



北京邮电大学
Beijing University of Posts and Telecommunications



Queen Mary
University of London

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.....	3
Keywords.....	3
Chapter 1: Introduction	5
1.1 Project Overview	5
1.2 Objectives	5
1.2.1 Development of a Comprehensive Analysis Framework	5
1.2.2 Implement skill assessment module based on NLP	6
1.2.3 Design and deploy real-time feedback mechanisms	6
Chapter 2: Background.....	7
2.1 Practical Limitations of Traditional Assessment.....	7
2.2 Technical Feasibility of Multimodal Data Integration	8
2.3 The development of AI technology	8
2.4 The use of AI in evaluating skills of students	9
Chapter 3: Design and Implementation	11
3.1 Core elements of effective collaboration	11
3.2 Key technology module realization	12
3.2.1 Data integration and preprocessing.....	12
3.2.2 Evaluation and analysis criteria	13
3.2.3 Format for equations	15
3.3 System development	15
3.3.1 Selection and deployment strategies of large Language models	15
3.3.2 Data processing and collaboration capability assessment process	17
3.3.3 Privacy protection and permission management system	18
3.3.4 System development and deployment environment	18
Chapter 4: Results and Discussion.....	20
4.1 Test planning and data preparation	20
4.2 Functional test situation	20
4.3 System reliability analysis	21
4.3.1 Test design	22
4.3.2 Result analysis.....	22
4.3.3 Format for equations	23
Chapter 5: Conclusion and Further Work.....	24
5.1 Conclusion.....	24
5.2 Reflection	24
5.3 Further work	25
References	27
Acknowledgement.....	28
Appendices	29
Disclaimer	29

AI-Enhanced Student Skills Development Tracker

Project specification	30
Early-term progress report.....	40
Mid-term progress report	44
Supervision log.....	51
Additional appendices (as needed)	53
<i>Risk and Environmental Impact Assessment</i>	<i>54</i>

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) [4] 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.

AI-Enhanced Student Skills Development Tracker

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) [6], 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) [5] 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

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:

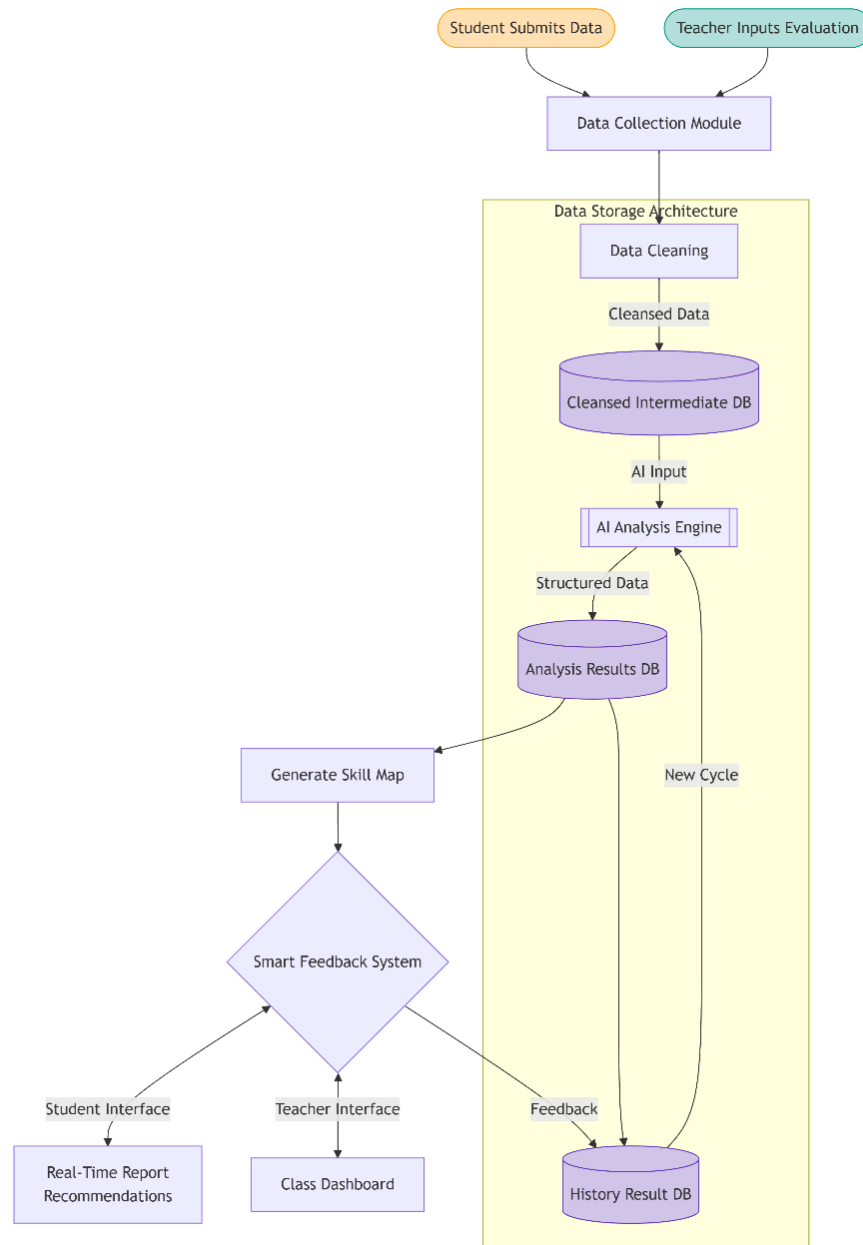


Figure 1 Architecture Diagram

3.2 Key technology module realization

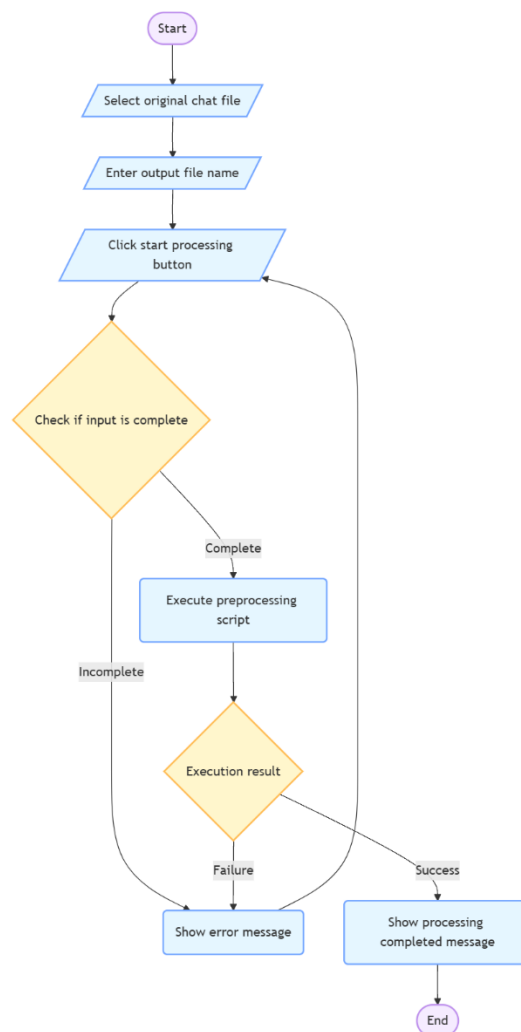
3.2.1 Data integration and preprocessing

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.

**Figure 2 Data Preprocessing flowchart**

3.2.2 Evaluation and analysis criteria

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

3.3.1 Selection and deployment strategies of large Language models

In the educational scenario, the selection of large language models (LLMs) needs to balance the three core requirements of performance, cost and privacy. This system adopts DeepSeek-R1-Distill-Qwen-7B as the core analysis engine.

As a domestic open-source model, DeepSeek-R1 outperforms similar models in tasks such as mathematical reasoning (with a score of 79.8% in AIME2024) and code generation (Codeforces Elo 2029). Its transparent thought chain feature can generate explainable collaborative ability analysis reports. Meet the requirements of educational assessment for process visualization 610. Compared with general dialogue models (such as ChatGPT), R1 is specifically optimized for inference tasks. Through the self-iterative training paradigm of reinforcement learning (RL), it avoids the possible human rating bias introduced by supervised fine-tuning (SFT), and is more

suitable for the strict requirements of objectivity in educational scenarios.

Although cloud-based large models (such as GPT-4) offer more abundant API functions, educational data involves sensitive interaction information between teachers and students and needs to comply with the requirements of the Personal Information Protection Law. Local deployment can completely avoid the risk of data leakage, and at the same time optimize the response speed through hardware customization.

At the level of model lightness, DeepSeek-R1-Distill-Qwen-7B compresses the parameter count from 671B of the original R1 to 7B through progressive knowledge distillation technology, reduces the memory usage to 1.2GB, and can run on consumer-grade GPU (such as NVIDIA 3060).

As shown in Table 1, this system compares the performance of mainstream open-source models in educational assessment tasks. DeepSeek-R1-Distill-Qwen-7B is significantly superior to Llama2-13B (requiring 70GB of video memory) and ChatGLM3-6B (with a 12% lower score in mathematical reasoning) in terms of overall cost performance. Its MIT open-source license even allows commercial applications to be free of licensing fees, providing a legal and compliant basis for possible school-enterprise cooperation. It is worth noting that although there are LAM models like Squirrel Ai specifically tailored for education, their closed-source nature and API call cost (\$3 per thousand times) are difficult to meet the high-frequency, low-budget deployment and development requirements within schools.

Table 3: Performance comparison of large models in educational scenarios

Model	Parameter quantity	MATH-500 score	Single inference cost	Open source license
DeepSeek-R1-Distill-Qwen-7B	7B	89.7%	0.002 dollars	MIT
Llama2-13B	13B	85.2%	0.008 dollars	Business restrictions
ChatGLM3-6B	6B	77.5%	0.005 dollars	Apache

LAM	130B	92.1%	3.0 dollars	Closed source
-----	------	-------	-------------	------------------

3.3.2 Data processing and collaboration capability assessment process

The data processing and evaluation architecture of this system is closely centered around the core requirements of the educational collaboration scenario. Through hierarchical analysis strategies and adaptive verification mechanisms, a complete technical chain from the original dialogue to the ability portrait is constructed. In the data processing stage, a semantic-enhanced cleaning scheme is adopted. Aiming at the characteristics of semi-structured data generated by instant messaging tools, multi-level standardization processing is carried out: Firstly, non-text interfering elements are stripped through the regular expression engine to retain the substantive content of the conversation; Then, the context association algorithm is used to reconstruct the conversation context and integrate discrete messages into discussion units with logical coherence. Specifically targeting the common cross-temporal dialogue features in educational scenarios.

The design of the dynamic assessment engine embodies the concept of process education evaluation and adopts a two-stage progressive analysis strategy. In the first stage, based on the predefined collaboration dimension framework (COLLAB_CRITERIA), the system conducts preliminary feature extraction of the dialogue content through a large language model and generates an initial score including basic indicators such as frequency statistics and response intervals. The second stage focuses on semantic deep mining and utilizes the model's contextual understanding ability to identify implicit collaborative behaviors. For instance, for messages containing guiding statements, the system activates the coordination ability evaluation submodule and verifies the actual influence of the suggestion in combination with the subsequent discussed evolution path. To ensure the robustness of the evaluation results, the system is equipped with a multi-layer verification mechanism. When there are logical contradictions in the model output (such as marking both "high participation" and "low response speed" simultaneously), the temperature coefficient progressive adjustment strategy is automatically triggered (the temperature parameter gradually increases from 0.3 to 0.7). Guide the model to re-examine the context association by changing the deterministic level of the model output.

At the result processing level, the system builds a standardized output pipeline. The original evaluation results were strictly formatted and verified. The key operations performed included:

unified character encoding conversion (to solve the problem of mixed use of Chinese and English punctuation), JSON structure integrity verification, and mandatory normalization of the score value range (to ensure compliance between 0-5 partitions). Through the `save_results` function, the system persistently stores the verified evaluation data indexed by timestamps. Its data structure strictly follows the schema version (`SCHEMA_VERSION`) definition and contains quantitative scores, qualitative feedback, and information of associated participants in five core dimensions. This design not only supports the vertical tracking of individual capabilities, but also provides a structured data basis for the comparative study of group collaboration patterns.

3.3.3 Privacy protection and permission management system

This system has established a multi-layer protection architecture, covering the entire life cycle of data collection, storage, processing and output, to ensure the privacy of educational data and the security of system operations. At the identity authentication level, a hierarchical encryption strategy is adopted - the user password is stored as an irreversible hash value generated through the SHA-256 algorithm. Even if the database is leaked, the original password cannot be restored. The role permission system is dynamically implemented through database constraints: when creating a group, the system automatically binds the affiliation between the creator and the group, eliminating the possibility of unauthorized operations at the structural level.

The hierarchical encryption strategy is implemented in the data storage link. Core sensitive fields such as user passwords are protected by high-strength hash algorithms, while behavioral data is processed and stored as JSON analysis results through field-level encryption technology. The database connection layer forcibly enables foreign key constraints to ensure the logical consistency of users, groups, and mapping relationships.

3.3.4 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;

AI-Enhanced Student Skills Development Tracker

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.

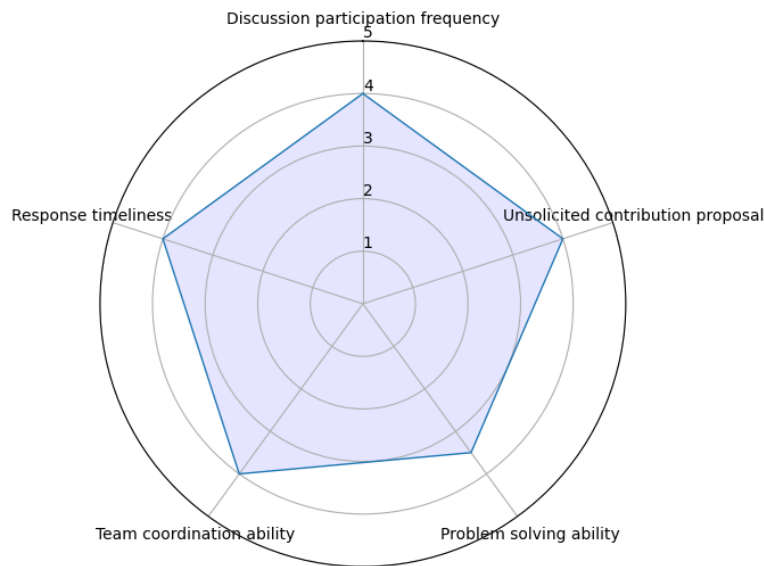


Figure 3 Capability assessment radar chart

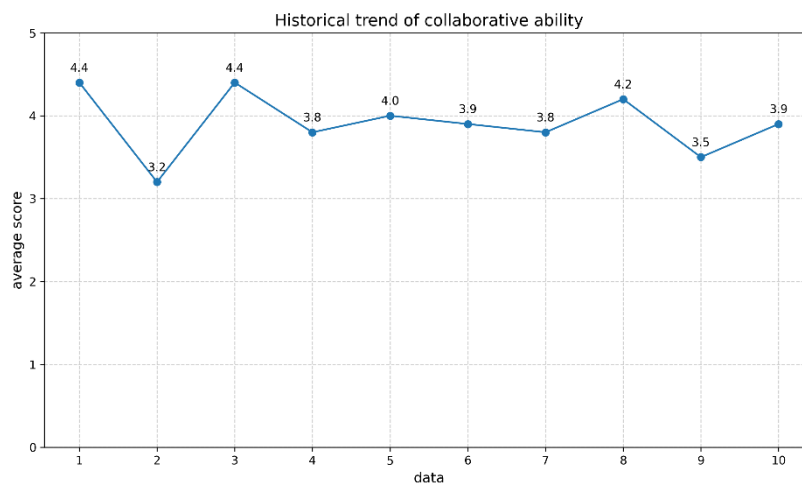


Figure 4 Capability assessment line chart

4.3 System reliability analysis

To verify the stability of the system in repeated tests, this study designed a reliability verification experiment. Through multiple analysis tasks under the same conditions, the consistency of the system's output results was evaluated. The experiment selected the chat

records of the course group containing 6 students as input data. The system analysis module was repeatedly run 10 times under the same environment (hardware configuration, model parameters, data preprocessing process). The score data of each skill dimension were recorded and the stability line graph was drawn.

4.3.1 Test design

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

Data input: Select historical chat records containing 200 text messages, covering collaboration scenarios such as group discussions, task allocation, and problem-solving.

Test indicators: Track the scores of the five core dimensions of "participation", "initiative", "problem-solving ability", "coordination ability", and "response timeliness" (0-5 points).

Repetition count: Run the analysis module independently 10 times, using exactly the same test process each time to eliminate interfering factors.

4.3.2 Result analysis

The line chart shows the score fluctuations of each skill dimension in 10 repeated tests. The results show that the scores of all dimensions fluctuated slightly within a reasonable range, with no significant deviation.

The statistical results show that the average variance of the intra-group scores of each dimension is approximately 0.1, proving that the analysis results of the system under the same conditions are highly consistent.

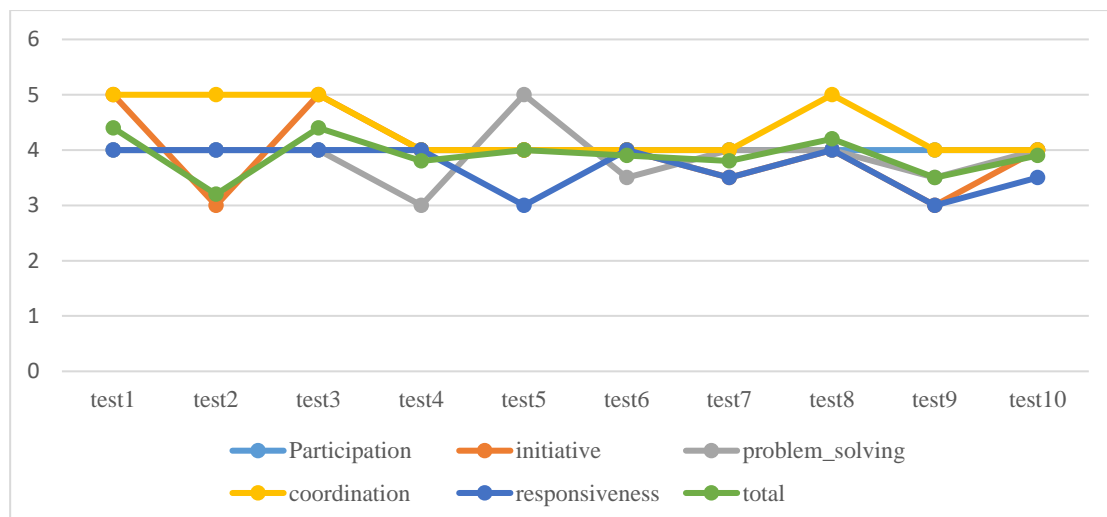


Figure 4 Line graph of system score distribution

4.3.3 Format for equations

Through repeated tests and verification, the analysis results of the system under the same input conditions show stable consistency. The fluctuations in the scores of each skill dimension are all controlled within a certain range, meeting the basic requirements for the reliability of assessment tools in educational scenarios. This indicates that the system can provide repeatable skill analysis results and has the basic usability to support the long-term tracking of students' skill development.

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.

AI-Enhanced Student Skills Development Tracker

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

References

Book Chapters:

[1] Hmelo-Silver CE, Chinn CA, Chan C, O'Donnell AM. Collaborative learning and skill development. In: *The International Handbook of Collaborative Learning*. New York, USA: Routledge; 2013. p. 123-156.

Journal Articles:

[2] Wang YY, Zheng YH. Technological pathways and educational applications of multimodal learning analysis [Chinese]. *China Educ Technol*. 2022;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>.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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:

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

[11] 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>.

[12] 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>.

[13] 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:

A handwritten signature in black ink, appearing to be the Chinese characters '管欣' (Guan Xin) written in a cursive style.

Date: 07-04-2025

Project specification

北京邮电大学 本科毕业设计（论文）任务书

Project Specification Form

Part 2 - Studentc

学院 School	International School	专业 Programme	e-Commerce Engineering with Law		
姓 Family name	Guan	名 First Name	Xin		
BUPT 学号 BUPT number	2021213030	QM 学号 QM number	210981128	班级 Class	2021215116
论文题目 Project Title	AI-Enhanced Student Skills Development Tracker				

<p>论文概述</p> <p>Project outline</p> <p>Write about 500-800 words</p> <p>Please refer to Project Student Handbook section 3.2</p>	<p>1. Introduction</p> <p>The AI-Enhanced Student Skills Development Tracker is designed to monitor and improve key skills in students, such as critical thinking, problem-solving, and collaboration. This project aims to create an AI-driven system that tracks the development of these skills over time. By analysing data from assignments, assessments, and classroom activities, the system will provide insights into each student's progress in specific skill areas. Educators will receive detailed reports that highlight strengths and areas for improvement, while students will benefit from personalised recommendations that guide their skill development. The system will be designed to adapt to each student's unique learning journey, offering targeted resources and activities to help them grow. The goal is to support educators in fostering well-rounded skill development and to empower students to take control of their learning.</p> <p>2. User Requirements</p> <p>Students</p> <p>Students will interact with the system mainly by uploading or linking their performance data, such as assignment scores, class participation records, and self-assessments. The system will analyse this data to measure skill levels in areas like critical thinking, problem-solving, and collaboration. Based on these analyses, the system will generate detailed reports and offer personalized feedback. Students will receive tailored recommendations and notifications about areas that need improvement, along with suggested resources or activities to enhance their skills [1].</p> <p>Educators</p> <p>Educators will have the option to view summaries of student skill assessments. They can also provide additional qualitative input, which the system will integrate into the analysis, offering a more holistic view of</p>
---	--

	<p>student progress. Educators can then review these insights and use them to adjust their teaching strategies.</p> <p>3. Experiment and research</p> <p>AI Model Performance Testing</p> <p>To ensure the skill tracker is both accurate and efficient, the AI models will undergo rigorous testing:</p> <p>Model Comparison: Different AI algorithms will be compared to find the most effective model for skill assessment. Performance will be measured in terms of accuracy, speed, and robustness.</p> <p>Skill Categorization: Validate the system's ability to categorize and evaluate different types of skills correctly. This will involve testing the model with a diverse set of input data and refining algorithms based on results.</p> <p>Accuracy Testing:</p> <p>Validate the accuracy of the AI models in evaluating student skills and generating appropriate recommendations.</p> <p>This module focuses on assessing the consistency between the AI's analysis and human evaluations. By comparing AI-generated insights with teacher assessments, we can identify areas where model adjustments may be required [2].</p> <p>Usability Testing:</p> <p>Ensure that the system is intuitive and easy for both educators and students to use.</p> <p>This module gathers user feedback on system usability. Through regular feedback loops, insights will be collected on interface clarity, ease of navigation, and understanding of feedback, which will be used to improve the user experience [3].</p>
--	---

	<p>4. Data Collections</p> <p>Data will be gathered from:</p> <p>Assignments and Tests: Used to evaluate critical thinking and problem-solving.</p> <p>Class Participation: Includes classroom activities like group work, discussions, and presentations to assess collaboration and communication skills.</p> <p>Feedback: Peer and teacher feedback on teamwork, communication, and other skills.</p> <p>Self-Assessment: Students will periodically evaluate their own skills.</p> <p>5. Tools and Technologies</p> <p>Programming Languages: Python for AI and backend development.</p> <p>JavaScript (React.js) for the frontend interface.</p> <p>AI Technologies: Use machine learning frameworks like TensorFlow or PyTorch to build and train models.</p> <p>Hardware:</p> <p>Cloud-Based Infrastructure: Services like AWS or Google Cloud to support scalable data processing and storage.</p> <p>GPU Service: Utilizing cloud-based GPU resources to accelerate AI model training and inference, making the system efficient and capable of handling large datasets.</p> <p>6. Expected Outcomes</p> <p>Comprehensive Skill Tracking: The system will offer real-time monitoring and analysis of student skill development, giving both students and educators valuable insights.</p> <p>Personalized Learning: Students will receive specific recommendations, making their learning experience more effective and focused.</p> <p>Support for Educators: Teachers will have a powerful tool to monitor and enhance student performance through data-driven insights.</p>
--	--

	<p>7. Conclusion</p> <p>The AI-Enhanced Student Skills Development Tracker is designed to provide a personalized and data-driven approach to education. By leveraging AI to measure and analyse student skills, the system will offer valuable insights to enhance the learning experience. The project will be executed in clear, structured phases, with opportunities for feedback and improvement at each stage to ensure it is practical, user-friendly, and impactful.</p> <p>[1] Zheng, H. Z. (2024). <i>Research and application of a learning resource recommendation model based on feedback information</i> (Master's thesis, Yunnan Normal University). Master's Thesis. https://link.cnki.net/doi/10.27459/d.cnki.gynfc.2024.001092</p> <p>[2] Zhang, H. C. (2024). Case analysis of AI-assisted autonomous learning in courses. <i>Electronic Technology</i>, 06, 302-303.</p> <p>[3] Zhang, H. Y., Huang, R., Li, Y., & He, J. G. (2024). Evaluation of AI-assisted English learning tools. <i>Computer-Assisted Foreign Language Education</i>, 02, 18-24, 103. https://doi.org/10.20139/j.issn.1001-5795.20240203</p>
道德规范 Ethics	<p>Please confirm by checking the box:</p> <p><input checked="" type="checkbox"/> I confirm that I have discussed ethical issues with my supervisor.</p>

<p>Please discuss ethical issues with your supervisor.</p> <p>Please refer to Project Student Handbook section 4.1</p>	<p>Summary of ethical issues:</p> <p>1. Data Privacy and Security:</p> <p>The system will use sensitive student data, including performance records, feedback, and possibly personal information. Ensuring the privacy and security of this data is crucial.</p> <p>2. Consent and Transparency:</p> <p>Students and educators should be fully informed about the data being collected and how it will be used.</p> <p>Obtaining informed consent from students and educators before data collection.</p> <p>3. Bias and Fairness:</p> <p>AI models can unintentionally reflect or amplify biases present in the training data, leading to unfair or unequal outcomes.</p> <p>Ensuring that recommendations and assessments do not unfairly disadvantage any student group.</p> <p>4. Impact on Learning:</p> <p>The system's feedback and recommendations should enhance learning without causing undue stress or negatively impacting students' self-esteem.</p> <p>Ensuring that the recommendations support a positive learning experience and encourage skill development constructively.</p>
--	--

<p>中期目标</p> <p>Mid-term target.</p> <p>It must be tangible outcomes, E.g. software, hardware or simulation.</p> <p>It will be assessed at the mid-term oral.</p>	<ol style="list-style-type: none"> 1. Basic Software Prototype A simple version of the software that allows data input and generates basic skill assessment reports using sample data. Integration of a pre-trained large model (e.g., GPT or Llama) to demonstrate initial natural language processing capabilities for analyzing text-based assignments or feedback. 2. Cloud and GPU Configuration Enabling GPU support to optimize the performance of the large model during training and analysis. 3. Data Processing and AI Model Setup Fine-tuning a pre-trained large model on sample educational data to provide early insights into skill development. 4. Simulation Demo A basic simulation to show how the system uses the large model to process and analyse data, then outputs assessment reports.
--	--

AI-Enhanced Student Skills Development Tracker

Work Plan (Gantt Chart)

Fill in the sub-tasks and insert a letter X in the cells to show the extent of each task

	Nov 1-15	Nov 16-30	Dec 1-15	Dec 16-31	Jan 1-15	Jan 16-31	Feb 1-15	Feb 16-28	Mar 1-15	Mar 16-31	Apr 1-15	Apr 16-30
Task 1 Skill Identification: Identify key skills to be tracked and develop metrics for assessing student progress in each area.												
Conduct research to identify essential skills, such as critical thinking, problem-solving, and collaboration, relevant to student success.	X	X	X									
Define clear, measurable metrics for assessing each identified skill, ensuring they are easy to track and understand.		X	X									
Develop a framework to categorize skills and associate them with specific data types.		X	X									
Task 2 AI Integration: Integrate AI algorithms to analyse student performance data and track skill development over time.												
Select the appropriate pre-trained AI model (e.g. GPT, BERT), set up the environment, and conduct preliminary functional testing.			X	X								
Find and use existing open source sample education data to fine-tune the model. Clean and preprocess the sample data, such as removing irrelevant information, normalizing text, and structuring data to be compatible with the AI models.				X	X	X						

AI-Enhanced Student Skills Development Tracker

Implement and test initial AI algorithms for analyzing student performance data, focusing on accuracy and reliability.				X	X	X	X					
Test the model to evaluate its accuracy and efficiency. Optimize the model based on specific performance metrics.					X	X	X	X	X			
Task 3 Recommendation Engine: Build a recommendation engine that suggests activities and resources to enhance specific skills.												
The ability to store information about learning resources and users.					X	X	X	X	X			
Implement a simple rule-based recommendation engine that suggests resources based on identified skill gaps.						X	X	X	X	X		
In the process of interacting with users, improve the suggestions provided by the system based on user feedback						X	X	X	X	X		
Task 4 Reporting System: Design a reporting system that provides detailed feedback to students and educators on skill progression.												
Develop the structure of the reports, outlining key metrics such as skill progression, strengths, and areas for improvement. Design a simple and user-friendly report.						X	X	X	X	X		
Implement simple data visualization features to make progress reports intuitive and easy to interpret.										X	X	X
Implement the function of allowing users to browse reports anytime and anywhere.								X	X	X	X	X

AI-Enhanced Student Skills Development Tracker

The system should be able to automatically compile the generated results into downloadable tables.								X	X	X	X	X
--	--	--	--	--	--	--	--	---	---	---	---	---

Early-term progress report

北京邮电大学 本科毕业设计（论文）初期进度报告

Project Early-term Progress Report

学院 School	International School	专业 Programme	e-Commerce Engineering with Law		
姓 Family name	Guan	名 First Name	Xin		
BUPT 学号 BUPT number	2021213030	QM 学号 QM number	210981128	班级 Class	2021215116
论文题目 Project Title	AI-Enhanced Student Skills Development Tracker				

1. Introduction

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. 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 [1].

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 [2]; 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 [3].

3. 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.

Student behavior can be monitored through AI where it can gather data such as grades of the assignments, students' interactions in classroom, online learning activities, etc., and the data collected can be processed in real-time by the learning analytics system. This data enables the AI model to gain a clear understanding of the performance of the student in different skills domains. For instance, AI systems can detect which aspects of critical thinking a learner is good at and which aspects he/she is not good at, and suggest appropriate materials and exercises to develop the aspect in question. Learning paths can not only engage students in learning, but also enhance specific learning skills at one's own pace [4].

One of the most widely used areas of applying AI technology in education is the adaptive learning and feedback provision. Students are offered individual learning pathways and feedback using learning analytics and machine learning that analyse student behavioral data, academic performance, and learning process in real time. The intelligent recommendation system can recommend learning resources to the students based on their preferences, aptitude, and learning achievement. For instance, intelligent recommender systems can suggest to students the best learning content they could possibly engage in based on their past performance and preferences. Moreover, AI can also adapt the learning content and frequency when the students face learning challenges to ensure that students can learn best way possible.

Sajja et al. (2023) [5] 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) [6] 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) [7] 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.

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) [8] 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.

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) [9], 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.

4. Ethical issues of Artificial Intelligence in Education

While AI has proven to be effective in education, its use has also raised numerous issues of ethical consideration especially in personalized learning environments and in automated testing. Olaf et al. (2019) [5] analyzed the problematic of ethical concerns in the use of AI and how to address the ethical concerns raised by the use of artificial intelligence in education and mitigate algorithmic bias in education. It is also important for AI systems in education not to focus too much on data, which would not pay attention to the need of each learner, and make sure that every learner will be given the right support and fair evaluation.

Another ethical concern that arises when using AI is the question of academic integrity. Due to the availability of AI tools like ChatGPT, the issue of how students can be prevented from cheating, and the ways to develop prevention strategies to make education fair has become a research question. Kasneci et al. (2023) [11] stress that AI based education solutions should include anti cheating methods to make the education process fair.

5. Conclusion

AI provides new possibilities for the education field, which allows monitoring and evaluating the student's progress in the process of acquiring important competencies in real time. Though some issues are still open, for example, data protection, choosing of the right algorithms, and models' effectiveness, as the AI technology progresses, it will be able to help learners achieve more significant progress in developing multidimensional skills.

References

- [1] 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>.
- [2] Naveed H, Humza et al. A Comprehensive Overview of Large Language Models. *ArXiv*. 2023 Jul 1. Available from: <https://arxiv.org/abs/2307.06435>.
- [3] 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>.

- [4] 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>.
- [5] 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>.
- [6] Crompton H, Song D. The Potential of Artificial Intelligence in Higher Education. Revista Virtual Universidad Católica del Norte. 2021; n. pag.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] Kasneci E, et al. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. Learning and Individual Differences. 2023; n. pag.

Completed task

The Llama 8B model is planned to be selected as the primary pre-trained model. Currently, a Linux virtual environment is running on the Windows system for development and testing purposes. The next step is to migrate to a cloud server based on the Linux system to handle large-scale computational tasks efficiently. The Llama model has been deployed, and all required libraries, including Hugging Face Transformers and PyTorch, have been installed. Version management is handled using Conda.

For the skill identification section, Further work is needed to further define the core skills and their assessment metrics beyond what is available, as well as to establish a corresponding classification framework to ensure that subsequent data processing and model fine-tuning align effectively with educational objectives.

是否符合进度？ On schedule as per GANTT chart?

YES

下一步 Next steps:

Integrate AI algorithms to analyse student performance data and track skill development over time.

Mid-term progress report

北京邮电大学 本科毕业设计（论文）中期进度报告

Project Mid-term Progress Report

学院 School	International School	专业 Programme	e-Commerce Engineering with Law		
姓 Family name	Guan	名 First Name	Xin		
BUPT 学号 BUPT number	2021213030	QM 学号 QM number	210981128	班级 Class	2021215116
论文题目 Project Title	AI-Enhanced Student Skills Development Tracker				
<p>是否完成任务书中所定的中期目标？Targets met (as set in the Specification)?</p> <p>NO –The section on fine-tuning models using datasets is not complete. Reason –There is some misunderstanding about the project in the proposal stage. At current stage of the mission, it is more important to build a reasonably system than to train fine-tuning models. The NLP required at this stage can be used directly by native models and API. Data set training models will begin as soon as the basic build of the system is complete.</p>					
<p>已完成工作 Finished work:</p> <p>1. Theoretical Justification for Collaboration Tracking</p> <p>At the early stage of the project, after many discussions with the tutor, we made clear the design direction of taking collaboration ability as the core evaluation index. The selection was based on the following consensus:</p> <p>1.1 Practical Limitations of Traditional Assessment</p> <p>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:</p> <p>First, the "silent contributor" value gap (Hmelo-Silver et al., 2013). 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" (Interactional Research Into</p>					

Problem-Based Learning, 2022).

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.

1.2 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).

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. Core elements of effective collaboration

In collaborative learning scenarios, the characteristics of effective collaborators can be systematically defined and evaluated by multidimensional indicators. Based on literature analysis, the following three dimensions are discussed, and the measurement path is illustrated in combination with specific quantitative methods.

(1) Engagement

Definition: The ability of collaborators to drive the achievement of team goals through substantial contributions (e.g., task division, idea generation).

Key Indicator: Task contribution rate

Data Sources: Version control systems (Git), task progress reports, member file upload records, task assignment tables, teacher grading sheets

Measurement: Proportion and diversity of individual workload in collaborative tasks (e.g., code submissions, document writing, data analysis, etc.)

(2) Communication Efficacy

Definition: The clarity and bidirectionality of information transfer.

Key Indicator: Information entropy value

Data Sources: Group discussion records (e.g., WeChat chat logs)

Measurement: Analysis using natural language processing (NLP), outputting positive/neutral/negative scores

(3) Accountability

Definition: The fulfillment of team commitments and respect for others' contributions, providing specific, actionable, and emotionally positive peer feedback.

Key Indicator: Specificity of feedback

Data Sources: Group discussion records (e.g., WeChat chat logs), peer evaluation sheets

Measurement: Use natural language processing (NLP) to analyze whether the feedback contains specific examples or step-by-step suggestions, outputting positive/neutral/negative scores.

3. Recommendation Engine Design

Based on the above indicators, the system design can be modular corresponding to ensure the "indicators - data - function" closed loop:

(1) Data Collection

The first step in the system is to collect data input. The system collects data through submitted jobs or other forms of data (such as text, job results, etc.), and the data collection module is responsible for integrating these inputs into the system after simple processing, and passing them to the subsequent data processing stage.

(2) AI Analysis & Skill Map Generation

The cleaned data is fed into an AI Analysis Engine, which uses algorithms to score students on different dimensions of collaboration.

(3) Feedback generation and Smart Feedback System

The system will generate personalized feedback based on the analysis results. This feedback can include specific suggestions for improvement, recommendations for learning resources, and more. The system distributes feedback to different user interfaces

(4) Data Storage

All data and feedback are stored in a database

New Cycle & Feedback

Once feedback is generated and distributed, the system enters a new evaluation cycle. New data and feedback will further refine the students' skill map, forming a closed loop of continuous improvement. This enables the recommendation system to track the development of students' skills.

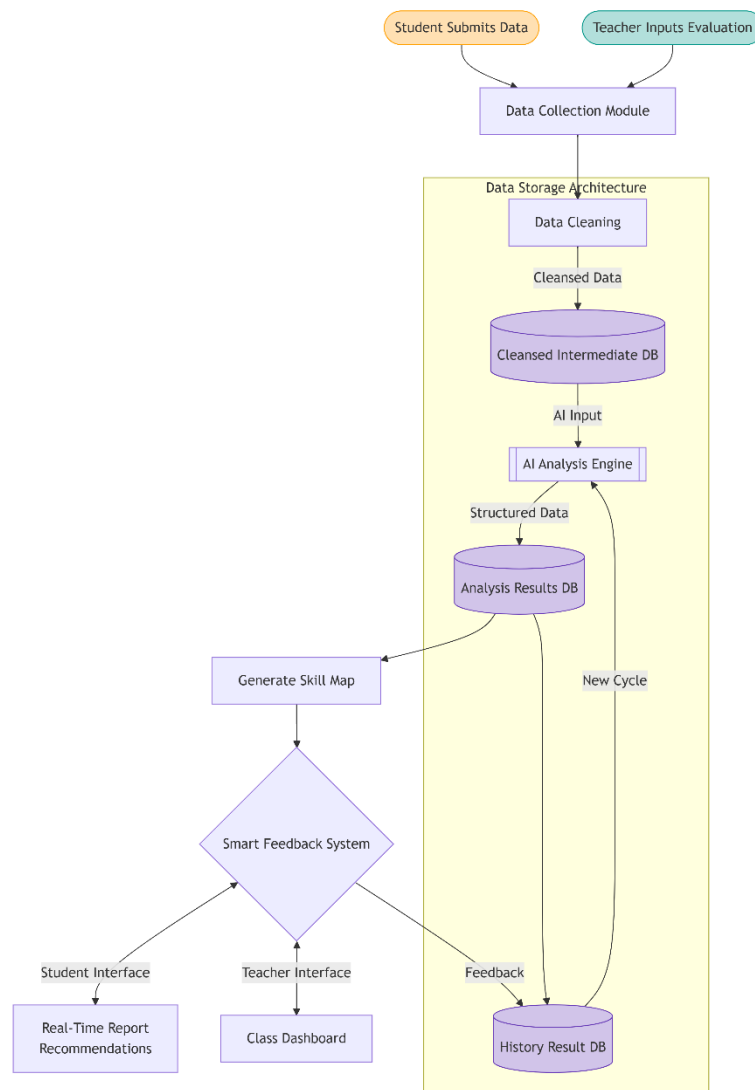


Figure 1 Architecture Diagram

4. Prototype Design

Based on a collaborative capability assessment framework determined above, a prototype system has been build based on NLP, which can automate the entire process of data acquisition, AI analysis, and feedback generation.

The DeepSeek-R1 has been selected for its superior understanding and cost-efficiency in educational contexts. Its distilled version (DeepSeek-R1-Distill-Qwen-7B) has been deployed on Alibaba Cloud. Future work includes fine-tuning this model using annotated collaboration data via LoRA to enhance behavior classification accuracy, targeting 90%+ precision for critical metrics like conflict resolution detection.

In the current phase of the project, the main goal is to design a functional prototype that can be used to achieve a simple collaboration capability assessment based on student participation in academic activities. The system is able to score from 0 to 5 on different dimensions based on pre-designed input text data and provide insights and feedback.

```
-- Summary of communication effectiveness evaluation --

Participant: Zhang SAN
Overall score: 4/5
Detailed feedback:
- Excellent clarity of information presentation (clear time, agenda)
- Effective coordination of multiple resources (technical department attended)
- Suggestions: Can increase the frequency of response to suggestions of other members
```

Figure 2 Simulation Demo show

The system code has been put on GitHub repository:

<https://github.com/bnd1970/Final-Year-Project-AI-Enhanced-Student-Skills-Development-Tracker>

References

- [1] Hmelo-Silver, C. E., Chinn, C. A., Chan, C., & O'Donnell, A. M. (2013). The International Handbook of Collaborative Learning. Routledge.
- [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]. <i>China Educational Technology</i> , 41(3), 45–53.
<p>尚需完成的任务 Work to do:</p> <ol style="list-style-type: none"> 1. Core task: Complete a fully functional circular recommendation system 2. Ability Map: Correlate collaborative ability data with academic achievement to build a visual growth path for students. 3. Adaptive recommendation engine: track and record students' growth, and automatically push resources according to weaknesses (such as weak conflict resolution ability → recommend to strengthen learning in this field).
<p>存在问题 Problems:</p> <ol style="list-style-type: none"> 1. Functional limitations of the system: the current system only realizes the feedback ability for the text quality of group assignments, and the current system has not realized long-term tracking 2. Data diversity Limitation: The existing system features only cover text document reading and writing 3. Lack of user interaction: The current system has not implemented user-friendly data visualization 4. Hardware limitations: System response time is slow
<p>拟采取的办法 Solutions:</p> <ol style="list-style-type: none"> 1. Function expansion: multi-modal data access to achieve comprehensive analysis of different types of data in group work; Design historical track database to realize long-term tracking 2. Report readability optimization: Design a feedback template engine to provide visual analysis results and reports 3. Cloud migration: Deploy the system to the cloud
论文结构 Structure of the final report: (Chapter headings and section sub headings)

<i>Abstract</i>	2
Keywords	2
Chapter 1: Introduction	3
1.1 Project Motivation.....	3
1.2 Objectives.....	3
Chapter 2: Background	4
2.1 Collaboration Metrics in Education	5
2.2 LLMs in Skill Assessment	5
Chapter 3: Design and Implementation.....	4
3.1 System Architecture	5
3.2 Model Training Workflow	5
3.3 Frontend-Backend Integration	5
Chapter 4: Results and Discussion.....	4
4.1 Simulation Outcomes	5
4.2 Model Accuracy Analysis	5
4.3 User Feedback Summary	5
Chapter 5: Conclusion and Further Work	4
5.1 Key Contributions	5
5.2 Limitations	5
5.3 Future Directions.....	5
References	4
Acknowledgement	4
Appendices.....	4

Supervision log

北京邮电大学 本科毕业设计（论文）教师指导记录表

Project Supervision Log

学院 School	International School	专业 Programme	e-Commerce Engineering with Law		
姓 Family name	Guan	名 First Name	Xin		
BUPT 学号 BUPT number	2021213030	QM 学号 QM number	210981128	班级 Class	2021215116
论文题目 Project Title	AI-Enhanced Student Skills Development Tracker				
<p>Please record supervision log using the format below:</p> <p>Date: dd-mm-yyyy Supervision type: face-to-face meeting/online meeting/email/messages/other (please specify) Summary:</p>					
<p>Date: 11-11-2024 Supervision type: face-to-face meeting Summary: Discuss and confirm initial phase goals.</p> <p>Date: 14-11-2024 Supervision type: email Summary: discuss how to write the report, received written feedback on the draft specification.</p> <p>Date: 10-1-2025 Supervision type: email Summary: discuss how to write the report, received the feedback about the report.</p> <p>Date: 14-2-2025 Supervision type: email Summary: discuss the feedback from the early-term progress report, decided to hold an online meeting for further discussion.</p> <p>Date: 17-2-2025 Supervision type: online meeting Summary: discuss the feedback from the early-term progress report and determine the next major tasks.</p> <p>Date: 20-2-2025 Supervision type: messages Summary: Report feedback and improvements made from the previous meeting.</p> <p>Date: 22-2-2025</p>					

Supervision type: messages

Summary: discussed how to write the mid-term progress report.

Date: 23-2-2025

Supervision type: messages

Summary: Report the revised mid-term progress report.

Date: 13-3-2025

Supervision online meeting

Summary: Feedback of the mid-term progress report.

Date: 18-3-2025

Supervision type: messages

Summary: Work progress report.

Date: 20-3-2025

Supervision type: messages

Summary: Work progress report.

Date: 21-3-2025

Supervision type: messages

Summary: Work progress report.

Date: 25-3-2025

Supervision type: messages

Summary: Work progress report.

Date: 26-3-2025

Supervision type: messages

Summary: Work progress report.

Date: 5-4-2025

Supervision type: messages

Summary: Work progress report.

Date: 10-4-2025

Supervision type: messages

Summary: Work progress report.

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.