OpenAI

OpenAI Principles for National Security Partnerships

OpenAI's mission is to ensure that AGI benefits all of humanity. As AI systems become more capable and eventually self-improving, their use by governments will become more consequential. These systems can help governments protect people, deliver services, enforce law, and defend societies. But they can also concentrate state power, reduce human judgment in consequential decisions, and amplify the harm caused by government mistakes. The core question, then, is who gets to decide how governments use AI, for what purposes, and subject to what limits and accountability. That question should be answered through accountable democratic institutions, not by regimes unaccountable to their populations or by a small group of private actors.

Democratic accountability requires democratic capacity. If government use of AI is to be authorized and governed by law, constrained by protections for human rights and civil liberties, and subject to democratic oversight, democratic governments need access to, and understanding of, powerful AI. They need that capacity both to govern AI responsibly and to defend against threats to national security, critical infrastructure, and public safety, including threats intensified by increasingly powerful AI. Democracies should not be left without these tools while adversaries use them without similar commitments to rights, accountability, or rule of law.

But putting powerful AI in the hands of any government risks concentrating power, threatening the checks and balances at the heart of democratic accountability. Our partnerships with governments should therefore help democracies protect their populations while also strengthening the oversight and institutional constraints that keep power accountable. OpenAI is not elected and is not a government oversight body. But we can help democratic oversight institutions build the technical understanding needed to scrutinize how public agencies use AI, especially in law enforcement and national security, and identify where checks and balances are not keeping pace with the technology.

The principles below provide a framework for ensuring national security and law enforcement uses of AI can help democracies defend their populations and institutions while strengthening democratic governance. They reflect an evolution in our approach to national security and law enforcement use of AI driven by changes in AI capabilities, safeguards, and the global threat environment.

In sum, the principles emphasize OpenAI's commitment to government uses of AI that:

- benefit people
- strengthen democratic governance and avoid concentration of power
- preserve appropriate human judgment and accountability in high-consequence national security and law enforcement settings

Based on these principles, we will not support use of OpenAI tools for:

- mass domestic surveillance
- high-stakes decisions — including decisions over the use of force — without appropriate human judgment and accountability
- uses that evade legal obligations, oversight, or accountability.

We will have a set process to apply this framework to national security and law enforcement opportunities. Over time, these recommendations will establish precedent. Like a common law system, the principles will become clearer through documented applications, creating a body of "case law" over time.

Our Principles

1. We support government uses of AI that benefit people.

If increasingly capable AI will shape public power, we should support uses that direct it toward public benefit, not merely state power. Governments are uniquely responsible for the common good and can use AI to that end: for example - to improve public safety, strengthen public health, or deliver essential services, and defend democratic societies.

In law enforcement and national security specifically, government use of AI can help protect people from coercion, exploitation, and other serious threats to safety and liberty. Keeping people safe is in the public interest, and government use of AI can serve that mission. We acknowledge that some protective uses involve hard tradeoffs, including the use of force and intrusions on privacy. As long as governments are accountable to their publics and are acting within democratic rule-of-law frameworks, those decisions belong to them.

2. We prioritize partners who share our values.

Increasingly capable AI may shift power within societies and among them. Because government use of AI can expand state power over citizens, we prioritize partners committed to using our technology lawfully, accountably, and for the benefit of their populations.

In this context, the United States has a distinctive role. It is our home jurisdiction, and its laws, institutions, and security architecture are the most immediate framework within which we operate, are held accountable, and pursue our mission. That gives us both a practical and civic reason to engage constructively with U.S. institutions.

The United States and many allies have been the leading democratic counterweight to authoritarian power and can continue to play that role in the global AI landscape. Our duty is to all of humanity, which means our interests are broader than those of any one government. But if the U.S. and its allies are denied access to advanced AI while authoritarian states develop or obtain it, the global balance of power could shift against democratic governance, human rights, and the rule of law. We therefore expect to work with the United States and its allies on national security uses of AI, including when that work supports the cause of democratic accountability, and satisfies our principled limits and safeguards.

We also expect to prioritize work with key U.S. allies and partners whose access to AI can help preserve a global balance favorable to democratic governance. But alliance or partner status is not enough on its own. Some important U.S. allies or partners may not fit neatly within a democracy-based framework. In those cases, strategic alignment may matter, but it should not decide the question by itself. We will also be guided by democratic governance, protection of individual rights, respect for human rights, meaningful oversight, the strength of applicable legal and accountability frameworks, and the specific risks of the proposed deployment.

Eligibility is only a threshold for engagement. We will not support uses that conflict with these principles, regardless of the partner. Sustained or systemic misalignment with a partner may lead us to narrow, suspend, or end a partnership, or to reassess eligibility.

3. We have an obligation to ensure AI strengthens democratic institutions and avoids concentration of power.

AI systems can help governments serve the public and protect democratic societies, but they can also amplify state power and create risks that existing legal, compliance, and oversight frameworks may not fully anticipate. Democratic institutions need to understand AI's promise and limits so they can govern, use, and oversee it responsibly.

OpenAI will help build that understanding through contractual negotiations, ongoing operational work, knowledge sharing, and public policy debates. Engagement with select government partners can serve the public good when it helps agencies, policymakers, and oversight institutions understand AI capabilities, limitations, and risks. It should also help identify uses for which AI is not yet ready, reliable enough, or consistent with critical human accountability.

That engagement should include the democratic institutions responsible for overseeing agencies that use our models, including in law enforcement and national security. OpenAI is not the right institution to oversee government use of AI. But we can help regulators, auditors, courts, lawmakers, and oversight bodies build the understanding and practical capacity to oversee AI use at speed and scale, including through the use of AI itself. And we can help the public, who are responsible for electing officials, understand the benefits and limitations of AI.

Where legal, compliance, or oversight frameworks are underdeveloped, we should help identify the gaps and advocate for improvements. In some cases, those frameworks may need to change before deployment is appropriate. In others, carefully structured engagement can help agencies identify risks, develop better guidance, and build capacity for responsible use, including in highly sensitive environments where our visibility may appropriately be limited.

4. In national security and law enforcement, AI should support human judgment and maintain human accountability.

In national security, law enforcement, and other high-consequence government settings, AI can help human decision-makers act with better information and precision. But AI must be reliable enough for the purpose, governable in practice, and used in ways that ensure human judgment over critical decisions and maintain human legal accountability and moral responsibility for the outcome. This is especially critical for the use of force, deprivation of liberty, or other actions with high and immediate consequences.

Human judgment over critical decisions must be meaningful in practice, not merely formal. As AI systems become more capable and more deeply embedded in operational workflows, human reviewers may defer to system outputs, lack the bandwidth to assess them, or be unable to understand and intervene in system behavior. When those risks are significant, we should help government partners understand them and design tools that make review, intervention, escalation, and deactivation practical and feasible. We will also therefore consider not only whether a human is nominally in the loop, but whether the system enables real and appropriate human judgment during the act of decision-making.

OpenAI

National Security and Law Enforcement Uses OpenAI Will Not Support

Based on these principles and our current understanding of the technology's capabilities, risks, and safeguards, we do not support the use of OpenAI technology for the following:

- Mass domestic surveillance. States generally recognize a need to collect information for foreign intelligence and law enforcement purposes under appropriate limits, but mass domestic surveillance can undermine the democratic institutions, public trust, and broader relationship between a government and the people under its jurisdiction on which democratic accountability depends. We do not support domestic use of AI for unauthorized or unconstrained collection, analysis, or monitoring; to infer sensitive traits about identifiable people and use those traits to disadvantage them; to retaliate against people for the lawful exercise of rights; or to fabricate, manipulate, or falsely substantiate evidence. By domestic mass surveillance, we mean mass surveillance (as defined above) by a government of its own population or other people within its jurisdiction.
- High-stakes decisions without appropriate human judgment and accountability. We do not support AI systems that make or automatically trigger decisions with direct consequences for an identifiable person's rights, liberty, safety, legal status, or denial of access to essential services. Humans must exercise appropriate judgment over such decisions and be accountable for the outcome. This prohibits, for example, automated determinations that a person meets the legal standard for surveillance, arrest, detention, or denial of essential services.
- Use of force without appropriate human judgment and accountability. One critical category of high-stakes decisions are those that involve the use of force. AI may help human decision-makers act with greater precision, reduce error, or identify relevant patterns more quickly, thereby helping democracies defend their populations more effectively. But we do not support systems that remove or limit users' ability to exercise appropriate human judgment over the use of force or that undermine their legal accountability or moral responsibility for the outcome, including systems that autonomously identify, select, and engage targets without appropriate levels of human authorization and human-defined operational parameters.

  Appropriate human judgment does not require a human decision on every discrete system action. It does require that humans make informed decisions about the conditions for deployment, including the type of target the system may to engage, limits on duration and geographical scope, and the constraints under which the system may be used. Targeting-related uses are therefore not categorically disallowed, but because targeting is one important subset of force-related decision-making, systems that automate substantial parts of the targeting process or perform critical targeting functions that bear directly on decisions about the use of force require heightened scrutiny. Review of systems intended for targeting should consider their role in the targeting process, proximity to the use of force, and whether humans can effectively manage their operation.

- Evading legal obligations, oversight, or accountability. We will not support uses of OpenAI technology to facilitate genocide, crimes against humanity, war crimes, or other gross violations of human rights or international humanitarian law. We also will not support using AI to violate the laws that govern public authority, or to circumvent the elections, oversight and

> audit requirements, approval requirements, or lines of responsibility intended to ensure that authority is exercised lawfully and accountably. That includes using AI to evade legal constraints, required oversight, or accountability.

These principles do not categorically exclude the use of OpenAI technology to conduct intelligence operations, investigations, or offensive and defensive military operations. We do not think the right line is a categorical distinction between offense and defense since, in practice, a capability or operation may be characterized as strategically or tactically offensive or defensive depending on its purpose, context, and intended effects. The relevant question, therefore, is whether those uses are consistent with the principles and remain within the limits outlined above.

Implementation Commitments

Implementing principles and restrictions require layered legal, policy, operational, and technical safeguards, such as contractual usage restrictions, model behavior specifications, and customer engagement and oversight. Sensitive partnerships may require multiple approaches depending on the purpose and context of the use, the risks involved, and the feasibility of different safeguards.

Oversight must be part of the partnership and system design. OpenAI should not substitute for lawful public oversight, but we also should not deploy powerful AI systems where no appropriate actor can meaningfully oversee their use. High-risk deployments should be structured so government oversight officials with appropriate authority, expertise, and access can oversee how the technology is used, and, where the risk warrants, be supplemented by additional technical oversight mechanisms such as audit logs, misuse classifiers, or AI-enabled oversight tools.

Applying the Framework Over Time

Over time, the application of these principles to particular opportunities will build precedent and clarify the framework in practice. Through that process, how we implement those commitments may evolve as AI capabilities, risks, and safeguards change, and as we gain new evidence, practical experience, and insight from difficult cases. Consistent with legal and security obligations, that precedent and learning will be shared with the company, and any resulting changes to these principles based on that learning will be communicated directly.