# AI-Driven Feedback Loops for Student Evaluation

## Contents

# Executive summary

## Key project outcomes

The AI-Driven Feedback Loops for Student Evaluation project aimed to use AI to streamline how student feedback is framed, collected, analysed and reported in the Nuffield Department of Primary Care Health Sciences. The project turned out to be significantly more complex than expected, and ultimately, AI was only used to explore feedback analysis. Process automation tools and workshops with the CTL turned out to be more appropriate in fulfilling other elements of the project brief.

Over six months, staff piloted the use of ChatGPT as a means for feedback analysis. The hope was that AI might significantly decrease time spent on managing feedback processes, enabling efficiencies and a consistent approach across all postgraduate courses in the portfolio. However, by the end of the project lifespan, using current ChatGPT models, staff resource required increased rather than decreased.

Staff found the current iteration of ChatGPT to have limited use for feedback analysis. Chief concerns were:

- inconsistent and oversimplified analysis: ChatGPT was able to identify high-level generalised themes from qualitative data, but could not simultaneously pick out important individual comments. It was not able to analyse both breadth and depth and was therefore unable to adequately analyse feedback data with the level of accuracy or intricacy required. ChatGPT was able to adequately analyse purely quantitative data on its own, but was unable to accurately consider qualitative and quantitative data alongside each other.
- differences in capacity of ‘chat’ vs ‘GPT’ vs various models (eg reasoning model vs language model)
- significant time input required to attain sufficient AI literacy to use ChatGPT for feedback analysis activity on current ChatGPT models
- current licensing frameworks hamper long-term record keeping for individual postgraduate course teams: licenses are linked to individual staff members rather than a programme email address.

The project did not result in producing the intended proof-of-concept which could be rolled out across the department’s programmes, or indeed the wider university. However, it did provide staff with a clearer understanding of AI’s capabilities and limitations, and a desire to revisit AI use for feedback analysis as LLMs iterate and improve. The project also generated ideas for non-AI interventions that could simplify and improve feedback mechanisms, eg

through reviewing feedback question frameworks and instituting automated written communications.

Overall, staff gained general confidence using AI through hands-on sessions, developing skills in prompting and experimenting with tasks beyond the project scope, such as proofreading and summarising documents. This upskilling contributes to longer-term AI literacy within the team, which has potential to catalyse innovation in teaching and learning, but it is not specifically related to the intended outcome of this project.

## Key learning for the team around using AI for developing teaching and learning

This project offered valuable lessons on using AI in teaching and learning. The key learning related to feedback analysis was the shift to viewing ChatGPT as a tool that requires precise guidance (prompts as well as layered contextual training) and post-analysis checks, rather than an autonomous tool:

- While ChatGPT could potentially accelerate simpler tasks like drafting survey questions and summarising pre-analysed feedback, it was not a reliable solution for primary feedback analysis. Trial outputs of primary data analysis required extensive human review to catch omissions, misinterpretations and hallucinations, negating its value for this purpose.
- The importance of clear, well-structured prompts. The team improved their “prompt engineering” skills, discovering that specific, contextual instructions on small data sets enhanced the relevance and accuracy of AI outputs, though human checks were still needed. Development of prompt libraries had some value, but would need ongoing development.
- While reference documentation went some way to inform data analysis capability, human understanding of the layered and complex environments associated with each course and within the department was essential to adequately interpret feedback.
- The interdisciplinary collaboration between educators, administrators, and technologists helped give the team a more realistic, nuanced understanding of AI’s role in feedback work.

# Project introduction

## Background and context

The Nuffield Department of Primary Care Health Sciences (NDPCHS) has seen rapid growth in its postgraduate programme portfolio over the last five years. Each new programme has implemented its own feedback processes,

meaning that there is currently no consistent administrative approach to collecting, analysing and responding to student feedback. Historically, processes have been manually implemented: surveys drafted by hand, results manually collated, and qualitative comments reviewed and reported back through a variety of mechanisms. In the hope of maximising the use of staff time, and facilitating the actioning of student input quickly and effectively, the AI Teaching and Learning Exploratory Fund enabled the department to explore whether AI—specifically large language models like ChatGPT—could streamline this process, reduce admin burden, and enhance teaching quality. The project also aligns with broader institutional priorities around digital innovation in education.

## Objectives

The project had four main objectives:

- **Automate feedback processes** by using AI to design surveys and analyse free-text responses.
- **Reduce admin burden** by shifting repetitive tasks (eg formatting and initial analysis) to AI tools.
- **Shorten feedback cycles t**o improve responsiveness and student satisfaction.
- **Upskill staff** in AI use through training and experimentation, building future capacity.

## Scope

The pilot ran over six months and focused on postgraduate programme feedback. It covered survey-based feedback only, using historical anonymised data. The team worked with readily available tools (ChatGPT, MS Forms), with no development of custom AI systems. The scope excluded direct student involvement and system-wide integration, aiming instead to test feasibility within a small, well-defined use case.

## Tools and technologies

ChatGPT Edu was the central tool, used for survey drafting, thematic analysis, and summarising feedback. Staff used the web interface with University-provided licences. Survey data came via existing platforms (MS Forms, PDF documents) and were manually fed into ChatGPT. Workflow automation was scoped with the Workflow Automation Competency Centre (WACC) but not yet implemented. Other tools included Microsoft Teams (for collaboration), Excel (for charting), and a shared folder for guidance materials and prompt libraries.

## Collaboration

The interdisciplinary team included NDPCHS staff, the Centre for Teaching and Learning (CTL), and the AI and Machine Learning Competency Centre (AICC). CTL advised on pedagogy and survey design; AICC provided AI training and technical support. Weekly drop-ins, Teams chat, and shared documentation supported real-time collaboration. This cross-functional model enabled faster problem-solving and ensured outputs were educationally and technically robust.

# Project outcomes and findings

## Evaluation results

A mid-project evaluation was conducted in April 2025 to assess progress and user experience with ChatGPT. Respondents rated overall satisfaction with the project at an average of 3 out of 5, indicating cautious optimism. Scores for how well the project met expectations averaged 3.25. Open-text responses echoed this – team members noted a slow start, uncertainty around direction, and the need for greater structure. This feedback prompted the team to introduce clearer task definitions for distinct project elements and led to the development of a structured framework for training on using ChatGPT for data analysis, including specific task example-based support. With this new framework in place the project team gained confidence in using ChatGPT and were able to more clearly evaluate the efficacy of AI for feedback analysis. A final evaluation survey is scheduled for late June 2025.
*Placeholder: Final evaluation results to be added.*

## Quantitative data

Survey data showed that team confidence in using ChatGPT was linked to training: all respondents said more practice or structured examples would help. When ranking task priorities, “analysing survey data” was consistently rated highest, followed by refining survey questions. Generating personalised feedback replies ranked lowest. Based on these results, training focused more on data analysis. In terms of training format, the majority of staff rated the drop-in sessions as “somewhat helpful,” with 25% saying they were “helpful.” This suggested room for improvement in session structure.

It should be noted that while “analysing survey data” was consistently the top priority for the project group, the practice of using ChatGPT for this purpose was complex and extremely time consuming and confidence in the tool at the time of this report remains low with current versions of ChatGPT. So, training increased confidence in use of the tool generally, and specifically in its

weaknesses related to feedback analysis.
*Placeholder: Final usage and confidence metrics to be added.*

## Qualitative insights

Open-ended feedback, which reflected the group's use of ChatGPT in general, i.e. including, but beyond the scope of the project, revealed that while ChatGPT was useful for generating and refining text, its analysis of free-text survey responses had limited value, and that close human supervision was essential. Participants consistently reported the AI misclassified, overlooked or fabricated feedback. Trust in the AI's analysis accuracy was low, and the team members cautioned that outputs should be treated at best as drafts requiring verification and not final answers.

More broadly, participants reported AI to be helpful for streamlining routine writing tasks, proofreading or making writing more concise, and became more confident using it for these purposes over time.

In terms of the process of the project itself, the project team members suggested the need for clearer early goals and more structured training from the start. Drop-in sessions were appreciated for peer learning, but a few preferred formal step-by-step guidance, where working through a set task assisted learning. The team also valued interdisciplinary collaboration, which allowed rapid troubleshooting and a balance of pedagogical and technical input. In summary, while ChatGPT added value for specific writing and summarising tasks, effective use depended on structured support, careful review, and clearly defined goals. Its use for feedback analysis was questionable.

# Lessons learned

## Challenges

The project surfaced a range of technical and organisational challenges.

From a feedback analysis point of view:

- ChatGPT variously ignored instructions, produced inconsistent outputs or struggled with multi-step tasks. This affected reliability—staff needed to iterate and/or reiterate prompts or break tasks into smaller steps.
- The tool also lacked nuance when analysing qualitative feedback, requiring considerable human oversight to avoid misinterpretations. This added layer of review consistently offset any theoretical efficiency gains in having ChatGPT prepare an initial draft feedback summary.

- The need for such extensive accuracy checks also significantly undermined user confidence in ChatGPT as a qualitative analysis tool.
- Limitations in sharing information from one ChatGPT account to another: while sharing of Custom GPTs was possible, and would facilitate easy knowledge exchange across programme teams, at the time of this project Custom GPTs did not have the functionality to set which model responses should use, a feature that was found to be essential in producing better outputs. This restricted users in focussing on chats as the mechanism for feedback analysis to achieve our aim, but made knowledge sharing complex.
- Limitations with the chat function: the team relied on building up long, iterative chat threads to get the desired outputs. These chats quickly became cumbersome, fragile, and difficult to maintain, particularly if token limits were reached and early context was lost. While sharing chats was technically possible, the dependence on large, user-specific chat sessions made knowledge sharing difficult and undermined the goal of creating a scalable, reusable framework.

From an organisational point of view:

- Team members had to balance the project alongside other responsibilities, which led to uneven engagement and difficulty maintaining momentum. Drop-in session attendance varied, and some participants fell behind. However, having four NDPCHS staff on the team meant we could share learning even though all team members could not attend all sessions.
- Final meetings of the wider project team were difficult to coordinate, given the number of people involved across three teams in the university (NDPCHS staff, CTL and the AICC).
- A broader challenge was defining success in an exploratory context. The initial project brief had ambitious goals but it became necessary to introduce more focused short-term targets and clearer evaluation criteria mid-way. This could potentially have been resolved earlier in the project with a tighter scoping conversation in the early stages of the project.
- Questions remain concerning licensing agreements. Programme teams need consistent access to protocols related to individual courses, yet licenses can currently only be linked to an individual email address. Retention of historical documentation therefore has challenges.

## Key takeaways

From these challenges, several key lessons emerged:

- AI is best applied to simple, well-defined, repeatable tasks—such as drafting survey questions or summarising verified data analysis, rather than complex analysis requiring contextual understanding of individual programmes, and NDPCHS as a department. Human-AI collaboration worked best when staff pre-processed input, but still required iterative testing and review, which was notably resource-intensive.
- Training mattered. Confidence grew through hands-on experimentation, shared prompt libraries, and support from peers. A mix of structured guidance and informal collaboration (eg Teams chat, shared docs) helped build a sustainable learning environment. Future projects should invest early in user onboarding, provide reusable templates, and schedule time for reflection and adjustment. Where exploratory work is ongoing, it is recommended that teams produce a shared log/activity tracker to assist with progress tracking, noting patterns, successes and failures, and to facilitate reflective learning.
- Interdisciplinary collaboration was a strength. The internal team from NDPCHS provided input from programme management, student experience, digital learning and portfolio strategy perspectives, helping to refine and scope the project appropriately from a number of angles. Additionally, pedagogical support from the Centre for Teaching and Learning (CTL) and technical advice from the AI & Machine Learning Competency Centre (AICC) allowed the team to troubleshoot and keep the project grounded in educational priorities.

## Advice for teams

- AI literacy levels: Provide basic ChatGPT training to upskill staff to a base level of user confidence prior to embarking on more complex activities such as feedback analysis.
- Educate staff on the strengths and weaknesses of ChatGPT for feedback related work - caution use of ChatGPT for analysis until models develop further and can more accurately process complex analysis approaches.
- If staff wish to use ChatGPT for feedback analysis with its current functionality, suggest that feedback is pre-processed and that AI be used for summary or presentation purposes only. Even then, verify its efficacy.
- Use small data sets in any one analysis run in a chat, perhaps by pre-sifting data on a specific topic before running through ChatGPT, to help optimise accuracy and minimise human verification time.
- Have a digital framework in mind for tracking and storing process documents and learning: create a framework for organising and sharing

information between license users and teams. Ensure that human checks are built into any workflow.

- Share findings and resources with others to build institutional knowledge, especially as models improve. Suggest storing any records for information sharing outside of ChatGTP itself to track how models are being used without 'confusing' AI analysis, and also to facilitate easier dissemination of prompt libraries etc.

# Appendices

## Digital artefacts

Includes core project outputs such as the prompt library used for ChatGPT interactions and anonymised examples of chatbot-generated summaries.
*Placeholder: The project team will attach the prompt library and selected anonymised chatbot transcripts.*

## Raw data

Covers anonymised student feedback, coded datasets (manual or AI-generated), and team evaluation responses used to inform findings.
*Placeholder: Links or references to securely stored raw data files will be added by the team (eg behind SSO if required).*

## Templates used

Includes templates used or developed during the project, such as survey forms, evaluation questionnaires, and feedback analysis guides.
*Placeholder: Sample templates and blank forms will be attached for reference.*

## Guidance for users

Resources to support AI use, including staff user guides, quick-start instructions, FAQs, and any relevant student-facing communications.
*Placeholder: The team will provide copies of staff guidance documents and any student communications where applicable.*

## Additional materials

Supplementary internal documents, such as planning notes, training slide decks, or interim reports developed during the project.
*Placeholder: An archive of additional project materials will be provided by the team.*

## SSO-protected information

Refers to sensitive materials not publicly shareable, including raw student feedback with identifiers or internal evaluation data.

*Placeholder: The team will list any SSO-protected materials without exposing their contents—for example, “Full student feedback data (SSO-protected).”*