

Meta Frontier AI Framework

META-FRONTIER-2024 · US · voluntary code

Source: <https://policywindow.org/wiki/meta-frontier-ai-framework>

Generated 2026-05-30T22:11:20 UTC

Summary

Meta's open-weight-frontier governance posture. Categorises frontier models into 'high risk' + 'critical risk' tiers; the framework's distinctive feature is its explicit defence of open-weight release as a governance posture (vs. the closed-model stance of Anthropic / OpenAI / DeepMind). Pre-release threat modelling + post-release monitoring; commits to halt training if critical-risk threshold reached without mitigations. Seoul Frontier AI Safety Commitments signatory.

At a glance

Adopted

2024-02-02

Status

in force

Effective

2024-02-02

Primary source

Meta Frontier AI Framework (Feb 2024)

How to cite this article

APA

Policy Window. (2024). Meta Frontier AI Framework [Wiki article — Instrument]. <https://policywindow.org/wiki/meta-frontier-ai-framework>

CHICAGO

Policy Window. 2024. "Meta Frontier AI Framework." Wiki article (Instrument). <https://policywindow.org/wiki/meta-frontier-ai-framework>.

HARVARD

Policy Window (2024) 'Meta Frontier AI Framework', Wiki article — Instrument, available at: <https://policywindow.org/wiki/meta-frontier-ai-framework>.

OSCOLA

Policy Window, 'Meta Frontier AI Framework' (Wiki article — Instrument, 2024)
<<https://policywindow.org/wiki/meta-frontier-ai-framework>> accessed [date].

BIBTEX

```
@misc{policywindow-meta-frontier-ai-framework,  
  title = {Meta Frontier AI Framework},  
  author = {Policy Window},  
  year = {2024},  
  howpublished = {Meta Frontier AI Framework (Feb 2024)},  
  url = {https://policywindow.org/wiki/meta-frontier-ai-framework},
```

```
note = {Primary source: https://ai.meta.com/responsible-ai/}  
}
```