Purpose This policy provides guidance on developing, deploying and using generative AI responsibly and aligning practices with the NIST AI Risk Management Framework (RMF). It aims to maximize benefits and minimize potential negative impacts to individuals, groups, organizations and society.

Scope This policy applies to all employees, contractors, systems and processes involved in the design, development, deployment or use of generative AI systems, including but not limited to, text, image, video and audio generation.

Policy Statements

Governance

1. A generative AI oversight team will be created, comprising diverse disciplines like engineering, human factors, audit, legal, and ethics.
2. Generative AI projects will follow documented processes for risk-based decisions on development, deployment and monitoring.
3. Staff will receive RMF and generative AI risk training on topics like safety, fairness, accountability, and regulatory compliance.
4. Reviews by the oversight team at major milestones will be required before deploying generative models internally or externally.

Mapping Risks 5. Intended use cases, target users, deployment contexts, and potential benefits and harms will be defined early and re-evaluated regularly.

6. Risks assessments will analyze and document safety, ethical, legal, reputational and technical risks for each intended use case.
7. Risk prioritization decisions will balance beneficial use cases and public access with responsible controls.

Measurement & Testing 8. Generative models will undergo rigorous testing to measure risks and evaluate trustworthiness characteristics before deployment.

9. Testing sets will cover a broad, representative set of use cases, be routinely updated, and model performance tracked over time.
10. Processes to continually monitor risks after deployment will be established, along with controls to address emerging issues.

Transparency & Accountability 11. Model details like data sources, training methodology and model versions will be documented to enable accountability if issues emerge.

12. Attribution indicating content is AI-generated will be clearly displayed for external uses.
13. Controls like human-in-the-loop oversight will be required where risks of harmful, biased or misleading outputs are higher.
14. Feedback channels will be created to allow reporting issues by users and affected groups, to improve models over time.
15. An ethics review board will evaluate high-risk use cases not covered by this policy, before approval.

Administration The Generative AI Oversight Team is responsible for administering this policy, establishing necessary procedures, guidelines and updates to align with regulations.