Supplementary material for:

Large language model used to simulate psychiatric OSCE scenarios: a medical student perspective

Zhaochu Geng,[1]* Craig S. Webster,[2,3] Yan Chen,[2] Lillian Ng,[4,5] Christian U. Krägeloh,[6] Angel Li,[1] and Marcus A. Henning.[2]

Affiliations:

[1]School of Medicine, University of Auckland, Auckland, New Zealand

[2]Centre for Medical and Health Sciences Education, School of Medicine, University of Auckland, Auckland, New Zealand

[3]Department of Anaesthesiology, School of Medicine, University of Auckland, Auckland, New Zealand

[4]Department of Psychological Medicine, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

[5]Auckland Regional Forensic Services, Health New Zealand, Te Whatu Ora, Auckland, New Zealand

[6]Department of Psychology, Auckland University of Technology, Auckland, New Zealand

*Correspondence to: zgen174@acuklanduni.ac.nz

Appendix

Table 5: Descriptive statistics for AI performance score across four clinical scenarios by prompt condition (three replicates per condition per scenario).

Scenario	Condition	Mean score	Median score	Standard deviation	Range
1	Original	46.7	48	2.3	4
	Context	48	48	0	0
	Variation 1	41.3	40	6.1	8
	Variation 2	36	36	4	8
2	Original	52.3	54	5.7	11
	Context	55.7	54	5.7	11
	Variation 1	44.3	49	6.1	12
	Variation 2	37.7	35	4.6	8
3	Original	53	53	3	6
	Context	58.3	58	2.5	5
	Variation 1	50	53	5.2	9

	Variation 2	42	42	0	0
4	Original	44	48	6.9	12
	Context	51.7	52	3.5	7
	Variation 1	43	45	3.5	6
	Variation 2	46	45	1.7	3

Note: AI performance was assessed by generating three replicates per prompt condition per scenario. These replicates were averaged to produce a single score prior to quantitative analysis. Possible scores ranged from 0 to 100.