Intelligent Tutoring Systeam

1st Yongru Qian
School of Computer Science
University College Dublin
Dublin, Ireland
yongru.qian@ucdconnect.ie

2nd Xiaochi Li
School of Computer Science
University College Dublin
Dublin, Ireland
xiaochi.li@ucdconnect.ie

3rd Alvaro Mauricio Patty Cruz
School of Computer Science
University College Dublin
Dublin, Ireland
alvaro.patty@ucdconnect.ie

4th Jiahao Liu
School of Computer Science
University College Dublin
Dublin, Ireland
jiahao.liu@ucdconnect.ie

5th Chih-Yu Huang
School of Computer Science
University College Dublin
Dublin, Ireland
CHIH-YU.HUANG@ucdconnect.ie

6th Weimin Sun
School of Computer Science
University College Dublin
Dublin, Ireland
weimin.sun@ucdconnect.ie

***Abstract*—In the current era, the rapid evolution of artificial intelligence has opened doors to combining education with cutting-edge AI capabilities. However, prominent online learning platforms like Coursera and Udemy have not yet tapped into this synergy. Our project aims to implement an intelligent tutoring system that leverages AI innovations such as ChatGPT to enhance the online learning experience and improve student performance. To achieve this, we use the OpenAI interface to interact with ChatGPT, employing a suite of frameworks including React, Spring Boot, and Flask for a seamless front-end and back-end service delivery.On the infrastructure front, we rely on the robust capabilities of AWS, Kubernetes, Docker, and Jenkins to facilitate continuous integration and continuous deployment (CICD). To assess the effectiveness of our system, we use a combination of A/B tests, tree tests, and questionnaires. These methods are complemented by stress tests and monitoring mechanisms to ensure reliability. While our system adeptly fulfills the majority of both functional and non-functional requirements, t-test results indicate that its impact on significantly enhancing student learning outcomes remains inconclusive.**

***Index Terms*—intelligent, ITS, chatGPT, openAI, learning**

I. Introduction

Education is fundamental to the socio-economic progress of societies and pivotal in alleviating poverty. Yet, millions globally remain devoid of this privilege. The United Nations Educational, Scientific and Cultural Organization (UNESCO) reports that by 2030, approximately 225 million children, adolescents, and youth will not be in school, while 750 million adults worldwide lack basic literacym [1]. These statistics emphasize the need for innovative interventions to close the educational divide.

In the digital era, a significant challenge that the global educational arena grapples with is the conspicuous disparity in educational resources and opportunities. Since the early 20th century, Intelligent Tutoring Systems (ITS), sophisticated platforms that offer personalized guidance and feedback to learners without the need for human intervention, have emerged as a potentially transformative solution to bridge this disparity [2].

The advent of ITS has revolutionized the pedagogical landscape, enabling educators to disseminate educational content online, engage interactively with students, and evaluate learners' data [3]. This paradigm shift ensures a remarkable reduction of educational inequities, granting students, regardless of geographical constraints or national boundaries, unfettered access to consistent educational resources via ITS platforms.

Recent advancements in machine learning and data analytics have greatly enhanced the capabilities of ITS, enabling a more personalized approach to instruction. These systems excel in autonomously analyzing students' academic progress and offering customized recommendations based on individual learning styles and knowledge levels. By evaluating a student's learning trajectory, these systems can provide targeted suggestions for improvement while considering their preferences and proficiency [4].

The introduction of models like OpenAI's GPT has demonstrated exceptional performance in tasks such as text generation, summarization, and question answering. Integrating GPT into ITS has the potential to significantly enhance the quality of interactions and personalization within the system. GPT's extensive training on diverse datasets equips it to fulfill precise personalization requirements [5]. Our primary challenge is that, by configuring specific attributes and metrics, ITS can offer a more detailed analysis of learners' behaviors and preferences, and thus utilize the GPT model to provide tailored feedback [6]. This amalgamation of advanced machine learning techniques and state-of-the-art NLP capabilities holds the promise of significantly elevating the overall quality of the learning journey.

The primary objective of this system is to establish a scalable linguistic learning intelligence platform. As it dispenses pertinent educational resources, the system meticulously constructs learning profiles based on the attribute of "learning styles" [7]. By integrating the GPT model, it offers an ITS that not only supports multilingualism and multimedia but also boasts personalized adaptive feedback and interactive mechanisms. Central to the platform's goals is the aspiration to cultivate a more inclusive educational environment. It endeavors to cater to the unique requirements of every learner, guaranteeing a tailor-made, efficacious learning journey for all.

The rest of the paper is structured as follows:

- **Section II:** Discussion on user scenarios, highlighting the role of Intelligent Tutoring Systems (ITS) in offering personalized and adaptive learning.
- **Section III:** A review of prior work on ITS, encompassing applications in education and impacts on individual learners.
- **Section IV:** Describes the implementation details of the project, including the full technical realization from development to deployment.
- **Section V:** Describes in detail the evaluation methodology and results of the project, including system performance evaluation and user evaluation.
- **Section VI:** Summarizes the successes and shortcomings of the project and plans the direction of future work.

II. User Scenarios

Tom, a college student with a keen interest in enhancing his English proficiency, aims to bolster his language skills for confident communication in international exchanges. However, during his studies, he often encounters questions that remain unanswered in real-time. Following a grammar practice session, he reviewed the answers, yet found himself puzzled by the rationale behind two incorrect responses. Unable to locate identical queries online, he had to wait for his English class three days later to seek clarification from his teacher. This disruption hampered his learning progress significantly.

As a user, Tom desires a feature to review his errors in tests, enabling him to grasp the underlying reasons for these mistakes.

Lily, a middle-aged woman, shares the aspiration of seamless communication during her travels abroad. Similarly, Mrs. Zhang, also middle-aged, seeks fluency in foreign conversations. However, they both face challenges in initiating their English journeys. Often, they invest days in online study materials only to realize the content's excessive difficulty, mismatched to their current level. This process not only squanders their time but also erodes their confidence.

A common thread among these users is their inability to afford private tutors to aid them in learning English.

Taking these user needs into account, our AI Tutoring System offers free access. Users can engage in a self-assessment test before embarking on their learning journey. The AI system subsequently tailors suitable study materials based on their self-assessment results. Additionally, each learning unit concludes with an evaluation test to gauge progress. Post-assessment, users can consult the AI for clarifications on any queries, with the AI providing pertinent explanations based on user responses.

III. Related Work

A. Related ITS Work

Abdolhossein and his team employed facial expression analysis to decode emotions related to learning, intervening, and engaging students through dynamic emotional avatars [8]. Wang introduced a unique learning map derived from blogs, streamlining students' searches for blog-related content and guiding them towards focused learning maps, thereby promoting rapid and meaningful learning [9]. In 2013, Tanner and Danielle integrated gaming elements into the iSTART system, enhancing students' motivation to learn. They harnessed the successful characteristics of the ITS for extended interactions, augmenting student engagement and enthusiasm [10]. Oscar's CITS, enhanced with AI capabilities, caters to individual learning styles that align with learners' existing preferences. This system provides cues and suggestions, encouraging learners to independently address challenges and offering tailored feedback when errors occur [11]. Jason utilized an advanced facial expression detection system (like FaceReader 5.0), combined with self-reporting tools (such as emotional value scales), and skin conductivity measurements (using Affectivas' Q-2.0 Sensor) to concurrently assess students' emotional responses during learning [12]. Certain researchers delved into conceptualizing and constructing an ITS focused on problem-solving techniques, expanding the existing ITS framework to create the iTutor [13]. David presented an approach centered around mathematical challenges, with the system identifying students' problem-solving approaches and providing adaptive responses based on the student's decisions and the problem's constraints [14].

B. Technical Challenges in Implementing ITS

The implementation of ITS brings the promise of a more personalized and effective learning experience, but it also introduces several technical challenges. These issues range from the integration of different technologies to ensuring the security of student data. Despite the benefits of ITS, it is essential to address these challenges to achieve successful implementation.

The integration and standardization of various technologies are two of the foremost challenges in the development of ITS. The need for seamless technology integration stems from the diversity of tools and systems in the educational technology landscape, leading to potential version incompatibility and interface mismatches [15]. OpenAI's GPT-4 offers APIs that support integration with different technologies, but specific requirements may necessitate custom solutions [16]. Standardization of interfaces is also a complex issue due to variations in system designs and implementations [17]. Creating standardized interfaces is essential for smooth interaction with external services, particularly within the broader ecosystem of educational technologies. ensuring smooth integration and standardization across the technology stack is crucial for the functionality and user experience of ITS.

As ITS may serve large numbers of users with fluctuating access volumes, auto-scaling becomes a key technical challenge [18]. While cloud solutions exist, balancing scalability with stability is difficult [19]. Costs, especially from APIs like GPT-4, are significant, requiring developers to balance educational benefits with expenses.

Security is vital when integrating external APIs into education systems. Developers must adhere to data protection laws and prioritize student data safety [20]. Ensuring robust security measures minimizes risks like data breaches [21]. Overcoming these challenges is crucial for ITS's successful implementation, ensuring enhanced learning experiences.

C. Dicussion of Technical Solutions

The primary target of our project is integrating the GPT model into the ITS. Studies indicate that AI-based technologies can enhance learning outcomes and motivation [22]. Specifically, AI-powered tutoring programs can boost students' academic performance and drive in educational settings [23]. In particular, ChatGPT, by offering tailored interactive support, can enrich the learning experience and heighten student engagement [24]. By exploring the integration of the GPT model with Intelligent Tutoring Systems (ITS), one can significantly address the limitations of ITS in personalized feedback and interaction. Additionally, this integration can support individualized tasks and activities aligned with students' needs and learning objectives.

In our implementation, we envisioned two technological solutions. The first approach centers on enabling the GPT model to autonomously generate questions. Based on the questions posed by GPT, tutoring sessions for students are facilitated. The second approach emphasizes the GPT model's capability in feedback and guidance. Here, appropriate questions and prompts are relayed to the GPT model via a backend database, ensuring optimal guidance and feedback. However, based on multiple pieces of evidence, the performance of the GPT-3 model reveals certain constraints. The GPT-3 model sometimes struggles with natural language generation, occasionally producing uncontrollable content [25]. Given that contemporary language learning and testing now incorporate rich multimedia contents like audio, video, and images, the GPT model's inability to autonomously generate multimedia-format questions can only yield basic textual queries, significantly dampening students' interest [26].

Thus, in this endeavor, we employed the second approach. Harnessing the GPT model's prowess in understanding and responding to natural language inputs, combined with its inherent self-learning capacities, we relay appropriate descriptors and metrics to the GPT model via a backend system [27]. This allows the model to offer tailored feedback and guidance to learners, thereby enhancing their overall educational experience.

IV. IMPLEMENTATION

This section offers an overview of the implementation of Intelligent Tutoring Systems (ITS), providing detailed insights into both the frontend and backend components, data source, the integration of the GPT model, and the Continuous Integration/Continuous Deployment (CICD) workflow. The structural organization of this section follows our technological stack, enabling a cohesive understanding of the separate elements that constitute the system.

A. Front-End

Continuing with this project, the frontend has been developed as a web service application to ensure portability and easy access for students across various operating systems and technological devices. To achieve this, we've employed the three primary programming languages for the web: JavaScript, HTML, and CSS.

JavaScript, as our core programming language, has been complemented by NPM, serving as both an architecture and package manager. NPM facilitated the installation, removal, and testing of libraries crucial to our project. Through command-line interactions, NPM has managed our development environment, facilitated project building for publishing, and enabled code testing.

We've harnessed the power of React, a free and open-source library, to craft a dynamic and interactive user interface. This approach has allowed us to modularize complex UI elements, creating reusable components and modules that enhance both efficiency and maintainability.

In conjunction with React, we've seamlessly integrated other libraries such as Axios. This library streamlines API integration by simplifying the process of making and managing requests. Additionally, Reactstrap has proven invaluable for incorporating new HTML and CSS elements. By utilizing its components, UI design becomes more intuitive. Furthermore, the implementation of translation into the UI has been facilitated by i18Next. This involved creating distinct JSON files for the various languages involved.

Internally, we've adopted tools like Prettier and ESLint to ensure our code is clean and well-structured. Throughout the development process, stringent control over components was exercised primarily within the JavaScript files. This approach has been pivotal in maintaining code quality and consistency across the platform.

B. Back-End

Within the scope of this project, the backend assumes the primary responsibility of receiving and dispatching client requests, implementing business logic, and facilitating interactions with the database. Given the specific technology stack chosen, which aligns with the nature of a concise MVP development effort, we've placed a strong emphasis on selecting open-source and cost-free frameworks and tools.

Our initial preference leaned towards the Spring Boot framework, serving as the cornerstone of our development. Its intrinsic features of automated configuration and embedded server capabilities have significantly streamlined our development process. By combining this with Maven, the focus narrows down to fundamental adjustments within the "pom.xml" configuration file and the ”application” configuration file. This endeavor not only ensures version control and lifecycle management but also fosters consistency across platforms and projects.

For data persistence, we opted for MySQL as our database solution. This decision rests on a twofold foundation. Firstly, MySQL’s InnoDB storage engine seamlessly aligns with the essential ACID properties - atomicity, consistency, isolation, and durability - ensuring transaction reliability and data integrity. Additionally, MySQL supports foreign constraints, reinforcing data integrity. Secondly, MySQL is optimized for performance, delivering attributes like high throughput and low-latency read and insert operations, making it well-suited for applications handling substantial workloads. To address the potential need for query optimization and customization, we turned to MyBatis as our Object-Relational Mapping (ORM) tool of choice. This selection is rooted in its exceptional flexibility in this domain, coupled with its capacity to handle intricate database structures and data transformations. Its smooth integration with Spring Boot further simplifies the implementation of data persistence.

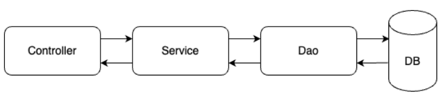

Fig. 1. three-tier architecture

Within the operational workflow of the system, as illustrated in the Fig. 1, a strategic adoption of a three-tier architecture (Controller-Service-DAO) has been implemented. The Controller tier serves as the entry point for user requests, invoking relevant services and providing the frontend with the required data. The Service layer manages business logic and orchestrates communication with the data access layer (DAO). The DAO tier interfaces directly with the data repository, usually a database, to facilitate Create, Read, Update, and Delete (CRUD) operations.

Throughout the business flow, communication between the frontend and backend occurs in JSON format. For instance, during the user login process, the frontend sends a JSON-formatted request containing the username and password. This request is initially received by the Spring Boot controller, which then forwards it to the appropriate service for processing. The service layer engages in vital business operations, such as verifying account existence and authenticating decrypted passwords. Concurrently, the service layer interacts with the DAO layer by invoking methods and utilizing XML configurations for result mapping. This interaction, facilitated by MyBatis and the MySQL database, involves executing customized SQL statements to retrieve the required data.

This architecture offers several advantages. Its modular code design, achieved through hierarchical division into well-defined modules and components, reduces inter-layer coupling. This allows targeted development on specific segments without affecting others, enabling parallel development and swift responsiveness within the team. Additionally, this architecture sets the stage for future scalability improvements. For instance, if increased data volume and the introduction of additional cache mechanisms (e.g., Redis) become necessary, these upgrades can be implemented without altering the Controller and Service layers, ensuring smooth transitions. This approach combines practical and theoretical principles, resulting in a comprehensive implementation plan that underscores the significance of the chosen architectural and technological paradigms.

C. Data Source

This project sources its data from three primary avenues: user information acquired directly from users; educational videos, which have been strategically selected from YouTube; and question data, which has been inspired by and adapted from official language examinations such as IELTS, HSK, and DELE, and then stored in our database.

While executing this segment, we encountered three primary challenges:

1) Database Design: Given the substantial volume of information at our disposal, it was imperative to adopt the Boyce-Codd Normal Form (BCNF) in the database design. Such an approach mitigates redundancy and dependency in the database tables, thereby upholding a high degree of data integrity and consistency.

2) Multimedia Information Retrieval: Our database is not equipped to directly house multimedia resources, be it video, audio, or image files. To circumvent this limitation, we retain references to the locations of these resources. For videos from YouTube, the frontend makes a request to the backend, which then retrieves the video URL from the database. This is subsequently accessed by the frontend through YouTube’s playback widget. As for image and audio resources, they reside on an AWS S3 server—a flat-storage device. We manually upload educational materials to the relevant S3 directories and record the paths and filenames in our database. When the frontend application requisitions specific learning materials, the backend supplies the stored path and filename, enabling the frontend to fetch the data from the S3 server via AWS authentication and the corresponding URL.

3) Transferring Multimedia Resources to the GPT Model: Given that the GPT model predominantly entertains text-based data for feedback, we’ve resorted to tagging our data. For every multimedia piece that necessitates an in-depth feedback from GPT, we append an appropriate descriptor, ensuring that GPT comprehends our inquiry optimally. Through this, we aim to extract GPT feedback tailored for multimedia content.

D. GPT

We have embraced a microservices architecture to implement intelligent feedback functionality through integration with the GPT model. This design approach offers the advantage of scalability and long-term maintainability. Given that our backend’s primary role involves the reception and routing of requests, a lightweight and streamlined framework for API communication is essential. To address this need, we have opted for the Flask framework. Leveraging Flask, we can swiftly configure an API interface that facilitates seamless interaction between the frontend and the GPT model. Upon the receipt of requests from the frontend, we can promptly engage with the GPT model to generate contextually appropriate responses.

Within this project, we harness the distinct attributes encapsulated by four dimensions - activist, reflector, theorist, and pragmatist - ascertained from the Learning Style test [7]. This strategic utilization enables the provision of tailored feedback in alignment with the individual learning styles of students. However, in light of the constraints associated with the GPT model, our interaction is confined to textual communication. In order to cater to the delivery of feedback in multimedia formats, a pragmatic solution has been implemented. We have associated non-textual data within our database with pertinent descriptions. This tactic enables us to manage queries and feedback pertaining to video and audio resources to a certain extent.

E. Implementation:

Our project uses AWS for cloud infrastructure, Docker for containerization and Kubernetes for managing docker containers. On top of that, we also use Jenkins for continuous integration/continuous delivery.

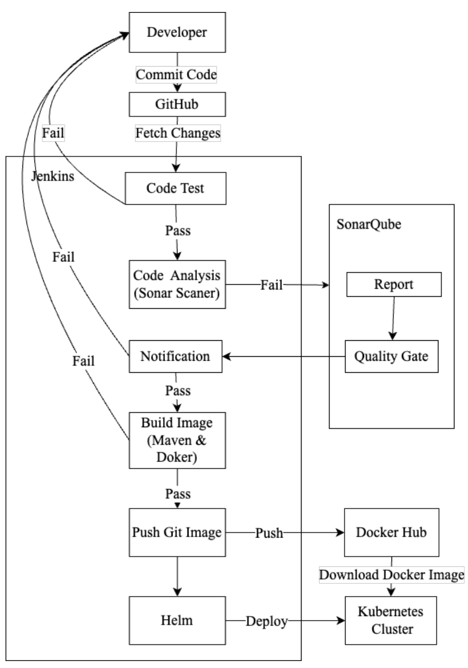

Fig. 2. cicd pipeline

The Fig. 2 is the CICD pipeline. It starts with a developer submitting the code that needs to be updated to Github, and the Jenkins server periodically checks Github for new commits. When a new commit is detected, Jenkins will start the corresponding pipeline process. Code testing is performed first, such as mvn test, and an exception is reported if a failure is detected. When the code testing is done, we use Sonar to analyze the code. The analyzed results of the code are uploaded to Sonar Qube for display. After the code is analyzed, the docker image will be built and uploaded to the docker hub. after the image is uploaded, Jenkins will call the slave node (Kops server) to update the new image to the Kubernetes cluster using a rolling update.

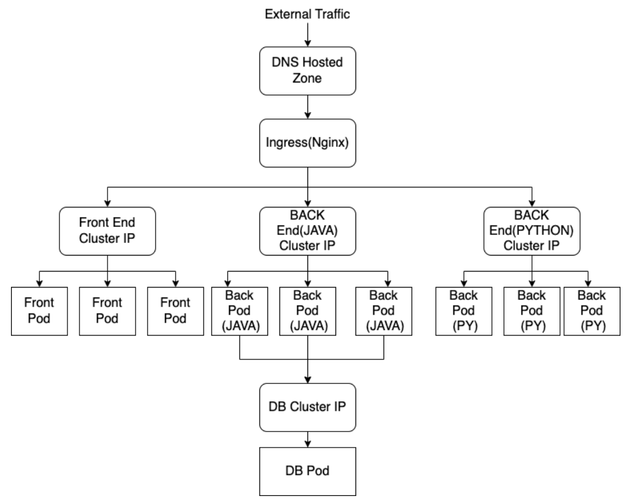

Fig. 3. Architecture of Kubernetes

Our Kubernetes cluster runs on AWS cloud servers and is controlled using Kops. When creating the Kubernetes cluster, Kops server will automatically request EC2 instances, loadbalancer, target groups and other resources based on the command. The Fig. 3 is a diagram of our Kubernetes architecture. First the user accesses it through our domain name. The domain is bound to our Kubernetes loadbalancer via Route 53, and we use a Nignx ingress controller to control which front-end and back-end services the request needs to access. Our services include a front-end service using the React framework, a Java back-end service using the Spring boot framework, a Python back-end service using the flask framework, and a MySQL database. For all front-end and back-end services, we use ClusterIP, Kubernetes Deployment and HPA to manage all the pods.ClusterIP is a Kubernetes service type that manages which traffic is sent to which pods. In our architecture, ClusterIP accepts traffic from the ingress controller and passes the request to the corresponding pod.Kubernetes Deployment is a resource object that is used to define and manage a set of pods.When using Kubernetes Deployment, we define the number of pods, and when a running pod terminates unexpectedly, or the When a service is updated, Kubernetes Deployment starts new or stops old pods to ensure that the number of pods running in the cluster remains the same. HPA (Horizontal Pod Autoscaling) is used to autoscale the pods within the cluster. Firstly the autoscaling within the cluster is categorized into horizontal scaling and vertical scaling. Horizontal scaling scales the number of pods but does not change the CPU and memory limits of the pods. Vertical scaling does not change the number of pods, but dynamically changes the CPU and memory limits of the pods. HPA monitors the average CPU utilization of the corresponding service and when the CPU utilization reaches the value we set (80%), new pods are started immediately to increase the load capacity of the cluster. In addition, when HPA detects low CPU utilization, it starts deleting excessive pods after five minutes. Similar to HPA, our cluster nodes are also equipped with an autoscaling feature, whereby new nodes are self-started when the cluster CPU utilization is too high. When the CPU utilization drops, the cluster will reschedule the pods running on the nodes that will be deleted to other nodes within 10 minutes.

V. Evaluation

The evaluation of the project was divided into a system evaluation and a functional evaluation.

A. system evaluation

The system evaluation consists of the following aspects: monitoring of basic metrics, scalability of the system and response time of users accessing the system.

1) basic metrics and scalability: With traditional architectures (no microservices architecture), it is important to develop evaluation criteria for basic metrics such as CPU, memory, etc. due to the lack of scalability. However, since we are using a microservices architecture, the role of these base metrics is more often used as a criterion for auto-scaling. The evaluation metrics for auto-scaling are the time required for auto-scaling and whether the scaling is perceived by the user. Autoscaling is divided into node autoscaling and pod horizontal scaling.Autoscaling of nodes takes about 2 to 3 minutes and is completely insensitive to the user since no traffic enters when node starts. For horizontal scaling of pods, the situation can be relatively complex. Each different service takes a different amount of time to start, for example, our Java backend service takes about 45 seconds to a minute or so to perform auto-scaling, which is extrapolated from the readiness time of the Http type probe. In addition, due to the use of probing technology, user traffic will be accessed after the probe detects that the service is ready, so the user will also not receive an incorrect return response. For pod scaling, some of the pods are removed when the average cpu usage is reduced to five minutes after which the pods can be reduced. The shortcoming of auto-scaling is that we set the CPU utilisation to start scaling to 80%, which is a resource-saving setting, but when a large number of requests come in, the system may perform a sudden scaling, during which the user will feel a significant increase in the response time of the service due to the queue stacking.

2) response time: The average response time is an important metric for measuring the system, and we use a stress test to measure the average response time of the system under normal conditions and when scaling up, respectively.

The table I shows the average response time under normal conditions.

TABLE I
NORMAL RESPONSE TIME

status	succ	avg_rt	error
OK	100.00%	0.011	

The table II shows the average response time when the server is experiencing auto-scaling. We can see that the average response time gets longer due to the stacking of the message queue, and it is not until the auto-scaling is complete that the average response time drops to a normal level.

TABLE II
BUSY RESPONSE TIME

status	succ	avg_rt	error
OK	100.00%	0.0046	

B. Functional evaluation

1) A/B Test: A total of six subjects were included in this test, all of whom were students of the School of Computing, UCD. Three of the subjects were assigned to the experimental group and the other three were assigned to the control group. Before the start of the experiment, members of both groups were pre-tested and their scores were recorded. During the experiment, the experimental group used our program for learning and the control group used the same learning materials as in the program, but did not use our service. At the end of the study, all members took a posttest and recorded their scores. Both the pre-test and the post-test consisted of five multiple-choice questions, and each subject answered the same questions.

We calculated the mean improvement of the subjects' pre-test and post-test and their variance for t-test. The data are shown in the table III,

TABLE III
T-TEST

	experimental group	control group
Average pre-test	0.5	0.66
Average post-test	3.33	3
Average improvement	2.83	2.33
variance	0.333	1

We chose the independent samples two-sided t-test for the following reasons: 1. The sample size was extremely small (n=3). 2. there was no overall standard (e.g., intelligence test) that could be used for comparison. 3. We did not know whether our system would have a positive or negative impact on students (it is possible that after using the system students would not be as good as if they were self-taught).

The (1) is the formula for the independent samples t-test:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (1)$$

The T-value of 0.75 was calculated by bringing the desired value into the formula. We chose 0.05 confidence interval as well as two sided t-test to get this critical value of 4.303 by looking up the table when the degree of freedom is 2. Since our t-value (0.75) is less than 4.303, it cannot be proved that our system has an improving effect on students' learning.

The shortcomings of this assessment method are the following two:

1) The number of participants in the experiment is too small, usually the group size for this type of experiment will be larger than five, in some cases, such as when schools promote new learning methods, a whole class will be used for the test. As we only had three people in each group taking part in the test, this made the t-value required to achieve confidence quite high. It is often difficult to prove statistically that the service we provide is effective.
2) Our test has only five multiple-choice questions, and with a small number of questions on the test, the random effects of subjects guessing their answers to questions will be greater than when there are many questions, which will make our final results inaccurate.

2) User Test: During a comprehensive user testing phase of our intelligent tutoring system, we engaged a diverse user group to ensure varied and in-depth feedback. This robust evaluation was driven by our commitment to inclusivity and a user-centred ethos.

Notably, our UI & UX design received widespread acclaim. Users lauded the intuitive design, swift page load speeds, and impeccable navigation aided by a well-structured information architecture. The comprehensive suite of tools and resources we offered was also recognized, showcasing the platform's robust functionality.

In addition, we have identified some important areas for improvement. Specifically, these encompassed shortcomings in accessibility for differently-abled users, the need for a more integrated social media experience, voiced demands for improved multilingual capabilities, and a consensus on refining the search function. In response to this invaluable feedback, we advocate a systematic accessibility review in compliance with WCAG standards, more seamless social media integration, the introduction of translation facilities aligned with user-focused localisation, and the advancement of the search module, spotlighting faceted search and predictive suggestions. With a foundation in iterative refinement, we hold every piece of feedback in high regard, using it as a beacon for our ongoing system enhancements.

3) Tree Testing: Tree Testing is an evaluation method in the field of user interface design and information architecture used to test whether the navigation and organization of information on a website or application is clear and easy to understand. The primary goal of tree testing is to verify that users can easily find specific information or perform specific tasks without being influenced by page design, style, or visual elements.

We tested our system with the assistance of Optimal Workshop. Before the test, we entered the sitemap of the AI Tutoring website into Optimal Workshop's website, set up tasks to be completed by the tester, and generated links for the test. The testers were asked to find the appropriate features,

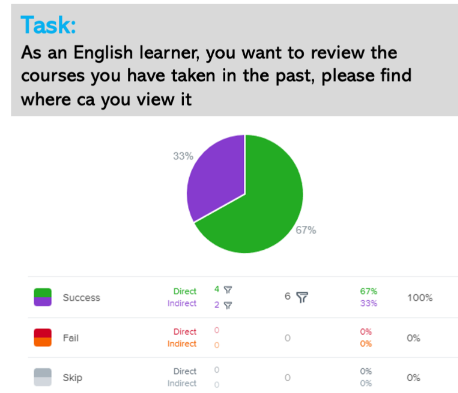

Fig. 4. Results of Tree test

and Optimal Workshop recorded and analyzed the data for each selection. As the Fig. 4 shown, we found that 100% of the users were able to find the correct feature, with 33% of the users not finding the feature directly.

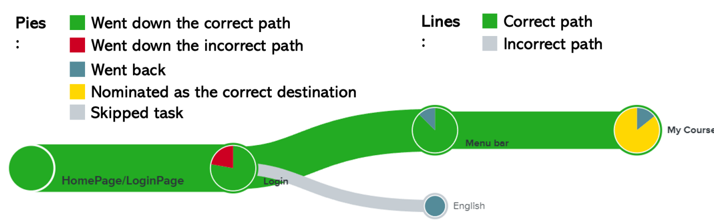

Fig. 5. View of pie tree

The users in this section went down the incorrect path to English before they find the correct feature. The Fig. 5 shown that our sitemap has a good recognition degree. This shows that the sitemap of our system has good recognition and users can use it without any training and find the corresponding features. However, there is still room for improvement, for example, the next version of the system to add a newbie guide function, for the use of the process to give the appropriate tips.

VI. Conclusion

Our project successfully realized the functional and non-functional requirements of an intelligent tutoring system. The unique feature of the project is to provide descriptions for all learning materials and quizzes, and make the GPT responses more relevant by sending the descriptions to the GPT API at the same time as the student asks a question. Our project still has some shortcomings. First, the combination of the Learning Styles Questionnaire and the GPT did not produce significant results; the GPT only mentioned learning styles in the responses, and there was not enough variation in the responses across learning styles. Second, the results of the t-test did not show in a statistically significant way that our system improves students' learning outcomes. The reason for this problem may be due to the small sample size; with a sample size of 3, only a large enough difference can be considered statistically significant. In addition, we still have several ideas that need to be realized in the future. First, we would like to add emotion recognition to the system for monitoring students' learning status. Alerting students when they are distracted or taking into account their learning state when they ask questions to the GPT to learn a particular segment. Secondly, we would like to include other kinds of learning style tests and combine learning styles with the student's learning process rather than the feedback provided by the AI. Finally, we want to make the system an open-source educational platform that offers the opportunity to upload custom courses, so that anyone can add courses that meet the specifications to our platform and make the AI service available to their students.

Acknowledgment

We extend our profound gratitude to *Dr. Simon Caton*, *Dr. Pavel Gladyshev*, *Dr. Timilehin Aderinola*, and *Dr. Honghui Du* for their invaluable guidance throughout our software project. Their weekly lectures were not only illuminating but also significantly accelerated our work's progress.

We also wish to thank *Team Super Saiyan* for their instrumental assistance during user testing. The queries they posed substantially alleviated the challenges we encountered with user experience and multilingual support.

References

[1] UNESCO Institute for Statistics (UIS). Literacy. Accessed Aug. 20, 2023. [Online]. Available: http://uis.unesco.org/en/topic/literacy

[2] A. Ni and A. Cheung, "Understanding secondary students' continuance intention to adopt ai-powered intelligent tutoring system for english learning," *Education and Information Technologies*, vol. 28, no. 3, pp. 3191–3216, 2023.

[3] B. Jacob, D. Berger, C. Hart, and S. Loeb, "Can technology help promote equality of educational opportunities?" *RSF: The Russell Sage Foundation Journal of the Social Sciences*, vol. 2, no. 5, pp. 242–271, 2016.

[4] V. Yuskovych-Zhukovska, T. Poplavska, O. Diachenko, T. Mishenina, Y. Topolnyk, and R. Gurevych, "Application of artificial intelligence in education. problems and opportunities for sustainable development," *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, vol. 13, no. 1Sup1, pp. 339–356, 2022.

[5] S. Grassini, "Shaping the future of education: exploring the potential and consequences of ai and chatgpt in educational settings," *Education Sciences*, vol. 13, no. 7, p. 692, 2023.

[6] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier *et al.*, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.

[7] P. Honey and A. Mumford, *Learning styles questionnaire*. Organization Design and Development, Incorporated, 1989.

[8] A. Rrafzadeh, S. Alexander, F. Dadgostar, C. Fan, and A. Bigdeli, "How do you know that i don't understand?" *A look at the future of intelligent tutoring systems. Computers in Human Behavior*, vol. 24, no. 4, pp. 1342–1363, 2008.

[9] K.-T. Wang, Y.-M. Huang, Y.-L. Jeng, and T.-I. Wang, "A blog-based dynamic learning map," *Computers & Education*, vol. 51, pp. 262–278, 08 2008.

[10] G. T. Jackson and D. S. McNamara, "Motivation and performance in a game-based intelligent tutoring system." *Journal of Educational Psychology*, vol. 105, no. 4, p. 1036, 2013.

[11] A. Latham, K. Crockett, and D. McLean, "An adaptation algorithm for an intelligent natural language tutoring system," *Computers & Education*, vol. 71, pp. 97–110, 2014.

[12] J. M. Harley, F. Bouchet, M. S. Hussain, R. Azevedo, and R. Calvo, "A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system," *Computers in Human Behavior*, vol. 48, pp. 615–625, 2015.

[13] D. Wang, H. Han, Z. Zhan, J. Xu, Q. Liu, and G. Ren, "A problem solving oriented intelligent tutoring system to improve students' acquisition of basic computer skills," *Computers & Education*, vol. 81, pp. 102–112, 2015.

[14] D. Arnau, M. Arevalillo-Herráez, and J. A. González-Calero, "Emulating human supervision in an intelligent tutoring system for arithmetical problem solving," *IEEE transactions on learning technologies*, vol. 7, no. 2, pp. 155–164, 2014.

[15] S. Ramaila and A. J. Molwele, "The role of technology integration in the development of 21st century skills and competencies in life scsiences teaching and learning." *International Journal of Higher Education*, vol. 11, no. 5, pp. 9–17, 2022.

[16] A. Tsikora, "A feasibility study of available natural language chatbot technologies," 2023.

[17] A. Lorenz, M. Raven, and K. Blind, "The role of standardization at the interface of product and process development in biotechnology," *The Journal of Technology Transfer*, vol. 44, no. 4, pp. 1097–1133, 2019.

[18] J. Yang, C. Liu, Y. Shang, B. Cheng, Z. Mao, C. Liu, L. Niu, and J. Chen, "A cost-aware auto-scaling approach using the workload prediction in service clouds," *Information Systems Frontiers*, vol. 16, pp. 7–18, 2014.

[19] D. C. Marinescu, *Cloud computing: theory and practice*. Morgan Kaufmann, 2022.

[20] M. N. Asghar, N. Kanwal, B. Lee, M. Fleury, M. Herbst, and Y. Qiao, "Visual surveillance within the eu general data protection regulation: A technology perspective," *IEEE Access*, vol. 7, pp. 111 709–111 726, 2019.

[21] S. Slade and P. Prinsloo, "Learning analytics: Ethical issues and dilemmas," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1510–1529, 2013.

[22] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[23] A. S. Patil and A. Abraham, "Intelligent and interactive web-based tutoring system in engineering education: Reviews, perspectives and development," in *Computational Intelligence for Technology Enhanced Learning*. Springer, 2010, pp. 79–97.

[24] K. Srinivasa, M. Kurni, and K. Saritha, "Harnessing the power of ai to education," in *Learning, Teaching, and Assessment Methods for Contemporary Learners: Pedagogy for the Digital Generation*. Springer, 2022, pp. 311–342.

[25] M. Zhang and J. Li, "A commentary of gpt-3 in mit technology review 2021," *Fundamental Research*, vol. 1, no. 6, pp. 831–833, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667325821002193

[26] J. Costley and C. H. Lange, "Video lectures in e-learning: effects of viewership and media diversity on learning, satisfaction, engagement, interest, and future behavioral intention," *Interactive Technology and Smart Education*, vol. 14, no. 1, pp. 14–30, 2017.

[27] M. Firat, "How chat gpt can transform autodidactic experiences and open education," *Department of Distance Education, Open Education Faculty, Anadolu Unive*, 2023.