

OpenAI Preparedness Framework

OPENAI-PREPAREDNESS-2023 · US · voluntary code

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

Generated 2026-05-30T22:11:15 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.

At a glance

Adopted

2023-12-18

Status

in force

Effective

2023-12-18

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}  
}
```