OpenAI Preparedness Framework

OPENAI-PREPAREDNESS-2023 · US · voluntary code

Source: https://policywindow.org/wiki/openai-preparedness

Generated 2026-11-19T01:06:07 UTC

Summary

Capability-tier risk evaluation regime with four categorical levels (Low / Medium / High / Critical) across four risk categories (cybersecurity, CBRN, persuasion, model autonomy). Pre-deployment evaluation against the framework gates release decisions; Safety Advisory Group + board-level Safety & Security Committee govern threshold determinations. Seoul Frontier AI Safety Commitments signatory. NOTE (iter-314): the Preparedness Framework is a versioned-evolving artefact; this row pins the originally-published Dec 2023 version, but OpenAI publishes updates on the safety/preparedness page. Citers tracking specific risk-category language or threshold definitions should confirm against the current published version — the catalog re-pins on the next Coverage Games event. Currency (2026-06-21): OpenAI published Preparedness Framework v2 (15 Apr 2025), superseding the Dec 2023 version this row pins — it collapsed the four capability levels (Low/Medium/High/Critical) to two gating thresholds (High/Critical), set three Tracked Categories (Biological and Chemical, Cybersecurity, AI Self-improvement), and moved persuasion out of the framework.

At a glance

Adopted

2023-12-18

Effective

2023-12-18

Status

in force

Primary source

OpenAI Preparedness Framework (Dec 2023)

How to cite this article

APA

Policy Window. (2023). OpenAI Preparedness Framework [Wiki article — Instrument]. https://policywindow.org/wiki/openai-preparedness

CHICAGO

Policy Window. 2023. "OpenAI Preparedness Framework." Wiki article (Instrument). https://policywindow.org/wiki/openai-preparedness.

HARVARD

Policy Window (2023) 'OpenAI Preparedness Framework', Wiki article — Instrument, available at: https://policywindow.org/wiki/openai-preparedness.

OSCOLA

Policy Window, 'OpenAI Preparedness Framework' (Wiki article — Instrument, 2023) <https://policywindow.org/wiki/openai-preparedness> accessed [date].

BIBTEX

```
@misc{policywindow-openai-preparedness,
title  = {OpenAI Preparedness Framework},
author = {Policy Window},
year   = {2023},
howpublished = {OpenAI Preparedness Framework (Dec 2023)},
url    = {https://policywindow.org/wiki/openai-preparedness},
note   = {Primary source: https://openai.com/safety/preparedness}
}
```