HEARING BEFORE
THE UNITED STATES SENATE SELECT COMMITTEE ON INTELLIGENCE

September 19, 2023

Testimony of Yann LeCun
Chief AI Scientist, Meta

Introduction

Chairman Warner, Vice Chairman Rubio, and distinguished members of the Committee, thank you for the opportunity to appear before you today to discuss important issues regarding AI. My name is Yann LeCun. I am currently the Silver Professor of Computer Science and Data Science at New York University (NYU). I am also Meta's Chief AI Scientist and co-founder of Meta's FAIR (Fundamental AI Research), where I focus on AI research, development strategy, and scientific leadership.

Overview of My Involvement in AI

I have worked in areas related to computer science, machine learning, and artificial intelligence for several decades. In 1987, I received my PhD in Computer Science from Universite Pierre et Marie Curie (known as the Sorbonne). Following a postdoctoral research position at the University of Toronto, I joined AT&T Bell Labs in 1988, where I focused on developing machine learning methods, including an image recognition method called convolutional neural networks. Over the last decade, convolutional networks have become the dominant method for image understanding in such applications as driving assistance, medical image analysis and optical character recognition, among many other domains.

Since 2003, I have been privileged to serve as a professor at NYU. In 2014, my colleague Yoshua Bengio and I established the research program on Learning in Machines & Brains (formerly Neural Computation and Adaptive Perception Program), sponsored by the Canadian Foundation for Advanced Research, which I continue to advise today.

I have long advocated for the responsible and ethical use of AI. That is why, alongside representatives from Google, Microsoft, Amazon, and IBM, I co-founded *Partnership on AI* in 2016, a non-profit that brings together academic, civil society, industry, and media organizations to address the most important and difficult questions concerning AI. Together, *Partnership on AI* conducts studies, shares insights, publishes guidelines, informs public policy, and advances public understanding of AI.

I have been fortunate to have received various recognitions for my contributions to the field of AI. Most notably, along with my colleagues Geoffrey Hinton and Yoshua Bengio, I received the 2018 ACM Turing Award for "conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing."

As evidenced by my experiences in academia, I care deeply about AI research and collaboration. In 2013, I joined Meta (then Facebook) to create and lead an AI research division known as FAIR. I built and directed FAIR until 2018, when I moved to my current role as Chief AI Scientist. In this role, I establish directions for AI research, advise the company on AI strategy, and provide thought leadership to the broader AI research and engineering communities.

I am proud of Meta's work to serve as a leader in AI and its contributions to help ensure that America continues to lead in developing this critical technology. While my work at Meta does not involve developing products or policies, I engage with dedicated colleagues who work tirelessly toward developing and deploying AI in ways that enhance economic and social benefits, while also executing on guardrails that anticipate and mitigate potential risks. A central part of Meta's mission has always been to connect people to each other. That's still the core of what we do, and we are consistently working to improve and to share our innovations and lessons learned in AI with each other and with others including companies in the United States and around the world.

Current State of AI

AI has progressed leaps and bounds since I began my research career in the 1980s. We've seen first-hand how making AI models available to researchers can reap enormous benefits. For example, AI is being used to translate hundreds of languages, reduce traffic collisions, detect tumors in x-rays and MRIs, speed up MRI exams by a factor of four, discover new drugs, design new materials, predict weather conditions, and help the visually impaired.

Society's ability to develop AI tools to defend against adversarial, nefarious, or other harmful content derives in large part from our social values. Meta, by way of example, has organized its responsible AI efforts around five key pillars reflecting these values:

- First, we believe that protecting the privacy and security of individuals' data is the responsibility of everyone and we have therefore established a cross-product Privacy Review process to assess privacy risks;

- Second, we believe that our services should treat everyone fairly and have developed processes to detect and mitigate certain forms of statistical bias;

- Third, we believe that AI systems should be robust and safe, which is why we established an AI Red Team to test our systems against adversarial threats to ensure that they behave safely and as intended even when they are subjected to attack;

- Fourth, we are striving to be more transparent about when and how AI systems are making decisions that impact the people who use our products, to make those decisions more explainable, and to inform people about the controls they have over how those decisions are made;

- Finally, we believe that we should be accountable for our AI systems and the decisions they make, so we have built governance systems to ensure we meet high standards.

Today, we are witnessing rapid advancements in the development of generative AI, and in particular large language models (LLMs). These systems are trained through self-supervised learning, or more simply, they are trained to fill in the blanks. In the process of doing so, the AI model learns to represent text, including the meaning, style, and syntax, in multiple languages. This internal representation can be applied to downstream tasks, such as translation and topic classification. It can also be used to predict the next words in a text, which allows LLMs to answer questions or write essays.

There's no question that people in the field, including me, have been surprised by how well LLMs have worked. Though they have great promise and potential, it is important to keep in mind that LLMs do have limitations. While LLMs can unlock a host of new possibilities in industries, from health care to logistics to manufacturing, they have limited abilities to reason, a prevalent feature of human intelligence. As the technology exists today, even the most powerful AI systems are quite far from approximating human intelligence. For example, a child can readily learn to clear the dinner table and fill the dishwasher, but AI tools have not reached the level of domestic robots. The language fluency of LLMs may suggest human-level intelligence, but it is far from that.

Future of AI & The Importance of Open Sourcing

The current generation of AI tools is different from anything we've had before, and it's important not to undervalue the far-reaching potential opportunities they present. However, like any new disruptive technology, advancements in AI are bound to make people uneasy. I can understand why. The development of AI is as foundational as the creation of the microprocessor, the personal computer, the Internet, and the mobile device. Like all foundational technologies, there will be a multitude of uses of AI technologies, some predictable and some less so, which can be alarming. And like every technology, AI will be used by people for good and bad ends. This will not be the first time that bad actors try to use developing technology to their own ends, like phishing scams or spreading misinformation.

As AI systems continue to develop, I'd like to highlight two defining issues. The first is safety. New technology brings new challenges, and everyone has a part to play here. Companies should make sure tools are built and deployed responsibly. And all of us – policymakers, academics, civil society and industry – should work together to maximize the potential benefits and minimize the potential risks. The second is access. Having access to state of the art AI will be an increasingly important driver of opportunity in the future for individuals, for companies, and for economies as a whole.

One way to start to address both of these issues is through the open sharing of current technologies. At Meta, we believe it is better if AI is developed openly, rather than behind closed doors by a handful of companies. Generally speaking, companies should collaborate across industry, academia, government, and civil society to help ensure that such technologies are developed responsibly and with openness to minimize the potential risks and maximize the potential benefits.

The concept of free code sharing in the technological ecosystem is not new, it started long ago. The 1950's and 1960's saw almost all software produced by academics and corporate research

labs like AT&T's Bell Labs working in collaboration. Companies like IBM, DEC, and General Motors set up user groups to facilitate sharing code among the users. The infrastructure of the internet and all cloud computing services run on open-source software (Linux, Apache, MySQL, JavaScript, for example), as do most web browsers and many of the apps we use every day.

An open source foundation is central to addressing both defining issues. It allows researchers and developers to test the systems, building better and safer products resulting in faster innovation and creating a flourishing market. That doesn't mean that every model can or should be open sourced. There's a role for both proprietary and open source AI models. But, giving businesses and researchers access to tools that would be challenging to build themselves, backed by computing power they might not otherwise access, can create vast social and economic opportunities. In other words, open sourcing democratizes access – it gives more people and businesses the power to access and test state-of-the-art technology to identify potential vulnerabilities, which can then be mitigated in a transparent way by an open community. An open source based model on top of which an industry can be built creates a vibrant ecosystem. Rather than having dozens of companies building many different AI models, an open source model creates an industry standard, much like the model of the Internet in 1992. Through this collaborative effort, AI technology will progress faster, more reliably, and more securely. It is no coincidence that so much leading work in AI is being done here in the United States, from foundational research to real world products. We have a dynamic economy of which the tech sector is a major part. Talented people want to build new things here and that helps our global competitiveness.

While AI will drive progress everywhere, I expect there will be particular benefits that accrue to the United States over time. We want to ensure that the United States and American companies lead in AI development, ahead of our competitors and adversaries, so that the foundational models are developed here and represent and share our values. By open sourcing current AI tools, we can develop and improve the foundational models faster than others – including potential adversaries – will be able to access and build on those models. With US leadership, we can cultivate this powerful technology based on our values, rather than relinquishing it to our adversaries. Leading the AI research and development effort puts us in a strong position to enhance the safety of our systems and to warn about potential risks.

Recommendations for What's Next

As AI technology progresses, there is an urgent need for governments to work together, especially democracies, to set common AI standards and governance models, with a focus on the spaces where there are gaps within existing regulation and frameworks. This is another valuable area where we welcome work with regulators to set appropriate transparency requirements, red teaming standards, and safety mitigations – and help ensure those codes of practice, standards, and/or guardrails are consistent across the world.

The White House's voluntary commitments are a critical step in ensuring responsible guardrails are established and they create a model for other governments to follow. We joined these commitments because they represent an emerging industry-wide consensus around the things that we have been building into our products for years. We believe these commitments strike a reasonable balance of addressing today's concerns and convening industry to address the

potential risks of the future. They enable the tremendous potential for AI while focusing on the greatest risks.

Continued US leadership by Congress and the White House is important in ensuring that there is a considered, collaborative approach to the regulation of AI so that society can benefit from innovation in AI while striking the right balance with protecting rights and freedoms, preserving national security interests and mitigating risks, where those arise. US-led frameworks for approaching these issues would help drive toward a global consensus that doesn't yet exist, and provide alternatives to approaches that are designed to curtail American innovation. As with other technological shifts, the government has an important role to play. The fact that Congress is willing to engage on these issues encourages us that guardrails will be put in place that help to ensure AI is developed and utilized in a way that promotes innovation and spreads the economic and social benefits, while anticipating and mitigating potential risks.

Conclusion

I'd like to close by thanking Chairman Warner, Vice Chairman Rubio, and the other members of the Committee for your leadership. At the end of the day, our job is to work collaboratively with you, with Congress, with other nations, and with other companies in order to drive innovation and progress, in a manner that is safe and secure and consistent with our national security interests.

We appreciate your attention to these important issues and look forward to continuing to work to find ways we can continue to improve our AI tools, processes, and collaborations.

Thank you, and I look forward to your questions.