Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/132/2024.20577

Bridging Educational Achievement Gaps with Generative AI: Personalized Curriculum for Targeted Learning Support

Yahan Cao[1,5,†], Chang Gu[2,6,†], Yixuan Li[3,7,*,†], Yanze Lyu[4,8,†]

[1]School of Architecture, Design and Planning, University of Sydney, Sydney, NSW 2006, Australia
[2]School of Informatics, Xiamen University, Xiamen, Fujian 361102, China
[3]Department of Psychology, Columbia University, New York, NY 10027, United States
[4]Faculty of Science, McGill University, Montreal, QC H3A 0G4, Canada

[5]13883937062@139.com
[6]guchang0208@stu.xmu.edu.cn
[7]yl4688@columbia.edu
[8]yanze.lyu@mail.mcgill.ca
*corresponding author
[†]co-first authors

Abstract. The inequitable distribution of educational resources plays a major role in widening achievement gaps, placing students from lower socioeconomic status (SES) backgrounds at a disadvantage due to systemic barriers that restrict access to these resources. Recent advancements in the development of artificial intelligence (AI), namely ChatGPT by OpenAI, showcased its ability to generate text responses in natural language format based on input prompts. The accessibility and convenience of ChatGPT hold promise for offering personalized learning support. In pursuit of this goal, the authors built Ligare – an AI-powered curriculum builder that integrates, optimizes, and presents generated responses with a user-friendly interface. The design process followed rigorous Human-Computer Interaction (HCI) protocols, including pre-development analysis, low- and high-fidelity prototyping, and subsequent evaluation. Although Ligare currently supports only math learning, the evaluation results demonstrate its potential for broader application, highlighting future directions for providing more accessible personalized education and addressing achievement gaps.

Keywords: ChatGPT, Human-Computer Interaction (HCI), Artificial Intelligence (AI), Personalized Education.

1. Literature Review

The inequitable distribution of educational resources is a pervasive challenge faced by educational systems worldwide. Among many factors that perpetuate this inequality, socioeconomic status (SES) has been extensively studied for its impact on education outcomes. First established by French sociologist Pierre Bourdieu, the Social Capital Theory (SCT) posits that the availability of three forms of capital– economic, cultural, and social– actively shapes education opportunities for children with

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/132/2024.20577

different SES backgrounds [1]. The disadvantages can manifest in the form of poorly funded school districts, less exposure to culturally enriching experiences, and fewer social connections who can provide mentorship. The lack of access to these resources often limits the educational prospects of students from low-SES families compared to their high-SES peers. Though the resulting disparity in educational outcomes varies in magnitude across countries, evidence from numerous studies confirms the consistent importance of SES on achievement [2]. COVID-19 dramatically amplified these vulnerabilities and widened the existing gap. No one was left unaffected by school closures and the difficulty of navigating remote learning, but marginalized communities were disproportionately impacted [3][4][5]. Quantitative studies have demonstrated that minority students and those coming from low-SES households experienced a more substantial decline in academic performance due to unequal access to resources and environments that effectively support their learning [6][7]. Even though the pandemic has subsided in 2021 and the schools have reopened, these disparities led to significant consequences. Limited access to high-quality educational resources creates additional obstacles as these students strive to recover from the decline [5][6].

On a positive note, the rapid development of large language models (LLMs) and their generative abilities in recent years may be a valuable asset in addressing the exacerbated inequalities in the aftermath of the pandemic. The most prominent LLM is ChatGPT developed by OpenAI. Originally created in 2018, it has evolved from a generative language model experimenting with a Transformer framework and unsupervised learning to an evermore powerful multimodal tool that revolutionized human-computer interaction [8]. The latest ChatGPT-4o (omni) outperforms its predecessors, with expanded context windows, optimized tokenizing and processing efficiency, and improved generation and comprehension of various forms of content [9]. Application of artificial intelligence in education (ALEd) has long been prevalent in the forms of chatbots, expert systems, intelligent tutors, machine learning, personalized learning systems or environments (PLS/E), and visualizations [10]. Thus, it should not be surprising that education researchers have discussed the possibilities of employing the model for teaching and learning across different subjects, suggesting positive prospects of its use as a supplementary educational tool [11][12]. Inspired by these research findings, the authors of this paper sought to realize the proposed possibilities by developing an educational application. The development focused on addressing the achievement gap by leveraging the personalizing and generative capabilities of ChatGPT-4o, responding to individual users' unique learning needs. It will generate customized curricula by structuring learning goals and schedules, reinforcing their funds of knowledge, and building competency in areas where they are falling behind compared to their peers. Additionally, it will curate open-access resources as guides and references and generate corresponding exercises to strengthen newly acquired knowledge. As a convenient, all-in-one learning tool, the application is designed to be easily accessible from mobile devices, lifting the barrier of requiring a less affordable computer.

2. Competitor Analysis

Prior to developing the application, we conducted a comprehensive competitor analysis of existing products in the market to ensure that there were no overlaps with the features and objectives of the envisioned design.

A preliminary search narrowed the analysis to 12 products, each of which was further evaluated for their purpose, target audience, features, and cost. The results of the evaluation are presented in the table below.

DOI: 10.54254/2755-2721/132/2024.20577

Table 1. Competitor analysis of existing AI curriculum builders.

Name of Product	Purpose	Audience	Features	Cost
Coursebox.ai	Automated course authoring, training and assessment	Adult content creators looking to build courses for revenue or to expand their influence	Files to course Al assessment Al tutor Al quiz generator Al course designer White label Sell courses Certificates Mobile apps	Begins with $0, up to $2,499 per month
AcademyOcean	Al Learning Management System (LMS) for employee growth	Corporate human resources (HR) learning and development (L&D) function	Knowledge management system LMS portals Cloud LMS Certificate builder LMS integrations Customizable LMS Quiz maker Webinar learning LMS analytics LMS administration AI-powered LMS	Begins with $255 per month, up to $2,380 per month
CourseAl.com	Streamline the process of creating online courses for financial independence	Adults looking to profit from course building	Al generated topic Al tailored course content Voice script generation Voice over generation All-in-one Email generation	Begins with $0 for 3 days, up to $29.99 per month
Courseau	Generator for effective, interactive courses for workplace training	Companies and small businesses	Outline generation Audio lessons Al editor Multi-lingual Embed/ SCORM/PDF Learning interface Latest Al advancements Extract source images Organizations	Begins with $99 per month, up to customized price based on enterprise size

DOI: 10.54254/2755-2721/132/2024.20577

Table 1. (continued).

Disco.co	Creator for cohort-based programs, upskilling courses, and community-based social learning	Adults who want to host a community of learners	Course templates Content library Custom onboarding Curriculum builder, course cloning Learner progress reports SCORM support Channels and DMs, bulk messaging Events Social feeds, groups and sub-groups Custom profiles Member directory Slack integration	Begins with $359 per month, up to customized price based on enterprise size
SC Training	Create and deliver onboarding, training, and learning content to employees	Companies and small businesses	AI generation as supportive feature for a lager product	Begins with $0 for under 10 learners of a course, up to customized price based on enterprise size
Heights	Create courses, website and scale a business	Adults looking to start a business	Content editing Payment processing Community building Student analytics Gamification Marketing tools White label branding Integrations Hosting and support	Begins with $19 per month, up to $399 per month
Kajabi	Content generation for revenue	Adults looking to monetize their knowledge	AI generation as supportive feature for a lager product	Begins with $55 per month, up to $319 per month
Lingio	Turn text into employee training course	Business managers	Gamification elements and storylines Language support and accessibility Customizable course options	Begins with $416 per month, up to customized price based on enterprise size

DOI: 10.54254/2755-2721/132/2024.20577

Table 1. (continued).

Mindsmith	AI Learning Management System (LMS) for employee growth	Companies and small businesses	Drag-and-Drop Interactive Elements Assessments Multimedia AI Audio Narration Export Options Robust Branding/Theming Real-Time Team Collaboration	Begins with $0, up to customized price based on enterprise size
Mini Course Generator	Enrich training, monetization, automated onboarding, in app education, lead generation	Adult individuals or organizations	AI course generation Learners' access Sharing Automation	Begins with $0, up to $25 per month
Teachable	Create and sell online courses to grow business	Adult content creators looking to build courses for revenue or marketing	AI generation as supportive feature for a larger product	Begins with $29 per month, up to $199 per month

As demonstrated by Table 1, we identified substantial gaps between the most successful existing AI curriculum builder tools and the application that we intend to create. The analyzed products are incomparable because 1) their purpose and interfaces are designed for adults; 2) they cater mostly to content creators rather than individual learners; 3) they are primarily intended for organizational use; and 4) their pricing is usually unaffordable. Given these critical differences, we recognized the need to develop our AI curriculum builder tailored for individual students to use at home for academic advancements.

3. Low-Fidelity Prototyping with Participatory Design

Building on these insights, we designed *Ligare*; powered by the ChatGPT-4o, this application generates unique learning plans for young learners according to their grade level, required learning content, and desired learning time. Compared to the language model itself, we enhanced the user interface to improve interaction and content presentation. To ensure the features align with user needs, we conducted participatory design testing on the low-fidelity prototype.

3.1. Participants

We invited five participants from distinct backgrounds with different learning needs and usage habits. These participants helped us examine the system's functions and user experience from multiple perspectives, as shown in Table 2.

Table 2. Participants' main concern in learning program planning from different backgrounds.

Participant	**Background**	**Need**
P1	Undergraduate student (computer science)	Reduce manual course content selection by recommending relevant content
P2	Master student (economics)	Enhance learning efficiency by automatically adjusting learning progress

DOI: 10.54254/2755-2721/132/2024.20577

Table 2. (continued).

P3	Project manager	A user-friendly interface for quick access to work-related courses
P4	Doctoral student (medicine)	Filter target medical content with optimized video categorization
P5	Undergraduate students (design)	A clean and aesthetically pleasing interface for enhanced user experience

3.2. Procedure

To help participants better understand the current state of the system, we first used paper and pen to show the preliminary design of the interface, including core modules such as learning plan generation, course video display, and user interaction functions, as shown in Figure 1. Then, we invited participants to brainstorm how to further optimize these functions and put forward suggestions for improvement on interface layout, function refinement, and personalization options.

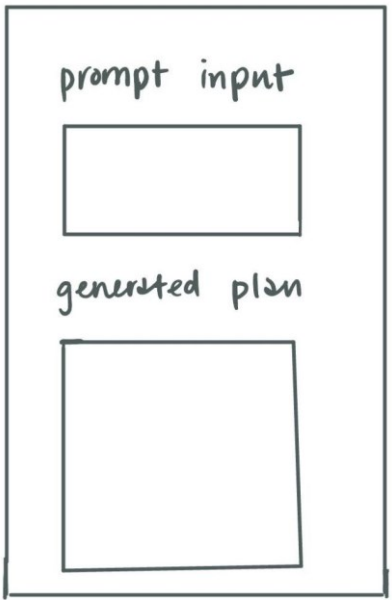


Figure 1. Preliminary design of the interface to guide participants through the system.

3.3. Key System Designs

Through collaboration with participants, we identified several key design features of the program to meet the requirements of different users and improve the overall user experience. These key designs include the following seven aspects:

1. Development of personalized learning plans with memory function;
2. Provide designated intake survey and customized learning methods;
3. Display of course learning videos;
4. Display of daily learning plans;
5. Flexibility to modify learning plans.

3.3.1. Development of Personalized Learning Plans with Memory Function

As a generative artificial intelligence, ChatGPT has powerful text generation capabilities, but it has some limitations when providing learning plans. The current GPT-4o model has a limited memory function, but facing many users from different backgrounds, it is still a big challenge to accurately identify and remember each user's learning preferences and needs, which leads to a lack of coherence and

DOI: 10.54254/2755-2721/132/2024.20577

personalization each time the learning plan is generated, and users need to repeatedly input personal information and adjust learning content. At the same time, the system cannot dynamically adjust the plan based on the user's past learning experience and performance, which reduces learning efficiency.

We have improved the system and added a memory function so that the system can record and track users' learning preferences. We designed a login interface as shown in Figure 2, set up a dedicated account for each user, and continuously saved the user's learning progress and preferences. The system can extract the user's historical data after the user logs in, providing users with a coherent learning experience and optimizing recommended content based on the historical data.

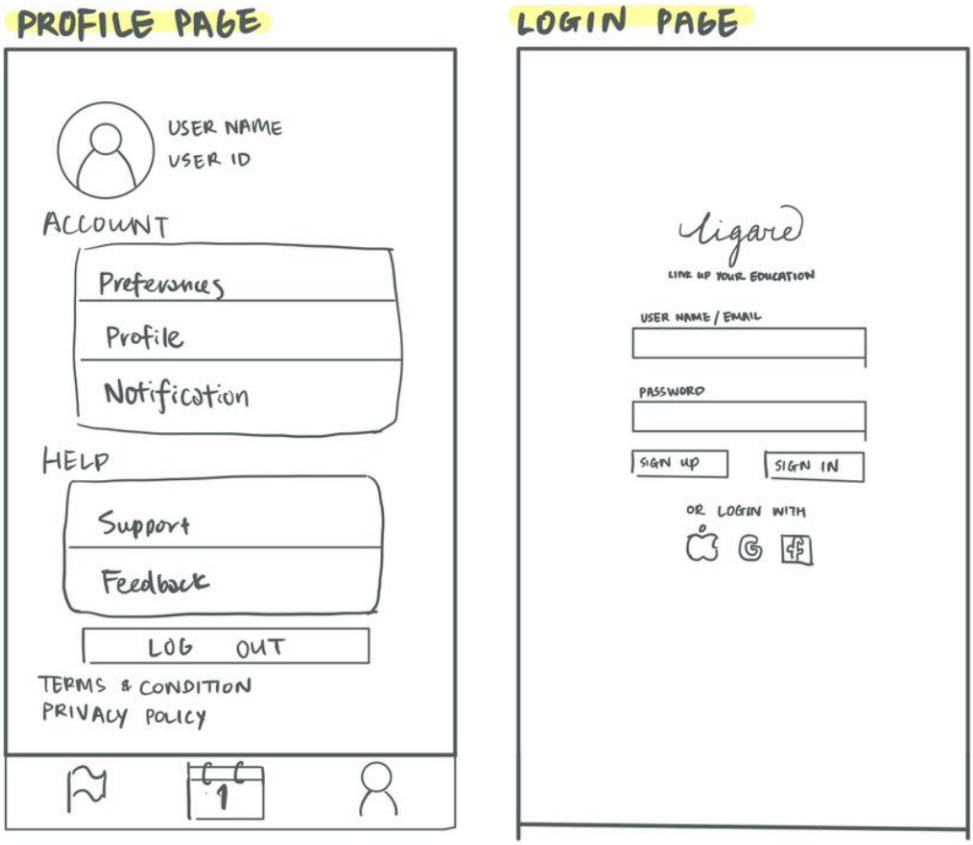


Figure 2. A login interface and personal profile page designed for feature #1.

3.3.2. Provide Designated Intake Survey and Customized Learning Methods

As a generative artificial intelligence, ChatGPT can generate learning plans based on input. ChatGPT relies heavily on users' input to generate text content, but different users have very different expressions of learning requirements and content, and some users cannot clearly express their needs in standardized language, which results in a certain degree of ambiguity. Therefore, the accuracy and fit of the generation of learning plans are greatly undermined.

In order to solve this problem, we designed an intake survey as shown in Figure 3 based on the common questions raised by the participants to understand the learning needs of each user, so that standardized language can be used for input, which not only improves the user experience but also improves the accuracy and fit of the generation of learning plans.

DOI: 10.54254/2755-2721/132/2024.20577

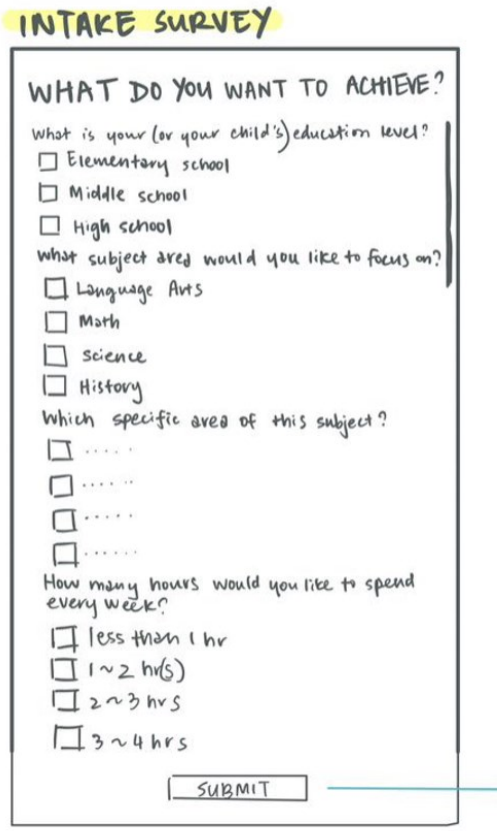


Figure 3. An intake survey designed for feature #2.

3.3.3. Display of Course Learning Videos

Although ChatGPT can generate corresponding course video links based on users' prompts, there are often problems with expired or unavailable course resources, and the matching degree with the learning content is low. Therefore, it is not always reliable to directly rely on ChatGPT to find available learning materials. To solve this problem, we first crawled learning videos from mainstream learning bloggers on YouTube through Python crawlers and stored them in the database in the format of keywords and video links. After generating the learning plan, the back-end code will perform precise matching based on keywords to improve the matching degree between the learning plan and learning resources. Users can view relevant learning videos directly on the interface, as shown in Figure 4. In addition, with the help of YouTube's related video recommendation function, the system can also personalize and recommend more learning resources to form a complete closed loop of the learning knowledge chain.

DOI: 10.54254/2755-2721/132/2024.20577

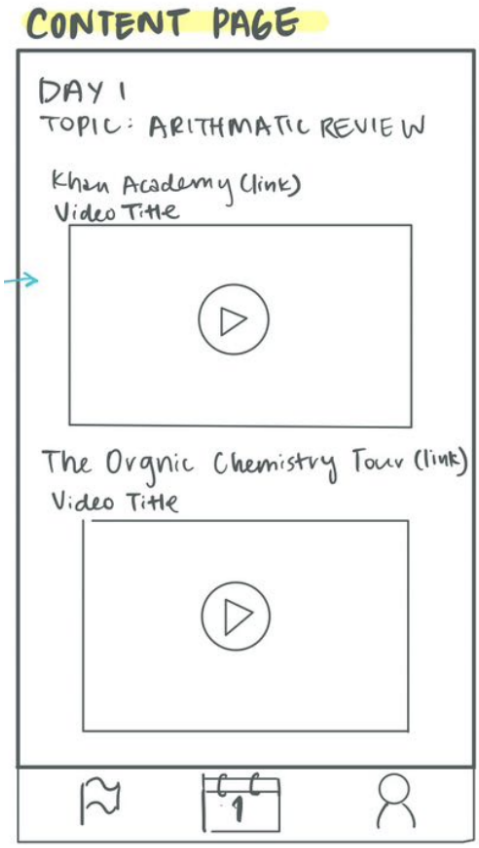


Figure 4. A content page directly displaying course learning videos.

3.3.4. Display of Daily Learning Plans

In the existing ChatGPT feedback results, it is usually presented as a complete text, lacking the necessary layout and structured organization. It makes the display of the learning plan not intuitive enough, and users need to spend more time to understand the key information when viewing it. In addition, the time schedule, task progress, and other information of the learning plan are not well highlighted, and it is difficult for users to quickly grasp the key points in daily use.

Therefore, we have improved the interface design shown in Figure 5 of the feedback results. We introduced a hierarchical layout structure, using clear titles, subtitles, and paragraph divisions to present the contents of the daily learning plan in an orderly manner. Elements such as tables and timelines are used to help users understand and track their learning progress more efficiently. At the same time, we adjusted the color scheme so that important information such as the day's tasks and urgent learning content can be highlighted. The system also supports customized views, and users can choose different display modes (such as concise mode and detailed mode) according to their personal preferences to meet different needs. Users can quickly browse and understand daily learning tasks without spending much time reading and analyzing long texts.

DOI: 10.54254/2755-2721/132/2024.20577

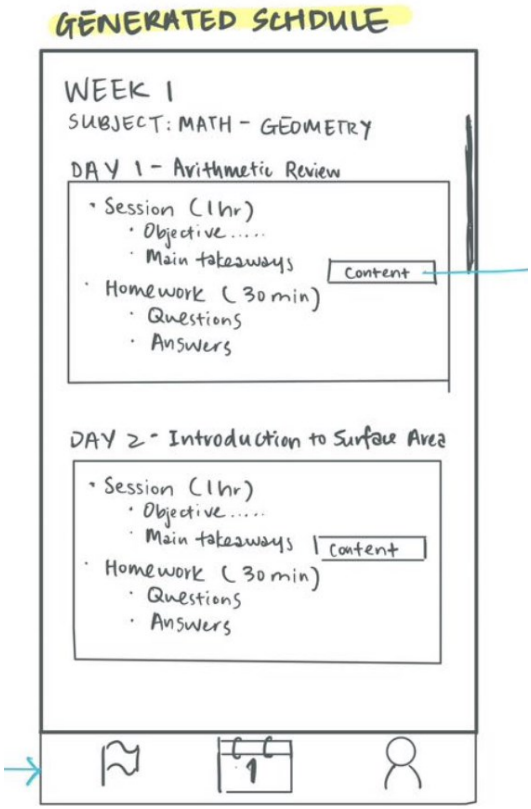


Figure 5. Optimized interface display of daily learning plans.

3.3.5. Flexibility to Modify Learning Plans

Traditional ChatGPT lacks a persistent memory function and cannot save or manage different learning plans for multiple users at the same time. This means that every time a user adjusts their learning plan, ChatGPT cannot automatically track and remember the previous learning situation, causing users to repeatedly enter a lot of the same information, increasing the complexity of the operation. Users often feel inconvenienced and lack a personalized experience when adjusting their learning plans, especially in long-term learning.

The user information memory function we introduced can combine databases and user identification to achieve personalized memory. The system automatically saves each user's learning progress and preferences and dynamically updates the learning plan based on users' feedback and adjustments. With this feature, users can modify learning tasks, adjust priorities, or add new learning content at any time, and the system will recalculate the learning progress and task allocation based on the latest plan. Moreover, the system provides intelligent adjustment suggestions to help users maintain the consistency and efficiency of learning while flexibly modifying plans. With this flexible modification mechanism, users no longer need to manually input all the adjustment details, and the system will provide them with a convenient and intelligent operation experience.

DOI: 10.54254/2755-2721/132/2024.20577

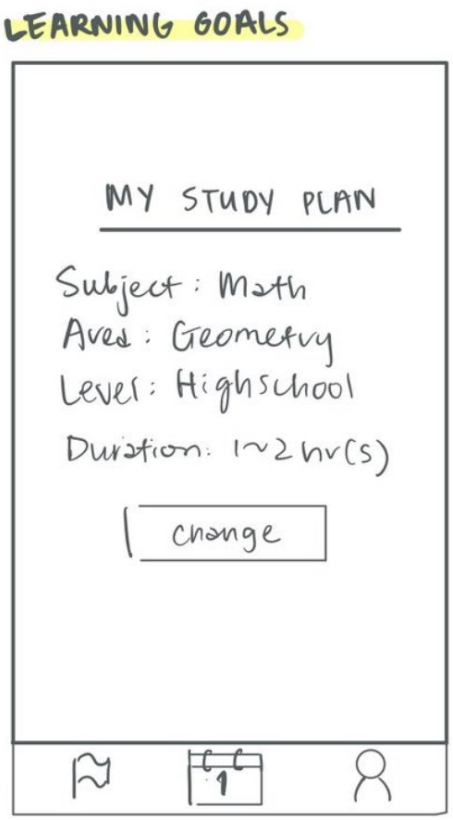


Figure 6. Learning goals page with change button.

4. Hi-Fidelity Implementation and Evaluation

We implemented the High-Fidelity prototype following the ideas generated within the Low-Fidelity prototype stage, with slight modifications on the user interface, benefiting from the feedback from the user tests on the Low-Fidelity prototype. The front end of the High-Fidelity prototype was implemented based on the Vue JavaScript framework, using HBuilderX IDE, and the backend was developed with Python, utilizing the FastAPI Python Web framework.

4.1. Prototype Walkthrough

The High-Fidelity prototype achieved the main functionalities of Ligare. After logging onto the system, the user will be able to answer several multiple-choice questions, covering the education level of the user, the subject of user interests and other information that are required to generate a personalized curriculum. After finishing the information collection, the user will be directed to the page with the generated curriculum. For each day within the curriculum period, the user will be delivered one or two videos that are closely relevant to the subject that the user is interested, and the videos will be arranged with the order of difficulty, designing a learning difficulty progression. The videos are shown directly on the page, providing benefits for the users to watch and study. For each video, there's one question generated accordingly to help the user understand and review.

DOI: 10.54254/2755-2721/132/2024.20577

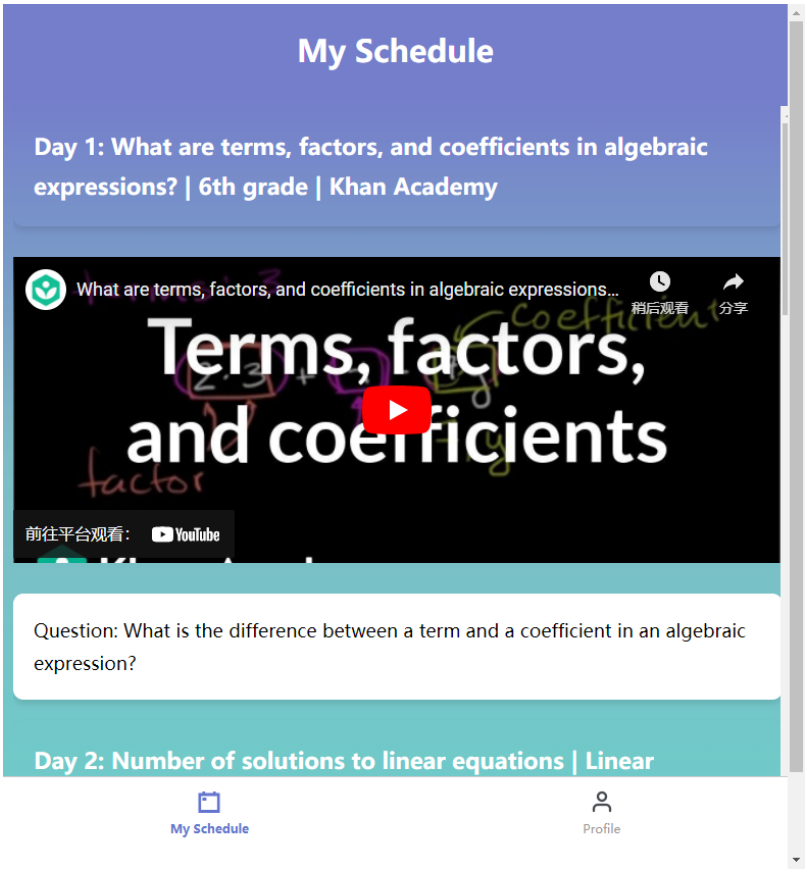


Figure 7. Curriculum generated by the high-fidelity prototype of Ligare.

4.2. Prototype Database design and incorporation of ChatGPT

According to the feedback from the design preliminary investigations and the results from the participatory design at the Low-Fidelity prototype stage as discussed above, the problem, that ChatGPT lacked accuracy and personalization, was discovered, and the need to provide a clearer and more convenient format for the curriculum demonstration was established. We achieved a clear interface in the front end based on the Vue framework, as demonstrated in the walkthrough and the figure above. We use a database approach to help increase the accuracy of ChatGPT response and provide a useful background for ChatGPT prompts.

4.2.1. Database with open source learning content

According to our investigation, we discovered that ChatGPT could hardly generate a clear and useful curriculum for the users, due to a limited range of accessible and its tendency to generate nonexistent resources. The invalid resources lead to links pointing to missing websites or websites with unrelated contents, due to either ChatGPT not being able to update the website data promptly, or it has a false understanding of the requirements provided by the users, and the insufficiency in the restrictions on the range of the resources origins leads to useless or incorrect resources provided by ChatGPT. Both deficiencies tremendously hurt the value of the AI-generated curriculum and create confusion with the user experience.

To address these issues, we implemented a database that allowed us to assemble the necessary resources beforehand. In our High-Fidelity prototype, the database was set up on the MongoDB Atlas Cloud Platform. Database on the cloud allows the user to access it in different regions. Video resources were stored inside the database in advance. We chose Khan Academy, a trusted resource providing free courses, as the test data stored in the database for the High-Fidelity prototype. In order to collect the

DOI: 10.54254/2755-2721/132/2024.20577

required information for each video of Khan Academy, we utilized the YoutubeAPI service provided by Google Cloud. Each video was extracted with its title, URL, and description. ChatGPT cooperates with the database establishment, by providing a generalized keyword for each video based on its title and description. Thus far, we have established a custom database with reliable video resources, each labeled with an AI-generated keyword. The figure below demonstrates the ChatGPT interaction to generate keywords for each video resource.

```python
def generate_keyword_for_video(title, description):

    # OpenAI prompt to generate a keyword for each extracted video information
    prompt = (
        f"Generate a relevant keyword for a video titled '{title}' with the following description: '{description}'. "
        "What you need to do is find the correct word inside the topic that describes the video, and usually the keyword "
        "lies before the word 'Khan Academy'. For example, Molecular structure of triglycerides (fats) | Biology | Khan Academy, "
        "with this topic you should return 'Biology'. Remember, only return the keyword, nothing else. Moreover, the keyword "
        "doesn't need to be only one word. For example, in the topic 'Squeeze theorem exercise example | Limits | Differential Calculus | Khan Academy', "
        "you should return 'Differential Calculus' as the keyword—don't only return half of it; this applies to other situations. "
        "If the topic doesn't fit the format, try to generate one on your own."
    )
    print(title)

    response = client.chat.completions.create(
        model="gpt-4o",
        messages=[
            {"role": "system", "content": "You are a helpful assistant."},
            {"role": "user", "content": prompt}
        ],
        max_tokens=10,
        temperature=0.3
    )

    keyword = response.choices[0].message.content.strip()
    print(keyword)
    print("\n")
    return keyword
```

Figure 8. Method to generate keywords for each video resource using ChatGPT.

4.2.2. ChatGPT Interaction with Database

With the custom database established as the trusted resource for Ligare, how to enable ChatGPT to interact with the database becomes a critical issue that needs to be addressed. An interaction flow was designed to facilitate the cooperation between ChatGPT and the program with the database, as demonstrated in the figure below.

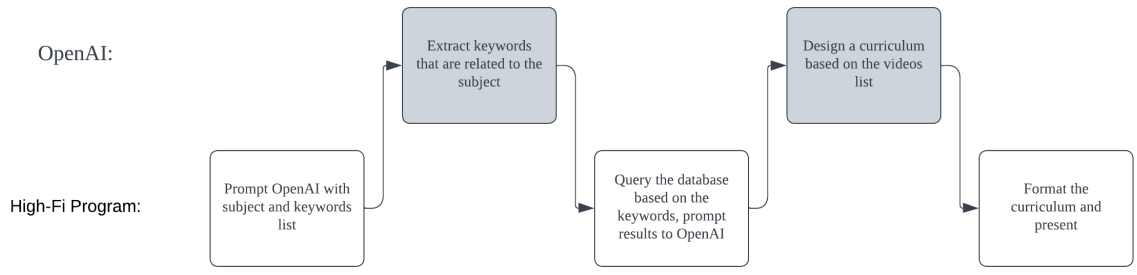


Figure 9. Interaction flow between ChatGPT and the database.

In order to solve the problem of finding the related topics based on the subject given by the user, we utilize the functionality of ChatGPT, which will be responsible for selecting a list of keywords that are closely related to the given subject within a keyword set. The results returned by ChatGPT will be used to query the custom database and extract resources that are labeled with one of the selected keywords. These extracted resources will construct the background information for ChatGPT to generate the final curriculum, based on the user's requirements. Inside the prompt to generate the curriculum, some instructions are added to achieve a better curriculum, for example, requiring ChatGPT to generate the

DOI: 10.54254/2755-2721/132/2024.20577

curriculum with an order of difficulty. Moreover, for each video selected, a question will be generated according to the video topic. The following code serves the functionality to generate the curriculum based on the extracted video resources from the database.

```python
def generate_learning_plan_with_questions(videos: list[dict]):
    """
    Generates a study plan using ChatGPT based on the extracted videos.
    ChatGPT selects 5 videos and organizes them from easy to difficult.
    Each day will include a video title, URL, and a short question.
    """
    video_info = "\n".join([f"{i+1}. {video['title']} ({video['url']})" for i, video in enumerate(videos)])

    plan_prompt = (
        f"Here is a list of videos with their titles and URLs:\n{video_info}\n\n"
        f"Please select 5 of these videos and organize them from easy to difficult. For each day, provide the following:\n"
        f"1. Video Title on the first line\n"
        f"2. Video URL on the second line\n"
        f"3. A short question based on the video's topic on the third line.\n\n"
        f"Please respond exactly in this format:\n"
        f"Day N\nVideo Title\nVideo URL\nQuestion\n\n"
        f"Where N is the day number."
    )

    response = client.chat.completions.create(
        model="gpt-4o",
        messages=[
            {"role": "system", "content": "You are a helpful assistant."},
            {"role": "user", "content": plan_prompt}
        ],
        max_tokens=1000  # Increased the limit to ensure enough space for questions
    )

    study_plan = response.choices[0].message.content.strip()

    if not study_plan:
        raise HTTPException(status_code=500, detail="Failed to generate study plan")

    # Split the response into blocks for each day
    plan_blocks = study_plan.split("\n\n")  # Each block corresponds to a day's plan

    formatted_plan = []

    for block in plan_blocks:
        day_lines = block.split("\n")
        if len(day_lines) >= 3:
            title, url, question = day_lines[1:4]  # We expect first line to be "Day N", then title, URL, and question
            formatted_plan.append(title)
            formatted_plan.append(url)
            formatted_plan.append(question)

    # Ensure we have 15 strings (5 days, 3 strings per day)
    if len(formatted_plan) < 15:
        raise HTTPException(status_code=500, detail="Incomplete study plan generated")

    return formatted_plan
```

Figure 10. Method to prompt ChatGPT to generate the curriculum based on the selected video list.

It is worth noting that within each step of interaction with ChatGPT, the response format needs to be strictly restricted to allow the backend program to interpret the result. However, this warrants careful consideration and caution since those soft restrictions provided inside the prompt are not enforced, leaving for the possibility that ChatGPT does not respond with the required format, leading to errors in interpreting the results and break down of the program.

DOI: 10.54254/2755-2721/132/2024.20577

4.3. High-Fidelity Prototype Evaluation

In Table 3, we present the results of ChatGPT and Ligare based on three key metrics: whether the output met user expectations, whether the AI continued asking questions without generating a concrete plan, and whether the generated learning plan was reasonable and accurate. We used user feedback to evaluate each model's performance. Only when the output meets the user's expectations, and the plan is both clear and aligned with the user's goals, do we consider the output successful. At the end of the table, we summarized the number of successful solutions for each model.

We observed that Ligare consistently outperforms ChatGPT in all three dimensions. Users reported that Ligare generated learning plans that closely matched their expectations, without unnecessary follow-up questions. By contrast, ChatGPT often required additional clarifications and failed to provide structured learning plans that met the users' needs.

Table 3. Results of ChatGPT and Ligare.

User ID	Model	Did the output meet user expectations?	Did the AI continue asking questions?	Was the learning plan reasonable and accurate?
User 1	ChatGPT	No (3 points)	Yes	Yes
User 1	Ligare	Yes (5 points)	No	Yes
User 2	ChatGPT	No (2 points)	No	Yes
User 2	Ligare	Yes (4 points)	No	No
User 3	ChatGPT	Yes (4 points)	Yes	No
User 3	Ligare	Yes (5 points)	No	Yes
User 4	ChatGPT	No (3 points)	No	No
User 4	Ligare	Yes (5 points)	No	Yes
User 5	ChatGPT	No (2 points)	Yes	No
User 5	Ligare	Yes (4 points)	No	Yes

As demonstrated in Table 3, Ligare generated learning plans that consistently met users' expectations across all tested cases. In particular, it avoided the issue of excessive questioning before producing a plan, a common problem encountered with ChatGPT. Users reported that Ligare produced clear, structured plans that aligned with their learning goals, while ChatGPT often failed to do so.

In conclusion, the results highlight the strengths of Ligare in terms of output accuracy, user experience, and plan structure. These advantages make Ligare a more reliable and user-friendly solution compared to ChatGPT in generating personalized learning plans.

4.4. Future Development

According to the high-fidelity prototype evaluation results, the prototype achieves the main functionality of Ligare and performs better compared to the ChatGPT response on the generated responses and the user experience. However, the interactions with ChatGPT still remain somewhat rudimentary and unstable. Better restrictions within the prompts generated by ChatGPT and more precise interpretations of the responses may help construct a more reliable connection between the program and AI. Moreover, AI's comprehension ability is still at its early stage, which limits the effectiveness of its functionality, for example, it struggles to understand the difficulty progression between each topic. At the current stage, manual classification helper functions might be able to resolve the confusion of ChatGPT.

In general, further progress to stabilize the interaction with ChatGPT and enhance its effectiveness needs to be made.

5. Conclusion

The authors of this paper introduce Ligare, an application developed to improve the generation of personalized learning plans compared to standard ChatGPT outputs. By refining the output format and minimizing irrelevant queries, Ligare consistently delivered learning plans that aligned with user

DOI: 10.54254/2755-2721/132/2024.20577

expectations. Through user testing and comparative evaluations, it was shown to provide more accurate, structured, and effective learning experiences. Although the scope of the project is limited to mathematics, the methodology can be applied across different subject areas, highlighting its potential for broader use. As a pioneering AI curriculum builder for students, Ligare exemplified the future possibilities of AI-powered education tools. By offering a more streamlined and efficient approach to personalized education, AI holds significant promise in helping to bridge the achievement gaps that have persisted in our society.

Acknowledgement
Yahan Cao, Chang Gu, Yixuan Li, and Yanze Lyu contributed equally to this work and should be considered co-first authors.

References

[1] Bourdieu, Pierre. "The forms of capital.(1986)." *Cultural theory: An anthology* 1, no. 81-93 (2011): 949.

[2] Broer, Markus, Yifan Bai, Frank Fonseca, Markus Broer, Yifan Bai, and Frank Fonseca. "A review of the literature on socioeconomic status and educational achievement." *Socioeconomic inequality and educational outcomes: Evidence from twenty years of TIMSS* (2019): 7-17.

[3] Assessment, NAEP NAEP Long-Term Trend. "Results: Reading and Mathematics." *The Nation's Report Card* (2022).

[4] Betthäuser, Bastian A., Anders M. Bach-Mortensen, and Per Engzell. "A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic." *Nature human behaviour* 7, no. 3 (2023): 375-385.

[5] Golden, Alexandrea R., Emily N. Srisarajivakul, Amanda J. Hasselle, Rory A. Pfund, and Jerica Knox. 2023. "What Was a Gap Is Now a Chasm: Remote Schooling, the Digital Divide, and Educational Inequities Resulting from the COVID-19 Pandemic." *Current Opinion in Psychology* 52 (August): 101632. https://doi.org/10.1016/j.copsyc.2023.101632.

[6] Kuhfeld, Megan, and Karyn Lewis. 2022. "Student Achievement in 2021-22: Cause for Hope and Continued Urgency." NWEA. https://www.nwea.org/uploads/2022/07/Student-Achievement-in-2021-22-Cause-for-hope-and-concern.researchbrief-1.pdf.

[7] Panagouli, Eleni, Androniki Stavridou, Christina Savvidi, Anastasia Kourti, Theodora Psaltopoulou, Theodoros N. Sergentanis, and Artemis Tsitsika. "School performance among children and adolescents during COVID-19 pandemic: a systematic review." *Children* 8, no. 12 (2021): 1134.

[8] Wu, Tianyu, Shizhu He, Jingping Liu, Siqi Sun, Kang Liu, Qing-Long Han, and Yang Tang. 2023. "A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development." *IEEE/CAA Journal of Automatica Sinica* 10 (5): 1122–36. https://doi.org/10.1109/jas.2023.123618.

[9] Islam, Raisa, and Owana Moushi. 2024. "EasyChair Preprint GPT-4o: The Cutting-Edge Advancement in Multimodal LLM GPT-4o: The Cutting-Edge Advancement in Multimodal LLM." https://easychair.org/publications/preprint/z4TJ/open.

[10] Zhang, Ke, and Ayse Begum Aslan. 2021. "AI Technologies for Education: Recent Research & Future Directions." *Computers and Education: Artificial Intelligence* 2 (100025): 100025. https://doi.org/10.1016/j.caeai.2021.100025.

[11] Pang, Samarnh, Engheang Nol, and Kimkong Heng. 2024. "ChatGPT-4o for English Language Teaching and Learning: Features, Applications, and Future Prospects." *Social Science Research Network*, January. https://doi.org/10.2139/ssrn.4837988.

[12] Wei, Xin. 2024. "Evaluating ChatGPT-4 and ChatGPT-4o: Performance Insights from NAEP Mathematics Problem Solving." *Frontiers in Education* 9 (September). https://doi.org/10.3389/feduc.2024.1452570.