OpenAI for Countries: Our Approach to Security

OpenAI's mission is to ensure that artificial general intelligence (AGI) benefits all of humanity. To advance this mission on the global stage through OpenAI for Countries, we are guided by three overarching principles:

(1) Extend and maintain our technology leadership
(2) Build a broad international coalition to promote U.S.-built democratic AI
(3) Maintain the world's most advanced AI infrastructure in America

OpenAI and the United States have a unique opportunity to lead the world in shaping one of the most consequential technologies of our time. As other nations look to us for guidance and partnership as the leaders on this technology, we can set the global standard for AI infrastructure rooted in democratic values, transparency, and security. This is a moment when we can support countries that would prefer to build on democratic AI rails, and provide a clear alternative to authoritarian versions of AI that would deploy it to consolidate power.

While OpenAI's mission is global, we believe working closely with the U.S. government best advances democratic AI. As we expand our international partnerships, we are further committed to ensuring that strong security standards and strong partner ecosystems remain at the core of how advanced AI is built and deployed. That's why we are proactively engaging with U.S. government entities—including those overseeing export controls — to ensure that our international partnerships meet the highest standards of security and compliance, and why our OpenAI for Countries initiative includes commitments from partner nations to invest in expanding our Stargate project here in the U.S.

We understand the unique security considerations these partnerships involve and the measures for long-term success as described here. Front of mind for us are:

- **Multilayered Model Security:** Safeguarding our models is a continuous commitment and a core pillar of our security posture. Every OpenAI model deployment is governed by a rigorous, continuously evolving security framework that spans information security, governance, and physical infrastructure. Our security measures are not static; they scale with the capabilities of our models and incorporate state-of-the-art protections, including hardware-backed security, zero-trust architecture, and strong cryptographic safeguards. We will continue to invest significantly in defense-in-depth measures that address physical security, insider threats, supply chain, and advanced cyber risks.

- **Strict Personnel Oversight:** OpenAI will maintain explicit and continuous oversight over all personnel with access to our information systems, intellectual property, and models. No individual or entity will gain such access without our direct approval. We invest deeply in technologies, monitoring, auditing, and governance processes to ensure that access is tightly controlled and rigorously enforced. This includes new and novel uses of our models and our technologies to provide critical oversight.

- **Pre-Deployment Reviews:** We recognize that some advanced models may present risks under our Preparedness Framework that are incompatible with certain

deployment environments. Each deployment of new models will undergo risk evaluation prior to deployment.

We are committed to an approach that is thoughtful and deliberate, implements strong safeguards, and is rooted in collaboration with the U.S. government. We believe that we have an opportunity to help shape a global AI ecosystem that serves the public good, strengthens alliances, and protects the technologies that underpin future prosperity.