Checklist for any Company using Generative AI

1. Red Team Testing (adversarial attack simulations)

- □ Have you been tested against adversarial attacks?
  - □ IP Exfiltration
  - □ Phishing
  - □ CAI Extraction (commercially available information)
  - □ Data poisoning
  - □ Training data exfiltration
  - □ PII Leakage
- □ Can you report the following results of red-team testing?
  - □ What software vulnerabilities and associated exploits have you discovered?
  - □ What measures have you taken to meet safety objectives
  - □ If you sell to critical infrastructure, what are your risk assessment results?

2. National Security Risk Mitigation

- □ Can your service generate outputs that may represent the following threats?
  - □ CBRN (Chemical, Biological, Radioactive, Nuclear)
  - □ Risks to critical infrastructure
  - □ Risks to Energy-Security
- □ What steps have you taken to mitigate CBRN threats?
  - □ Evaluation by DOE, private AI labs, Academia, or Third-party model evaluators
  - □ How have you lowered the barrier to entry for CBRN threats from non-state actors?

3. Employee Data Collection & Usage

- □ Do you collect data about employees or the work they do?
- □ Do you use that data to monitor or augment their work with AI?
- □ How do you ensure transparency of that usage under worker-protection law?

4. American Consumer Protections

- □ How do you protect against: fraud, discrimination, risks to financial stability?
- □ How do you ensure privacy for consumers using your product?
- □ What due diligence and monitoring do you conduct for third-party AI services?

Disclaimer: This document is intended solely as an informational guide and should not for any purpose be constituted as legal advice. We advise all parties to refer to the executive order published on www.whitehouse.gov as the sole source of truth. Further this document does not make any claims as to regulations or mandates by the US government and should be read as a subjective interpretation. For any and all inquiries please contact info@promptarmor.com

PromptArmor