Exploring Human+LLM Feedback Systems: Observing Student Behaviour in Software Engineering Education

UNSW Team: Mr Yiwen Liao, Dr Yuchao Jiang, Dr Yuekang Li, , Miss Iromie Samarasekara, A/Professor Zixiu Guo, Dr George Joukhadar, Professor Fethi Rabhi,
UTS Team: Dr Madhuhi Bandara, Mr Drishtant Leuva , Professor Asif Gill

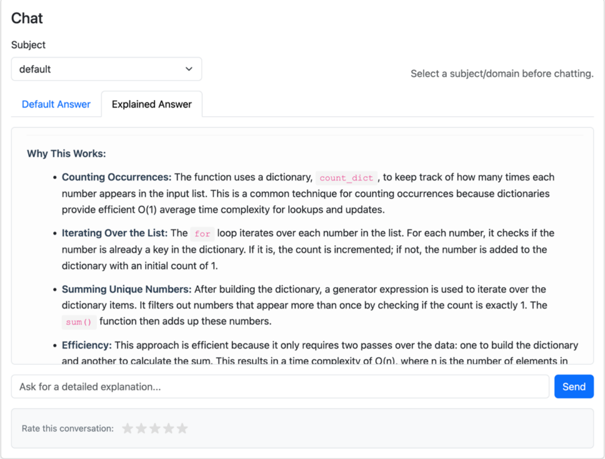
Trustworthy
Digital Society

Overview

Despite automated feedback generation at scale, offering the potential to improve equity and consistency in formative assessment [3], the following challenges inhibit its adoption in practice:

students consistently perceive human feedback as more credible, actionable, and fair [2]. Trust in AI-generated feedback often declines once the source is disclosed, regardless of quality [1].

- Students consistently perceive human feedback as more credible, actionable and fair [2]
- Trust in AI-generated feedback often declines once the source is disclosed regardless of the quality



Chat

Subject

default

Select a subject/domain before chatting.

Why This Works:

- **Counting Occurrences:** The function uses a dictionary, `count_dict`, to keep track of how many times each number appears in the input list. This is a common technique for counting occurrences because dictionaries provide efficient O(1) average time complexity for lookups and updates.
- **Iterating Over the List:** The `for` loop iterates over each number in the list. For each number, it checks if the number is already a key in the dictionary. If it is, the count is incremented; if not, the number is added to the dictionary with an initial count of 1.
- **Summing Unique Numbers:** After building the dictionary, a generator expression is used to iterate over the dictionary items. It filters out numbers that appear more than once by checking if the count is exactly 1. The `sum()` function then adds up these numbers.
- **Efficiency:** This approach is efficient because it only requires two passes over the data: one to build the dictionary and another to calculate the sum. This results in a time complexity of O(n), where n is the number of elements in

Rate this conversation: ★★★★★

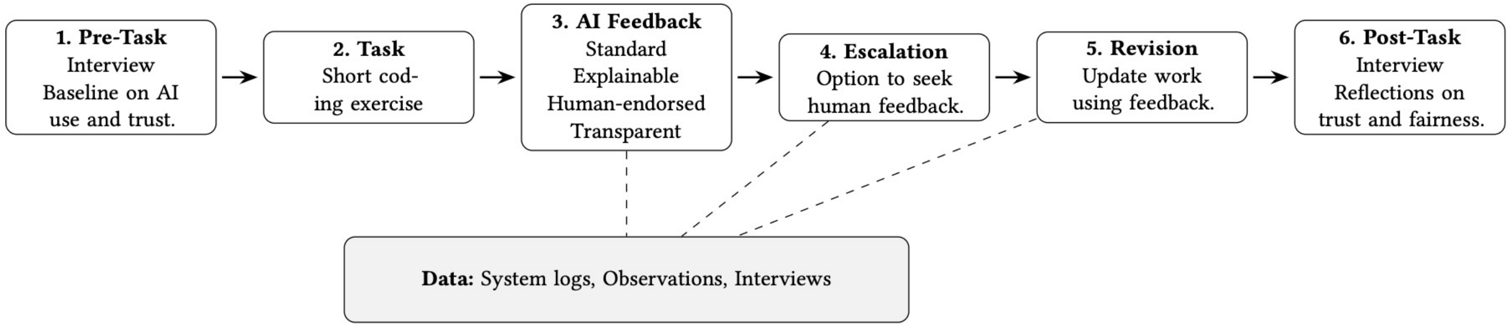
Fig: Hybrid Feedback Platform.

Motivation

- To understand when and why students escalate from AI to human support
- To identify how design interventions such as transparency, explainability, and endorsement affect trust and uptake
- To evaluate whether hybrid AI+human feedback models improve learning outcomes compared to AI-only or human-only systems

Research Questions

1. When and why do students choose to escalate from AI-generated feedback to human feedback after receiving initial guidance?
2. How do factors such as explainability, endorsement, and transparency in AI feedback influence students' trust, engagement, and learning outcomes?
3. Under what conditions are hybrid AI+human feedback models more effective than AI-only or human-only systems in supporting student



1. Pre-Task
Interview
Baseline on AI use and trust.
2. Task
Short coding exercise
3. AI Feedback
Standard
Explainable
Human-endorsed
Transparent
4. Escalation
Option to seek human feedback.
5. Revision
Update work using feedback.
6. Post-Task
Interview
Reflections on trust and fairness.
Data: System logs, Observations, Interviews

Fig: Experiment flow and data sources.

Proposed Techniques

A hybrid feedback platform in which students first receive AI-generated feedback on a short programming or writing explanation task.

- Students may choose to escalate to a human tutor for additional feedback, allowing the analysis of both behavioural decision points (when escalation occurs) and attitudinal responses (how trust and credibility are formed).
- Data will be collected from three complementary sources:
- system logs, researcher observations, and structured interviews to provide a comprehensive view of engagement and trust formation.

Expected Outcomes

- **Escalation patterns**: Identifying when and why students choose to request human input after receiving AI feedback.
- **Impact of design interventions**: Evaluating whether features such as explainability, human endorsement, or transparency increase trust and uptake of AI feedback.
- **Learning behaviour**: Assessing how the above different feedback types influence revision quality, learner confidence, and perceptions of fairness.

Contributions

- **Framework**: A replicable experiment design combining system logs, observations, and interviews to capture behavioural and perceptual aspects of hybrid feedback use
- **Theoretical groundwork**: A critical review highlighting the unresolved gap in empirical evidence on the influence of design interventions on trust and escalation in hybrid AI–human feedback systems.
- **Design Implications**: Empirically informed guidelines for integrating AI feedback in ways that balance scalability withhuman credibility in educational contexts.

References

1. Erkan Er, Gökhan Akçapınar, Alper Bayazıt, Seyyed Kazem Banihashem, and Omid Noroozi. 2024. AI or human? Evaluating student feedback quality in higher education. Assessment & Evaluation in Higher Education 49, 8 (2024), 1012–1026.
2. Erkan Er, Gökhan Akçapınar, Alper Bayazıt, Omid Noroozi, and Seyyed Kazem Banihashem. 2025. Assessing student perceptions and use of instructor versus AI-generated feedback. British Journal of Educational Technology 56 (2025), 1074–1091.doi:10.1111/bjet.13558