PREPARED TESTIMONY OF NAVRINA SINGH, FOUNDER AND CEO, CREDO AI

BEFORE THE HOUSE COMMITTEE ON SCIENCE, SPACE AND TECHNOLOGY SUBCOMMITTEE ON RESEARCH AND TECHNOLOGY

HEARING DATE/TIME: SEPTEMBER 29, 2022 10:30 A.M. EST

HEARING TITLE: TRUSTWORTHY AI: MANAGING THE RISKS OF ARTIFICIAL INTELLIGENCE

Introduction

Madam Chair, Ranking Member Feenstra, and Members of the Subcommittee on Research and Technology, thank you for the opportunity to testify today and to be a part of this distinguished panel of witnesses. My name is Navrina Singh, and I am the Founder and Chief Executive Officer of Credo AI.

A "credo" is a statement or system of beliefs or principles to guide actions. I founded Credo AI in March 2020 with one key goal in mind: to enable organizations to deliver Responsible AI (RAI) at scale. RAI includes assessment, reporting, and governance to the highest of ethical standards in order to ensure a fair, transparent, compliant, and auditable environment for the development and use of artificial intelligence (AI).

Credo AI is a software company, and our core product is the Credo AI Responsible AI Governance Platform™. Our platform is designed to help organizations consistently translate principles into actionable metrics, assessments and benchmarks throughout the entire AI lifecycle. We recognize that enterprises are at different levels of maturity in their development and use of AI. Our mission at Credo AI is to realize comprehensive Responsible AI governance by providing software tools to enterprises wherever they are in their AI Governance journey.

What Is Responsible AI?

At Credo AI, we define the phrase "Responsible AI" as AI that is human centered. That means that AI systems need to be performant, fair, transparent, safe and secure, privacy-preserving, and auditable. These tenets are aligned with the ways that many other organizations, regulatory bodies, and standard-setting bodies define the phrase "Responsible AI." Our customers aren't only concerned with making sure their systems are accurate or performant; they want to know if their systems are fair, transparent, and robust, and they want to know how regulators are defining these parameters. Measuring and managing each of these tenets is very complex and context-dependent—each must be aligned, based on the context of their use, with the values both of society and the organization developing or using the AI.

Credo AI's RAI Governance Platform is built to help organizations map, measure, manage, and mitigate AI risk and compliance for all their AI use cases. The Platform provides organizations with context-driven requirements for their AI use cases in the form of "Policy Packs." Policy Packs provide specific technical and process requirements that an AI system must meet, based on regulations, laws, standards, frameworks, guidelines, an organization's internal guardrails, and industry best practices. Our platform connects to our open source Responsible AI assessment framework, Credo AI Lens, which can be used to assess machine learning (ML) models and datasets based on the requirements coming from Policy Packs, allowing our customers to programmatically generate governance artifacts like model cards, assessments reports,

transparency reports or audit reports. The platform standardizes governance activities, promotes multidisciplinary collaboration among technical and business stakeholders, and reduces the burden of governance on technical teams, making it easier for organizations to govern their AI systems more effectively and gain confidence in their AI use.

How to Create an Environment that Fosters RAI

AI is a transformative technology that is rapidly evolving. There is a significant opportunity to encourage the development of trustworthy technology and set effective policy, and as the Subcommittee studies this important issue, Credo AI respectfully offers the following key points for your consideration:

- **RAI Requires a Full Lifecycle Approach**: At Credo AI, we believe managing risk is only one part of delivering on the promise of Responsible AI—it demands a full lifecycle approach. AI systems cannot be considered "responsible" based on one point-in-time snapshot, but instead must be continuously evaluated for responsibility, and transparently reported on throughout the entire AI lifecycle, from the design, to development, to testing and validation, to production and use.

- **There Is No One-Size-Fits-All Approach to AI Governance**: We believe that 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.

- **Transparency Reporting and System Assessments Can Deliver Trustworthy and Accurate AI:** The importance of transparency reporting and system assessments cannot be overstated as a critical foundation for RAI governance for all organizations. Reporting allows policymakers to start to evaluate different approaches, and potentially opens the door for benchmarking—reporting is the step that gets us to standards that can be enforced. We have seen firsthand how comprehensive and accurate assessments of the AI applications and the associated models/datasets, coupled with transparency and disclosure reporting, encourage responsible practices to be cultivated, engineered, and managed throughout the AI development life cycle. Fundamental to this is access to compliant and comprehensive data for assessments.

Companies Are Seeking Guidance

In our experience, organizations understand that Responsible AI is a competitive advantage for them in this age of AI. Organizations know there is a need for RAI governance, and welcome a collaborative approach to developing it. The notion that AI regulation will cause U.S. companies to offshore or cause AI to stagnate is a *false* premise. Based on our experience in the field working with companies that develop and deploy AI, we repeatedly hear a desire to have those systems work well in a compliant, safe, fair and auditable fashion. This leads to an important synergy: the more that policymakers can do to help companies understand how to develop trustworthy systems, the easier it will be for those companies to maximize the value of those systems. Thoughtful policy making and governance via public-private partnership can create conditions for innovations in AI for these companies.

Key Challenges to Overcome in the Development and Use of Responsible AI

While there is reason for optimism, there is much work to be done. Credo AI has experience working with customers across industries, and we have observed that they are all working to set up processes to foster RAI. The key challenges that we have observed and that we hope policymakers will consider when it comes to more effectively promoting the responsible development and use of AI include:

- **Standards and benchmarks for RAI are still emerging.** We 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. While many emerging regulations set "fairness," "transparency," and other RAI dimensions as key requirements for compliance, there are not yet clear standards or benchmarks for what it means for an AI system to be "fair" or "transparent." That is because there are many ways to define these terms. Without clear standards and benchmarks, organizations are left having to develop and justify their own measures for different technical dimensions of their AI systems. Standards and benchmarks should also try to account for the challenges of operationalizing such requirements and frameworks depending on the size and reach of the organization. Expecting a small or mid-sized business to operationalize new standards as quickly as major multinational companies would present its own challenges.
- **AI regulations must include reporting requirements to foster transparency and drive towards standards.** We urge policy makers to establish requirements that mandate disclosures and transparency reporting around the procurement, development, and use of AI. Because of the lack of standards today, many organizations are reluctant to share results about the behavior of their AI systems externally—because they have no idea how their results might compare with those of their competitors, or whether they are "good" or "bad" for external stakeholders. We are strong supporters of reporting requirements,

therefore, that promote and incentivize public disclosure of AI system behavior and operation as a key driver of the establishment of standards and benchmarks.

Context is Critical: Metrics for Each Tenant of RAI Vary

We strongly believe that AI is industry-specific, application-specific, and context-driven and needs to continuously assessed —factors that should be reflected in its governance.

For example, when considering what definition to use for fairness, we feel that there is no one-size-fits-all answer or approach. There is not a single definition of algorithmic fairness accepted across industry sectors and use cases.

Algorithmic fairness is a field of research aimed at understanding and correcting the ways that historical societal biases show up in AI systems. An AI system can be considered to be "fair," in the sense of algorithmic fairness, if it does not perpetuate or amplify harmful societal biases in its operation.

When data scientists are evaluating whether their AI systems are fair, they look at specific technical measures of *bias* in their AI systems—to understand if these systems are perpetuating harmful societal biases. There are two primary ways that we measure bias in our AI systems: evaluating **parity of performance** and **parity of outcomes**.

- Parity of performance is about evaluating whether your ML model performs equally *well* for all different groups that interact with it. For example, does your facial recognition system detect Black women's faces at the same or similar accuracy rate that it detects white men's faces?
- Parity of outcomes is about evaluating whether your ML model confers a benefit to different groups at the same rate. For example, does your candidate ranking system recommend Black women get hired at the same or similar rate as it recommends white men?

We do not have a singular definition of fairness — nor should anyone who is thinking about algorithmic fairness — because fairness is incredibly context-dependent.

Here's an example to illustrate why you cannot have a "one size fits all" definition of algorithmic fairness. Let's say that you have an AI system that is going to be predicting whether someone should be given a loan (a credit risk prediction system), and you have another AI system that is going to predict whether somebody has cancer by analyzing a CT scan for tumors. For your credit risk prediction system, the system is considered "unfair" if it predicts that Black women are credit-worthy (and therefore should be given a loan) at a much lower rate than white men; we want to make sure that our credit prediction system is conferring the benefit of getting a loan

relatively equally across groups, regardless of gender or race. This is an example of parity of outcomes. For the cancer detection system, however, the parity of outcomes isn't the primary concern; we don't care if the system is predicting that women have breast cancer at a rate that is significantly higher than men. This is because for this cancer detection system to be considered "fair," we want to make sure that it is *equally accurate* for all groups that interact with it. The issue here is parity of performance: our cancer detection system will be considered fair if it has the same performance rate across all groups.

The metrics that you use to measure parity of performance are different from the metrics that you use to measure parity of outcomes—and even within these two categories, there are many different metrics that you can pick, depending on what is most important based on the use case context.

Similarly, when considering the question of what metric or measure to use for algorithmic performance: there is no single metric for performance. Depending on your use case context, model type, and data type, you may select from a wide variety of different metrics that are all reasonable and accepted ways to evaluate performance of an AI/ML system.

For a cancer detection system, assessors might care more about a system that has relatively equal *false negative rates* across groups, because incorrectly diagnosing someone as healthy who actually has cancer is a life-threatening mistake (the cost of making an incorrect "negative" prediction is very high). For a facial recognition system that is going to be used to grant access to a device—say, your phone—assessors may care more about *false positive rate*, however, because they want to ensure that this system doesn't accidentally grant access to your phone to someone who should not have access (the cost of making an incorrect "positive" prediction is very high).

These examples are all intended to show that there is no one definition for fairness when it comes to AI systems, and context is a key factor in determining what is fair. At Credo AI, we provide tools to our customers to help them determine how fair their AI system is by working with our customers to align on the exact metrics that should be used to assess fairness **based on their use case context**. This work is informed by the industry best practices that the customer's use case is aligned with. Our policy team also focuses on bringing in requirements from regulations, laws, standards, guidelines, and frameworks—and our data science team partners with our customers to understand exactly what their ML models are designed to do, and how they do it; we then create a technical assessment plan designed to evaluate the exact dimensions of the system that are most relevant for understanding whether it is fair in the context it will be deployed.

Given the context-driven nature of AI governance, we advise policymakers to develop context-specific guidance and rules, and transparency reporting will help industry to arrive at the

right standards and rules based on this context - through an iterative process of revealing benchmarks and best practices.

Addressing Risk Now Ensures Leadership in the Long Run

AI is a multi-trillion dollar industry. For the United States to lead in its development, it is crucial to understand the economic and societal outcomes for our nation. RAI is about better, more effective outcomes—it is about producing *more* value for AI builders and consumers by building trust in technology.

RAI is a core competitive differentiator, not just for companies, but for countries. Any government helping to set up RAI requirements on testing and metrics now will have a competitive advantage in first creating and developing accurate methods for assessment and alignment to create trustworthy AI. The work to build trustworthy AI is not *just* about "doing the right thing" and setting "values" that make people feel good. It is about building systems that work better - systems that do not have unintended harmful consequences.

The MIT Sloan Management Review and Boston Consulting Group Report[1] published this month (September 2022) reported that, "RAI Leaders can realize measurable business benefits from their RAI efforts…[which] include better products and services, improved brand differentiation, accelerated innovation, enhanced recruiting and retention, increased customer loyalty, and improved long-term profitability, as well as a better sense of preparedness for emerging regulations. RAI leaders are nearly three times as likely to realize business benefits from their organizations' RAI initiatives than non-RAI leaders." This is just one illustration of how investing in trustworthy AI pays off.

When we consider what the effect of algorithmic bias can be on economic contributions to society, we should look at real-world examples, such as an AI system that was deployed in the market and automatically granted lower lines of credit to women than to men. The biased AI system allocated differential lines of credit to a husband and wife with the same address and joint home income, with the woman being granted a much lower line of credit by the AI system than the man. In this scenario, there is no reason a wife should have less credit than her husband, and the system's decreased accuracy resulted in a loss of economic contributions that women bring to society - an outcome that is unfairly impacts the individuals and is also bad for the business

Another example of algorithmic bias illustrates a loss to the workforce—an AI system that was trained mainly on men's resumes deprioritized the word "Women's" when it appeared in resume

[1] Elizabeth M. Renieris, David Kiron, and Steven Mills, "To Be a Responsible AI Leader, Focus on Being Responsible," MIT Sloan Management Review and Boston Consulting Group, September 2022.

search results. As a result, the AI system missed out on a stellar talent. This example is not just about hurting people, but about creating a system with failures that will have a negative economic impact on the workforce at large.

If we truly want to be a global leader in AI, then our focus should not be on building the most powerful system the fastest, but rather on building responsible technology and support systems that will serve us best in the long run. We will sacrifice the opportunity to lead if we are simply moving quickly for the sake of getting ahead in a way that is not aligned with our societal values.

Conclusion

Credo AI is grateful for the opportunity to appear at today's hearing, and we applaud the Subcommittee's focus on how best to empower organizations to create AI with the highest ethical standards in order to deliver Responsible AI at scale.