CSET CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

Written Testimony of Dr. Dewey Murdick

Executive Director
Center for Security and Emerging Technology, Georgetown University

For a House Committee on Science, Space and Technology hearing on "Artificial Intelligence: Advancing Innovation Towards the National Interest."

June 22, 2023

Thank you, Chairman Lucas, Ranking Member Lofgren, and members of the committee for this opportunity to discuss how we can make AI better for our country.

There are many actions Congress can take to balance supporting innovation in AI (**run faster**), limiting how AI technology is misused by authoritarian governments (**tech protect**), and improving user safety for AI-enabled systems (**consumer safety**). I want to highlight three points to keep the U.S. at the forefront of AI:

1. **We must get used to working with AI** — we need to learn when to trust our AI teammates and when to question or ignore them. Understanding what the AI can and cannot do (or should and shouldn't do) is very important.
2. **We need skilled people** — both to build future AI systems and increase general AI literacy.
3. **We need to keep a close eye on our policies** — every policy we put into action should be monitored to see if it is working or if adjustments need to be made. This is especially true when facing peer innovators because it will allow us to act and react quickly to any moves made by rivals like China.

Let's talk a bit about China's role

China is highly committed to advancing its AI capabilities. It uses legal, extralegal, and occasionally illegal means to leverage its state powers, and has made substantial research and development investments to strengthen its position. China will remain a significant peer to the United States and European Union in many critical areas.

China focuses on different areas in AI than the United States and our allies do. For example, in my opinion, China's technical innovation priorities can be simplified (perhaps overly so) in terms of:

- Mass surveillance technology (e.g., facial recognition, gait detection, information operations) — a top priority that enables China to maintain control over information dissemination, monitor the population, and ensure political stability.
- Command and control technology (e.g., centralized decision-making, enhanced situational awareness, military force coordination compliance) — an essential capability for a country whose leaders want to maintain control over military operations and may or may not trust leaders across the command structure to faithfully execute the central authority's intentions.

Interestingly, despite China's strong commitment to AI advancement, deploying large language models (LLMs) doesn't appear to be a high priority. It seems LLMs may be at odds with the interests of the Chinese Communist Party, which can be attributed to concerns surrounding the risk of providing the nation with tools that could generate politically-sensitive outputs.

Nevertheless, we need to remember that China isn't 10 feet tall. (Interestingly this risk goes both ways as there are signs China tends to overestimate U.S. capabilities, too.) The United States should not let a fear of Chinese AI advancements deter oversight of the AI innovation. Instead, we should concentrate on developing strategies for managing the risks of AI, protecting people's rights, and guiding the advancement of AI technology. This will ensure that the United States leads in setting the global rules of the road, and that those rules reflect values we hold dear.

Get used to using AI via human-machine teaming

I believe it's important to teach people when to rely on AI systems and when to disregard their output or avoid using them. This *learned trust* is a theme that will be central to AI use over the next decade.

- The worst-case scenario is when people trust AI systems when they shouldn't (like driving into a lake because your GPS said so), don't trust them when they should (missing out on massive productivity gains), or systems manipulate their human counterparts to achieve some goal.[1]
- The best outcome is a balanced trust between humans and AI. This comes from good design, testing, implementation, and lots of training so everyone knows what these systems can and cannot do.

I believe this theme should drive our innovation and training standards.

[1] We may have already seen hints of this when GPT-4 AI managed to convince a TaskRabbit worker to solve a CAPTCHA or when a large language model named Eliza allegedly encouraged a man's recent suicide in Belgium.

The recent leap forward in AI technology is difficult to ignore. Many among us have interacted with advanced AI tools such as OpenAI's ChatGPT, Bing Chat, Google's Bard, Midjourney image generation tool on Discord, or Github's Co-Pilot. Each one of these tools is a testament to the rapid advancements we've seen since 2018. These changes have caught 100+ million people's attention, and policymakers are rightly asking what we do in the AI-infused era to maximize the positives these tools offer and address their inherent risks.

It is hard to predict where the next innovation will come from, which is why it is wise to invest in the full range of sources of future innovation — talent, plus the triad of algorithms, compute, and data. Future innovation may come in any number of areas from new types of algorithms.

Based on the research trend forecasts that I've seen (for example, here), innovation will likely drive the convergence of AI with new sectors and areas. Examples include medical imaging (better scans for wiggling kids), AI-enabled control of nuclear fusion research into new energy sources, AI advances in signal processing on the battlefield, a transformed manufacturing ecosystem, and so much more.

This means an increasingly large cross-section of the population will encounter AI daily. It is imperative that everyone has a baseline AI literacy, not only to be economically competitive, but also to be an informed citizen that contributes to the difficult societal conversations we must have about the development, deployment, and use of these novel tools.

Develop Talent

The importance of people has been CSET's top, most consistent finding since we started doing analysis in 2019.

First, we need to prioritize general AI awareness and how to effectively operate in these new human-machine teams. We need to work together as a society to decide how to use these systems — using our democratic processes. There is an urgent need for widespread training and awareness about these systems. Robust training has given our firefighters and military personnel a baseline level of trust in one another, an understanding of the team's capabilities, and an expectation of where and how a teammate will act. We need similar expectations for AI.

As Helen Toner from CSET said, these new AI systems are like actors in an improvisation. I have found tools like ChatGPT incredibly powerful in certain contexts, but through measured use over time, I've come to understand where it can help and where it is inclined to hallucinate. People need to realize when they have a useful partner and when their improv teammate is making things up and refusing to end the scene.

Second, we need to focus on getting more people skilled in creating today's and tomorrow's AI tools. The AI workforce, comprising both technical and nontechnical roles, is vital for effective AI development and deployment. AI is not just a field that can be left to the elite PhDs and engineers building these systems.

We cannot afford to underutilize any talent pool — we need to include talented individuals regardless of where they come from, what they look like, or their socioeconomic statuses, and ethnicities. And we need to make sure we keep attracting the world's talent to the United States with effective immigration programs and policies. AI talent development and retention of diverse domestic and international talent is essential for both technical and nontechnical professions.[2]

Usually this can take a long time, but Congress has the power to help right away by facilitating a targeted high-skilled immigration program. Congress can also incentivize efforts to advance general AI literacy, grow more STEM talent, and promote certification programs within the United States:

1. **Initiate High-Quality K-12 AI Education:**
    - Collaborate with states to develop K-12 AI education strategies and computer science standards.
    - Fund and partner with nonprofits to develop AI curricula and class materials for K-12 education to promote general AI literacy and encourage specialized study.
2. **Boost AI Talent Development Beyond Four-Year Degrees:**
    - Enhance community college involvement in the AI talent pipeline, focusing on fostering diversity and bridging gender gaps.
    - Encourage development of AI and AI-related programs, focusing on stackable credentials and facilitating two-year to four-year institution transfers.
    - Develop a federal grant program or tax credits promoting collaborations between industry and academia in AI and related fields, including technical and community colleges.
    - Strengthen non-degree programs in AI-related fields through reform and expansion of existing federal programs.
3. **Promote the Use of AI Certifications:**
    - Set federal hiring criteria to give credit for AI-related certifications to boost their utilization and to serve as a model across the industry.

[2] China has a widely adopted K-12 curriculum and produced roughly four times more science, technology, engineering, and math (STEM) graduates in 2019 (China graduated 1.6 million undergraduate students, compared to the U.S.'s 412,000). It is also rapidly growing its AI-related PhD capacity.

4. **Enhance Institutional Infrastructure and Support:**
    - Establish a dedicated office within the White House to strategically support AI education in community and technical colleges.
    - Secure state and federal financial backing to improve AI and AI-related programming.
    - Commission entities like NIST to create and regularly update a comprehensive framework for technical and nontechnical AI roles and competencies (similar to the NICE framework).

Monitor the Effectiveness of Policy Actions

I hear three strategic goals being implicitly voiced as U.S. leaders and representatives navigate issues at the intersection of emerging technology (like AI), national or economic security, and public policy. How do we simultaneously:

1. Drive technological innovation (**run faster**)?
2. Impede our adversaries' progress in crucial areas (**tech protect**)?
3. And promotes safe, secure, values-driven deployment of AI-infused capabilities (**consumer safety**)?

Each of these three goals reflects a set of policy levers that we have at our disposal. However, there are often tensions between these three goals and the levers that can be used to implement them.

- For example, when we impose export controls on certain chips to slow down our adversaries' advances in AI, could this disincentivize U.S. companies from developing and producing the most advanced semiconductor technology? Or more fundamentally, do we know if the controls have slowed down the adversary in question? A similar situation arose for the satellite industry in the 1980s.
- Similarly, does a blind race to out-innovate all others lead us down the road paved by software development, where "move fast and break things" puts people at risk because the fear of getting behind outweighs a paradigm that drives security?

Actions focused on one goal will affect the others, sometimes adversely. Balancing these three goals requires carefully crafted use of the full array of policy levers (read more here), but more importantly it requires constant monitoring & evaluation, and a rapid adaptation to lessons learned.

It is critical that we monitor and evaluate our actions in all three of the policy goals (run faster, tech protect, and consumer safety). If we establish export controls, we ought to evaluate whether they are cutting off our adversaries' access to the capabilities we hope to control, stifling U.S. competitiveness or leading to workarounds that are slowing AI innovation as a whole. Each time we decide to pull one of the many levers, we should be evaluating its effects and recalibrating as needed. As of now, we don't have this comprehensive capability in the U.S. government. It's a new capability that we need to account for when implementing new policies. The proposed *Office of Global Competition Analysis* and other models are under consideration in the House and Senate. This analytic capability could expand our monitoring ability and could increase our agility when facing peer innovators.

Gather Information and Upgrade Shared Resources

In the short-term, there are steps we can take to promote the development of AI that aligns with U.S. values. We should find ways to get the information we need to make better decisions, including by:

- Tracking AI harms in incident reporting (including voluntary and required reporting);
- Inquiring into the data and models used in existing high-consequence applications to increase accountability (e.g., sentencing algorithms); and
- Encouraging the development of the third-party auditing and red teaming ecosystem.

And, improve the quality of shared resources, such as open source training and pretrained models that form the backbone of many of today's AI systems.

Update Authorities and Oversight Capabilities

In the longer-term, we need to rethink our approach to AI software development and deployments. Unlike with food, drugs, or finance, software is often released without thorough testing — this needs to change. There is a spectrum ranging from our current approach to more rigorous testing procedures, such as those used in clinical trials. We need to think about where we should be on that scale, but right now we're at one extreme end. We need to make changes over time to the rules about who is responsible when things go wrong with software, especially those end-user license agreements that let the companies providing the service off the hook for any or all problems.

If we think about how to license AI software and make sure it's used safely, we might need to update the authorities for existing agencies or even create a new one to oversee this. This new, semi-independent body could:

- Check how AI is being used in and overseen by existing agencies;
- Be the first to deal with problems, directing them to the right government or private resources; and
- Fill any gaps where existing sector-specific agencies don't cover.

However, I believe there are a number of questions that we need to be asking here before we take this step:

- With the infusion of AI, which existing authorities need clarification or expansion? Which agencies should wield these updated authorities?
- What should the scale of these authorities be and how would we pay for them (e.g., system licensing, organizational certification, or appropriate government agency utilization)? How do we help avoid regulatory capture by major private sector organizations?
- Are there good organizational models that can be adapted? Can an existing agency be repurposed or does it require something entirely new? Which model would best help departments navigate within a rapidly evolving application environment?
- How long will it take to get new authorities or a brand-new agency in place? How much will potential political and technical challenges slow down the establishment of a new organization?
- How can we empower sector-specific regulators to apply their expert knowledge, while ensuring cross-sector issues are addressed?

Many have proposed an international regulatory body for AI given its global nature. While international engagement is crucial, the United States can ensure leadership through our established and respected domestic approach. Figuring out a model that allows a lead organization to seamlessly work with organizations as diverse as the Federal Aviation Agency (FAA), Food and Drug Administration (FDA), Federal Trade Commission (FTC), and the Financial Industry Regulatory Authority (FINRA) is essential. Starting domestically is also a good way to preserve the democratic processes that are essential for a representative democracy.

This is not a problem we will "solve" in a few years and move on. AI is transformative and will require dedicated and consistent attention to get it right. I appreciate this committee's dedication to that effort.

Thank you,

Dewey Murdick, Ph.D.
Executive Director
Center for Security and Emerging Technology (CSET)
Georgetown University