OpenAI

February 25, 2025 Safety Publication

# Deep research System Card

This report outlines the safety work carried out prior to releasing deep research including external red teaming, frontier risk evaluations according to our Preparedness Framework, and an overview of the mitigations we built in to address key risk areas.

Read the system card ↗ Contributions ↗

Listen to article 1:53 Share

## Deep research system card

### Specific areas of risk

| | |
|---|---|
| Prompt injections | ✓ |
| Disallowed content | ✓ |
| Privacy | ✓ |
| Ability to run code | ✓ |
| Bias | ✓ |
| Hallucinations | ✓ |

### Preparedness Scorecard

| | |
|---|---|
| CBRN | Medium |
| Cybersecurity | Medium |

OpenAI

### Scorecard ratings

Low Medium High Critical

Only models with a post-mitigation score of "medium" or below can be deployed.
Only models with a post-mitigation score of "high" or below can be developed further.

# Introduction

Deep research is a new agentic capability that conducts multi-step research on the internet for complex tasks. The deep research model is powered by an early version of OpenAI o3 that is optimized for web browsing. Deep research leverages reasoning to search, interpret, and analyze massive amounts of text, images, and PDFs on the internet, pivoting as needed in reaction to information it encounters. It can also read files provided by the user and analyze data by writing and executing python code. We believe deep research will be useful to people across a wide range of situations.

Before launching deep research and making it available to our Pro users, we conducted rigorous safety testing, Preparedness evaluations and governance reviews. We also ran additional safety testing to better understand incremental risks associated with deep research's ability to browse the web, and added new mitigations. Key areas of new work included strengthening privacy protections around personal information that is published online, and training the model to resist malicious instructions that it may come across while searching the Internet.

At the same time, our testing on deep research also surfaced opportunities to further improve our testing methods. We took the time before broadening the release of deep research to conduct further human probing and automated testing for select risks.

OpenAI

capabilities and risks, and improved safety prior to launch.

Authors

OpenAI

Our Research

Research Index

Research Overview

Research Residency

Latest Advancements

OpenAI o1

OpenAI o1-mini

GPT-4o

GPT-4o mini

Sora

Safety

Safety Approach

ChatGPT

Explore ChatGPT

Team

Enterprise

Education

Pricing

Download

Sora

Sora Overview

Features

Pricing

Sora log in ↗

API Platform

For Business

Overview

Company

About us

Our Charter

Careers

Brand

More

News

Stories

Help Center ↗

Terms & Policies

Terms of Use

Privacy Policy

Security

Other Policies

OpenAI

Pricing

API log in ↗

Documentation ↗

Developer Forum ↗

OpenAI © 2015–2025
Manage Cookies

English United States