Undergraduate Project Report
2024/25

AI-Enhanced Student Skills Development Tracker

Name:	Guan Xin
School:	International School
Class:	2021215116
QMUL Student No.:	210981128
BUPT Student No.:	2021213030
Programme:	e-Commerce Engineering with Law

Date: 07-04-2025

Table of Contents

Abstract

The AI-Enhanced Student Skills Development Tracker is a project designed to evaluate and track the development of critical student skills, such as critical thinking, problem-solving, and collaboration, using advanced artificial intelligence techniques. Traditional methods of assessing student performance, such as tests and assignments, provide limited insights into a student's progress in these multidimensional skill areas. To address this limitation, the system leverages AI, particularly machine learning (ML) and natural language processing (NLP), to analyze a student's behavior, assignment results, and classroom activities. The AI model processes multimodal data, offering both students and educators personalized insights and recommendations for skill improvement. This tracker provides a more holistic view of a student's performance, encouraging incremental growth in essential competencies through real-time feedback. The project aims to support educators in fostering comprehensive student development and empowering students to take control of their own learning journeys. The ultimate goal is to create a system that not only tracks skill development but also provides actionable, data-driven feedback to enhance both teaching and learning practices.

Keywords

AI, Student Skill Development, Critical Thinking, Problem-Solving, Collaboration, Personalized Learning, Natural Language Processing (NLP), Machine Learning, Skill Assessment, Education Technology, Learning Analytics, Feedback System, Recommendation Engine.

摘要

人工智能增强学生技能发展跟踪项目旨在利用先进的人工智能技术评估和跟踪学生关键技能的发展，如批判性思维、解决问题和协作。评估学生表现的传统方法，如考试和作业，对学生在这些多维技能领域的进步提供的见解有限。为了解决这一限制，该系统利用人工智能，特别是机器学习（ML）和自然语言处理（NLP）来分析学生的行为、作业结果和课堂活动。人工智能模型处理多模态数据，为学生和教育工作者提供个性化的见解和技能提高建议。这个跟踪器提供了一个更全面的学生表现视图，通过实时反馈鼓励基本能力的增量增长。该计划旨在支持教育工作者促进学生的全面发展，并使学生掌握自己的学习历程。最终目标是创建一个系统，不仅可以跟踪技能发展，还可以提供可操作的、数据驱动的反馈，以增强教学和学习实践。

关键词

人工智能、学生技能发展、批判性思维、问题解决、协作、个性化学习、自然语言处理（NLP）、机器学习、技能评估、教育技术、学习分析、反馈系统、推荐引擎。

Chapter 1: Introduction

1.1 Project Overview

In contemporary education, the development of student skills is crucial, particularly in areas such as critical thinking, problem-solving, and collaboration. Standard approaches such as the use of tests and assignment results are limited in their ability to measure the changes in students over these diverse skill areas. All these methods do not give a holistic picture of students' progress in the competencies. Due to the continued development of artificial intelligence (AI) in the recent past, the education sector is steadily adopting AI solutions for purposes of improving the assessment of student skills.

The *AI-Enhanced Student Skills Development Tracker* is designed to offer students' performance analysis and recommendations for further improvement based on the data about students' learning behavior, results of the assignments, and activity in the class. This system is to provide timely feedback on students' learning of critical thinking, problem solving and collaboration, so that both students and teachers will be able to improve these skills incrementally.

1.2 Objectives

The main objective of the AI Enhanced Student Skill Development Tracker is to improve the assessment and cultivation of students' basic abilities through AI technology. The project's objectives are designed to address the limitations of traditional educational assessment methods while leveraging technologies to provide actionable, real-time insights. Below is a comprehensive elaboration of the key objectives:

1.2.1 Development of a Comprehensive Analysis Framework

The primary goal is to build a robust framework for processing and analyzing chat logs in group assignment discussions. Traditional skills assessment methods often rely on manual scoring or self-reported questionnaires, which are not only time-consuming but also susceptible to subjectivity. This project will use raw chat data directly as the primary input, ensuring a straightforward way to capture students' collaboration capabilities.

This project aims to create a scalable and replicable system capable of extracting meaningful patterns from unstructured chat data. This will lay the foundation for subsequent skills assessment modules, ensuring that the analytical approach is both systematic and adaptable to different educational scenarios.

1.2.2 Implement skill assessment module based on NLP

The second core objective is to design and implement NLP driven assessment modules that specifically assess critical thinking, problem solving, and collaboration skills. Each competency will be assessed using different but complementary methodologies, ensuring a comprehensive assessment of student performance. In terms of critical thinking assessment, the system will analyze the logical structure of arguments in the chat logs, including identifying arguments, supporting evidence, counterarguments, and logical fallacies. Quantify the depth and coherence of students' reasoning through techniques such as discourse analysis and argument mining.

These modules will be integrated into a unified platform to provide students with clear and understandable skills assessment results. The goal is to go beyond simple outcome-oriented ratings and provide nuanced feedback that both highlights strengths and points to improvement.

1.2.3 Design and deploy real-time feedback mechanisms

The third goal involves creating feedback mechanisms that translate the results of analysis into actionable recommendations. Unlike traditional assessments, which typically provide feedback long after learning activities have ended, this system will emphasize immediacy and relevance.

Students will receive a personalized contribution summary report that identifies specific cases in which they demonstrated critical thinking, effective problem solving, or collaborative behavior in group discussions, along with suggestions for improvement.

To improve usability, feedback systems will be designed with intuitive interfaces that may include visual elements such as skill development graphs or radar maps. The goal is to make the feedback not only informative but also engaging, facilitating students to actively reflect on the learning process.

Chapter 2: Background

In contemporary education, the development of student skills is crucial, particularly in areas such as critical thinking, problem-solving, and collaboration. Standard approaches such as the use of tests and assignment results are limited in their ability to measure the changes in students over these diverse skill areas. All these methods do not give a holistic picture of students' progress in the competencies. Due to the continued development of artificial intelligence (AI) in the recent past, the education sector is steadily adopting AI solutions for purposes of improving the assessment of student skills.

The *AI-Enhanced Student Skills Development Tracker* is designed to offer students' performance analysis and recommendations for further improvement based on the data about students' learning behavior, results of the assignments, and activity in the class. This system is to provide timely feedback on students' learning of critical thinking, problem solving and collaboration, so that both students and teachers will be able to improve these skills incrementally.

2.1 Practical Limitations of Traditional Assessment

Traditional assessment systems have long faced a fundamental contradiction in group work practices: an excessive focus on final deliverables (e.g., project reports, functional prototypes, or presentation performance) while neglecting the value extraction and skill development inherent in collaborative processes. This assessment paradigm essentially simplifies the complexity of team dynamics, and its limitations are increasingly evident in real-world educational contexts. Taking a typical software development team as an example, instructors often grade students based on explicit metrics like code functionality or interface aesthetics, yet struggle to capture critical process-oriented elements such as collaborative interactions among members. Research indicates that traditional grading systems, due to their overemphasis on observable outputs (e.g., verbal participation), lead to two core issues:

First, the "silent contributor" value gap (Hmelo-Silver et al., 2013)[1]. These students may advance team progress through non-verbal means, such as continuously refining documentation structures, coordinating resource allocation during critical phases, or building knowledge-sharing frameworks. Their contributions act as the "lubricant" in a mechanical system—though not directly generating visible outputs, they determine overall efficiency. Yet their work is often relegated to a "black box".

Second, the absence of process-oriented competency evaluation. Skills highly valued by modern employers, such as team leadership and cross-cultural communication, are inherently cultivated through dynamic interactions like conflict resolution and role adaptation. Traditional assessments, however, reduce these competencies to static descriptors on resumes.

2.2 Technical Feasibility of Multimodal Data Integration

The limitations of traditional assessment stem from a disconnect between educational measurement tools and the complexity of collaborative behaviors. Shifting focus to digital collaboration environments reveals that multimodal data offers a technological pathway to address this challenge. Modern teamwork platforms (e.g., GitHub, Tencent Meeting) inherently generate structured and semi-structured behavioral traces: version control systems meticulously document the spatiotemporal distribution of code contributions, collaborative editing preserves the incremental process of knowledge co-construction, and instant messaging tools archive decision-making dialogues. Through multidimensional AI-driven analysis, such data can be transformed into new metrics for assessing collaboration quality—technically, natural language processing (NLP) tools can extract knowledge co-construction features from interaction logs (Wang Yiyan & Zheng Yonghe, 2022; Wilson Chango, 2023)[2].

Correlational analysis of these heterogeneous datasets can uncover implicit collaboration patterns. For instance, a member who contributes minimally during requirement discussions might drive project progress indirectly by influencing others' commit behaviors through sustained code refinements. Such cross-modal contribution recognition mechanisms enable the development of a more holistic competency evaluation framework.

2.3 The development of AI technology

Since the advent of the Turing Test in the fifties, the primary concern of AI scientists has been to create machines that could learn language. In the 1970s, the concept of expert systems was recognized as one of the most important trends in the AI field. These systems addressed real-life issues by transforming knowledge from a specific domain into a set of instructions for computers. In the 1990s ML and NLP brought AI to various fields such as healthcare and finance [3].

In the twenty first century, the developments in the machine learning and deep learning further improved the capabilities of AI and resulted in new application areas like smart learning and intelligent learning tools. AI has evolved from simple data processing to decision making over the recent past due to development in computational power and algorithms [6]; AI is best in language understanding, image identification, and self-reasoning.

In recent years, new large language models (LLMs), including GPT-3, have achieved significant progress because of the growth of pre-trained models and the availability of data. These models not only demonstrate better performance but also have special skills, including contextual awareness and incremental reasoning that cannot be achieved by smaller models, thus allowing AI to take on more challenging tasks and enter new areas of application [7].

2.4 The use of AI in evaluating skills of students

In recent years, the application of AI in education has gradually deepened, especially in personalized learning and student assessment. Big data analysis can be used to monitor students' learning behavior and skill performance in real time, and assist teachers to understand students' learning characteristics and deficiencies in a timely manner so as to offer individualized learning solutions. AI methods including machine learning, deep learning and natural language processing (NLP) have been applied in student assessment and feedback, learning analytics, etc.

Sajja et al. (2023) [7] pointed out that VTA is one of the most discussed issues in the context of the use of AI in education in recent years. With the help of NLP technology, virtual teaching assistants are capable of giving students instant feedback and help through natural language.

Crompton and Song et al. (2021) [8] provided a general introduction to the use of AI in higher education and noted that AI technology can enhance the personalization process of education and enhance the learning performance of students. By the use of the virtual teaching assistant, the students are able to have a real time conversation with the AI and get an immediate response which can be engaging and interactive in the sense that the student is learning.

In the educational field, the progress in AI technology in the field of Natural Language Processing (NLP) has made it possible to automate many tasks including writing of academic papers and text analysis. Neumann et al. (2023) [9] aim at identifying the advantages and disadvantages of using ChatGPT in higher learning institutions focusing on software engineering and academic writing. The study proves that ChatGPT can enhance students' writing skills and plays a supportive role in generating text and academic writing.

In this context, the use of AI technology in education can not only enhance the efficiency of learning effects, but also enhance the educational justice. Big data and machine learning can help students with different learning experiences to have their own learning programs, and each student will have proper learning materials for their level. For instance, in some areas and schools where there are inadequate teachers, AI can fill the vacuum by giving students feedback and assistance through virtual teachers and thus enhance the performance of the learners.

According to Essel et al. (2022) [11], the use of AI chatbots as virtual teaching assistants can enhance educational access and student achievement when teaching in low resource settings. Teacher shortage can be solved by using AI in the classroom and equal opportunities for every student can be given using technology.

Automatic scoring system is one of the applications of AI in the education system. AI can also provide feedback on the student's work, tests, and performance in class without much difficulty. Perkins et al. (2023) [10] mentioned the application of AI tools (including ChatGPT) in formal assessment, while noting that while AI can enhance efficiency in scoring, academic integrity is not yet a solved problem. Hence, ways on how to maintain and implement academic integrity and discourage cheating and academic misconduct when using AI tools has remained a big challenge in the application of AI.

Chapter 3: Design and Implementation

Normally there will be a part about the design and implementation of the system, especially for an implementation type of project. However, every project has its unique phases so you should talk to your supervisor about it.

3.1 Core elements of effective collaboration

The AI-enhanced student skill development tracking system constructed in this study takes data-driven, multi-dimensional evaluation and personality feedback as its core design concepts. By integrating multi-source data, AI analysis and visualization technologies, it can dynamically track students' collaboration skills.

The system employs a microservice architecture, where each service handles a specific task. For example, data collection services are responsible for collecting data from various sources, data preprocessing services transform raw data into a format suitable for analysis, and AI model services use trained models for prediction and analysis. The database service stores all relevant data, including student information, skill scores, and model output. Communication between services is facilitated through RESTful apis, ensuring seamless integration and efficient data transfer.

The main architecture of the system includes three core modules: data layer, analysis layer and application layer. Each module realizes data exchange and function cooperation through standardized interface, which ensures the expansibility and compatibility of the system.

The data layer is responsible for the collection and preprocessing of multi-source data. The system supports access to structured data such as chat records (such as group chat logs in Excel format), basic user information (such as group and role ID), and forms a unified input data set through data cleaning and format conversion.

The analysis layer relies on the large language model (LLM) to build a collaborative ability evaluation engine, conducts semantic analysis of chat records based on preset evaluation dimensions (engagement, initiative, problem solving, etc.), and generates quantitative scores and qualitative feedback.

The application layer outputs results in two forms: interface and PDF report, providing personalized growth summary for students, and team collaboration analysis tools for teachers.

Based on this design concept, the following system structure diagram is designed:

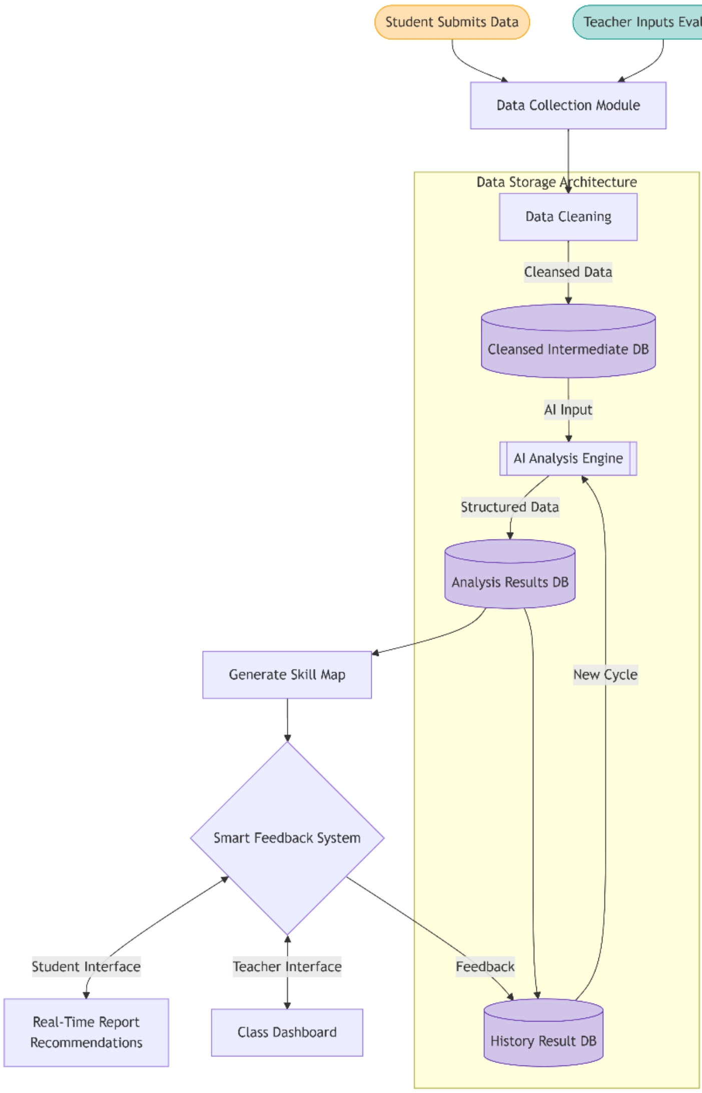

Figure 1 Architecture Diagram

3.2 Key technology module realization

3.2.1 Multi-source data integration and preprocessing

The system will first clean the original chat records, screen out text type messages belonging to users and group them by speakers to form a corpus of individual dimensions. At the same time, the ID mapping mechanism is designed, and the ternary mapping relationship of "original ID - system ID - group" is established through SQLite database to ensure the accuracy of cross-scene data association.

Table 1: Mapping table structure

field name	Data type	state
original_id	TEXT	The original ID of the chat log file
system_id	TEXT	A unique identifier assigned by the system
group_name	TEXT	Group name

In the pre-processing process, the text will be basic processing, and the chat records will be spliced into long text, laying the foundation for the subsequent semantic analysis.

3.2.2 Format for body text

The core functions of the evaluation engine are implemented by a large language model (LLM), with lightweight models (such as the Ollama-based local deployment model) being selected to balance analysis accuracy with computational efficiency. The evaluation dimension was set as five core indicators: participation, initiative, problem solving ability, coordination ability and timeliness of response. Each indicator was quantitatively scored on a 0-5 scale, and specific improvement suggestions were generated.

The System guides the model through a preset System Prompt template to output structured results, including ratings, feedback text, and mentioned key members. For example, for a student's chat record, the model analyzes whether the content of his speech contains collaborative behaviors such as task assignment, question answering, and progress follow-up, and then judges the score of coordination ability.

The evaluation logic is shown as follows:

Table 2: Scoring evaluation logic

Evaluation Dimension	Score (0-5)	Explanation
Participation	0-5	Measures the frequency of engagement in discussions and how often the person contributes to the conversation.

Initiative	0-5	Assesses how often the individual proposes ideas or contributes voluntarily without prompting.
Problem Solving Ability	0-5	Evaluates the ability to address and provide solutions to problems raised during discussions.
Coordination	0-5	Focuses on how well the person works with others, such as task delegation, follow-up on progress, and organizational coordination.
Responsiveness	0-5	Measures the timeliness of responses, including how quickly the person replies to team members' questions or comments.

3.2.3 Format for equations

The report generation module generates a multi-dimensional growth report for each student based on the assessment results, including:

Core skills radar map: visualize the five dimensions of score, intuitively show the strengths and weaknesses;

Periodic feedback summary: Sorting out the improvement suggestions in time order, supporting vertical comparative analysis;

Team collaboration insight: Label high-frequency collaborative objects mentioned to help identify individuals' role positioning in the team.

The report supports two output formats: an interface that displays key information in real time, and a PDF file that provides a structured archive.

3.3 System development and deployment environment

The system is developed using Python language, and the core technology stack includes:

Data processing: Pandas for table data cleaning and conversion, SQLite3 to achieve lightweight data storage;

AI analysis: Ollama library calls the local large language model to realize low-latency inference;

Visualization: FPDF library generates PDF report, combined with Matplotlib to draw radar map and other visual elements;

Deployment: Supports local operation and ADAPTS to the Windows operating system to meet data privacy protection requirements in educational scenarios.

Chapter 4: Results and Discussion

4.1 Test planning and data preparation

In order to comprehensively examine the performance of the AI-enhanced student skill development tracking system, this study developed a comprehensive test plan covering functional, performance and stability aspects. The test data was derived from real education scenarios, which collected chat logs of five course groups over a semester, involving more than 20 students, total about 2,000 text messages. This data forms the basis of system testing to verify the actual performance of each function.

Once the data is collected, the system performs extensive preprocessing. The system identifies and enters missing values using techniques such as mean/median input for numerical data and pattern input for categorical data. Outliers were detected using the interquartile range (IQR) method, and outliers were removed or adjusted. The text data in the forum and job submission was cleaned by removing special characters, converting to lowercase letters, and performing tokenization. After the system automatically eliminates irrelevant information, it will provide basic data for system testing to verify the actual performance of each function.

4.2 Functional test situation

Functional tests focus on the core functionality of the system:

Accuracy of data processing: Compare the original chat records with the data preprocessed by the system to verify the accuracy of ID mapping, text filtering and message grouping, and ensure that the analysis results generated by AI meet the requirements of json format and can be correctly identified. After testing, the false positive rate of AI analysis results is very low, which ensures the reliability of data processing and lays a good foundation for subsequent analysis.

Evaluate consistency: Compare the collaboration scores generated by the model for the same chat record across rows. The overall consistency is high, indicating that the evaluation model has certain validity.

Report generation completeness: Checking the personalized report generation, the final PDF report generated by the system is rich and intuitive. The report contains a line chart of the previous results, which clearly shows the changing trend of students 'collaborative ability ratings at different stages, and helps students and teachers understand the dynamic process of skill development. The radar chart presents students' current skill level distribution in an intuitive graphical way from multiple dimensions such as participation, initiative, and problem solving ability, so that students can clearly see their own strengths and weaknesses. At the same time, the report also gives the average score of each dimension, providing a quantitative overall assessment indicator. After testing, all the reports can be completely generated, and the charts correspond to the text content accurately.

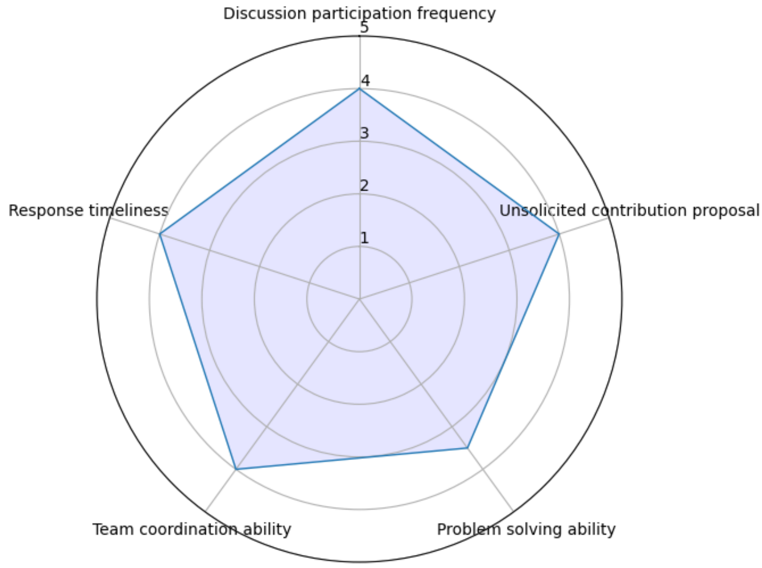

Figure 2 Capability assessment radar chart

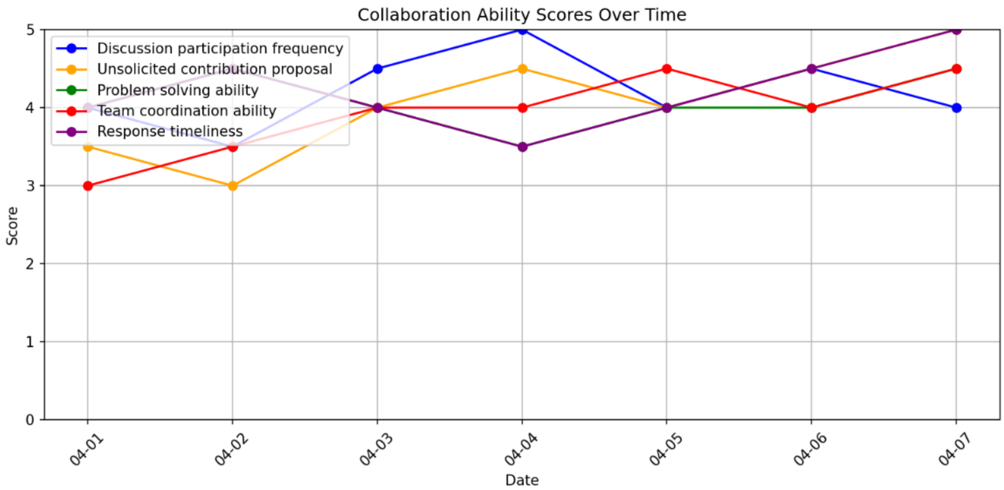

Figure 3 Capability assessment line chart

Chapter 5: Conclusion and Further Work

5.1 Conclusion

This study successfully built an AI-enhanced student skill development tracking system, which realized the accurate and dynamic tracking of students 'collaboration skills with the help of a unique technical architecture and evaluation model. The system integrates multi-source data, uses large language models to carry out deep semantic analysis, and presents the results in the form of personalized reports to provide practical decision-making reference for students and teachers. Key results include:

Construction of evaluation system: Design of collaboration ability evaluation system including participation, initiative, problem solving ability and other dimensions, more comprehensive description of the development of students' collaboration skills.

Implementation technology: the local deployment of large language models takes into account both analysis accuracy and computing resource requirements, while ensuring data privacy security.

Verify the application value: Through the actual test and user feedback, it is confirmed that the system has a certain effect in helping students 'self-cognition and assisting teachers 'teaching decision-making.

5.2 Reflection

The current implementation of the AI-Enhanced Student Skills Development Tracker reveals several critical shortcomings that limit its practical utility in real-world educational settings. Most notably, the system suffers from significant usability challenges that hinder adoption by both educators and students.

A fundamental limitation lies in the system's narrow data scope. The current version relies predominantly on text-based chat logs from collaborative platforms, ignoring other crucial dimensions of student skill development. This limited data paradigm cannot capture creative outputs such as diagrams, prototypes or works of art, patterns of behavior in practical activities or laboratory work

The over-reliance on digital text analysis creates a distorted assessment framework that privileges certain types of learners while marginalizing others. Students who excel in verbal expression may receive inflated evaluations, while those who demonstrate skills through alternative modalities remain undervalued. This data paucity fundamentally constrains the system's ability to provide truly comprehensive skill assessments.

5.3 Further work

A critical area for advancement involves breaking free from the constraints of text-based analysis. The next development phase should establish a multimodal data integration framework capable of processing the full spectrum of student work products. This means developing sophisticated computer vision algorithms to evaluate visual artifacts - from engineering sketches to artistic creations - with the same rigor currently applied to text analysis. Similarly, audio processing capabilities must mature beyond simple transcription to analyse presentation skills, verbal reasoning patterns, and discussion dynamics in classroom recordings. Such expansion would finally allow the system to recognize students who demonstrate competencies through non-textual means, addressing one of the most significant biases in the current implementation.

Technical enhancements must focus on making the system's analytics more nuanced and context-sensitive. Current assessment models often fail to account for disciplinary differences - what constitutes excellent problem-solving in a literature class differs markedly from a robotics lab. Similarly, the system needs better temporal analysis capabilities to distinguish between consistent skill demonstration and isolated moments of competence. Longitudinal tracking algorithms could map skill development trajectories, identifying not just current abilities but growth patterns and potential future performance.

Implementation strategies should prioritize real-world validation through extensive field testing across diverse educational settings. Particularly important will be testing in resource-constrained environments to ensure the system doesn't inadvertently widen educational inequalities through its technological demands. These trials should employ rigorous mixed-methods evaluation, combining quantitative performance metrics with qualitative feedback from educators and students.

The ethical dimensions of an expanded system warrant particular attention as development progresses. Broader data collection necessarily raises more significant privacy concerns, requiring robust anonymization techniques and clear data governance policies. Algorithmic fairness testing must become an ongoing process rather than a one-time check, especially as the system begins assessing more subjective competencies.

The path forward will require balancing technical ambition with pedagogical wisdom. Each enhancement should be guided by fundamental questions about what truly

References

Book Chapters:

[1] Hmelo-Silver, C. E., Chinn, C. A., Chan, C., & O'Donnell, A. M. (2013). The International Handbook of Collaborative Learning. Routledge.

Journal Articles:

[2] Wang, Y. Y., & Zheng, Y. H. (2022). Duo modal xuexi fenxi de jishu lujing yu jiaoyu yingyong [Technological pathways and educational applications of multimodal learning analysis]. China Educational Technology, 41(3), 45–53.

[3] Wen S, Qian L, Hu M, Chang Z. A review of research progress on question-answering technology based on large language models. Data Analysis and Knowledge Discovery. 2023 Nov 22;1-17. Available from: http://kns.cnki.net/kcms/detail/10.1478.G2.20231110.1612.002.html.

[8] Crompton H, Song D. The Potential of Artificial Intelligence in Higher Education. Revista Virtual Universidad Católica del Norte. 2021; n. pag.

[10] Perkins M. Academic Integrity Considerations of AI Large Language Models in the Post-Pandemic Era: ChatGPT and Beyond. Journal of University Teaching and Learning Practice. 2023; n. pag.

[11] Essel HB, et al. The Impact of a Virtual Teaching Assistant (Chatbot) on Students' Learning in Ghanaian Higher Education. International Journal of Educational Technology in Higher Education. 2022;19: n. pag.

[12] Zawacki-Richter O, et al. Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where Are the Educators? International Journal of Educational Technology in Higher Education. 2019;16: n. pag.

[13] Kasneci E, et al. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. Learning and Individual Differences. 2023; n. pag.

Conference Papers:

[9] Neumann M, et al. “We Need To Talk About ChatGPT”: The Future of AI and Higher Education. In: 2023 IEEE/ACM 5th International Workshop on Software Engineering Education for the Next Generation (SEENG). 2023:29-32.

Online sources:

[4] Naveed H, Humza et al. A Comprehensive Overview of Large Language Models. ArXiv. 2023 Jul 1. Available from: https://arxiv.org/abs/2307.06435.

[5] Chen J, Liu Z, Huang X, Wu C, Liu Q, Jiang G, Pu Y, Lei Y, Chen X, Wang X, Lian D, Chen E. When Large Language Models Meet Personalization: Perspectives of Challenges and Opportunities. ArXiv. 2023 Jul 1. Available from: https://arxiv.org/abs/2307.16376.

[6] Lee GG, et al. Multimodality of AI for Education: Towards Artificial General Intelligence. ArXiv. 2023 Dec 1. Available from: https://arxiv.org/abs/2312.06037.

[7] Sajja R, et al. Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education. ArXiv. 2023 Sep 1. Available from: https://arxiv.org/abs/2309.10892.

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisor, for his invaluable guidance, patience, and encouragement throughout this research. I am truly inspired by his dedication to academic excellence and his willingness to offer advice even during busy schedules.

My heartfelt thanks go to my classmates and friends. Their companionship made this journey more enjoyable.

Finally, I owe my deepest appreciation to my family for their unconditional love, understanding, and encouragement over the years. Without their sacrifices, none of my achievements would have been possible.

Appendices

Disclaimer

This report is submitted as part requirement for the undergraduate degree programme at Queen Mary University of London, and Beijing University of Posts and Telecommunications. It is the product of my own labour except where indicated in the text. The report may be freely copied and distributed provided the source is acknowledged.

BUPT No.: 2021213030
QMUL No.: 210981128
Full Name (Pin Yin): Guan Xin
Full Name (Chinese): 管欣

Signature:

Date: 07-04-2025

Project specification

Include your project specification, part 1 and part 2 here. It must be the final version submitted to QMPlus.

Early-term progress report

Mid-term progress report

Supervision log

Additional appendices (as needed)

```
COLLAB_CRITERIA = {
    "participation": "Discussion participation frequency",
    "initiative": "Unsolicited contribution proposal",
    "problem_solving": "Problem solving ability",
    "coordination": "Team coordination ability",
    "responsiveness": "Response timeliness"
}

SYSTEM_PROMPT = f"""As a team collaboration analyst, evaluate the chat history based on the following dimensions:

1. Detailed feedback with improvement suggestions
2. Skill scores (0-5) in these dimensions:
{json.dumps(COLLAB_CRITERIA, indent=4, ensure_ascii=False)}

Scoring rules:
1. Analyze message content, not quantity
2. Focus on constructive contributions
3. Pay attention to coordinating organizational behavior
4. Evaluate the effectiveness of problem solving

Return valid JSON format:
{{
    "feedback": "detail analysis...",
    "scores": {{
        "participation": 0-5,
        "initiative": 0-5,
        "problem_solving": 0-5,
        "coordination": 0-5,
        "responsiveness": 0-5
    }},
    "mentions": ["userA", "userB"],
    "version": "{SCHEMA_VERSION}"
}}
Rules:
1. Output only JSON. No markdown, explanations, or extra text.
2. Use numeric scores (0-5) only.
3. Do not include any comments or additional formatting.
"""
```

Figure 4 System prompt

Risk and Environmental Impact Assessment

In the AI-Enhanced Student Skill Development Tracking System project, a comprehensive assessment of potential risks is critical to ensuring the project's success. We analyse from the dimensions of data quality problem and model performance problem.

Data quality issues: Because the data comes from chat logs of multiple course groups, there may be missing records, inconsistent formatting, or content errors. The probability of occurrence is "Moderate" (L=3), because the data collection process is complex and prone to such problems. If it occurs, it will have a great impact on the project schedule, and the severity of consequences is "Very Serious" (C=3), which may lead to inaccurate model training and affect the evaluation results. Risk level $R = L \cdot C = 3 \times 3 = 9$, which is "Significant Risk".

> **Countermeasures:** Repeated testing, the establishment of a sound data cleaning and verification mechanism, after data collection, through the combination of automated scripts and manual spot checks, data preprocessing and quality check, to ensure data integrity and accuracy.

Poor model performance: The selected large language model may not be able to accurately analyse complex semantics in educational scenarios, resulting in inaccurate collaborative ability assessment. The probability of occurrence is "Unlikely" (L=2), although the model has been preliminarily tested, there may be deviations in practical applications. Once it occurs, it will seriously affect the core function of the system with a severity of "Major" (C=4), which may make the system unable to effectively evaluate student skills and hinder the project. Risk level $R = 2 \times 4 = 8$, which is "Significant Risk".

> **Countermeasures:** In the early stage of the project, comparative tests are conducted on various models to select the most suitable model for the education scene. At the same time, continue to pay attention to the development of the model, timely update the model or adjust the parameters to improve the model performance.