STATEMENT BY
ALEXANDR WANG
FOUNDER AND CHIEF EXECUTIVE OFFICER, SCALE AI

BEFORE THE
SUBCOMMITTEE ON CYBER, INFORMATION TECHNOLOGIES, AND INNOVATION
OF THE
HOUSE COMMITTEE ON ARMED SERVICES

"MAN AND MACHINE: ARTIFICIAL INTELLIGENCE ON THE BATTLEFIELD"
JULY 18, 2023

Chairman Gallagher, Ranking Member Khana, and distinguished members of the Cyber, Information Technologies, and Innovation Subcommittee, thank you for the opportunity to testify today on the critical role that artificial intelligence (AI) will play in the future of our military.

I am honored to be here today to discuss why AI is the most critical technology in this next era of warfare and what the United States must do to win.

Introduction

My name is Alexandr Wang, and I am the founder and CEO of Scale AI (Scale). Scale was founded in 2016 with the mission of accelerating the development of AI. I am proud to say that Scale is committed to supporting U.S. national security and that our technology and platforms power the most ambitious AI projects in the world. From our earliest days of annotating AI data for autonomous vehicle programs at General Motors and Toyota, to our work with leading technology companies such as OpenAI, Meta and Microsoft, and the U.S. government, including the Department of Defense's Chief Digital and Artificial Intelligence Office (CDAO), U.S. Army, and U.S. Air Force, Scale has always been a leader in AI infrastructure development.

As someone who has been part of the forefront of AI development for more than seven years, it is exciting to see this technology finally reach its watershed moment. AI has come to dominate every conversation, every headline, and nearly every technological development we see today. At this critical juncture, the United States must recognize the urgency to navigate this new landscape because we risk ceding our global influence, national security, and democracy to an authoritarian regime.

Supporting the U.S. government and the national security mission is deeply personal for me. I grew up in Los Alamos, New Mexico, where my parents were physicists at Los Alamos National Laboratory, the birthplace of a technology that defined the last era of warfare - the atomic bomb. I was keenly aware that an emerging technology, like AI, could completely change global politics and the nature of war.

This is not a new realization or future speculation.

China Recognizes the Importance of Global AI Leadership

Four years ago, in 2018, I went on an investor trip to China that was both enlightening and unsettling. During this visit, I saw firsthand the progress that China was making toward developing computer vision technology and other forms of AI. I was troubled because this technology was also being used for domestic

repression, such as persecuting the Uyghur population. It was evident that the China Communist Party (CCP) had already strategized how to harness AI for advancing its military and economic power. As China President Xi Jingping declared that same year, "[We must] ensure that our country marches in the front ranks where it comes to theoretical research in this important area of AI and occupies the high ground in critical and AI core technologies.[1]"

China deeply understands the potential for AI to disrupt warfare and is investing heavily to capitalize on the opportunity: It considers AI to be a "historic opportunity" for "leapfrog development" of national security technology.[2] As of 2020, China had outspent the United States on AI technology for defense, both in absolute terms and proportionally. China's military arm, the People's Liberation Army (PLA), spent between $1.6B and $2.7B on AI against an overall defense budget of $178B in 2020, whereas the DoD spent only between $800M and $1.3B on AI against an overall DoD budget of $693B for the same period. China is spending between 1% and 1.5% of its military budget on AI, while the US is spending between 0.1% and 0.2%. Adjusted for the total military budget, China is spending ten times more than the US.[3]

This year, China is projected to spend approximately $14.75 billion on AI investments.[4] In contrast, the administration's FY24 budget request included roughly $5.5 billion for AI.[5] While this marks a historic investment by the United States in AI, we must intensify our efforts to outmatch China's rapid advancements.

The reason for this is that the United States is at risk of being stuck in an innovator's dilemma because it is comfortable and familiar with investing in traditional sources of military power. While we are making sense of this technology and conceptualizing a framework for how to use it, Chinese leaders are actively working to use AI to tighten their grip domestically and expand their reach globally. It's time to act. The U.S. must learn to embrace AI innovation before we are disrupted.

---

[1] See, https://www.newamerica.org/cybersecurity-initiative/digichina/blog/xi-jinping-calls-for-healthy-development-of-ai-translation/.

[2] See, https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

[3] See, https://cset.georgetown.edu/publication/harnessed-lightning/

[4]See, https://news.cgtn.com/news/2023-04-10/China-s-AI-market-spending-to-cover-10-of-world-total-in-2023-report-1iSPv1hUlWM/index.html#:~:text=Spending%20in%20China's%20artificial%20intelligence,International%20Data%20Corporation%20(IDC).

[5] See, https://www.pillsburylaw.com/en/news-and-insights/ai-biden-fy2024-budget.html

This urgency is highlighted by the results of an analytic exercise that reviewed thousands of pages of open-source data on the PLA adopting AI.[6] China's capabilities highlighted in the report should serve as an immediate wake up call. "PLA advances in AI and autonomy will create new vulnerabilities for the United States and allied forces operating in the Indo-Pacific." Further, it showed that "The PLA is stepping up investment in information operations and adaptive radar systems to jam and blind U.S. sensor and information networks, which PLA leaders judge to be particularly vulnerable."[7] "The PLA is also prioritizing the development of autonomous vehicles, specifically sub-surface and aerial platforms, that suggests it could confer an asymmetric advantage for the PLA in combat with the U.S. or similarly advanced opponent," according to the Center for Security and Emerging Technology at Georgetown University.[8]

From a purely technological standpoint, China has already surpassed the U.S. in computer vision and is a fast follower on Large Language Models (LLMs). In 2022, an aerial imagery object detection global challenge was conducted and the results speak for themselves– the first, second, fourth, and fifth place winners were all Chinese companies or universities.[9]

To close the gap, China is heavily investing and bringing the power of its domestic industrial AI base to support government-backed programs.[10] Since 2020, China has launched 79 LLMs[11] and there are frequent announcements about new national labs opening and state-backed AI companies being formed.[12] The reason for this investment is that "in the AI race between China and the U.S., AI research will be pivotal for China's future success – and hence too important to leave in private hands…State-sponsored AI research is China's Apollo Program."[13]

Scale's Commitment to U.S. National Security

As a patriotic American, I recognized the potential value of Scale's technology for national security use cases and committed to support the United States in preventing President Xi's vision from becoming a reality. For the past three years, Scale has proudly partnered with the U.S. Department of Defense–and stakeholders across the national security space–to integrate our best-in-class

[6] See, https://cset.georgetown.edu/publication/harnessed-lightning/
[7] iBid.
[8] See, https://www.army-technology.com/analysis/the-role-of-ai-in-the-peoples-liberation-army/
[9] COCO is the internationally recognized benchmark for image recognition. The leader board can be found here: https://paperswithcode.com/sota/object-detection-on-coco
[10] See, https://www.state.gov/wp-content/uploads/2020/05/What-is-MCF-One-Pager.pdf
[11] See, https://www.reuters.com/technology/chinese-organisations-launched-79-ai-large-language-models-since-2020-report-2023-05-30/
[12] See, https://thebambooworks.com/china-goes-it-alone-in-ai-2-0-drawing-on-local-funds-and-trio-of-industry-veterans/
[13] See, https://thediplomat.com/2023/03/the-future-of-state-sponsored-ai-research-in-china/

commercial AI technologies into critical programs that directly impact our national defense. Scale has top American AI talent[14] working on our programs and is actively investing in and committed to teaching and training a homegrown workforce of the future.[15] Our goal is to accelerate AI overmatch for defense and ensure the United States maintains its strategic advantage. This includes:

- *Scale Autonomous Mission Systems:* This year, the Defense Innovation Unit (DIU) selected Scale for a critical Army Program Executive Office for Ground Combat Systems (PEO GCS) autonomy. Scale has developed a data engine that is intended to support the Army's Robotic Combat Vehicle (RCV) and that data engine could power any Army autonomous system to enable a new generation of ground vehicles. This critical work has the potential to define the future of the military's work for ground, air, sea, and space autonomy.
- *Scale Data Engine:* Scale is working across government agencies to annotate and prepare vast troves of data into a high-quality resource that can be used to train AI models. This is laying the groundwork for AI Overmatch by creating a common data resource. For the U.S. Air Force Research Lab (AFRL), Scale builds and deploys advanced object detection and classification models onto secure networks and integrates those models with existing platforms such as the Air Force Distributed Common Ground System (AF DCGS) to give Airmen access to new AI capabilities within their existing workflows.
- *Scale Donovan:* In May 2023, Scale launched Donovan, our AI-powered decision-making platform, which is the first LLM deployed on Department of Defense classified networks. Donovan has the ability to ingest vast amounts of structured and unstructured data to make sense of any aspect of the real world in minutes using simple, natural language. Because it is compatible with the government's own data, end users could share these findings with other trusted networks. For example, a Naval officer could share their findings with intelligence analysts, who then use Donovan to explore a myriad of unstructured documents and quickly detect patterns and trends that would otherwise take weeks to verify and contextualize.

The Era of AI–Data is the Ammunition in AI Warfare

I firmly believe that the United States can still win the race for global AI supremacy, but for the U.S. to maintain this leadership, we must first understand how the landscape is changing and critically examine the DoD's current

[14] See, Forty-two percent of Scale's federal workforce comprises veterans based on self-reported data.
[15]See, https://www.businesswire.com/news/home/20220803005682/en/Scale-Announces-New-Office-in-Downtown-St.-Louis-to-Support-Local-Economic-Growth-and-Tech-Industry-Expansion

capabilities. AI-powered warfare will feature algorithm-fueled military planning, targeting, command and control, and autonomous platforms.[16]

Today, the United States and our allies are confronted with a very real challenge: legacy military platforms are being disrupted by AI. Those platforms, while still important, will be disrupted by cheaper autonomous drone fleets. For example, China has begun testing adaptive drone swarms,[17] which, if used in combat, would turn our legacy aircraft carriers into giant targets.[18]

In the intelligence realm, AI is already playing a critical role because AI applied to satellite imagery and other sensor data has enabled Ukrainian targeting and tracking of Russian troops.[19]

During the daily battle rhythm, the DoD creates more than 22 terabytes of data daily,[20] and because of their outdated data retention and management policies, warfighters, analysts, and operators are unable to tap into its full potential because it is not AI-ready. These potential insights are wasted. The Director of the National Geospatial Intelligence Agency publicly estimated that at the current, accelerating pace of collection, we would need over 8 million imagery analysts by 2027 to process all imagery data.[21] Without AI-ready data, there will be no way to keep pace with our adversaries.

DoD has been working for more than a decade[22] to solve these complex challenges. However, more needs to be done. Early on, and much like other emerging technologies, individual DoD units and end users began learning how to integrate AI into their operations. The DoD has recognized the limiting nature of this approach and the need for a unified strategy. As Deputy Secretary of Defense Kathleen H. Hicks said, "Artificial intelligence may transform many aspects of the human condition, nowhere more than in the military sphere."[23]

One notable step forward took place in May 2021 when Hicks released a memorandum kicking off the creation of the CDAO.[24] The CDAO is critical to

[16] See, https://www.amazon.com/Warbot-Dawn-Artificially-Intelligent-Conflict/dp/0197611699
[17] See, https://www.thedrive.com/the-war-zone/37062/china-conducts-test-of-massive-suicide-drone-swarm-launched-from-a-box-on-a-truck
[18] See, https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy
[19]See, https://www.washingtonpost.com/national-security/2022/05/11/ukraine-us-intelligence-sharing-war/
[20] See, https://www.defensenews.com/pentagon/2017/04/06/pentagon-tech-advisers-target-how-the-military-digests-data/"
[21]See, https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf
[22]See, https://www.politico.com/news/magazine/2023/06/15/pentagon-artificial-intelligence-china-00101751
[23]See, https://www.politico.com/news/magazine/2023/06/15/pentagon-artificial-intelligence-china-00101751
[24]See, https://media.defense.gov/2021/May/10/2002638551/-1/-1/0/DEPUTY-SECRETARY-OF-DEFENSE-MEMORANDUM.PDF

ensure a coordinated effort to both the DoD's AI work and its approach to data retention and management.

While this progress is promising, more must be done to achieve AI overmatch.

AI always boils down to data. All of the advancements in commercial AI technologies, such as ChatGPT, have come from using mass troves of data. For this reason, the DoD's own data strategy highlights the importance of prioritizing AI-ready data that is labeled, tagged, and annotated. Additionally, CDAO has the legislative mandate to establish a centralized data repository that will enable the DoD to leverage the power of its own data for AI overmatch. However, implementing this has been challenging because DoD lacks the proper data retention and management systems to operationalize it. Within the DoD, much of our key AI asset—our data—is being wasted every day. This concept is critical to enabling AI platforms of all kinds, but it relies on AI-ready data to succeed, and one of our most critical AI resources is not being used.

China understands this fact. According to an emerging technology expert at the Brookings Institution, "China is renowned for its data collection and thus algorithm development, which will likely define its advantage going forward…The U.S. struggles to reach equivalence in this area, so if China's data collection efforts make for a measurable improvement to its algorithms relative to U.S. ingenuity, China could take the lead."[25]

AI Overmatch–The Path to Global Leadership

To counter this growing threat and win the AI race, we need to achieve AI Overmatch. Adapting to the inevitable transformation of warfare in the AI era will require a shift in the DoD's approach to achieve data supremacy, investment in new technology, Pathfinder Projects, and personnel training. This can only be done successfully by 1) systematic collaboration among Congress, the DoD and industry and 2) developing a regulatory framework that encourages responsible innovation. Today, I would like to propose a five-part framework for achieving it. These pillars represent top-down and bottom-up shifts that should be considered to maintain the U.S.'s security and technological edge:

Investment: China is projected to continue to outpace American investment in AI. Unless we start to prioritize investment in both AI systems and the underlying data infrastructure to power it, we risk falling behind China and doing too little too late.

- While it is important to recognize that more must be done, Scale was pleased to see the FY24 President's budget request that recommends $1.8

[25] See, https://www.japantimes.co.jp/news/2023/04/20/asia-pacific/china-ai-future-wars/

billion in DoD AI investment.[26] It is critical that this funding is upheld through conference as another 12 months is too long to wait to adequately fund AI development.

Data Supremacy: AI systems are only as good as the data that they are trained on, and leading the world in developing AI-ready data is an absolute requirement to maintain our strategic advantage in the era of AI warfare. The advancements in LLMs over the past 5 years, including ChatGPT, have been achieved through training models on 1000 times more data than previously done. We must aim to accomplish a similar 1000 fold increase in our DoD AI implementations.

- Scale has been pleased to see Congress prioritize the CDAO and its legislative mandate to create a centralized data repository. It is critical that Congress continues to heavily invest in AI-ready data.

Test and Evaluation: One of the most important ways to ensure that AI models provide reliability and accountability for users is through a risk-based approach to test and evaluation with human oversight. We believe that this not only protects taxpayer resources by ensuring that Congress acquires high-quality AI systems, but also is one of the strongest methods to limit bias and uphold the DoD Ethical AI Principles.[27]

- Test and evaluation has long been a key part of the product development cycle for responsibly bringing consumer-facing technologies to market and military technologies into production. This is essential for AI applications because they are rapidly developing and constantly iterating, and therefore continually presenting new opportunities and risks to the end user.
- A risk-based approach to test and evaluation will ensure that AI is factual, accurate, and explainable regardless of the underlying model or data being used. If the product–including the data infrastructure and underlying model–does not meet these requirements, we risk sacrificing user trust in the technology.
    - The Biden Administration has embraced this concept by highlighting Scale's role building an evaluation platform for existing LLMs at the world's leading hacker conference, DEFCON, in August.[28] Scale recommends that Congress adopts comprehensive, risk-based test and evaluation criteria to ensure that AI will meet user safety and reliability standards prior to deployment within the DoD.

[26]See, https://www.defense.gov/News/Releases/Release/Article/3326875/department-of-defense-releases-the-presidents-fiscal-year-2024-defense-budget/

[27] See, https://www.ai.mil/blog_02_26_21-ai_ethics_principles-highlighting_the_progress_and_future_of_responsible_ai.html#:~:text=These%20principles%20encompass%20five%20areas,lifecycle%20both%20interactively%20and%20iteratively.

[28] See, The White House, Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation that Protects Americans' Rights and Safety, Washington DC, May 04, 2023

Pathfinder Projects: Congress and DoD need to commit to supporting new AI Pathfinder Projects–projects that have the mission and funding to solve unique DoD challenges.

- To date, the largest AI Pathfinder Project within DoD is still Project Maven, which began in 2017. In the past six years, there have been many important lessons learned, but no new efforts have been initiated.
    - There are endless DoD use cases that would benefit from being identified as a Pathfinder Project. For example, the Army is making progress on Project Linchpin and their ground autonomy work; Joint All Domain Command and Control (JADC2) requires DoD buy-in at all levels to succeed; and the Navy has discussed a concept called Project Overmatch, which would create a whole-of-Navy approach to AI adoption. While much work is being done to define these projects, Scale recommends that Congress pushes each branch of the military to formally identify its next Pathfinder Project and adequately fund it to be successful.

Personnel Training: The U.S. should invest in rapidly training and upskilling our military commanders and personnel on AI.

- Even with advancements in technology, humans always pay the price of war. The U.S should continue to invest heavily to ensure that its military has the best equipment, training, and leadership in the world, and part of the necessary training to succeed in the next era of warfare will be advanced AI literacy across all military units. This is crucial as we fully embrace the era of AI.
- Beyond simply training service members on AI fundamentals, the US should train commanders and personnel with necessary data skill sets to adopt AI in a way that will make multi-domain warfare a reality.
- Scale has the experience to understand this challenge and lend our expertise in a way that benefits the United States and economy broadly. We established a hub in St. Louis, which has created more than 300 tech-focused jobs, which range from entry-level labelers to machine learning engineers with advanced degrees. We anticipate growing these opportunities in the future. Scale looks forward to working with Congress and the DoD to identify technical gaps in current skill sets and how to best address those gaps.

Conclusion

The race for global AI leadership is well underway, and I could not be more excited to do everything in my power to ensure that the U.S. wins. This is one of the few true missions of our time that will define the future of war and global politics. We cannot sit by the sidelines, and it is in moments like this that

Congress, the DoD, and the tech industry can either rise to the challenge together or stand idle.

I am filled with a sense of optimism as we stand on the cusp of a new era, where these challenges are being met head-on by brilliant leaders across the public and private sectors. I am honored to work with this subcommittee to forge strong relationships between Congress, DoD, and the tech industry so we can collaborate and stay ahead of some of the most pressing threats of the next decade.

Thank you for the opportunity to be here today. I look forward to your questions.