

GenAI 24x7 Tutor — University of Wollongong's Human-Simulated Benchmark of ChatGPT, Wolfram GPT and Tutor Me GPT for Engineering and Maths

AI in Education · Good practice · Australia

Quality score: 20/100

Evidence strength: Cautionary / limited

Summary

University of Wollongong researchers had assistants role-play average and struggling students across 35 engineering/maths topics to benchmark ChatGPT-4, ChatGPT-4o, Wolfram GPT and Tutor Me GPT as tutors, finding frequent errors and near-zero accuracy on some mechanical topics.

Key dimensions

Dimension	Score
Evidence of learning impact	0/3
Equity & inclusion	0/3
Transparency, ethics & data protection	3/3
Teacher capability & pedagogy	0/3
Scalability & sustainability	0/3

Evaluated by C-NAPSE

Data sources

[C-NAPSE web research](#) — accessed 2026-08-24

[University of Wollongong, Australasian Artificial Intelligence in Engineering Education Centre \(AAIEEC\)](#) — accessed 2026-08-24

[AIMS Press — STEM Education journal article](#) — accessed 2026-08-24

[SSRN — companion pilot study \(SEFI 2024\)](#) — accessed 2026-08-24

Cite this — Evidoria. *GenAI 24x7 Tutor — University of Wollongong's Human-Simulated Benchmark of ChatGPT, Wolfram GPT and Tutor Me GPT for Engineering and Maths*. Persistent ID: 551d1ff6-ee91-45f7-9237-a491a5ea8cf9.