

Anthropic's new usage policy covers deceptive campaigns, weapons software and model abuse

Source copy kept by wheresthe.ai for the story "Anthropic's new usage policy covers deceptive campaigns, weapons software and model abuse" (<https://wheresthe.ai/d/anthropic-usage-policy-2026-update-2026-10/>).

Source: <https://www.anthropic.com/news/2026-usage-policy-update>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 473dff2392405bbc527b07e029823718d5d32895de18b5632cba074c0b685e21

Announcements 2026 Usage Policy update Oct 8, 2026

Each year, Anthropic updates its Usage Policy in response to the evolving capabilities of our models, and the feedback we've received from our customers. We're publishing a new version of the policy today. In this post, we summarize the changes we've made. Most of the updates in the latest version are intended to clarify existing rules. In the year since our last refresh, Claude has taken on longer, more independent work. This update provides new examples that show how our rules apply to Claude's new capabilities. It draws on new patterns of misuse in influence operations, weapons development, and surveillance that we've observed in the past year (as we've documented in our latest threat intelligence report). We've also made updates to clarify our requirements for high-risk use cases in areas like health and finance, to add controls for when Claude is used to autonomously take physical actions, and to address abusive behavior toward our models. The updated policy takes effect on November 12. The full Usage Policy and Supported Regions Policy can be found on their respective pages. A new section on deceptive activity As we wrote about last month , we've seen examples of state media outlets, government propaganda offices, and commercial firms using Claude to run networks of fake accounts and fabricated news sites. Our policy already prohibited this activity, but the restrictions were scattered across our elections, fraud, privacy, and disinformation sections. We've consolidated those rules into a new section titled Do Not Engage in Deceptive Campaigns or Artificial Activity, which applies to deceptive activity of any kind (whether political or commercial). It covers efforts to obscure who is behind a message or amplify content through fake accounts or posts, along with building the tools and infrastructure for running influence operation campaigns. Refocusing our elections section We've refined our political content restrictions in each version of our Usage Policy in order to ensure that our models can be used appropriately in democratic processes. In this version, we've further sharpened our section on elections, focusing it specifically on disallowing the use of Claude to deceive voters or disrupt elections (for instance, by spreading false information about candidates or how to vote, impersonating candidates or election officials, or trying to suppress turnout). We've renamed it Do Not Undermine Democratic Processes to reflect that scope. For a similar reason, we've also removed our blanket prohibition on personalized vote and campaign targeting. This prohibition covered legitimate civic work, such as nonprofits using Claude to write information for voters in other languages, or election officials sending ballot cure notices. The behaviors that motivated the original rule (like voter targeting that relies on deception or the misuse of voters' personal data) remain prohibited under our deceptive campaigns, surveillance, and privacy sections. Clarifying our weapons prohibitions Our Usage Policy has always prohibited using Claude to develop weapons. Recently, we've seen people attempt to use our models to build guidance and control software for weapons. The updated section makes clear that our prohibitions include the software and components that make weapons work, as well as actions like arming drones and other autonomous vehicles. These changes reflect how we've enforced our previous policy. More precise language on surveillance and law enforcement Our September threat intelligence report documented the growing use of AI to build surveillance systems that identify and track political dissidents. We've rewritten our surveillance and criminal justice section to state more precisely what we prohibit. For instance, tracking people without their consent is prohibited, whether it happens in real time or through analysis of previously collected data. Claude cannot be used to decide or recommend who to investigate, arrest, or charge in a law enforcement or criminal justice process. We also prohibit Claude from being used to build or improve tools designed for surveillance. As with the previous update to our weapons prohibitions, this update does not reflect a change in what we enforce in practice: it is just intended to make our policy more explicit. The new section also spells out which uses remain permitted, including tracking that people have agreed to (such as fraud monitoring), content moderation, journalism, and legal research. Clearer requirements for high-risk use cases When our products are used in ways that can affect someone's health, legal rights, finances, livelihood, or access to essential services, we require there to be a qualified "human in the loop" (that is, someone who has the authority to review and, if necessary, change Claude's recommendations), and for the individual affected to be told that AI

was used. These requirements haven't changed with the updated version of our Usage Policy. However, users have often asked us whether a given use case has these requirements, so we've rewritten our High-risk Use Case Requirements section to answer that more directly. It now lists which kinds of recommendations are covered and which are not. Following the launch of our Model Hardware Standard, this section also adds requirements for the use of our models when they're connected to hardware that takes autonomous physical actions and which might be capable of causing injury. A qualified operator must be able to observe the equipment and stop it if needed; the equipment must also be able to hold a safe state if Claude is disconnected. Addressing abusive behavior toward our models We've added a prohibition on sustained and needless abusive or cruel behavior toward our models. The policy update is meant to apply only in extreme cases, where users repeatedly act cruelly toward our models, with no discernible purpose. It does not apply to common versions of user frustration, pushback, dark creative themes, or model testing and research. This addition aligns with a step we've already taken, allowing Claude models to end rare conversations with persistently abusive users on Claude.ai and Claude Code. Such abuse is the main focus of this update; Claude's ability to end these interactions will remain the primary enforcement mechanism. Clarifying our Supported Regions Policy Last year, we updated our Supported Regions Policy to restrict access to Claude for companies who are majority-owned by entities headquartered in our unsupported regions. We've updated our Supported Regions page to clarify how we enforce this policy. Our policy prohibits the use of Claude by people who are physically located in an unsupported region, by entities who are incorporated or headquartered in one, and by entities who are majority-owned or controlled by persons or entities in those regions. Looking ahead We'll keep updating our Usage Policy as Claude's capabilities and risks change. As with this and our prior updates, we'll seek input from across Anthropic, and from policymakers, subject matter experts, civil society, and the people who use our products.

Related content

An update on independent evaluators To evaluate the safety of our systems, Anthropic is giving employee-like access to a team from METR, a nonprofit that evaluates frontier AI models, and we're in early discussions with several other organizations, including Hugging Face. [Read more](#) **Preparing for the 2026 US midterms** Earlier this year, we shared an update on our election safeguards. Here, we outline how we've continued to strengthen these safeguards ahead of this year's US midterms. [Read more](#) **Anthropic opens Singapore office and signs MOU with the Government of Singapore** Anthropic has opened an office in Singapore and signed an MOU with the country's government to advance safe, responsible AI adoption. [Read more](#)