

AHEAD Trial — AI-Generated Medical Exam Questions Match Human-Written Ones on Most Measures, With a Flagged Equity Gap

AI in Education · Good practice · Canada

Quality score: 47/100

Evidence strength: Moderate

Summary

A blinded RCT of 258 first-year UBC medical students found AI-generated exam questions were made 5.6x faster than student-written ones, with similar scores and acceptability, though human items discriminated slightly better and a gender gap emerged only on the AI exam.

Key dimensions

Dimension	Score
Evidence of learning impact	1/3
Equity & inclusion	2/3
Transparency, ethics & data protection	3/3
Teacher capability & pedagogy	0/3
Scalability & sustainability	1/3

Evaluated by C-NAPSE

Data sources

[C-NAPSE web research](#) — accessed 2026-09-06

[University of British Columbia Faculty of Medicine](#) — accessed 2026-09-06

[Zenodo - sample question dataset](#) — accessed 2026-09-06

[BMC Medical Education - full text](#) — accessed 2026-09-06

[ClinicalTrials.gov registration NCT07481162](#) — accessed 2026-09-06

Cite this — Evidoria. *AHEAD Trial — AI-Generated Medical Exam Questions Match Human-Written Ones on Most Measures, With a Flagged Equity Gap*. Persistent ID: 6e5d3873-5ba1-4b93-b6a1-329127631218.