CASE STUDY

Smarter, not harder: AI-enhanced assessment design, marking, and feedback

Ria Symonds, School of Mathematical Sciences, University of Nottingham. Nottingham, UK.
Ria.Symonds@nottingham.ac.uk

Abstract

In today's fast-paced academic environment, Generative AI (GenAI) offers a powerful opportunity for STEM educators to work smarter – not harder. This article explores how GenAI can streamline the entire assessment lifecycle: from designing complex, discipline-specific assessments and creating detailed marking rubrics to automating feedback and using transcription tools like Whisper to deliver rich, personalised responses efficiently. Drawing on hands-on experience developing curriculum and assessments for a Data Scientist Degree Apprenticeship at the University of Nottingham, we'll discuss examples that show how GenAI can uphold academic rigour while significantly reducing time spent on repetitive, manual tasks. Rather than replacing academic judgment, these tools enhance it—freeing up time to focus on teaching, innovation, and student support.

Keywords: Apprenticeship, Data Science Generative AI, Assessment.

1. Introduction

Academic practice is entering a transformative era driven by Generative AI (GenAI). It is no longer a futuristic concept, but it is a daily reality for students and educators, reshaping how students learn and how academics teach, design assessments, and deliver feedback. In apprenticeships, on-the-job learning is central, and the shift is especially visible: learners are already deploying AI tools in real workplace contexts, and institutions must bridge policy, pedagogy, and practice to keep pace.

At the University of Nottingham, our Level 6 Data Scientist Degree Apprenticeship integrates mathematics, statistics, software engineering, AI, and machine learning, with authentic, work-based projects at its core. Over the past two to three years, GenAI has become a persistent theme in our teaching, employer reviews, and assessment conversations. In particular, we have observed that apprentices are encouraged to use AI responsibly at work, provided they follow organisational guidance and protect confidentiality. Against this backdrop, a persistent tension emerges: How do we maintain academic standards while staying aligned with the rapid changes shaping industry and society?

It is increasingly likely that graduates will need fluency in AI to remain competitive and drive innovation in their fields. In a recent survey from the Digital Education Council, a global alliance of universities and industry representatives focused on education innovation, it was found that 86% of students globally are regularly using AI in their studies, with 54% of them using AI on a weekly basis (Digital Education Council, 2024). In addition, AI-related jobs are growing 3.6 times faster than all UK jobs, and 64% of CEOs expect major reskilling within three years (PwC, 2024). This is not limited to STEM courses either, 50% of UK business leaders say Arts and Humanities graduates will play a critical role in harnessing AI tools, highlighting the need for all disciplines to embed data and AI skills (Universities UK, 2024).

In this article, we argue that GenAI should support, not replace, academic judgment. Used well, it can cut down repetitive tasks in creating teaching resources and assessments. We show this through examples from our programme: designing assessments that reflect real workplace tasks; creating rubrics that ensure consistent, clear marking; and teaching apprentices to document AI use, justify methods, and assess model risks. Ultimately, graduates need AI fluency to lead innovation ethically, which means embedding AI in the curriculum while strengthening critical thinking and professional judgment.

2. Policy, Practice, and the Responsible Use of AI

At the University of Nottingham, the use of Generative AI in teaching and assessment is still evolving. Currently, staff adoption tends to be ad hoc, with no single source of truth or shared repository of best practice. While there are policies in place, practical examples of how to integrate GenAI effectively remain limited, normally shared through in-house seminars and workshops.

Our Quality Manual addresses AI under "Academic Integrity". It defines False Authorship as submitting work generated or improved by software not permitted for that assessment. This includes tools such as ChatGPT, Copilot, DeepSeek, DALL-E, and similar. The Quality Manual also states that a student should be able to find the permitted use of GenAI for each assessment they encounter but, ultimately, students are responsible for checking and seeking clarification.

Staff are encouraged to use Generative AI to improve marking and feedback, but with caution. Current guidance emphasises three principles: data privacy, meaning only Microsoft Copilot is approved for university use; transparency, requiring staff to inform students how AI has been applied in marking or feedback; and consistency, with a mandatory common statement on AI use included in all coursework assessments from September 2025. An example used on the Data Science apprenticeship can be seen below:

> *"You may use AI (including proof reading tools / large language models, coding copilot tools) during the completion of this coursework. However, you must clearly explain how you used the tools.*
>
> *The tools should be used to improve your development of the knowledge, skills and behaviours and to make your coursework more efficient to deliver, so you can do more than would otherwise have been possible. You should not use AI to avoid the need to develop the knowledge, skills and behaviours.*
>
> *For avoidance of doubt, failure to acknowledge when and how you have used AI will be considered to be 'false authorship' under the university's misconduct regulations, and you risk having a misconduct case brought against you."*

These policies aim to balance innovation with integrity, but the challenge remains: moving from compliance to confident, consistent practice. To support this, we share examples from our programme that demonstrate practical ways to integrate GenAI responsibly into teaching and assessment.

3. Embedding GenAI into the assessment cycle

Our approach to embedding Generative AI within the Data Scientist programme was guided by the existing assessment cycle. Rather than treating AI as an isolated tool, we considered how it could add value at each stage of the process: assessment design, assessment marking, assessment feedback, and student reflection. A summary of this approach can be found in Figure 1.

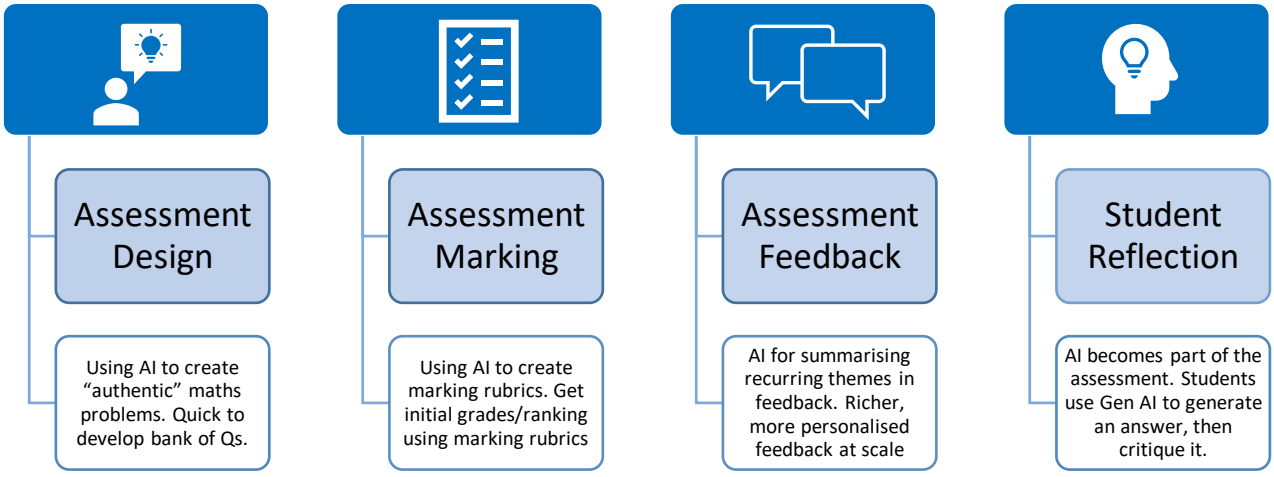

Figure 1: Embedding AI in the assessment cycle

3.1. Assessment Design

On the degree apprenticeship programme, we have adopted the use of authentic assessments that replicate real workplace tasks. Using this style of assessment offers an approach to bridge the gap between academic learning and professional practice while enabling learners to apply subject knowledge and develop essential workplace behaviours and skills (Symonds & Mott, 2024). Instead of traditional exams and coursework, we use synoptic, progressive assignments that reflect real job tasks—such as business case reports, presentations, and portfolios. A key challenge is ensuring these assessments are relevant across diverse sectors where apprentices work, including pharmaceuticals, finance, utilities, e-commerce, and logistics. One way we've addressed this is by using GenAI to transform abstract, highly mathematical questions into applied business problems. For example, instead of asking students to complete a truth table using logical operators, we frame the task around detecting fraudulent transactions: if certain conditions are met—such as 'X and Y' or 'P or Q'—the system flags potential fraud. This approach helps apprentices apply mathematical logic in realistic scenarios while developing problem-solving skills. Once a single scenario is created, GenAI makes it straightforward to generate equivalent questions in different contexts, such as pharmaceuticals or logistics. This means we can assess the same underlying learning outcomes while tailoring the framing to the apprentice's sector, ensuring both authenticity and consistency.

Another way we have used GenAI is to create multiple-choice questions for the apprenticeship End Point Assessment (EPA). One component of the EPA is a 30-question Knowledge Test which we also replicate at the end of each year as a summative assessment. This requires a large, varied question bank that avoids repetition and includes questions at different levels for each stage of study (Year 1, Year 2, Year 3, EPA). To populate this bank, we used Microsoft Copilot to generate questions by feeding in sample items written by the programme team and prompting the tool to cover all "Knowledges" covered by the occupational standard. Creating this bank manually would demand significant staff time, but GenAI has made the process quick and efficient and, moreover, it allows us to easily update and expand the bank as the curriculum evolves.

While GenAI significantly reduces the workload, some refinement of outputs is still necessary. Human verification and quality assurance remain critical to maintain academic integrity. For example, when generating multiple-choice questions, Copilot frequently selected "option b)" as the correct answer. If left unchanged, this pattern could allow students to guess answers without applying

knowledge or reasoning. Similarly, in creating context-based mathematical questions, GenAI often produced numerical values that were too obvious or simplistic, reducing the level of challenge and relevance. To ensure appropriate complexity and authenticity, manual adjustments to values were required before finalising the questions. These examples highlight that while AI accelerates the process, academic oversight is essential to ensure fairness, rigour, and alignment with learning outcomes.

3.2. Assessment Marking

While we have explored AI for assessment design, its role in marking is currently supportive rather than autonomous. We have not yet used AI to grade student work, academic judgment remains essential, but we have found it invaluable for creating detailed marking rubrics that promote consistency and clarity. This is particularly important on apprenticeship programmes, where assessments are authentic to the workplace and can involve complex, multi-step tasks. Apprentices need to understand exactly how learning objectives will be evaluated, and clear rubrics make this possible. On the Data Science programme, we use grid-based rubrics that map each learning outcome to specific performance levels, making 'what good looks like' explicit. This approach helps students interpret feedback, self-assess, and focus their efforts effectively, while ensuring fairness and transparency across modules.

Criteria	Hard Fail (<30%)	Fail (E Grade, 30%)	Pass (D Grade, 40%)	C Grade (50%)	B Grade (60%)	A Grade (70%)	A+ Grade (80%+)
Understanding of Organisation	*Shows no understanding of organisation or data science role.*	*Limited or incorrect understanding of organisation; lacks clarity in purpose of data science.*	*Basic understanding of organisation and purpose of data science pipeline; lacks depth or specificity.*	*Clear understanding of organisation and pipeline purpose; some gaps in specifics.*	*Solid understanding of organisational context, pipeline purpose, and goals.*	*Thorough understanding of organisation, with clear relevance of pipeline to goals.*	*Insightful, detailed understanding; aligns data science pipeline with organisational goals in a sophisticated manner.*
Description of Data Science Pipeline	*Lacks pipeline description; does not address stages or sequence.*	*Vague, inaccurate description; misses most pipeline stages.*	*Basic description of pipeline stages with major gaps or inaccuracies.*	*Adequate description covering main pipeline stages; lacks detail or depth in one or two areas.*	*Detailed, accurate description of each pipeline stage, though some minor gaps.*	*Comprehensive, accurate description; clear sequencing of stages and roles.*	*Well-articulated and sophisticated pipeline description; connects stages seamlessly, reflecting real organisational flow and purpose.*
Application to Specific Scenario	*No scenario provided or scenario lacks relevance to organisation.*	*Scenario barely relates to organisation; limited connection to pipeline.*	*Scenario provided with basic relevance to organisation; some limited connection to pipeline stages.*	*Relevant scenario with clear application to each pipeline stage; some lack of depth in specifics.*	*Scenario is relevant, applied effectively across all pipeline stages with examples.*	*Highly relevant scenario with specific, well-integrated examples across pipeline.*	*Richly detailed scenario that fully exemplifies pipeline; integrates specific methods/tools with strong organisational insight.*
Stakeholder Engagement	*No mention of stakeholders or their role in the pipeline.*	*Minimal reference to stakeholders; unclear role or relevance.*	*Basic identification of stakeholders; limited explanation of their influence or interaction.*	*Clear mention of key stakeholders; some explanation of their influence on pipeline stages.*	*Detailed description of stakeholders and their roles; connection to each stage of pipeline.*	*Insightful analysis of stakeholder roles; demonstrates how stakeholder needs influence pipeline decisions.*	*Excellent, discussion of stakeholders; shows a deep understanding of stakeholder impact and how needs are integrated into pipeline stages.*

Figure 2: Example of marking rubric created autonomously with GenAI

GenAI has helped to streamline the creation of these rubrics. We use AI tools to produce an initial draft with clear, outcome-aligned statements, reducing the time required to create them manually. These are then reviewed and refined by staff to ensure that the statements reflect each grading boundary appropriately. This ensures consistency across the programme and frees staff to focus on higher-value tasks. Students appreciate the transparency this provides, as they can see exactly how their grade was determined and benefit from consistent expectations across all assessments.

3.3. Assessment Feedback

Recently, we have introduced oral feedback as an alternative way to provide richer, more personalised responses to students, recognising that students prefer audio feedback and find it more comprehensive and global than written comments (Cavanaugh & Song, 2014). However, oral feedback can be difficult for students to revisit and reflect on later. To address this, we use Whisper AI to transcribe spoken feedback and summarise key points, ensuring accessibility and creating a permanent record for review. Beyond individual feedback, we combine these transcripts with Copilot to identify common themes across the cohort, producing summary reports that highlight recurring strengths, frequent challenges, and priority areas for development. Shared alongside individual feedback, these reports help students benchmark their performance and focus on improvement. This approach not only saves time compared to writing feedback manually but also enhances its quality, accessibility, and impact.

3.4. Student Reflection

To bridge academic learning with workplace practice, we have asked students to use GenAI to generate solutions to specific mathematical assessment problems and then critically evaluate the output's accuracy and suitability. This exercise goes beyond technical critique: students must decide whether the output is trustworthy, identify limitations, and suggest improvements. They then write a short statement outlining how they would present this solution to a line manager—explaining whether it should be used as-is, adapted, or rejected, and why.

This mirrors real-world scenarios where professionals increasingly rely on AI tools for data analysis and decision-making but remain accountable for the quality and integrity of outputs. By requiring students to justify their recommendations, we develop skills in risk assessment, ethical use of AI, and clear communication—competencies essential for STEM roles. Ultimately, this reflective approach reinforces to our students that AI is a powerful assistant, not a substitute for human judgment. It prepares students to engage critically with technology, ensuring they can harness the power of AI effectively and responsibly in professional contexts.

4. Key Takeaways

We have trialled several ways of using AI to streamline the teaching and assessment cycle in the Data Science Degree Apprenticeship, and so far, the results have been highly positive. These approaches have saved staff time and resource while enhancing student success and engagement. Students report that AI-supported feedback and authentic assessments improve clarity and relevance, and staff value the efficiency gains without compromising academic rigour.

In the workplace, our apprentices report that AI use is permitted and often encouraged, which means our responsibility extends beyond teaching technical skills. We must equip students to use AI responsibly, critique outputs, and communicate best practice. This ensures they can not only apply AI effectively but also educate their organisations on ethical and productive use. Institutional guidance will continue to evolve, so staying informed and sharing best practice is essential. The message is simple: don't be afraid to experiment. By embracing AI thoughtfully, we prepare graduates for a future where human expertise and technological innovation work hand in hand.

5. References

Cavanaugh, A J. & Song, L., 2014. Audio Feedback versus Written Feedback: Instructors' and Students' Perspectives. Merlot Journal of Online Learning and Teaching, 10, 122-137.

Digital Education Council., 2024. Digital Education Council Global AI Student Survey 2024. [online] London: Digital Education Council. Available at: https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024 [Accessed 2 Dec. 2025].

PwC (2024) UK AI Jobs Barometer. [online] PwC UK. Available at: https://www.pwc.co.uk/services/technology/generative-artificial-intelligence/uk-ai-jobs-barometer.html [Accessed 2 Dec. 2025].

Symonds, R. and Mott, L., 2024. Creating authentic assessment in mathematics. MSOR Connections, 22(2), pp. 37-46.

Universities UK., 2023. *Jobs of the Future*. [online] Universities UK.