TEACHING AND LEARNING CONFERENCE 2026

Evaluating GenAI Potential and Integrating it in Engineering Education

From Cognitive Performance to Authentic Peer Assessment

Dr Salman Shahid — Discipline Head of Teaching, University of Manchester
Student Contributors: Antonis Theodorou, Tayyib Majumder, Shaun Walmsley, Odeneho Agyemang-Duah

MANCHESTER 1824

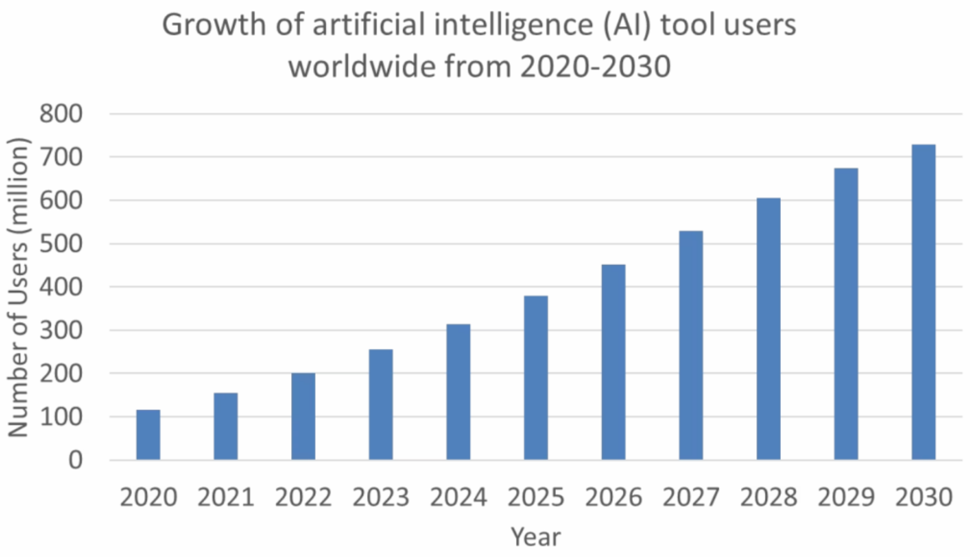
The University of Manchester

CONTEXT

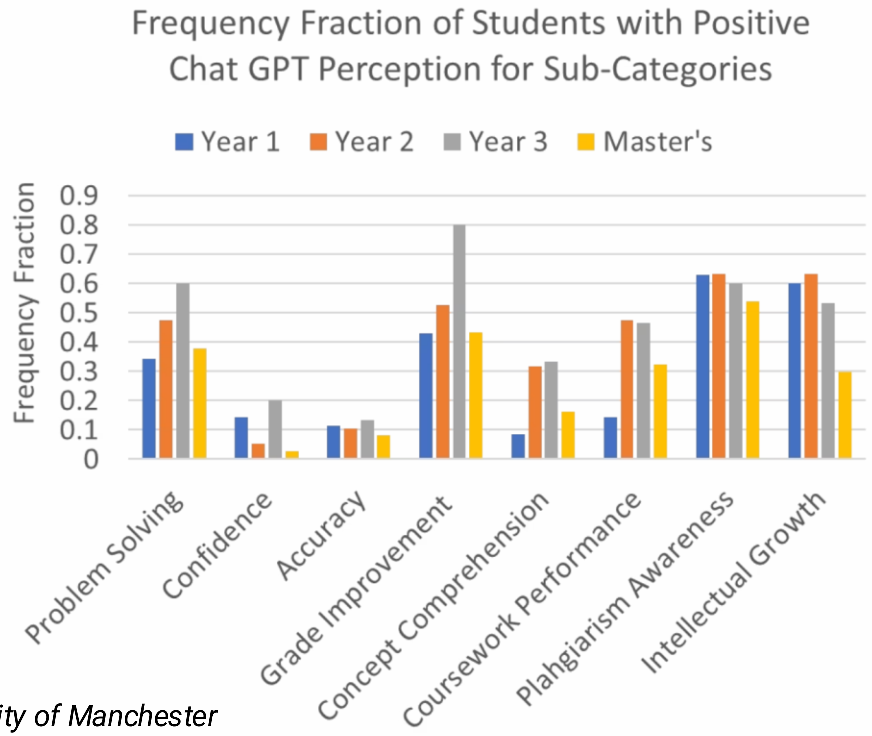
The AI Revolution in Higher Education

UK undergraduates are increasingly turning to AI tools for academic support. A Higher Education Policy Institute (HEPI) survey revealed that **53% of students utilise AI to generate content for their coursework** — a figure that demands urgent pedagogical attention.

The global AI user base is projected to grow from approximately 100 million in 2020 to over 700 million by 2030. This rapid adoption makes it essential that engineering educators move beyond prohibition and instead design learning experiences that develop responsible, critical AI use.



Source: Statista (2022).



Student survey data from University of Manchester

RESEARCH PROGRAMME

From Evaluating AI to Designing AI-Enhanced Learning

This research follows a deliberate two-study structure: we first rigorously evaluated what AI can and cannot do before designing any educational intervention. Evidence precedes design.

Study 1: Evaluate

Bloom's Levels

Identify Limits

Study 2: Design

AI Literacy

This evidence-based sequence ensures that the educational design directly addresses the specific cognitive gaps identified in Study 1, rather than adopting AI tools uncritically or speculatively.

STUDY 1

ChatGPT's Cognitive Performance: Bloom's Taxonomy Analysis

Bloom's Level Performance

Bloom Level	Accuracy
Remember	95%
Understand	88%
Apply	78%
Analyse	64%
Evaluate	53%
Create	41%

- A diverse dataset of 100 undergraduate level problems (exam questions, courseworks, problem sets etc) spanning six cognitive domains was used to assess the ChatGPT 3.5 's reasoning across ascending levels of cognitive complexity.
- Each response was evaluated for accuracy and categorized into five error types.

What the Data Tells Us

ChatGPT's performance **decreases systematically** with cognitive complexity. The model is strong for recall and explanation, but weaker at higher-order thinking.

- **Strong performance** at Remember and Understand
- **Moderate performance** at Apply
- **Significant decline** at Analyse, Evaluate, and Create

The drop from 95% (Remember) to 41% (Create) represents a 54-percentage-point gap — a critical finding for curriculum design.

Three Distinct AI Capability Tiers

Tier 1
Remember / Understand

95–88%

Tier 2
Apply

78%

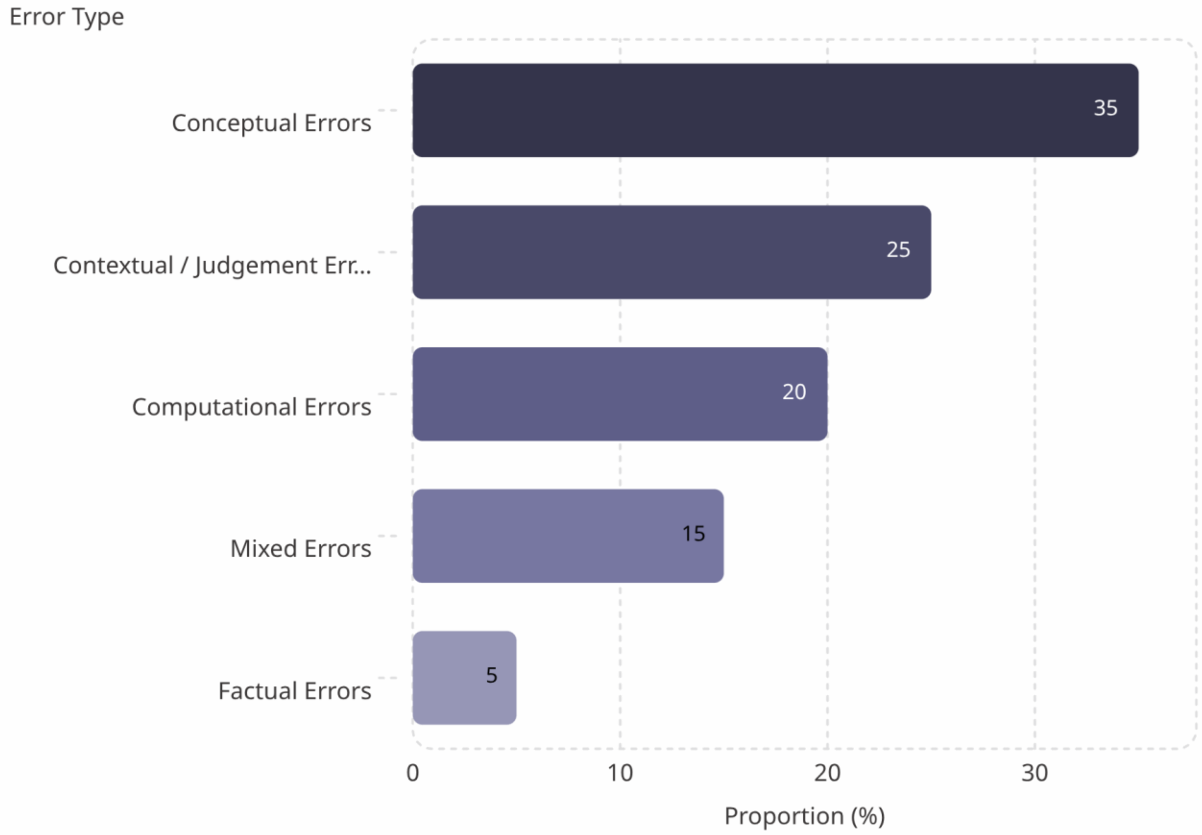
Tier 3
Analyse / Evaluate / Create

64–41%

Source: Shahid S, Walmsley S. Evaluating ChatGPT's Cognitive Performance in Chemical Engineering Education. Information. 2026; 17(2):162.

STUDY 1 · ERROR ANALYSIS

What Types of Errors Does ChatGPT Make?



The Judgement Problem

Factual errors are least common at 5%. The main failure modes are conceptual and contextual — the kinds that matter most in engineering design.

Hollow coherence is the key issue: responses can sound polished and convincing while still containing flawed engineering reasoning.

⚠ Students who mistake fluent AI responses for correct engineering reasoning are at significant academic and professional risk.

This finding directly motivates Study 2: students need structured opportunities to *critique* AI rather than simply consume it.

THE EDUCATIONAL CHALLENGE

"Hollow Coherence" — The Pedagogical Problem

> Students often mistake fluent AI responses for correct engineering reasoning.

AI Responses Sound Convincing

Well-structured, professional prose gives an impression of authority and correctness regardless of technical validity.

AI Responses Look Professional

Formatted outputs with numbered steps, equations, and technical vocabulary create false confidence in the material presented.

AI Responses Can Still Be Wrong

Particularly at Bloom Levels 4–6, outputs may contain conceptual or contextual errors with no visible signal of uncertainty.

Therefore: students require structured experiences that develop the habit of critically interrogating AI outputs — not simply accepting or rejecting them wholesale. This is the foundation of Study 2.

RESEARCH-INFORMED DESIGN

From AI Weaknesses to Educational Design Principles

The intervention was not designed in the abstract. Every design principle maps directly to a weakness identified in Study 1. This explicit linkage is what makes the approach research-informed rather than merely technology-enhanced.

AI Weakness Identified (Study 1)	**Educational Design Response (Study 2)**
Contextual and judgement errors (25%)	Structured peer review to expose flawed reasoning in context
Weak design evaluation (Evaluate: 53%)	Cross-group critique with explicit evaluation criteria
Weak creativity (Create: 41%)	AI-supported brainstorming as scaffold, not substitute
Conceptual errors (35%)	Structured reflection on where AI helped vs. misled

Design Principles: ✓ AI as co-pilot, not autopilot · ✓ Human judgement remains central · ✓ Peer review exposes AI limitations · ✓ Critical evaluation prioritised over answer generation

Methodology: A Three-Stage Framework

1

Stage 1 · Traditional Problem Solving

Students formulate the problem, define assumptions, and develop a solution strategy entirely *without* AI. This establishes a baseline and ensures foundational competence before AI is introduced.

2

Stage 2 · AI-Assisted Redesign

Students use structured ChatGPT prompts to review, critique, and refine their Stage 1 solution. They explore alternatives, test assumptions, and perform conceptual simulation with AI guidance.

3

Stage 3 · Cross-Group Peer Review

Each group critically evaluates another group's AI-enhanced design, identifying actionable improvements. Feedback is used to make final refinements, mirroring professional engineering review practice.

Student Partnership at the Heart of the Research

Student Research Partners

This work was not done *to* students — it was developed *with* them. Student partners were actively involved at every stage of the research design and implementation:

- Antonis Theodorou
- Tayyib Majumder
- Shaun Walmsley
- Odeneho Agyemang-Duah

Their Contributions

01

Designing Activities

Co-created the structured prompt sequences and peer-review rubrics used throughout the unit.

02

Testing Prompts

Iteratively refined ChatGPT prompts to ensure they scaffolded critical thinking rather than bypassed it.

03

Developing the Peer-Review Framework

Contributed domain knowledge to construct evaluation criteria that reflect authentic engineering judgement.

04

Evaluating Implementation

Provided reflective feedback on student experience that directly informed iterative improvements.

STUDY 2 · CASE STUDY

The AI Co-Pilot Philosophy in Practice

Not Auto-Pilot

ChatGPT does not provide ready-made solutions or spoon-feed designs. Instead, it functions as an intelligent co-pilot that poses strategic guiding questions, offers alternative design perspectives, highlights critical trade-offs and safety implications, and challenges students to justify their engineering decisions.

This approach develops process design skills, safety awareness, and critical thinking that extend well beyond simple model answers.

Where AI Added Value

- Generating alternative reaction routes and process configurations
- Hazard identification and preliminary safety analysis
- Sustainability considerations and heat integration suggestions
- Rapid brainstorming of separation sequences

Where Human Judgement Was Essential

- Final design decisions and flowsheet selection
- Safety evaluation and inherently safer design (ISD)
- Economic trade-off analysis and justification
- Validation of AI-generated data against authoritative sources

Process Synthesis: Beyond Simple Reactions

Example: AI Co-Pilot

Scenario: Students are asked to design the cumene process from benzene + propene.
Student role: Identify main and side reactions.
ChatGPT co-pilot role:

- Pprompts:
  - *“List all possible products when benzene reacts with propene using zeolite catalysts.”*
  - *“What secondary reactions produce DIPB/TIPB, and how can these be exploited instead of treated as waste?”*
- Explains: DIPB/TIPB can be recycled in a **transalkylation reactor** to recover cumene
- Asks reflective questions:
  - “Would you design a process to minimise DIPB formation or to recycle it? Which is more sustainable?”

Outcome: Students don’t just write “reaction produces cumene” — they **design a reaction system with recycle integration**, building realistic process synthesis skills.

Exploring Alternative Process Routes

Example: AI Co-Pilot

Scenario: Explore different ways to achieve high selectivity to cumene.

Student role: Draft a conventional fixed-bed design.

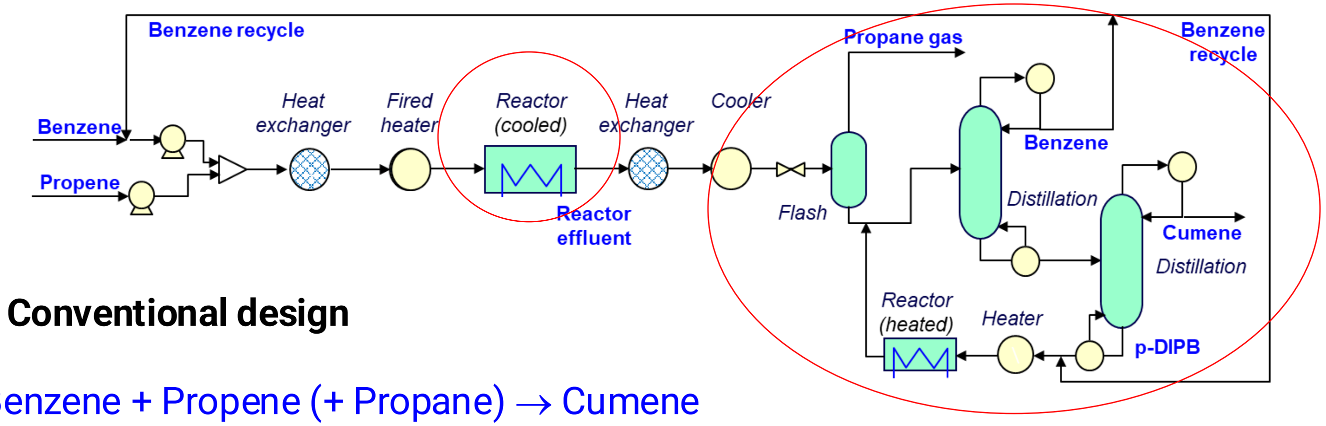
ChatGPT co-pilot role:

- Offers alternatives:
    1. **Multi-bed fixed reactor with inter-bed cooling** – controls exotherm, avoids runaway.
    2. **Tubular reactor with external cooling loop** – allows tighter heat control.
    3. **Reactive distillation** – couples reaction + separation, reducing benzene recycle load
- Probes trade-offs:

  "Reactive distillation reduces separation units but adds operability complexity — how will you control reaction hotspots inside a distillation column?"

Outcome: Students compare **conventional vs intensified designs** and articulate why industry prefers multi-bed packed beds despite novel alternatives.

Example: Cumene production alternative flowsheet and process improvements

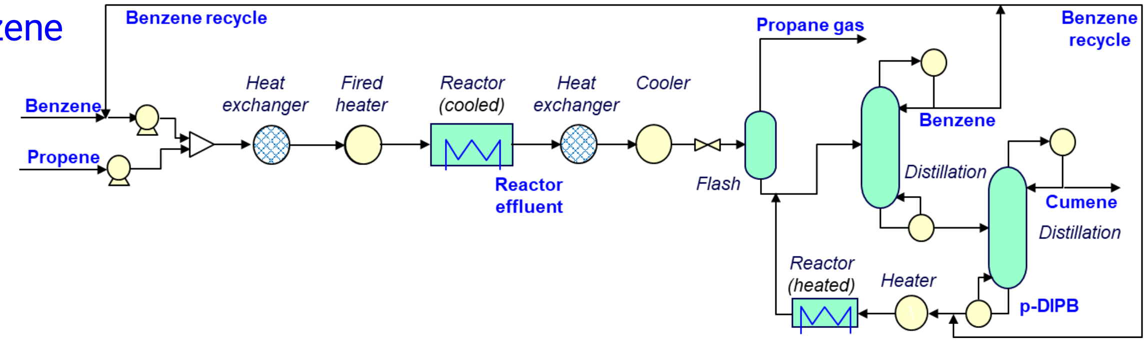


Conventional design

Benzene + Propene (+ Propane) $\rightarrow$ Cumene

Cumene + Propene $\rightarrow$ *p*-Diisopropylbenzene

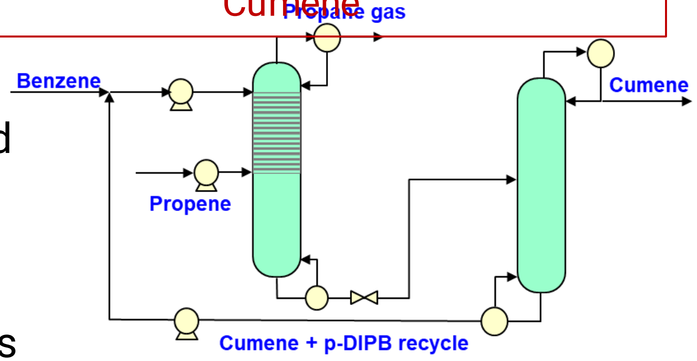
e.g.,
Reactor type
Reactor T & P;
separation operating conditions and hazard control?



Chemical/product process requirements

- Selection of reaction pathway
- Selection of conventional reactor technology
- Selection of reactor feed and operating conditions
- Propose and support separation technology (/Sequence) and removing and recycling reversible by-products
- Identify hazards and propose flowsheet with increased inherent safer design
- Explain economic trade-offs associated with design variables

p-Diisopropylbenzene + Benzene $\rightleftharpoons$ 2 Cumene



Transalkylation reactor

PERFORMANCE IMPACT

AI-Assisted Design: Measurable Performance Improvement

62%

Traditional Approach

Mean peer review score (±8% std). Designs converged on similar conventional flowsheets with limited exploration of alternatives.

81%

AI-Assisted Approach

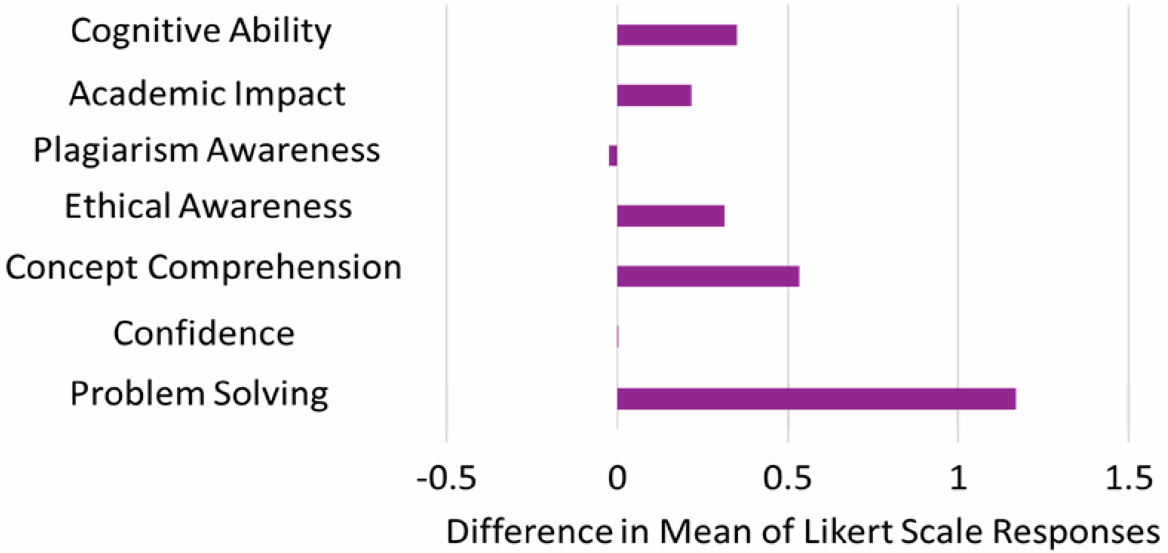
Mean peer review score (±5% std). Significant improvement with *reduced variability* — indicating more consistently strong performance across the cohort.

Why Did Students Improve?

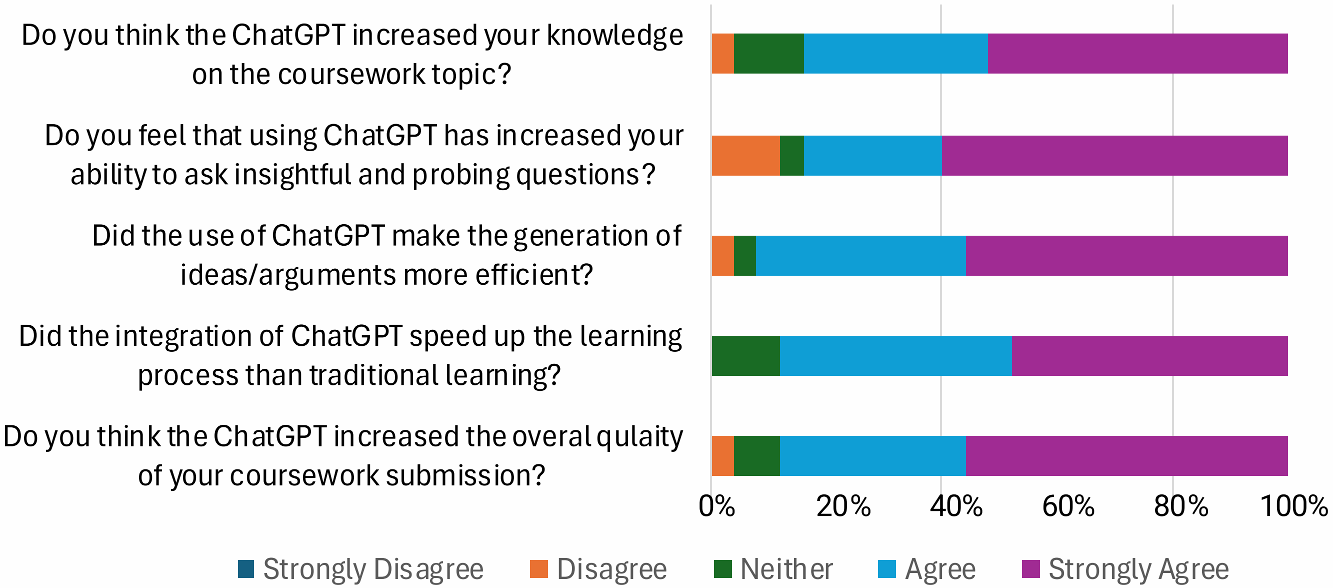
The performance gain reflects genuine development of engineering capability. Students improved because they:

- → Compared alternative solutions rather than settling for the first feasible design
- → Critiqued AI outputs and identified where they were misleading or incomplete
- → Engaged in structured peer review that required articulating engineering reasoning
- → Justified design decisions against economic, safety, and sustainability criteria

Student Feedback Analysis



Comprehensive student survey analysis reveals significant positive impacts of integrating AI alongside important concerns requiring attention.



The distribution of student experience of using ChatGPT shows how it has helped them throughout the project.

STUDENT FEEDBACK

Student Voices: Learning to Critique AI

Over **80% of students expressed positive feedback** regarding the integration of ChatGPT into their project workflow. Crucially, students demonstrated nuanced recognition of both AI capabilities and limitations — precisely the AI literacy outcome the intervention sought to develop.

"ChatGPT integration offered valuable insights into alternative approaches and best practices, enhancing our understanding of the design space."

"We learned to identify errors in ChatGPT suggestions, which improved our critical evaluation and decision-making skills considerably."

"ChatGPT's effectiveness varied depending on the stage of design. It excelled at brainstorming but its utility diminished in more complex tasks requiring judgement."

"Peer marking allowed us to see our work from different viewpoints, understand the strengths and weaknesses of our design, and ultimately grow as engineers."

These qualitative responses directly mirror the quantitative findings of Study 1: students independently identified that AI is strong for brainstorming and weak for complex reasoning — validating the research-informed design approach.

Challenges and Mitigation Strategies

Key Challenges Encountered

Balancing Independence and AI Assistance

Ensuring AI scaffolds rather than replaces independent critical thinking requires careful pedagogical design and explicit guidelines from the outset.

Accuracy and Fabrication

ChatGPT occasionally generated fabricated equipment specifications and cost data — reinforcing why validation is non-negotiable in engineering contexts.

Over-Reliance Risk

Some students were tempted to copy AI-generated text directly, bypassing the critical engagement the activity was designed to develop.

Mitigation Strategies Deployed

- Position AI explicitly as co-pilot, not autopilot, from the first session
- Use structured comparison-based prompting that requires students to evaluate alternatives
- Require peer review to cross-check assumptions and flag AI errors
- Train students in database fact-checking against authoritative engineering sources
- Explicit discussion of AI ethics, academic integrity, and validation responsibilities
- Require a reflective note: *"Where did AI help, and where did it mislead?"*

These strategies reduced blind trust in AI and embedded the habit of checking and justifying every design decision.

Change starts with questions

Thank You

Contact us:

Dr Salman Shahid
Chemical Engineering Department
salman.shahid@manchester.ac.uk