Humanity's Last Exam

HLE · knowledge benchmark · 2025

Source: https://policywindow.org/wiki/humanitys-last-exam

Generated 2026-11-19T01:06:22 UTC

Summary

3,000+ frontier-difficulty expert-curated questions across all academic disciplines. Designed to remain unsaturated through 2026+.

At a glance

Score range

0–100 % accuracy

Contamination risk

low

Methodology

https://lastexam.ai/

Saturation

active

Details

Center for AI Safety + Scale AI collaboration. Frontier models 8-22% at launch. Replaces MMLU as the de-facto knowledge ceiling. Currency (2026-06-21): HLE SOTA has climbed past the article's framing — Artificial Analysis (June 2026) shows Claude Fable 5 ~53.3%, Claude Opus 4.8 ~45.7%, Gemini 3.1 Pro Preview ~44.7% (no-tools), and with-search results (e.g. Qwen3-Max Thinking) report ~49-58%; the article's top milestone (Gemini 3 Pro Preview 37.5%, "well above 40%") now understates the frontier and "active/unsaturated through 2026+" is strained.

How to cite this article

APA

Policy Window. (2025). Humanity's Last Exam [Wiki article — Benchmark]. https://policywindow.org/wiki/humanitys-last-exam

CHICAGO

Policy Window. 2025. "Humanity's Last Exam." Wiki article (Benchmark). https://policywindow.org/wiki/humanitys-last-exam.

HARVARD

Policy Window (2025) 'Humanity's Last Exam', Wiki article — Benchmark, available at: https://policywindow.org/wiki/humanitys-last-exam.

OSCOLA

Policy Window, 'Humanity's Last Exam' (Wiki article — Benchmark, 2025) <https://policywindow.org/wiki/humanitys-last-exam> accessed [date].

BIBTEX

```
@misc{policywindow-humanitys-last-exam,
title  = {Humanity's Last Exam},
author = {Policy Window},
year   = {2025},
```

```
howpublished = {HLE (2025)},
url    = {https://policywindow.org/wiki/humanitys-last-exam},
note   = {Primary source: https://lastexam.ai/}
}
```