credo | ai

U.S. Senate AI Insight Forum

Transparency, Explainability, Intellectual Property & Copyright

Written Statement of Navrina Singh
Founder and CEO, Credo AI

29 November, 2023

Leader Schumer, Sen. Rounds, Sen. Heinrich, Sen. Young, and distinguished members of the U.S. Senate, I am grateful for the opportunity to speak to you today.

My name is Navrina Singh, and I am the founder and CEO of Credo AI, an artificial intelligence (AI) governance startup helping organizations to responsibly build, adopt, procure and use AI at scale.

Founded in 2020, Credo AI's pioneering software AI governance platform helps enterprises from Global 2000s to small- and medium- sized enterprises measure, monitor and manage AI risks, while ensuring compliance with emerging global regulations and standards, such as the European Union (EU) AI Act, the NIST AI Risk Management Framework (RMF), and the work of international standard setting bodies such as the ISO and IEEE. It is exciting to see how much the discussions on AI governance and guardrails have evolved over the past 18 months, and we are encouraged by the fact that our mission and work at Credo AI aligns closely with the goals of the SAFE innovation framework. Our enterprise customers provide Credo AI with a unique vantage point which we endeavor to bring to this Forum - we are exposed to a wide range of use cases in generative AI and classical machine learning (ML) across diverse organizations and industries who are adopting AI to drive innovation.

The United States has always been a world leader in driving transformational innovation to advance progress and prosperity. AI represents a historic technological shift - one with the power to further uplift our economy and society. We believe the key to realizing its full promise is leading in the pursuit of Trustworthy AI. This moment presents an opportunity for American researchers, companies, policymakers and citizens to usher in an era of Trustworthy AI by placing equal focus on capability building and guiding principles.

The U.S. is uniquely poised to blaze a trail where AI development and ethical application co-evolve from the start. Trustworthy AI is a core competitive differentiator, not just for companies, but for countries. Any government helping to set up requirements on AI transparency and explainability now will have a competitive advantage in creating and

developing accurate methods for assessment and alignment that foster a robust ecosystem of trustworthy AI.

In order to make AI transparency a reality, we encourage the U.S. Congress to 1) prioritize AI transparency and disclosures, injecting transparency, evaluations, and disclosures throughout the *entire* AI value chain; 2) invest in contextual regulations, benchmarks, and evaluations with requirements based on the specific use case; and 3) operationalize standards, AI assurance sandboxes, and dedicated AI oversight. I look forward to discussing these ideas with the distinguished members of the U.S. Senate and the esteemed participants of this Forum.

Prioritizing AI Transparency & Disclosures

Transparency is the cornerstone of trust and accountability in AI. At Credo AI, we have had the opportunity to work with a vast range of enterprises, from the largest global corporations to nimble startups. This has allowed us to develop what we call "policy intelligence" - an understanding of how to effectively translate legislative policy into operational code across critical industries and use cases to inform governance.

The AI ecosystem, intricate and multi-layered, demands transparency and disclosure at every juncture—from foundational models like GPT-4 and StableDiffusion, to the infrastructure layers and application interfaces. At Credo AI, we recognize that generative AI, while a beacon of innovation, also brings with it an array of governance challenges. Each layer of the “genAI stack” presents unique risks and governance opportunities. The foundation models are just the beginning; the infrastructure on which these models are trained and the applications that interface with end users are equally critical. It is across this spectrum that we must inject robust governance to maintain control and ensure safety.

To achieve “transparency in AI,” it is necessary to inject disclosures throughout the entire AI value chain. Disclosures can take the form of several different types of “governance artifacts,” including comprehensive AI use case risk reporting, model cards, data set cards, and algorithmic impact assessments (AIAs). Imagine a world where every AI system is accompanied by a report card—a dossier that includes its performance, robustness, and fairness metrics, along with an evaluation of the risks it may pose.

We can achieve this. Before deployment, AIAs can serve as a litmus test, providing metrics on the system's performance, robustness, and fairness, as well as explainability documentation and a thorough assessment of risks (both intended and unintended). Such assessments are the foundations of trust in other domains, safeguarding our environment, privacy, and human rights, all while nurturing innovation and ensuring safety. AIAs are indispensable

transparency tools for entities throughout the entire AI value chain, from foundational model providers to the developers who build upon them, to the end users using AI in their daily lives.

As the U.S. Congress considers legislation related to privacy and artificial intelligence, it is valuable to consider **requirements for private companies to provide both governance artifacts that include risk- and impact-informed evaluations of their AI system, and transparency into the governance actions organizations are taking pre- and post-deployment.** We also believe it is essential to mandate Responsible AI disclosure reporting that should encompass investment (percentage) in AI safety and governance, risk- and impact-informed evaluations of their AI systems, and elucidate the steps enterprises are taking to govern AI pre- and post-deployment.

Such public disclosures become even more potent when aligned with industry-specific benchmarks, allowing downstream procurers to gauge the safety and efficacy of AI vendors in the marketplace effectively. Fundamentally, transparency reporting and disclosures are the building blocks of trust in the AI Ecosystem as a whole. **By standardizing disclosures, we can expedite the application of AI across various industries.**

One mechanism to deliver this transparency throughout the AI value chain is the work that Credo AI has recently pioneered with AI Trust Reports, a transparency tool that details the use case risks, governance controls, and compliance protocols. These reports are critical for generative AI startups to demonstrate their commitment to responsible AI practices, enabling them to innovate rapidly, enhance their marketability, and build enduring trust with customers.

We must foster an environment where transparency is not an afterthought, but a foundational aspect of AI development. Such a framework will not only engender trust, but also catalyze innovation in a manner that is responsible, equitable, and aligned with our collective values.

Investing in Contextual Regulations, Benchmarks, & Evaluations

Investing in use case specific regulations and technology-informed standard evaluations and benchmarks is pivotal. Last year I testified on these topics before the House Committee On Science, Space and Technology Subcommittee on Research and Technology for a hearing entitled “Trustworthy AI: Managing the Risks of Artificial Intelligence.” Achieving trustworthy AI depends on a shared understanding that AI is industry specific, application specific, data specific and context driven. There is no one-size-fits-all approach to “what good looks like” for most AI use cases. For example: there is no single definition of algorithmic “fairness,” because the concept of fairness is incredibly context-dependent. Similarly, when considering what metric or measures to use for the performance of an AI system, assessors should be able to select from a wide variety of different metrics that take into account use case context, model

type, and data type. The organization building the AI system should be consulted about "acceptable" performance metrics. This requires a collaborative approach to assessments, and we advocate for context-based tests for AI systems with reporting requirements that are: specific, regular, and transparent.

Then and now, we continue to advocate for rigorous evaluations of capabilities and risks at both the development and utilization stages of AI systems. **Responsibility permeates the entire value chain, and while liability may be more narrowly defined, the capacity for ensuring safety is a shared duty.**

In concert with these methods for accountability and improved evaluations, the development of a comprehensive taxonomy of risk is essential. This taxonomy will offer consistent and clear guidance for AI developers and deployers to assess the potential impacts of their AI systems effectively. We at Credo AI urge the U.S. Government to prioritize resources for the U.S. Department of Commerce to support the expansion of NTIA and NIST's ongoing work to develop such a taxonomy. This will not only inform risk-based transparency disclosures, but also elevate the level of oversight and understanding of AI systems.

We have seen firsthand how comprehensive and accurate assessments of the AI applications and the associated models and datasets, coupled with transparency and disclosure reporting, encourage responsible practices to be cultivated, engineered, and managed throughout the AI development life cycle.

Operationalizing Standards, Sandboxes & Oversight

We also urge policymakers and standard-setting bodies to **prioritize establishing context-focused standards and benchmarks—that are globally interoperable—that can help take some of the guesswork out of compliance with AI regulations.** Standardizing evaluations - built on industry-driven and consensus-led global standards from international standard-setting organizations like ISO, IEEE, and NIST, will be key to building a robust evaluation and assurance ecosystem for both traditional machine learning and generative AI evaluations. Investing in improved evaluations on a host of dimensions (like propensity for misinformation) can help control and align AI systems, providing the infrastructure for a *proactive* accountability system.

It is crucial that these standards consider the diverse capabilities and resources of organizations of different sizes, recognizing that smaller enterprises may face distinct challenges in adopting new standards compared to their larger counterparts.

In addition to the advancement of AI assurance standards, Credo AI advocates for the establishment of specialized environments known as AI assurance sandboxes. These would be **designated spaces where companies can rigorously test and refine transparency and**

mitigation requirements for their AI systems within a secure and controlled setting. The purpose of these sandboxes would be to significantly enhance the quality of transparency and explainability methods for AI technologies. By providing a space for iterative testing, we can build a solid foundation of public trust and confidence in these systems.

These sandboxes also play a crucial role in regulatory compliance, offering companies a "grace period" in order to align with new regulations without the immediate risk of penalties. Within these sandboxes, organizations can receive direct feedback from regulatory bodies on areas that require further compliance efforts. These sandboxes should be considered as a "safe harbor," where companies can proactively work on transparency and risk mitigation strategies based on voluntary guidelines, without the pressure of financial repercussions for initial (and inevitable) missteps in a nascent area of research. This approach encourages sincere and earnest efforts by companies to meet regulatory standards before they are fully enforced.

Lastly, we agree that the United States would greatly benefit from "making every agency an AI agency" by equipping existing agencies with the expertise, staffing, funding, and other relevant resourcing needed to provide dedicated oversight of AI systems. However, in our view, the United States also needs proactive guardrails that decentralized agencies may not be able to provide, especially for the accelerated innovations in Generative AI. Credo AI sees value in allocating funding and resourcing for an independent federal agency or Cabinet-level position that would oversee R&D investments, coordinate a national AI strategy for the United States, and have the authority to regulate AI including Foundation Models and Generative AI. This consolidation and enforcement will be net positives for the development of safety standards and mandatory requirements for industry.

Conclusion

With the momentum sparked by recent global advancements in AI guardrails like the White House Executive Order, White House AI Commitments, G7 AI Code of Conduct, Bletchley Declaration, and the expected EU AI Act, the urgency and significance of translating AI governance into tangible actions with the AI Insight Forums could not be clearer. The United States stands at the cusp of leading the global AI revolution. We have an opportunity to welcome a new era of Responsible AI guided not by fears of the unknown, but by our highest shared hopes for humanity's progress. Our global leadership depends on it at this defining moment for the 21st century and beyond.

Credo AI is ready and willing to proactively collaborate with the United States Congress in this endeavor. We thank you for this initiative.