

MMLU-Pro

MMLU-PRO · general reasoning benchmark · 2024

Source: <https://policywindow.org/wiki/mmlu-pro>

Generated 2026-07-09T20:34:54 UTC

Summary

Successor to MMLU with 10-option multiple-choice (up from 4), more reasoning-focused tasks, and removed leaky / ambiguous items.

At a glance

Score range

0–100 % accuracy

Methodology

<https://arxiv.org/abs/2406.01574>

Contamination risk

medium

Saturation

satürating

Details

Less saturated than MMLU. Frontier models ~70-80%.

How to cite this article

APA

Policy Window. (2024). MMLU-Pro [Wiki article — Benchmark]. <https://policywindow.org/wiki/mmlu-pro>

CHICAGO

Policy Window. 2024. "MMLU-Pro." Wiki article (Benchmark). <https://policywindow.org/wiki/mmlu-pro>.

HARVARD

Policy Window (2024) 'MMLU-Pro', Wiki article — Benchmark, available at: <https://policywindow.org/wiki/mmlu-pro>.

OSCOLA

Policy Window, 'MMLU-Pro' (Wiki article — Benchmark, 2024) <<https://policywindow.org/wiki/mmlu-pro>> accessed [date].

BIBTEX

```
@misc{policywindow-mmlu-pro,  
  title = {MMLU-Pro},  
  author = {Policy Window},  
  year = {2024},  
  howpublished = {MMLU-PRO (2024)},  
  url = {https://policywindow.org/wiki/mmlu-pro},  
  note = {Primary source: https://arxiv.org/abs/2406.01574}  
}
```