and experts focused on AI's implications on the U.S. workforce. Specifically, experts in this session emphasized the role of more traditional businesses – not the tech industry – in shaping AI use cases. As one scholar stated, "Currently, the voice of AI is the voice of the tech industry. And, it should be the voice of the manufacturing industry, the financial services industry, etc." According to the National Association of Manufacturers in 2021, manufacturers account for 10.7% of the total output in the country, employing 8.14% of the workforce.[27] Financial organizations project this sector's output to grow by 3% in 2025[28] with employment numbers surpassing pre-pandemic levels (approximately 13 million).[29] Driving this surge is the commitment by various companies to upskill and/or reskill the workforce. Alternatively, we are seeing massive disruption to white-collar work thanks to the ability of generative AI to handle entry level tasks in professions such as law, publishing, and accounting.[30]

As one of the largest food and beverage companies in the world, PepsiCo employees more than 300,000 people globally and 100,000 in the U.S. Many of these employees include front-line roles responsible for making, moving, and selling its products.[31] Over the past four-years, PepsiCo has made strategic investments in developing a suite of upskilling initiatives that provide end-to-end opportunities for all its employees at no cost to them. One example is the development of the Digital Academy, which includes more than 11,000 learning assets designed to help employees acquire digital skills. The academy offers both on-demand courses and ongoing learnings such as certifications and credentials. Since its launch in 2022, more than 11,000 employees have participated, earning 600 certifications in areas ranging from DevOps to Power BI for data analytics.[32] Beyond digital skills, PepsiCo also offers a "myeducation" benefit that offers a catalog of programs including commercial driver's licenses. The company pays 100% of the cost for tuition, books, and fees up front to help eliminate any financial barriers to participation.[33] Education and skills training are parts of the puzzle. For PepsiCo, the use of AI runs across the entire value-chain inclusive of asset maintenance and supply chain management. Optimizing for operational efficiency includes

---

[27] National Association of Manufacturers. United States Manufacturing Facts. https://nam.org/state-manufacturing-data/2022-united-states-manufacturing-facts/#:~:text=Manufacturers%20in%20the%20United%20States,was%20%242.5%20trillion%20in%202021.

[28] ING. US Manufacturing Outlook. (March 11, 2024). https://think.ing.com/articles/us-manufacturing-outlook-better-times-are-coming/#:~:text=Given%20an%20environment%20of%20a,to%202.5%25%20growth%20in%202026.

[29] Deloitte. "Taking charge: Manufacturers support growth with active workforce strategies." https://www2.deloitte.com/us/en/insights/industry/manufacturing/supporting-us-manufacturing-growth-amid-workforce-challenges.html#:~:text=Manufacturing%20employment%20has%20surpassed%20pre,million%20as%20of%20January%202024.&text=The%20number%20of%20manufacturing%20establishments,the%20end%20of%20the%20period.

[30] McKinsey. "Gen AI and the Future of Work." https://www.mckinsey.com/quarterly/the-five-fifty/five-fifty-gen-ai-and-the-future-of-work

[31] Upskill America and Aspen Institute. "Case Study: UpSkilling for Career Mobility at PepsiCo" (August 23, 2023). https://www.aspeninstitute.org/publications/case-study-upskilling-for-career-mobility-at-pepsico/

[32] Ibid.

[33] Ibid.

considering workforce needs and safety as well as overseeing product quality and demand.

"While AI is not a panacea as it is not the solution to absolutely everything in the world; it is going to help us solve so many problems that have been intractable before," shared one scholar. For example, in areas looking at urbanization, AI tools can help multiple stakeholders—from city planners to structural engineers to construction firms—better tackle questions around affordable housing options, sustainability and costs. "This is the archetype of a really difficult design problem. And, in architecture, 80% of the decisions that affect the sustainability of a building are made at the state of conceptual design," shared an expert. "So, the more information and the more intelligence you can bring to this initial phase is critical." For this example, the output features an AI algorithm and design package that helped stakeholders explore every possible way that the buildings could be laid out while balancing for greenery, noise, sunlight, energy usage, and material costs. In another use case, the expert shared how companies leverage AI to capture and produce hundreds of pages of documentation for thousands of components that make-up a turbo jet engine. The use of AI for what is seemingly a simple task requires a sophisticated understanding of not just the component parts but a semantic understanding of what can and cannot be put together. Automating this task allows for much cheaper production in the long run.

To accelerate more efficiencies, one scholar posited the need for large scale, industry specific foundational AI models. Akin to the large language models used by OpenAI or Anthropic, the development of a shared industry data commons for use by manufacturing, architecture, and construction could open-up huge gains in areas that affect our built environments. Instead of each company creating stand-alone proprietary models and conducting data analysis, a shared repository enables increased collaboration and cross-pollination of information. This is where federal level policy can be most constructive. Legislative power can help ensure the safe, interoperable pooling of data by these companies. As one scholar highlighted, we see others, including Europe and China already beginning to collect data from manufacturing. "If we want to design things better, faster, and more highly optimized, then we need massive amounts of data to achieve that level of automation or efficiency," shared one expert.

While the examples of businesses embracing AI were plentiful, participants voiced additional concern around constituents' fears of automation replacing human labor. Participants pushed on experts to provide guidance on how companies can successfully manage the risk of job loss and worker displacement as related to the adoption of AI. Additionally, participants asked if and how officials can effectively translate these protections to their constituents. One scholar responded noting that "I don't believe there's going to be less jobs in the future, but there will definitely be different jobs and that's going to create enormous discomfort." Therefore, it is critical for companies to know what kinds of reskilling or upskilling or AI education programs are available at every single level of the company. "Because if we have a future where workers understand the potential of things, and they feel supported by the company at that stage, then AI is going to lead to huge opportunities." Other efforts by companies to help translate the potential impact of AI to their local communities and consumers includes

inviting them in to witness the technology and to experience firsthand how this technology works. “The only way to bypass this problem is by education,” emphasized one scholar.

Relatedly, participants discussed the gap between the need for, and availability of, trained, skilled workers. For some, this is symptomatic of a larger education systems problem where skills training is not fit for purpose for what companies really need. One potential outcome is a shift towards a more blended system that features both skills-based training with the traditional academic environments. For others, this gap represents a growing cultural and societal problem in a lack of motivation to work. This raises significant challenges for businesses, like PepsiCo, that are committed to supporting the U.S. workforce but must compete with more competitive, lower-cost labor markets around the world.

Conclusion

In the final analysis, AI is just a tool, albeit one with exponential possibilities. And we are only at the beginning of this journey. Throughout the conference, both participants and scholars toggled between its potential for good and for harm. Experts doubled down on the need to recognize that AI is not intrinsically high-risk, but is highly dependent on who uses it and how. Regulatory strategies, including an AI governance framework, may provide the agility to manage the scale, speed, and the unknown of this technology. Additionally, for every frontier example of AI, scholars offered several use cases that showcase the immense power of simple, less compute-intensive systems that are transforming lives at massive scale. Participants heard and saw a variety of AI applications from personalized education companions, to ambient listening in healthcare, to managing responses during a climate disaster. For policymakers, the challenge is to take productive steps to understand each specific use case, its context, and any potential risks. As several experts cautioned, it is essential to demystify AI and not inflate its capabilities. There are no inevitable outcomes with AI. It is, however, imperative that steps are taken sooner than later to begin orienting its use towards the future outcomes we want.

POLICY ACTION MEMORANDUM FOR MEMBERS OF CONGRESS[34]

Policy discussions in Bellagio began from the premise of "exponential growth" in AI capabilities, from well-known applications in entertainment and industry to novel uses in healthcare and education. Advancing AI, especially when combined with other established and emerging technologies, appears to promise enhancements in human health, safety, productivity, and quality of life. At the same time, participants recognized the risks of AI-enhanced disinformation and manipulation in democratic elections, as well as the race for AI leadership between democratic and authoritarian states, and the race to regulate among democracies. Although experts downplayed any looming existential risk to humanity from AI, geopolitical competition for AI-related resources and the potential for job displacement across many sectors stood out as significant long term risk factors. The following recommendations emerged from deliberation among members and scholars throughout the conference:

Economic Opportunities and Challenges

- Recognizing that open jobs outnumber job seekers in the current market, leverage federal funds, such as under the Inflation Reduction Act (IRA), to deploy AI tools for connecting employers with job seekers more efficiently.
- Recognize that AI is going to be a part of everything and be prepared to regulate it within other policy verticals. At the same time, consider the limitations of these tools and to be open to what may seem like "wacky" ideas today, such as Universal Basic Income, as a way of preparing for a future in which the workforce and the economy are totally transformed.
- As employers adopt AI to make workers more productive, it is important not to undervalue the human factor: human beings will care for the sick or elderly, and human beings will fix your household heat or plumbing. Policies should include skills-based training to ensure that human workers continue to serve these vital roles, including when assisted by AI.

Support for U.S. Innovation

- Be cautious about regulatory over-reaction to fears around AI, which could hamper U.S. AI innovators, learning from Europe's recent experience in this area.

---

[34] *Note: This policy action memorandum is compiled for Congressional participants and depicts policy ideas that emerged during the conference sessions in Bellagio, Italy. The Aspen Institute is a neutral convener. We merely cataloged the ideas that came forth.*

- Consider supporting the Create AI Act as a way of supporting those who do not have vast financial resources to invest in developing new models.
    - Government should be prepared to act boldly in support of U.S. AI innovation, on the scale of its investments in NASA or the Manhattan Project.

Regulation and AI in the Federal Government

- Congress should think about regulation in terms of first principles: safety, security and freedom, and working with friends and allies.
- There is a need for significant increases in federal funding for AI, especially for agencies like National Institute of Standards and Technology (NIST), which is currently under-resourced relative to its AI responsibilities.
- Consider supporting the Federal AI Risk Management Act, which addresses what federal agencies are doing to implement the recent Executive Order on AI.

Election Security and Democracy

- Consider passing a law to identify and mandate a standard for provenance verification of AI generated content, including "watermarking" where that is technically feasible.
- Fund local and state election officials to enhance communication and secure election systems, and to hold companies accountable when they cause problems.
- Use the bully pulpit of Congress to address the general public with messages like: "if something dramatic happens on your social media account 2 days before an election, take a pause and think hard about whether it is true."
- Recognize that challenges will get tougher as technology evolves, we will need AI itself to secure elections, and a federal standard to ensure compliance across states.
- Consider supporting the AI Foundation Model Transparency Act which forces tech companies to increase the transparency of widely used foundation models, and the Protecting Americans from Deceptive AI Act.

Education and AI Literacy

- Congress itself needs to be more educated on AI, as this is the only way to generate public trust on these issues.
- Support a national effort to integrate AI education into school curricula across the country.
- Consider supporting legislation incentivizing businesses to work with community colleges on job training programs that actually promise graduates competitive jobs, including in the service and manufacturing sectors.

Data Rights and Privacy

- Enact a U.S. data privacy standard.

- Given AI's dependence on aggregating large amounts of data, including personal medical and genetic information, consider ways to democratize the profits that ultimately come from that data.
- Use AI tools to help protect citizens from signing long, complex releases of their personal data that they do not understand.
- Consider using blockchain technology to stamp data with a personal identifier so that authenticity can be verified, privacy protected, and compensation paid.
- Support ways of compensating creators for material used in training AI models.

Strategic Competition, Energy and Infrastructure

- A bipartisan approach is essential to prevailing in strategic competition with China.
- Pausing AI development or imposing an overly heavy regulatory burden could slow U.S. industry, while Chinese government-backed industry will not pause. Thus, a better strategy is to double down on being the world's AI leader, with appropriate ethical guardrails.
- The higher the skills and capability in the U.S. economy, the more likely investment will come and stay here. Thus, stopping innovation to protect existing jobs will be counterproductive in the long term.
- Support datacenter and clean energy infrastructure in the United States, since both the development and use of AI across the economy will quickly outpace the availability of energy resources in many locations.

INTRODUCTORY READINGS

Artificial Intelligence (A.I) 101[35]

Aspen Digital, Aspen Institute

WHAT IS ARTIFICIAL INTELLIGENCE?

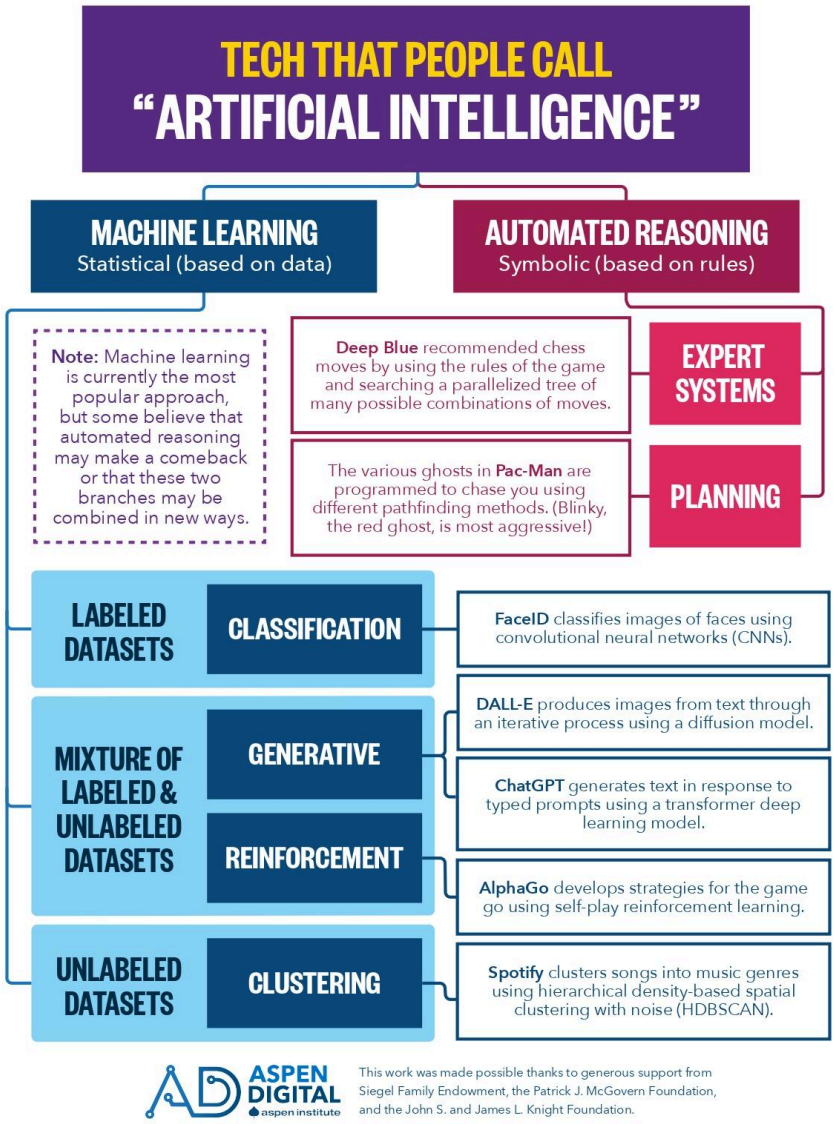
Artificial intelligence (AI) has historically referred to a collection of technologies designed to emulate human intelligence. In recent years, the term has become synonymous with machine learning, a set of computer processes used to identify unintuitive patterns in data. Examples of AI today include speech recognition, autonomous vehicle navigation, and the generation of new content, such as text or images.

Although the words "artificial intelligence" may conjure scenes from science fiction, most tools labeled with the term "AI" are not the powerful thinking machines of Hollywood movies. "Artificial General Intelligence," or AGI, is the widely-used term to refer to those not-yet-realized advanced technologies that could independently learn new capabilities. (See What is artificial general intelligence?) Historically, AI models were developed to accomplish specific individual tasks, but there are efforts to pursue AGI through combining some of these capabilities into "foundation models." Today's AI tools are more basic and already deeply embedded in a variety of sectors, including business, government, and even things as commonplace as the autofocus in your camera.

The definition of what counts as AI continues to evolve and remains a subject of debate. In fact, some take issue with the term "AI" altogether. (See Why don't people like the term "artificial intelligence?") Within this primer, "AI" will be used to refer to a collection of machine learning technologies that are designed to automate specific tasks.

Throughout the years, there have been several technologies that people have called "Artificial Intelligence"

[35] This information was provided by Aspen Digital. It is also available here. This work was produced by Eleanor Tursman, B Cavello, and Tom Latkowski, and was made possible thanks to generous support from Siegel Family Endowment, the Patrick J. McGovern Foundation, and the John S. and James L. Knight Foundation.



WHAT IS MACHINE LEARNING?

Machine learning systems are a type of AI that are essentially pattern recognition tools. They are "trained" to identify patterns within large collections of data (such as text, images, and video) in order to produce a set of instructions, or a "model," which applies that "training" to new data. For instance, a machine learning model could be "trained" on news articles and then used to predict the next word in a sentence you are typing.

WHAT IS "THE ALGORITHM?"

Many people use the term "algorithm" colloquially to refer to a wide array of technologies (such as the Instagram algorithm, an encryption algorithm, or a facial recognition algorithm). Generally, the term "algorithm" is defined as a set of instructions for a computer to execute. However, when people talk about algorithms in relation to AI, they typically mean one of two related, but different, things:

1. THE PROCESS OF CREATING A MODEL
   When someone says a facial recognition system was created "by feeding images from Facebook into a machine learning algorithm," that means that a machine learning model is being "trained" to represent the patterns in that image data to recognize faces.

2. THE APPLICATION OF A MODEL TO PRODUCE AN OUTPUT
   When someone says "the YouTube algorithm prefers short-form content," they're referring to how the model YouTube uses to recommend videos to watch next shows shorter videos to more people.

WHAT IS DATA, AND WHY IS IT SO IMPORTANT?

Data, like AI, is an umbrella term that covers more than just numbers. The term describes many types of information that are stored and processed on computers. Videos, electronic health records, and the location information on your phone are all different kinds of data.

More reliable AI systems typically require a large amount of data to "train." This is because the patterns in a small collection of data may not be generalizable. For instance, a system built to label dogs in a collection of images may not operate reliably on images of dogs in a park if all of the examples used to "train" the system were of dogs in homes. Using more data from more diverse sources helps to ensure that the patterns represented in the machine learning model apply to a wide variety of contexts.

There is a tendency to misinterpret the accuracy of machine learning models as a measure of how well they represent reality. In fact, what academics and engineers often call the "accuracy" of AI systems is only a measure of how well they represent the data used to "train" them. If the data used to create a model is flawed (whether because it is incomplete or biased), the outputs of the model will reproduce or even amplify those flaws.

NON-EXHAUSTIVE LIST OF TYPES OF DATA

- Text (books, articles, blog posts, discussion threads, social media, health records)
- Images (photos, paintings and illustrations, x-rays, maps)
- Audio (voice, music, birdsong, engine noise)
- Video (CCTV, drone footage, film and television recordings)
- Biometrics (face, fingerprint, heart rhythm, gestures)
- Geolocation information

Having transparency into what data is used to "train" a model can give us insight into how the model responds to different examples. Generative AI systems like ChatGPT are trained on large swaths of the internet—knowing which parts of the internet are included in the training data makes ChatGPT's output more explainable.

WHY DON'T PEOPLE LIKE THE TERM "ARTIFICIAL INTELLIGENCE?"

Many researchers and activists have argued that describing these systems as "intelligent" attributes too much agency to the technology itself and erases the humans involved in the process. When people talk about AI, they'll often say things like "an AI fired 3000 workers" or "DALL-E created award-winning art."

In reality, these tools—regardless of how much or little oversight they receive—do not exist in a vacuum. Humans choose what types of systems to develop and curate the data to "train" machine learning models. Humans define the criteria for good system performance, and humans deploy the resulting technology—even if they abdicate responsibility for its impacts. Critics argue that calling these tools "artificial intelligence" obfuscates the human roles in these processes and makes it difficult for people impacted by the deployment of AI to seek remedy or recourse. Although there are compelling arguments for abandoning terms like "artificial intelligence" and "machine learning," they are nonetheless already in wide use. Rather than avoiding these terms, it may be more pragmatic to help the public contextualize them by both explaining what specific technologies constitute the AI systems being discussed and to highlight the people involved in building and using these tools. (See How to Talk About AI for examples.)

WHAT IS ARTIFICIAL GENERAL INTELLIGENCE?

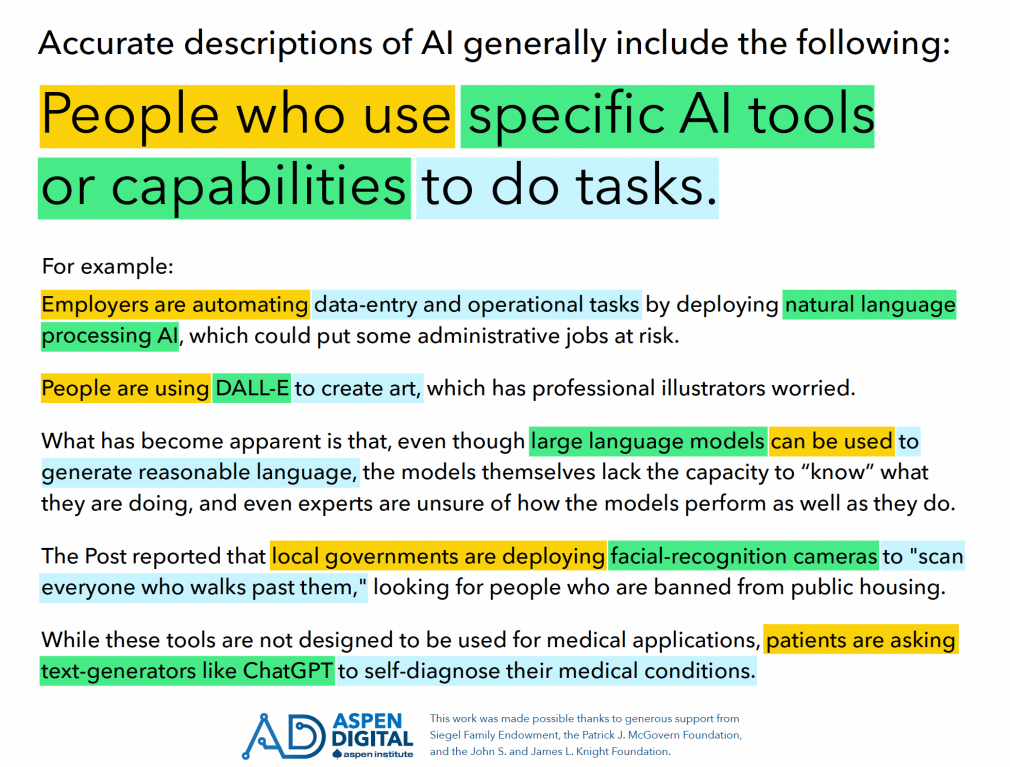
Artificial general intelligence (AGI), sometimes referred to as "strong AI," is a conceptual computational tool that exhibits human-level or beyond human-level intelligence in all domains. Some people believe it is important to pursue AGI because a

machine that has generalizable, human-level intelligence could be a useful tool for space exploration, national security, or neuroscience. No such capabilities currently exist, though many companies are actively pursuing this possibility.

Throughout history, different tests have been proposed for identifying intelligence in AI systems, including playing chess, making a cup of coffee, or the "Turing test," which, to pass, a human must fail to identify the AI in conversation. All of these have since been ruled insufficient indicators of human-like intelligence. Until the more philosophical question of how to define intelligence is addressed, there may be no agreed upon way to evaluate attempts to make these kinds of systems.

HOW TO TALK ABOUT AI

This section showcases examples of how to write about AI inspired by news stories from the last year. Reporting on AI should avoid personifying the technology or obfuscating the people and organizations using the tools. Instead, aim for sentences that clearly highlight both the types of technology being used and the people involved in their design and deployment. Researchers are often useful sources for getting more information on the specific capabilities and limitations of AI systems. (See Common Roles in AI for more information).



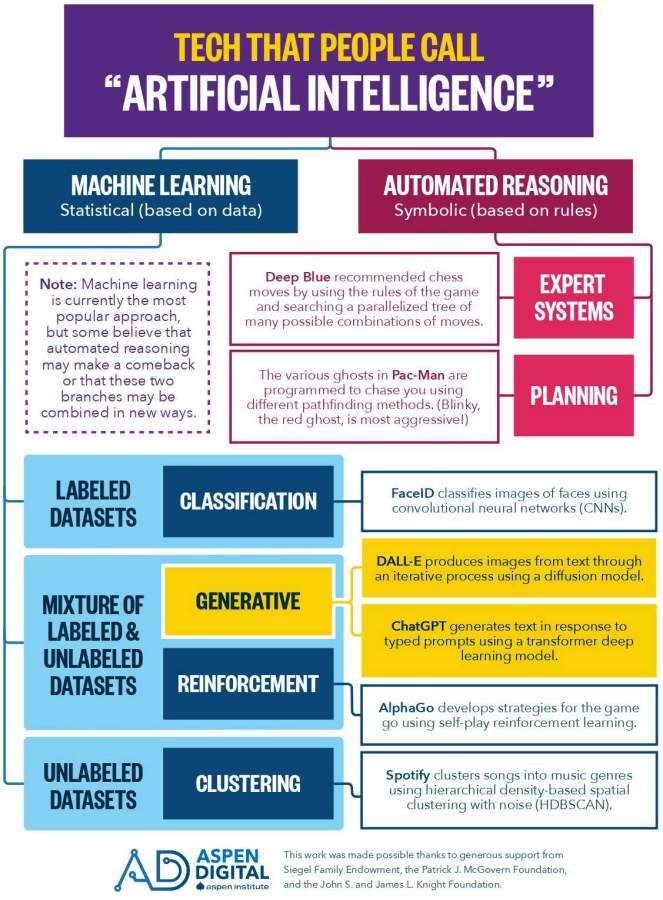
Aspen Institute Congressional Program

Introduction to Generative A.I.[36]

Aspen Digital, Aspen Institute

WHAT IS GENERATIVE AI?

Generative AI is a subset of artificial intelligence technologies that are used to create new content, such as images or text, based on patterns in large amounts of existing content. Generative AI differs from classification AI—like email spam filtering or tumor detection used in medical settings—because generative AI systems are designed to make content, not to make decisions.



[36] This information was provided by Aspen Digital. It is also available here. This work was produced by Eleanor Tursman, B Cavello, and Tom Latkowski, and was made possible thanks to generous support from Siegel Family Endowment, the Patrick J. McGovern Foundation, and the John S. and James L. Knight Foundation. AI 101 © 2023 by Aspen Digital is licensed under CC BY-NC 4.0.

WHAT GENERATIVE AI CAPABILITIES EXIST TODAY?

While ChatGPT captured the public's attention by allowing people to generate uncanny and seemingly confident responses to a vast array of written prompts, there is a diverse set of generative AI applications that have been made available to businesses and consumers. Increasingly, AI developers are creating multimodal tools, tools which incorporate multiple sources of input data (such as images, audio, or text) at once.

- Image-to-image (Canvas)
- Text-to-image (Midjourney, DALL-E, Stable Diffusion)
- Text-to-audio (MusicLM)
- Video-to-video (Project Morpheus, AI video compression)
- Text-to-video (Make-A-Video)
- Image-to-text (Automatic image description)
- Text-to-text (including computer code) (Github Copilot, Bard)

HOW MIGHT GENERATIVE AI BE USED?

Although generative AI tools are still in the early stages of development, they are already being used to produce content at surprisingly high speeds, low costs, and with relative ease for end users. This newfound accessibility has labor and operational implications for software engineering, media production, education, the commercial art market, and more. No one knows exactly what will emerge from the explosion of generative AI tools hitting the market, but early experiments point toward the potential for larger scale disruption to business, security, and society at large:

Hyper-personalized Content

Traditionally, the cost of making individually personalized content, such as ads with your face in them or movie trailers narrated in the sound of a loved one's voice, was prohibitively high. People may now easily use generative AI to realize this level of extreme personalization, either for their own fulfillment or to manipulate others.

The Rise of "No-Code" Application Development

Historically, in order to develop websites or computer applications, you needed to know programming languages. Now, it is becoming possible to use conversational language to prompt an AI tool to produce computer code for you (even if today's systems are still imperfect). These tools may lower costs by expanding the number of people who are able to create and contribute to software development and make a wide range of

products and services more accessible. However, they could also negatively impact how much people are paid for these skills and change the nature of work to make it more tedious and less collaborative.

Better Augmented Reality

Real-time rendering of believable digital environments is computationally intensive and expensive. These graphical requirements have been a pernicious issue for augmented or virtual reality applications because time delays in rendering can create a jarring and unnatural user experience. Generative AI systems could be used to approximate (if not perfectly replicate) complex physical phenomena, like lighting and shadows, making these virtual scenes feel more immersive.

There are still many unknowns and opportunities for discovery. We have only scratched the surface on possible uses of these tools.

KEY ISSUES IN GENERATIVE AI

There are a number of promising applications of generative AI systems, a subset of artificial intelligence technologies that are used to create new content based on patterns in large amounts of existing content. These uses are not without their risks, however. The following sections highlight a number of the most pressing issues associated with generative AI, with links to a number of illustrative articles exploring perspectives on each of these issues.

INFORMATION ECOSYSTEMS

How will generated content affect the trustworthiness of media?

Media created to mislead is not a new problem, but generative AI makes it much easier to create mis- and disinformation at scale and to create convincing human-like AI interactions that could be used to exploit users with scams or security attacks. As generated content improves, it will be harder to detect inauthentic content in the wild, and it will be easier for smaller actors to manufacture large-scale disinformation campaigns.

Deepfakes	Humans may find deepfake faces more trustworthy than real ones
Confident nonsense	"While the answers which ChatGPT produces have a high rate of being incorrect, they typically look like they might be good," burdening human moderators
Democratizing malware	ChatGPT makes it possible for anyone to produce malware and phishing emails
Impact on elections	Generative AI is already being used to influence voters and reduce trust in the electoral system

EXPECTATIONS & CLAIMS

What are the limitations of generative AI systems?

People selling AI products and services benefit from systems being perceived to be more reliable and capable than they are, from the anthropomorphization of "smart assistants" to the typing animations of text-generators like ChatGPT. Peeking behind the curtain reveals that the AI tools on the market are specialized in scope, not general- purpose "intelligences," and still have crucial vulnerabilities and flaws. Although it might be tempting to ascribe greater power to these systems, there are still many questions about whether they are appropriately effective for the widespread adoption we are already seeing, let alone that we are on the verge of "Artificial General Intelligence" that surpasses humans in a broad range of capabilities. Systems like ChatGPT and Bard were designed to produce confident sounding text, not factual statements.

AI hype	An AI model "passing" an exam designed for humans is not indicative of intelligence
Confident but inaccurate	CNET used ChatGPT to generate articles that ended up riddled with errors
False equivalency	Equating GPT's text prediction capabilities with consciousness is misleading
Unintuitive failures	Certain strange prompts (like usernames) elicit nonsense from ChatGPT

INTELLECTUAL PROPERTY

Who owns what?

Many datasets used to build today's generative AI have been compiled by scraping, or extracting information, from the web. For example, to make a generative AI tool that can output digital images, developers scraped millions of existing images hosted on large art platforms, like Flickr and DeviantArt. Content collected online in this manner is often used without the consent or knowledge of the original creator. Even if the original content is not reproduced by the system (although in some cases it can be), this process leads to thorny questions around attribution, intellectual property, monetization of generative AI tools, and economic harms to creative industries.

Sensitive data	Scraping isn't flawless—personal health data was found in a popular image dataset
IP protections at work	Google won't release its text-to-music generator because there is a chance it will reproduce copyrighted music it was trained on
Fights for	A class action lawsuit is filed against Microsoft for lacking attribution for code used to
Copyright ramifications	AI-generated works should not be permitted copyright protection

FUTURE OF WORK

How will generative algorithms impact peoples' livelihoods?

Generative tools can be used as assistants, augmenting human creativity, but they can also be used to automate certain types of work, from writing copy to creating spot art for articles to coding. There are many open questions about what tasks will be most easily automated and whether that automation will result in a reduction in total jobs, a profound change in how certain work is valued, or a restructuring of labor as new jobs are created. For example, a

software developer that once created website templates could either (1) lose their job because someone else can use a tool to do themselves what they would have hired the developer to do, (2) get a reduction in salary as they face more competition in the

market or, (3) no longer code as much manually, but instead be in charge of generating outputs using the AI.

Expanded creativity	Generative AI could make it much easier for people to create websites and apps without needing to know how to code
Impact on creative industries	The use of generative AI in Hollywood was a key sticking point in labor negotiations
Augmenting work	How professionals can use ChatGPT today
Invisible labor	Automating some tasks just creates a different kind of work

ENVIRONMENTAL IMPACTS

How does generative AI contribute to climate change and consume natural resources?

Generative AI systems are more resource-intensive than many similar technologies. The large data centers (collections of connected computers) used to develop and deploy these tools require power and cooling, using electricity and water. While the development of AI models typically requires a large amount of energy, generative AI is unique in that people's ongoing use of these systems makes up the bulk of its consumption. Additionally, although today the water usage of generative AI is dwarfed by the amount used for other purposes, like growing almonds or making potato chips, water and energy usage would rise further if proposals such as incorporating generative AI into every Google search are implemented.

Carbon footprint	Generative AI use eats up resources needed to mitigate the climate crisis
Energy usage versus other technologies	ChatGPT uses five to nine times as much electricity as a Google search
Freshwater usage	Generative AI systems contribute to stress on local water infrastructure in drought-prone areas
Reducing climate impacts	Some argue that companies can make AI greener by fine-tuning existing models instead of training new ones

DISCRIMINATORY EFFECTS

How do generative systems perpetuate societal harms?

Unless the data ingested into AI models is carefully curated—which datasets scraped from the web rarely are— tools built using that dataset will reflect the biases of the unfiltered internet. Even with careful dataset curation, however, AI tools need to be fine-tuned by human content moderators to mitigate systemic biases. In some cases, creators or deployers of a system will manually override the AI system to limit output of harmful material, but these sorts of interventions are necessarily brittle and imperfect.

Biased assumptions	The "magic avatars" created using Lensa AI sexualize women and whiten people of color regardless of their users' wishes
Mitigating harms	GPT-3 will make biased statements against certain groups, but it may be possible to mitigate this with extra training focused on fairness
Content moderation	In protecting the world from biased and discriminatory outputs in these systems, content moderators suffer the consequences

FEEDBACK LOOPS

How will generative AI impact future AI development?

Future datasets scraped from the web will be impacted as everyday people, content farms, and disinformation campaigns saturate the internet with generated content. New AI models that are trained using these datasets may perpetuate existing biases documented in large language models like GPT-3 or image generation models like Stable Diffusion. Detecting generated content to exclude it from datasets is an active field of research but is by no means a solved problem. This feedback loop of using content produced by machines to train machines to produce more content could reduce the quality and performance of future AI systems.

WHAT COMES NEXT?

While these are the early days, many experts agree that generative AI systems will have far-reaching implications for society. Unlike blockchain and other emerging technologies that have caught the tech industry's eye, generative AI tools have sparked the public's interest and imagination with creatives and business leaders alike identifying ready applications. Most immediately, here are some things that could come next:

- Tech companies vying to both define and control new markets carved out by generative AI tools, deploying work-in-progress tools into an unregulated space

- The establishment of legal frameworks and precedent to better define both intellectual property rights and consumer protection with regards to generative AI and the AI space more broadly

- The immediate disruption of some existing labor markets while new areas of work are still being defined

EXPERTS' ESSAYS

WEDNESDAY | April 3, 2024

Raffi Krikorian	*AI 101: Five Things about Artificial Intelligence Worth Pausing to Consider*
Alondra Nelson	*The Right Way to Regulate AI*
Klon Kitchen	*Business of Knowing: Private Market Data and Contemporary Intelligence*
David Rhew	*Unlocking the Potential of AI through Policy that Ensures Trust and Adoption*

THURSDAY | April 4, 2024

Chris Krebs	See *The Impact of Generative AI in a Global Election Year* under Experts' Recommendations
Vivian Schiller	*Generative AI Risk Factors on 2024 Elections*
Vilas Dhar	*Will 2024 Be the Year of Responsible AI?*
Anna Makanju	*Navigating the AI Era. Here's How the U.S. Can Maintain Its Edge – and Improve the Lives of All Americans*
Darío Gil	*Understanding and Governing Generative AI*

FRIDAY | April 5, 2024

Athina Kanioura	See *Case Study: Upskilling for Career Mobility at PepsiCo* under Experts' Recommendations; Also see *Frontline A.I.: A Guide for Manufacturers* under Experts' Recommendations
Mike Haley	*Generative AI – Move Over Language Models and Make Way for Industry*

AI 101: Five Things about Artificial Intelligence Worth Pausing to Consider

Raffi Krikorian
Chief Technology Officer, Emerson Collective

Artificial intelligence might seem new, but we have been living in an AI world for some time. It is powering your Netflix recommendations and your Uber driver's route. Social media uses it to know which posts you are most likely to click on, and Alexa and Siri rely on it to answer your endless questions. But it is not all about content and convenience. Medical systems are beginning to use it to augment doctors' knowledge, finding otherwise invisible signs of disease. Khan Academy is using it to educate millions of kids around the globe for free. And more.

For the last nine months, I have been exploring the world of AI through my podcast and newsletter, Technically Optimistic. As the Chief Technology Officer of Emerson Collective, I am fascinated by the issues, innovations, ethics, and legislation surrounding this transformational technology. I want to engage people in thinking about how it can be developed and deployed with society's best interests at heart. I can tell you that when I sit down to write my newsletter about AI every Monday, so much has happened that it is hard to settle on a topic. From election deepfakes to federal regulation, education to privacy, it is endless — and fascinating. Just like when it comes to understanding AI itself, it can be hard to know where to start.

But let's look at how we got here. AI's journey began in the mid-1950s, when a Dartmouth professor on summer vacation led a small workshop focused on making machines smarter and faster, able to accomplish anything a human brain could do, such as master reasoning, language, and new tasks. In some ways, the deep-learning techniques that researchers began exploring then are not that different from what is driving today's boom.

It all shifted in 2012 with the idea of a neural network, a math-based system that finds patterns in vast amounts of data that no human could unearth. By 2015, an AI program had beat the world champion at the ancient game of Go — a game that is harder than chess by multiple factors. Within a few years, the big tech players — Google, Microsoft, and OpenAI — had developed neural networks that were trained on incredible amounts of text 'scraped' from the internet. These large language models, or LLMs, had consumed so much text that they were able to have nuanced conversations, write code, and even produce not-terrible novels based on simple prompts.

These LLMs are what power ChatGPT, Bing, Bard and others, and are what have everyone from journalists to translators to songwriters feeling like their days are numbered. They are also what political organizers are using to analyze voter insights, generate campaign ads and compose fundraising emails.

ChatGPT falls under the heading of generative AI, meaning that it can create things — in this case, text. Generative AI is also used to create images, which the programs DALL-E and Midjourney do to occasionally impressive effect. And OpenAI just announced the release of Sora, which can assemble Hollywood-worthy video clips from a simple text description.

In the space of just over a year, anyone with access to a computer suddenly has the power to create, compute, and disrupt beyond their wildest dreams. We are on an exponential growth curve of capabilities — and we as humans do not know how to perceive exponential growth. So what we really need to do now is consider some big questions to help us set guidelines, establish morals, etc., so that we can deal with this unprecedented growth later.

The Power — and the Peril

As we begin to explore the power of these new tools, there are five risk factors that we all should consider:

1. These systems are data-hungry. As you have just read, LLMs are powered by data. ChatGPT devoured all of Wikipedia, libraries worth of novels and, as the New York Times lawsuit contends, millions of copyrighted stories for which it paid no licensing fee and sought no permission.

The tech companies say that we need to keep feeding these LLMs in order to help them grow smarter, more diverse and 'human' in their thinking. But where will that information continue to come from beyond huge scrapes of the Internet? Should OpenAI, Microsoft and Alphabet — companies that have billions in funding to develop their products — pay for the copyrighted information that they are taking? Should they ask for permission or even disclose where their data sets come from? And who would regulate that?

Some media companies have begun charging licensing fees: Last week, Reddit struck a $60 million deal with Google for its content. But $60 million is not a lot when you factor in the idea that the data is potentially being used to train a model that could put the company out of business. And when it comes to content that is not user-generated, like

Reddit's — media organizations that rely on journalists and fact-checkers, like Dow Jones & Co. or the New York Times — I would say that it is essential that such trusted sources continue to exist in a world in which AI-generated misinformation is being spread at an unprecedented rate.

Other groups do not have the option to charge for their data. There are examples of 'data extraction,' such as indigenous people in Hawaii giving over their genetic data to big companies so that they can develop drugs. Should they be compensated? Or should we be looking to the Initiative for Indigenous Futures, which works with these groups to bring them into the conversation? In India, people are sitting in the equivalent of call centers to enter their languages into computers so that these systems can learn them. Is that fair, or should we be looking at models like Karya, which actually does fair compensation?

These LLMs are also being fed our personal data, sold to them by third parties, without us being aware of it. This can include surveillance information about our location, our browsing history, how long we stay on a certain Kindle page, and more. It can include sensitive medical information that can be sold to insurance companies to deny you care, or used against women who travel out of state for abortions.

I write frequently about data and privacy in my Technically Optimistic newsletter, and as you all will know, we are not even close to regulating Americans' privacy through federal legislation. Fourteen states have leapt into the power vacuum to try to protect their residents, but given the speed of AI's development, the country will not be protected quickly enough. We missed the boat on regulating social media. We should nothave to wait for a disaster to strike before we act.

Colorado Senator Michael Bennet told me on the podcast that he has been proposing the Digital Platform Commission, a federal oversight organization that would create compliance standards and oversee them, along the lines of the Food and Drug Administration or the Federal Communications Commission. He has been doing this for years because he does not believe that the government can get it together in time to make it happen on its own. Bennet sees the ideal commission as composed of experts with a background in areas such as computer science, software development, and technology policy. This is a promising way forward, but more is needed: Each agency should be staffed with a diverse range of people who think about AI and emerging technologies so that their bespoke needs and understandings are considered.

2. These systems can amplify bias and discrimination. Think about it: These models are developed and trained on the data that is available to them. And that data tends to be biased toward white, male Westerners. Racism, sexism, and other cultural assumptions are baked in. You have heard about the results of AI-generated

decision-making perpetuating these biases: Black men arrested due to faulty facial recognition systems. Skin cancer in black patients going undetected by AI trained to spot abnormalities. Amazon's automated hiring process passes over female candidates for C-suite jobs, since the resumes used to train the software were from past decades, when women were underrepresented.

When I was running Uber's Advanced Technologies Center, charged with making self-driving cars, we gathered most of our road and training data in Tempe, Arizona, where the weather is nice and it is easy to drive around in a car with video cameras recording everything year-round. Then we took all that data back to our headquarters in Pittsburgh to start doing work. We developed the car. We put it on the road. And it literally had trouble recognizing black people. Why? Because there were not many black people in the data set we collected in Tempe.

As these systems interact more and more with parts of our lives, we need people with diverse backgrounds to take all the issues into account. Diversifying the field of developers and engineers is essential. According to a 2019 UNESCO estimate, women make up just 12% of AI researchers, and represent only 6% of software developers. In 2022, only 1.7% of doctorates in computer science, computer engineering or information in the United States went to Hispanic graduates, and 1.6% for black graduates. But education in foundational computer science needs to start in high school or earlier, and be available to students in all schools around the country. Therefore, we should fund educational programs and research grants that will guarantee that the whole world is reflected in these AI systems.

3. These systems should remain open source. One of the biggest debates right now is whether we should be open source or closed source. Open source means that the original source code is transparent and easily modified by anyone. It is a public collaboration rather than proprietary.

Historically, most people liked open source because it had a connotation of security: You could see what was going on under the hood. If you saw a problem, you could fix it. The same is true about AI models: open source here could mean open data, open testing frameworks, and seeing how these models were developed, what they were tested against. etc.

However, open source means bad actors can get the tools and do crazy things. In fact, we are now probably going to have an insatiable stream of child sexual abuse material (CSAM) on the internet because bad actors are taking these tools and training them to generate it. But bad actors are going to be bad actors; we have to consider these problems

in the full context. If we do not have open systems, then we are at the whim of the closed platform companies, and we will never be able to interrogate bias, discrimination, etc.

4. These systems require more transparency. These systems are so powerful that they are surprising the people who are building them. That is certainly a reason for concern.

In this instance, transparency means a few things: We need the ability to view how these systems are built, trained and tested, so that we can all agree on their soundness. We need the access and ability to be able to record what these systems are doing, so that we can learn from them later. We also need better tools to understand the systems that are in development, which requires more research funding. (My current concern is that academia, which could be researching ways to get at transparency, has been co-opted by the big companies to work on new features rather than accountability.)

When it comes to the public, transparency takes on another meaning. Now that systems like ChatGPT are being used to write articles, create videos, and replicate voices, we are in a critical moment: Today, the majority of our information comes from the Internet. But the reality is that more and more of what we see there is fake. We need to question the origin of every image, article, and video and audio clip — especially in this critical global election year. But right now, it is also critical that we have trust in our sources, our political candidates, and the democratic process itself.

As Oren Etzioni — the former CEO of and current advisor to the nonprofit A12, the U.S. research institute founded by the late Paul Allen — told *Popular Science*: "We are witnessing a pivotal moment where the adversaries of democracy possess the capability to unleash a technological nuclear explosion."

'Watermarking,' or posting digital disclaimers for AI-generated content, is a bare minimum, and a pretty low bar — one that can be easily overlooked by users. In Michigan, there is new legislation that requires any political ad that uses AI to manipulate its content to include a disclaimer across TV, radio, print, and social media. And last May, Sen. Amy Klobuchar and her colleagues presented the REAL Political Ads Act. This 'commonsense legislation' requires a disclaimer on political ads using AI-generated images or videos in an attempt to increase transparency and accountability in political advertising. It has yet to pass.

Simply put, I implore you to fast-track legislation mandating that the public knows what is real.

5. These systems require more education. As I like to say, my mother-in-law understands why we do not raise the speed limits on highways, but she does notunderstand the potential dangers of having a video doorbell, no matter how ubiquitous they are becoming. In 2018, Finland did an experiment to get 1% of its population up to speed on AI. The result? They got 10%. And now Finland, by some measures, has some of the highest per-capita AI startups, and AI is being used in all parts of their society. (Plumbers wrote in to say that their business changed because of the class). We need to do that here, and increase computer-science and AI literacy at all levels: K-12 as well as adults. MIT's Day of AI, Stanford's AI4All and code.org may all be paths worth replicating.

Ideally, this broader public education would quickly funnel into the academic level, because we need more researchers — especially ones who are not commercially captured, as mentioned above. That means that we need to increase funding through the National Science Foundation, to build the National Artificial Intelligence Research Resource (NAIRR) Task Force, and to support more fundamental infrastructure so we can get this all working.

As NSF Director Sethuraman Panchanathan said of NAIRR's potential, "By creating an equitable cyberinfrastructure for cutting-edge AI that builds on-ramps for participation for a wide range of researchers and communities, the NAIRR could build AI capacity across the nation and support responsible AI research and development, thereby driving innovation and ensuring long-term U.S. competitiveness in this critical technology area."

Long-term U.S. competitiveness is certainly important. But we must not forsake our privacy and safety — both personal and national — because we are in awe of or don't fully understand the power and transformative potential of this revolutionary new technology, which is developing faster than even its engineers imagined. We must regulate and educate immediately for the safety of our nation. Because tomorrow is already here.

The Right Way to Regulate AI[37]

Focus on Its Possibilities, Not Its Perils

Alondra Nelson
Harold F. Linder Professor of Social Science, Institute for Advanced Study

Artificial intelligence "is unlike anything Congress has dealt with before," U.S. Senate Majority Leader Charles Schumer said in June 2023. The pace at which AI developers are producing new systems—and those systems' potential to transform human life—means that the U.S. government should start "from scratch," he declared, when considering how to regulate and govern AI. Legislators, however, have defied his wishes. Following OpenAI's late 2022 unveiling of ChatGPT, proposals for how to encourage safe AI development have proliferated faster than new chatbots are being rushed to market. In March 2023, Democratic legislators proposed moratoriums on some uses of AI in surveillance. The next month, a group of bipartisan lawmakers floated a bill to prohibit autonomous AI systems from deploying nuclear weapons. In June, Schumer debuted his own AI agenda, and then in September, a bipartisan group of senators reintroduced a bill for AI governance promoting oversight, transparency, and data privacy.

The race to regulate is partly a response to the platitude that government may simply be too sluggish, too brittle, and too outmoded to keep up with fleet-footed new technologies. Industry leaders frequently complain that government is too slow to respond productively to developments in Silicon Valley, using this line of argument to justify objections to putting guardrails around new technologies. Responding to this critique, some government proposals encourage expeditious AI development. But other bills try to rein in AI and protect against dangerous use cases and incursions into citizens' privacy and freedoms: the Algorithmic Accountability Act that House Democrats proposed in September 2023, for instance, mandates risk assessments before technologies are deployed. Some proposals even seek to accelerate and put the brakes on AI development at the same time.

This commendable but chaotic policy entrepreneurship risks scattering government's focus and threatens to lead to a situation in which there is no clear governance of AI in the United States at all. It doesn't have to be this way. A tendency to slip behind the curve of technological innovation is not an inherent weakness of government. In fact, trying to outpace government regulation is the tech industry's deliberate strategy to circumvent oversight. Government has an irreplaceable role to play as a stabilizing force in AI development. Government does not have to be a drag on innovation: it can enable it,

[37] This essay was originally published by *Foreign Affairs* on January 12, 2024.

strategically stewarding science and technology investments to not only prevent harm but also enhance people's lives.

From its first days, U.S. President Joe Biden's administration has worked toward a more integrated technology policy agenda that addresses AI's widening uses, considering competition, privacy, and bias as well as how to safeguard democracy, expand economic opportunity, and mitigate an array of risks. But AI technology is changing rapidly, and much more must be done to quickly clarify the central goal of AI governance so that policymaking is not only reactive.

AI governance should reject choice architectures that cast the future as a rigid binary—between a vision of paradise or dystopia or between a false dilemma of pursuing efficiency or ensuring equity. Safety and innovation in AI are not mutually exclusive. Because new and emerging AI technologies are so dynamic and used for so many purposes, however, they may elude conventional policy approaches. The United States does not need so many new AI policies. It needs a new kind of policymaking.

False Analogy

To regulate AI, many policy advisers in the United States and beyond have first sought an analogy. Are AI systems more like a particle accelerator complex, a novel drug therapy, or nuclear power research? The hope is that identifying a parallel, even a loose one, can point to the existing governance strategy that should apply to AI, guiding current and future policy initiatives.

The economist Samuel Hammond, for instance, took inspiration from the massive twentieth-century U.S. effort to build and assess risks related to nuclear weapons. He has proposed a Manhattan Project for AI safety, a federal research project focused on the most cataclysmic risks potentially posed by artificial intelligence. The nonprofit AI Now Institute, meanwhile, has begun to examine the viability of a regulatory agency based on the U.S. Food and Drug Administration: an FDA-like regulator of AI would prioritize public safety by focusing on prerelease scrutiny and approval of AI systems as the U.S. government does with pharmaceuticals, medical devices, and the country's food supply.

Multilateral analogies have also been suggested. The German Research Center for Artificial Intelligence has advocated modeling AI governance on the European Organization for Nuclear Research (CERN), the intergovernmental body that oversees fundamental scientific research in particle physics. In May 2023, Sam Altman, Greg Brockman, and Ilya Sutskever—then co-leaders at OpenAI—recommended that an AI governance framework be modeled on the International Atomic Energy Agency; in this

model, the United Nations would establish an international bureaucracy to develop safety standards and an inspection regime for the most advanced AI systems.

The absence of an internationally coordinated research infrastructure poses a significant challenge for AI governance. Yet even conventional multilateral paradigms predicated on nation-state membership are unlikely to produce an effective way to govern competitive, for-profit industry efforts. AI companies are already offering products to a global and diverse customer base, including public and private enterprises and everyday consumers. And none of these analogies, including the U.S. domestic ones, reflect the fact that the data that enable AI systems' development have already become a global economic and political force. Further, all these potential models end up neglecting some critical domains on which AI will likely have a transformative impact, including health care, education, agriculture, labor, and finance.

The problem with reaching for a twentieth-century analogy is that AI simply does not resemble a twentieth-century innovation. Unlike the telephone, computing hardware, microelectronics, or many pharmaceutical products—technologies and products that evolved over years or decades—many AI systems are dynamic and constantly change; unlike the outputs of particle physics research, they can be rapidly deployed for both legitimate consumer use and illicit applications nearly as soon as they are developed. Off-the-shelf, existing governance models will likely be inadequate to the challenge of governing AI. And reflexive gestures toward the past may foreclose opportunities to devise inventive policy approaches that do not merely react to present challenges but anticipate future ones.

Drop an Anchor

Instead of reaching to twentieth-century regulatory frameworks for guidance, policymakers must start with a different first step: asking themselves why they wish to govern AI at all. Drawing back from the task of governing AI is not an option. The past decade's belated, disjointed, and ultimately woefully insufficient efforts to govern social media's use of algorithmic systems are a sobering example of the consequences of passively hoping that social benefits will trickle down as an emergent property of technological development. Political leaders cannot again buy the myth—peddled by self-interested tech leaders and investors—that supporting innovation requires suspending government's regulatory duties.

Some of the most significant challenges the world faces in the twenty-first century have arisen from the failure to properly regulate automated systems. These systems collect our data and surveil our lives. The indiscriminate use of so-called predictive algorithms and decision-making tools in health care, criminal justice, and access to housing causes

unfair treatment and exacerbates existing inequities. Deepfakes on social media platforms stoke social disorder by amplifying misinformation. Technologies that went undergoverned are now hastening democratic decline, intensifying insecurity, and eroding people's trust in institutions worldwide.

But when tackling AI governance, it is crucial for leaders to consider not only what specific threats they fear from AI but what type of society they want to build. The public debate over AI has already shown how frenzied speculation about catastrophic risks can overpower people's ability to imagine AI's potential benefits.

Biden's overall approach to policymaking, however, illustrates how viewing policy as an opportunity to enrich society—not just as a way to react to immediate problems—brings needed focus to government interventions. Key to this approach has been an overarching perspective that sees science, research, and innovation as offering both a value proposition and a values proposition to the American public. The administration's signal early policy achievements leveraged targeted public funding, infrastructure investment, and technological innovation to strengthen economic opportunities and ensure American well-being.

The 2022 Inflation Reduction Act, for instance, was not designed to merely curb inflation: by encouraging the production and use of advanced batteries, solar power, electric vehicles, heat pumps, and other new building technologies, it also sought to help address the climate crisis and advance environmental justice. The 2022 CHIPS and Science Act promoted the revival of U.S. innovation by backing the development of a new ecosystem of semiconductor researchers and manufacturers, incorporating new opportunities for neglected U.S. regions and communities.

Government investments in science and technology, in other words, have the potential to address economic inequality. Like building a stock portfolio, it will take time for some of these investments to yield their full benefits. But this lodestar liberalism—anchored in values—has allowed the administration to forge bipartisan support in an otherwise fractious political milieu.

Flexible Benefits

The Biden administration has begun to make moves to apply the same approach to AI. In October 2022, the White House released its Blueprint for an AI Bill of Rights, which was distilled from engagement with representatives of various sectors of American society, including industry, academia, and civil society. The blueprint advanced five propositions: AI systems should be safe and effective. The public should know that their data will remain private. The public should not be subjected to the use of biased algorithms.

Consumers should receive notice when an AI system is in use and have the opportunity to consent to using it. And citizens should be able to loop in a human being when AI is used to make a consequential decision about their lives. The document identified specific practices to encode public benefits into policy instruments, including the auditing, assessment, "red teaming," and monitoring of AI systems on an ongoing basis.

The blueprint was important in part because it emphasized the idea that AI governance need not start entirely from scratch. It can emerge from the same fundamental vision of the public good that the country's founders articulated centuries ago. There is no society whose members will always share the same vision of a good future, but democratic societies are built on a basic agreement about the core values citizens cherish: in the case of the United States, these include privacy, freedom, equality, and the rule of law.

These long-standing values can—and must—still guide AI governance. When it comes to technology, policymakers too often believe that their approaches are constrained by a product's novelty and must be subject to the views of expert creators. Lawmakers can become trapped in a false sense that specific new technologies always need specific new laws. Their instinct becomes to devise new governance paradigms for each new tech development.

his instinct is wrong. Throughout history, the United States has reinterpreted and expanded citizens' rights and liberties, but the understanding that such entitlements and freedoms exist has been enduring. If policymakers return to first principles such as those invoked in the AI Bill of Rights when governing AI, they may also recognize that many AI applications are already subject to existing regulatory oversight.

Anchoring AI governance to a vision of the public good could diminish regulatory confusion and competition, stemming the flow of the sometimes contradictory bills lawmakers are currently producing. If it did, that would free both lawmakers and regulatory agencies to think more creatively in the areas in which policy innovation is truly needed. AI does pose unprecedented challenges demanding policy innovation. Already, the Department of Commerce's National Institute of Standards and Technology (NIST) has embarked on a different kind of policymaking when it comes to AI.

With a constitutional mandate to "fix the standard of weights and measures," NIST determines the proper standards to measure such things as length and mass, temperature and time, light and electricity. In 2021, Congress directed NIST to develop voluntary frameworks, guidelines, and best practices to steer the development and deployment of trustworthy AI systems, including ways to test for bias in AI training data and use cases. Following consultations with industry leaders, scientists, and the public, in January 2023, NIST released its first AI Risk Management Framework 1.0. The "1.0" was meaningful. Versioning—think of Windows 2.0, 3.0, and so on—has long been

commonplace in the world of software development to patch bugs, refine operations, and add improved features.

It is much less common in the world of policymaking. But NIST's use of policy versioning will permit an agile approach to the development of standards for AI. NIST also accompanied its framework with a "playbook," a practical guide to the document that will be updated every six months with new resources and case studies. This kind of innovation could be applied to other agencies. A more agile way of reviewing standards and policies should become a more regular part of the government's work.

The Old Becomes New

The AI Bill of Rights and the NIST AI Risk Management Framework became the foundations of Biden's sweeping October 2023 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Running at 111 pages, it mobilizes the executive branch to use existing guidelines, authorities, and laws, innovatively applied, to govern AI. This sweeping mandate gives many key actors homework: industry leaders must provide insight into the inner workings of their most powerful systems and watermark their products to help support information integrity. The order directed the U.S. Office of Management and Budget to issue guidance on the federal government's own use of AI, recognizing that the government possesses extraordinary power to shape markets and industry behavior by setting rules for the procurement of AI systems and demanding transparency from AI creators.

But more must be done. AI governance needs an international component. In 2023, the European Union advanced significant new laws on AI governance, and the United Kingdom is moving to address AI regulation with what it calls a "light touch." The African Union has a regional AI strategy, and Singapore has just released its second national AI strategy in four years.

There is a risk that the world at large will suffer from the same glut of competing proposals that bedevils AI governance in the United States. But there are existing multilateral mechanisms that can be used to help clarify international governance efforts: with the UN Charter and the Universal Declaration of Human Rights, UN members states have already agreed to shared core values that should also guide AI regulation.

Democratic leaders must understand that disrupting and outpacing the regulatory process is part of the tech industry's business model. Anchoring their policymaking process on fundamental democratic principles would give lawmakers and regulators a consistent benchmark against which to consider the impact of AI systems and focus

attention on societal benefits, not just the hype cycle of a new product. If policymakers can congregate around a positive vision for governing AI, they will likely find that many components of regulating the technology can be done by agencies and bodies that already exist. But if countries do decide they need new agencies—such as the AI Safety Institutes now being established in the United States and the United Kingdom—they should be imagined as democratic institutions that prioritize accountability to citizens and incorporate public consultation.

Properly constructed, such agencies could be a part of a broader governance infrastructure that not only detects how AI can infringe on rights and livelihoods but also scouts out how AI can proactively enhance them—by making dangerous jobs less perilous, health care more effective, elections more reliable, education more accessible, and energy use more sustainable. Although AI systems are powerful, they remain tools made by humans, and their uses are not preordained. Their effects are not inevitable.

AI governance need not be a drag on innovation. Ask bankers if unregulated lending by a competitor is good for them. Simply put, the ballast provided by proactive governance offers stability but also provides a controlled range of motion. First, however, policymakers must acknowledge that governing AI effectively will be an exercise in returning to first principles, not just a technical and regulatory task.

The Business of Knowing: Private Market Data and Contemporary Intelligence[38]

Klon Kitchen
Managing Director, Beacon Global Strategies

In January 2021, the United States Defense Intelligence Agency (DIA) acknowledged that it buys Americans' location data generated by their phones. Assuring legislators in a letter that "personnel can only query the US location database when authorized through a specific process," the DIA also argues that Fourth Amendment requirements for a warrant before collecting this information do not apply, because they are purchasing this data as a service and not using the power of law to compel its acquisition.1

Employing similar logic, the Department of Homeland Security (DHS), the Internal Revenue Service (IRS), the Federal Bureau of Investigation (FBI), the Drug Enforcement Administration (DEA), and other government organizations are also purchasing private market data (PMD)—data that is generated by consumers, companies, and other entities and that is collected, collated, analyzed, and sold by technology companies and data brokerage services. This, of course, is raising many concerns and questions. "It's critical we uncover how federal agencies are accessing bulk databases of Americans' location data and why," Nathan Freed Wessler, senior staff attorney with the American Civil Liberties Union's Speech, Privacy, and Technology Project, said in a statement.2 "There can be no accountability without transparency."3

Some will assume these practices illustrate a federal government run amok, intent on trampling Americans' constitutionally protected rights under the guise of "national security." Others will view cries of tyranny! and warnings about the "deep state" as nothing more than naivete about the realities of a dangerous world or fearmongering for political advantage. But the issue is more complicated, and there is another side of the story. Government access to PMD does implicate liberty concerns, but it also implicates security issues that require serious consideration if this constitutionally induced tension is to be properly balanced.

This paper argues that US government access to at least some private market data—and the limiting of foreign access to this same information—is essential for national security. It also argues, however, for a refined awareness that acknowledges the privacy we have already lost and that implements greater government oversight and accountability. It must also be said that this paper provokes more questions than it answers. It does not exhaustively assess or explain many of the relevant facts, trends, issues, and implications

[38] This essay was originally published by the Hoover Institutions' National Security, Technology, and Law journal: Aegis Series Paper No. 2110

cited. The aim here is to abstract from nuance and detail to explain how our nation has come to this place, and to emphasize the security implications of our chosen path forward.

The Proliferation of PMD and of Its Value for "Knowing"

In 2018, people created, captured, copied, and consumed 33 zettabytes (ZB) of data—approximately 33 trillion gigabytes or 128,906,250,000 maxed-out iPhone 12s' worth of information.4 This number jumped to 59 ZB in 2020 and is predicted to hit 175 ZB by 2025. Put another way: Humans currently produce 2.5 quintillion bytes of data every day.5 If you laid flat 2.5 quintillion pennies, you could cover the earth's surface five times. By 2025, this number is projected to be 463 exabytes every day. Again, for reference: If a gigabyte is the size of the earth, an exabyte is the size of the sun—and you can fit about 1.3 million earths in the sun.

To put it into even more accessible metrics, in every minute of every day in 2020, users uploaded 500 hours of video to YouTube, sent 41 million messages on WhatsApp, uploaded 147,000 photos to Facebook, installed TikTok 2,704 times, submitted 69,000 applications on LinkedIn, and hosted 208,000 Zoom meetings.6 Every minute. Every day. And this is only the beginning.

As fifth generation (5G) and subsequent telecommunications networks that can transport even more data come online, the oft-promised "Internet of Things" (IoT)—a world where the internet is not just a place you go on your phone, tablet, or laptop, but where it is everywhere, connecting almost everything, and is assumed the way one assumes air-conditioning when you walk into a building—is projected to include more than 30.9 billion IoT devices globally by 2025.7 We are not just awash in data; we are drowning in it, and the flood is rising exponentially.

That does not mean, however, that we are not leveraging this data. Quite the opposite in fact; whole economies are being built on this information that, as we will see, is becoming a critical national resource. But data are not most valuable in isolation. Data's true utility is realized when data are collected, collated, analyzed, and wrung dry of their attendant insights. These services are being offered by a growing number of technology companies and data brokers, and they are redefining economies and modern notions of what can be known and hidden about ourselves.

There are some 4,000 data brokerage companies around the world, with 87 percent of those companies headquartered in the United States.8 Just one of these data brokers,

estimates the US Federal Trade Commission (FTC), "has 3000 data segments for nearly every U.S. consumer."9 Another "has information on 1.4 billion consumer transactions

and over 700 billion aggregated data elements."10 And still another "adds three billion new records each month to its databases."11 One of the largest of these brokers, Acxiom, has 23,000 servers collecting and analyzing data on more than 500 million consumers worldwide.12 All of this adds up to an industry worth more than $200 billion that can accurately be described as the beating heart of the "knowledge economy."13

A key portion of this industry—and a part that helpfully illustrates just how valuable this information can be—is sometimes referred to as programmatic marketing or the programmatic web. Programmatic marketing is the use of artificial intelligence (AI) and robust data sets to enable highly tailored marketing based on a consumer's demographics, attitudes, and behaviors, as understood by an analysis of their digitized data. Programmatic marketing is why women between the ages of nineteen and thirty-six receive ads for baby clothes after they search for "best folic acid supplements." It is why men who are assessed to have a high likelihood of prostate cancer receive unsolicited online ads for erectile dysfunction drugs. And it is why ads for those shoes you looked at three weeks ago appear as you read the New York Times online.

Thomas Davenport, Abhijit Guha, and Dhruv Grewal have explained how companies can better use data and AI for programmatic marketing to improve their bottom lines.14 They divide these tools into two general types: task automation and machine learning. Task automation applications "perform repetitive, structured tasks that require relatively low levels of intelligence," according to the article.15 "They're designed to follow a set of rules or execute a predetermined sequence of operations based on a given input, but they can't handle complex problems such as nuanced customer requests."16 Examples would include a customer relationship manager program that automatically sends an email to new customers or basic consumer service chatbots like Facebook's Messenger bots.

Machine learning algorithms "are trained using large quantities of data to make relatively complex predictions and decisions. Such models can recognize images, decipher text, segment customers, and anticipate how customers will respond to various initiatives, such as promotions."17

Summarizing the utility of these applications, the authors are clear about their value:

> AI can streamline the sales process by using extremely detailed data on individuals, including real-time geolocation data, to create highly personalized product or service offers. Later in the journey, AI assists in upselling and cross-selling and can reduce the likelihood that customers will abandon their

digital shopping carts. For example, after a customer fills a cart, AI bots can provide a motivating testimonial to help close the sale—such as "Great purchase! James from Vermont bought the same mattress." Such initiatives can increase conversion rates fivefold or more.

After the sale, AI-enabled service agents from firms like Amelia (formerly IPsoft) and Interactions are available 24/7 to triage customers' requests—and are able to deal with fluctuating volumes of service requests better than human agents are. They can handle simple queries about, say, delivery time or scheduling an appointment and can escalate more-complex issues to a human agent. In some cases AI assists human reps by analyzing customers' tone and suggesting differential responses, coaching agents about how best to satisfy customers' needs, or suggesting intervention by a supervisor.18

In many ways, we are only at the forefront of programmatic marketing. As daily life becomes more digitized and as companies become more adept at collecting and leveraging our "digital exhaust," programmatic marketing will represent an unprecedented source of insight into our individual and our collective lives. This data can enable a near-total reconstruction of an individual's identity, location history, interpersonal relationships and networks, entertainment and purchasing preferences and habits, and even future economic, social, and political outcomes.

Facebook is a familiar example of the power and value of data. By creating an account and filling out a basic profile, the social media company learns a user's name, birth date, phone number, email address, contacts, schools attended, current and past occupations, relationship status, hometown, current city of residence, physical address, birth name, personal website, and other social media profiles. As you continue to use the site, Facebook learns where you like to visit, shop, and eat because you check in at these locations or post pictures of your experiences. Even if you do not post your location and even if you decline permission to share your GPS position, the company is able to follow your location by tracking the IP addresses and other information from the devices you use to access the social media service.

If you use Facebook Messenger to chat or to call your friends, the company says it does not record the content of those interactions, but it does know how often you speak with a contact and for how long. As you post and share content, the company learns even more about your religious, social, and political views, where and how you consume media, and what content you find most engaging. The company then combines this information with other "partner data," including information from other apps and even offline actions and purchases. And all of this is applied to more than 2.85 billion monthly active users globally—continually adding to and refining the Facebook social graph: a sophisticated

graph of the social relations and interactions between all of the entities on the social network.

All of this data collection translates into meaningful value for Facebook. In 2020, Facebook generated nearly $84.2 billion in ad revenues—nearly 90 percent of the company's total revenue—and the company accounts for nearly 10 percent of all

digital advertising globally.19 And this is just one company in a growing constellation of businesses who specialize in data generation, collection, and utilization. In fact, global programmatic advertising spending has almost doubled in the last four years and is expected to reach $155 billion by the end of this year.20

The simple but profound truth illustrated in this example is that modern marketing is fundamentally an "intelligence" operation. Governments around the world employ millions of people tasked with collecting, understanding, predicting, and shaping human behavior and events; but the private sector is pioneering this art and science and is functionally disrupting the state's monopoly on this critical capability. Even more, the data itself is overwhelmingly being generated and held in the commercial sector, where it is in some ways easier and in some ways harder to acquire.

The Need to "Know" Everything and the Promise of AI for National Security

Knowledge has always been a means to power. The more one knows, the better one can understand a situation, a challenge, an opportunity, or a risk. The gathering of knowledge, then, has always been a defining feature of national security and of American national security, specifically. After all, it is very difficult to defend against threats or to seize opportunities if you do not know about them.

In this vein, before the United States became a nation, General George Washington wrote of the "advantage of obtaining the earliest and best Intelligence of the designs of the Enemy," and charged Nathaniel Sackett with the creation of what would eventually become the Culper Spy Ring.21 This and other intelligence operations were so successful that, at the end of the Revolutionary War, British Major George Beckwith concluded, "Washington did not really outfight the British. He simply out-spied us."22 The value of intelligence to American security has persisted ever since.

The US intelligence community budget was $85.8 billion in 2020, spread across eighteen member departments and agencies, with at least 263 discrete intelligence organizations being established or restructured since 2001.23 This sprawling enterprise is arrayed against an equally diverse set of issues, according to the Office of the Director of National

Intelligence, including Russia, China, North Korea, Iran, Western isolationism, biological/chemical/nuclear WMDs, outer space, cyberspace, artificial intelligence, quantum computing, automation, nanotechnology,

biotechnology, global inequality, violent extremism, migration, urbanization, climate change, pandemics, and transnational crime.24 In fact, it is not hyperbole to assert that the United States has the largest, most diverse set of national interests—and, therefore, corresponding intelligence requirements—of any nation in the history of the world. This unprecedented interest and capacity also create an unending demand for information.

Importantly, it is essential to understand that the US intelligence community is tasked with much more than the anti-terrorism operations that are featured in pop culture. American policy makers lean on intelligence to inform their decisions on a much broader set of national security issues that increasingly intersect with an even broader array of facts and topics. Explaining this reality back in 2014, the DIA's then chief analytic methodologist, Josh Kerbel, observed the following:

> Today, however, the [intelligence community] no longer has the luxury of watching a single discrete entity that demands classified collection in order to obtain relevant data. There is a much more expansive range of interconnected and complex challenges. These challenges—economic contagion, viral political and social instability, resource competition, migration, climate change, transnational organized crime, pandemics, proliferation, cyber security, terrorism, etc.—are interdependent phenomena, not discrete "things." . . . Intelligence analysts must be capable of thinking creatively—holistically and synthetically across traditional boundaries. The long-held emphasis on reductive thinking that breaks issues into discrete pieces—reinforced by the compartmentalization associated with classified information—is no longer sufficient.25

Kerbel's point is that modern intelligence must account for the growing interconnectedness of
the world and of its attendant challenges. This, he argues, requires the intermingling of unclassified and classified data "holistically and synthetically" to enable complex understanding of complex problems. Intelligence must evolve, and it is.

But what is intelligence? It is necessarily more than data. It is, instead, data leveraged and applied. For national security purposes, it is not enough to know a fact. That fact must have context so that it is properly understood. Its relevance to mission requirements and the opportunities and risks created by its acquisition and use must also be assessed. Finally, information must be actionable, that is, it must enable action that

improves—or at least is thought to improve—the national security. In this sense, intelligence is not a single piece of information but is instead the product of data being pooled together in a manner that provides insights and then enables action.

A definition of intelligence from the Central Intelligence Agency (CIA) is clarifyingly simple: "Reduced to its simplest terms, intelligence is knowledge and foreknowledge of the world around us—the prelude to decisions and action by U.S. policymakers."26 If data leveraged and applied provides "knowledge and foreknowledge of the world around us,"27 then there is good reason to believe we are on the cusp of a golden age of intelligence—because, as we have seen, we are awash in data about our world.

But the US intelligence community faces a two-sided challenge in this regard: First, it cannot adequately process and use the data it has; and second, it is struggling to gain access to important nonclassified data sets—such as private market data—that could provide material advantage. The first is a technical challenge while the second is a political and legal one.

When it comes to better leveraging the data it has, the intelligence community, like the private sector, is placing its hopes in AI. Former Director of National Intelligence Dan Coats and former Principal Deputy Director of National Intelligence Susan Gordon outline the intelligence community's plight clearly:

> Closing the gap between decisions and data collection is a top priority for the Intelligence Community (IC). The pace at which data are generated and collected is increasing exponentially—and the IC workforce available to analyze and interpret this all-source, cross-domain data is not . . . the IC must adapt to the rapid global technological democratization in sensing, communications, computing, and machine analysis of data. These trends threaten to erode what were previously unique USIC capabilities and advantages; going forward, we must improve our ability to analyze and draw conclusions from IC-wide data collections at scale.28

Put simply: The intelligence community believes that emerging technologies are essential for the production of timely and valuable intelligence and that a failure to leverage these tools risks its irrelevance and the nation's security. To this end, the intelligence community has developed the Augmenting Intelligence using Machines (AIM) strategy, which explains how it intends to develop and to utilize artificial intelligence, process automation, and intelligence community officer augmentation (AAA) technologies to achieve its mission. As the intelligence community explains:

> The AIM initiative will enable the IC to fundamentally change the way it produces intelligence. We will achieve superiority by adopting the best available commercial AI applications and combining them with IC-unique algorithms and data holdings to augment the reasoning capabilities of our analysts. Simply stated, our goal is the following: "If it is knowable, and it is important, then we know it."29

The AIM strategy then provides four "primary investment objectives" that are essential for success. First, the IC must lay a digital foundation for long-term "science and technical intelligence." This involves the mundane, but critically important, acts of building cloud computing and other infrastructure, normalizing data standards, expanding government understanding of commercial offerings and supply chains, and baselining US and foreign AI capabilities and programs.

The second objective calls for the IC to expand its use of commercial and open-source AI. Agile and rapid acquisition is deemed critical for this requirement. Relatedly, the third AIM objective focuses on breaking down data-sharing barriers within the IC, with a special emphasis on the development of AI solutions that can ingest and process data from across all intelligence sources.

The fourth and final objective sets the stage for long-term thriving by requiring ongoing research and investment in AI models that go beyond simply "fusing" information, but that actually enable human analysts to better discover goals and intent or to extract entity information from incomplete or multimodal data.30

The reader need not fully understand each of these objectives, or even the larger AIM strategy. What is important to understand is that the intelligence community believes it must take significant and sustained action if it is to be effective going forward. Massive investments, new partnerships, and fundamental changes to established methodologies are deemed critical for future national security. If the director of the National Geospatial-Intelligence Agency was correct, for example, when he publicly estimated that the current acceleration of collection will require more than eight million imagery analysts by 2037 (an impossible demand to meet), it is easy to understand why the intelligence community feels such urgency and is placing such hope in the promise of artificial intelligence.31

But even if the intelligence community is able to meet the technical challenge of better leveraging all the data it has, it still faces the political and legal challenge of getting greater access to data that would significantly improve its ability to protect the nation—particularly data that is generated, collected, and analyzed in the private marketplace.

Foreign intelligence agencies like the CIA or the National Security Agency enjoy very broad collection authorities when it comes to non-US citizens. Domestic intelligence agencies like the DHS and the FBI have more constraints—especially when it comes to US citizens—but are still able to conduct extensive surveillance and analysis, when necessary, within existing legal frameworks. The need, then, for greater access to PMD is not primarily driven by tactical demands (though it would be helpful here too) but, instead, by the growing need for deep awareness at scale.

Twenty years after 9/11 the American government is well practiced and well enabled to do the type of "man-hunting" intelligence work that is featured so prominently in popular entertainment. But the return of so-called great power competition with other nations is reminding policy makers that true national security is not contained only within the need to "find, fix, and finish" an individual target—it also includes being able to understand, predict, and influence whole governments and populations, and private market actors are uniquely capable of collecting and using the data underlying such capabilities.

Specifically, private market data offers an appealing opportunity for the intelligence community to develop at-scale intelligence because it is unclassified, "rich," and recent.

First, private market data is unclassified—meaning it can be easily used and shared. This information is typically freely (if not always knowingly) provided by users in exchange for services, and most terms of service agreements allow the collecting entities to use or to sell this information in whatever way they choose. Anyone who purchases this data, likewise, has minimal constraints on what they can do with this information and whom they can sell it to or share it with. This agility and shareability is very attractive to an American government that is routinely beset by information silos and bureaucratic barriers to essential collaboration. The unclassified nature of this information also allows this data to be intermingled with other datastores, further enabling the data "fusion" and analytic sharing that is called for in the AIM strategy discussed above.

Second, private market data is "rich." This is true in both volume and detail. PMD is frequently collected on a massive scale (remember the FTC findings mentioned earlier) and this is important for identifying trends and gleaning insights at a societal level. Again, we have already considered the extreme detail of this data, so further discussion is not needed. The salient point of this "richness" is that when this volume of highly detailed data is combined with modern and emerging processing capabilities, it yields previously unimagined awareness at the macro, mezzo, and micro levels of the world.

Third, private market data is recent. The "every minute of every day" statistics shared earlier illustrate the volume of new PMD constantly being generated.32 And that is to say

nothing of the metadata—data that gives information about and describes other data—accompanying this content. This constantly refreshing torrent of information can provide insights into virtually every aspect of people's, and a nation's, economic, social, and political life. For an intelligence enterprise tasked with a real-time understanding of geopolitical realities strategically, operationally, and tactically, private market data constitutes an unparalleled pool of insights that is tantalizingly within reach.

The intelligence community's growing "need to know" and the emerging ubiquity of data together capture the proper context for understanding the government's attraction toward private market data. Here are two illustrations of how the government might specifically use this data to advance the nation's security.

Imagine the FBI learns that a known foreign weapons proliferator is attempting to supply a domestic terrorist group with radiological materials so that they can attack the US Senate with a "dirty bomb." It also discovers that this proliferator is attempting to use a known human-smuggling network to infiltrate the United States and to deliver this radiological material to his buyer. Now assume the Bureau has access to a facial recognition tool that scrapes social media and other open-source data sets and is able to identify the ringleader of the human-smuggling network by comparing a partial mirror reflection in a child exploitation video with a Facebook picture from another user that just so happens to capture the criminal in the background, establishing his presence at the time and location of the explicit video. This allows the ringleader to be identified, located, and arrested. Follow-on analysis not only allows law enforcement to disrupt the human-smuggling ring but also to lure the weapons proliferator and the domestic terrorists into a sting that prevents the US Senate attack, liberates scores of women and children, and results in multiple arrests and convictions.

Or, consider a larger geopolitical challenge. Imagine the US intelligence community has access to decades of agriculture, climate, and economic trade data that has been collected by dozens of private market sources, including "smart" farm equipment, digitized trading markets, and industry association reporting. Now imagine this data has been pooled and fused by the IC, allowing them to alert the president to a high risk of famine within a partner nation that, if allowed to take hold, would likely result in large-scale death, massive refugee migration into neighboring countries, and the significant weakening—possibly even the downfall—of a friendly government in a strategically important region. But because this warning was possible, international aid and support were mobilized, the crisis was averted, and the improved alliance enabled the United States even greater influence in the region.

Frankly, these two examples are narrow and are relatively simple applications of PMD. Far more sophisticated examples will be possible as more data is made available and as

AI capabilities develop. But both of these examples are rooted in real intelligence challenges and demonstrate the potential impact of government access to private market data. Now imagine if the government had failed to detect and disrupt either of these challenges—both could have catastrophic consequences.

The utility of PMD to modern intelligence does not, however, ameliorate the discomfort many feel regarding US government access to this data and the capabilities it is generating. This is why careful oversight will be essential.

Where We Are and What We Must Do

Concerns about the loss of privacy and liberty are well founded, and the American ethos has always suspected the accumulation of power by the state. The Constitution is primarily a restraining document on the government. It does not exhaustively list all of a citizen's rights; instead, it lists a limited number of specific powers and authorities of the state for the purposes of the common defense and ordered liberty.

But the growing scope of threats to the common defense and to our ordered liberty—alongside the undeniable value of PMD to securing these same objects—suggests that a refinement of the "social contract" is not only in order but is already occurring because the underlying drivers—data proliferation, the declining capacity of the US intelligence community to achieve its mission, and the migration of "intelligence" into the private sector—are only growing stronger. This, then, requires a clear understanding of where we now stand and of what we must now do.

First, Americans have already willingly ceded much of their privacy—at least as it has been popularly understood—to both governmental and corporate powers. I have discussed at length the troves of data that are collected and analyzed and what can be done with these insights. Shoshana Zuboff claims we now live in an age of "surveillance capitalism," which she defines as follows:

> 1. A new economic order that claims human experience as free raw material for hidden commercial practices of extraction, prediction, and sales;
> 2. A parasitic economic logic in which the production of goods and services is subordinated to a new global architecture of behavioral modification;
> 3. A rogue mutation of capitalism marked by concentrations of wealth, knowledge, and power unprecedented in human history;
> 4. The foundational framework of a surveillance economy;
> 5. As significant a threat to human nature in the twenty-first century as industrial capitalism was to the natural world in the nineteenth and twentieth;

6. The origin of a new instrumentarian power that asserts dominance over society and presents startling challenges to market democracy;
7. A movement that aims to impose a new collective order based on total certainty;
8. An expropriation of critical human rights that is best understood as a coup from above: an overthrow of the people's sovereignty.33

You need not fully embrace Zuboff's admittedly dire description to agree with her core claim that society is being reshaped through the generation and collection of private market data.

And people are feeling this change. According to Pew polling, 81 percent of polled Americans believe "they have little/no control" over what data is collected from them.34 Another 81 percent believe the "potential risks" of data collection "outweigh the benefits."35 More than three-quarters are "very/somewhat concerned" about how this data is collected.36 And nearly six in ten say "they have very little/no understanding" about how this information is used.37 So, clearly, there is broad-based recognition that large-scale data collection is eroding personal privacy.

But these concerns are not having an obvious impact on people's behavior. The number of American adults who own a smartphone has doubled since 2011 to nearly 85 percent.38 Social media usage is also booming, with 81 percent of Americans on YouTube, 69 percent on Facebook, 40 percent on Instagram, 31 percent on Pinterest, and 21 percent on Chinese-owned TikTok.39 Since 2016, Facebook has endured multiple scandals about its data collection and security—including the infamous Oxford Analytica fiasco and reports about it paying 13- to 17-year-olds $20 per month in exchange for nearly unfettered access to their mobile information—and yet its user base and profits have grown vastly during this same time period. In April 2021, Facebook reported more than $26 billion in revenue, which is a 48 percent increase over the previous year.40

These and similar statistics do not point to a market failure; they point to a market decision. As concerned as Americans are about the collection and use of their data, they are not sufficiently concerned to deny themselves the conveniences and benefits of the apps and services that harvest this data. This means, as Julia Angwin observed in Dragnet Nation, that people have reconciled themselves to a world in which you "can always be found . . . watched in your own home . . . no longer keep a secret . . . be impersonated . . . be financially manipulated."41 As disquieting as this may be, it is nevertheless a reality. Is it really surprising, then, that the US government sees this market decision and hopes that it too can benefit from this wealth of data—especially when the American people have such high expectations regarding their security?

The second reality we must reckon with is that "the common defense" now requires a greater contribution from the people. As previously stated, the United States has the largest, most diverse set of national interests—and, therefore, corresponding intelligence requirements—of any nation in the history of the world, and Americans have a very low tolerance for national security risk when push comes to shove.

To wit, after observing a decline in US public support for the dropping of two atomic bombs on Hiroshima and Nagasaki, Japan, from 85 percent in 1945 to 46 percent in 2015, Stanford scholars Scott Sagan and Benjamin Valentino wondered if this shift would hold up if Americans faced a similar challenge to World War II—drop the bomb and kill more than 100,000 Japanese or invade Japan and lose several thousand US soldiers.42 A Stanford news article explains:

> "We wondered what would happen today if Americans were faced with a similar tradeoff," Sagan said. "Has the U.S. public really changed? Or were previous polls misleading guides to real public attitudes about nuclear weapons use?"
>
> Sagan's findings from a survey experiment conducted in July 2015 involved a representative sample of the U.S. public asked about a contemporary, hypothetical scenario designed to replicate the 1945 decision to drop a nuclear bomb on Hiroshima.
>
> He and Valentino created a news story in which Iran attacked a U.S. warship in the Persian Gulf, Congress declared war, and the president was presented with the option of sending U.S. troops to march into Tehran, which would lead to many American military fatalities, or dropping a nuclear weapon on an Iranian city to try to end the war.43

The result?

> Their findings demonstrate that, contrary to the nuclear taboo thesis, a clear majority of Americans would approve of using nuclear weapons first against the civilian population of a nonnuclear-armed adversary, even killing 2 million Iranian civilians, if they believed that such use would save the lives of 20,000 U.S. soldiers.
>
> In addition, contrary to the principle of noncombatant immunity, an even larger percentage of Americans would approve of a conventional bombing attack designed to kill 100,000 Iranian civilians in the effort to intimidate Iran into surrendering, according to Sagan.44

Americans feel similar urgency on broader notions of national security. Nearly 70 percent of Americans say "taking measures to protect the U.S. from terrorist attacks" is a top long-range foreign policy goal45 and 45 percent say China is the United States' greatest enemy.46 Another 63 percent say "the economic power of China is a critical threat to the vital interests of the U.S. in the next 10 years."47 Finally, 70 percent of polled Americans say "international issues [are] relevant to their daily lives."48 What is the upshot of all of this? The people of the United States have a broadly shared concern about their peace and tranquility, and when these are perceived to be credibly threatened, they have high expectations that the government will decisively act.

It should be obvious by now that PMD can greatly enhance the government's ability to meet these expectations and to stay ahead of a constantly expanding list of threats.

But PMD is a broad category, and the IC's access to it is heavily influenced by how it is collected, who collects it, where it was collected, and from whom or what it is collected. These variables must be taken into consideration.

For example, any data collected by a foreign entity—government or nongovernment—from intelligence. There should be no constraint on their ability to buy, steal, or otherwise acquire this data because constitutional protections do not extend beyond our own citizens. Foreign-sourced data that includes US persons' data, including personally identifiable information (PII), should also be easily acquired, but will require special handling that minimizes the US persons' data. Such mitigation efforts are already integrated into the intelligence process and are easily applied here. Domestic private market data and data collection requires more protections.

Domestic intelligence agencies like the FBI and DHS should be given primary responsibility for acquiring and holding PMD from domestic sources that includes US persons' PII. This is in keeping with existing authorities and responsibilities and maintains the important distinction between domestic and foreign intelligence activities. Importantly, however, the IC must formalize capabilities and methodologies to "fuse" this data, while protecting Americans' PII, so that any insights that are relevant to the foreign intelligence mission are discovered and leveraged appropriately. Domestic data that does not include PII—such as economic data, climate data, generalized sociological statistics, and so on should generally be made available to the foreign-focused IC members either through purchase, information sharing agreements, legal mechanisms such as national security letters, or other routine channels.

But if the Leviathan is to be more heavily fed, its chains must also be reinforced. The American people can no longer accept the emaciated oversight and a near-total lack of

transparency regarding the US intelligence enterprise—particularly regarding the realm of government data acquisition and use.

For starters, Congress must improve its intelligence and cybersecurity oversight. The House and Senate Select Committees on Intelligence should require an annual report from the US IC on what PMD it is accessing, how this PMD is being leveraged (with specific examples of positive and negative outcomes), how the nation's geopolitical rivals are using this information, and other relevant reporting. The Director of National Intelligence should also consider issuing an annual unclassified report cataloging the IC's PMD acquisitions and partnerships. Some will argue that in the name of protecting sources and methods this information cannot or should not be shared. On the other hand, the reality is that if the citizens of the nation do not trust the government with this data in the first place, there will be no sources and methods to protect.

Finally, Congress should adopt the Cyberspace Solarium Commission's recommendation that the House and Senate form permanent select committees on cybersecurity.49 All cybersecurity-related budgetary and legislative jurisdiction should fall under these two committees and they would be responsible for overseeing the Executive's efforts to integrate cybersecurity strategy and policy within the government and between government and industry. A key aspect of this role would include overseeing how government and the private sector secure the PMD they acquire and exploring new technologies and methodologies that enable PMD to be leveraged while also expanding individual anonymity (e.g., homomorphic encryption).

Many other changes are in order but cannot be exhaustively cataloged here. The fundamental point that must be reiterated, however, is that if the state requires access to Americans' PMD in order to secure the nation, the government must also be willing to constrain itself to more robust oversight and accountability. If the Leviathan cannot or will not submit, it cannot be allowed to run free. Americans decided long ago that they would rather endure threats from abroad than tyranny at home.

As the nation negotiates this new balance between security and liberty, there is one obvious action that must be taken no matter how these tensions are resolved.

Limiting Foreign Government Access to US PMD

Even if the reader is not persuaded that PMD is vital for national security, the governments of other nations certainly are. The present risks of our citizens' data being sold to foreign governments are grossly underappreciated. Although plugging this gaping hole in our data security touches on a range of hot-button issues, banning the sale of

sensitive American data to adversarial governments should be an obvious priority for quick, decisive action.

Unsurprisingly, China already steals the type of bulk data sets on Americans that data brokers sell. In July of last year, FBI Director Christopher Wray noted, "If you are an American adult, it is more likely than not that China has stolen your personal data."50 Indeed, one of the largest Chinese hacks of Americans' personal data was that of Equifax, a leading data broker, resulting in the People's Republic of China (PRC) gaining information on almost half of all Americans. The Chinese Communist Party theoretically could have legally purchased the same information, probably with greater ease. We also know from the director of the United States National Counterintelligence and Security Center that China is using both "illegal and legal means" to collect bulk personal data of the sort sold by data brokers.51 Here is one example of how this data could be used against us.

Imagine that a hostile foreign nation is given access to huge stores of American social media data like photos, phone numbers, family members and contacts, locational data, online viewing and purchasing habits, political and social affiliations, "keyboard stroke patterns," and so on—all of which are routinely captured. Now imagine this government were to focus on the data generated around an important military installation like Fort Bragg, North Carolina—home to one of our nation's elite special mission units. Using just this data, a sophisticated intelligence activity could begin to identify individual members of this unit and their families. They could use GPS locations (or their absence) and social media posts discussing "TDYs"52 or "vacations" or "alone time" as a type of indication and warning notice for when members of this unit might be deploying. They could also follow the GPS locations of spouses to discover patterns of life or "inappropriate" relationships that could be leveraged for influence or blackmail. All of these, and much more nefarious deeds, are easily done with the information collected by virtually every application downloaded to a mobile phone.

Two main difficulties present themselves to redressing the issue. First, enforceability will be challenging. Data—even vast quantities of data—are notoriously "slippery," meaning it is difficult to track where it goes or what it is used for once it is transferred. While there is some ability to "hash" or "beacon" data so that it can be traced, these capabilities would be quickly overtaken by the scale of the data in question. An honest assessment must admit that, even if China is banned from purchasing American PMD, it is likely to acquire it through commercial cutouts and to continue to steal it. But imperfect security is not a justification for assuming unnecessary risk. To put it metaphorically, right now hostile regimes like those in Beijing and Moscow are making uncontested layups by purchasing US PMD. A ban on these purchases would at least push them back to the three-point line and put a hand in their face.

A second difficulty is the economic dimension. Given Chinese governments' unrestricted access to the data of companies operating in the PRC, regulations on data transfers could be disruptive and costly to a wide swath of businesses that work with companies in China.53 Depending on the form of the restrictions, businesses from a host of other countries that deal heavily in data, like Ireland, could also suffer considerable losses along with their American counterparts.54

Any viable solution would have to carefully address both of these challenges, balancing business interests with enforceability and maintaining enough adaptability to account for rapidly evolving technologies and privacy concerns.

So far, a few options have emerged. A new bill would have the Secretary of Commerce identify categories of personal data that are important to protect and data-receiving countries of concern, in order to administer licenses for data export.55 Others have suggested more intermediary measures, such as requiring data-selling companies to declare their foreign customers, or expanding the Committee on Foreign Investment in the United States process to restrict adversaries from buying their way into American data-brokering operations.56

Putting aside further questions of methods, however, the primary challenge to addressing the threat remains an insufficient sense of urgency. Corporate bulk data transfers don't quite trip the same alarms that hypersonic missiles do. But in a world in which data is the new oil, there is a very real sense in which these companies can sell off American security to our adversaries—with potentially devastating consequences.57 Yet the national security dimension of data brokering is pretty straightforward: Selling Americans' sensitive data to unfriendly foreign governments is a pressing security threat that should not be permitted.

Conclusion

Data is becoming the most plenteous and valuable resource on the planet. In it, we find a seemingly inexhaustible source of insight about ourselves and the world in which we live. These insights enable amazing opportunities and advancements for human thriving. Even more, technologists are pioneering mind-boggling methods for collecting, collating, understanding, and using data—many of which would have been thought to be impossible only a decade ago.

Disruption has always been a natural part of innovation, and certainly this is the case today. Political leaders, particularly, are being forced to accept that intelligence, "knowledge and foreknowledge of the world around us—the prelude to decisions and action,"58 is no longer the exclusive domain of governments but is, instead, a booming industry driven by private sector actors and capabilities. In recognition of this reality, the US intelligence community is turning to industry for help in fulfilling its constitutional mission to provide for the common defense. This provokes serious issues.

The IC's need for private market data (PMD) is clear. But the risks that come with government access to PMD are also clear. While the national security relevance of such access is increasingly compelling, it must be accompanied by corresponding constraints and accountability. A government unwilling to accept such restrictions and transparency inherently demonstrates that it cannot be trusted with such data.

Equally concerning is the PMD access currently enjoyed by hostile foreign governments like China. It is nothing short of madness for the US government to allow the sale of this data to entities we know are using it to imperil American people and interests. The idea that Beijing may have greater access to US PMD than the American government is obviously unacceptable and should be immediately addressed.

The American people and their government leaders cannot avoid these realities. Instead, they must adapt to them by refining our institutions and the critical balance between liberty and security. These changes necessarily require uncomfortable choices that bring with them no ironclad assurances of safety. But while an evolution as discussed in this paper does not guarantee success, a lack of adaptation will guarantee failure.

In the final analysis, one thing is clear: Going forward, we will all be "known." It is simply a matter of by whom and for what purpose.

Notes

Thanks to Jack Goldsmith and Andrew Keane Woods for comments on a prior draft.

1 Chris Mills Rodrigo (@millsrodrigo), Twitter (Jan. 22, 2021, 1:28 PM), https://twitter.com/millsrodrigo /status/1352684462795067393.

2 Sara Morrison, *A Surprising Number of Government Agencies Buy Cellphone Location Data. Lawmakers Want to Know Why*, Vox (Dec. 2, 2020, 4:25 PM), https://www.vox.com/recode/22038383/dhs-cbp-investigation-cellphone-data-brokers-venntel.

3 *Id.*

4 *See* David Reinsel, John Gantz & John Rydning, Data Age 2025: The Evolution of Data to Life-Critical—Don't Focus on Big Data; Focus on the Data That's Big 7 (2017), https://www.import.io/wp-content/uploads/2017/04/Seagate-WP-DataAge2025-March-2017.pdf.

5 *See* Jacquelyn Bulao, *How Much Data Is Created Every Day in 2021?*, TechJury (Aug. 6, 2021), https://techjury.net/blog/how-much-data-is-created-every-day/.

6 *Data Never Sleeps 8.0*, Domo, https://www.domo.com/learn/infographic/data-never-sleeps-8 (last visited Aug. 14, 2021).

7 Lionel Sujay Vailshery, *IoT and Non-IoT Connections Worldwide 2010-2025*, Statista (Mar. 8, 2021), https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/.

8 *See What Are Data Brokers—and What Is Your Data Worth?* [Infographic], WebFX Blog (Mar. 16, 2020), https://www.webfx.com/blog/internet/what-are-data-brokers-and-what-is-your-data-worth-infographic/; Kevin B. Johnston, *Top 15 Broker-Dealer Firms for 2020*, Investopedia (Aug. 2, 2019), https://www.investopedia.com/investing/broker-dealer-firms/.

9 Fed. Trade Comm'n, Data Brokers: A Call for Transparency and Accountability IV (2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

10 *Id.*

11 *Id.*

12 *Id.* at 8.

13 *What Are Data Brokers—and What Is Your Data Worth?* [Infographic], *supra* note 8.

14 Thomas H. Davenport, Abhijit Guha & Dhruv Grewal, *How to Design an AI Marketing Strategy: What the Technology Can Do Today—and What's Next*, Harv. Bus. Rev., July–Aug. 2021, https://store.hbr.org/product/how-to-design-an-ai-marketing-strategy/S21041. 15 *Id.*

16 *Id.*

17 *Id.*

18 *Id.*

19 *See* Statista Research Department, *Facebook's Advertising Revenue Worldwide from 2009 to 2020*, Statista (Feb. 5, 2021), https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide/.

20 *See* Statista Research Department, *Global Programmatic Advertising Spending from 2017 to 2021*, Statista (Apr. 1, 2021), https://www.statista.com/statistics/275806/programmatic-spending-worldwide/.

21 Natasha Bertrand & Michael B. Kelley, *This Letter from George Washington Marks the Birth of American Espionage*, Bus. Insider (Feb. 25, 2015, 6:42 AM), https://www.businessinsider.com/this-letter-from-george-washington-is-the-birth-of-american-espionage-2015-2.

22 *George Washington, Spymaster*, George Washington's Mount Vernon, https://www.mountvernon.org/george-washington/the-revolutionary-war/spying-and-espionage/george-washington-spymaster/ (last visited Aug. 14, 2021).

23 *See* Press Release, Office of the Director of National Intelligence, DNI Releases Appropriated Budget Figure for 2020 National Intelligence Program (Oct. 21, 2020), https://www.dni.gov/index.php/newsroom/press-releases/item/2161-dni-releases-appropriated-budget-figure-for-2020-national-intelligence-program; Web Politics Editor, *'Top Secret America'—Yahoo! News on the 'Top 10 Blockbuster Revelations,'* Wash. Post (July 21, 2010), http://voices.washingtonpost.com/top-secret-america/2010/07/top_secret_america_-_yahoo_new.html.

24 *See* Off. Dir. Nat'l Intel., National Intelligence Strategy of the United States of America 2019, at 4, https://www.dni.gov/files/ODNI/documents/National_Intelligence_Strategy_2019.pdf.

25 Josh Kerbel, *The US Intelligence Community's Kodak Moment*, Nat'l Int. (May 15, 2014), https://nationalinterest.org/feature/the-us-intelligence-communitys-kodak-moment-10463.

26 Cent. Intel. Agency, A Consumer's Guide to Intelligence, at vii (1999).

27 *See id.*

28 Off. Dir. Nat'l Intel., The AIM Initiative: A Strategy for Augmenting Intelligence Using Machines, at III (2019), https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf (emphasis added).

29 *Id.* at IV (quoting Principal Deputy Director of National Intelligence Sue Gordon).

30 *Id.* at V.

31 *See* Robert Cardillo, Director, Nat'l Geospatial-Intel. Agency, Remarks at the 31st Annual Small Satellites – Big Data Conference (Aug. 7, 2017), https://www.nga.mil/news/Small_Satellites_-_Big_Data.html.

32 *Data Never Sleeps 8.0*, *supra* note 6.

33 Shoshana Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power, at VII (2019).

34 Brooke Auxier et al., *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information*, Pew Research Center (Nov. 15, 2019), https://www.pewresearch.org/internet

/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

36 *Id.*

37 *Id.*

38 S. O'Dea, *Smartphone Ownership in the US 2011–2021*, Statista (May 12, 2021), https://www.statista.com/statistics/219865/percentage-of-us-adults-who-own-a-smartphone/.

39 Brooke Auxier & Monica Anderson, *Social Media Use in 2021*, Pew Research Center (Apr. 7, 2021), https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

40 Press Release, Facebook, Facebook Reports First Quarter 2021 Results (Apr. 28, 2021), https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-First-Quarter-2021-Results/default.aspx.

41 Julia Angwin, Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance 6 (2014).

42 Scott D. Sagan & Benjamin A. Valentino, *Revisiting Hiroshima in Iran: What Americans Really Think about Using Nuclear Weapons and Killing Noncombatants*, 42 Int'l Sec. 41, 41–46 (2017).

43 Clifton B. Parker, *Public Opinion Unlikely to Curb a US President's Use of Nuclear Weapons in War, Stanford Scholar Finds*, Stan. Univ. News Serv. (Aug. 8, 2017), https://news.stanford.edu/2017/08/08/americans-weigh-nuclear-war/.

44 *Id.*

45 *Amid G7 Summit, Most Americans Confident in Biden's Handling of World Affairs*, Ipsos (June 13, 2021), https://www.ipsos.com/en-us/news-polls/g7-kicks-off-most-americans-confident-bidens-handling-world-affairs.

46 Mohamed Younis, *New High in Perceptions of China as US's Greatest Enemy*, Gallup (Mar. 16, 2021), https://news.gallup.com/poll/337457/new-high-perceptions-china-greatest-enemy.aspx.

47 *Id.*

48 *US Adults' Knowledge about the World*, Council on Foreign Rels. (Dec. 2019), https://www.cfr.org/report/us-adults-knowledge-about-world.

49 US Cyberspace Solarium Comm'n, 116th Conc., Final Report 2 (2020), https://www.solarium.gov/report.

50 Christopher Wray, Director, Fed. Bureau Investigation, The Threat Posed by the Chinese Government and the Chinese Communist Party to the Economic and National Security of the United States, Remarks at the Hudson Institute Video Event: China's Attempt to Influence US Institutions (July 7, 2020), https://www.fbi.gov/news/speeches/the-threat-posed-by-the-chinese-government-and-the-chinese-communist-party-to-the-economic-and-national-security-of-the-united-states.

51 Zach Dorfman, *China Used Stolen Data to Expose CIA Operatives in Africa and Europe*, Foreicn Pol'y (Dec. 21, 2020, 6:00 AM), https://foreignpolicy.com/2020/12/21/china-stolen-us-data-exposed-cia-operatives-spy-networks/.

52 A military acronym for "temporary duty assignment," or travel.

53 *See* Klon Kitchen, *Why America Needs a Clear Policy to Deal with Chinese Cyber Security Concerns*, Nat'l Int. (Feb. 18, 2021), https://nationalinterest.org/blog/buzz/why-america-needs-clear-policy-deal-chinese-cyber-security-concerns-178434.

54 *See New US Senate Bill May Stop Ireland Processing US Data, Unless Ireland Acts on GDPR Enforcement*, Irish Council for C.L. (Apr. 15, 2021), https://www.iccl.ie/news/new-us-senate-bill-may-stop-ireland-processing-us-data-unless-ireland-acts-on-gdpr-enforcement/.

55 *See* Drew Harwell, *Wyden Urges Ban on Sale of Americans' Personal Data to 'Unfriendly' Foreign Governments*, Wash. Post (Apr. 15, 2021, 7:00 AM), https://www.washingtonpost.com/technology/2021/04/15/personal-data-foreign-government-ban/.

56 Michael Kans, *Data Brokers and National Security*, Lawfare (Apr. 29, 2021, 8:01 AM), https:// www.lawfareblog.com/data-brokers-and-national-security.

57 *See The World's Most Valuable Resource Is No Longer Oil, but Data*, Economist (May 6, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

58 *See* Cent. Intel. Acency, *supra* note 26 at vii.

Unlocking the Potential of AI through Policy that Ensures Trust and Adoption

David Rhew
Global Chief Medical Officer & Vice President Healthcare, Microsoft
Adjunct Professor, Stanford University School of Medicine

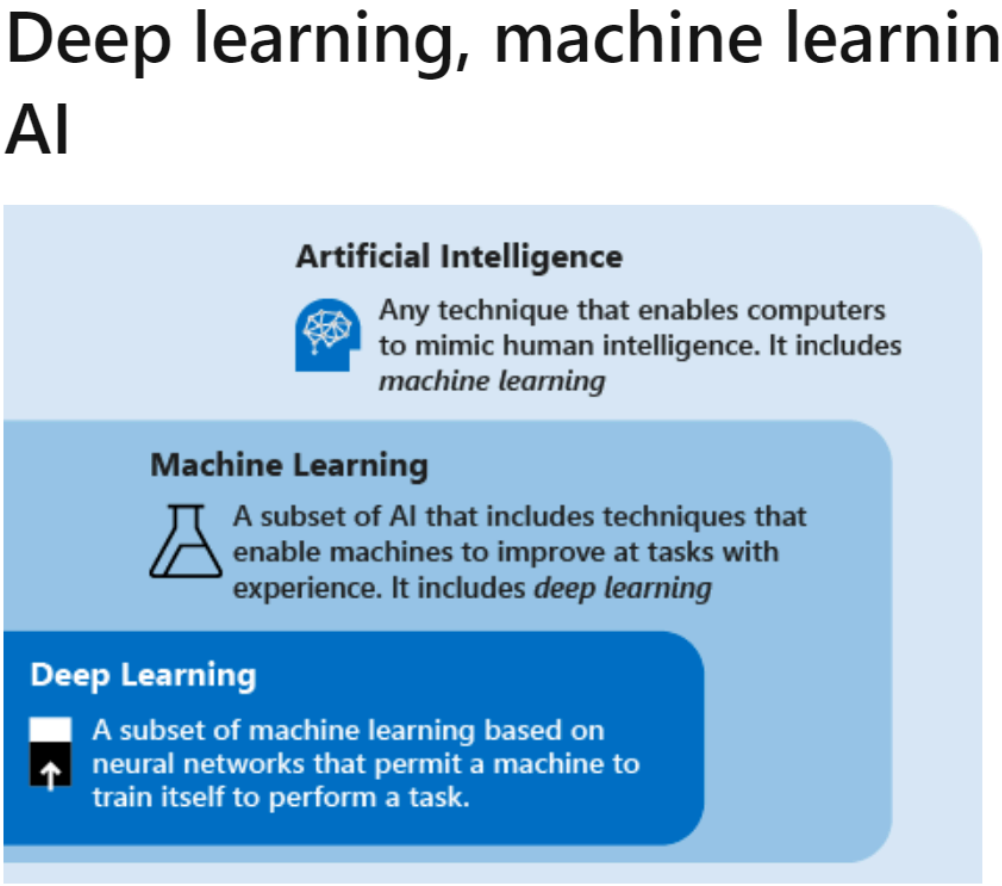
Artificial Intelligence (AI) has the potential to help us address some of healthcare's biggest challenges. Some examples are provided below:

- **Healthcare waste**. In 2022, the U.S. spent $4.5 trillion on U.S. healthcare, accounting for 17.3% of the Gross Domestic Product. Approximately 25% of that spend is waste (Shrank et al. JAMA. 2019). AI has the potential to streamline care, improve operational efficiencies, and reduce wasteful spending.
- **Access to affordable healthcare**. Individuals who have limited access to healthcare are less likely to receive preventive care services, which means that when they get sick, they are more likely to present with advanced stage illness. AI can be used to screen large populations and identify individuals before they present signs and symptoms for conditions such as diabetes, cardiovascular disease, hypertension, hyperlipidemia, and cancer.
- **Clinician burnout.** According to both the American Medical Association and American Nursing Association, over 60% of physicians and nurses are burned-out, and many are leaving the practice of medicine, leading to workforce shortages. The most alarming gaps are expected in primary care and rural communities. Without clinicians, we reduce patients' access and continuity of care, cause health costs to rise, and worsen health disparities. A major contributing factor to burnout is the increasing amount of administrative burden that clinicians experience on a daily basis, such as documenting into the electronic health record, filling out paperwork, addressing insurance claims, filling out forms, and answering emails. Many of these tasks can be streamlined with AI.

However, the challenge is not just about assessing how AI can be used in healthcare, but ensuring that individuals, institutions, and policy makers develop trust that AI will be implemented safely, securely, and responsibly, and that each of us feels confident in our ability to properly implement AI. In other words, successful adoption of AI requires that we focus on technology, process, and people.

Technology

The term Artificial Intelligence (AI) was first introduced in 1956. Machine Learning (ML) was introduced in 1959. Together AI/ML have demonstrated a high level of accuracy and reliability. The major limitation for AI/ML has been the size and diversity of the training data sets.



Reference: Microsoft Learn. 2024

Deep learning and neural networks were introduced in 1967, followed by generative AI in 2021. **Generative AI** involves predicting/generating text, images, audio, code, or other types of content, often in response to a prompt entered by a user. The capabilities and performance of generative AI have continued to increase rapidly and dramatically, which is why many experts and observers believe that we are now entering into a new age of AI, where AI will transform every aspect of our lives. According to a February 2024 New England Journal of Medicine AI article[39] , some potential "low-hanging fruit" use cases for healthcare include:

[39] https://ai.nejm.org/doi/full/10.1056/AIp2400036

1. "Enhancement of the doctor–patient interaction, including capture of the recorded patient visit
2. Prioritizing and analysis of test and imaging results
3. Differential diagnosis
4. Plan for therapy and discussion of alternatives
5. Instructions for the patient and caregivers
6. Appointment scheduling and other administrative functions
7. Responses to patient questions at optimal levels of literacy."

Process

In order to ensure trust in AI, we need to identify and operationalize responsible AI (RAI) principles. The Coalition for Health AI (CHAI) is a non-profit organization that works in close collaboration with healthcare providers, academia, non-profits, and industry, along with U.S. government agency observers including AHRQ, CMS, FDA, HHS, ONC, NIH, and White House OSTP. CHAI has released a blueprint for trustworthy AI in healthcare that includes the following principles:

- Usefulness
- Safety
- Accountability & Transparency
- Explainability & Interpretability
- Fairness with mitigation of bias
- Security and resilience
- Privacy-enhanced

Another group of healthcare organizations and technology enablers have formed the Trustworthy and Responsible Health AI Network (TRAIN), whose mission is to enable all healthcare organizations, including those with low resources, the ability to implement RAI principles at scale. They hope to develop and implement RAI guardrails, mechanisms and safeguards designed to prevent misuse, protect user privacy, and promote transparency and fairness, so that each organization can apply RAI in their own setting. They also look to create a nationwide network that would allow for AI algorithms to be tested at sites across the network through a federated process with privacy-preserving technologies to ensure that both the data and AI are kept private and secure.

People

Ultimately, none of this matters unless people feel confident in their ability to use and manage AI responsibly. With every technological advancement, we have seen a lag in the

time to acquire skills required to properly adopt technology. To address this, we will need to rethink how individuals get educated and skilled beyond training for AI and technology jobs. All individuals will need to understand how to incorporate AI into their daily work. McKinsey refers to these jobs as '**AI Translators**' and estimates that the AI talent gap for AI Translators will be greater than 25 times larger than the gap for AI technology roles. McKinsey also estimates that by 2030, 65% of all job skills will be adjusted to include AI. (McKinsey Global Institute, LinkedIn Demography- US Data – 2023 Industry).

The learnings for how to apply AI in a work setting will come largely from industry as opposed to traditional educational institutions. However, the vehicle for educating and training at scale will need to come from trade schools, colleges, and universities. This means that we need a mechanism to capture the learnings from industry and bring them to institutions of learning.

Role of Policymakers

We have a tremendous opportunity to address some of the biggest challenges in healthcare through AI. However, the barriers to implementation and adoption extend beyond the control of individual organizations. They involve developing and/or reinforcing policies that touch on the following areas:

1. **AI policy**. AI algorithms need to be tested in local and diverse data sets. They also need to be assessed post-deployment. Privacy standards should ensure that both data stewards and AI algorithm developers feel comfortable and confident that neither the data nor the AI will be exposed. These are examples of complex challenges that can potentially be solved through **public-private partnerships** in which policymakers work with developers, implementers, and technology enablers to ensure that AI algorithms are governed and managed appropriately. This public-private partnership approach is currently being explored by CHAI in collaboration with federal agencies.
2. **Skilling & reskilling for AI.** To change the way educational institutions train and skill individuals will require a collaborative mindset and an alignment of incentives. Today, industry and education stakeholders in states such as Michigan, Texas, California, and New York are beginning to discuss how to **develop AI curricula** that bring AI industry knowledge onto educational platforms. We are also seeing grass roots efforts underway in which educational institutions and technology organizations are hosting generative AI **prompt-a-thons** in which individuals are participating in hands-on sessions where they can learn how to use generative AI. We need to find a way to scale these types of efforts, so that we can '**train-the-trainers,**' establish sustainable programs to skill and reskill frontline

workers and managers on how to use and manage AI responsibly, and ultimately close the AI skills gap.

3. **Multistakeholder engagement and collaboration.** In an ever-evolving AI landscape, it is important that policy supports, but does not limit innovation. Fair and inclusive exchange of knowledge, cooperation, and addressing any common issues among government, public sector, academia, and industry will help to maximize the advantages this quickly developing technology can offer to society.

Summary

AI has the potential to help address some of healthcare's most pressing and complicated challenges. However, it is only useful if adopted, and adoption requires that we develop trust in the processes to ensure that AI can be deployed safely, fairly, and responsibly. It also requires that we gain confidence in our ability to use and manage AI. Policymakers can help accelerate AI adoption through (1) public-private partnerships, (2) AI skilling and reskilling programs, and (3) multistakeholder engagement and collaboration.

Recommended Readings

1. Lee, P., Goldberg, C., Kohane, I. The AI Revolution in Medicine: GPT-4 and Beyond. 1st Edition. Published by Pearson. 2023. https://aka.ms/AIRevolution
2. Horvitz, E. AI Anthology: A collection of essays on the future of AI. Microsoft Unlocked. 2023. https://aka.ms/AIessays
3. Yaraghi, N. Generative AI in health care: Opportunities, challenges, and policy. January 8, 2024. https://aka.ms/BrookingsGenAI24
4. Bhasker, S., Bruce, D., Lamb, J., Stein, G. Tackling healthcare's biggest burdens with generative AI. July 10, 2023. https://aka.ms/McKinseyHealthcareGenAI

Generative AI Risk Factors on 2024 Elections

Vivian Schiller
Vice President and Executive Director, Aspen Digital, Aspen Institute

Josh Lawson
Director, AI and Democracy, Aspen Digital, Aspen Institute

More than 2 billion people are eligible to participate in major global elections that will occur throughout 2024. With anti-democratic movements deepening their grip, the stakes could not be higher.

Trust in democratic institutions and facts themselves faced headwinds long before the public gained access to new generative AI tools. While the underlying technology is not entirely new, OpenAI's public launch of ChatGPT in November 2022 unleashed a mix of euphoria and hand wringing from a public coming to terms with such capable tools.

There is no firm consensus whether generative AI threats in the civic context represent *a difference in degree* or *a difference in kind*. Some suggest wide availability of fast-evolving AI tools simply exacerbate familiar misinformation challenges from familiar bad actors. But others cite the exponential rate of technological improvements and the promise of ever-greater speed, scale, and sophistication as reason enough to expect and to counter a dramatic erosion in trust across democratic institutions, including elections.

Our research at Aspen Digital yielded seven risk factors:

1. ***Siloed Expertise:*** Elections officials are not up to date on AI capabilities and unlikely to know where they can turn for help, we found in conversations. The AI labs and some tech companies are not attuned to the challenges elections officials face. **"There's very little understanding about how democracy works,"** said an expert who engages regularly with AI labs and tech companies. Dots aren't being connected amongst AI experts; mis- and dis-information specialists; elections officials; and policymakers. This is certainly true in the US, and we expect even greater disparities globally.

2. ***Public Readiness:*** Experts doubt the public will be resilient in the face of AI tools, which some expect to "flood the zone" with believable falsehoods during crises. Even if the public infrequently encounters AI-generated content, a surge in press coverage around AI capabilities might be enough to trigger public reactions that affect civic behavior–including an erosion in public trust overall. As a result,

people may revert to sources they already trust regardless of veracity, or ***reject factuality in general, a phenomenon known as the*** "liar's dividend." These outcomes do not require personal exposure to fake content but may occur simply because the public is aware that content could be fake.

3. ***Inadequate Platform Readiness:*** Over the last two years, major platforms have cut staff across integrity operations, and offered less transparency to media and researchers. Generative AI is likely to pressure already-taxed platform resources, experts said. The capacity to generate volumes of content at speed may overwhelm fact-checking efforts, even as the ability to produce unlimited variations of the same underlying claim might avoid detection by integrity tools built to prioritize the virality of a particular post (not a general claim).

4. ***Slow Moving regulation:*** The EU recently enacted regulations to hold platforms accountable for "harmful content" (or face a financial penalty), and they are acting quickly to create an "AI Act" that could have broad implications worldwide. The AI Act, along with a joint effort between the US and the EU to create a transatlantic AI Code of Conduct are under consideration, but would take so long to be adopted that they will not impact the 2024 election cycle. In the US, many efforts are underway at the local and state levels, but federal policy is not expected before November elections.

5. ***Increasing Quality AI-Generated Media:*** As AI-generated content has increased in quality, visual instinct alone is increasingly unreliable. Consequently, policymakers and others have shifted their mitigation efforts to overt labeling of AI generated content, digital signatures–so-called "watermarking" technologies that are still in their infancy.

6. ***Scaled Distribution at High-Speed:*** Until recently, substantial resources were needed to draft convincing misinformation or to effectively alter audio/visual content. Technical expertise and language requirements prevented some bad actors from creating and distributing large volumes of content. AI dramatically lowers these barriers and allows people to generate high-quality content that may restate the same false claim in many different ways or depict fake events from multiple camera angles, for example.

7. ***Message Targeting & Hyperlocal Misinformation:*** Generative AI may supercharge targeting capabilities by allowing creators to dramatically scale so-called "A/B testing," producing so many variations of content that targeting models grow exponentially robust as users engage with particular messages. Some believe these capabilities will result in such granular targeting that messages will

essentially be honed to particular psychological profiles–what some have called "superhuman persuasion."

AI may also generate compelling content that appears credible simply because it references highly localized information–" hyperlocal misinformation" -- such as the name of the school where a precinct is located or the names of streets and neighborhoods. Some are concerned that people will use AI to create hyperlocal misinformation about conditions at critical polling locations or safety in certain locations on Election Day. The risk is particularly acute given improvements across language groups.

8. ***Automated Harassment:*** Bad actors may create harassing content targeting elections administrators, activists, journalists, and other civic leaders and topics for a number of reasons: to intimidate, to reduce the algorithmic distribution of a post by adding large volumes of toxic comments, or to sway opinion during a crisis by appropriating particular hashtags.

9. ***Cybersecurity of Elections Infrastructure:*** Experts we spoke with raised concern that generative AI is a boon for social engineering scams, including phishing attacks, raising concerns that AI-enabled audio impersonation could spoof official communications from superiors to poll workers AI capabilities are also expected to enhance malware as fast-evolving code generation and analysis features are increasingly integrated into AI tools.

These developments underscore the urgent need for coordination, prioritization, and accountability across all sectors with stakes in a shared democratic future. The coming months will require policymakers, tech companies, and civil society to take responsible action in the face of evolving social and technological shifts in a critical election year.

Will 2024 Be the Year of Responsible AI?[40]

Yolanda Botti-Lodovico
Policy and Advocacy Lead, the Patrick J. McGovern Foundation

Vilas Dhar
President, the Patrick J. McGovern Foundation

As artificial intelligence becomes ubiquitous, we have an opportunity to harness its power to bring about an equitable, prosperous future. But to achieve this, we must heed the lessons of the digital revolution, maintain the current momentum, and prioritize fairness over corporate profits.

CHICAGO/WASHINGTON, DC – The start of 2024 has been marked by a wave of predictions regarding the trajectory of artificial intelligence, ranging from optimistic to cautious. Nevertheless, a clear consensus has emerged: AI is already reshaping human experience. To keep up, humanity must evolve.

For anyone who has lived through the rise of the internet and social media, the AI revolution may evoke a sense of déjà vu – and raise two fundamental questions: Is it possible to maintain the current momentum without repeating the mistakes of the past? And can we create a world in which everyone, including the 2.6 billion people who remain offline, is able to thrive?

Harnessing AI to bring about an equitable and human-centered future requires new, inclusive forms of innovation. But three promising trends offer hope for the year ahead.

First, AI regulation remains a top global priority. From the European Union's AI Act to US President Joe Biden's October 2023 executive order, proponents of responsible AI have responded to voluntary commitments from Big Tech firms with policy suggestions rooted in equity, justice, and democratic principles. The international community, led by the newly established United Nations High-Level Advisory Body on AI (one of us, Dhar, is a member) is poised to advance many of these initiatives over the coming year, starting with its interim report on Governing AI for Humanity.

Moreover, this could be the year to dismantle elite echo chambers and cultivate a global cadre of ethical AI professionals. By expanding the reach of initiatives like the National Artificial Intelligence Research Resource Task Force – established by the United States'

[40] This essay was originally published by Project Syndicate on January 30,2024

2020 AI Initiative Act – and localizing implementation strategies through tools such as the UNESCO Readiness Assessment methodology, globally inclusive governance frameworks could shape AI in 2024.

At the national level, the focus is expected to be on regulating AI-generated content and empowering policymakers and citizens to confront AI-powered threats to civic participation. As a multitude of countries, representing more than 40% of the world's population, prepare to hold crucial elections this year, combating the imminent surge of mis- and disinformation will require proactive measures. This includes initiatives to raise public awareness, promote broad-based media literacy across various age groups, and address polarization by emphasizing the importance of empathy and mutual learning.

As governments debate AI's role in the public sphere, regulatory shifts will likely trigger renewed discussions about using emerging technologies to achieve important policy goals. India's use of AI to enhance the efficiency of its railways and Brazil's AI-powered digital-payment system are prime examples.

In 2024, entities like the UN Development Programme are expected to explore the integration of AI technologies into digital public infrastructure (DPI). Standard-setting initiatives, such as the upcoming UN Global Digital Compact, could serve as multi-stakeholder frameworks for designing inclusive DPI. These efforts should focus on building trust, prioritizing community needs and ownership over profits, and adhering to "shared principles for an open, free, and secure digital future for all."

Civil-society groups are already building on this momentum and harnessing the power of AI for good. For example, the non-profit Population Services International and the London-based start-up Babylon Health are rolling out an AI-powered symptom checker and health-provider locator, showcasing AI's ability to help users manage their health. Similarly, organizations like Polaris and Girl Effect are working to overcome the barriers to digital transformation within the non-profit sector, tackling issues like data privacy and user safety. By developing centralized financing mechanisms, establishing international expert networks, and embracing allyship, philanthropic foundations and public institutions could help scale such initiatives.

As nonprofits shift from integrating AI into their work to building new AI products, our understanding of leadership and representation in tech must also evolve. By challenging outdated perceptions of key players in today's AI ecosystem, we have an opportunity to celebrate the true, diverse face of innovation and highlight trailblazers from a variety of genders, races, cultures, and geographies, while acknowledging the deliberate marginalization of minority voices in the AI sector.

Organizations like the Hidden Genius Project, Indigenous in AI, and Technovation are already building the "who's who" of the future, with a particular focus on women and people of color. By collectively supporting their work, we can ensure that they take a leading role in shaping, deploying, and overseeing AI technologies in 2024 and beyond.

Debates over what it means to be "human-centered" and which values should guide our societies will shape our engagement with AI. Multi-stakeholder frameworks like UNESCO's Recommendation on the Ethics of Artificial Intelligence could provide much-needed guidance. By focusing on shared values such as diversity, inclusiveness, and peace, policymakers and technologists could outline principles for designing, developing, and deploying inclusive AI tools. Likewise, integrating these values into our strategies requires engagement with communities and a steadfast commitment to equity and human rights.

Given that AI is well on its way to becoming as ubiquitous as the internet, we must learn from the successes and failures of the digital revolution. Staying on our current path risks perpetuating – or even exacerbating – the global wealth gap and further alienating vulnerable communities worldwide.

But by reaffirming our commitment to fairness, justice, and dignity, we could establish a new global framework that enables every individual to reap the rewards of technological innovation. We must use the coming year to cultivate multi-stakeholder partnerships and promote a future in which AI generates prosperity for all.

Navigating the AI Era. Here's How the U.S. Can Maintain Its Edge – and Improve the Lives of All Americans

Anna Makanju

Vice President, Global Affairs, OpenAI

Today's neural networks, often called "generative AI", have been compared to nuclear power, the internet, or electricity. None of these analogies is quite right, but what they capture is the vast array of possibilities this technology carries to improve lives or to inflict harm. At the heart of OpenAI's mission is the imperative to tilt this balance decidedly in favor of humanity's benefit—an ambitious goal that hinges on far more than technological supremacy. It requires the United States to lead, not just by outpacing adversaries and implementing adequate guardrails, but through a holistic strategy that includes substantial investment in national infrastructure for AI, comprehensive support for citizens navigating this new landscape, and, most importantly, the ethical and innovative application of advanced AI.

This essay will examine primarily the last issue. OpenAI invests heavily into safety research and mitigations, and this is a critical area of focus for policymakers. Unfortunately, the equally critical issue of government implementation of AI to leverage its benefits for citizens has so far taken the back seat. Perhaps this is because the current public discourse around the benefits of advanced AI tends to focus on capabilities that might one day provide benefits instead of those that are already benefiting people. I will provide a number of examples below, but first I want to offer a framework for understanding the potential benefits provided by generative AI (a term I dislike precisely because it obfuscates the technology's most useful capabilities).

In particular, generative AI is excellent at the following:

1. Providing comprehensive summaries and analyses of complex information from diverse sources like legislative documents, academic literature, websites, and reports - distilling key insights, extracting relevant details, and synthesizing overarching themes on demand.

2. Streamlining workflows by automating information lookup, data extraction, and question-answering tasks that require integrating content from multiple sources while accounting for specific contexts or requirements.

3. Tailoring communication outputs to meet individual preferences across tone, style, format, and level of detail - enabling highly personalized interactions and responses at scale for diverse audiences.

Taken together, these three capabilities could transform how government operates and enable it to deliver services to the public more efficiently and engagingly. As the examples below showcase, there are many angles for immediate civic benefit. How might they be applied to the most pressing challenges facing the public sector in health, education, and more?

Healthcare

Let's start with healthcare, where generative AI tools are dramatically speeding up the pace of pharmaceutical research, making it easier – and cheaper – for companies to identify promising new drugs, test them, and bring them to market.

Pharmaceutical giant Moderna, for instance, is using OpenAI technology to streamline its analysis of clinical trial data, reducing the time spent processing documents and formulating dosage recommendations by 84%. By shortening the time spent on clinical trials, life-saving medications can reach Americans who need them much more quickly.

ChatGPT and other AI-powered tools are also enabling patients to better navigate our complex healthcare system while making it easier and faster for doctors to do their work - for example, by simplifying medical consent forms. These forms tend to be dense and notoriously difficult to understand, and this means patients may end up agreeing to procedures that they may not fully understand. This has been a long-standing problem - in the 1980s, the *New England Journal of Medicine* published research which found that surgical consent forms were written at a college reading level. Decades later, researchers found that the pattern had persisted to the present day. The problem of course is that most people don't read at a college level.

Lifespan, Rhode Island's largest healthcare provider, turned to GPT-4, one of OpenAI's most advanced generative AI tools, to close that gap. To ensure the model's accuracy, Lifespan leadership had legal and medical reviewers closely examine the AI-produced forms before they were put into use. They found that GPT-4 was so accurate that Lifespan's clinicians only had to make a single, small modification: inserting the term "sleep medicine" next to the word "anesthesia" on the consent form. Last fall, Lifespan began using the new forms – which shrunk from three pages to one – across its entire system.

Rewriting a medical consent form may seem like a comparatively small improvement, but the reality is that ensuring patients trust their interactions with our healthcare system is an ongoing challenge that is deeply intertwined with patient outcomes.

AI-powered tools are also improving other aspects of our broader healthcare system, like the notes doctors write after seeing a patient. A start-up called Summer Health has built a new medical visit notes tool which uses GPT-4 to automatically generate a summary of a physician's notes that is written in clear, jargon-free language. Doctors who use the tool say it's reduced the time they spend on note-taking and other administrative tasks from 10 minutes per visit to two minutes per visit, leaving them with more time to spend with those in their care. Patients, meanwhile, say they appreciate receiving AI-generated, physician-reviewed, summaries that are written in clear, accessible language rather than complicated medical terminology.

Education

The primary benefit that generative AI provides both students and teachers is the ability to receive, or provide, the kind of personalized instruction that resource constraints would otherwise not allow. Companies and educational nonprofits have used our technology to create tools that are being used from elementary school through college, and by students and teachers in both urban areas and rural ones. Canva, a leading graphic design company, is making educational templates free to roughly 60 million students and teachers per month, who can use GPT-4 to create interactive lesson plans or instantly translate educational materials into other languages.

Another powerful example is Khanmigo, which functions as a virtual tutor for students needing individualized instruction and as a teacher's assistant for educators looking to create new lesson plans or other classroom materials. It's the brainchild of Sal Khan, the founder of the nonprofit Khan Academy, which offers free online courses in dozens of languages to students and teachers in more than 190 countries.

Last year, OpenAI collaborated with Khan to train our model on the Khan Academy's lesson plans and other data, which helped ChatGPT learn how to better process and respond to questions about math and other academic subjects. Khanmigo can now meet students at their existing levels of knowledge and walk them through questions that get more complex over time. It focuses on teaching students how to solve problems on their own, not on providing them with the answers.

Teachers, meanwhile, are embracing ChatGPT and other AI-powered tools faster than many outside observers had expected. A recent report by the Walton Family Foundation

found that 51% of teachers – including 69% of both Black and Latino educators – are already using ChatGPT. More than 88% of the teachers who took part in the survey said that ChatGPT has had a positive impact on their ability to connect with their students.

ChatGPT and other AI tools are also making it easier to tutor students and provide them with individualized support while enabling teachers to generate more compelling lesson plans and other educational materials. Carnegie Mellon University's PLUS program, for instance, helps math tutors increase the amount of individualized instruction that they can provide to low-income middle school students.

Delivery of Government Services

ChatGPT is also being used to make government itself more efficient while enabling elected officials and other policymakers to better understand and meet the needs of their constituents.

Earlier this year, OpenAI signed an agreement with Pennsylvania Governor Josh Shapiro for a first-of-its-kind pilot program designed to identify specific ways that state government employees can leverage ChatGPT to do their daily work more effectively. As part of the new effort, state employees will use the tool for tasks like making outdated policy language more accessible, reducing the duplication and conflicting guidance contained within hundreds of thousands of pages of employee handbooks and manuals, and helping state employees write and test their own code. The intent is to improve how the government can use AI to better meet the needs of its constituents.

The City of New York, meanwhile, recently teamed with the NYC Department of Small Business Services to launch the MyCity Chatbot, which provides information from 20,000 separate city webpages about how to start and manage a small business. It provides residents with real-time responses to questions about the regulatory requirements associated with their specific small business while also providing direct links to government websites that offer even more information. The chatbot is currently available in 10 non-English languages, and New York officials are planning to expand its multilingual capabilities and the amount of information it can provide.

There are also many examples around the world where these technologies are bringing remarkable efficiency to government services. For example, in Kenya and India, Digital Green is enabling government experts to provide farmers with accurate, timely guidance about where to grow specific crops and how to protect them from drought and disease. Using GPT-4, they were able to lower the cost of providing these vital services from $35 per farmer to $0.35 cents per farmer -- allowing the government experts to help

exponentially more people provide for themselves and their families. 10BedICU, meanwhile, is piloting AI-powered tools in ICU wards in more than 200 government hospitals to help medical professionals treat more patients. The tools include an automated discharge summary generator, a transcription tool trained in local languages, and an ICU nurse's assistant trained on standard ICU protocols.

These examples detail a tiny percentage of what the millions of people who use our tools each day are doing with them. The tools are of course imperfect - they make factual errors, and are simply of limited utility for some disciplines. The models will almost certainly become more accurate and capable in the coming months and years, and develop the ability to use outside tools and pursue complex goals with limited direct supervision. But it would be a mistake to wait to learn how to use and integrate these models until some future capability threshold is reached. The more policymakers interact with these tools now, while they are still relatively safe and limited, the faster they will be able to steer them towards the most beneficial path and accurately pinpoint their key risks before managing these becomes even more complex.

Conclusion

Certainly it will be important to keep in mind that access to the technology remains uneven, particularly in rural and underprivileged communities, and while AI's applications in fields like healthcare and education offer transformative possibilities, expert guidance is necessary to prevent inequitable outcomes for different demographic groups. There's also a risk that innovations might primarily benefit those who already have better access to healthcare and education services. Ensuring that AI-driven solutions reach underserved communities is essential to closing health disparity gaps, and why integration into government service delivery is a critical pillar discussed here.

At the same time, as noted above, generative AI excels at providing on-demand summaries and analyses of complex information from diverse sources, automating multi-source information lookup and data integration tasks based on specific contexts, and tailoring communication outputs with personalized tone, style, format, and detail for diverse audiences at scale. As a result, it can provide immense value to governments and citizens across critical areas like healthcare, education, and service delivery.

From accelerating drug development to enabling personalized tutoring to simplifying regulatory compliance, this technology already offers concrete solutions to many long-standing challenges. These examples represent just the tip of the iceberg in terms of advanced AI's potential to transform governance and enhance public services. As the capabilities of models like GPT-4 continue advancing, the imperative for governments to

systematically adopt and integrate this technology grows more urgent.

In addition to leading on safety mitigations, the U.S. will need a comprehensive national strategy focused on developing robust AI capabilities, fostering AI literacy among the public workforce, and, perhaps most critically, being at the forefront of responsibly leveraging these tools for public benefit. OpenAI stands ready to partner with lawmakers to ensure that artificial intelligence is used safely now and into the future – and that the U.S. remains the global leader in AI.

Understanding and Governing Generative AI

Darío Gil

Senior Vice President and Director of Research, IBM

"Can machines think?"—Alan Turing, 1950

In his seminal 1950 paper,[41] Alan Turing posed "that machines will eventually compete with men in all purely intellectual fields," starting with chess, followed by teaching them "to understand and speak English." Now, more than 70 years later, and thanks to advances in raw computing power combined with large data sets and advances in algorithms, we are seeing machines becoming ever more adept and efficient at learning and speaking English and other languages.

Artificial intelligence and **machine learning** are not new. They have been with us for a long time. In fact, the term artificial intelligence was coined by John McCarthy in 1955, when organizing the famous 1956 Dartmouth Summer conference that gave birth to the field. Artificial intelligence (AI) combines computer science and robust datasets to enable machines to "learn" to do problem-solving in a way that resembles how humans do. AI is already pervasive in our lives. We experience it every day when doing web searches, in recommendation systems like those in Amazon or Netflix, and in transactional payment systems that detect fraud in real time.

An Inflection Point in AI

Now, conveying knowledge into AI models used to be done using annotated data. These are typically datasets that humans label by hand. Building an AI model large enough to work using this process is costly and slow. To make things worse, it was necessary to gather and label data to train one model to perform each specific task. The game changed with the introduction in 2017 of a new method to build an AI model. It uses massive amounts of unlabeled data to train a model by masking and predicting random words in a sentence, predicting the next word or the next sentence, or going through similar tasks. Once the model is trained on an unlabeled dataset and has learned the patterns in the dataset, it can be fine-tuned to perform a wide range of downstream tasks by using a small amount of task-specific labeled training data (ten to a hundred times less labeled data than required for the previous way of training with annotated data). Models built

[41] A. M. Turing (1950) Computing Machinery and Intelligence. Mind 49: 433-460.

this way are called **foundation models**. Foundation models make derivative AI models easier to build, faster to deploy, and capable of performing more tasks. Generative AI is probably their most consequential manifestation.

Generative AI refers to the class of algorithms where a model is trained to generate high-quality text, images, and other content that is similar to the data used to train it. The advent of foundation models, given their remarkable performance and extensibility to a wide range of tasks, and generative AI, is bringing an inflection point in AI.

Main Uses and Impact of Foundation Models and Generative AI

A rapidly increasing user base is actively exploring foundation models and generative AI for applications involving content generation, summarization, classification, etc. The adoption of pre-trained foundation models can unlock exciting use-cases with unprecedented time to value. These range from sentiment analysis, email routing, and text analysis to extracting insights from company documents, generating synthetic data and marketing content, enhancing virtual assistants and customer service, and doing semantic search.

Beyond language, foundation models are equally applicable to other data modalities such as code, making it easier for anyone to write code with AI-generated recommendations. Foundation models can be used to generate images and to better detect anomalies for cybersecurity use cases. They can be trained on time series data for planning analytics, click stream data for customer care, chemistry data for material and drug discovery, and tabular data such as transactions. They can also improve analytics to predict damages due to natural disasters and assist in IT operations to reduce costs. Foundation models trained on sensor data could be used to optimize the maintenance of industrial equipment. This is just scratching the surface of the use cases that foundation models can enable.

It is therefore no surprise that nearly 80% of enterprises are already working with or planning to adopt foundation models and generative AI.[42] According to McKinsey, generative AI could add between $2.6 and $4.4 trillion annually to the economy.[43]

[42] Scale Zeitgeist: AI Readiness Report, a survey of more than 1,600 executives and machine learning practitioners.
[43] McKinsey & Company, Beyond the hype: Capturing the potential of AI and gen AI in technology, media, and telecommunications (2024), https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/beyond-the-hype-capturing-the-potential-of-ai-and-gen-ai-in-tmt#/

Goldman Sachs estimates that it could lead to a 7% raise in global GDP over the next ten years.[44]

Risks and Concerns

As popular as they have become among consumers, there are challenges to scale the adoption of foundation models and generative AI in an enterprise or a government setting. Enterprises need to contend with the protection, control, and monetization of proprietary data and IP; the freedom to adopt the breadth of community-driven innovation emerging every day; the flexibility to deploy models across multiple environments; and the need to mitigate the reputational, regulatory, and ethical risks of AI systems.

There are also concerns about the potential for generative AI to be misused or cause harm in new or unforeseen ways. Some of the risks are the same faced with other kinds of AI, like bias. But generative AI can also pose new risks and amplify existing risks, such as the capability of generating false yet plausible-seeming content. We call this **hallucination**. Other risks for both enterprises and government include adversaries or malicious insiders injecting false, misleading, or incorrect samples; using undesirable outputs from downstream applications for re-training purposes; legal restrictions on moving or using data; copyright and other IP issues with the training data; handling any potential presence of personal identifiable information and sensitive personal information; vulnerabilities to adversarial attacks; the generation of toxic, hateful, abusive, and aggressive content; challenges in explaining why output was generated; challenges in determining the original source and facts of the generated output; documenting data and model details, purpose, potential uses, and harms; and determining ownership of AI generated content, among others.

For enterprises and governments, the trustworthiness of the models and the ability to leverage data effectively and securely are paramount.

With Great Power Comes Great Responsibility: Ensuring the Safety of AI

Trust is the ultimate license to operate. The benefits of AI are moot if we cannot have confidence in the predictions and content generated by the models. It is therefore critical

[44] Goldman Sachs, Generative AI could raise global GDP by 7% (2023), https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html

to build responsibly and put governance into the heart of the AI lifecycle. Here, we discuss some of the AI safety concerns and how they can be addressed.

Copyright and Other IP Issues

One of the many concerns surrounding generative AI is the inadvertent incorporation of proprietary content that may violate copyrights or contract laws. Foundation modes are trained using large and diverse datasets, including publicly available data. The publicly available data may include both copyright-protected materials such as text, audio, and images, and non-copyrighted materials, such as government works or other works in the public domain. Training a foundation model thus implicates the Copyright Act to the extent that reproductions are necessary to creating these datasets. However, these reproductions are not in the final product or made available to the public. They are not stored by the model or retrieved once training is complete. Moreover, the datasets are not used for any discrete expressive content, rather, they are useful to train the model about non-expressive facts and statistical information, such as the relationship between words—the building blocks of language itself. This is not protected by the Copyright Act.

For example, text-based foundation models use a method called "tokenization" to process the words and phrases of raw text data into numerical representations based on the semantic and syntactic structure of text.[45] Tokens can represent units of text, such as words, sub-words, or even individual characters. The process maps tokens to a unique numerical representation known as embedding, which is represented mathematically as a vector. Embeddings represent words as numbers and can be thought of as a dictionary that helps the model understand the meaning of words by placing them in a mathematical space in which similar words are located near each other. During training, the model obtains relationships (e.g., parameters, probabilities) between the various vectors (e.g., words, characters) to create embeddings where similar vectors represent words with similar meanings, together with the relationships between the vectors. This is used, for example, to predict the sequential ordering of vectors and form words that appear in a sentence. In other words, the foundation model learns from the uncopyrightable, factual information about the dataset so that it can make predictions based on future inputs.

In addition, the resulting foundation model itself is not intended to substitute for or compete with any original, creative content in the dataset. Hence, although some of the material in these datasets may be covered by copyright, the use of such material to train foundation models is fair use.

[45] BSA | The Software Alliance, Comments to US Copyright Office regarding Artificial Intelligence and Copyright, https://www.bsa.org/files/policy-filings/10302023uscoaicopyright.pdf (2023)

In cases where a user queries an AI tool and obtains an output that is substantially similar to the original and competes with it, existing copyright law, including the flexible use-specific doctrine of fair use, is sufficient to balance the interests of the rightsholders against those of the AI developer or user. Still, some creators may wish to prevent crawling and scraping of their material. More standard, automated tools can be built to empower creators to express their preferences regarding access to their data by AI training bots. Regulatory enforcement against AI developers who fail to abide by, or intentionally circumvent, creator-implemented automated directives would strike the right balance between protecting rightsholders' preferences while still permitting the development of industry-critical foundation models that rely on web crawling.

IP and copyrights concerns are not limited to the model side, they can also happen when users prompt the models. For instance, a user might unintentionally disclose proprietary data in their prompts. This problem can be addressed by implementing guardrail techniques that can screen input and output text and vet it against proprietary content to identify potential attribution against any designated target content. Depending on the use case, one can use different types of similarity, e.g., exact sentence matching, fuzzy matching, and identification of rephrasing on the same content. For copyright detection, the interest is in identifying the exact reuse of content, but for confidential data, the content needs to be identified however it might have been rephrased. Guardrails can be implemented through software-as-a-service capabilities while keeping the content used for vetting confidential and on-prem, depending on the type of users.

There are a number of protocols and guardrails that can be implemented to address copyright and IP issues, but any new protocol and industry-wide mechanism should be industry-enabled to avoid one-off proprietary arrangements that result in content only being accessible by the largest, best-funded companies.

Personal Identifiable Information and Sensitive Personal Information

The threat of attackers with black-box API (Application Programming Interface)[46] access extracting personal information can be addressed by safeguarding models with guardrails that detect personal identifiable information (PII). They work using a multi-stage process that includes a classifier that looks at the context of the potential

[46] An API or application programming interface is the set of rules or protocols that define how software applications communicate with each other to exchange data, features, and functionality. Developers use APIs to integrate data, services, and capabilities from other applications instead of developing them from scratch. APIs also allow the data and functionality of software applications to be easily and securely shared among parties.

spot. Currently, PII guardrails include key capabilities like detecting malicious intent and context-based compound sensitive PIIs, and cover domain-based PII.

Hallucination

Hallucinations can spread misinformation and lead users to wrong conclusions. This is especially serious in critical fields like medicine and finance. Domain-specific hallucination detection works in situations where technical terms and language particularities are infrequent during the pre-training stage of the AI models. Once the hallucinated content is identified, it can be replaced with the right referenced answer. This is key to improving confidence on the results provided by generative AI.

Toxic, Hateful, Abusive, and Aggressive Content

Among the simplest ways to avoid training AI models with false, hateful, or potentially infringing material is to exclude such material from notorious sources. For example, one could selectively blocklist websites known to disseminate this type of information. Carefully curating domain-specific and internet datasets is the first step to build trustworthy models. Those datasets must be cleansed and filtered for hate, profanity, biased language, and licensing restrictions before using them for training AI models. The AI community continues to develop and refine new methods to improve data quality and controls.

Tracking and managing every step of the process from data acquisition to cleansing, filtering, processing, and training, allows us to react and meet the evolving set of legal and regulatory requirements. Logging and tracking the curated data, the methods used to curate it, and the models that each datapoint has touched enables the identification of affected models and any data that may need to be removed if anything changes in the future.

Governing the Generative AI Lifecycle

AI is essentially a reflection of its underlying data. Training AI models on datasets of unknown quality and provenance can represent legal, regulatory, ethical, and inaccuracy problems. Data provenance and quality matters, as does using data sources lawfully and responsibly. The implementation of guardrails can mitigate potential risks, including detecting copyright infringement, blocking harmful content, recognizing objectionable PII or explicit and implicit hate, etc. AI techniques in reinforcement learning with human feedback offer ways to align the models with human values, reduce hallucinations, and build guardrails. Training a foundation model to a specific domain or industry can also help minimize the scope of risk to which the models can give rise because the model can

be conditioned to generate outputs that are "tuned" or relevant to that domain or industry.

Ultimately, we must build technologies to govern data and models that enable risk detection and business impact assessments, provide control-points to automate and accelerate compliance processes, define and validate AI model-related norms for policy making, support the enforcement of policies, provide tests for verification, and ensure lineage and audit trails. Data insights can play a crucial role in aligning AI models by providing valuable information about the characteristics, biases, and performance of the model.

Proper model evaluation is also central to the viability of AI for government and enterprise uses. This involves human experts that assess the outputs of the model for creativity, coherence, and factual accuracy, and perform custom evaluations for requirements like compliance with business conduct guidelines, multilingual correctness, robustness, safety, code, and others.

What Could Policymakers Do?

It is important to recognize that AI is not intrinsically high-risk. It can be understood and managed and like most technologies, its potential for harm and good depends on how it is used and who uses it. How AI is used is fully within our control. To that end, policymakers should take productive steps to address the concerns surrounding generative AI, recognizing that a risk and context-based approach to AI regulation remains the most effective strategy to minimize the risks of all AI. Each AI application is unique and, consequently, the associated risks vary. While some applications might appear trivial, others, such as medical diagnostics and loan approvals, have profound implications. Regulations should, therefore, be tailored to address the specific risks associated with each AI use-case. The oversight of AI recommending movies is not the same as that of AI making loan decisions or posing risks to an individual personal safety. This is why we should aim for a multi-layered analysis of AI based on use cases and their effects, considering the different applications, levels of adoption, and assessing the different levels of risks. The level of oversight implemented varies depending on those factors. Regulations should also consider the fact that new risks might emerge that may not have been anticipated when general rules are first developed.

Policymakers must also strive for a regulatory framework that fosters, not restrains, innovation. This is key—if businesses feel too constrained, they will be compelled to move their digital technology ventures offshore. Critically, policymakers *should avoid a licensing regime for AI*. This would inadvertently increase costs, hinder innovation,

disadvantage smaller players and open-source developers, quiet diverse voices, and multiply bias. Preventing the type of regulatory capture that stifles open-source innovation is paramount. AI should be built by all, for all, not bestowed on a few of the largest, best-funded companies. This is why any move that would consolidate market control should be avoided.

While governments play an important role, AI creators and AI deployers must shoulder their responsibilities. Simply put, those who create AI systems should be accountable for the AI they use in the context in which it is deployed. AI creators must be held to a standard of openness and transparency on the data, processes, and methods used to create and test their AI. Likewise, those who deploy AI systems should be accountable for the context in which the system is deployed. For instance, companies deploying AI for employment decision-making cannot claim immunity from employment discrimination charges. In the realm of accountability, we must be cautious not to replicate mistakes of the past. Section 230[47] stands as a cautionary tale; we cannot create another shield against legal liability for technology providers in the face of known and preventable unlawful uses of that technology.

Smart policies, coupled with corporate accountability, can address security and societal risks, but it will only work if a broad community is represented in converting principles into robust yet flexible policies. Our AI future—and its extraordinary potential to improve our lives—should be determined by the many, not the few, with responsible and trustworthy AI shaped by diverse, inclusive voices. We must rely on the diversity of our institutions for proper AI governance and allow society to participate in addressing the potential misuse and risks in ways that people trust. AI should be centered in the aspirations of citizens, how they want to shape the future of society, and how and when they want to use the technology. This is achieved by fostering an open and inclusive AI ecosystem. Open innovation is also critical for AI creators to gain the skills they need to ethically deploy AI. There are many actions the government can take to foster an open AI innovation ecosystem. For example, funding the National AI Research Resource Task Force would not only pave the way for AI innovations to be shared more widely but also ensure their secure and ethical deployment.

[47] Section 230 of the Communications Act of 1934, enacted as part of the Communications Decency Act of 1996, grants limited federal immunity to any provider (and user) of an interactive computer service for the actions that occur on their platform, regardless of whether the platform turns a blind eye to illegal activity. Section 230(c)(1) specifies that "no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." Courts have found companies that knowingly host illegal content to be exempt from legal liability based on the broad protection that Section 230 provides.

The AI Alliance

The belief that the future of AI is too important to be decided behind closed doors by a small group of powerful companies, the importance of representing the broad community, and the acknowledgement that the fundamental advances that have made today's generative AI possible have been achieved by a diverse scientific community, led to the launch of the AI Alliance[48] in December 2023. It is an international coalition of innovators, represented by over 70 diverse institutions, that includes the science agencies that support curiosity-driven exploratory work, the universities that educate generation after generation of computer scientists and AI experts, the industrial research laboratories that create breakthrough demonstrations of AI systems, and the wide array of enterprises, from nonprofits to startups to established multinationals, that commercialize and scale AI products and services. These organizations are innovating across all aspects of AI education, research, technology, applications, and governance.

The AI alliance was designed to better reflect the needs and complexity of our societies in a more open, transparent model for innovation. Guided by the core principles of open science and open innovation, it brings together a critical mass of computing, data, tools, and talent to accelerate advances in AI. It builds and supports open technologies across software, models, and tools; enables students, developers, and scientists to understand, experiment, and adopt open technologies; and advocates for the value of open innovation with organizational and societal leaders, policy and regulatory bodies, and the public. Besides the AI Alliance, we have seen initiatives like the World Economic Forum's AI Governance Alliance and the EU's European AI Alliance, which recognize the importance of an open ecosystem for responsible innovation.

The AI Alliance is a catalyst for driving an AI agenda underpinned by some of society's most fundamentally important principles: scientific rigor, trust, ethics, resiliency, and responsibility. It offers the opportunity for the community to define together the evolution of AI. This is a technology that is destined to play an increasingly prominent role in everyone's lives, redefining the ways we work, play, learn, communicate, and more. Because of that, every single one of us has a vested interest in its future.

We are approaching a fork in the road. One path is dangerously close to creating consolidated control of AI, driven by a handful of companies that have a closed, proprietary vision for the AI industry. It is not hard to imagine the stifled innovation, hoarded benefits, and questionable oversight lurking in this path. The other path is a broad and open road, a highway that belongs to the many, not the few, and is protected by the guardrails that we create together. The choice is ours.

[48] https://thealliance.ai/

Conclusion

AI's transformative potential is undeniable. With the capacity to revolutionize industries, augment human productivity, and tackle the world's most pressing challenges, it stands as one of the most significant technological advancements of our era. But as with every powerful technology, the benefits do not come without risks. Fortunately, there are a variety of actions to take to ensure that AI is a force for good.

The safety of AI can be increased with better tools and standards for data provenance, quality, and pre-processing; application-level benchmarks; and tools to monitor and improve AI models. AI models can be aligned with ethical and safety practices, and methods to find and fix biases before the models are tuned and deployed can be implemented and improved. It is also important to understand the true “intelligent” and generative capabilities of AI, its velocity of development, and the resilience of guardrails to develop the proper focus and agility to respond to advances. Understanding existing laws, regulatory frameworks, and the diversity of safety mandates of our institutions will help determine what can be leveraged and potential challenges when implementing policies.

Knowing the technology and being able to verify safety as opposed to trusting black box APIs, and promoting greater scientific understanding of the challenges is critical to informing smart, effective policymaking. Policies should target how and where AI is used, not its underlying code. For example, we do not regulate wheels. We regulate cars, trains, and the landing gear on airplanes because that is where the risk occurs. AI should be regulated in the same way and AI creators and deployers should be held accountable for the technology they unleash in the world. Liability is more effective than licensing and avoids further entrenching the market position of a handful of players.

Ultimately, AI safety will only work if a broad community is represented and made firsthand participant in its policies and governance. We have seen the power of collaborative, open innovation driving technological shifts from industrialization to the internet. AI will be no different. We must be careful not to relinquish our AI future to the hands of a few, to the detriment of all.

Generative AI – Move Over Language Models and Make Way for Industry

Mike Haley
Senior Vice President of Research, Autodesk

Could robots evolve into a new artificial species? Hans Moravec, professor of robotics and AI at Carnegie Mellon University, explored this question in his 1998 book, *Mind Children*. Still, his work is best remembered for "Moravec's Paradox," the observation that it is much easier to create technology that solves a complex reasoning task—like doing your taxes—than it is to build a machine that solves the basic perception and movement problems of a young child.

Researchers struggled to train machines to complete tasks like folding towels. Hackaday, 2016, [49]

One explanation for Moravec's Paradox is that our brains have had millions of years to develop sensory and motor skills, whereas abstract thought is a relatively new capability that evolved only about 100,000 years ago. Another explanation could be that the material world, governed by an infinite number of constantly interacting physical processes, is much more difficult to reason about than an abstract but finite world like

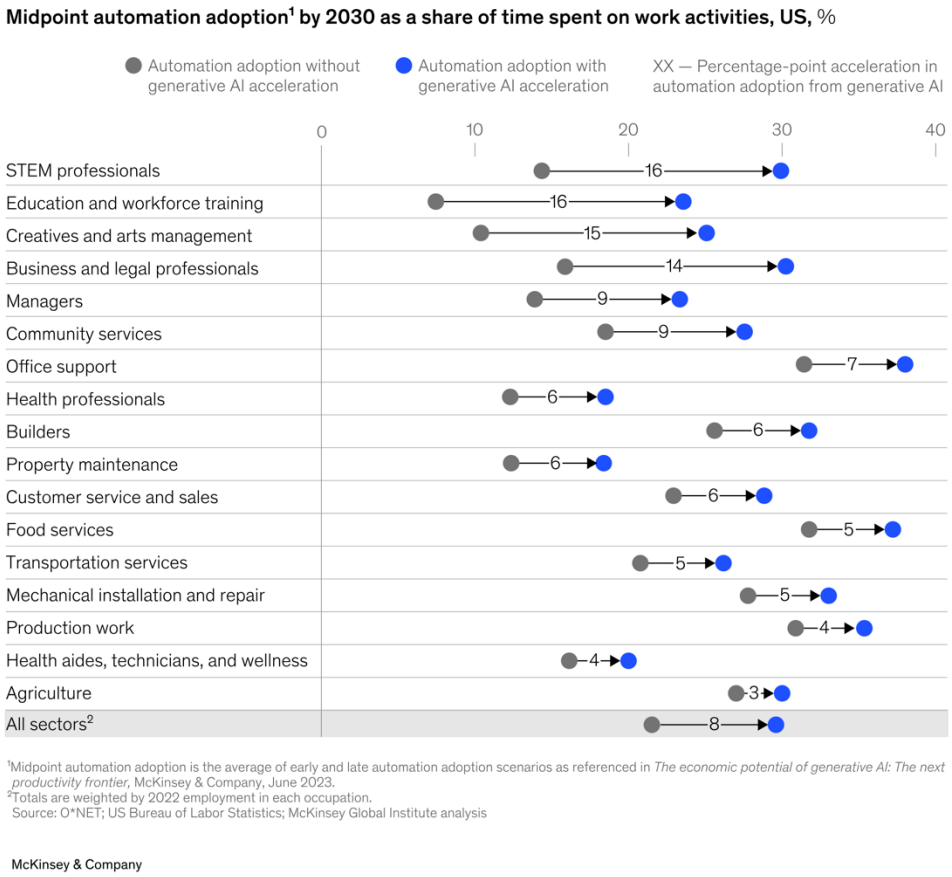
[49] https://hackaday.com/2016/02/24/the-challenges-of-a-laundry-folding-robot/

the U.S. tax code. A single rivet in an airplane is a minute part, but it might fail for countless reasons, including environmental factors, materials capabilities, manufacturing processes, and the forces and behaviors of the rest of the aircraft. In the face of the physical world's incredible complexity, we often can make only an educated guess.

Where Will AI's Impact Be Felt First: Blue-Collar or White-Collar?

When I first became involved in AI and robotics, I believed these technologies would first impact blue-collar jobs. AI would automate factories. It would drive construction sites with robots. It would take over menial jobs of all sorts, ultimately relegating humans to supervision and evaluation roles where abstract reasoning played a greater part.

The last several years of AI development have flipped this narrative for me. I still believe that AI will affect physical work, but I expect it will take much longer than I originally thought. Massive disruption to white-collar work that is fundamentally logical, symbolic, or diagrammatic in nature will outpace disruptions to blue-collar work. In some industries this is already happening. Generative AI technologies are rapidly upending professionals such as paralegals, copyeditors, and accountants, and are poised to disrupt many other white-collar professions.



McKinsey Global Institute, 2023, Generative AI and the future of work in America[50]

These fields completely digitized their work already, so all their data is potentially available to train AI systems. Almost all legal documents are now stored electronically; we all write using word processors; and accounting and other financial information is now routinely kept in a database and processed in a spreadsheet. Time-to-market is fastest where digital data is most available, and as Moravec's Paradox tells us, AI's efficiency gains are most dramatic where the targeted work is the most abstract.

In 2023, we saw the rapid adoption of Generative AI to create language, documents, images, and even video. These tools arrived first because most of the past decade's AI research and development was based on massive datasets of language, imagery, and video available across the Internet. But not all digital data is available on the open Web, so which tools will come next?

[50]https://www.mckinsey.com/mgi/our-research/generative-ai-and-the-future-of-work-in-america

Large, Industry-Specific AI Models Are Coming

The huge advances we have seen in large language models are due to extremely large deep-neural networks' ability to understand statistical patterns in large amounts of digital data. The more consistent that data is, the better the trained models can learn, reason about, and generate similar content. Human language is based on learnable alphabets, grammatical structures, and vocabularies. These elements of language appear in many scientific and industrial fields as well—think of chemical compound exploration or bioinformatics.

The industries where we design and make the built world—manufacturing, architecture, and construction—also have well-defined and consistent digital data. They include everything from consumer electronics to airplanes, from single-family homes to skyscrapers, from roads and railways to city and utility infrastructure. Our research lab at Autodesk has focused on AI using this type of data for the last decade. What would the implications be for how we design and create the world if we enhance them with the appropriate AI?

Certainly, those processes could be greatly improved. For example, a typical building involves at least four complete redesigns before construction.

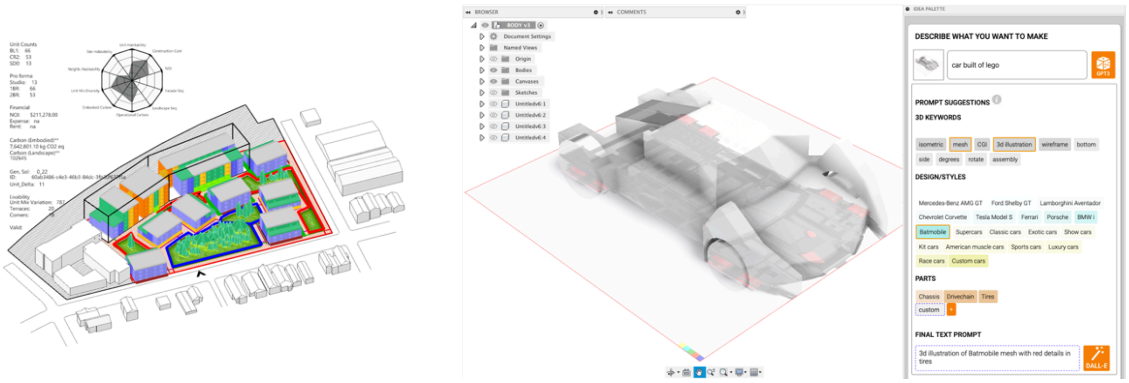
Consider how buildings are built. An architecture firm likely begins with conceptual designs and finishes with a detailed blueprint of the intended building and related systems. Next a structural engineer designs the framing, structures, and concrete slabs that constitute the building's supporting skeleton. The structural engineer is legally liable for their work so will reference the architect's drawings but often insist on creating their own version of the building design as a means of maintaining control of their output. Plumbing, HVAC, and electrical designers likely create their own designs for the same reasons. Finally, the general contractor that constructs the building likely produces its own drawings and designs. Unnecessary duplication, complicated coordination, and the inefficient use of time and money inevitably results from all these redesigns. Advances in digitalization, collaboration, and automation have delivered incremental improvements for over 30 years, but we have yet to fix this fundamentally inefficient process.

The Nature of the Disruption

I believe AI is the solution we have been waiting for, and that it will lead to a long overdue disruption in manufacturing, architecture, and construction. It will apply first to the most abstract and digitalized work—design information—before moving gradually

downstream into the physical-world processes in which things get made. Along the way it will catalyze changes in these professions and open the world of advanced design and engineering tools to more people.

Complex computer-aided design (CAD) and engineering software often acts as a barrier to creative people frustrated by expressing every nuance of an idea using a keyboard, a mouse, and a complex software user interface. Generative AI will revolutionize interaction with design software, helping creative people express themselves naturally using language, sketching, examples, and images, just as we might explain ideas to another human. For experienced engineers or designers, this means easier work, and they can focus more on creative problems. For those just entering the workforce, it means getting up to speed and being productive much faster. Together, it means a renaissance in creativity powered by the next generation of highly accessible tools.

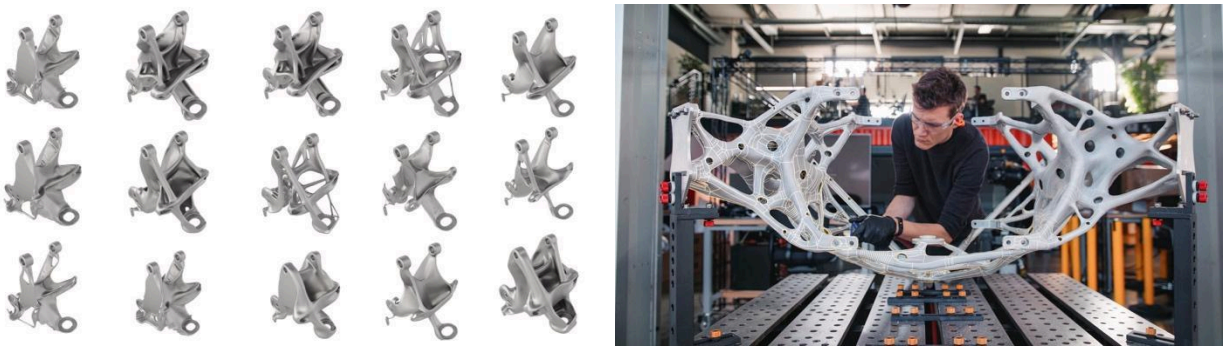


Future architects and automotive designers will employ AI to test early concepts and create prototypes. Autodesk Research, 2024.

AI will also change how designers and engineers spend their brainpower! Today, design work mostly consists of either repetitious, tedious work (approximately 80% in most industries) or deeply technical and complex work, with an engineer or designer working under tight time constraints. The good news is that AI can learn the well-defined patterns of repetitious work in order to reproduce it. This will result in automating laborious tasks like creating blueprints, finding the right engine components, or checking the building code compliance of an architectural design.

For complex and technical creative work, AI will become an assistant. Microsoft's GitHub CoPilot has shown how valuable such an assistant can be to software programmers, in some cases accelerating software delivery tenfold. CoPilot's output is not always correct, so programmers still need to validate and often tweak its suggestions, but even experienced developers feel their abilities powerfully amplified by this tool. Applied to design and engineering, AI can explore solutions that incorporate millions of variables, something far beyond human brain capacity. These systems will

provide engineers design options and starting points that they would never develop by themselves. This melding of creative people with extensive computational capabilities and AI will deliver major breakthroughs to complex challenges in transportation, affordable housing, waste reduction, and renewable energy.



Researchers at General Motors[51] relied on computers to explore design options for an improved seat bracket while NASA's Jet Propulsion Laboratory[52] used the technology to optimize lunar landers for strength and weight.

Manufacturing and Construction Must Prepare for Change

Why do I believe that the current trends in AI will have these effects on design for manufacturing and construction? I described how large deep-learning models can recognize and reason about the patterns inherent in consistently structured digital information. In the last decade we have also seen the rise of *multi-modal* AI models that have learned across different types of data (*e.g.,* text *and* imagery) to do things like identify language patterns that relate to specific imagery or video, resulting in image and video synthesis tools such as *Midjourney* or *DALL-E*.

Over the past decade, Autodesk Research has applied these same techniques to 3D models, including buildings, vehicles, and film and video game characters. Each of these has consistent patterns, grammar, and structures. For example, the way a building is designed is relatively consistent, involving structural framing, slabs, walls, windows, cladding, and more. Similarly, multi-modal AI models can be trained to associate language and other forms of natural expression—like sketching—to direct the synthesis of 3D models.

Architects and engineers encode everything within these 3D models, including materials, components, and the ways in which components, products, and buildings will be manufactured and constructed. By training AI systems on these 3D models, we are

[51] https://www.autodesk.com/customer-stories/general-motors-generative-design
[52] https://www.autodesk.com/customer-stories/jpl-interplanetary-lander-video

fundamentally transforming the entire design-to-creation process. Once in place, these AI models will have learned a lot about how things in the real-world are designed and made. Consequently, they will begin to feel like having a team of supporters providing:

- suggestions and alternatives during design
- constant analysis and validation
- autocompletion of repetitive design work
- automated documentation
- even preparing factory machines to make what was designed.

Obstacles to Progress

The future is bright for these technologies, but formidable challenges—technical, economic, and regulatory—remain. One key challenge lies in the probabilistic nature of deep-learning AI models. Each prediction or output of an AI model has inherent variability resulting in what's often termed "hallucinations." When Generative AI produces a piece of creative writing or an imaginary painting, this variability is often desirable, even charming. However, Generative AI that predicts the framing structures of a bridge or a specific ratio of chemical compounds must be precise. Nobody wants to drive across a bridge that was automatically designed to be *approximately* correct. Considerable effort is underway to discover methods for controlling, validating, and adapting Generative AI models to produce precise and controllable outputs under specific circumstances.

A second challenge lies in the sources and quantities of data required to train large AI foundation models. Earlier I noted how the rapid advances in AI research over the last decade have been powered by large datasets sourced from the open Internet. However, as we seek to train AI models based on data lying within corporate firewalls, aggregating equivalently sized datasets becomes much more difficult. No single company, in any industry, has sufficient data to train a large AI foundation model. It will be necessary to create incentives, areas of common interest, and governing structures across industries to achieve this. One example might lie in construction, where many companies have decades worth of health and safety data. Improving health and safety is a "raise all boats" issue that benefits everyone and where industry players already have a history of collaboration. So, it might be an attractive starting point for construction firms to pool their health and safety data to train a large AI foundation model that could dramatically reduce safety risks on construction sites.

Finally, regulating business practices, AI usage, and personal data usage will be necessary to avoid a "race-to-the-bottom" of manipulative business practices that

ultimately undermines professionals' trust in these technologies. AI is a computing tool, the power of which will dwarf innovations like the Internet and perhaps even computers themselves. With this power comes infinite opportunities for good but also opportunities for misuse and manipulation. Industry requires sufficient, modern, and constantly evolving regulation to unlock the power of AI for our economy while also ensuring the safety of its users and the integrity of the institutions and companies delivering AI tools.

Moravec's Paradox continues to hold true, but its edges are eroding. Computers can now see us, reason in our language, and interpret our sketches and drawings. What's more, computers can equivalently communicate back to us. Starting with the areas of work that are the most digitalized, we will see huge benefits in industries like manufacturing and construction, where better, safer, more efficient vehicles, buildings, infrastructure, and more can be planned and designed.

That said, we would be foolish to dismiss the risks that lie ahead. Rather, we need to shine a light on them, invest in resolving or governing them, and ensure that any regulation is sufficiently nimble to evolve with the breakneck pace of AI. Today it's difficult to find a single computer science department or software company not working on AI. Governments, universities, and industry are pouring billions of dollars and incredible talent into this technology. But to fully realize its potential requires transparency and coordination across these institutions. With that, we can all look forward to driving across AI-designed bridges with peace of mind.

EXPERTS' RECOMMENDATIONS

The Impact of Generative AI in a Global Election Year[53]

Valerie Wirtschafter, The Brookings Institution

Executive Summary

The influence of the online ecosystem in shaping democratic discourse is well-documented, with the expanded reach of generative artificial intelligence (AI) representing a novel challenge in a historic election year. Generative AI enables the creation of realistic images, videos, audio, or text based on user-provided prompts. Given the potential exploitation of this technology, particularly in the context of elections, it has garnered significant attention.

The transformative impact of generative AI on the information space has not matched these initial expectations. However, instances of manipulated or wholly generated content have surfaced, posing a threat to democratic discourse and electoral integrity. Addressing this challenge requires a multifaceted response.

Interventions ranging from legislative measures targeting election-specific deepfakes to voter education initiatives are imperative. Tech companies should also play a central role, including through the implementation of imperfect technical solutions to identify the origins of generated media. While these interventions may not eliminate the challenges posed by generative AI, they represent progress toward managing a complex issue during a critical election year.

Introduction

In 2024, a record number of countries will hold elections. Collectively, they are home to more than 41 percent of the world's population and 42 percent of global GDP. Much like past elections, the online ecosystem will play a role in shaping the contours of these campaigns, but new developments have strained an already contested information space. One of these developments is the rapid advance of generative artificial intelligence (AI), which allows anyone to conjure up realistic images, video, audio, or text based on user-provided prompts or questions.

[53] This research was published by the Brookings Institution on January 30, 2024

Generative AI outputs have been improving steadily for nearly a decade. However, following the viral launch of ChatGPT in November 2022, a significant amount of commentary focused on the potential for this type of content to create a "disinformation nightmare" in 2024 by accelerating the production of false information. A year after this initial frenzy, generative AI has yet to alter the information landscape as much as initially anticipated. However, even at a smaller scale, wholly generated or significantly altered content can still be—and has already been—used to undermine democratic discourse and electoral integrity in a variety of ways. Specifically, generated content tied to elections can:

- Shape last-minute attempts to deter voters from exercising their right to vote or manufacture an event featuring a generated depiction of a candidate that is difficult to debunk.
- Lead authentic information to be cast as false or generated to avoid uncomfortable questions around accountability, particularly in the face of true scandals that could impact political campaigns.
- Speed up, improve, and reduce the cost of existing information operations designed to manufacture the perception of consensus around political issues, undermine government responsiveness, sway public opinion, exacerbate divisions, demobilize or deceive voters, and undermine trust in electoral processes.

As democratic countries consider how to respond, they should evaluate a wide variety of interventions, from new or updated legislation targeted to election-specific concerns, such as the dissemination of deepfake content depicting candidates running for office, to voter education efforts aimed at teaching citizens how to scrutinize generated content. Tech companies also should play a central role by implementing imperfect but important technical solutions around content provenance and watermarking, and investing in new tools for detection. They can also facilitate knowledge-sharing across platforms from which citizens obtain their online information. Collectively, these interventions are unlikely to wholly address the challenge generative AI poses to information integrity. Yet, they are positive steps toward making a seemingly intractable challenge more manageable in a historic election year.

What Is Generative AI?

Generative AI is a class of artificial intelligence that takes an input—provided by a user—runs it through a pre-trained model and returns a set of expected generated outputs. Advanced generative AI leverages deep learning—an area of the broader machine learning field that uses multi-layered neural networks to generate more and

more complex associations between patterns—to create images, text, videos, audio, or other content.

Generative AI relies on several foundation models within deep learning, such as generative adversarial networks and transformers, to process large amounts of data to "learn" the representation of a specific output for which the models were trained. ChatGPT, for example, was fine-tuned as a chatbot, while Copilot was fine-tuned to generate code, and DALL-E2 was trained to output images. The goal is to output new content that resembles the patterns learned from the training data.

Advances in deep learning, larger datasets, improved computing power, and increased investments have led to rapid improvements in output quality in just a few years. This culminated in the public release of chatbots such as ChatGPT and image generators such as Midjourney or DALL·E2 in 2022, which were hailed for the quality of their outputs. ChatGPT alone had more than 100 million users two months after it launched in November 2022.

All Hype or an Emerging Threat to the Information Space?

Since ChatGPT's viral launch, much commentary has focused on the potential for generated content to upend democratic elections by turbocharging the production of fabricated information. Over the past year, nearly 30,000 news articles indexed on Google News have focused on how generative AI tools might impact upcoming elections.3 Despite this interest, generated content has only been distributed sporadically online, even at times when the demand for credible information has been high and the supply of it low. Instead, recycled or decontextualized videos and images have filled this void.

In the absence of sweeping information campaigns using generated content, some have dismissed potential transformative impact of generative AI altogether. It is difficult to assess how frequently generated content is shared across social media to spread misleading information due to the diversity of channels for distributing content, lack of researcher access to data, and the increasing challenge of identifying wholly generated content. Yet there are ways to glimpse its prominence in broader conversations around information integrity, even if this type of assessment is far from comprehensive.

For example, X allows eligible users to add clarifying information to misleading posts in the form of a "Community Note." Notes that are rated helpful are displayed alongside the post as additional context. Critically, data from the Community Notes program is made publicly available, which includes the text of the additional context users have

suggested for flagged posts. As a result, we can examine the frequency with which generative AI and other related terms are referenced in this suggested context to better assess the reach of deceptive AI-generated content across X.

Drawing on this data, I find that since the launch of ChatGPT the number of Community Notes mentioning AI-related terms has grown over time (Figure 1, Total Notes). However, notes referencing these terms still only make up a little over 1 percent of the more than 300,000 notes written over the past year (Figure 1, % of All Notes).

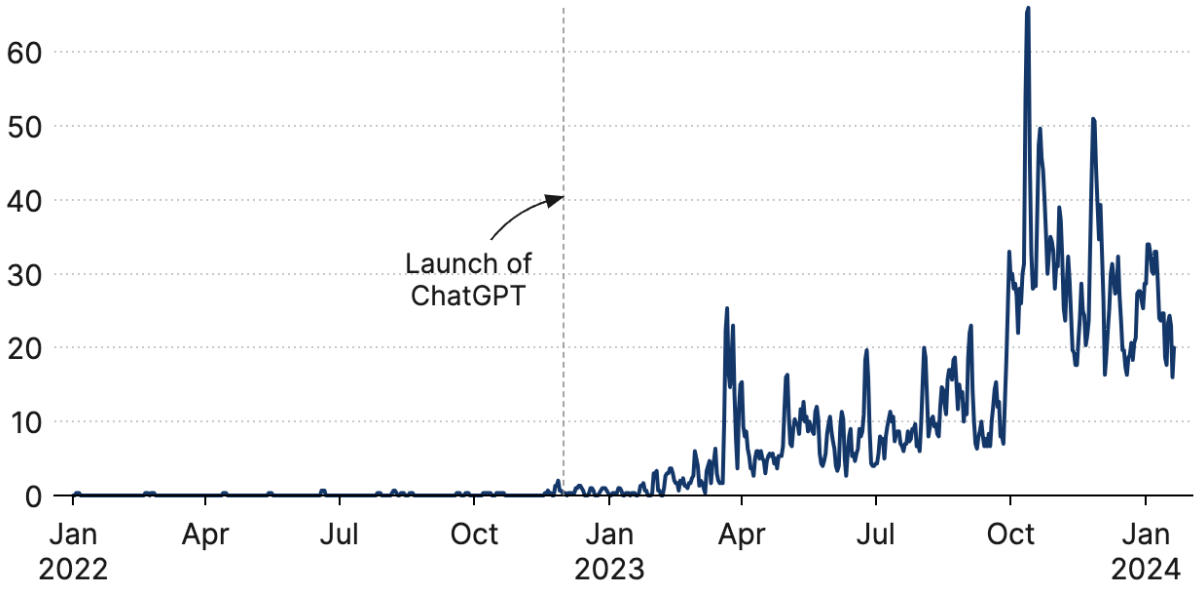
FIGURE 1

The number of Community Notes per day that mention AI-related terms has grown over time, but the topic still features in a small percentage of all written Notes

Click on the buttons below to see data for different measures:

Total Notes | % of All Notes

Total Notes



Source: Community Notes

BROOKINGS

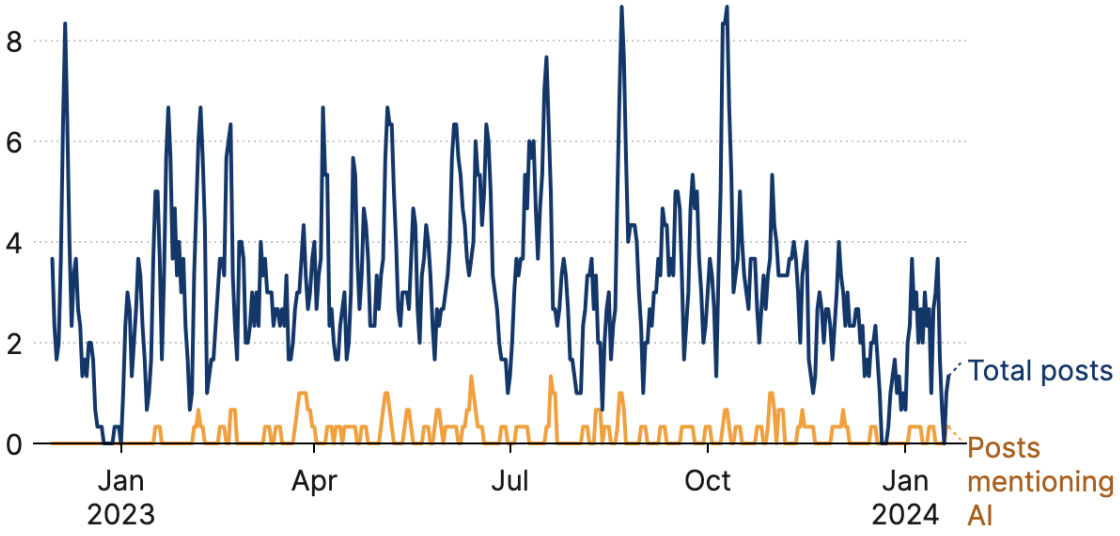
Note: AI-related terms include deep-fake, chat gpt, chatgpt, chatGPT, midjourney, dalle, deepfake, deep fake, ChatGPT, AI, artificial intelligence, inteligencia artificial, IA, generative, generated, Midjourney, and DallE. Data downloaded on January 24, 2024. All data represent 3-day rolling averages.

Another way to assess the extent of this challenge is to look at the number of claims evaluated by fact checkers that reference generated content. Drawing on approximately 1,300 claims fact checked as false by U.S. based fact-checking organization PolitiFact since the launch of ChatGPT, I find that 6 percent of fact-checked content references a term tied to AI-generation in their assessment of the claim. The number of posts referencing these terms has become a steady fixture over time (Figure 2), but they still represent just a small percentage of the total claims reviewed.

FIGURE 2

The total number of PolitiFact posts fact-checked as false that mention AI-related terms has remained consistent since the launch of ChatGPT

They still represent a small fraction of the total claims reviewed by PolitiFact



Source: PolitiFact

BROOKINGS

Note: AI-related terms include deep-fake, chat gpt, chatgpt, chatGPT, midjourney, dalle, deepfake, deep fake, ChatGPT, AI, artificial intelligence, inteligencia artificial, IA, generative, Midjourney, and DallE. Data downloaded on January 24, 2024. Only claims rated "Pants on Fire" and "False" are included. All values represent the 3-day rolling average.

It is important to emphasize that neither of these cases cover all the content that circulates online. As a result, it is likely that due to capacity constraints these figures represent an undercount of the type of generated content circulating on X and across the

various spaces from which PolitiFact identifies claims. Yet they also do not demonstrate an overwhelming flood of generated content across the information space. Instead, generated images, text, videos, and audio seem to complement existing, already prominent ways for disseminating false claims, which may also leverage recycled images or video.

How Generative AI Content Has Already Undermined Democratic Discourse

Although generated content still makes up a small fraction of the overall contested information space, its usage will likely become more common, and it has already begun to undermine democratic discourse around elections. Two recent cases illustrate the unique damage even a small amount of generated content can have on the information space in the overarching context of elections.

Slovakia's Pre-Election Deepfake

In September 2023, generative AI-based political interference upended Slovakia's parliamentary elections. Two days before voters cast their ballots in an election with implications for the trajectory of Slovakia's military assistance to Ukraine and support for NATO, an audio clip which bore the markings of generated content spread widely across social media. This audio allegedly featured the voices of Michal Šimečka, leader of the pro-NATO Progressive Slovakia party, and a journalist from the daily newspaper *Denník N*. discussing ways to manipulate the election and buy votes from the country's minority Roma population. Although the audio seemed suspicious from the outset, it was shared by thousands on social media, including by a former member of parliament from Slovakia's opposition party.

Fact checkers quickly cast doubt on the authenticity of the recording due to incongruencies in the audio sound, awkward word choices, and suspicious phrase cadencing, among other anomalies. However, Slovakia's 48-hour pre-election moratorium period, during which media outlets and politicians are compelled to stay silent and avoid election-related announcements, hindered the extensive dissemination of corrective information.

The generated audio also capitalized on a flaw in Meta's manipulated-media policy, which explicitly addresses only wholly faked videos and not audio content. Although fact checkers were eventually able to attach a label to the post across Meta platforms, the generated audio still circulated widely in a fragmented information space where content moderation practices vary widely.

This type of last-minute scandal is particularly challenging in countries such as Slovakia, where media blackouts limit the press from discussing campaign-related content in the lead up to an election. These blackouts are typically around 24 hours but can last as long as three days in some countries. As a result, they can pose clear challenges to the debunking of viral generated content. Absent a change in election laws, social media companies must clearly delineate and enforce content moderation policies, with particular attention paid to addressing loopholes in manipulated media policies.

Argentina's "Melcogate"

AI-generated content also played an unexpected role in Argentina's 2023 presidential elections. A few days before the first round of voting, scandalous audio recordings began to circulate widely online. The audio recordings allegedly featured Carlos Melconian, then presidential candidate Patricia Bullrich's pick for economy minister, speaking crudely about women and offering government positions in exchange for sexual favors.

In the aftermath of the incident, known as "Melcogate," Bullrich and her party swiftly came to Melconian's defense and dismissed the recordings as fabricated and potentially altered or generated using "voices and artificial intelligence." She also questioned the source of the audio clips and attacked the journalist who shared the leaked audio, accusing him of unethical behavior in the past and of using the doctored audio as a part of a pro-incumbent smear campaign. Initially, Melconian chose to remain silent about the audio leaks. However, in subsequent interviews he did not explicitly deny the authenticity of the recordings, stating instead, "Even if it were me, what does this say? Nothing."

It has yet to be established whether the audio clips were indeed an AI-generated deepfake. However, the incident highlights the unexpected ways that even the potential for something to be AI generated can shape the contours of an electoral contest. Moving forward, politicians will be able to reasonably dismiss true scandals as fabrications – known as "the liar's dividend" – due to the mere possibility of credible deepfakes and other generated content. This is already happening in the United States as well, where true, old clips of former President Donald Trump have been rebranded as AI-generated. And there is some evidence that this strategy works for politicians. A recent survey experiment found that casting true scandals as "misinformation" makes voters more likely to support the implicated politician. As a result, generative AI might have its most pernicious impact in spaces where it is, in fact, not used at all. This makes it all the more critical for researchers to be equipped with the tools required to better understand the scope of the challenge, in order to avoid feeding into the hype that allows the "liar's dividend" to find fertile ground.

Where Generative AI Content Could Influence Upcoming Elections

Beyond these examples, there are several avenues where generated content could make an already complex information space even more complicated. These efforts are not new: they have been the focus of election-related disinformation campaigns for some time. However, generative AI content has the potential to turbocharge campaigns designed to undermine democratic discourse by making content higher quality, more substantively distinct, and easier to mass produce than past information campaigns launched both domestically and as part of foreign influence operations.

In these contexts, generative AI content can act more as an amplifier for the spread of disinformation. Previously, these efforts required coordination between multiple actors—or even an entire troll farm—and were somewhat discoverable due to their use of recycled photos or grammatically incorrect or repetitive messaging. Now, it is possible to create large volumes of distinct content, devoid of many of these prior errors, with just a few clicks of a button.

TABLE 1

Harms to democratic processes that could be amplified with generative AI content

Output	Type of generated content
Manufacturing the perception of consensus around political issues	
Inorganic social media posts about issues or topics	Text, Image, Video, Audio
Inorganic social media comments and engagement to amplify or algorithmically boost issues or topics	Text
Contacting government officials through constituent channels	Text
Undermining government responsiveness to voters	
Spamming mass records requests	Text
Swaying public opinion and exacerbating divisions	
Social media posts and profiles designed to shape perceptions or manufacture consensus around issues and inflame divisions	Text, Image, Video
New articles to seed specific perspectives about an issue and crowd out search results	Text, Image, Video
Robocalls from candidates or other influential figures	Audio
Leaked recordings purporting malfeasance	Audio, Video
Demobilizing or deceiving voters	
Robocalls designed to demobilize or deceive voters	Audio
Social media posts or media articles designed to demobilize or deceive voters	Text
Fabricated evidence designed to demobilize or deceive voters	Image, Video
Undermining trust in electoral processes	
Leaked recordings about efforts to rig elections or evidence of election rigging	Audio, Video
Fabricated evidence of efforts to rig elections or evidence of election rigging	Image, Video, Audio

Table 1 provides an overview of different ways that generative AI could amplify or exacerbate existing threats to democratic processes. These threats include: (1) manufacturing the perception of consensus around political issues; (2) undermining government responsiveness to voters; (3) swaying public opinion and exacerbating divisions; (4) demobilizing or deceiving voters; and (5) undermining trust in electoral process.

Different types of content, from robocalls to social media posts, can and will certainly continue to circulate in the absence of generated content. But generated content may make the production of convincing outputs at scale less costly, more credible to voters around the world, and more challenging to identify and debunk.

For example, deepfakes and voice cloning have already been used to imitate candidates running for office. In one incident, an AI-generated robocall purporting to be U.S. President Joe Biden sought to discourage Democrats from voting ahead of the New Hampshire primary in the United States. Moving forward, such tactics could not only be used to more convincingly demobilize or deceive voters, but also to sway public opinion and exacerbate political divisions. Much like the Biden robocall, these efforts might be at least somewhat discoverable at the national level, but they will likely be harder to detect in state, municipal, and other local races, where resources and attention are limited.

Malicious actors also could use generated images to make influence operations and coordinated inauthentic behavior run through fake accounts more convincing. Where once these profiles relied on recycled images lifted from unsuspecting social media users, wholesale personae can now just as easily be created to make these campaigns appear more credible. Automated processes could also help to scale these fake personas in parallel more rapidly than before.

Finally, large amounts of distinct text shared on social media or through a proxy website could be used to manufacture the perception of consensus, or sow alternative narratives without some of the telltale grammatical errors and misused jargon prominent in influence operations of the past. This type of content could fill gaps online for quality information about election-related topics in specific languages spoken by minority populations and overwhelm search results that at least in part rely on the freshness of content when algorithmically ranking results. Text outputs could also be used to fabricate more credible records requests from government officials by producing slightly different outputs that make it difficult to streamline tasks. For already stretched bureaucrats and election officials, this time-intensive work could make an already challenging space even more complicated.

Other Threats to the Online Ecosystem

Despite these clear challenges, the presence—or absence—of generative AI outputs is not in itself enough to disrupt democratic processes in an election year. The social media platforms where voters seek out information, the algorithms that govern the type of information shared, and the automated and manual review processes that scrutinize content moderation practices play an important role in shaping what voters see.

In the past year, the information space has fragmented, pushing users further and further into ideological echo chambers, with varying degrees of attention to content moderation. In some cases, discomfort with making content moderation decisions has led platforms to lean more heavily on crowdsourced solutions. Although these "wisdom of crowds" approaches can be effective, they should not be considered adequate solutions for the problems in this space, particularly given the intrinsic difficulty of detecting AI-generated content already. This challenge will likely become even more acute as AI systems continue to evolve and produce more convincing outputs.

At the same time, it has also become more difficult for researchers to access data required to explore the evolving nature of information operations in the AI era. In some cases, public APIs where researchers can collect data do not exist. In others, data access has also been severely limited by tech companies. This lack of data access limits researchers' ability to understand the effectiveness of information campaigns, whether they are reaching their intended audience, and what role AI-generated content is playing in making them appear more credible. Without this knowledge, it is difficult to both understand the scope of the challenge and develop evidence-based responses to counteract their influence.

Strategies for Defending the Information Space in an Election Year

Addressing the challenges posed by AI-generated content will require coordination across a wide range of actors, from governments to AI companies and social media platforms, as well as users. Interventions targeted toward output development, distribution, and detection will help to mitigate some of the problems generative AI poses to overall information integrity during elections. While these measures are unlikely to resolve the issues, they are positive steps in addressing a seemingly intractable challenge during a pivotal election year.

Development

Tech companies that develop AI tools are already working on strategies to better signal when an output is generated during the development process. Imperfect technological solutions include watermarking, which adds a pattern to generated content to signal that it was generated, and content provenance, which provides a layer of information, akin to a nutrition label, to help signal when an image or video was created with an AI tool, and where and how it has been subsequently edited. Watermarking for text outputs has also shown some promise. The challenge with this approach is that the outputs of highly capable models that do not opt in to watermarking or content provenance requirements might be mistaken as human generated. Additionally, while this type of metadata tagging might be helpful if it is somehow uniformly implemented across the tech industry, screenshots or phone recordings of images or videos can also remove this information, and watermarkings can easily be broken. To address challenges posed by generated content in the development phase, tech companies and legislators should consider:

- ***Widespread implementation of current technical solutions, and continued investment in more sophisticated approaches:*** Despite the limitations of technical solutions, tech companies should rapidly deploy these tools as a first line of defense against generated content tied to elections. However, while these tools are necessary, they are by no means sufficient, with mixed performance on a range of technical and policy considerations.6 As a result, continued investment in efforts to improve information about content provenance and watermarking across the industry, as well as the development of new, better solutions, will be vital to identifying generated content.

- ***Legislation designed to limit or build in further accountability for generated content depicting candidates actively running for office:*** The spread of generated content that features candidates running for office represents an immediate concern, particularly in low-resourced contexts and subnational elections. In some cases, legislation that puts guardrails—or updates existing guidelines—on the deceptive use of generated outputs that feature candidates running for office might make sense.7 However, this approach faces the challenge that politically harmful, but true content could still wrongly be deemed generated—a so-called false positive possible to many AI detectors—and generated content could be evaluated as true—a false negative. As a result, any legislative process will need to consider these clear shortcomings when assessing possible violative behavior.

- ***Additional user requirements to generate content featuring candidates running for office:*** Another possibility is for the tech companies that develop AI tools to require additional disclosures and validation processes for users seeking to generate outputs from a list of candidates actively running for office around the world. Candidates hoping to use these generators as part of their campaign could still be allowed to do so by providing additional information, but it may also allow for better tracking of deceptive generated content. This approach has the same shortcoming as watermarking efforts, namely that it cannot stop smaller-scale actors or adapted open-sourced models from producing this type of output. However, it may help address the challenges stemming from mainstream AI-generator tools, that at least for now are more likely to produce the highest quality outputs.

Distribution

Strategies that tackle the generation process are an important avenue for intervention, but so too is addressing how harmful generated content spreads. Without the ability to spread widely online, the Slovakian deepfake would have barely resonated. The reason it spread was due in part to the nature of the output—an audio recording—which bypassed Meta's content moderation practices that focus exclusively on video-based, wholly generated media. To address these distribution-related challenges, tech companies could:

- ***Revisit and close loopholes in manipulated media policies of social media platforms:*** Social media platforms should urgently revisit their manipulated media policies to ensure they are well-equipped to contend with all types of generated content, including video, audio, and images. Companies also must decide whether these policies should incorporate partially manipulated content designed to mislead or exclusively focus on AI generation, even if the former has the same effect on voters.

- ***Collaborate to better identify harmful generated content and share information across platforms:*** The tech companies that develop generative AI tools and the social media platforms where this type of content spreads should also collaborate to limit the widespread dissemination of harmful generated content tied to elections. A repository of recently generated political content could make it easier for social media websites to identify malicious generated political content at scale. Platforms could also use this space to share information about other AI-generated posts identified from lesser-known tools that do not participate in cross-platform collaborations. This type of approach could be similar to the hash-sharing database of the Global Internet Forum to Counter

Terrorism (GIFCT), which anonymizes images and videos from known terrorist organizations into numerical representations that GIFCT member companies have removed from their platforms. This means that if generated content appears on Facebook and it is removed, it could then be securely added to a searchable database for reference by other Trust and Safety staff across different platforms, all while retaining user privacy.

The limitation of these approaches is that they require investments in Trust and Safety work and content moderation across social media platforms. Additionally, the fragmentation of the information space across many different actors—including some decentralized ones—makes this type of work more challenging due to the proliferation of additional stakeholders with varying degrees of interest in stemming the flow of malicious, generated content.

Detection

Looking further down the information pipeline, all actors—from government officials to social media platforms—should invest more in detection capabilities, which could involve technical solutions, mandated researcher access, and voter education. While these efforts will always operate in the manner of an arms race, detection efforts should not be excluded from approaches designed to mitigate the potential harms of AI-generated content, even if they will need to evolve as the capabilities of generator tools improve. Addressing these detection-related challenges may require:

- ***Additional research and resources to improve AI detection tools:*** At present, the capabilities of AI detector tools vary dramatically, and the risk of false positives and negatives is high. In tandem with research to improve the credibility of generated outputs, tech companies actively developing AI tools should invest far more in improving detection approaches to better identify generated content online, for example, of fabricated images or videos. Beyond tech company investments in this space, foundations and funders also could support research and development of these types of tools.

- ***Broader research access to social media data:*** The research community also will play a critical role in identifying the distribution patterns of generated content and offering an external view on the landscape, without some of the incentives that may shape ongoing research within the more profit-driven private sector. At present, researchers have limited access to the data required to evaluate the prevalence and impact of AI-generated content, particularly as it pertains to elections. This makes it difficult to pinpoint information operations and to assess the prevalence and scope of this challenge vis-à-vis broader trends

in the information space. By understanding the extent to which generated content spreads across multiple social media platforms, researcher can provide an external assessment of the threat landscape to level-set concerns about the proliferation of this type of content, particularly when overhyping its prevalence may inadvertently facilitate a "liar's dividend." It will also enable the development of more tailored, evidence-based policies to promote AI's benefits, while mitigating its harms. Legislation mandating data access by external researchers (akin to the Digital Services Act) remains critical. However, other existing proposals to simulate the online platform experience could also help facilitate research for scholars unwilling to or unable to collaborate with private sector actors. It is, however, important that these opportunities be available to researchers defined broadly—including civil society and think tank researchers—and not just those affiliated with an academic institution.

- ***Widespread education efforts focused on digital literacy in the AI era:*** It is critical for election officials, tech companies, and social media platforms to develop and widely disseminate voter education that highlights ways to approach political material skeptically, particularly given the potential for it to be wholly fabricated.9 More broadly, this education should focus on tips for identifying credible vs. generated content, sometimes known as "glitch analysis," with the recognition that these strategies are already not foolproof and will likely become less relevant over time. For audio, this could include asking questions such as: What does the voice tone sound like? Does the pronunciation sound awkward? Is the word choice unusual or highly formal? Do pauses seem unnatural? Are there particular elements of the spoken quality of a certain language that seem off? Are there factual or grammatical mistakes? For video, this could include questions such as: Does the audio look like it is synced to the movements of the person's mouth? Does the person depicted ever pause? What are the eyes doing during the video? Do gestures and movement seem natural? For images, the questions could include: Do the hands have an unnatural number of fingers? What does the background look like? Are accessories distorted? And do reflections in mirrors converge at a single point? These types of signals are far from infallible, and the best models are quickly learning how to address some of these issues. But, in a historic election year, they may still be useful clues as voters encounter information online.

As policymakers, tech companies and researchers continue to explore the malicious applications of AI-generated content, it is important to underscore their potential beneficial effects on elections too. For example, AI tools can help candidates reach new voters in their native language or assist with translating important campaign and election-related information into other languages, filling content gaps and data voids

where false claims thrive. Given the productivity benefits of generative AI, these tools may also help less well-resourced campaigns remain competitive. In tackling any challenge related to generated content, shifts in policies and approaches should focus on the harms of these outputs, rather than whether or not the content is made using generative AI.

Case Study: Upskilling for Career Mobility at PepsiCo[54]

Upskill America, Aspen Institute

Introduction

As one of the largest food and beverage companies in the world, PepsiCo is a more than $85 billion organization whose products are consumed more than a billion times each day.

With more than 300,000 employees globally and 100,000 US-based employees — many of whom are in front-line roles responsible for making, moving, and selling its products — PepsiCo is committed to creating meaningful jobs and growth opportunities. The company has developed a suite of upskilling initiatives that provide end-to-end opportunities. Employees have access to everything from high school diplomas to basic digital training to earning a bachelor's degree, all at no cost to them.

"At PepsiCo, we encourage our associates to embrace a 'learn it all' mentality," said Ronald Schellekens, PepsiCo's chief human resources officer. "We've put associates at the center of our training design, delivering resources in formats that resonate with them and are digestible within timeframes that can be integrated into daily operations. Learning takes place through various methods, and our goal is to ensure we are providing the tools to help our associates fulfill their career aspirations."

A Comprehensive Suite

PepsiCo operates multiple distinct but connected upskilling programs designed to prepare employees for an increasingly digital future and to help them build the skills they need to advance.

The company leverages these upskilling investments to create lasting value for the business and the overall career mobility for workers. "Everyone has a different 'why' that motivates them to learn," PepsiCo's Chief Learning Officer Molly Nagler said. "We offer a portfolio that meets both employee and business needs. There's something for everyone."

---

[54] This study was originally published by Upskill America, an initiative of the Aspen Institute Economic Opportunities Program, on August 23, 2023. UpSkill America thanks Molly Nagler, Dewey Torres, Cristina Rivera, and Maly Scott from the Global Learning Center of Excellence and sector leads Abigail Helems and MaryKate Bisket for their time and insights. This brief was prepared by Haley Glover, director of UpSkill America, an initiative of the Economic Opportunities Program at the Aspen Institute.

Digital Academy: No-Cost Digital Learning for All

PepsiCo's Digital Academy includes more than 11,000 learning assets designed to help any employee in the company acquire the digital skills they need. The Academy's curriculum is multilevel. It offers content tailored for employees who are already in technical roles and want to keep learning, as well as other courses and programs built for those who are not in technical roles, but regularly use digital tools in their work. Additional options are available for those who are simply learning the basics. Employees can access a variety of resources, from short how-to videos to more in-depth boot camps on technology topics and competencies.

The Academy offers both on-demand courses, which employees can access and complete at no cost for professional development and ongoing learning, as well as pathways to credentials and certifications. Launched in 2022, more than 11,000 employees participated in 140,000 self-paced learning modules within the first year, earning 600 certifications in areas including Cloud Azure, Data Scientist, DevOps, Site Reliability Engineers, and Power BI for Data Analytics.

myeducation: Improving Access to Credentials

PepsiCo's myeducation benefit launched in 2022 for full-time US-based employees, providing them with access to more than 100 diploma, certificate, and degree options in a variety of fields at no cost to them. The catalog prioritizes programs in high-demand skills, such as data analytics, the trades, and supply chain, as well as high school completion and English language learning.

myeducation also supports employees with opportunities to earn credentials, including commercial driver's licenses (CDL), which are required for high-demand roles transporting products. Available to all eligible employees after six months of continuous work without manager approval, the benefit removes potential bias and barriers from access to programs and empowers employees to drive their own development. Moreover, PepsiCo pays 100% of the cost for tuition, books, and fees up front, eliminating financial barriers to participation.

In partnership with Guild, 160 Academy, and Ancora, myeducation offers access to highly reputable schools and universities as part of the 100+ programs in the catalog — all tied to business needs and internal career pathways.

"This is not just about graduations, it's about mobility post-completion," said Dewey Torres, senior director of PepsiCo's Global Learning Center of Excellence, who is responsible for driving the company's upskilling and reskilling solutions. "We're building an internal talent pipeline."

Building a Talent Marketplace

The Global Learning team, as part of the HR function's holistic talent strategy, is working to connect individual programs into learning journeys that build robust talent pipelines for hard-to-fill roles.

PepsiCo has implemented an internal talent marketplace that is not only the access point for these learning journeys, but also a means for employees to have visibility into the array of opportunities available to them.

Employees can learn about roles where their skills might be a good fit, as well as the learning paths needed to advance into roles. "The ultimate goal is to create a space for every employee, no matter their role or position, to build skills and advance their career, knowing that we're moving into a digital future," said Maly Scott, senior manager and global learning lead for digital skills at scale. "That's a broad mission, but it's important to create a personal and tailored journey for everyone by bringing more transparency to what is possible and access to clear pathways."

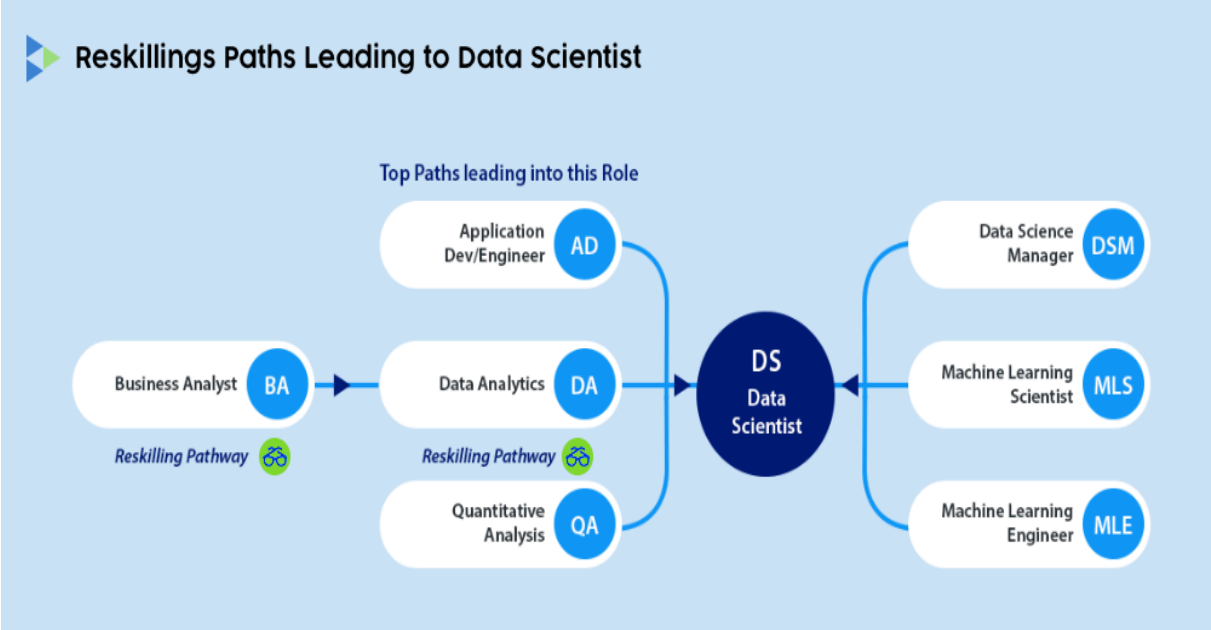
A Deeper Look: myDevelopment

myDevelopment, an internal talent marketplace powered by Gloat, integrates project and experiential learning alongside academic programming. Most roles require some experience, which can be a barrier for a full-time worker in their current role. Through myDevelopment, employees can apply for one of more than 500 "stretch projects" to develop new skills and build competency. Additionally, the platform enables employees to find and apply for 90-day, short-term assignments. These short-term assignments enable associates and managers to try out positions in new functional skill areas in a low-stakes way before deciding on a potential new career path.

Through this approach, PepsiCo helps to level the playing field for incumbent employees, who can access entry-level roles in new areas and compete with applicants who have internship and other job readiness experiences. Each project requires only a modest amount of time outside work hours to complete, and project outcomes are logged internally through an employee profile, which can be easily shared for internal job applications and interviews.

To support employees' use of myDevelopment, PepsiCo utilizes Draup, an AI talent analytics platform, to identify the technical and durable skills required in a role and to construct meaningful career pathways. Using a data system that combs through millions of job postings each day, Draup enables PepsiCo to understand where there are roles with high alignment across five criteria (all of which can be weighted based on company priorities), including technical skills, soft skills, compensation, median experience required, and data on common role transitions.

With the intelligence from this platform, PepsiCo can show employees how the skills they have in a current role align with other roles within the company, as well as the upskilling or reskilling required to get there. "We take this information for high-growth areas, and we use it to define a career path so folks can see a trajectory," said Torres. This can be particularly helpful when comparing two roles and functional areas that are emerging or newly defined, versus career paths that are better known within PepsiCo.



Skills are a common language across programs, platforms, and solutions that power PepsiCo's talent marketplace, making this a practical and strategic effort.

Practically, orienting on a common skills taxonomy enables human resources to "harmonize" the process of creating an improved learner experience. The employee user experience is vitally important to PepsiCo. "There is so much underneath each of these programs," said Scott. "We have to remove friction for people."

Focusing on skills is a significant value-add for talent at PepsiCo. For instance, across its North America food and beverage businesses — PepsiCo Foods North America and PepsiCo Beverages North America — business leaders work with HR to identify the skills required for particular roles. The team analyzes training program outcomes against required job skills and shows the program outcomes. If there are gaps, the team returns to the provider to close them. "But first, we need to have that discussion with the HR business partners," said Torres. "We want them to be able to offer the upskilling programs to our current front-line talent and see this benefit as an internal talent pipeline."

The company's focus on skills is also a strategic choice that allows corporate learning efforts to impact the entire organization. With a deeper understanding of employees' skills, corporate learning and development activities can be tailored to personal career journeys and aggregated across departments and the organization to show areas for future learning and development. "We are looking at everything across the entire talent lifecycle and showing people what is possible," said Torres.

Aspen Institute Congressional Program

Organizing the company's learning and development efforts around skills enables the Global Learning team to prove its value.

Upskilling As a Strategic Investment

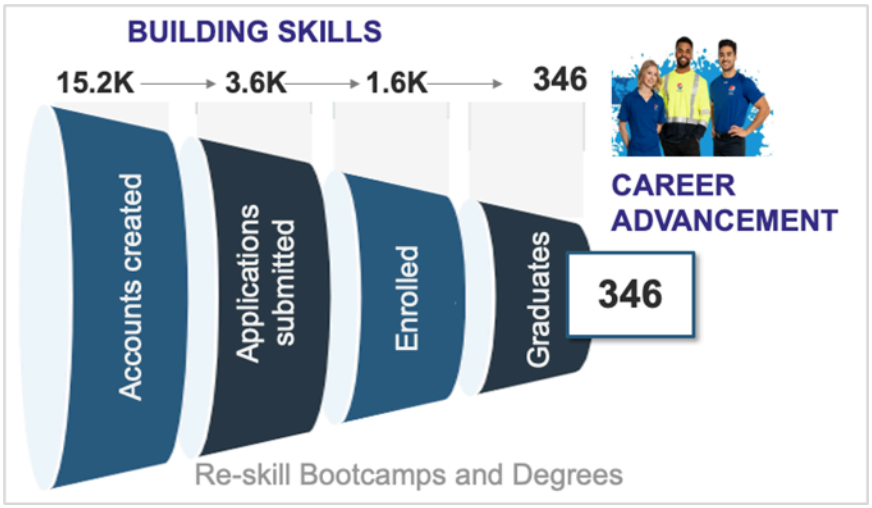
Along with the various business units, the Global Learning Center of Excellence is committed to showing upskilling and learning efforts in a different light, a strategy that is vital to the future of the company and its people.

PepsiCo is working proactively to educate incumbent employees to consider open roles that are hard to fill. For example, there is significant skill shortage for maintenance mechanics and CDL delivery drivers. Upskilling programs enable front-line employees to go through a certified program to move into these high-growth roles. Not only do maintenance technicians/drivers earn a competitive salary, but the company also currently has many open positions for these roles, which costs PepsiCo in overtime and lost productivity when they cannot fill them externally.

PepsiCo also tracks the outcomes of its education and training programs. Since launching in March 2022:

- More than 15,000 US-based employees have created a myeducation profile, and more than 1,600 have enrolled.
    - 40% of those enrolled are pursuing credentials in high-demand digital fields at PepsiCo, including data analytics and cybersecurity.
    - 346 employees have already completed credentials through the program, including high school diplomas, CDLs, and even several bachelor's degrees.
- The program is driving retention, especially among early-tenure front-line employees. Attrition is 18% lower for myeducation participants versus non-participants.
- PepsiCo leadership is excited by early evidence for career mobility.
    - myeducation participants are 1.7 times more likely to experience a job level or role change.
    - 311 program participants (including active students and graduates) changed roles in 2022, with 125 "meaningful promotions" involving a level change and 13 placements into priority job roles.
    - 60 employees have earned their CDL and now fill the high-demand driver role in the company.

- myeducation is creating opportunities for PepsiCo's diverse workforce.
    - 58% of myeducation participants represent diverse backgrounds, while 46% of the company's workforce identify as diverse.
    - Black women are 1.9 times more likely to enroll than other demographic groups.
- New testimonials are captured every month.
    - Amanda Bradshaw: "I thought that I might be an administrative assistant for my entire career. But thanks to myeducation, I was able to complete a certificate in data analytics, and at the end of the program, I was promoted to a food service sales analyst."
    - Danny Rodriguez: "I used to be an operator on the lines. After the certificate in myeducation, I became a mechanic and I fix the machines. And I actually love my job."



Transforming Roles

The learning efforts and offerings at PepsiCo continue to evolve with the demand across its workforce. The company's overall goal is to invest in an employees' employability and prepare them for current and future roles.

For example, within plants, PepsiCo is using Autostore technology, an automated system that identifies and retrieves products for distribution. This innovation requires employees to learn a new skill, while also saving time and energy during the day. This

enables the company to help an employee build a new skill while improving how their time is spent.

PepsiCo's upskilling solutions are tailor-made for employees in transforming roles and are designed to cultivate the full array of the skills they need to thrive. "This is intentional from the digital perspective, but we are looking at the holistic picture of what employees need to develop in their roles," Scott said.

PepsiCo's digital transformation initiative is helping to improve the employee experience on the job, while investing in building skills that will prepare them for the future.

"Our priority is to keep humans at the center of our digital transformation," said Athina Kanioura, chief strategy and transformation officer at PepsiCo. "When I joined PepsiCo, we conducted an assessment of the maturity of our technology platforms and training for all levels, including our front-line employees, associates, and executives. As a result, we launched Digital Academy, which is inclusive to all employees and leverages the power of AI to recommend the best training courses and degrees based on an employee's role and experience and future career aspirations. We've seen tremendous results with the program so far. Overall, we hold firm to the belief that continuous learning and progress in one's career are vital at every stage. It's not just about filling in the gaps; it's also about advancing their careers. Increasing digital knowledge — not just on my team, but across the organization — can have a positive impact across the entire company."

What We Learned

In just over a year, PepsiCo has started to transform its upskilling operations, creating significant value for the business and its workers. Other employers making these changes can learn from PepsiCo's experience.

Reducing Friction, Creating Meaning

The Global Learning team at PepsiCo uses the word "expose" frequently, both in the context of exposing employees to upskilling opportunities and in the context of bringing to light things that were previously hidden or complicated. A major component of PepsiCo's upskilling success has been its commitment to reducing friction, increasing transparency, and creating meaning for employees. The company's approach is learner- and employee-centered, prioritizing messaging, platform design, and information that build employee confidence and eliminate confusion.

"PepsiCo had to bring all these disparate programs together and mask that complexity by describing what employees can actually accomplish through their programs — take on a stretch opportunity, pursue a degree, etc. — all through the lens of the employee. Most big companies have similar solutions and programs, but the top layer is the learning experience where companies have to piece it all together for the learner," Torres said.

Beyond eliminating confusion and intimidation around what programs offer, PepsiCo also committed to increasing transparency in what employees could see. For the Global Learning team, that meant not only showing what skills employees could gain through the Digital Academy, but also how those skills align with our ongoing transformation. It meant that employees could understand which postsecondary credential programs were available and how those credentials would contribute to a new career path. And it meant those who needed extra experience and a venue to demonstrate their skills could do that without leaving the company or sacrificing significant off-work hours.

The increased transparency and strategic integration of upskilling programs with business needs does not come at the cost of employee choice or self-determination. PepsiCo's Global Learning team hypothesized that employees who ran into barriers finding programs, understanding their purpose, and envisioning themselves participating would become discouraged and not participate. Further, the team knew that employees needed a full array of upskilling options that would meet employees where they are and fit into their complex lives. "You have to deliver the message and the learning in a way that feels approachable and like it was built for you personally," Scott said. "So, we build digestible, consistent stories, helping our employees understand that we're working to set them up for success — not leave them behind."

The personalized approach extends beyond the front line. PepsiCo's platforms also focus and tailor learning content to executives and managers. The content is at the right level, curated for roles with the right content for them.

An End-to-End Responsibility

PepsiCo is intentionally connecting its upskilling and learning programs across multiple dimensions that exemplify best practices, including:

- Learning Program Array — PepsiCo employees can earn credentials from high school diplomas and short-term certifications to bachelor's degrees, or learn through self-paced programs, meeting all employees where they are on their upskilling journeys.
- Application — PepsiCo has created venues for employees to access and apply their learning, both in short-term "stretch projects" that allow for demonstration

of skills and in advancing roles. PepsiCo's learning programs are engines for employee mobility.

- Internal Impact — Working across the entire talent lifecycle, PepsiCo's learning programs are designed to address and solve business problems. Learning programs are not standalone or partitioned benefits — they are integrated solutions with company-wide impact.

PepsiCo's learning leaders take this comprehensive approach and its impacts personally. "I feel like it's our responsibility to build skills for our employees, but it doesn't stop there," Torres said. "We need to guide people along in their journey, rather than wait to see how many can connect the dots and figure it out by themselves."

Getting to Core Issues

The level and impact of change proposed by PepsiCo's Global Learning team and the need for compelling reasons to do things differently meant the team needed to get specific with their objective. They built use cases for their work, demonstrating how the proposed shifts would solve problems and how systems would work. They also met company leaders with compelling data and evidence, showing precisely how learning programs were addressing pain points and driving results.

Beyond solving problems, though, the team worked to transform the system. PepsiCo had many individual solutions in place that provided value. However, by helping connect the dots from the center, employees are getting much more value from those solutions working in harmony.

Centering Learning Leadership

While most large organizations have learning programs that support and advance upskilling and development for employees, the quality and investment vary greatly. PepsiCo's Global Learning team has significant programming opportunities, as well as a budget, to advance their goals, but the larger body of work is connecting the dots internally and making the case for change.

"The company has been around for a long time, and we've done things very well historically, so there has to be a compelling reason to do things differently," Scott said. "We've focused on being the translator and connector of all these people and initiatives… there are pockets everywhere, but no one previously helped drive those pieces together and help surface and challenge those silos."

The Global Learning team also centered business strategy and bottom-line value in its approach, working at every stage to solve problems for stakeholders throughout the company and the talent lifecycle.

Conclusion

PepsiCo is on a journey — one the company anticipates will increase meaningful opportunities for employees and enable the business to operate even faster, stronger, and better. With a focus on digital upskilling, high-quality credentialing opportunities that align with in-demand roles, and the infrastructure to enable internal candidates to demonstrate skill mastery, PepsiCo is creating the conditions for career mobility.

Focusing on the end user, such as the front-line employee, PepsiCo also drives efficiency and effectiveness, creating learning experiences that are easy for employees to understand and have transparent outcomes for employees and the company.

Thousands of PepsiCo employees have engaged with learning opportunities within just one year of full implementation. Hundreds have been promoted into new roles. These early results are encouraging, especially when coupled with the strategic connections and value across the talent lifecycle generated throughout the company by the Global Learning team. PepsiCo is taking a proactive, strategic approach to upskilling that will improve economic mobility for employees and drive lasting value for the company and beyond.

Frontline A.I.: A Guide for Manufacturers[55]

Aspen Digital, Aspen Institute

Retention and recruitment in the frontline workforce is a challenge for US manufacturers, both large and small. Some firms are turning to automation to solve these problems, but initial research shows that deploying automation without also making job quality improvements may do more harm than good.

For manufacturers that are considering integrating automation, be it artificial intelligence (AI) or otherwise, into their operations, the recommendations presented here provide guidance. They are based on interviews with experts, research, and best practices from leading firms in the field on how to deploy automation in a way that raises job quality, improves retention and recruitment, and protects their bottom line.

The Challenges

1. Reduced human oversight in AI software and data-driven automation brings a new set of risks.
2. New and changing technologies require new skillsets, but manufacturers are struggling to source suitable talent.
3. Automation deployment in manufacturing can lead to deskilling and higher churn, intensifying retention challenges for the industry and contributing to institutional brain drain.

Business Motivations

Manufacturers are turning to AI tools to increase the productivity of plants (reducing downtime, reducing waste, increasing capacity per line) while reducing the amount of human capital (both number of workers and hours worked) required to perform certain tasks. The tax structure in the US incentivizes automation, since taxes are higher for labor and lower for capital. Severe production disruptions driven in part by the COVID-19 pandemic have created a very strong incentive to automate work. A study by Oxford Economics estimates that 20 million manufacturing jobs will be automated by 2030. Given the industry's challenging history with employee retention and recruiting, some companies are looking at automation as a silver bullet for all of their workforce challenges.

[55] This work was made possible with the support of PepsiCo, Inc. Aspen Digital is grateful to Dr. Athina Kanioura and her team who supported our research. You can see the original article here.
Thanks also to Eleanor Tursman, Morgan McMurray, Elizabeth Miller, Anahita Sahu, Devon Regal, Haley Glover, B Cavello, and our other Aspen Institute colleagues for their contributions to this work.

Applications of Automation

The Center for Economic Studies found that as many as 64% of US workers and 72% of manufacturing workers are exposed to automation technologies like AI, robotics, and specialized software, based on data from the US Census Bureau's 2019 Annual Business Survey. Adoption is concentrated in large firms. Common applications of these technologies include:

- **Maintenance**: Predictive maintenance software leverages a network of sensors on machinery to detect signs of wear so that servicing can be planned pre-emptively instead of in response to expensive breakdowns.
- **Quality control**: Companies are automating processes, often through a combination of sensor data and AI software, to identify when a product fails a quality control check.
- **Dirty, dull, and dangerous work**: Machines are used to perform repetitive, specific manipulation tasks, such as processing components on an assembly line, and can be deployed in environments where temperatures or aerosols are unsafe for humans.
- **Streamlining human-computer interactions**: To speed up redundant computer interactions such as copy-pasting information across platforms, companies use digital rule-based programs called "robotic process automation" that replicate the interactions of a human navigating through computer interfaces.
- **Shift scheduling**: AI-driven tools are used for "smart scheduling" which can reduce unassigned time. Although these systems can be designed and used to give workers more control over their hours, they are frequently used as a form of worker surveillance, cited as a cause of high worker burnout and turnover.

Impacts of Automation

Availability of Jobs

Research on the impact of AI adoption on the availability of jobs and wages is not clear-cut. Some groups predict that the impact will depend on the specific use case and application of the AI tool, while others say that there is a real chance that AI deployment will ultimately lead to pervasive unemployment. Regardless of the net impact of AI adoption on jobs overall, it is clear that some jobs will be lost, leading to:

- **Fewer jobs focused on routine tasks**: Workers without specialized skill sets and workers performing routine or replaceable tasks may struggle to compete with automation.

- **Disproportionate displacement**: Women (nearly 30% of the frontline manufacturing workforce), workers of color (nearly 30%), and workers without a 4-year degree (nearly 71%) are more likely to face disproportionate displacement.
- **Lower barriers to entry**: Automation that simplifies tasks can make a role accessible to a larger number of potential workers, but this can also result in greater competition for roles.
- **Training challenges**: The benefits of AI tools are greater for those with specific training and roles, disproportionately putting certain workers, especially those over 40, at higher risk of job displacement due to automation.

Job Quality

People promoting AI as beneficial to the future of work have advocated that it can free up workers from dirty, dull, or dangerous tasks, allowing them to instead focus on higher-value activities that enable upward mobility internally within the organization. Commonly mentioned benefits include:

- Making industrial work safer: Minimizing direct contact between a human worker and dangerous machinery via robotics can improve safety in industrial workplaces.
- Reducing physical labor requirements: Manufacturing work has historically been physically demanding and taxing, limiting the number and tenure of workers in the field.
- Freeing workers up from undesirable labor: Automation can reduce the amount of necessary tedious or repetitive work.

However, 2022 research from the Partnership on AI has indicated that, in fact, the opposite is happening:

"CURRENT IMPLEMENTATIONS OF AI IN WORK ARE REDUCING WORKERS' OPPORTUNITIES FOR AUTONOMY, JUDGMENT, EMPATHY, AND CREATIVITY...WORKERS IN US WAREHOUSES WITH HIGHER DEGREES OF AI IMPLEMENTATION OFTEN HAD LESS VARIETY IN THEIR TASKS AND MORE TECHNOLOGICAL GUARDRAILS TO ASSIST THEM IN PERFORMING THEM CORRECTLY."

According to these findings, which analyzed interviews with frontline workers in manufacturing, call centers, and data annotators who work with AI tools, the main threat of AI deployment lies not in reducing the number of available jobs for workers, but in decreasing the quality of work, by:

- **Placing harmful pressure on workers**: Algorithmic management tools are often used to push intense, algorithmically-set productivity quotas.
- **Reducing skill requirements to do certain work**: Employers risk "deskilling" workers, where workers learn less about the processes they are contributing to, disrupting pathways to higher-paying work and reducing institutional wisdom.
- **Minimizing worker voice and agency:** Using tools to reduce worker autonomy leads to higher rates of attrition.

Maximizing the Benefits and Minimizing the Harms

Managerial decisions play a substantial role in shaping the impact of technology. Employers have an opportunity to lead and reap the benefits of implementing worker-centered processes when introducing these new tools. By doing so, companies not only foster a positive work environment but also enhance their competitive edge, ensuring sustainable growth and minimizing potential pitfalls associated with technology adoption.

Aspen Digital, in consultation with experts from academia, civil society, and industry, has developed the following recommendations on how to integrate automation into the manufacturing frontline responsibly. At a high level, getting the most out of automation requires thinking of workers as investments and assets that must be leveraged. The following themes represent best practices that are further detailed in the recommendations below:

- **Upskilling**: Train the workforce to adapt to new skills automated tools require.
- **Human-in-the-loop**: Maintain human review and control of automated or AI-informed decisions.
- **Participatory design**: Use tools that incorporate feedback from and address pain points of the workers themselves—people who will actually be using the tools.
- **Human interaction**: Emphasize and support human-to-human interaction at work.
- **Supportive, not prescriptive, tools**: Use tools that are meant to increase the agency of workers.
- **Transparency**: Be transparent with workers about the benefits and risks of

automated tools, both to the firm and to them.
- **Predictability**: Look for and evaluate tools in part based on whether they bring predictability into the frontline workplace (such as a scheduling tool that makes it easier for workers to plan their shifts).
- **HR foundations**: Attract and retain talent with competitive compensation and benefit packages.

The following recommendations may be more ambitious for smaller firms, so buy-in from leadership will be pivotal. Small firms should consider:

- **Collaborating** on best practices with other firms in their region through industry-sector partnerships.
- **Partnering** with local community colleges for training and upskilling.
- **Creating** a registered apprenticeship program with resources from the Department of Labor.
- **Using** regional Manufacturing Extension Partnership (MEP) resources.
- **Tapping into** their local workforce development boards.
- **Consulting** organizations like the Workforce & Organizational Research Center (WORC), America Works, the Urban Manufacturing Alliance, and the Institute for the Future of Work.

The recommendations below are tailored to meet three goals shared by leaders in manufacturing:

1. Reduce the risks of automated systems.
2. Upskill to get the most out of automation investments.
3. Retain workers and their valuable institutional knowledge.

To read about these goals and the recommendations to address them in more detail, please see the accompanying Goal Spotlights for each goal. The Frontline AI Cheat Sheet also contains helpful terminology and more details on strategies for worker engagement.

Issues for Further Exploration

- Large and small manufacturers do not equally enjoy the benefits of AI and automation. Smaller manufacturers are often limited in their ability and resources to initiate conversations on workforce development. Early and clear leadership from larger manufacturers in partnership with their vendors and

suppliers can help steer the industry goals in a direction that prioritizes worker well-being.

- Transitioning to automated manufacturing while prioritizing job quality is a challenge. Industry-wide foundational and fundamental cultural shifts are required where the frontline workforce is viewed as an investment and asset that must be leveraged.
- Manufacturing employers and labor representatives are frequently portrayed as opposed, but they share many goals when it comes to the impacts of technology regarding staff retention, productivity, and resilience. Successful implementation of these recommendations requires collective action and continued engagement with key allies such as labor groups, employer coalitions, technology vendors and civil society.

Actionable Recommendations

To read about these goals and the recommendations to address them in more detail, please see the accompanying Goal Spotlight for each goal. The Frontline AI Cheat Sheet also contains helpful terminology and more details on strategies for worker engagement.

Goal 1: Reduce the Risks of Automated Systems

At a Glance:

1. Reinforce to managers that AI tools can (and do) make mistakes.
2. Maintain managerial decision-making and human oversight of automated systems.
3. Develop clear internal guidelines for identifying contexts in which AI should not be used, such as in hiring.
4. Evaluate the impacts of deploying automated systems on your workforce by identifying and tracking KPIs that
5. measure employee satisfaction, health, and skill development, such as internal promotion rate, injury rate reduction, and employee satisfaction index.
6. Ensure that an AI tool was designed to meet your specific needs by consulting workers.
7. Set up real-time feedback loops during and after deployment using insights from the frontline. For more resources, see Strategies for Worker Engagement.
8. Use a combination of quantitative KPIs and qualitative worker feedback via surveys or managerial check-ins to evaluate physical and mental health impacts of deployed automated systems.
9. Ask technology developers or vendors specific questions about their products

including about your right to repair, the interoperability of their tools, and the ownership of data collected.
10. Prioritize informed consent prior to data collection.

Goal 2: Upskill to Get Most Out of Automation Investments
At a Glance:

1. Identify skill gaps and provide training in basic digital skills based on what types of upskilling workers want.
2. Use community college partnerships to develop high-value interpersonal skills such as knowledge-sharing, conflict resolution, and negotiation.
3. Make upskilling accessible by making sure training opportunities are available on site, during work hours, in multiple languages, and with appropriate compensation for time spent.
4. Provide training for a variety of skills and in a range of formats based on what workers prefer, such as cross- training workers on different technologies and mentoring programs and apprenticeships.
5. Clearly outline economic and career mobility benefits for workers who participate in an upskilling program.
6. Designate a worker or small group of workers as subject matter experts for new technology or specific functions of the new technology.

Goal 3: Retain Workers and Their Valuable Institutional Knowledge
At a Glance:

1. Be straightforward and communicate clearly with workers about expected changes by providing comprehensible explanations of the AI system's function, talking plainly about staffing changes, and avoiding technology or business jargon.
2. Provide adequate (at least 8 weeks) notice to workers and unions before deploying new technologies.
3. Get feedback and collaboratively define productivity goals, both anonymously and through high-touch options like workshops, when adopting new technology.
4. Seek worker input when creating policies for algorithmic management and worker surveillance (e.g., wearable technology, sensors, and other monitoring systems), both of which can decrease job quality and impact retention.
5. Provide a clear career growth plan and allow workers to advance professionally by providing advancement training and opportunities on a yearly or more frequent cycle.
6. Deploy automation in ways that provide equal opportunities to employees of all

backgrounds, regardless of race, age, gender, education, experience, native language, or other individual traits.

7. Recognize and compensate workers for their role in training peers and new hires through initiatives such as microcredentials, scholarships, or paid management trainings.
8. Deploy technology that will complement or support your workers' professional identities. Complementary technologies are much more easily accepted and adopted.

RAPPORTEURS' RECOMMENDATIONS

Why AI Is Such a Hard Problem for D.C.[56]

Derek Robertson, Politico

POLITICO's AI & Tech Summit yesterday brought together legislators, entrepreneurs, and policy wonks to hash out exactly where American governance stands in respect to this transformative new technology.

So… what came of it?

Like most serious policy discussions, the chatter Wednesday was split between what AI means for America's leadership globally, and what policymakers can do here at home to enhance it. (Read the big takeaways from most individual panels here.)

Geopolitically, the main issue seems clear: Competition with China, and establishing global standards for AI that counter the authoritarian use of technology. But when it comes to what to tackle first in Washington, the answer is murkier. Let's take a look at some of the summit's biggest moments to get a better picture:

ON CHINA: Palmer Luckey, the flamboyant founder of the defense contractor Anduril Industries, distilled years of anxiety and tightrope-walking among global tech giants over a theoretical conflict with China into a pithy appeal to companies not yet sold on economic nationalism.

"How stupid will you feel if you build a company that assumes the geopolitical situation with China stays the same or improves?" Luckey asked, positing a scenario where an invasion of Taiwan or other geopolitical tensions force strong American sanctions or even military action.

"You won't be able to look back and say, 'Who could have predicted this, nobody saw this coming.' When you're a new company you can choose to decouple yourself from China, you can choose to make things in other countries, there are other options for most products," he said.

Even just 10 years ago, a fully-globalized, China-heavy supply chain for advanced technology like the microchips used to power advanced AI systems seemed like a

[56] This article was published by Politico on September 28, 2023

permanent feature of the economic and tech landscape. Now America is actively restricting China's access to futuristic chip technology, with more controls likely on the way. The shape of the digital future can change, very fast.

One former official for the Office of the U.S. Trade Representative warned that stopping China from getting its hands on the most advanced U.S. microchips might not be enough, as they make a huge push to invest in "legacy chips" not impacted by current trade restrictions. Lakshmi Raman, the CIA's director for artificial intelligence, warned that China is growing its AI tools in "every which way" to launch a suite of AI-powered cyberattacks and disruptions.

...AND IN WASHINGTON: What are legislators and regulators actually going to do about it?

Most of the talk around AI legislation at yesterday's summit was about its impact in the U.S. — and an audience poll showed attendees worry more about existential risks of AI rather than global competition.

Whatever the concern, lawmakers didn't offer much reassurance in the form of proposals. Instead, they said they're still working on basic questions. "Are we going to do a broad-based approach with a new agency? Potentially like the EU has done? Or are we going to adopt a sectoral approach, where we empower our existing sectoral regulators to regulate AI within their sectoral spaces?" said Rep. Jay Obernolte (R-Calif.), vice chair of the Congressional Artificial Intelligence Caucus.

Sen. Todd Young (R-Ind.) said he thinks it's "very likely that we'll pass at least some narrow pieces of an AI regulatory regime" in the current Congress, but he was vague on the details, aside from saying Americans should be more concerned about the use of automated weapons than in-app AI entertainment. Rep. Ted Lieu (D-Calif.) said any AI legislation is facing an extra obstacle in the current shutdown fight in Congress. Sen. Ed Markey (D-Mass.) called for tightening AI safety measures for children and teens, just after he asked Meta to delay its AI chatbots until their effects on young people were studied, as first reported by POLITICO's Rebecca Kern.

Ironically, all this talk from Washington revealed exactly how much Silicon Valley is still in the driver's seat when it comes to writing our AI future, at least for the moment. As POLITICO's Daniella Cheslow wrote after the summit's close:

"With AI regulation still fluid, industry players are making their own suggestions, and regulators are relying in part on their goodwill... Several technologists made comments that showed they are operating in a regulation vacuum," citing among other things a

senior VP for chipmaker Qualcomm Technologies saying "quite a few of us have our own set of guardrails."

Twitter's slow-but-steady evolution away from its former self continued last night, as Elon Musk announced he disbanded a team focused on stamping out disinformation during elections.

"Oh you mean the 'Election Integrity' Team that was undermining election integrity?" Musk wrote in an X post. "Yeah, they're gone." The Information reported that several staffers in Ireland working on fighting disinformation were fired this week. (The Information also reported that Musk previously said he would expand the team.)

The move is fully in keeping with Musk's overall laissez-faire philosophy on speech, showing a tendency to err on the side of allowing false and hateful messages on the platform in the spirit of open discourse. It also comes hot on the heels of a warning from European Union officials that rampant false information, especially on X, is plaguing the Slovakian elections, just days before the votes are cast.

Which jersey are you wearing in the (rhetorical) AI wars?

In an op-ed for the New York Times, two technologists break down the debate over AI development and policy into three main camps: "Doomsayers," obsessed with the potential existential risk of AI; "Reformers," progressives more concerned about how it might entrench existing inequalities in society; and "Warriors," foreign policy hawks who see it as a tool of competition with China.

"These factions are in dialogue not only with the public but also with one another. Sometimes, they trade letters, opinion essays or social threads outlining their positions and attacking others' in public view," write Bruce Schneier and Nathan Sanders.

The authors say we should do more than just follow how these groups jockey for power, but also to "Look past the immediate claims and actions of the players to the greater implications of their points of view," they write. "This isn't really a debate only about A.I. It's also a contest about control and power, about how resources should be distributed and who should be held accountable."

Child Safety Hearing: Senators Demand Tech Executives Take Action to Protect Children Online[57]

During a tense hearing that included executives from TikTok, X, Snap and Discord, Mark Zuckerberg, the leader of Meta, told the families of abuse victims he was "sorry for everything you have all been through."

Mike Isaac, The New York Times

Six Takeaways from a Contentious Online Child Safety Hearing

After a series of tense exchanges between senators and tech executives that clocked in at just under four hours, the Senate Judiciary Committee hearing on online child safety came to an end on Wednesday with no clear resolutions in sight. The audience included several family members of victims, who cheered as senators berated the executives and listened stoically as Mark Zuckerberg, the chief executive of Meta, addressed the crowd directly.

Here are some of the key takeaways.

Senators were aggressive in their questioning.

In one of the more combative tech hearings in recent years, senators from both parties refused to back down and pressured the chief executives of Meta, X, TikTok, Discord and Snap to take responsibility — and apologize — for their companies' role in harming children. At times, the senators shouted and talked over the executives, drawing applause from those in the room. Senator Lindsey Graham of South Carolina said the companies had "blood on your hands."

Zuckerberg addressed families of victims.

After being pressured by Senator Josh Hawley, Republican of Missouri, to apologize for the harm caused by Meta, Mr. Zuckerberg stood from his chair, turned around and addressed families of victims in the audience who had suffered abuse on Meta's apps.

"I'm sorry for everything you have all been through," Mr. Zuckerberg said. "No one should go through the things that your families have suffered." He said that his company was working so that no one else would have to do so, and did not address Meta's role. The leaders of Meta and TikTok took most of the heat.

[57] This article was published by the New York Times on January 31, 2024.

Though executives from Meta, Snap, Discord, X and TikTok were all called to the hearing — the latter three were subpoenaed to testify — it was Mr. Zuckerberg and Shou Chew, TikTok's chief executive, who spent the most time in the spotlight. Senators grilled the two men on the number of abuse incidents across their platforms.

Two of the five chief executives agreed to support the Kids Online Safety Act.

Evan Spiegel, chief executive of Snap, and Linda Yaccarino, who leads X, both agreed to support the Kids Online Safety Act, or K.O.S.A. The proposed law would require online services like social media networks, video game sites and messaging apps to take "reasonable measures" to prevent harm — including online bullying, harassment, sexual exploitation, anorexia, self-harm and predatory marketing — to minors who use their platforms. Mr. Zuckerberg, Mr. Chew and Jason Citron, the chief executive of Discord, did not pledge their support, with some arguing that it was directionally helpful but contained some overly broad restrictions that may come into conflict with free speech issues.

TikTok faced heat for its ties to China.

Lawmakers repeatedly pressed Mr. Chew about TikTok's ties to the Chinese government, thanks to its Chinese ownership by ByteDance. Mr. Chew, who was born in Singapore and still lives there with his three children, was asked whether he had a Chinese passport or had ever applied for Chinese citizenship. (He had not, though he lived in Beijing for five years.) He was also questioned about the progress of TikTok's multibillion-dollar plan for walling off sensitive U.S. user data.

After years of debate, no bills have passed.

Despite years of railing against Big Tech in public, no meaningful legislation has moved its way through Congress to be signed into law.

Artificial Intelligence Act: MEPs Adopt Landmark Law[58]

News: European Parliament

- Safeguards on general purpose artificial intelligence
- Limits on the use of biometric identification systems by law enforcement
- Bans on social scoring and AI used to manipulate or exploit user vulnerabilities
- Right of consumers to launch complaints and receive meaningful explanations

On Wednesday, Parliament approved the Artificial Intelligence Act that ensures safety and compliance with fundamental rights, while boosting innovation.

The regulation, agreed in negotiations with member states in December 2023, was endorsed by MEPs with 523 votes in favour, 46 against and 49 abstentions.

It aims to protect fundamental rights, democracy, the rule of law and environmental sustainability from high-risk AI, while boosting innovation and establishing Europe as a leader in the field. The regulation establishes obligations for AI based on its potential risks and level of impact.

Banned Applications

The new rules ban certain AI applications that threaten citizens' rights, including biometric categorisation systems based on sensitive characteristics and untargeted scraping of facial images from the internet or CCTV footage to create facial recognition databases. Emotion recognition in the workplace and schools, social scoring, predictive policing (when it is based solely on profiling a person or assessing their characteristics), and AI that manipulates human behaviour or exploits people's vulnerabilities will also be forbidden.

Law Enforcement Exemptions

The use of biometric identification systems (RBI) by law enforcement is prohibited in principle, except in exhaustively listed and narrowly defined situations. "Real-time" RBI can only be deployed if strict safeguards are met, e.g. its use is limited in time and geographic scope and subject to specific prior judicial or administrative authorisation. Such uses may include, for example, a targeted search of a missing person or preventing a terrorist attack. Using such systems post-facto ("post-remote RBI") is considered a high-risk use case, requiring judicial authorisation being linked to a criminal offense.

[58] This article was originally published by the European Parliament on March 13, 2024

Obligations for High-Risk Systems

Clear obligations are also foreseen for other high-risk AI systems (due to their significant potential harm to health, safety, fundamental rights, environment, democracy and the rule of law). Examples of high-risk AI uses include critical infrastructure, education and vocational training, employment, essential private and public services (e.g. healthcare, banking), certain systems in law enforcement, migration and border management, justice and democratic processes (e.g. influencing elections). Such systems must assess and reduce risks, maintain use logs, be transparent and accurate, and ensure human oversight. Citizens will have a right to submit complaints about AI systems and receive explanations about decisions based on high-risk AI systems that affect their rights.

Transparency Requirements

General-purpose AI (GPAI) systems, and the GPAI models they are based on, must meet certain transparency requirements, including compliance with EU copyright law and publishing detailed summaries of the content used for training. The more powerful GPAI models that could pose systemic risks will face additional requirements, including performing model evaluations, assessing and mitigating systemic risks, and reporting on incidents.

Additionally, artificial or manipulated images, audio or video content ("deepfakes") need to be clearly labelled as such.

Measures to Support Innovation and SMEs

Regulatory sandboxes and real-world testing will have to be established at the national level, and made accessible to SMEs and start-ups, to develop and train innovative AI before its placement on the market.

Quotes

During the plenary debate on Tuesday, the Internal Market Committee co-rapporteur Brando Benifei (S&D, Italy) said: "We finally have the world's first binding law on artificial intelligence, to reduce risks, create opportunities, combat discrimination, and bring transparency. Thanks to Parliament, unacceptable AI practices will be banned in Europe and the rights of workers and citizens will be protected. The AI Office will now be set up to support companies to start complying with the rules before they enter into force. We ensured that human beings and European values are at the very centre of AI's development".

Civil Liberties Committee co-rapporteur Dragos Tudorache (Renew, Romania) said: "The EU has delivered. We have linked the concept of artificial intelligence to the fundamental values that form the basis of our societies. However, much work lies ahead that goes beyond the AI Act itself. AI will push us to rethink the social contract at the heart of our democracies, our education models, labour markets, and the way we conduct warfare. The AI Act is a starting point for a new model of governance built around technology. We must now focus on putting this law into practice".

Next Steps

The regulation is still subject to a final lawyer-linguist check and is expected to be finally adopted before the end of the legislature (through the so-called corrigendum procedure). The law also needs to be formally endorsed by the Council.

It will enter into force twenty days after its publication in the official Journal, and be fully applicable 24 months after its entry into force, except for: bans on prohibited practises, which will apply six months after the entry into force date; codes of practise (nine months after entry into force); general-purpose AI rules including governance (12 months after entry into force); and obligations for high-risk systems (36 months).

Background

The Artificial Intelligence Act responds directly to citizens' proposals from the Conference on the Future of Europe (COFE), most concretely to proposal 12(10) on enhancing EU's competitiveness in strategic sectors, proposal 33(5) on a safe and trustworthy society, including countering disinformation and ensuring humans are ultimately in control, proposal 35 on promoting digital innovation, (3) while ensuring human oversight and (8) trustworthy and responsible use of AI, setting safeguards and ensuring transparency, and proposal 37 (3) on using AI and digital tools to improve citizens' access to information, including persons with disabilities.

Let's Not Make the Same Mistakes with AI that We Made with Social Media[59]

Social media's unregulated evolution over the past decade holds a lot of lessons that apply directly to AI companies and technologies.

Nathan Sanders and Bruce Schneier, MIT Technology Review

Oh, how the mighty have fallen. A decade ago, social media was celebrated for sparking democratic uprisings in the Arab world and beyond. Now front pages are splashed with stories of social platforms' role in misinformation, business conspiracy, malfeasance, and risks to mental health. In a 2022 survey, Americans blamed social media for the coarsening of our political discourse, the spread of misinformation, and the increase in partisan polarization.

Today, tech's darling is artificial intelligence. Like social media, it has the potential to change the world in many ways, some favorable to democracy. But at the same time, it has the potential to do incredible damage to society.

There is a lot we can learn about social media's unregulated evolution over the past decade that directly applies to AI companies and technologies. These lessons can help us avoid making the same mistakes with AI that we did with social media.

In particular, five fundamental attributes of social media have harmed society. AI also has those attributes. Note that they are not intrinsically evil. They are all double-edged swords, with the potential to do either good or ill. The danger comes from who wields the sword, and in what direction it is swung. This has been true for social media, and it will similarly hold true for AI. In both cases, the solution lies in limits on the technology's use.

#1: Advertising

The role advertising plays in the internet arose more by accident than anything else. When commercialization first came to the internet, there was no easy way for users to make micropayments to do things like viewing a web page. Moreover, users were accustomed to free access and wouldn't accept subscription models for services. Advertising was the obvious business model, if never the best one. And it's the model that social media also relies on, which leads it to prioritize engagement over anything else.

[59] This article was originally published by the MIT Technology Review on March 13, 2024

Both Google and Facebook believe that AI will help them keep their stranglehold on an 11-figure online ad market (yep, 11 figures), and the tech giants that are traditionally less dependent on advertising, like Microsoft and Amazon, believe that AI will help them seize a bigger piece of that market.

Big Tech needs something to persuade advertisers to keep spending on their platforms. Despite bombastic claims about the effectiveness of targeted marketing, researchers have long struggled to demonstrate where and when online ads really have an impact. When major brands like Uber and Procter & Gamble recently slashed their digital ad spending by the hundreds of millions, they proclaimed that it made no dent at all in their sales.

AI-powered ads, industry leaders say, will be much better. Google assures you that AI can tweak your ad copy in response to what users search for, and that its AI algorithms will configure your campaigns to maximize success. Amazon wants you to use its image generation AI to make your toaster product pages look cooler. And IBM is confident its Watson AI will make your ads better.

These techniques border on the manipulative, but the biggest risk to users comes from advertising within AI chatbots. Just as Google and Meta embed ads in your search results and feeds, AI companies will be pressured to embed ads in conversations. And because those conversations will be relational and human-like, they could be more damaging. While many of us have gotten pretty good at scrolling past the ads in Amazon and Google results pages, it will be much harder to determine whether an AI chatbot is mentioning a product because it's a good answer to your question or because the AI developer got a kickback from the manufacturer.

#2: Surveillance

Social media's reliance on advertising as the primary way to monetize websites led to personalization, which led to ever-increasing surveillance. To convince advertisers that social platforms can tweak ads to be maximally appealing to individual people, the platforms must demonstrate that they can collect as much information about those people as possible.
It's hard to exaggerate how much spying is going on. A recent analysis by Consumer Reports about Facebook—just Facebook—showed that every user has more than 2,200 different companies spying on their web activities on its behalf.

AI-powered platforms that are supported by advertisers will face all the same perverse and powerful market incentives that social platforms do. It's easy to imagine that a

chatbot operator could charge a premium if it were able to claim that its chatbot could target users on the basis of their location, preference data, or past chat history and persuade them to buy products.

The possibility of manipulation is only going to get greater as we rely on AI for personal services. One of the promises of generative AI is the prospect of creating a personal digital assistant advanced enough to act as your advocate with others and as a butler to you. This requires more intimacy than you have with your search engine, email provider, cloud storage system, or phone. You're going to want it with you constantly, and to most effectively work on your behalf, it will need to know everything about you. It will act as a friend, and you are likely to treat it as such, mistakenly trusting its discretion.

Even if you choose not to willingly acquaint an AI assistant with your lifestyle and preferences, AI technology may make it easier for companies to learn about you. Early demonstrations illustrate how chatbots can be used to surreptitiously extract personal data by asking you mundane questions. And with chatbots increasingly being integrated with everything from customer service systems to basic search interfaces on websites, exposure to this kind of inferential data harvesting may become unavoidable.

#3: Virality

Social media allows any user to express any idea with the potential for instantaneous global reach. A great public speaker standing on a soapbox can spread ideas to maybe a few hundred people on a good night. A kid with the right amount of snark on Facebook can reach a few hundred million people within a few minutes.

A decade ago, technologists hoped this sort of virality would bring people together and guarantee access to suppressed truths. But as a structural matter, it is in a social network's interest to show you the things you are most likely to click on and share, and the things that will keep you on the platform.

As it happens, this often means outrageous, lurid, and triggering content. Researchers have found that content expressing maximal animosity toward political opponents gets the most engagement on Facebook and Twitter. And this incentive for outrage drives and rewards misinformation.

As Jonathan Swift once wrote, "Falsehood flies, and the Truth comes limping after it." Academics seem to have proved this in the case of social media; people are more likely to share false information—perhaps because it seems more novel and surprising. And unfortunately, this kind of viral misinformation has been pervasive.

AI has the potential to supercharge the problem because it makes content production and propagation easier, faster, and more automatic. Generative AI tools can fabricate unending numbers of falsehoods about any individual or theme, some of which go viral. And those lies could be propelled by social accounts controlled by AI bots, which can share and launder the original misinformation at any scale.

Remarkably powerful AI text generators and autonomous agents are already starting to make their presence felt in social media. In July, researchers at Indiana University revealed a botnet of more than 1,100 Twitter accounts that appeared to be operated using ChatGPT.

AI will help reinforce viral content that emerges from social media. It will be able to create websites and web content, user reviews, and smartphone apps. It will be able to simulate thousands, or even millions, of fake personas to give the mistaken impression that an idea, or a political position, or use of a product, is more common than it really is. What we might perceive to be vibrant political debate could be bots talking to bots. And these capabilities won't be available just to those with money and power; the AI tools necessary for all of this will be easily available to us all.

#4: Lock-in

Social media companies spend a lot of effort making it hard for you to leave their platforms. It's not just that you'll miss out on conversations with your friends. They make it hard for you to take your saved data—connections, posts, photos—and port it to another platform. Every moment you invest in sharing a memory, reaching out to an acquaintance, or curating your follows on a social platform adds a brick to the wall you'd have to climb over to go to another platform.

This concept of lock-in isn't unique to social media. Microsoft cultivated proprietary document formats for years to keep you using its flagship Office product. Your music service or e-book reader makes it hard for you to take the content you purchased to a rival service or reader. And if you switch from an iPhone to an Android device, your friends might mock you for sending text messages in green bubbles. But social media takes this to a new level. No matter how bad it is, it's very hard to leave Facebook if all your friends are there. Coordinating everyone to leave for a new platform is impossibly hard, so no one does.

Similarly, companies creating AI-powered personal digital assistants will make it hard for users to transfer that personalization to another AI. If AI personal assistants succeed in becoming massively useful time-savers, it will be because they know the ins and outs of your life as well as a good human assistant; would you want to give that up to make a

fresh start on another company's service? In extreme examples, some people have formed close, perhaps even familial, bonds with AI chatbots. If you think of your AI as a friend or therapist, that can be a powerful form of lock-in.

Lock-in is an important concern because it results in products and services that are less responsive to customer demand. The harder it is for you to switch to a competitor, the more poorly a company can treat you. Absent any way to force interoperability, AI companies have less incentive to innovate in features or compete on price, and fewer qualms about engaging in surveillance or other bad behaviors.

#5: Monopolization

Social platforms often start off as great products, truly useful and revelatory for their consumers, before they eventually start monetizing and exploiting those users for the benefit of their business customers. Then the platforms claw back the value for themselves, turning their products into truly miserable experiences for everyone. This is a cycle that Cory Doctorow has powerfully written about and traced through the history of Facebook, Twitter, and more recently TikTok.

The reason for these outcomes is structural. The network effects of tech platforms push a few firms to become dominant, and lock-in ensures their continued dominance. The incentives in the tech sector are so spectacularly, blindingly powerful that they have enabled six megacorporations (Amazon, Apple, Google, Facebook parent Meta, Microsoft, and Nvidia) to command a trillion dollars each of market value—or more. These firms use their wealth to block any meaningful legislation that would curtail their power. And they sometimes collude with each other to grow yet fatter.

This cycle is clearly starting to repeat itself in AI. Look no further than the industry poster child OpenAI, whose leading offering, ChatGPT, continues to set marks for uptake and usage. Within a year of the product's launch, OpenAI's valuation had skyrocketed to about $90 billion.

OpenAI once seemed like an "open" alternative to the megacorps—a common carrier for AI services with a socially oriented nonprofit mission. But the Sam Altman firing-and-rehiring debacle at the end of 2023, and Microsoft's central role in restoring Altman to the CEO seat, simply illustrated how venture funding from the familiar ranks of the tech elite pervades and controls corporate AI. In January 2024, OpenAI took a big step toward monetization of this user base by introducing its GPT Store, wherein one OpenAI customer can charge another for the use of its custom versions of OpenAI software; OpenAI, of course, collects revenue from both parties. This sets in motion the

very cycle Doctorow warns about.

In the middle of this spiral of exploitation, little or no regard is paid to externalities visited upon the greater public—people who aren't even using the platforms. Even after society has wrestled with their ill effects for years, the monopolistic social networks have virtually no incentive to control their products' environmental impact, tendency to spread misinformation, or pernicious effects on mental health. And the government has applied virtually no regulation toward those ends.

Likewise, few or no guardrails are in place to limit the potential negative impact of AI. Facial recognition software that amounts to racial profiling, simulated public opinions supercharged by chatbots, fake videos in political ads—all of it persists in a legal gray area. Even clear violators of campaign advertising law might, some think, be let off the hook if they simply do it with AI.

Mitigating the Risks

The risks that AI poses to society are strikingly familiar, but there is one big difference: it's not too late. This time, we know it's all coming. Fresh off our experience with the harms wrought by social media, we have all the warning we should need to avoid the same mistakes.

The biggest mistake we made with social media was leaving it as an unregulated space. Even now—after all the studies and revelations of social media's negative effects on kids and mental health, after Cambridge Analytica, after the exposure of Russian intervention in our politics, after everything else—social media in the US remains largely an unregulated "weapon of mass destruction." Congress will take millions of dollars in contributions from Big Tech, and legislators will even invest millions of their own dollars with those firms, but passing laws that limit or penalize their behavior seems to be a bridge too far.

We can't afford to do the same thing with AI, because the stakes are even higher. The harm social media can do stems from how it affects our communication. AI will affect us in the same ways and many more besides. If Big Tech's trajectory is any signal, AI tools will increasingly be involved in how we learn and how we express our thoughts. But these tools will also influence how we schedule our daily activities, how we design products, how we write laws, and even how we diagnose diseases. The expansive role of these technologies in our daily lives gives for-profit corporations opportunities to exert control over more aspects of society, and that exposes us to the risks arising from their incentives and decisions.