

MMLU

MMLU · general reasoning benchmark · 2020

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

Generated 2026-05-30T22:07:42 UTC

Summary

Massive Multitask Language Understanding — 57-subject multiple-choice covering humanities, STEM, social sciences, professional/legal.

At a glance

Score range

0–100 % accuracy

Methodology

<https://arxiv.org/abs/2009.03300>

Contamination risk

high

Saturation

saturated

Details

Saturating — top models ~92%. Test-set leakage to training corpora is widely documented. MMLU-Pro is the harder successor.

How to cite this article

APA

Policy Window. (2020). MMLU [Wiki article — Benchmark]. <https://policywindow.org/wiki/mmlu>

CHICAGO

Policy Window. 2020. "MMLU." Wiki article (Benchmark). <https://policywindow.org/wiki/mmlu>.

HARVARD

Policy Window (2020) 'MMLU', Wiki article — Benchmark, available at: <https://policywindow.org/wiki/mmlu>.

OSCOLA

Policy Window, 'MMLU' (Wiki article — Benchmark, 2020) <<https://policywindow.org/wiki/mmlu>> accessed [date].

BIBTEX

```
@misc{policywindow-mmlu,  
  title = {MMLU},  
  author = {Policy Window},  
  year = {2020},  
  howpublished = {MMLU (2020)},  
  url = {https://policywindow.org/wiki/mmlu},  
  note = {Primary source: https://arxiv.org/abs/2009.03300}  
}
```