Meta Frontier AI Framework

META-FRONTIER-2024 · US · voluntary code

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

Generated 2026-11-19T01:06:06 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. Currency (2026-06-21): On 2026-04-08 Meta released the Advanced AI Scaling Framework v2.0, superseding/renaming the original Frontier AI Framework — it adds a 'Loss of Control' risk domain alongside Cybersecurity and Chemical & Biological, strengthens deployment-decision criteria, and introduces public Safety & Preparedness Reports (per ai.meta.com/blog/scaling-how-we-build-test-advanced-ai). Note: the framework was first published 2025-02-03 (not Feb 2024 as recorded).

At a glance

Adopted

2025-02-03

Effective

2025-02-03

Status

in force

Primary source

Meta Frontier AI Framework (Feb 2025)

How to cite this article

APA

Policy Window. (2025). Meta Frontier AI Framework [Wiki article — Instrument]. https://policywindow.org/wiki/meta-frontier-ai-framework

CHICAGO

Policy Window. 2025. "Meta Frontier AI Framework." Wiki article (Instrument). https://policywindow.org/wiki/meta-frontier-ai-framework.

HARVARD

Policy Window (2025) '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, 2025) <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   = {2025},
howpublished = {Meta Frontier AI Framework (Feb 2025)},
url    = {https://policywindow.org/wiki/meta-frontier-ai-framework},
note   = {Primary source: https://ai.meta.com/responsible-ai/}
}
```