Confronting Bias:
BSA's Framework to Build Trust in AI

Tremendous advances in artificial intelligence (AI) are quickly transforming expectations about how the technology may reshape the world and prompting important conversations about equity. While AI can be a force for good, there is a growing recognition that it can also perpetuate (or even exacerbate) existing social biases in ways that may systematically disadvantage members of historically marginalized communities. As AI is integrated into business processes that can have enormous impacts on people's lives, there is a critical need to ensure that organizations are designing and deploying these systems in ways that account for the potential risks of unintended bias.

The Framework is a tool for ensuring that AI is accountable by design and can be used by organizations of all types to manage the risk of bias throughout a system's lifecycle. Built on a vast body of research and informed by the experience of leading AI developers, the Framework:

> Outlines a process for performing impact assessments to identify and mitigate potential risks of bias

> Identifies existing best practices, technical tools, and resources for mitigating specific AI bias risks that can emerge throughout an AI system's lifecycle

> Sets out key corporate governance structures, processes, and safeguards that are needed to implement and support an effective AI risk management program

The Framework is a playbook organizations can use to enhance trust in their AI systems through risk management processes that promote fairness, transparency, and accountability. It can be leveraged by organizations that develop AI systems and companies that acquire and deploy such systems as the basis for:

- **Internal Process Guidance.** The Framework can be used as a tool for organizing and establishing roles, responsibilities, and expectations for internal risk management processes.

- **Training, Awareness, and Education.** The Framework can be used to build internal training and education programs for employees involved in developing and using AI systems, and for educating executives about the organization's approach to managing AI bias risks.

- **Supply Chain Assurance and Accountability.** AI developers and organizations that deploy AI systems can use the Framework as a basis for communicating and coordinating about their respective roles and responsibilities for managing AI risks throughout a system's lifecycle.

- **Trust and Confidence.** The Framework can help organizations communicate information about a product's features and its approach to mitigating AI bias risks to a public audience. In that sense, the Framework can help organizations communicate to the public about their commitment to building ethical AI systems.

- **Incident Response.** Following an unexpected incident, the processes and documentation set forth in the Framework can serve as an audit trail that can help organizations quickly diagnose and remediate potential problems.