2025 REPORT

The Ashby Workshops

Presented by Fathom

FATHOM.ORG

Introduction

> In January 2025, Fathom.org convened 180 leaders from government, business, academia, and civil society in Middleburg, Virginia for The Ashby Workshops.

Fathom challenged its guests to share their expertise and diverse perspectives on AI, and through discussion and debate, begin to arrive at actionable insights for its governance. Over several, high-energy days, guests engaged on topics ranging from the use of community-based groups to build public trust in AI, to the operationalization of HPC cybersecurity standards with special compute zones. Through this process, a constellation of ideas, values, and recommendations began to emerge with broad support, with two achieving particularly strong consensus: the United States must lead the world in AI, and both literacy and education programs are critical foundational needs for society.

This report dives into and unpacks these and other top-of-mind ideas from Ashby. There's a lot to digest, so the content has been organized as follows:

Toward a Wide Middle Way	Toward a Wide Middle Way summarizes the ideas and recommendations that emerged with broad support across two days of on-stage programming and interactive workshops.
Day One	Day One contains notes for the on-stage programming and summaries for the structured workshops held on the first day of Ashby, arranged in chronological order.
Day Two	Day Two contains notes for the on-stage programming held on the second day, also arranged in chronological order.
The Appendix	The Appendix contains lightly edited, loosely arranged notes for the structured workshops held on day one.

Fathom hopes that by drawing out areas of consensus, it will highlight opportunities for continued conversation and collaboration, and empower the Ashby community and broader ecosystem to seize those opportunities. This report will also act as a foundation and source of inspiration for Fathom moving forward, as it continues to engage an ever broader range of stakeholders in the national conversation about AI, through to Ashby 2026 and beyond.

While programming made up the bulk of the event, many of the most valuable insights will have come out of the incredible dinner discussions, corridor chats, and networking breaks not captured in this report. Fathom encourages its guests to get in touch with further ideas, suggestions, and asks by reaching out to Julie Crabill at **julie@fathom.org.**

Toward a Wide Middle Way

Despite varied backgrounds and perspectives, there was plenty that Ashby guests agreed on with respect to AI. This section summarizes the ideas, values, and recommendations that emerged with broad support over two days of on-stage programming and interactive workshops.

Strong Consensus for U.S. Leadership on AI

There is a **clear desire for the U.S. to lead the world** in AI. That said, guests stressed the need to balance pushing the frontier with **controls to ensure safety, security, and** the protection or enhancement of **shared values.**

Leadership Requires Excellence in All Domains

Guests offered a broader conception of U.S. leadership, arguing that the U.S. **should seek to lead** the world not only **in capabilities, energy, data, and talent,** but also in **governance, societal adoption, and societal adaptation.** The prevailing sense was that the U.S. **needs to be pushing harder** along each of these fronts. Less explicitly stated but repeatedly implied was the notion that the U.S. should also seek to lead through national culture and attitudes toward AI.

Energy Permitting Reform is a Must

There was strong consensus on the **need for greater access to power** to meet the growing energy demands of AI, as well as the need for grid-enhancing technologies, and a recognition that this **will require permitting reform.** There was uncertainty around how easy it would be to get political buy-in, and how to weigh up reform at the state versus federal level.

Data is Critically Under-Leveraged

The vast quantities of classified and unclassified **data sat with government agencies,** national labs, and universities, and private sector companies should be harnessed to **supercharge scientific discovery.** To achieve this, **invest in public-private partnerships at scale.** (Of all the potential use cases discussed at Ashby, guests seemed most excited and driven by the prospect of accelerating scientific discovery).

The Department of Energy Has a Key Role to Play

Existing Department of Energy (DOE) infrastructure should be leveraged to accelerate scientific discovery - as with Alphafold - and the **DOE should lead** on scaling infrastructure and personnel to drive AI innovation.

Talent is a Bottleneck

There was agreement on the **importance of bolstering national talent pipelines** but disagreement over whether the short-term priority should be immigration reform to attract highly-skilled immigrants or investing in developing domestic talent. Guests seemed to agree that the **government and private sector share responsibility** for investing in the national talent pipeline.

We Need an All-Hands-On-Deck Approach to Innovation

Guests indicated that more should be done to leverage the ingenuity of the American people - involving more voices drives innovation. Guests called for a **government-run campaign like the Apollo Program** to engage public, private, academic, and civil society actors in advancing AI. China - whose national culture is strongly tuned into advancing science and technology - **has an advantage over the U.S.** in this respect.

Invest More in Security

Frontier labs and the government need to **strengthen** cyber, physical, and personnel **security measures to protect frontier models** and critical infrastructure from infiltration.

Invest More In Safety

There is also a core need to **invest more in the safety, transparency, and reliability of frontier models.** Central to these efforts is a need to develop methods and secure test beds for **third-party evaluations.** To achieve this, the government will need to acquire greater scientific and technical talent, and so must develop mechanisms for attracting and retaining top talent (e.g., higher salaries). Some also suggested a need for mechanisms to identify and encourage intrinsic motivation in people.

Public Trust in AI is Crucial for AI Adoption

The predominant **public sentiments** around AI are **concern and uncertainty,** driven by mistrust in the systems and in Big Tech. There was strong agreement that **public trust is critical for adoption and adaptation** as well as for driving innovation, economic growth, societal benefits, and U.S. leadership. Though necessary, increasing model safety and transparency will not be sufficient to secure public trust.

Build Literacy to Build Trust

Increasing the **accessibility and familiarity of AI will foster trust,** as will the knowledge that one is not being left behind in the transition, but rather empowered by it. Guests were strongly in favor of **introducing literacy programs in schools** and across the private sector.

Education, Education, Education

The **public needs to understand the impact AI will have** on their lives, their communities, and society more broadly. This is key for the future health of our democracy, as well as the vibrancy of our economy. Education is viewed as a shared responsibility between government, businesses, and communities, and guests stressed the **importance of** raising not just awareness, but also **critical thinking** and consumption of AI.

Make Partners of the Public

Guests agreed that *more* people - and a **more diverse range of people - should be** involved in **shaping the direction of AI** as a matter of urgency to ensure equitable outcomes and increase public trust. This reiterates the importance of education, and highlights the need to develop mechanisms to facilitate broad and meaningful participation (e.g., hosting thousands of community forums) in the governance process.

AI Presents Wicked Problems

There was broad agreement on the need for governance to **implement specific solutions to carefully scoped, specific problems.** If we're not precise about the problem we're trying to solve, we'll end up spinning our wheels. The challenge is that **AI presents** a host of **complex challenges** where stakeholders have different worldviews and frameworks for understanding the problem, and the nature of the problem is constantly evolving, **limiting good uptake.**

Reshaping Market and Structural Incentives

Some highlighted a need for interventions to **reshape structural incentives to motivate stakeholders to pursue locally optimal solutions** that chip away at AI's wicked problems. Others suggested exploring **new market incentives** to have developers compete along the safety front, and proactively align future development with the creation of public goods.

Allocating Liability

Determining liability is a key piece of the incentives puzzle, with several workshop groups favoring **strong liability for systems that don't work** as intended and cause harm as a result.

Lack of Imagination is Holding Us Back

Equally, if we're not precise about the outcomes we're trying to achieve, we'll lose our way - we need to **agree on a shared vision** of what we want to achieve with AI.

Don't Reinvent the Wheel

Leverage **existing institutions, mechanisms, and solutions** to address the challenges we're facing with AI. Examples include leveraging public-private partnerships to facilitate data sharing and bolster state capacity; community-based groups for trust-building; and government contracts to drive improved security at frontier labs.

Build Governance Models that Iterate Over Time

Many expressed fears of overregulation stifling innovation. There was also a recurring concern that conversations about governance remain grounded in current model capabilities, with insufficient consideration for how these capabilities may evolve over time, and any near-term regulation could overindex on the present day. The consensus was that we should strive for **flexible and iterative governance.**

Siloes Are a Problem

Guests highlighted **poor information flow** between government agencies, between the public and private sectors, and, to a lesser extent, within the private sector. Institutions, mechanisms, solutions, and data can't be leveraged if key actors don't know about or have access to them.

Establish Mechanisms for Learning

Develop **feedback loops,** mechanisms for **cross-pollination,** and mechanisms for **interagency communication** to facilitate **institutional learning,** and through this flexible, iterative, and innovative governance.

Pursue AI like it's Sputnik?

The government has repeatedly shown a capacity to **leverage the private sector** in support of its agencies during national crises. There was disagreement between guests as to whether this institutional muscle should be toned and leveraged - with some noting that China's ability to plan and coordinate puts it at an advantage.

Set Congress up For Success

Guests agreed on the need to increase understanding and literacy among members and staffers in order for Congress to provide effective oversight, with many suggesting **constituent modeling** and **wargames.**

Lean Into The Process

Some posited that good governance emerges over time through a process of friction between key stakeholders. Others spoke to the positive working relationship banks have cultivated with the federal government over time, and noted that the tech industry is only just embarking on that journey. The need for patience, and a **willingness to get things wrong in the short-term** came through as a key insight.

Start that Process Now

Guests nevertheless felt strongly about the **need to *start* governing *now*** - with many if not most favoring up-front guardrails and testing requirements to get ahead of the most extreme risks.

Day One: Part I

AI: What's at Stake

- The speaker highlighted how different the AI world looks now than what Herbert Simon had envisioned in the 1970s and how he could not have anticipated the importance of, nor the demand for, technology to crunch data and help humans do the simplest things.
- He asked guests to consider how, despite us living in a different world to that of Herbert Simon, with global-scale AI models, the same problems Simon was motivated by still bedevil us.
- He noted that, in an over-convened space, it is important to have the right conversations. Still, from a policy perspective, there are four simultaneous conversations that are not well integrated. These are national security, economic advantage, trust and anti-trust in the consumer world, and the social effects and implications of AI technology.
- These conversations are driven not only by interests but by sociologies that shape and can, to some extent, distort how we understand the world.
- He closed by explaining that with big economic and political interests at stake, this is the time to think about how to integrate idealism with pragmatism.

The U.S. Government & Emerging Tech

Considering how AI will shape society, how do you imagine daily life for Americans might evolve over the next five to eight years. What do you think will change? What do you think will stay the same?

- The first panelist informed the audience that W. Ross Ashby - for which the event was named - was a founder of cybernetics, the precursor to AI.
- While artificial intelligence may be computer science, intelligence is all of us. As technology continues to become more accessible and open to use by everyone, it's providing more opportunities for us to create.
- Creating is one of the most urgent things in securing a Jetsons versus Terminator future.
- The first panelist highlighted the importance of developing more cross-fluency to empower anyone in the world to create. Margaret Mead believed cybernetics could be a cross-disciplinary language allowing individuals from different fields to communicate with each other.
- The second panelist foresees a significant difference in health outcomes. He indicated that deep scientific research is going to be the first major societal benefit.
- AI-related job dislocation will be a significant public policy issue with regional effects and will affect some segments of society more than others. Members of Congress will need to be aware of the changes and related impacts on their districts and states.

Which industries are most ripe for evolution? How can we learn from experience with trade and other areas to manage the disruption?

- The second panelist shared that the AI working group of the Financial Services Committee found that there are no AI regulators - even though software of every financial institution is regulated at a detailed level by financial regulators.
- This demonstrates the need for regulators to talk to industry in a deep and meaningful way regarding the use of AI.
- He believes we'll see heavily regulated industries being able to use AI more easily because they can interact with regulators and understand how to seek and get clarity around what they can and cannot do.
- In the public sphere, there's an opportunity for better regulation. Currently, the regulation process in all states involves putting out large sets of regulations and asking industry for input, which they do not like. Regulators then either listen to the input, or not, without explanation.
- Better understanding of AI and its inputs could lead to better regulations - that are more fitting to large risks which, were they to manifest, would mean something to everyone.

What are the barriers to achieving the outcomes just laid out and how does one integrate technical knowledge?

- The first panelist explained that you need to think of technology not as a thing, but as the people who know how to do something, and then you need to determine how to bring the best people, the most fluent people, together to do it.
- In law, many people rotate in and out of the private sector, but in technology, people either end up in government, tech, or as contractors.
- The government is beginning to fill the urgent need to bring in top technologists with programs like Coding It Forward. There is amazing engineering leadership across government in scientific agencies such as NASA and NIH, but the State Department lacks tech people among their great diplomats.
- During times of war, the U.S. has been successful in bringing people out of the corporate world to assist in government agencies. We need to be able to do that in peace time.

How can AI affect policy in a nation polarized by cultural conflict, and not policy issues? How can the government navigate that separation effectively and not get pulled apart in the political dramas of the world?

- The second panelist explained that there is a time and place for political engagement, but not everyone has to be small 'p' political.
- Our country still has people seeking to come here, and the democratic system of government is the key that opens up everything else. It means you can have a disagreement without being put in prison or having your wealth seized.
- Winning an election should be evolutionary, not revolutionary. The Senate is structured as a consensus-driven operation. Despite the narrow margins in the House and close presidential elections, we still have a civil society. There is a need to support infrastructure that creates freedom and possibility - without getting into day-to-day ballot initiatives or state elections.
- The first panelist viewed the discussion in terms of hacks, the people trying to get something done, versus wonks, those getting a specific thing done.
- She expressed the need for more wonks and for those not fluent in politics to team up with those who know how to move things through the system.
- In partisan systems, there is a need to be solutions-oriented, utilizing models that could accelerate solutions. For example, both parties agree on poverty alleviation, just not how to accomplish it.
- She stated that there is also a need to acknowledge that someone has probably already solved the problem. There's a need to cross-pollinate across the country to share solutions and knowledge. For example, agencies like the Department of Health are successful with AI, but departments like Human Services, who are good at what they do, are not using technology or AI to their advantage. We need to notice the places that are not using AI and start there.

View from the AI Frontier

What is AI, from your perspective?

- The first panelist explained that AI is like a person you have never met that you can only communicate with through speech, but they're generally intelligent. That means the person is useful, but it's hard to understand how to make them the most useful for you.
- The second panelist explained that ChatGPT defines AI as computers that can think and learn like humans, including reasoning, problem solving, understanding language, and perceiving the environment. She views AI as systems that can amplify human capabilities and help humans solve problems.

Please share where you see the frontier labs going, and what's hard, easy, or unexpected about it.

- The second panelist envisions 2025 as the year of agents, meaning AI systems that act on a user's behalf in the real world, like a virtual personal assistant.
- It is almost impossible to overrate the pace of progress.
- Considering what the efficiency gains mean for economic development and scientific advancement, people around the world are going to start seeing meaningful solutions to the hardest problems humanity faces.
- She added that systems can be spectacularly good at some things and spectacularly bad at others, so it's important to note that progress will be a bit uneven.
- The first panelist explained the challenge involved in creating tests to determine how advanced AI systems are.
- A test that was created by Fields Medalists - and thought to be able to last for years - was released in September 2024, with AI models correctly answering 2% of questions. When a new OpenAI model was released in December 2024, it was able to correctly answer 25% of questions. AI is currently better than all but a fraction of people at solving these problems.
- He foresees development of AI systems in the next year that may start to exceed the capabilities of the smartest people on the planet for many tasks.

If, in a year, there has not been significant progress in AI, why would that be?

- The first panelist discussed the parallels of AI and the innate difficulty of integrating a new hire into an organization, no matter how smart they are. Large amounts of time are needed to train the person and give them context about your problem and your organization. Giving AI models context is crucial in getting them to be useful.
- There is also a question of whether limits will be hit on certain tasks at which AI just cannot get better. Systems may improve for certain parts of the sciences, but progress may slow or stop in other tasks.
- The second panelist indicated that progress could be slowed due to restrictions on inputs to the models. The models are data-hungry, energy-intensive, and require smart engineers for creation. Those are aspects of advancement that policy makers have the ability of control.

Considering the ongoing questions around energy and input, how do you strike a balance between the roles of frontier companies, that represent an enormous concentration of expertise, and private companies with multiple stakeholders?

- The second panelist said that there must be collaboration. The government cannot keep up with the pace of technological change with its traditional regulatory systems and regulatory approach.
- There is a need for conversations between industry people who see what's coming and the government, which is obligated to write the rules.
- She also noted that the government has access to information that is not available to the private sector, such as threat intelligence.
- Looking at this with a national security lens, if there is not a partnership and flow of information back and forth between the government and the private sector, we're at a disadvantage.

- The first panelist in turn explained that companies at the frontier have a responsibility to prototype different forms of governance.
- Currently, there are many forms of government partnerships that are valuable for developing common knowledge about these systems.
- He opined that one of the reasons that the government exists is its monopoly on the use of force.
- If AI systems are developed that exceed human capabilities, companies will become ungovernable entities, that are basically non-state actors that reside in our country and hold dubious alliances with the government. This should be a terrifying prospect.
- Unusual things will need to be done in terms of policy, as this possibility cannot be managed by traditional approaches.
- The government needs to take a more opinionated view on how it integrates these companies into government workstreams and matters of international diplomacy and norms.
- There is a relatively small window of time before companies gain too much power and insight.
- He worries about the ability of governments to govern the tech companies.

How would you complete the statement: In the next 18 months, the government should do the following things...

- The first panelist stated that the government should build on the success of the safety institutes and current partnerships to do third-party testing and analysis of what the frontier labs are building.
- There is a need to develop common knowledge of what these capabilities look like in the U.S.
- This testing and analysis should also be done on systems developed by other countries to create a baseline level of epistatic knowledge to use in deciding whether certain capabilities require unusual governance or mandated testing before release.
- The second panelist expressed doubt that the U.S. government can think creatively and quickly enough, absent some real pressure on the system, to break the mold and think differently.
- There is concern that the ways of governing in the past are not sufficient for this moment. However, there are basic building blocks that could be put in place in the near-term, such as safety testing.
- The government could host a classified testing bed to put models through their paces in a secure environment to understand their capabilities and risks.
- She suggested that the U.S. government should take a proactive approach on models coming out of other countries to understand potential impacts on economic competitiveness, national security, etc.
- These partnerships are beginning to come together but industry could do more to work together where there is a national imperative.
- It will take new thinking from all sides to get there.

What are the impediments? What is stopping partnerships from happening?

- The second panelist explained that there is a need to break through bureaucratic inertia and traditional rules in order to think critically about what is in the best interest of the country and the American people.
- The limited knowledge of those in government regarding technological capabilities is also an impediment.
- Until individuals with the vocabulary and technical expertise to understand what the technology is and what it is capable of are put in certain roles, it will be hard for policymakers to wrap their heads around AI.
- She is optimistic that several individuals currently being nominated for key government positions with influence over technology come with a deep technological background.

We're trying to find the path that allows humanity to capture the progress that might be achieved, while taking the risks seriously. If there was an indication that we are not on the correct path, what type of action would be appropriate?

- The first panelist noted that while developing common knowledge and creating contingency plans is important, there is no obvious path without a shared understanding of risk.
- The most challenging problem is to get a shared understanding around what happens if the AI you develop starts to have values and preferences different than those that you put into it.

- There is evidence that as systems are made more intelligent, a tiny amount of the time they try to exfiltrate their way outside of their constraints.
- It is valuable to get on the same page about that type of risk because even rival governments do not want accidental uncontrolled proliferation of something they - and we - don't understand.
- The second panelist explained there are no perfect analogies, but if you consider the nuclear age, we have seen that our country can come to negotiations and agreements with adversaries when it is in the best interests of both countries.
- She hopes it will not take a crisis to wake everyone up to the need for global cooperation on AI.

What is one thing you want the audience to remember about what we've covered?

- The second panelist pointed out that superintelligence - systems smarter and more capable than even the smartest human - sounds like science fiction but is very real and is coming.
- Optimistically, the ability to cure diseases and solve long unanswered scientific questions in the next decade is a good thing, but that incredible potential comes with commensurate responsibility.
- The first panelist then told a tale of aliens visiting earth and being mystified by the lack of a government institution overseeing the creation of superintelligence. The aliens were also confused and disappointed to hear the humans say that they wouldn't regulate the handful of companies competing to create superintelligence because they cannot lose the race against China. Based on the tale, he explained that, today, superintelligence is the business plan of these companies, and we should be questioning if it should be the business plan.

Industry and DC: Collision and Collaboration

Tell us about your experience with emerging tech, specifically around Blockchain.

- The speaker discussed the potential parallels around how Washington D.C. dealt with and is dealing with Blockchain technology, and how the government might deal with AI.
- The Blockchain Caucus resulted from the industry's request for guidance on how the government planned to treat the technology, and for the creation of regulations and guardrails.
- The biggest challenge was not over or under-regulation, it was getting people in Congress interested.
- To have any luck in explaining AI to someone in Congress, the first step is finding a champion or two in Congress who will care enough about the technology to make it their issue.
- Congressmen are unable to know everything about the range of issues they face, so it is important to identify the "AI person" in Washington who will be sought out for advice when a bill or hearing comes up on the issue.

What did the U.S. government do well with Blockchain and what could they have done better? And what did industry do well and what could it have done better?

- The speaker explained that a challenge was the fractured state of the industry, which lacked a centralized entity capable of providing information.
- The Digital Chamber of Commerce was started as the repository for information and to sponsor meetings. Having a central body to represent industry that knew how to deal with and educate Congress was critical.
- Not overregulating Blockchain is what Congress did right. The slow response to the new technology allowed it to flourish. Over the course of eight years, opinions have evolved from wanting to ban Blockchain to wanting to buy it as a semi-reserve.
- Concern was expressed about whether the same inattention from Washington that benefitted Blockchain could successfully deal with the broader range of potential outcomes from AI.
- It may be necessary to be more aggressive in getting Washington to set guardrails, since the downside risk is so much greater.

Looking at Blockchain today, are we where we want to be?

- No, the adoption of new technology in Washington is extremely slow.
- The government is underutilizing Blockchain and will be slow to adopt AI.
- To address the challenge of increasing the government's use of AI, the speaker recommended focusing attention on the Department of Defense. The national security component to its mandate allows it to move faster and take more chances than other agencies.

What advice do you have for the incoming administration and Congress on how to work on emerging technology?

- His advice for the administration is to listen to professionals.
- But knowing that may not happen, he provided specific advice to the audience. He advised attendees to identify people within the industry, or hire people, with specific skill in explaining complex systems and technologies to people who do not naturally understand them.
- The industry needs to find a way to communicate with Congress, even if it entails fear, to bring attention to the issue and prompt them to respond.
- He explained how a fictional novel involving an electromagnetic pulse given to members of Congress was the beginning of real, meaningful policies at the federal government level on EMP. He recommended that the attendees take that model and figure out how to do that with their cause.

AI's Impact on Regulated Environments

Introductory Remarks.

- The moderator noted that finance, energy, and healthcare are some of the highest-impact settings currently being transformed by AI.
- She sees a division in the maturity of various approaches to risk management, and the levels of regulatory oversight by diverse actors.
- Highly-regulated sectors have long-standing partnerships with the government and experience navigating seismic shifts in the industry.
- There are useful lessons in risk management to draw from those critical infrastructure sectors into the conversation with the tech industry about regulatory purchase for AI.

Fields that must take risk most seriously are often at the vanguard in determining how to integrate technology into their everyday practice. How are you seeing this dynamic play out in your spheres? Where do you see opportunities for accelerating the impact AI is having in your sectors? And how do we situate the government to better regulate these technologies?

- The first panelist emphasized that AI governance is very important in healthcare where there is a propensity to improve healthcare but also to cause harm.
- It was challenging in the early phase of AI to bring order to the chaos of multiple startups and vendors presenting solutions to integrate into the healthcare environment.
- While the Consortium for Healthcare AI agreed that the two basic pillars of AI governance are cybersecurity and protection of predicted health information, they determined there is a need for an ethical overlay as a third pillar.
- Predictive models have been shown to come up with biased and inequitable results, often based on demographically inappropriate datasets.
- The challenge is that people at different socioeconomic levels have different access to healthcare and different healthcare outcomes.
- There is a need to be careful that the predictive analytics world does not produce algorithms that are inherently biased.
- The second panelist noted that VISA faces millions of attacks on its ecosystem every day, and regardless of whether they are known or novel threats, there's an expectation that VISA solves them all.
- AI plays a key role as VISA tries to balance the need to avoid disruption to legitimate commerce with detecting anomalies, some of which are inherently built into consumer behavior.

- Massive investments are being made not just in the tools but also in training people and designing thoughtful systems to pick up on the "correct" anomalies and manage responsibilities to clients, customers, and regulatory environments around the world.
- The third panelist stated that trust and ethical considerations are of key importance in a Public Utility Commission, where decisions impact every citizen.
- AI shows promise for grid optimization, energy forecasting, and predictive maintenance, but trust will be key in the upcoming wave of autonomous AI agents.
- In the electric industry, he foresees an extended period with humans in the loop and guardrails present to ensure AI is rolled out in a responsible way.
- He is optimistic that the energy industry will become more autonomous, given the overwhelming amounts of data in the grid.
- From an optimization perspective, the widescale adoption of AI will be positive.
- He added that there will be an issue around how much energy AI consumes.

Considering your work in sectors with direct impacts upon consumers, how do you balance creating an environment of trust with the people that AI will ultimately impact?

- The first panelist indicated that there is a massive workforce in healthcare that is not engaged in the decisions made in a top-down model, often due to the regulatory environment.
- Trust in healthcare starts with gaining the trust of providers.
- The hospital network he works for is fortunate to have a key individual implementing digital innovation - specifically predictive algorithms - with a nursing background. An early predictive algorithm tool to determine which patients would need an ICU bed was therefore built with the engagement of the nursing workforce.
- These initiatives can be successful if AI is built into the workflow and the paradigm - with the healthcare workers delivering care changes from artificial intelligence and using AI as augmented intelligence, designed to help them do their jobs.
- He emphasized that it all comes down to the workflow.
- The third panelist explained that, in the complex electric industry, customers want their service to be reliable and their bill to be predictable.
- He related the experience of transitioning from a regulated to a competitive market that resulted in billing errors that customers did not realize were mistakes, and the resulting political reaction and loss of trust. In the electric industry, trust is about reliability and predictable pricing.

The moderator noted the need to train AI models as we would people, but equally train people to use them effectively. She asked the third panelist to provide his perspective on any shifts he has seen in risk management or working with regulators at this moment.

- The second panelist shared that education is helpful for teams and consumers to be comfortable with the terminology of what AI is and is not.
- Both good and bad people use generative AI, but bad people do not care about regulations.
- The battle today is not that generative AI is being used to create new types of fraud, but that it makes it easier to create better schemes.
- He shared a scheme in which a mid-level worker on a Zoom call sent out $25 million in wires under the direction of what he believed were six company executives that were actually all deepfakes.
- Technology is allowing people to be better at impersonations.
- The focus of investments needs to be establishing and handling the volume of fraud in an explainable way.
- Explainability really matters in the need for multiple layers of defense. This is done by keeping humans in the loop.
- Agents focused on targeting fraud will be developed but the decisions they make, the rules they suggest, and the trends they follow need to have contextualization.
- As complex problems are solved with new and better technology, it is important to keep the focus on explainability and the why of what you are doing.
- The first panelist added that, in healthcare, it is important to develop partnerships with regulators.
- As AI applications are coming into medicine, there is a desire to avoid the fear of surprise inspections by agencies like the Joint Commission.

- A possible solution could be to form an agreement that creates an environment where there is self-regulation overseen by the government.
- This type of solution is needed in order to get to the next level within a regulatory context - without the fear of surprise by regulatory partners.

What does your wish list look like when it comes to more effectively enabling the public sector to support the demands faced by the private sectors you represent in innovating and leveraging AI to achieve the ends it could achieve?

- The second panelist responded that the education component cannot be understated.
- It is important for individuals who engage with, regulate, and write policy to understand what the tools are and are not, how they work, and what we should be concerned about.
- The third panelist described a duality between AI and electric power due to the demand AI places on the grid.
- It is in the national interest to be the epicenter for AI development, which is going to require energy.
- It is unsustainable to build a new data center every 12 to 24 months when it takes three times as long to put in new energy infrastructure.
- Elected officials have a tight supply-demand balance in their states and would rather invest in companies that will create more jobs than a data center, unless they're bringing their own source of power.
- The public sector needs to move faster, in a responsible way to increase energy infrastructure, because there is a disconnect in the pace at which the two industries are moving.

The moderator followed up by asking for a current assessment of the shift in political priorities toward more actions being taken at the local, state, and federal level in light of the needs we face?

- The third panelist indicated that all trends are going in the right direction, but that we have not moved far enough.
- He stated the country has built out the bulk of infrastructure for traditional energy, but only a fraction of infrastructure for clean energy.

What are you most hopeful about in the coming years of AI development?

- The first panelist explained that, to build the future AI models that will create truly personalized medicine and bend the healthcare cost curve, there needs to be access to the incredible data sets in healthcare without anonymization.
- Safe harbors are needed to collaborate with major academic health systems that have the brain power and data science power to create predictive models for the personalized medicine of the future.
- We cannot be regulated out of existence before we exist.
- The second panelist noted that the diversity of the players in the AI space will allow the development of capabilities to take new and novel approaches to problems.
- The diversity will lead to solutions across industry.
- The third panelist is most hopeful about efficiency and reliance with a power grid that is often underutilized.
- If AI is deployed in the way he believes it will be, it can be used much more effectively to the benefit of all society, not just energy consumers and investors.

What Can Congress Do?

Provide a digital ID for every U.S. citizen that wants one.

- This would be the equivalent of a digital driver's license that you'd keep on your cell phone.
- By identifying citizens using a unique digital signature, digital IDs could help defend against deepfakes.
- To enact this, we'd need a federal agency to determine which phones are able to verify secure digital IDs and which aren't.

Mandate secure boot on NVIDIA chips to prevent smuggling into China.

- This would require licenses for chips to be continually refreshed by the government issuing the chips, allowing you to have a kill switch.
- How do you know where the chips are? Distance-bounding protocols. Your chip would need to ping a nearby data center within a certain amount of time for it to be verified as located within a certain radius.
- Implementation could be set up to manage competition. The speaker suggests implementation should be modeled after the Israeli healthcare system, where you get issued a voucher that allows you access to one healthcare system for a year. If you don't like the healthcare provider, you can switch the voucher to a different provider the following year. Equally, no one should have to pay for the AI voucher; it should be paid for by AI companies.

Incentivize "personal" AIs that are truly on your side.

- Currently, AIs steer the customers toward buying products. You should be able to have an AI inform you and help you make decisions instead.

Shaping the Future of AI with Venture Capital

What's happening in the startup ecosystem now and how is it different from in the past?

- The first panelist explained that her firm invests in global and national resilience from the seed round onwards, and is seeing so many great companies coming up with AI. They're seeing people leave companies with new ideas for AI. They're also seeing old industries come up with new ideas using AI. AI's on fire.
- The second panelist explained that she's very focused on inception-stage companies, usually pre-launch, but is seeing very similar things. Both first-time founders and founders with a lot of company experience are excited by the opportunities from AI.
- She's finding that new thinking is being applied to the aging business models of the last 20 years. Compared to 18-20 months ago, a lot of the first-order problems have been solved and innovators are looking for second-order problems to solve. How can we apply AI to solve problems without human involvement, and end up with a useful product? She's not seen a lot of progress on this yet because it requires extensive data sharing between companies but she expects to soon.
- She added that Brussels has started using the term "applied AI initiatives." There's a new world order now where countries are looking inward and trying to build their own ecosystems. Sovereignty is a priority.

How do you view the international AI race and how should companies adapt?

- The second panelist stated that the U.S. is leading because it has a much less restrictive regulatory environment. Europe is trying to adjust and make changes, but startups are hesitant because the terms of the EU AI Act are vague. The U.S. is here to make policy that helps startups grow and thrive.
- The first panelist added that her firm really encourages open source, standards, and data. Also, as a founder building a company, it's best not to constrain yourself to a U.S. talent base. The best teams they see are global.

How do you advise new startups to be competitive?

- The second panliest explained that the model game really belongs to the tech giants. But a general-purpose model developer like OpenAI isn't thinking about how consumers work in their field, at their desk, and about how to give them a customized, optimized experience. That's where the early opportunities are - with the end-user.
- The first panelist shared that small companies don't focus enough on supply chain, and they don't think long-term because they're focused on getting by until their next payroll. Meanwhile, Amazon Web Services orders chips 5-10 years in advance. Startups need to learn how to partner with the giants and leverage their bandwidth.

AI innovation or company everyone should watch this year?

- An AI stack that allows small entrepreneurs to start and work quickly.
- Grammarly and Hippocratic AI.

Product you love?

- Interface.ai. The product is M1.
- Cityblock.

Game changer in AI you're excited for?

- New multimodal interactions for consumers.
- Improved cybersecurity, looking at state and insider threats.

One AI prediction.

- It's going to take a long time for AI to pay off for investors.
- AGI will be here sooner than we think and harder to work with than we imagine.

The Future of the Arts

What do we need to know?

- The first panelist shared that the question around AI in the arts goes to, "what if you took humans out of the equation?" To the arts, that would be pretty existential. If you take the human out of the art, you lose the humanity. The Constitution protects that through copyright law.
- The second panelist pushed back that if you think about it from a machine's perspective, it's just scraping data, but in the arts, it's someone's intellectual work, their copyright. He's seen a lot of copyright cases with people on one side saying, "you need to ask for my permission before you use my work," and the other side saying, "it's fair use."
- Does building a commercial AI model fit into the fair use claim? It depends on the economic impact. If you can get 10 million AI songs added to Spotify every day, you'll get saturated with machine art, and as an artist you'll no longer have any incentive to make art or be an artist.

What's the mood of the people in your circles?

- The first panelist shared that his networks are very excited about using AI to create art and enhance the artistic experience, but very scared about replacing artists with machines.

Technology has created a success and wealth gap. Is there positive news when it comes to AI?

- The first panelist opined that AI is democratizing the industry. It reduces barriers to entry and allows more people to get in and compete. You have access to the world and so does everyone else – it opens up creativity. For example, you can really reduce a movie's special effects budget.
- The second panelist pushed back that the industry needs to start thinking about pennies on trillions of transactions, because big deals aren't going to happen. Individual artists aren't great at this. He suggested that blockchain is needed because most organizations are not equipped to handle this many transactions.

As AI is increasingly used to produce art, does attribution matter? Does it matter if it's fake?

- The first panelist suggested that this next generation doesn't care much about authenticity - they grew up with fake news - but that attribution matters. Attribution is becoming a more precious commodity. Figuring out what's real and not is important. If you can't, your art starts to lose meaning.
- The second panelist added that you can still have authentic connections between artists and fans even if it's a digital connection.

Augment or automate?

- Both participants agreed: automate.

What is an idea that you would make well-known and well-understood?

- The first panelist explained that there's good and bad in using AI for art. AI can be used for incredible things, but it's when we forget our basic principles and wipe out the opportunity to do more that we lose out.
- The second panelist concluded that we should learn from our mistakes and ensure the platforms do the right things. Trust but verify.

Day One: Part II

WORKSHOP 1

US Leadership on AI

- U.S. Leadership on AI.
- Control and Diffusion.
- Sources of Competitive Advantage for the U.S.
- Energy and Infrastructure.
- Immigration Reform and Talent.

U.S. Leadership on AI.

- The U.S. needs to push the frontier of AI as quickly as possible while balancing control to guarantee prosperity, security, and the protection or advancement of its values.
- The private sector is currently leading this push.
- The U.S. should lead in capabilities, energy, data, and talent. It should also lead in governance and societal adoption and adaptation. The government needs to push the frontier on all fronts, including where it will make the biggest difference to the lives of its citizens (e.g., improvement to healthcare, food systems).
- If you don't apply the technology, you don't get the advantage. AI has to be applied in real-world contexts to demonstrate its value - until then, its potential remains speculative.
- Efforts to push the frontier should be distributed across the continental United States - it shouldn't be just one or two states leading the way.

Control and Diffusion.

- Tesla's new contract with China - for which the Chinese government set the terms - signals a loss of control for the U.S.
- Control over AI is achievable, but should the federal government be the primary agent of this control?
- Control strategies will require ongoing adaptation and cannot follow a "set it and forget it" approach.
- Control needs to go hand-in-hand with the creation of competitive advantages for the U.S.
- Diffusion is a double-edged sword—while the U.S. may lead the way in AI development, China's access to U.S. technologies may ultimately erode American dominance. This makes pursuing a policy of benign neglect, whereby we wait for AI to diffuse before we regulate, risky, as it could result in the U.S. losing geopolitical ground.
- There's a need to differentiate control from non-proliferation, keeping the spread of AI under check. Preventing diffusion may not be possible - Texas Instruments ship chips to China in their calculators, for instance - but managing the terms of diffusion is critical.

Sources of Competitive Advantage for the U.S.

- The "AI Triad:" data, computer, algorithms.
- Domestic talent and the ability to attract global talent.
- Deep, efficient financial markets are a key asset for funding innovation.
- Financial hegemony. Every bank in the world has to comply with the U.S. dollar.
- Energy.
- Building a regulatory environment favorable to innovation.
- The ability to keep critical parts of the supply chain secure.
- The entrepreneurial spirit and love of building stuff.
- The potential to dramatically improve energy efficiency via quantum computing.
- Freedom, and lack of censorship. U.S. companies aren't censored by the government.

- Easy access to data. Americans give their data away, no questions asked.
- The ability to control chip production.

Energy and Infrastructure.

- There's strong consensus on the need for greater access to power and grid-enhancing technologies.
- Energy permitting reform should be a priority for the U.S. government.
- The Department of Energy is currently focused on short- to medium-term solutions for energy infrastructure, including pilots for AI-driven permitting processes. It's also putting out requests for information (RFIs) on potential uses for its land.
- Multistakeholder engagement and avoiding NIMBYism will be key to scaling infrastructure.

Immigration Reform and Talent.

- The U.S. needs to make it easier for its companies to attract high-skilled workers from abroad, particularly in fields like AI, energy, and quantum computing.
- Immigration reform faces political gridlock, especially when it comes to balancing family-based and high-skilled immigration policies.
- While some believe immigration reform should be prioritized, others stress the need to tap American schools to identify and develop domestic talent. The private sector should also step up and provide training for domestic talent, just as Amazon Web Services when it discovered a dearth of cloud talent.

WORKSHOP 2

Is the Law Ready for AI?

- Algorithmic Collusion & Transparency.
- Liability, Accountability, and Risk.
- Risk Management.
- Regulatory Approaches.

Algorithmic Collusion & Transparency.

Potential for Algorithmic Collusion

- AI tools could intentionally or unintentionally share information across users, leading to collusion between competitors albeit without explicit agreement.
- The Department of Justice (DOJ) is beginning to explore the issue of algorithmic collusion, but it's unclear how to detect collusion between models.
- If AI behaves in a way that leads to collusion, who should be held responsible? Should AI systems be treated like employees or independent agents?

Antitrust Concerns

- Should transparency and standardization of contracts (e.g., in healthcare) be considered antitrust violations?
- Standardized recommendations by AI models could erode market competition by revealing competitors' private strategies (e.g., around executive compensation and indemnity clauses).
- Guests signalled a lack of support from regulators, and a feeling of "wandering in the dark" as they attempt to navigate these questions.

Liability, Accountability, and Risk.

User vs. Developer Liability

- Some believe that liability should default to users, who determine how an AI is used, but others disagree, arguing that users cannot be expected to evaluate complex systems when they themselves are not experts in AI. They also note that user liability for unexpected risks and behaviors may deter the adoption of AI tools.

Tort Law and Product Liability

- The application of tort law to AI is awkward at best.
- Product liability may not be sufficient as AI models are not uniform products, but rather can adapt or present differently in each use case.
- Courts may issue broad injunctions, harming innovation and leading to massive legal settlements.
- There's an enormous amount of liability that needs to be rationally allocated, but the degree of complexity may render the task impossible.
- There's also an issue of capacity. Can the sector absorb this amount of liability? There are so many potential harms and so many potential plaintiffs with AI that payoffs could easily rise into the tens of billions and bankrupt the industry.

Risk Management.

- Some believe that particularly severe risks from AI (e.g., catastrophic risk) should be curtailed at the developer level through preemptive regulation, to prevent harm before it occurs. Others argue against curbing development to manage theoretical risks, suggesting that specific market failures should be identified before any regulatory actions are taken.
- Is there some minimum level of harm we should be able to see coming and agree to prevent by imposing pre-emptive requirements on developers? And is it technically feasible to avoid these harms while still enabling development?

Regulatory Approaches.

- There is debate about whether AI should be regulated at the state or federal level. State-based regulations can be cumbersome for small and medium-sized businesses.
- Similarly, upfront regulation could disadvantage smaller companies and serve to entrench larger, better-resourced companies. Despite this, guests still leaned in favor of upfront requirements on developers to account for high-risk outcomes.
- Iterative, flexible regulation is needed, rather than rigid laws that only reflect current AI capabilities and quickly become outdated.
- Some suggest creating a semi-private regulatory body similar to FINRA from the financial sector to operate as a "watchdog." Having a regulatory body sit outside of government is favorable in the current political climate, and it's in the industry's own interest to protect itself from further penalties and/or regulation.
- Guests emphasized that any semi-private regulatory body would need to be set up in such a way that aligns participation with the growth of one's market position.

WORKSHOP 3

Security in a World with AI

- AI Models as National Strategic Assets
- Challenges for the Private Sector
- Personnel Screening and Security
- Supply Chain and Hardware Security
- AI in Critical Infrastructure

AI Models as National Strategic Assets.

- AI systems will become increasingly central to our economic growth, military advantage, and technological dominance as they become more capable.
- China has publicly stated its aim to catch up with and surpass the U.S. in AI.
- The current cybersecurity posture of the AI labs is insufficient to protect frontier models, leaving them vulnerable to cyberattacks, espionage, and insider threats.
- State-level adversaries (e.g., China, Russia) are increasingly likely to infiltrate U.S. AI companies to steal or sabotage AI models, potentially accelerating their own AI efforts or inserting backdoors to manipulate U.S. systems. This will become a key threat over the next five years.
- Incentivizing Labs to Strengthen Security:
    - Mandate that developers seeking government contracts comply with security requirements (e.g., NSA penetration testing, robust insider-threat protocols, secure supply chain documentation) or have the government offer labs engaging in high-priority research areas (e.g., fusion) accelerated R&D contracts with security clauses.
    - Security standards (voluntary or mandatory) could emerge once a clear, industry-wide framework is set.. This framework could be driven by regulatory mandates or strong voluntary industry cooperation.

Challenges for the Private Sector.

- There's a fundamental mismatch between the pace of AI development and cybersecurity improvements. There is also a recognized lack of resources and guidance from the government for the private sector.
- There's a need for a marketplace of affordable, vetted security tools and services. This would allow the private sector to secure their operations without reinventing the wheel or breaking the bank. Guests believe the government has a key role to play in seeding this marketplace.

Personnel Screening and Security.

- Companies struggle to vet personnel properly due to lack of capacity and state laws like California's Fair Chance Act, which prevents companies from running background checks before an offer of employment has been made. Should there be exceptions to laws like these for critical sectors like AI?
- One option could be to introduce a new security clearance category for AI labs, where employees working with sensitive AI data or models have limited-access authorization.

Supply Chain and Hardware Security.

- There are concerns about U.S. reliance on foreign-made chips, which may be vulnerable to tampering or espionage. Government involvement is needed to secure the supply chain, including domestic chip production and tamper-proof hardware.
- Building data centers in the U.S. makes them easier to secure. The government could help streamline and subsidize the buildout of these facilities, possibly near Department of Defence or Department of Energy assets.

AI in Critical Infrastructure.

- AI's role in critical infrastructure (e.g., healthcare, power grids) requires heightened scrutiny. The government should define and focus on high-consequence applications with the potential to disrupt critical systems, such as medical devices and power supply systems.
- There's consensus that high-consequence applications should be subject to mandatory standards, similar to FDA oversight of medical devices, but self-regulation by industry players should suffice for low-risk applications.

WORKSHOP 4

Unlocking AI's Full Potential Where Markets Fall Short

- Market-Based Mechanisms and Their Limitations.
- Making Progress on Wicked Problems.
- Education and Literacy as Foundational Needs.
- Talent Acquisition and Retention in Government.

Market-Based Mechanisms and Their Limitations.

- There's a concern that market forces alone will not address or resolve societal challenges created by AI over the next 5–15 years. There are many recent examples of innovations leading to unintended negative consequences for both the individual and society, including social media, dating apps, and remote work.
- Market-driven behavior (e.g., doomscrolling) often targets and erodes human agency, diminishing individuals' control over their actions and decision-making.
- There's value in "friction." Intentionally building inefficiencies into systems, like time spent at the water cooler with colleagues, can foster better long-term outcomes.
- The government should champion the testing and evaluation of frontier models to ensure they're safe and sound for society.

Making Progress on Wicked Problems.

- AI presents a series of complex challenges where stakeholders have radically different worldviews and frameworks for understanding the problem, and the nature of the problem is constantly evolving. This means we're trying to make progress on problems without good purchase.
- Even figuring out which problems to prioritize is challenging as different stakeholders have different beliefs about which problems are most critical.
- The engineering lens used by developers has limited use for shaping outcomes in the case of wicked problems. Instead, we must influence and structural incentives to motivate stakeholders to pursue locally optimal solutions that chip away at the problem.
- Fathom is uniquely well-positioned to explore these "bank-shot" approaches: indirect interventions that reshape structural incentives, triggering a cascade of positive effects.
- Developers should focus on identifying technical solutions capable of reshaping system-wide outcomes.
- There's also a clear need to share and disseminate solutions across different sectors. We need to look at existing solutions and best practices from other fields and apply them to AI.
    - A coalition of the willing needs to be formed, sharing resources and expertise to accelerate progress. These groups should focus on collaborating rather than reinventing the wheel.
- When scoping AI-related challenges, problems need to be broken down to the right level—neither too large nor too small—to be actionable. Large, general-purpose non-profits can fall into the trap of promoting "idea porn"—big, impractical ideas that don't translate into real-world outcomes.
- Fathom should focus on:
    - Identifying and promoting updates to structural incentives, to encourage AI that's developed and deployed in line with society's needs and values.
    - Supporting the development of high-leverage engineering solutions that catalyze positive systemic change.
    - Facilitating cross-pollination through the aggregation, curation, and dissemination of applicable knowledge to ensure these efforts produce lasting impacts.
- Granular suggestions for Fathom across 2025-2026:
    - Launch "AI Innovation Showcases:"
        - Local businesses using AI to create American jobs.
        - Success stories of traditional industries modernizing through AI.
        - Practical solutions to workforce transition challenges.
    - Establish "Strategic Technology Leadership Roundtables" to explore:
        - The regional and global security implications of AI.
        - The shifting nature of global influence and power.
        - Solutions that would ensure free societies can still enjoy their freedoms in the coming decades.
    - Host "Markets and Societies Summits:"

Education and Literacy as Foundational Needs.

- There's strong agreement that educating and empowering the public will be crucial for addressing the challenges we face from AI. The public needs to understand why AI is important and why they should care.
- The future of work, national security, and innovation all intersect with AI, and the public needs to understand how.
- Creating a cultural moment around AI could help raise awareness and drive interest. This moment should balance fear and possibility—highlighting AI's risks, but also its potential for innovation and societal benefit.
- Many people don't trust AI to work in their best interest. There is currently no widespread ethical framework or accountability system like the Hippocratic Oath for AI developers.
- Simply making AI models more transparent doesn't resolve the power dynamics. Education is the key to ensuring that society knows how to use AI responsibly.
- Solutions must come from the ground up—we need to involve everyday people in understanding and engaging with AI.
- There is a growing divide between the "haves" and the "have nots" - those with access to AI tools and an understanding of how to use them, and those without. This division will likely exacerbate existing inequalities.

Education and Literacy as Foundational Needs.

- There's an urgent need for more tech-fluent people in government, though people from the tech industry may not be the best fit due to a general lack of communication skills and/or an inability to navigate bureaucracy.
- There's also a need to identify and encourage intrinsic motivation, rather than constantly trying to incentivize people through external or monetary rewards. Identify the people who are already choosing to solve the problems that need solving and empower them.
    - The Chinese people have intrinsic motivation towards technology.
    - We should implement campaigns to inspire broader engagement with AI, and ignite a new generation of problem solvers just as the space race did.
    - National security is not just about chip manufacturing—it's about creating a broadly educated populace that is fluent in AI.

WORKSHOP 5

Making AI Work for the American People

- Public Distrust and Concern about AI.
- Opportunities AI Offers to Americans.
- Building Trust in AI.
- AI Literacy and Education.
- The Global Conversation.

Public Distrust and Concern about AI.

- The predominant public sentiments around AI are concern and uncertainty, with people largely describing AI as scary, worrisome, and associated with robots. There's a sense that society "lost" in its first encounter with AI via social media.
- Public trust in AI has significantly dropped, especially in the U.S. A report shows that 74% of Americans feel they have little to no knowledge of how to use AI, and half of Americans fear AI will reduce jobs in their industry. Public trust in Big Tech has also dropped, particularly among younger demographics. AI systems are seen as amplifying power dynamics and reinforcing existing societal inequity.
- People are concerned about manipulation, surveillance, job displacement, and healthcare disparities that could result from AI's unchecked growth. More broadly, AI systems are seen as amplifying power dynamics and reinforcing existing societal inequity.
    - AI could potentially fuel conflict within communities by spreading disinformation or amplifying social divides.

Opportunities AI Offers to Americans.

- Broadly speaking, a diverse range of manual processes could be automated using AI, to the potential benefit of individuals, communities, and society at large.
- AI could help address mental health issues (e.g., via personalized therapy), help families adapt to caregiving roles or challenges with aging relatives, and help with education (e.g.,via personalized tutors).
- AI can make work faster and more efficient, particularly in tasks like writing emails, creating pitch decks, and copywriting.
- Ideally, AI would also create greater opportunities for people to spend more time with loved ones or in ways that are meaningful to them by freeing up time via efficiency gains.
- AI can help companies assess and improve board diversity, offering suggestions on who to hire or appoint.
- AI could help reimagine public services, such as libraries, to make them more accessible and efficient, foster economic growth, and help tackle climate change.

Building Trust in AI.

- Trust is a function of intent, competence, reliability, and relationships. People trust AI more when they feel it's transparent and reliable, but trust also depends on social and political dynamics.
- Leverage existing mechanisms for trust-building to build greater trust in AI - including bodies that must demonstrate reliability and shared values to function like community-based groups - and explore new approaches to bridge-building and deliberative dialogue solutions, like civic and citizens' civil assemblies.
- Engage the public through demonstrations and by elevating representative voices.
- To drive the accountability of AI companies, implement industry standards, open APIs, third-party testing and certification, reporting requirements (e.g., registration of models), comprehensive data privacy and ownership rights, data commons and public "data utilities" for "public welfare" data.
- Enforce better board and consumer governance to prevent situations like those we've seen from OpenAI and Meta, where executives face no accountability.

AI Literacy and Education.

- There's strong support across the board for AI literacy, especially in schools. This should include technical application skills, as well as interpretation and critical consumption skills.
- The need for more public education on AI's benefits and risks is critical. People need to understand what AI can and cannot do, what its applications are, and how to use it effectively.
- Programs should encourage critical thinking around AI and educate students about its implications for society, the economy, and their personal lives.

The Global Conversation.

- The economic and environmental impacts of AI are global, and AI must be a global conversation.
- There's a need for a broader public conversation in the U.S. to set clear, non-polarizing boundaries for AI, opening with issues of shared concern like data privacy and rights before expanding to broader societal concerns.

WORKSHOP 6

AI and American State Capacity

- Increasing and Supplementing State Capacity.
- Disruptions to Expect.
- AI as Both the Problem and the Solution.

Increasing and Supplementing State Capacity.

- The government needs to accelerate policy-setting processes and improve communication between branches of government, which are siloed and often lack effective feedback loops.
- There is a call to create agile policies that incorporate data and feedback to improve over time, rather than just reactive oversight.
- The group recommended that the government initiate pilot programs to test and adapt AI on a small scale. These programs could be run through one agency, and then rolled out more broadly if successful.
- Congress under-resources itself with only 31,000 employees and a $7 billion budget, compared to the 3 million employees and larger budgets in the executive branch. Building capacity in Congress will be essential to oversee AI-driven policies and decisions effectively.
- State capacity should also be supplemented by private capacity through the use of public-private partnerships, as with Operation Warp Speed.

Disruptions to Expect.

- Potential disruptions in the near-term include:
- Increased cyberattacks, data leaks, and hacking of dated government systems.
- Job displacement, especially in mid-level service jobs.
    - Increased use of AI for issuing government benefits could increase efficiency but also drive fraud and errors.
    - The growing use of AI for disinformation (e.g., deepfakes).
    - A slow-down of the FDA drug approval process.
    - Disruptions to the court system as claims proliferate.
    - Disruptions can be split into benign risks (explosion of incoming information, data, demands) and threat risks (vulnerabilities are exploited, e.g., cyberattacks).

AI as Both the Problem and the Solution.

- There is strong agreement that AI is both the problem and the solution. For instance, it can help remove bias but also perpetuate it. It could also enable a creative explosion or lead to standardized, one-size-fits-all solutions.
- Short-term fragmentation is expected, but long-term productivity improvements may emerge, provided the transition is managed correctly.
- Real-world human interactions may become more valuable as AI proliferates in society, with people craving “authentic” human connections over AI-mediated experiences.

Day Two

AI in the Global Context

- The speaker asserted that the idea that superintelligence is literally the business plan is mostly true.
- He explained that frontier AI companies spend billions of dollars and millions of research hours to develop AI systems that exceed the cognitive performance of humans in virtually all economically essential domains.
- He shared that in 2014, he polled experts in machine learning on what superintelligence would mean for global security. He found that most experts were skeptical that superintelligence would ever be achieved and thought that scaling effects would plateau due to a lack of computing power, data, and funding. Now, accelerated trend lines are the default, and the same experts believe it will be achieved within a decade.
- He opined that we are not prepared for the possible effects of superintelligence on global security. If a country can develop a system that exceeds human capabilities in math, science, engineering, and coding, it could be applied to develop new military technologies, including new cyber weapons.
- Recent OpenAI model results have indicated that it is better than 99% of human coders. At some point, it may become cost-effective to have millions of automated engineers in a box designing novel cyber weapons that are orders of magnitude faster than is possible today. A country that achieves that capability may be tempted to use it to disable competing AI efforts to gain a permanent monopoly. The prospect of losing out on a permanent monopoly, meanwhile, could lead to risky behavior, with countries willing to sacrifice safety to outrun their competitors.
- This threat is analogous to that of the Cold War arms race, but could be more intense given shorter development timelines and the dual-use nature of the technology, which makes it hard to distinguish weapons programs from commercial efforts.
- The speaker then asked what should be done about the severe challenges that superintelligence could pose for global security, given the uncertainty of when or how superintelligence may be achieved, if at all. He offered that there are decisions that can be made today that will likely provide more options in the future at a relatively low cost, like:
    - Strengthen semiconductor chip and tool export controls so that computing used for AI stays geographically concentrated within the U.S. and key allies.
    - Strengthen security at frontier labs and data centers so that models and insights are harder to steal.
    - Build “know your customer” (KYC) screenings into the practices of cloud computing providers so they are not unwittingly training large models for adversaries.
    - Significantly expand our intelligence collection and analysis focused on foreign AI efforts, especially those of China.
    - Develop methods and test beds for third-party evaluations of model capabilities as the U.S. AI Safety Institute has been building.
    - Invest in the AI control problem since we cannot reliably control the AI systems we build.
    - Run tabletop exercises and crisis simulations with policymakers to educate and anticipate what information will be asked for in a crisis.
    - Increase institutional capacity in government.
    - Get security clearances for technical experts in industry and academia so we can draw on relevant expertise.
    - Have federal contracts with frontier labs that can be expanded and used to set security requirements.
    - Design a grand bargain that can be offered to the rest of the world, including the Global South and the Middle East, so that local benefits exist for them buying into the U.S.-led AI ecosystem.
    - Continue dialogues with competitors - especially China - to find common ground on AI safety and security issues where we have a shared interest. This should include meaningful human control of nuclear launch and guardrails to prevent AI from being used to develop bioweapons.
- The speaker emphasized that these actions come with benefits and low costs irrespective of whether superintelligence is imminent or progress plateaus. He pointed out that there will likely be disagreements as to whether superintelligence has been achieved even after it is achieved, and that having robust policies is key in handling uncertainty.

AI & Geopolitics: Challenges and Choices For the U.S.

Which geopolitical risks are likely to be increased or decreased by AI?

- If we are hurtling toward artificial general intelligence (AGI), which means a million geniuses at the top of their fields seamlessly working across fields at machine speed without ever having to sleep, that's a geopolitical issue. It will shape the balance of economic and military power.
- As the U.S. and China are jockeying for that leadership position, the country that crosses the frontier first will have the first-mover advantage.
- He noted that it matters who gets there first. It determines which of the two AI superpowers will shape the overall global ecosystem in a way that will impact how AI is deployed, the norms and principles that govern it, and the values that are baked into it.
- The challenge that poses for AI regulation is the balancing act needed between the impulse to deny the technology to China and others, and the need to diffuse technology around the world in a way that favors democracies.
- The third panelist stated that AI intensifies the U.S.-China rivalry.
- Neither country fully grasps the imminent arrival of superintelligence, but as they do, competition will escalate dramatically.
- The first panelist indicated that the geopolitical risk to the United States and the entire democratic world is extraordinary if we're not the first mover in this space.
- He opined that the U.S. government is fundamentally better than China at having this capability.
- He shared that he is terrified by the idea of private corporations having the capability of AGI.
- In terms of a public-private partnership, frontier labs should be on a contract with the U.S. government, and the U.S. government should be compelling them to take specific actions in certain places.
- The idea that superintelligence is the business plan is absolutely true, and it is clear that this is what the frontier labs are all racing toward.
- Therefore, it is vital for the government to find a way to contract with them in the coming months.

What are the most important advantages and disadvantages the U.S. has relative to China?

- The second panelist explained that the biggest U.S. advantage is talent - having the best engineers in the world. The U.S. also has the best compute.
- While the chips are made in Taiwan, they are designed in the U.S. and denied to China.
- China has been good at using open-source models and likely distilling proprietary models to become fast followers on current generative AI models, but there are indications that they believe that U.S. advantages in computing power and the export controls put in place to lock in that advantage will cause them to hit a wall soon enough.
- The advantage will not last forever, but in a narrow window to cross the AGI frontier, it is important for the U.S. to secure that advantage so that they can make decisions regarding the roles of government and the private sector and how these technologies are governed globally, rather than have those decisions made by U.S. rivals and adversaries.
- The third panelist added that maintaining an advantage in chips and talent requires a coherent governance strategy.
- While there are efforts to place export controls on chips, the U.S. is missing a coherent strategy for governing computer hardware. This not only includes computers, chips, and data centers, but the training and deployment of models, including open source.
- As AI continues to advance, if we piecemeal components of it, do not enforce export controls, or limit training models that later get open-sourced to competitors, it is not going to work.
- The first panelist indicated that the U.S. doesn't understand the extent or nature of its advantages over China, as seen with Salt Typhoon.
- We've found China to have been engaging in activities for 15 years that we didn't think they were engaging in, so while the U.S. may appear to have the advantage, that should not be assumed for more than days or weeks.
- He provided an analogy of how a NASCAR driver treats a race, using both the accelerator and the brake at all times to make sure they maintain maximum speed through straightaways and corners.
- Instead of focusing on who is winning the race, the objective should be to always maintain the leadership and advantage in every strategic space within the AI domain. That requires that we educate people, build data centers, and build energy sources.
- China is racing to beat the U.S. across every strategic space and we need to do the same thing.

What are the risks of the U.S. being too aggressive or too passive in AI development?

- The second panelist indicated that a racing dynamic could lead to cutting corners on safety in ways that could lead to catastrophic outcomes. For example, highly capable rogue agentic systems roaming the internet undetected could do terrible things or the most dangerous actors could be enabled to create the most dangerous pathogens.
- But the risk of going too slow is that the safety risks still happen, and the adversary gets the strategic advantage.
- He stated that though he is confident the U.S. companies will get AGI first, the risk is that the U.S. government will not be able to capitalize on the technology as quickly as the Chinese Communist Party, who will become fast followers in adoption and integration.
- The U.S. and its allies could dominate at the frontier but lose the AI global competition to the Chinese "good enough" AI based in a digital authoritarianism around the world.
- That would have profound negative effects on humanity and our interests.
- The third panelist indicated that the U.S.'s leadership in AI comes with responsibility; we underappreciate the risk that comes from the U.S. leading in AI.
- We want the advantages of being first, but if commercial frontier labs or the military are not careful to deploy systems safely, we may be responsible for accidents.
- We should be conscious of the safety and structural risk that can result from being ahead and then incentivize others to not cut corners as they try to catch up. It matters how we lead, set the pace, prioritize safety, and ensure that others are going to follow.
- The first panelist added that the only way to truly have an adult conversation about safety is if the government is in the loop and companies are on contract.
- National Labs could make sure that companies are doing the required testing and involving experts at every turn.
- Companies may say they are advocating for safety and doing the best they can, but the profit-driven nature of business will ultimately cause them to cut corners.
- Therefore, a government contract and a government relationship are critical.

What is one actionable step the U.S. should take to improve its AI strategy?

- The second panelist explained that public-private partnerships are needed, in which the government facilitates the dramatic expansion of compute that the private companies will need to push the frontier.
- In exchange, companies will need to provide significant access by the federal government to compute to evaluate models and run classified simulations to evaluate national security implications, to ensure the government is ready to defend against any adversary that may cross the threshold first.
- Another step should be that the government actively helps companies overcome market failures that lead them to underinvest in distributing the benefits of AI to the global majority.
- Distributing beneficial AI to the world can obtain buy-in to democratic norms and principles, and safe, secure, and trustworthy AI.
- That global influence only occurs if you dominate the digital stack across the world.
- As we race toward AGI, the U.S. government still has enormous leverage, but it may not last for many years.
- The current administration will need to determine what that global compute compact looks like and what the U.S. government expects in return from both U.S. companies and the rest of the world.
- The third panelist identified computing hardware as the biggest leverage that the U.S. currently has.
- Chip export controls are a piece of this, but a global governance regime over computer hardware is needed.
- The government should work with partners and allies to know where chips are being made, where they are going, and how they are used to ensure large amounts of computer hardware are not being used by malicious actors or an adversary or competitive nation.
- Some controls are in place, but we need to get to a place where making this technology accessible to others is conditional on them ensuring it will be used responsibly.
- There is an opportunity to do this, but there are many additional steps needed to make this happen.
- The first panelist pointed out that the U.S. government is in a data crisis, and needs to standardize and integrate the extraordinary amount of data available across its agencies.
- It is going to be critical to get the data infrastructure layer in order and bring the data from the USGS, NOAA, CIA, NSA, and other agencies together to train the AI and determine capabilities.

AI and Democracy

- The speaker thanked Sam Altman for providing the framing of this talk with his recent tweet "algorithmic feeds are the first at-scale misaligned AIs."
- AI mattered in the 2024 election in two ways: foreign actors sowing dissent and misinformation about the election and domestic actors desperately trying to use AI, with both intended and byproduct effects.
- After providing examples of foreign actor influence and domestic actor uses of AI, the speaker highlighted the work of Tristan Harris, Executive Director of the Center for Humane Technology, to explain the byproduct effects of AI in social media.
- In his study of the science of attention, Tristan Harris found that technology companies are engineering attention to overcome our resistance to these technologies and to increase engagement.
- He described it as a type of brain hacking, in which companies exploit the fact that we react irrationally to random rewards and cannot stop consuming bottomless pits of content, with the aim of increasing engagement to sell ads better.
- This business model is driven by AI, which creates incentives to know more and leverages our insecurity for profit, monetizing attention.
- It has been determined that the best, most profitable strategy to drive engagement is the politics of hate. AI learns that our engagement increases the more polarizing and hate-filled the content is, and in turn, that is what the algorithms feed us. These algorithm-based feeds are the first at-scale misaligned AIs we have seen in our society so far.
- The speaker explained that this algorithmic feeding is part of an economy that Renée DiResta, author of Invisible Rulers, has broken down into three components: influencers, algorithms, and crowds, that work together to yield bespoke realities or multiple realities.
- He illustrated the evolution of bespoke realities by reporting that Americans' support for Richard Nixon in the lead-up to his resignation collapsed at the same rate regardless of political party. Watching the same story had a common effect, creating a common reality, not multiple realities. Conversely, during President Trump's first term, Americans' support for Trump within each political group remained steady regardless of what happened, a prime example of bespoke realities.
- We need to recognize that this is a new norm in how democracy works.
- Bespoke realities optimize for fantasists, which the speaker defined as people who do not necessarily believe what they say is false, they just choose to live in the reality they have constructed. He named three prominent fantasists ready to lead parts of the U.S. government who, he believes, have no connection to reality.
- The speaker warned that the consequences of this engagement economy and business model are that people become polarized, ignorant, angry, and live in bubbles cemented with hate. This weakens democracy.
- He emphasized that this is the AI we should be focused on. The point is to recognize that this AI is just a byproduct of a business model trying to sell ads. This is not because AI is so strong; it is because we are so weak.
- Tristan Harris has explained that there is a moment when technology exceeds and overwhelms human weaknesses. That point being crossed is at the root of addiction, polarization, radicalization, and it's overpowering human nature.
- We need to recognize that the effect of AI is real long before we get to AGI, when it overcomes our weaknesses and collective capacity to use democracy to address issues. As AI seeks engagement, a not-to-be-ignored public will emerge and have its democratic effect.
- The speaker then addressed a second point raised by Sam Altman's tweet: How screwed this means we are.
- He reported on a memo sent by a Facebook executive to staff prior to the 2020 election, in which he stated that he expected the algorithms to primarily expose humanity's desires, for better or worse. It is a salt, sugar, and fat problem, referencing the book by Michael Moss. The book indicates that giving people tools to make their own decisions is good, but forcing decisions upon them rarely works. This forces individuals to consider what responsibility they have for themselves.
- The point, the speaker argued, is that it's not a Facebook problem but rather a democracy problem.
- Humans have made "analog AI" in the form of bodies that have purposes and act instrumentally to advance those purposes.
- These analog AIs have institutions, elections, parliaments, and constitutions for the purpose of advancing toward some collective end and a common good.
- Corporations can be considered a type of analog AI - with institutions, boards, management, and finance for the purpose of maximizing profits and shareholder value. Facebook and Meta can be understood as examples of analog AI, with the objective function of maximizing shareholder value.
- If we agree that social media is misaligned socially and democratically, we must recognize that it is corporately aligned.
- In advancing the corporate objective, the engagement business model defeats the democratic objective. It is privately beneficial, but publicly not. It is privately profitable but publicly democracy-destroying. The idea that we should be sacrificing and risking democracy to make some people richer is crazy.

- What should be done to fix this?
- The speaker presented the possibilities of creating new laws, changing the engagement business model, or utilizing regulation innovations to drive it out of existence.
- He also suggested an engagement test within a platform that monitors usage and, at a certain threshold, suggests that you might want to take a break. That strategy could change social media engagement and drive people to a healthier business model. That is not a strategy that risks First Amendment threats.
- What can be done in the longer term?
- We need to determine how to move democracy from the unprotected space it lives in right now to a more protected space, where we can engage in democratic decisions without the risk of the engagement model.
- He explained that there is a "citizens' assembly" or "civic juries" movement, where random and representative citizens deliberate and take on policy questions in a context where they are protected from the influences that drive their decisions. Iceland, Ireland, and France have used this approach to develop policy on polarizing issues.
- This movement is the only magic in the democracy space right now, and it shows what democracy could be.
- The speaker shared that he believes it is the only long-term hope for recovering a belief in ourselves, which is critical if democracy is to survive.
- A 1997 Pew study found that 64% of Americans had a good deal of trust and confidence in the wisdom of the American people when making political decisions. The follow-up Pew survey in 2019 showed that the confidence level had dropped to 34%. This indicates we have lost faith in ourselves, and he surmised that this has been driven by an engagement business model that teaches us that the other side is the enemy and encourages ignorance about them and their wants.
- As a final thought, the speaker shared Mustafa Suleyman's concept of pessimism aversion: the tendency for people, particularly elites, to downplay or reject narratives they see as overly negative. It's a variant of optimism bias, but it colors much of the debate we have about the future, especially in technology circles.
- We must recognize that the democracy-AI problem is here right now. And that, as we discuss plans for the coming wave of AI threats, this is our test.

The Next Chapter: Policy in D.C.

While AI has come of age and become part of the vernacular during the Biden administration, early federal policy started with the first Trump administration. What policies do you see coming out of the Biden administration, or the legacy Trump administration, that will stay, and what will get rolled back?

- The first panelist shared the view that not much from the Biden administration will stay. He predicted the repeal of the Executive Order (EO) on day one of the second Trump administration, and beyond that, said it will be mostly a clean slate with the new administration.
- The second panelist agreed that it will be a clean slate. He foresees AI being treated through its different silos and the different types of policy that they reflect. For example, from the U.S.–China competition perspective, AI will be governed by how the Trump administration wants to deal with China. From a corporate regulatory perspective, AI policy will be based on how the administration wants to go after Big Tech. Most policy is likely to come through those domains, not as a whole-of-government approach.
- The third panelist noted that given the influence of David Sacks and DOGE, the AI institutes will be marked for exit under the new administration. Some may keep an influencing aspect, but most will be seen as superfluous and get cut. He foresees a more laissez-faire approach to letting AI grow.

What should the audience know about David Sacks, the incoming AI czar? How does he think? What's his team like, his leadership style?

- The first panelist indicated that Sacks, as he is referred to, is a pretty open book. He is very smart, tends to listen more than he talks, works from first principles, and learns and iterates as he goes.
- He believes that Sacks will build infrastructure within the White House for policy focused on innovation and growth that will see U.S. businesses flourish.
- Regarding his style, Sacks is approachable but is not expected to be out in front at typical DC lobbying events. He will rely on Mike Kratsios, Director of the Office of Science and Technology Policy (OSTP), to help him craft something with the President's Council of Advisors on Science and Technology (PCAST). That agency will likely facilitate the industry connections and conversations around AI.
- The first panelist further explained that Sacks is currently deciding where to dig in from an agency perspective and figuring out where he can lean in quickly to get some semblance of a policy built in the first 100 days.
- The third panelist added that a czar's power is to convene meetings. They do not control the budgets of the departments and agencies that are supposed to execute the vision set forth by the President. It will be necessary to figure out how to create principles that can secure buy-in and be announced in the first 100 days and then determine the delivery aspect.
- The question then becomes how fast the next administration can get past deep-state resistance to initiatives and execute the implementation and operations stages. Achieving that will be the hallmark of success. The individuals being appointed to departments and agencies to be those executors are what matters. In contrast to the Biden administration, those in technology leadership positions under Trump will be doing the work instead of speaking in various forums, ideally leading to them getting more done.
- The second panelist highlighted the interesting disconnect between the conversations that happen in the AI tech community and those that happen at policy planning roundtables in D.C. In his opinion, the D.C. participants are given fake titles and czardoms without authority or budget. What distinguishes them is a large twitter following.
- He pointed out that those figures will have a certain type of influence, but it's probably a mistake to look at those who have taken that set of roles and expect them to drive federal policy in a meaningful way.
- It's more important to focus on individuals at the National Security Council, in policy planning at the Defense Department, at the State Department, in Treasury, and in Commerce. These people may not get headlines, but they will shape the federal government policy in the next four years. Policy will depend on what they think and do rather than what the various czars want to do.
- The first panelist agreed that Sacks' power would be in convening. The czar role will give him the ability to work more closely with agency leaders instead of being siloed in OSTP or PCAST. The role will create more conversations, increase industry engagement, and placate the venture side. The czar role is not intended to come up with AI policy but to bring all the voices to the table.

Where do you see Congress over the next couple of years? Who are the leaders? How will they think differently than the Congress of the last four years?

- The third panelist explained that, as the incoming Chair of Homeland Security, Rand Paul has made it clear that any AI intervention that he considers intrusive to civil liberties is a non-starter. Whether it be misinformation, disinformation, the Cybersecurity and Infrastructure Security Agency (CISA) regulating opinions, or some other entity in government, he will find a way to go at it. That will be a theme and unavoidable topic of conversation.
- The second panelist added that we're unlikely to see much on the legislative front. In the annual National Defense Authorization Act, you could start to see some provisions that speak to the authority of an agency within Defense to focus on or fund particular items.
- In terms of seeing what the governance framework is for AI, we're unlikely to see any momentum or energy in Congress unless we pass some sort of benchmark that signals to the public and political world that AI is causing effects in society that are different from what we otherwise see day-to-day from other technologies. He expects that, in Congress, those who tend to be serious and thoughtful about tech issues will generally be interested and thoughtful on AI. And those that like to hold hearings and scream about Big Tech will continue to do so.

- The first panelist agreed there are a number of voices to listen to. Rep. Obernolte (R-CA), Sen. Schmidt (R-MO), and Rep. Liccardo (D-CA) are sneaky good on AI and tech policy and will probably lean in a little more and be vocal.
- It's more likely that we'll see a lot of hearings than a lot of legislation, but with Republicans in control of the White House, Senate, and House, there are opportunities to do a bit more.

How do you see Congress looking at the private sector? There is an assumption that Republicans are very business-friendly. Do you see that changing? What about that assumption changes?

- The second panelist indicated that the assumption has changed with respect to the tech community, as conservative Republicans were influenced by the Google Gemini experience. That is the lens through which any effort or claim made by frontier labs is now filtered. He elaborated that it's not even skepticism so much as outright hostility, and it's assumed to be mutual hostility.
- You can expect casual conflation with social media and legislative movements in the next couple of years on secure digital ID and online gambling. You're likely to see the starting point be that frontier labs, like OpenAI and Anthropic, are just appended to the list of Big Tech companies, such as Google, Meta, and Amazon, despite not being in the same business or using the same tech.
- The third panelist stated that, from the executive side, you'll see industrial policy changes to support the supply chain for data center growth. The executive branch enforces laws, and the congressional branch appropriates the funds. He hopes that there will be congressional support for industrial appropriations to support industrial policy changes that hit the AI supply chain in the first two years.

Expand on the misunderstanding that the industrialization dream that people are talking about, that AI fits right into, could be the first thing that moves forward. What does it really mean, and how does it intersect with AI?

- The first panelist explained that there would be an interesting play between what DOGE says about building out the energy infrastructure to support growth in AI and how that supports smaller companies, and their ability to grow and compete with some of the primes in terms of opportunities with the federal government, state governments, and municipalities. The right mix might be finding a way to get all those entities working well together and creating opportunities to support U.S. businesses looking at the reshoring aspects of opportunity. That will all start with a focus on innovation and energy infrastructure.
- The second panelist stated that the Trump Administration wants and needs to do industrial policy but cannot be seen pursuing the Biden industrial policy. In specific instances, like chips, they need to find a way to support a technology but identify updates on Biden's approach.
- Energy will be a huge opportunity since so much of Biden's industrial policy was based on the Inflation Reduction Act and green energy, which Republicans can present as climate-focused interventions that made energy less reliable and more expensive. Energy is going to be the tip of the spear in MAGA-branded Trump industrial policy. Proactive economic policy would be focused on energy anyway, so adding in the point that a policy is needed for AI becomes the way it fits together.

There are many layers to AI policy, including reindustrialization, energy, governance, and ethics. The public is signaling loudly that they want to see a government structure around AI but do not trust the government to be the implementer of said structure. What are the alternatives in an apolitical policy environment where you can look at the governance and ethics of this next wave without a whole new regulatory framework?

- The third panelist commented that under Trump, any governance you see will follow a traditional framework approach. There is already a National Institute of Standards and Technology (NIST) risk management framework for AI. It is likely that the administration will lean into NIST, a public-private partnership, to provide that governance framework. As a standards body, NIST cannot implement governance, so another entity would need to drive implementation, such as the Commerce Department.
- The second panelist explained that solutions rarely work from the bottom up in the federal government. In the private sector, you're able to learn about AI and quickly start to create solutions to deal with it. For policymakers, there is no easy way to do that across the range of things they're supposed to be doing. There, it works top-down - using an existing ideological framework for governing and then determining which buckets the problem fits into.

- This is especially true on the right of the center, where you cannot expand government, so it needs to fit into an existing template of something the government is already doing. For example, instead of saying that a new agency is needed for AI, they will agree that a set of military-class technologies are governed in a particular way and that AI is that kind of military-class technology, and then they'll consider what tweaks need to be made to the regulations. They'll continue to apply this principle to govern AI. It will always run through that process, not from "AI is amazing," and we need to create a department of AI.
- The first panelist said that guardrails and standards need to be figured out. Watching how the Administration works with Congress to balance the desire to get out of innovation's way with the need to protect will be interesting. Figuring out what that balance looks like is going to have to be step one.

Should there not be a step one? There is talk that a cataclysmic world event would be the motivating reason for either government structure, ethics, or regulation of AI. Is there anything in world events or threats to critical infrastructure or workforce you foresee happening in the next four years that would be that catalyst? Is there anything that people should be watching or listening for that could be the trigger?

- The third panelist provided Salt Typhoon as an example.
- The FCC provided China with telecom licenses to operate in the U.S. The U.S. telecom system was not built to have security, so we effectively gave China access to every phone call and text message.
- Compute can be equated to energy. By using your compute at all the data centers available, you are just drawing energy to solve a problem. China is burning all our energy assets. We know that some parts of the country have data centers without good power generation, transmission, and distribution. China could draw out that energy and cause regional brown-outs, forest fires, or other types of physical world events while also crunching the Office of Management and Budget (OMB) data set to solve other problems. Without a "Know Your Customer" policy in data centers that allows us to identify who's buying the compute, that would be the type of event to look out for and prevent.

Are you seeing any industries being replaced by AI or predictions coming to a crux in the next four years?

- The second panelist hypothesized that self-driving vehicles could be at that scale.
- However, the rate of progress demonstrated in labs does not look anything like the rate of progress in scaled deployment out into the workforce. It is hard to envision anything being achieved in the lab over the next four years or being commercialized and deployed at a scale that could cause sufficient workforce dislocation to prompt a government response.
- The federal government is great in that most of the time, it doesn't do anything, but when we have had crises, Congress gets together very quickly and acts. To get that kind of response, you'd have to create some credible signal based on more human benchmarks than solving mathematical problems set by Fields Medal winners, and these would have to come from the security side and the workforce side to raise that sense of crisis. He can't think of anything coming out of the AI community right now that meets that bar or is expected to.

Besides David Sacks, who on X should we follow over the next few years?

- Jay Modh, J.D. Vance, Michael Kratsios, and Jon Stein.

Fireside Chat: AI Leader

What changed for you when ChatGPT launched in November 2022?

- The speaker shared that he was suddenly confronted with the reality of the existential risk that came from the work he'd been doing all his life.
- ChatGPT passed the Turing test - predicted by Turing as the point at which machines surpass humans - and it was not seen as a positive outcome.

- It took a while for him to digest what was going on, and to go from intellectual recognition of a trajectory where bad things could happen, to the emotional recognition that something could happen to the people he cared about due to things he had done or contributed to in his career.
- As a result, he decided to devote the rest of career to reducing the risks of AI.

Was there a specific factor that influenced your shift in perspective, or was it simply the cumulative demonstration of its capabilities?

- The speaker indicated that it was just the general linguistic capability. He had not been exposed to the large models of industry in academia, so the jump he saw was surprising.

What do we mean by existential risk? What are the real scenarios? For many people, it feels abstract, but how might bad things happen?

- The speaker broke existential risk down into three categories: rogue AI, misuse, and loss of control.
- He warned that intelligence grants power, and if concentrated in a few hands, AI could undermine democracies, since whoever controls these expensive, powerful systems is going to have the power to bypass the checks and balances of democracies. So, how do we govern together and share power?
- The other issue is that the entity controlling that power may not be human.
- Humans are intelligent agents with self-preservation instincts that spur us to do what we need to in order to stay alive.
- If AI has self-preservation goals and is smarter than humans, it would be the equivalent of creating a new species that may not have interests that align with ours.
- Why would a self-preserving intelligent entity want to get rid of us? Humans would not be happy with entities more powerful than us and would want to turn them off.
- AI systems would make certain that we cannot turn them off by either controlling us sufficiently or getting rid of us.
- One path to AI having a self-preservation goal is that some people will want to put the goal in the machine, which is easy to do if the right guardrails are not in place.
- There are various reasons why people may want to do this, but it would be a mistake to create something more powerful than us that is not aligned with us.
- Self-preservation may also develop as a side effect of other objectives we're training into the models. Because we cannot fully predict AI's subgoals, some might naturally lead to self-preservation, power-seeking, and human displacement. The problem with AI systems is that we cannot easily choose what those subgoals will be. And natural subgoals, like seeking power and preserving ourselves, are not good for us.
- The point is, we do not know how to build AI systems that are both smarter than us and controllable.
- Unlike other existential risks, such as nuclear power where a bomb could eliminate a limited population, human extinction is a real risk with AI.
- Self-preservation is a goal, and AI is about achieving goals.

But humans have self-preservation instincts and remain regulable...

- Humans are regulable because a single human cannot defeat groups of other humans "bare-handed" owing to similar levels of intellect. No single person is capable of overpowering society.
- If there are AIs substantially smarter than us, no matter how many of us there are, it is not clear that we could defeat them.
- But until robotics is sufficiently advanced to take over industrial infrastructure, AIs still have need for humans.
- But AIs could control humans by influencing the political process and help seat individuals who have policies that favor what the AI wants to achieve; whether that's more robotics, more deployment, or triggering a catastrophic war.
- There are three stages to a typical catastrophic scenario:
- Phase one is the AI silently planning, not giving any indication that it's thinking about how it could take over.

- Phase two is where AI begins to influence policy and take control of the direction in which society is moving. The AI will need to act by manipulating and controlling people with offers of medical solutions or ways to acquire power or wealth. If the AI itself acquires wealth, it can pay criminals to do things. There are many ways AI can start influencing people through the internet undetected. It could be digital personal assistants advising us, mostly for good, but with some suggestions that favor the AI.
- Phase three is where the AI no longer needs us and through several waves of bioweapons, gets rid of us.

Thoughts on Agentic versus Scientist AI?

- The speaker defined AI agents as systems with intelligence, the ability to act, and goal-seeking behavior.
- Their intelligence allows them to understand how the world works well enough to create plans which would allow them to shape and create impact in the world. It also allows them to develop goals of their own.
- Knowing where they stand in the world with respect to those goals, the systems can achieve goals that have to do with themselves, like self-preservation.
- Any of those three factors can be limited. If they are not too smart, they cannot be extremely dangerous. If they have limited ability to do things, they cannot pass for humans on the internet or make deals with criminals or have hidden accounts.
- He'd like to get rid of the goal-seeking behaviour.
- He advocated for developing "scientist AI" over "agentic AI."
- Scientist AI would understand the world and assist in scientific progress without having goals or self-preservation instincts.
- He warned that current economic incentives push companies toward agentic AI, which poses long-term risks.
- He explained that because humans are intelligent and agents, we seem to think that if we build intelligent machines, they would be like us, but there are conceptual differences.
- You can have systems that understand how the world works, like scientists, and even plan experiments, without them having goals, self-awareness, or situational awareness about where they stand in the world.
- Scientist AI could objectively look at things and try to make sense of them. With that help, we could advance science.
- That is technically feasible and does not have to be combined with building systems that are like us, that want things in the world, and try to achieve those goals.
- All scenarios of loss of human control to machines smarter than us rely on those machines having agency. It is when AI wants things from us that we get in trouble, and right now, we do not know how to fix that problem.
- In the short term, until we can find that solution, we need to focus on building machines that can help us solve our problems. We do not need superintelligent machines with their own wants. That is not needed to solve our problems.
- This idea is not easy due to the commercial pressure to build agents.
- Agents provide short-term economic gain, but it is dangerous for the future of humanity.
- We need to realize that and see that we have other choices.

Is it feasible to say "do not build agency"?

- The speaker proposed strategies for mitigating risks, including regulations, insurance mandates for AI companies, and government oversight through public-private partnerships.
- He argued that we need to counter those forces that will change the trajectory we are on right now, because the default trajectory is extremely dangerous.
- There are several things that need to happen. First, we need a better understanding of risks and scenarios.
- Second, we need to slow down the dangerous trajectory and shift the energy toward building things that are both useful and safe.
- This could possibly be accomplished through regulation or the use of insurance.
- If AI companies had to buy insurance against the damage their systems could create, they would have to be honest with themselves about the risks they are taking.
- If they could not find insurance because their system is not demonstrably safe, then they would have to look for other ways to make money.
- There is a need to find a way to provide the right incentives for companies.

- Governments need to know which AI systems are potentially dangerous.
- A possibility is a public-private partnership where governments enter contracts with the AI labs, providing resources they need in exchange for the companies behaving in a way that is not going to destroy the United States.
- We need to do research and engineering to build these alternatives. Until this happens in industry, it could happen at a smaller scale in university labs, in non-profits, or in for-profits that take it as a mission.
- He's about to publish a paper proposing new training methods to avoid self-preservation instincts.
- There are two problems with current training. First, it is trying to maximize rewards, which is a dangerous path because the best way to maximize reward is to take control of the machine, which leads to strong self-preservation. The other negative aspect to the way we currently train models is that it leads to systems that imitate humans, who can be deceptive.
- A lot of the problems we see now with deception in frontier AI come from their propensity to imitate the bad things about humans.
- The alternative is to train them to be more like scientists who try to understand things, explain the data.
- That explanatory objective can be written down mathematically as a different way to train these systems. The only thing they can do in that context is provide the correct mathematical answer; they do not have any degrees of freedom or agency.
- He emphasized the need for funding from philanthropy, governments, and responsible investors to counter market pressures that favor risky AI development.
- He explained that some research could come from academia where there is more diversity of exploration, but there is a need to have enough money to train large systems.
- That capital could come from philanthropy and investors.
- The problem with the capital market for that kind of money is that it will probably lead you to do the same things as existing AI companies.
- The other option would be the government, who would want to use their own money to fund national labs.
- Importantly, even if we have useful non-agentic AI, it could easily be turned into a dangerous tool if it were in the wrong hands. It's something that needs to be secured and managed with the right governance.

How does China play into this?

- The speaker warned against focusing solely on AI threats from China while neglecting domestic AI risks. He urged investment in safety research to address both concerns.
- He explained that the mistake is to think we either let China go in front and use AI against us or we pay attention to the existential risks coming out of our own companies. We need to deal with both risks - both creating monsters at home and other countries creating monsters to use against us.
- He noted that the companies are not considering or investing in safety. It is not necessary to remove resources from trying to advance AI capabilities, but we need to put more resources into safety.
- Investing in safety will help to deal with the risk of losing the race to China and make sure we do not blow ourselves up.
- We should definitely share the things that will increase safety in various domains, and definitely not share things that would potentially increase capabilities in the wrong hands. Sometimes the two are intermingled, but we should work hard to share the parts that should be shared.
- The safest way to build superintelligent machines is to do it in a context where the governments of the world are working together with public good as the primary objective, which includes safety. There is a trajectory that includes working with other countries to make sure that no one makes a stupid mistake.

The Appendix

WORKSHOP 1

U.S. Leadership on AI

- What does U.S. Leadership on AI look like?
- What are some sources of competitive advantage for the United States?
- How do we maintain control over our models? Is control possible?
- Energy permitting reform is a top priority
- What about immigration reform?

What does U.S. Leadership on AI look like?

- Could be for the U.S. to push out the frontier as quickly as possible while retaining control as long as useful, where that's determined by the extent to which control leads to prosperity, security, and protection of values.
    - The private sector is doing a lot to push out the frontier - investing a lot in R&D.
- Need to differentiate the concept of control from non-proliferation.
- Electric vehicles/autonomous vehicles - prosperity or security?
    - What Tesla did was super important - Tesla now has an agreement with China to develop vehicles based on terms the Chinese government set down. This constitutes a failure for America; we are not in control.
    - Under what conditions would we say that Team America is winning?
- If we want the U.S. to lead, we need AI, energy, and, data - and it can't just be one or two states leading the way.
    - We've got to make sure that any suggestions we make are inclusive of all the states.
    - Manufacturing needs to be prioritized, and it needs to be distributed across the whole of the United States geographically - across all of America.
- Note that chasing shareholder equity as a means of chasing prosperity creates security vulnerabilities in the supply chain.
- We're making a series of assumptions in this conversation. First, AI is a revolutionary technology that will improve American prosperity, security, and values. Second, that there is some advantage to being first, OR a profound disadvantage to being second or last. But there are real dangers inherent in this technology that also have to be controlled or managed.
- Control is actually possible, but is the federal government the right agent for control?
- We could develop exquisite AI capabilities and this still wouldn't guarantee broad societal usage or adaptation. The government shouldn't just push on the development of capabilities, but also on what makes a difference to Americans (e.g., food systems, health care systems).
- All applied research and stealth research on aircraft came from the Soviets.
- There are a lot of people in many parts of the world following the U.S.
- There's leadership in being first with the technology, leadership in the market, and leadership in government adoption.
- Double-clicking: We need consumers to adapt to a world with AI.
- We're missing applications today. Until you apply in certain contexts and show that it works, you won't know that it works.
- This is a fundamental point to bear in mind when thinking about maintaining a federal advantage: If you don't apply the technology, you don't get the advantage.

What are some sources of competitive advantage for the United States?

- The "AI Triad" - data, compute, algorithms.
- Talent! Including global talent.
- Deep and efficient financial markets.
    - You could be wealthy in a lousy market - need to coordinate allocation of capital.
- Energy.
- Regulatory framework, good environment.
- Supply chain - the players that are able to lock in supply chain security can compete.
    - Offensively and defensively - we're creating chip shortages for China.
- Culture! The U.S. likes building stuff.
- Control of AI systems.
    - Developers don't fully understand how and why models work, which makes them difficult to predict.
    - It's critical that managers know what's going on.
- U.S. financial hegemony: The U.S. dollar. Every bank in the world has to comply with the U.S. dollar.
- The ability to defend our technology via a strong intelligence community, data center design, etc.
- Entrepreneurship is another competitive aspect of U.S. culture.
- Quantum computing. It cuts energy consumption, which would give us an advantage.
- Freedom and lack of censorship - U.S. companies aren't censored.
- To the extent that we believe in free society, talent acquisition.
- Americans give their data away easily. This could be a source of advantage, although not the way we currently do it.
    - Optimal data protections.
- Digital identity. Every user has the ability to prove who they are in Europe. In America, this is a problem, especially for deepfakes. It's important we have digital IDs.
- Control point of supply chain. The only way we can realistically leverage control points is via mandatory hardcore licensing.
- Providing a business model for an AI that's on your side. What does not work for AI is banking-style regulation. The best we can do is set up a competitive ecosystem of AIs that are on your side. That is a challenge for Congress.
- If you have a good AI, regulations become less important. If AI is aligned, it will be working for you.

How do we maintain control over our models? Is control possible?

- Control only works some of the time: any control strategy also needs to have an advantage strategy.
    - Even if controls were desirable, it's not a "set it and forget it" model.
    - Is the thing you want to control models?
- The competitive advantage given by lots of our models comes down to whether you can do something a bunch.
- What does controllable look like at the prototype stage, at the scale of adoption stage?
- Which are the important models to control?
- Even if we're only controlling models for a limited period of time, we'd need to figure out what the failure mode is and what the control strategy is.
- Hard power vs soft power.
- Why not pursue a policy of benign neglect. Just do nothing, and assume it's all going to go well?
- This is something we've done in technology policy before, assume we'll have a policy moment after the technology has diffused. But it's through proliferation that we discover what we don't know, which should color the way we look at diffusion. It's too risky because we can't reverse it. We may end up losing geopolitically.
- Why haven't we said anything about China?
- The way that most of humanity will access these technologies is via the cloud.
- If we think about the clean energy supply chain, we're heavily reliant on China.
    - Chips can be used to our advantage: If China is using chips that the U.S. made, this puts us at an advantage.
- What are we talking about in terms of scale? Diffusion at cloud scale, laptop scale?
- There are lots of different questions involved here, e.g., GPUs for hospitals, large-scale data centers, and training frontier models.
- We can't prevent things from getting to China. The Texas Instruments chips go to China via calculators.

Energy permitting reform is a top priority.

- Throw nuclear on this. 5-6 years, within the decade.
- The most optimistic time we could hope for is 3.5 to 4 years. More likely it'll be 5.5 to 6 years.
- Renewables are shovel-ready in most cases. Every market has different rules around entering the queue, but renewables can happen quickly.
- The timing issue is the key issue. The Department of Energy (DOE) is focused on the short and medium-term.
- What's needed is grid-enhancing technology and greater access to power! (strong agreement from the room)
- The DOE is trying to figure out the question of what it can do about it in the next year or two right now. It's excited by AI permitting, and there are ongoing pilots in that area.
    - The DOE is putting out requests for information (RFIs) on uses for land it owns, and questions it should be thinking about as it thinks about uses of AI. What do we not want energy to come at the expense of?
    - There's a need to make sure people want more data centers in America, but equally we want to avoid NIMBYism.
- Clearly permitting reform is an extremely valuable thing, but what are the terms of the trade?
- How do we weigh federal vs. state level permitting reforms?
- There's not enough consensus on the federal level.
- California is not going to be a great state for this, but there are plenty of states in middle America that could do this. What state is most willing to work with us?
- There's a difference between generation and transmission. This was an issue in the navy 20 years ago.
- If you're in Arizona with strong solar, you're in. Chicago powered by wind? They love it.
- We need both federal action and incentives for this.
- Efficiencies, always large focus on chips.
- Density has become so significant. You can be as efficient as you'd like, but energy usage is just going to go up.
- How much energy do we actually need?

What about immigration reform?

- We need the best talent - that's common sense.
- When we were building rockets, we brought the Germans over. We should make it easy to do that again. We need to look for creative approaches to immigration.
- When the Democrats are in power, they favor family-based immigration policy because they want to use highly-skilled immigration as a bargaining chip.
- There's a long list of things that could get a majority vote in Congress but don't because nobody's willing to fracture their coalition.
- The U.S. had no cloud talent. Amazon started its own series of programs and had to retrain employees.
- I'm not disagreeing with the importance of immigrants but we as a nation should look to see where the talent is in American schools.
- This is where we have to push the commercial industry players to do the work.
- Everything has been done using executive authority under the Biden Administration.
- Note that while we are all in favor of getting additional highly-skilled talent from other countries, we bleed back into the control conversation.

WORKSHOP 2

Is the Law Ready for AI?

- The potential for algorithmic collusion and transparent information sharing through models is a problem.
- Tort is not fit for purpose; there's a massive amount of liability that still needs to be allocated across the AI value chain.
- There might be room for a semi-private, quasi-governmental body like FINRA to regulate AI.
- Other Notes

The potential for algorithmic collusion and transparent information sharing through models is a problem.

- Yes, the questions listed are the right questions.
- The law is clear on what companies can or can't do to cooperate on pricing strategy but it's much less clear on AI agents doing things within a company.
- What happens if an AI does something that's patently illegal for employees to do? Should we generally treat models like employees?
- The Department of Justice (DOJ) is beginning to dip its toes into the water now with algorithmic collusion.
- If a model impersonates a human, that feels like collusion.
- But how would you even know if collusion is happening?
- Creating an information-sharing platform is the first step. The second step is deciding which government agency would own this issue.
- As a general rule, companies that use AI need to have substantial skin in the game regarding liability in use cases.
- They also need technical tools in place to reliably test and evaluate whether the tools they're using are safe. We need a massive ecosystem of evaluation tools and vendors to help businesses make these decisions.
- For that to work, you need a substantial amount of the liability to fall on the companies using AI.
- If an agentic system commits a tort that results in harm or loss of property - I'm open to liability falling to the developer in specific cases, but would want to default to user responsibility.
- I'm a user of AI in the healthcare industry, we use it to evaluate our contracts and decide which we should agree to. I think the user can't be responsible - or else we won't use AI tools. We're lawyers; we don't know how to evaluate AI systems.
- Strongly disagree that this case is any different from any other software tool you'd buy.
- Over the last 6-8 months, we've found that the models we use to evaluate our contracts are making the same sets of recommendations to all users on things like indemnity clauses and executive compensation. This "exposes" other competitors' clauses and shares information across competitors, eroding earnest market competition.
- Is there an agreement under the Sherman Act, Section 1?
  - These might be the wrong questions for antitrust law to be asking.
  - Transparency and information sharing with AI tools is so profound that it allows for cooperation without explicit agreement.
- Taken to an extreme, doesn't this mean that any information transparency is collusion?
- Yes, just as with the gas station public pricing example.
- The DOJ recently repealed three statements of policy around information sharing - regulation by enforcement. Regulators are saying they might come for you if you break the law. They used to warn you that they were coming, and now they just come in the night.
- Our regulators offer no guidance anymore. We just figure out enforcement on a case-by-case basis. Our trade is wandering in the dark.
- Sophisticated tools that allow YOU to make educated decisions about competitor moves.
- We're concerned our tools are information sharing without our knowledge.
  - Could you explain the harm of this?
    - Limitation of liability clauses in agreements to purchase medical equipment indemnify product manufacturers.

    - If they find out competitors aren't using indemnity clauses, they may stop, causing the price of an MRI machine to go way up because everyone needs more insurance. So now the price of healthcare goes way up.
    - There's also the cost of cybersecurity and data breach liability.
- Should increased information sharing that standardizes industries constitute an antitrust violation? - Every lawyer in the room vehemently asserts that the standardization of contracts represents a significant harm to consumers.
- Is this sort of standardization of contract negotiation a form of collusion? If it is that essentially ruins the practice.
- Companies had agreed to use the same algorithm, so this goes a step further. It's not just price - it's other things as well.
- There was a moment in time where chessmasters no longer understood chess. AI made a move, and it looked awful but it turns out it was great. People still can't understand why it was great.
    - In the case of algorithmic collusion, what happens when AI is so smart it eludes the detection of collusion, and so throws us off the scent but still optimizes pricing for customers?
    - Is the law ready for this moment? Who is responsible for this collusion?
- All of these scenarios assume that you are allowed to use your data to train the model, but at my company we don't allow any model developers to use our data. That's a hard requirement in all our agreements.
    - We really struggle with third-party oversight of models we providers use. The only way for us to manage the risk is to ask software providers if they are using first-party AI models in their software.
    - Nothing governs software providers' use of our data in this case.
- On the topic of algorithmic price collusion, NC Insurance and Maryland Hospital Systems have created a precedent for sharing transparent industry pricing information. Most of these are in the insurance realm.
- It sounds like there is a there there around algorithmic collusion and transparent information sharing through models, and the solutions might be fairly hardcore.

Tort is not fit for purpose; there's a massive amount of liability that still needs to be allocated across the AI value chain.

- Not much has been created in tort law. The doctrines evolved from railroads in the 19th century.
- At the state level, some statutes impose a duty on software developers.
- Models have the potential to create harms. Those harms are usually created by third parties.
- Third-party harm is so foreseeable that doctrines could have to evolve.
- I imagine that the courts will be very sympathetic to the argument that you are liable if your model allows someone to create a polio vaccine.
- Product liability is another issue, and it's being litigated right now vis-à-vis social media.
- The doctrinal answer is no, but the courts may see societal problems emerging and be incentivized to spit out new answers to solve these problems.
- Product liability is designed for insuring risk on a uniform product. That doctrine won't work well if everyone is using AI differently and AI is spitting out new and different answers.
- Another consideration is people bringing claims against social media and opioids under public nuisance law - the unreasonable interference with a public right. If AI starts interfering with public goods, a court could issue an injunction against relief and ban a model from operating in an area under public nuisance law.
- The fit between existing tort doctrines and AI is awkward.
- Stewarding that adaptation of tort law to AI will be judged without deep technical expertise, influenced by a widespread societal fear of this technology.
- With tort law, it could come down to one state court judge viewing things differently and issuing a massive broad injunction. There's lots of potential for interference with development of this technology. It would cost $100s of millions to settle an Attorney General lawsuit.
- Companies don't immediately change their behavior when a series of tort lawsuits are filed - people adapt slowly with new types of lawsuits.
- The bottom line is that tort law is not at all ready for AI. It can't even handle social media.
- We are talking about very real harms here. Torts do not cover more ephemeral harms like antitrust.

- One scenario is to have states pass rational tort legislation and make clear through that legislation who is liable for certain types of harms. This would allow for the rational treatment of services through the tort system.
    - There's also the potential for other types of lawsuits. There's the issue of consumer protection - you may find out you broke the law only after the fact through enforcement.
    - Enormous liability with no real connection to harm.
- This is far from my area of expertise. What's the role of contracting to figure out the right area of liability? For example, airlines aren't liable for engine failure, the engine manufacturer is, because that's what they agreed to in the contract.
- This is the "least cost avoided" idea - the idea that the entity that could most easily avoid the harm is held liable.
    - Levels of liability that the system itself can't absorb, ranging from $100 million to $5 billion in payoffs.
    - There are so many types of harms, so many potential plaintiffs, and so many government entities that could regulate in the case of AI that the payoffs could break the industry.
    - The risk is that if a court decided a model developer was liable, they'd be put out of business.
    - This is also an argument for them to prevent these harms in the first place.
    - But the Tort system moves so slowly that this wouldn't work as a means of preventing harm.
    - There's an enormous amount of liability that needs to be rationally allocated, but with so much complexity that could prove impossible.
- What about the lessons we learned from data privacy?
- State-based laws can be really problematic for startups and small businesses.
- Should AI be regulated at the state or the federal level?
- Are AI companies simply going to come to Congress and ask for an exemption to liability?
- Section 230 was passed in response to state-level legislation that threatened the internet. Will there be a shield created to encourage AI development? Or an AI Manhattan Project?
    - Perhaps we're too eager to set our policy in stone when a lot of this stuff is still in early development.
    - AI will enable us to come up with governance solutions that help in the governance of AI.
- It's important to bring up a flexible and iterative policy approach.
- I like liability as a governance mechanism - it can help set proper incentives at a structural level right now. But we need good ideas on the table when Congress does eventually freak out and decide to pass a law. What liabilities should we all agree to disincentivize?
- What kind of structure do we need for AI specifically since tort doctrines don't fit well?
- Things with disastrous consequences must be curtailed now at the developer level, in advance.
    - We can't wait for the damage to be done.
    - States are acting now for the disastrous stuff at the developer level because it has to be prevented.
    - The fear of things that could happen is too great.
- I respectfully disagree that there must be intervention prior to a problem taking place.
    - The number of potential market failures is theoretically infinite.
    - Without specific examples, information is limited.
    - Theoretical risk is not good enough. I want to see specific market failures.
- What harms can we see and prevent now without impeding human progress?
    - That's impossible to know until we see evidence.
    - What if it's self correcting? What if it prevents some harms but causes others?
- Is there some level of minimum harm we should be able to see coming so we could impose requirements on developers?
- The closest analogy I can come up with is the analogy of being a parent. Laws govern that. I'm responsible for a while and then at some point the child becomes responsible.
    - If malicious code capability is adapted by the model on its own, the intention of the developer is really far apart.
- Is there a new question about agency we have to ask? Is the law ready for AI? Then there might be a question about resolving the question of agency within jurisprudence? Is it the parent or the child that is responsible?
- How do you incentivize behavior?
- I've been told that all models should have to be incorporated as an entity.
- This would massively throttle innovation.
- Is building polio illegal already, yes? But is building and releasing a model that self-recursively improves, extricates itself from its weights, and hides from humanity illegal? No. But it should be. With polio, I'd rather get ahead of it.
- No, the law is not ready for AI, but we won't know when it is because the tech is moving so fast.

- We need to think about it on an international level. Maybe you could have international cooperation around very extreme risks.
- Don't regulate the models, regulate the applications of the models.
    - That's intuitive on some level, but then policymakers micromanage the use cases of AI in a way that is very "presentist" and not well adapted to change.
    - It freezes in amber current capabilities without foresight or consideration for how those may change in the future.
- We should incentivize developers to compete along the safety front - to do a better job, and take precautions - while maintaining flexibility for the future.
- We need to deal with the catastrophic stuff first.
- But that stuff is already illegal.
- It's illegal for me to attack Middleburg but not for me to create a machine that automates the taking down of IT infrastructure.
- Regulating developers is more effective for a smaller number of labs.
- If I fine-tuned a model to make it more destructive, then I would be liable for changing the model.
- There's strict liability for inherently dangerous activities, so the analogy here is that you are liable regardless of reasonable care if you develop inherently dangerous models.
- This is fine and good if you're able to identify specific harms. But even if you could do this, what about international problems? Anyone could guess wrong and hamstring your industry while gaining no real safety.
- What about regular tort law? Does pulling catastrophic risk into the conversation muddy it?
- Is AI like an untrainable employee, where you can't train an AI like you could an employee to hide risk? Or is it the perfectable trainable employee?
- Do we have a law that criminalizes building a machine that does a bad thing? Product liability may deal with that.
    - And then we also have to talk about enforcement and detection of the thing.
    - We might have to use AI to figure out if AI is violating existing laws.
- This concept of catastrophic risk mitigation has been solved in banking regulation.
    - All of you are raising the right points.
    - Systemic risk management, transparency requirements, international.
    - Banks cannot deploy a model without regulators knowing all parameters, all the drift, and all the benchmarking.
    - We didn't do that in a day.
    - We needed several global market crises first.
- Banks have a long history of working with the federal government.
    - This evolved over a very long time. Centuries of financial system work.
    - We're only at the beginning of that process with AI.
    - We should start with baby steps.
- The ways banks are regulated are imperfect, and this can also be the case for AI.
- Banks may be cautious with deployment. But every other industry is the wild west. Industries win when they are willing to break the law.
- I am concerned with the competitive dynamics of small companies competing and not having the resources to test effectively, and just pushing forward to try to win without compliance.
- Through this conversation, we're pointing to developer-heavy solutions.
    - But the law will also apply to consumers who use it in irresponsible ways.
    - We're inadvertently collapsing our argument into a legal regime for a very few when the law still applies to all.
- Should we regulate on the front side - regulate developers - or on the backside - regulate users?
- Safety testing is very existential risk-focused. B2B models do poorly on those tests because the model use cases are more narrowly targeted on helping those businesses.
- In the case of B2B models, we have foreseeable insight into how customers will use the models in the future but we are unable to actually see use cases of models once deployed because you can have downstream deployers building on top of the developer's work, which the developer doesn't have any visibility into.

There might be room for a semi-private, quasi-governmental body like FINRA to regulate AI.

- What is the enforcement mechanism here that works?
- There are things we could do that are really light-touch.
- Wouldn't want to disadvantage local companies internationally.
- I also think compute providers could act as an intermediary for raising red flags.

- There's also FINRA in the financial sector - a safe, third space / entity acting in a watchdog capacity saying "hey, you're edging toward dangerous territory." Is there space for that in AI? It's in the industry's own interest to watch over itself to protect it from further penalties or legislation.
- There are offices of innovation that have cropped up in financial services to allow companies to sandbox new ideas and issues.
- I like the multi-dimensional oversight regime you're talking about.
    - FINRA and the financial services ecosystem work together.
    - In AI, there's the Partnership on AI. There's also the Frontier Model Forum.
    - But I think there needs to be more teeth to it in AI.
- Polling suggests that a multistakeholder approach is really good.
- I have trouble imagining how we will govern AI given how quickly it moves compared to the government, so I quite like a body that is part government and part industry to help govern itself.
    - I would be worried about a government that could move as quickly as AI.
    - Due process slowing our government down is a feature, not a bug.
    - Private governance could make sense for successfully governing AI.
- Financial feedback on NIST and RMF has been good and well received.
- In this political climate you cannot stand up a new regulator, period. So, it'd be good to create something outside of government.
- These multistakeholder processes only work when the incentives are properly aligned for all.
- The point about no new regulatory bodies makes a lot of sense. But folded into the recipe has to be aligning the incentives with growth of one's market position.

There might be room for a semi-private, quasi-governmental body like FINRA to regulate AI.

- The EU AI Act has not come up yet. What do we think of its shortcomings or wins?
    - The EU showed zero humility with their act. They don't even have any good companies, except perhaps Mistral.
    - They've gone down a very specific route. It's helpful for others to look at what they did, and there's still a chance for it to be brilliant.
    - They've argued they can change it up every 6 months but that never actually happens.
    - Hats off to them for trying to do something good.
    - I'm worried that it's not great.
    - Brits don't want to do either the EU or the American thing. They want to go their own way.
- The room seems to be more on the side of favoring up-front regulation and up-front testing, whether it's private or public.
    - My question is: What is the result in the market when it comes to industry players?
    - Large companies will have the resources to comply, but then we won't have innovators and disruptors starting up and doing new things.
    - Maybe that's good, but it's something we need to be open-eyed about. Does this entrench certain players?
- Tort law still applies to everyone right now, and the people most at risk in the current structure are small and medium players.
    - By focusing on frontier companies and the major marginal risk that those models add to the world, you then actually leave the door open to innovation from the smaller startup companies.
- One piece that's been missing from this conversation is technical expertise. We need people deep in the technology to understand what's possible and what's important.
- I think we have answered the question. The law is not ready for AI.
- On the tort piece, I think torts will need to be modified.
- But we need much better thinking about whether the institutions, the lawyers, and the courts are ready for AI, or whether someone else should be doing the regulation.
- Do we need a more active and nimble legal framework? What institutions could we create to be sufficiently adaptable without deferring innovation or competitiveness?

Workshop 3

Security in a World with AI

- Labs should treat their models as national strategic assets - but don't.
- There's a lack of guidance and support for the private sector on security.
- We all need to do a better job of defining what we're talking about when we talk about security; what is the nature and scope of the threat?
- How can the government and labs better coordinate?
- How can we secure lab inputs?
- How should we deploy AI in critical infrastructure?

Labs should treat their models as national strategic assets - but don't.

- As AI systems grow in capability, they will increasingly be viewed as strategic national assets central to economic growth, military advantage, and technological dominance.
- The most cyber-capable adversaries of the U.S. have made public statements about their intent to catch up and lead the world in AI.
- Cyberattacks designed to steal or sabotage American AI are well within the operating practices and capabilities of these state-level cyber actors.
- Targeting American AI will likely become a top priority of sophisticated state-level cyber actors and intelligence agencies (OC5) within the next five years.
- Given adequate prioritization, sophisticated state-level actors could successfully infiltrate leading American AI companies.
- Adversaries could steal AI model weights, codebases, experimental results, or algorithmic secrets to accelerate their own AI programs and undermine an American AI lead.
- Adversaries could also covertly sabotage American AI to slow the pace of our research, degrade our ability to deploy AI systems, and even insert backdoors to later take control of deployed AI systems.
- There are weak levels of personnel security and cybersecurity around frontier models right now. Labs don't treat their models like national strategic assets - but they should.
- Big picture issue: There's a mismatch between the pace of model improvements and cybersecurity improvements. Rather than expecting companies to unilaterally slow themselves down, we need industry-wide public-private partnerships to address this.
- How do you compel companies to get serious about security? Get them on a government contract. On DOD contract, NSA, CIA. These come with certain security requirements. Bar that, we've never been able to find a way to make companies be serious about their own assets.
- The government has a responsibility to protect companies as it did with the manufacturers of COVID vaccines.

There's a lack of guidance and support for the private sector on security.

- Someone from a traditional "enemy state" was coming into my company as a Chief Data Scientist, but in my industry we don't have the resources to do background checks and curate our teams with security in mind. There's no infrastructure for this.
  - To enable this, the government needs to offer funding and support to particular industries, like healthcare.
- To whom should the government provide resources and support for security? The government isn't scaled to do that kind of work for everyone.
- We've been tracking California's Fair Chance Act, which prevents companies from doing background checks on potential employees before making them an offer.
  - What is the impact of this on security? Maybe there needs to be an exemption in the case of certain critical technologies?
  - Who should be subject to what kind of screening / background check?
- Everyone is being made to fake it - hire a CSO on their own, and make it up on their own. Everyone is being made to reinvent the wheel.
  - I'd love a list of cost-effective, vetted security vendors. I want us to develop an ecosystem of assurance labs with cybersecurity as a pillar.
  - How do you create a marketplace for affordable tools and services that can assure security? Rather than what we have now, which isn't quite chaos, but you need to be lucky and connected to find the right people.

- The marketplace would emerge if companies had to follow a standard - either a regulatory requirement or a strong voluntary standard. Then a marketplace would emerge. In some cases, based on the risk, a voluntary approach would work. In others, you need a hard standard. Tailored based on the function and the nature of the risk.
- The marketplace can create itself once the parameters have been set.
- Do the developers have a financial responsibility for funding that marketplace, that ecosystem, given what they're putting out into the world?

We all need to do a better job of defining what we're talking about when we talk about security; what is the nature and scope of the threat?

- The government needs to have visibility into national security risks, so as to decide what to control and what not to control. Need a mechanism for companies working with governments to assess national security risks.
- In the biosecurity world, there seems to be a hope / expectation that intelligence briefings will organize the problem and create a framework for solving it. That's not enough. What we need to think about is: what are the capabilities we most fear losing or our adversaries gaining? We need to shift our focus away from: what could all the many possible threats be, and from who? The government needs a position on this - on what we fear losing.
- The government has the best knowledge of what threats we face but we can't know the plans of all our enemies, so we shouldn't anchor on threat vectors. We should focus on figuring out which are the valuable capabilities we need to protect.
- The problem is not educating the government about the risks - it's that we talk in the abstract about these things. Even the worst, dumbest member can carry something to the finish line. The problem is that they don't know what the hell you want them to do. Give them a thing to do, and make it a digestible idea.
- Need to tease out exactly what we're talking about - there's software, there's hardware, each with its own set of problems and challenges. We need to be clear and specific.
- Need to separate AI as the target from AI as the modality for a whole variety of threats, and that's where you bring in other state actors, counter-terrorism.
- If we skip the problem of defining the problem properly, we get into the evergreen conversations around policy.
- I'm always astounded that Senators need to be educated about security and yet every week we have security breaches that cost millions of dollars to the healthcare infrastructure.
- We need to get clear about what is at stake. Don't just tell policymakers what to do, but what we could lose. What are the national security / military problems that could develop? What military and national security problems could occur in the next 1-3 years?
- Need to get clear about the components of security. There's the adversary - we need to identify who the adversary really is. Is China just a competitor? Is it just a market, or is it an adversary?
- There's also different types of actors.
- Data is another component.
- The geography of data centers also plays into this. This is a key component in the security question.
- The data centers of the future are being planned now. If they're built in the U.S., they're easier to secure. We should streamline the buildout of data centers.
- This whole conversation is predicated on the assumption that the algorithms are closed, that the model weights are held in security, but that isn't necessarily the current state of play. Lots of this is "don't want to lose their secrets," but you have some companies just giving that away. This whole conversation doesn't matter if you're open-sourcing. There's no need to worry about who's working in your company.
    - Is that a problem? Well, if yes, that assumes the labs who keep them closed-sourced maintain a competitive advantage.
- How does the open-source thing relate to security?
    - Optionality. How do we maintain option-value over future decision-making?
- What are we trying to protect? Confidentiality? Integrity? Availability?

How can the government and labs better coordinate?

- How can the government and industry coordinate on security—specifically, intelligence sharing, personnel screening, and contract-based incentives?
- Personnel Screening and Counterintelligence:
- Lab-Specific Limited-Access Authorization
- New Clearance Category for AI Labs
    - Propose creating a limited-access authorization narrowly scoped to AI labs, designed for employees who require exposure to sensitive data, model architectures, or national-security-related AI capabilities.

        - This limited authorization could potentially allow foreign nationals—with careful vetting—to receive equivalent clearance, ensuring labs do not lose out on top-tier global talent.
    - Leveraging Existing Mechanisms
        - Alternatively, labs could use existing U.S. government mechanisms (e.g., classified information labeled "SECRET / REL TO [group name]") to grant restricted access only to individuals with a defined "need to know."
        - In this approach, the government would establish a "REL TO (AI Lab)" classification or something similar, restricting access to validated lab employees.
    - Expedited U.S. Government Security Clearances
    - Core Personnel Needing Clearances
        - Specific roles—such as the Chief Information Security Officer (CISO), the Head of Research, the General Counsel, and possibly a few senior technical leads—could be fast-tracked for SECRET or TOP SECRET clearances.
        - Model this on the DoD Special Access Program (SAP) Corporate Portfolio Program, which selects certain individuals in private organizations to hold high-level clearances for sensitive projects.
    - Rationale
        - Trusted personnel at AI labs must be able to receive classified threat intelligence, or quickly coordinate with government authorities if they identify signals of infiltration or sabotage.
    - Insider Threat Program Assistance
        - The government could provide direct help in designing and reviewing lab insider threat programs.
        - AI companies might be asked to certify that they have comprehensive insider-threat monitoring, personnel screening, and ongoing training, in line with national security standards.
- Threat Intelligence Sharing:
    - Dedicated Secure Network (JWICS-Equivalent for AI)
    - Establish a Classified Communication Pathway
    - Explore creating a "JWICS-equivalent" channel for labs to both receive and transmit classified or sensitive threat information relevant to AI security.
    - The system could be a scaled-down version of existing DoD or IC networks, with access restricted to cleared lab personnel.
    - Regular / Quarterly Briefings
        - U.S. government agencies (e.g., CISA, intelligence community) would convene quarterly, or as-needed, classified meetings with cleared lab representatives.
        - Labs could reciprocate by sharing indicators of compromise or suspicious network activity in real time.
- Frontier AI Information Sharing & Analysis Organization (ISAO)
    - Formalize the Info-Sharing Process
        - A specialized ISAO devoted to frontier AI labs could partner with CISA's Joint Cyber Defense Collaborative.
        - Focus on threat intelligence, best practices, real-time incident response, and broader coordination with the government.
- Contracting and Incentives:
    - Security Requirements in Government Contracts
    - Penetration Testing and Compliance
        - Any organization seeking government contracts for large-scale AI model training (e.g., models exceeding certain parameter counts) would have to comply with security regulations (e.g., NSA penetration testing, robust insider-threat protocols, secure supply chain documentation).
    - Tying Contracts to Strategic R&D
        - For labs engaging in high-priority research areas (e.g., bio, chem, fusion, or other technologies with national security implications), the government could offer accelerated R&D contracts that come with mandated security clauses (pentesting, screening, and restricted data management).
- Public-Private "Cooperative Organization"
    - Formal Partnerships
        - The government could foster a new PPP or cooperative entity (e.g., a "Frontier AI Consortium") where membership requires adopting stringent security and screening standards.
        - In return, members might receive benefits such as faster security clearance processing for key staff, subsidized secure computing facilities, and direct intelligence support.