

Cambridge Study Finds AI Not Yet Reliable for Grading University Essays

AI in Education · Good practice · United Kingdom

Quality score: 40/100

Evidence strength: Moderate

Summary

Cambridge psychologists and AI researchers tested Claude and ChatGPT grading 750 real essays from three UK universities. The systems matched human examiners' grade bands only 35–65% of the time, rewarding essay length and vocabulary over academic substance.

Key dimensions

Dimension	Score
Evidence of learning impact	1/3
Equity & inclusion	1/3
Transparency, ethics & data protection	2/3
Teacher capability & pedagogy	1/3
Scalability & sustainability	1/3

Evaluated by C-NAPSE

Data sources

[C-NAPSE web research](#) — accessed 2026-09-01

[University of Cambridge](#) — accessed 2026-09-01

[University of Cambridge](#) — accessed 2026-09-01

[EurekAlert](#) — accessed 2026-09-01

[Phys.org](#) — accessed 2026-09-01

Cite this — Evidoria. *Cambridge Study Finds AI Not Yet Reliable for Grading University Essays*. Persistent ID: d938f37e-9a5c-47de-a2f4-cadb5bdbcbf3.