Generative AI Detection and the Validity of Professional Competency Assessment

Ms Phillippa Hulme & Dr Jenny Macaskill

Introduction and Aim

Generative Artificial Intelligence (GenAI) has become popular within everyday life, and a recent study shows 37% of students use GenAI frequently to support studies (Kiezebrink 2024). Whilst the benefits of GenAI in education are increasingly reported (Beckingham *et al*., 2024), ethical concerns around its use to demonstrate professional standards remain. With 25% of stage 1 students referred for academic misconduct for professional competencies in 2024/25, this has been highlighted as a key issue within the Biomedical Science courses at Robert Gordon University (RGU) and tools for detecting GenAI use are required to support academic decision making. This research aims to identify the reliability of the Turnitin AI detection tool for responses generated by two popular GenAI platforms, ChatGPT and Microsoft Copilot.

Methods

Step 1: Question Selection

HCPC Standards of Proficiency demonstration questions selected:
- Professional and Regulatory Standards (PRS)
- Equality, Diversity and Inclusion (EDI)
- Data Protection and Confidentiality (DPC)

Step 2: GenAI Generation

ChatGPT Copilot

Step 3: Three output conditions

Version 1: Raw GenAI output

Version 2: AI output with formatting & spelling corrections

Version 3: AI output with added referencing & expanded information

PRS

EDI

DPC

Step 4: Turnitin Analysis

Figure. 1: Flow chart demonstrating the steps to analyse AI generated work through Turnitin

Results

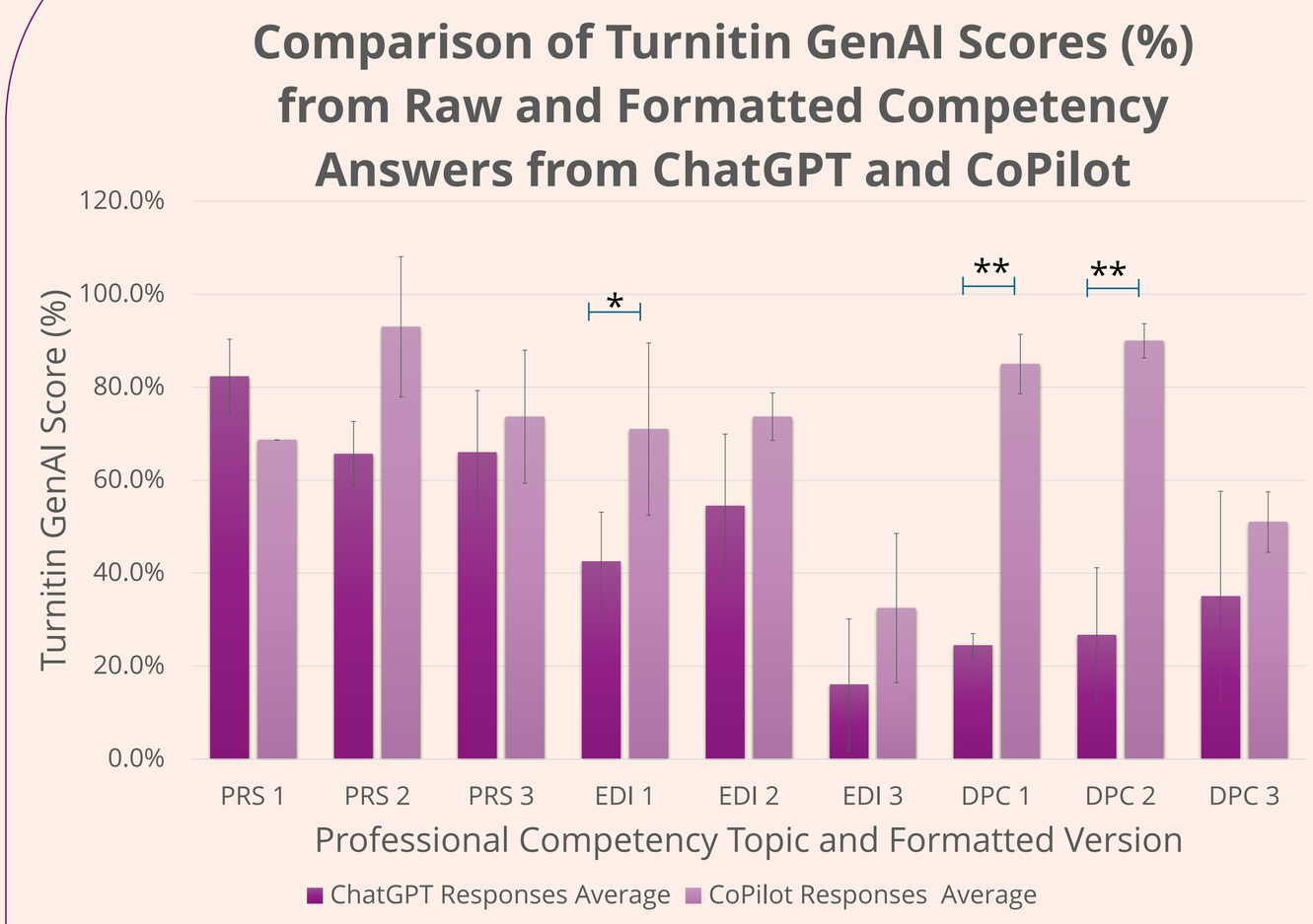

Figure. 2: Turnitin GenAI likelihood scores for three evidence of achievement: Professional and Regulatory Standards (PRS), Equality, Diversity and Inclusion (EDI) and Data Protection and Confidentiality (DPC). n=3, 2 tailed T-test *p<0.05, **p<0.01

Observations of the GenAI text

- Inconsistent referencing style
- Use of American guidelines
- Writing in bullet point style

Figure. 3: Key observations of GenAI use in assessments

Discussion

- The percentage of AI detected using Turnitin is significantly higher using Copilot versus ChatGPT for the EDI and DPC competencies (Figure 2, *p<0.05, **p<0.01).
- Formatting answers reduced the AI percentage detected using ChatGPT and Copilot suggesting student alterations after using GenAI can make detection more difficult.
- Updating assessments (figure 4) has led to a reduced ability to use GenAI and 0% academic misconduct.
- The pass rate increased from 55% at first attempt in 2024/2025 to 88% at first attempt in 2025/2026.

Conclusion

- This work highlights the need for low AI scores to be investigated further.
- Alternative and innovative assessment methods have reduced the ability for students to use GenAI through in class assessments and group work.

Changes implemented to competency assessment

- **Professional & Regulatory Standards** Presentation
- **Equality, Diversity & Inclusion** Group concept map & reflection
- **Data Protection & Confidentiality** MCQ & short answer class test

Key outcomes of alternative assessment

- Varied assessment types reduced ability to rely on GenAI output
- 88% pass rate after switching to varied assessments
- Zero AI-related allegations following redesign

Figure. 4: Key changes and outcomes of the alternative assessments

References:
Beckingham, S. et al (2024) Using Generative AI Effectively in Higher Education. Sustainable and Ethical Practices for Learning, Teaching and Assessment. Routledge.

Kiezebrink, K. *et al*. (2024) GenAI use in higher education: stakeholder perceptions and attitudes. *AdvanceHE*. Available at: GenAI_use_in_HE_stakeholder_perceptions_Dec2024_1734357471.pdf (Accessed 15 January 2026).