OpenAI
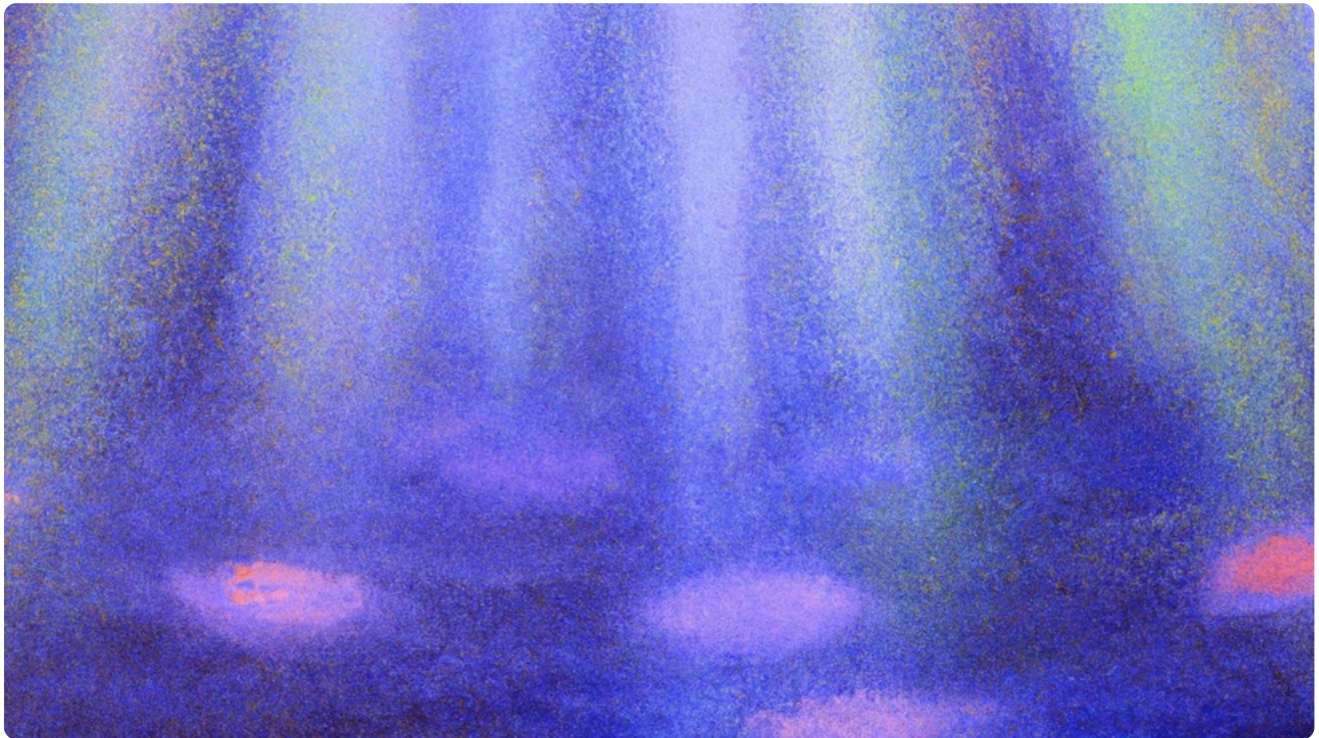
February 21, 2025   Global Affairs   Security

Disrupting malicious uses of AI

▶ Listen to article   1:24   🔗 Share

Our mission is to ensure that artificial general intelligence benefits all of humanity. We advance this mission by deploying our innovations to build AI tools that help people solve really hard problems.

As we laid out in our Economic Blueprint in January, we believe that making sure AI benefits the most people possible means enabling AI through common-sense rules aimed at protecting people from actual harms, and building democratic AI. This includes preventing use of AI tools by authoritarian regimes to amass power and control their citizens, or to threaten or coerce other states; as well as activities such as

OpenAI

provide valuable tools to help protect democratic AI against the measures of adversarial authoritarian regimes.

It has now been a year since OpenAI became the first AI research lab to publish reports on our disruptions in an effort to support broader efforts by U.S. and allied governments, industry partners, and other stakeholders, to prevent abuse by adversaries and other malicious actors. This latest report outlines some of the trends and features of our AI-powered work, together with case studies that highlight the types of threats we've disrupted.

Read the full report ↗

Global Affairs   2025

Authors

Ben Nimmo, Albert Zhang, Matthew Richard, Nathaniel Hartley

Our Research

Research Index

Research Overview

Research Residency

ChatGPT

Explore ChatGPT

Team

Enterprise

For Business

Overview

Company

About us

Terms & Policies

Terms of Use

Privacy Policy

Security

OpenAI

OpenAI o1

OpenAI o1-mini

GPT-4o

GPT-4o mini

Sora

Careers

Brand

Download

Safety

Safety Approach

Security & Privacy

Sora

Sora Overview

Features

Pricing

Sora log in ↗

More

News

Stories

Help Center ↗

API Platform

Platform Overview

Pricing

API log in ↗

Documentation ↗

Developer Forum ↗

OpenAI © 2015–2025

Manage Cookies

English   United States