

PAPER

A Lightweight LINE-Based Knowledge Chatbot for Mobile Technical Certification Learning: An Empirical Study with ChatGPT as a Baseline

Cheng-Hsiu Li  (✉)

National Taitung Junior
College, Taitung,
Taiwan, R.O.C.

lcs@ntc.edu.tw**ABSTRACT**

This study developed a lightweight LINE-based mobile knowledge chatbot for the Level B Computer Hardware Repair certification subject and examined its instructional effectiveness in comparison with ChatGPT. The system was deployed through the LINE mobile messaging platform, allowing learners to access certification-related explanations on their own smartphones through natural-language interaction. Unlike large language model (LLM)-based systems that rely on retrieval-augmented generation (RAG), vector embedding, and index construction, the proposed chatbot was built around a Knowledge Base Recommendation Algorithm (KBRA), which generates instructional responses through structured keyword matching and fuzzy search over a domain-specific knowledge base. This mobile deployment was designed to provide a low-cost, easily accessible, and interactive learning support tool for bounded technical certification domains. A nonequivalent pretest-posttest quasi-experimental design was adopted. The experimental group used the LINE-based KBRA chatbot for mobile self-directed learning, whereas the control group used ChatGPT. Both groups participated in an eight-week intervention and completed pre-test and post-test assessments consisting of 10 open-ended questions. The results showed that both groups achieved substantial learning gains. After controlling for pre-test scores, analysis of covariance (ANCOVA) revealed no significant difference in post-test performance between the two groups, and the effect size associated with group membership was negligible. In addition, the two one-sided tests (TOST) equivalence analyses indicated that the learning gains of the two groups were practically equivalent within a ± 10 -point margin. These findings suggest that, in this bounded certification-learning context, the LINE-based KBRA chatbot may provide learning support comparable to ChatGPT under the conditions examined. However, this comparability should be interpreted cautiously because the study was conducted with a small sample, a single certification domain, and a specific eight-week self-directed learning setting. The proposed chatbot should therefore be regarded as a feasible and deployment-friendly learning support option for structured technical certification learning, rather than as a general alternative to ChatGPT.

Li, C-H. (2026). A Lightweight LINE-Based Knowledge Chatbot for Mobile Technical Certification Learning: An Empirical Study with ChatGPT as a Baseline. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(18), pp. 23–45. <https://doi.org/10.3991/ijim.v20i18.62472>

Article submitted 2026-05-27. Revision uploaded 2026-06-29. Final acceptance 2026-06-29.

© 2026 by the authors of this article. Published under CC-BY.

KEYWORDS

knowledge-based chatbot, LINE-based chatbot, mobile learning, technical certification learning, ChatGPT, learning effectiveness, equivalence analysis

1 INTRODUCTION

Since the 1970s, behaviorism has been widely recognized in academia as a theoretical framework capable of explaining learning processes related to knowledge acquisition and memory [1]. Drill-and-practice digital learning, grounded in behaviorist principles, is a widely adopted instructional approach, particularly suited for test bank-based cognitive learning. This approach is built on the stimulus–response (S–R) association model, in which repeated exposure to stimuli and corresponding responses are reinforced through immediate feedback. Positive reinforcement is provided through praise and rewards, while punishment suppresses incorrect responses (law of effect). The closer the feedback follows the response, the more effective the reinforcement (law of contiguity). Additionally, mastery is achieved through constant repetition and practice, emphasizing the strengthening of S–R connections (law of exercise) [2].

Drill-and-practice digital learning is widely adopted in the theoretical component of technical certification examinations to support classroom instruction. Through S–R–feedback learning cycles, students can repeatedly practice large-scale item banks to reinforce knowledge recall. The integration of gamification elements has been shown to reduce the monotony of repetitive training, improve intrinsic motivation for self-directed learning, and thereby enhance learning performance [3]–[6]. However, drill-and-practice approaches primarily promote memorization of question banks rather than fostering an understanding of the underlying concepts and meanings of the items.

Generative chatbots (e.g., ChatGPT and Gemini) are rapidly emerging as important instructional tools for supporting subject-matter learning in professional and technical domains. Regarding learning effectiveness, multiple empirical studies have shown that students who interact with chatbots demonstrate better academic performance than control groups engaging solely with teacher-led instruction, and they also exhibit positive attitudes toward AI-assisted learning [7]–[10]. In terms of personalized and self-directed learning, generative AI can automatically provide diverse learning materials, practice exercises, and real-time feedback tailored to individual learner needs, thereby enhancing self-assessment and learning motivation [9], [11]–[13]. For innovation in assessment and instructional design, AI enables automated grading, the generation of multiple assessment formats, and supports teachers in designing personalized and interactive coursework [9], [10], [13], [14]. However, due to the risks of factual inaccuracies, biased content, and inappropriate pedagogical use during implementation, careful teacher supervision and pedagogical guidance remain essential [15]–[17], [29]. These findings suggest that generative chatbots have considerable potential for supporting self-directed and domain-specific learning, but their use in certification preparation still requires attention to accuracy, response control, and implementation conditions [24], [27].

In technical certification examinations, generative AI can assist candidates in practicing item banks, providing instant responses to subject-related inquiries, and improving learning efficiency across multiple domains [18]. Moreover, AI-assisted learning has been shown to enhance learners’ motivation, knowledge mastery, and exam confidence, making it particularly suitable for self-directed and distance learning contexts [19]. Nevertheless, when certification knowledge is bounded and highly structured, it remains necessary to examine whether simpler and more transparent

chatbot approaches, including rule-based or knowledge-based tutoring systems, can provide comparable learning support with lower implementation costs [24], [26].

Retrieval-augmented generation (RAG) is an artificial intelligence technique that integrates the generative capabilities of large language models (LLMs) with information retrieval mechanisms over external knowledge sources, such as documents or databases. By retrieving relevant information prior to response generation, RAG enables AI systems to produce more accurate, specific, and up-to-date answers. This approach can mitigate common limitations of LLMs, including outdated knowledge and factually inconsistent responses, making it suitable for domain-specific applications. The construction of a RAG vector database typically consists of four major stages: document parsing and cleaning, text chunking, embedding generation, and index construction. These stages require substantial computational resources, including CPU (or GPU) processing power, system memory, and disk I/O. In educational settings with limited computational resources, these preprocessing, embedding, and indexing requirements may create practical barriers to system construction, maintenance, and updating. Prior studies on vector search and indexing further show that retrieval infrastructures involve important trade-offs among construction cost, retrieval efficiency, scalability, and accuracy. Hierarchical Navigable Small Worlds (HNSW) constructs its index incrementally by inserting vectors into a multi-layer proximity graph, which introduces additional construction overhead due to neighbor selection and multi-layer maintenance. However, the absence of training or coarse search structures and the possibility of parallel implementation make the construction cost manageable and suitable for offline indexing scenarios [20]. By optimizing k -selection and fully exploiting GPU parallelism, the proposed approach significantly reduces vector index construction time, enabling billion-scale k -NN graph construction within practical offline time constraints while maintaining high retrieval accuracy [21]. Milvus is designed to efficiently manage large-scale and dynamic vector data by supporting incremental index construction, heterogeneous CPU–GPU acceleration, and distributed indexing, thereby reducing the need for expensive full reindexing while maintaining scalable performance [22]. ANN-Benchmarks systematically demonstrate that index construction time is a critical yet algorithm-dependent factor in approximate nearest neighbor (ANN) search, revealing inherent trade-offs between construction overhead, retrieval accuracy, and query efficiency and underscoring the importance of offline indexing in large-scale vector search systems [23]. Taken together, these studies indicate that vector indexing and retrieval infrastructures can be powerful for large-scale AI applications, but they also involve construction, indexing, and maintenance considerations that may be disproportionate for small, bounded certification-learning domains. This motivates the present study's focus on a lightweight knowledge-base-centered chatbot that avoids embedding generation and vector index construction while supporting mobile, domain-specific learning interaction [25], [26], [30], [31].

Recent work on educational chatbots and mobile learning has also emphasized their value in supporting flexible access, immediate interaction, and learner-centered guidance in higher education contexts [25], [28], [30], [31]. In particular, recent bibliometric evidence indicates that mobile learning has become an increasingly important research direction in higher education, with growing attention to flexibility, accessibility, and technology-mediated learning environments [28]. This study takes the Level B Computer Hardware Repair certification subject as a case example and aims to develop a lightweight knowledge-based chatbot for supporting learners' conceptual understanding of technical certification content. Although the integration of LLMs with RAG enables natural language interaction with learners and provides domain-specific expertise, such approaches are associated with substantial computational and maintenance costs. In particular, the runtime requirements of LLMs,

together with the construction time and indexing overhead of vector databases, impose significant demands on hardware resources. Moreover, system responsiveness is highly dependent on the specifications of the underlying computing platform, and such systems may still generate unsupported or factually inconsistent responses.

To address these limitations, this study developed a lightweight LINE-based mobile knowledge chatbot for supporting technical certification learning. The system was designed for smartphone-based access through the LINE mobile messaging platform, allowing learners to ask certification-related questions and receive immediate instructional responses during self-directed learning. Rather than relying on LLM–RAG integration, the chatbot uses a structured knowledge base and a Knowledge Base Recommendation Algorithm (KBRA) to retrieve and deliver domain-specific explanations through conversational interaction. In this sense, the study is positioned within the field of interactive mobile learning technologies: it examines how a lightweight mobile chatbot can provide accessible, low-resource, and domain-bounded learning support in a technical certification context.

The objectives of this study were threefold: (1) to design and implement a LINE-based mobile knowledge chatbot using a structured knowledge base and KBRA; (2) to evaluate whether learners using the mobile KBRA chatbot achieved learning outcomes comparable to those using ChatGPT; and (3) to examine the practical implications and limitations of using lightweight mobile chatbot systems for bounded technical certification learning.

2 MATERIALS AND METHODS

2.1 System architecture and mobile deployment

The proposed chatbot was implemented as a LINE-based mobile learning system to support self-directed technical certification learning. LINE was selected as the user-facing platform because it is a widely used mobile messaging application in Taiwan and allows learners to interact with the chatbot through a familiar smartphone-based interface. Learners did not need to install a separate educational application; instead, they accessed the chatbot by adding the LINE Bot as a learning contact and submitting questions through ordinary text messages. The overall system architecture and learner interaction flow are shown in Figure 1.

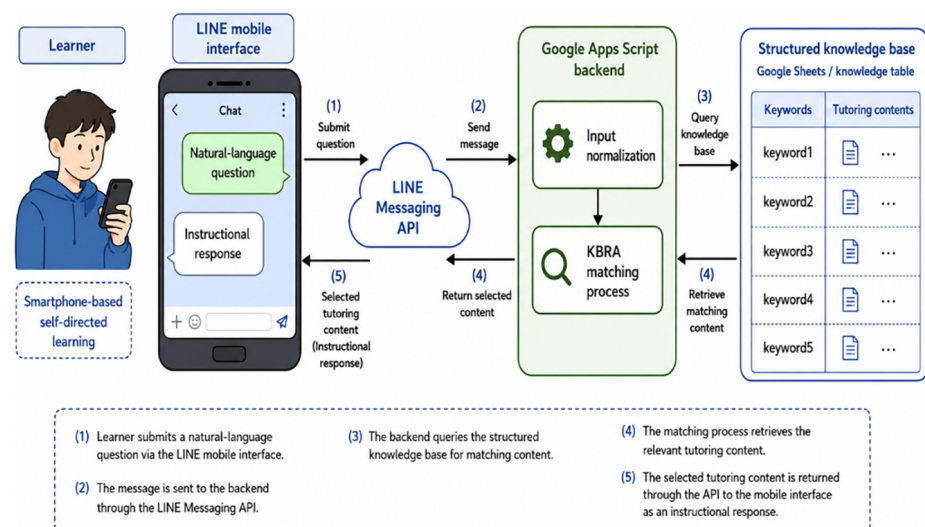


Fig. 1. Architecture and interaction flow of the LINE-based mobile KBRA chatbot

The system consisted of three main components: (1) the LINE mobile user interface, through which learners entered natural-language questions and received instructional responses; (2) a Google Apps Script backend, which received user messages through the LINE Messaging API and executed the KBRA matching procedure [5]; and (3) a structured knowledge base, implemented as a table containing keyword fields and corresponding tutoring contents. As shown in Figure 1, when a learner submitted a question through the LINE mobile interface, the message was first sent to the backend service through the LINE Messaging API. The backend then normalized the input, queried the structured knowledge base, and executed the KBRA matching procedure to identify the most relevant knowledge entry. Finally, the selected tutoring content was returned to the learner through the LINE chat interface as an instructional response.

The mobile interaction design emphasized accessibility, immediacy, and low deployment cost. Learners could ask questions at any time during the self-directed learning period, including outside the classroom, and receive immediate responses through the same mobile communication environment they used in daily life. The chatbot interaction was therefore not limited to a desktop computer laboratory or a fixed learning management system. This design was intended to support flexible mobile learning, particularly for students preparing for technical certification examinations who may need repeated access to short explanations while reviewing question-bank content.

2.2 Materials

The KBRA was used in this study as a lightweight rule-based recommendation procedure for selecting tutoring content from a structured domain-specific knowledge base. It was not intended to replace more advanced retrieval or generative AI methods; rather, it was designed to provide a transparent, easily maintainable, and low-resource mechanism suitable for bounded certification-learning content.

To ensure that the subject-matter chatbot can accurately respond to users' open-ended queries, the system incorporates a KBRA designed on keyword matching and fuzzy search mechanisms. The core logic of the system is grounded in a structured knowledge base, where each entry contains five keyword fields (keyword1 to keyword5) and a corresponding instructional response.

When a user submits a query, U , the system first normalizes the input by converting all English text to lowercase, thereby eliminating case-sensitivity issues during keyword matching. Subsequently, for each knowledge base entry B_i , the system retrieves its associated keyword set

$$K_i = \{k_{i1}, k_{i2}, k_{i3}, k_{i4}, k_{i5}\},$$

and computes a relevance score S_i between the entry and the user input. The scoring procedure is defined as follows:

$$S_i = \sum_{j=1}^5 \delta(k_{ij}, U) \quad (1)$$

The indicator function is defined as follows:

$$\delta(k, U) = \begin{cases} 1, & \text{if } k \in U (\text{case-insensitive and } k \neq \emptyset) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

In other words, the system performs a fuzzy match between each non-empty keyword and the user input U . Every occurrence of a keyword contributes one point to the relevance score. This matching mechanism allows users to phrase their queries in natural language while enabling the system to perform tolerant, partial-keyword matching to improve flexibility and robustness.

After computing all relevance scores, the system identifies the entries that achieve the highest score. Let S_{max} denote the maximum score, and define the candidate set as:

$$C = \{B_i \mid S_i = S_{max}, S_i > 0\} \quad (3)$$

If multiple entries share the same highest score, the system randomly selects one from the candidate set as the final response:

$$B^* \sim \text{Uniform}(C) \quad (4)$$

The final output returned to the user is therefore:

$$\text{Response} = B^* . \text{tutoring_contents} \quad (5)$$

If no entry obtains a positive relevance score (i.e., $S_i = 0$ for all i), the system returns a default message informing the user: "Sorry, I am still learning. I will record your question and try to answer it next time!".

The pseudocode for the proposed KBRA is presented below:

```

Function handleMessage(userInput):
  totalbank ← get_rows_knowledge_bank_main()
  message ← "Sorry, I am still learning. I will record your question and try to answer it next time!"
  maxScore ← 0
  topMatches ← empty list
  userInput ← userInput.toLowerCase()
  For each bankdata in totalbank:
    score ← 0
    keywords ← [bankdata.keyword1, keyword2, keyword3, keyword4, keyword5]
    For each keyword in keywords:
      If keyword is not null/undefined/empty:
        keywordLower ← keyword.toLowerCase()
        If userInput contains keywordLower:
          score ← score + 1
    If score > maxScore:
      maxScore ← score
      topMatches ← [bankdata]
    Else if score == maxScore and score > 0:
      Append bankdata to topMatches
  If topMatches is not empty:
    randomIndex ← random number between 0 and length(topMatches)-1
    message ← topMatches[randomIndex].tutoring_contents
  replyMessage ← {
    "type": "text",
    "text": message
  }
  lineReplyMessage(replyToken, [replyMessage])

```

For KBRA, let N denote the number of entries in the knowledge base and $|U|$ the length of the user query. Since each knowledge base entry contains a fixed number

of keywords ($K = 5$), the relevance computation for a single entry requires at most $O(|U|)$ time for fuzzy keyword matching. The algorithm scans the entire knowledge base once to compute relevance scores, resulting in an overall time complexity of $O(N|U|)$. As the keyword dimensionality is constant, the computational cost scales linearly with the knowledge base size and the input length. This lightweight complexity enables efficient real-time responses without requiring high-dimensional vector indexing or large-scale model inference.

Through the aforementioned matching mechanism, the system incorporates fuzzy-search capabilities and a diverse response strategy. The verification results of the open-ended question-answering function are shown in Figure 2. When students encounter difficulties with items in the question bank, they may pose their questions to the subject-matter chatbot. The chatbot then applies the KBRA to suggest the most relevant answer.

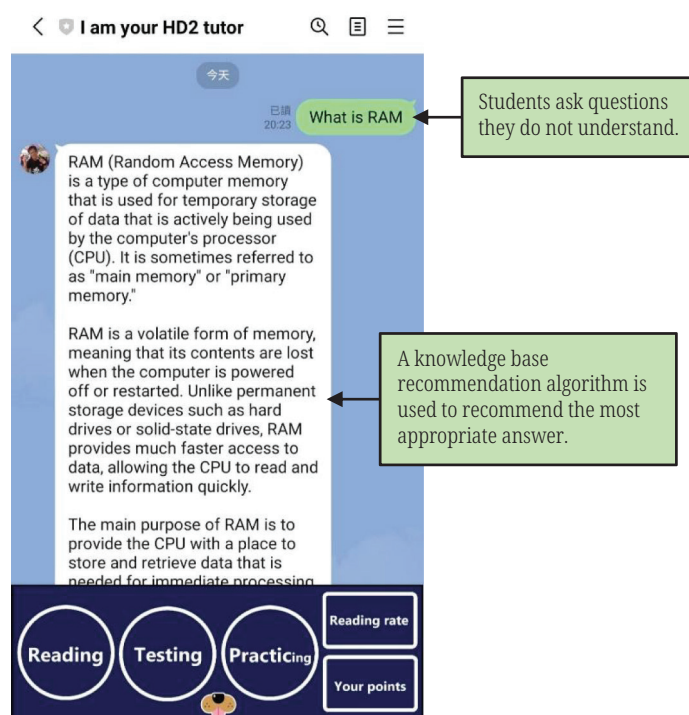


Fig. 2. Verification of the KBRA function

Next, a supplementary expert-scored response-quality comparison was conducted to compare the responses generated by the KBRA-based chatbot and ChatGPT within the Level B Computer Hardware Repair certification domain. This comparison was not intended to serve as a general benchmark of chatbot performance, nor was ChatGPT treated as the ground truth. Instead, the purpose was to provide a limited, domain-specific comparison of response quality based on expert judgment, focusing on conceptual accuracy, completeness, relevance to the expected certification content, and the presence of unsupported or unnecessary elaboration. The scope of this comparison was limited to the Level B Computer Hardware Repair certification question bank, within which 10 open-ended questions were designed for testing. The same set of prompts was submitted to both systems. ChatGPT was accessed through the standard web interface available at the time of the study, and no external files, customized knowledge sources, or additional retrieval tools

were used. Because the exact backend model version was not fixed or controlled by the researchers, the study reports the ChatGPT access condition rather than treating a specific model version as an experimentally controlled variable. The test responses were evaluated by a domain expert based on the accuracy of key concepts, completeness of explanation, relevance to the expected certification content, and presence of unsupported or unnecessary elaboration. Each item was scored on a 10-point scale, yielding a maximum possible score of 100. The scoring results are presented in Table 1.

Table 1. Expert-scored response-quality comparison between the KBRA-based chatbot and ChatGPT on a 100-point scale

No.	Topic Category	KBRA-Based Chatbot Score	ChatGPT Score	Key Remarks
1	Definition of Karnaugh Map	95	95	Both systems accurately explained the role of Gray code and the mechanism of Boolean logic simplification.
2	Optical Disc Data Transfer Rate	100	95	The KBRA-based chatbot provided a concrete numerical example, whereas ChatGPT presented a general formula.
3	Series Resistance Calculation	100	100	Both systems correctly applied the additive rule for total resistance (R_t).
4	Soldering Temperature	90	90	Both systems correctly indicated that a 30W soldering iron can provide a stable temperature of approximately 300–350°C.
5	Odd Parity Check	95	98	ChatGPT provided a more detailed explanation of the conditions for assigning 0s and 1s.
6	One's Complement	90	92	Both systems described bitwise inversion, while ChatGPT further elaborated on the concept of two's complement.
7	PCB Component Labeling	100	100	Both systems correctly identified that the prefix "R" denotes a resistor.
8	IDE Ribbon Cable Standard	95	95	The KBRA-based chatbot accurately defined the 40-pin standard.
9	BASIC Operator Precedence	85	95	ChatGPT provided a more comprehensive hierarchy, including modulus (Mod) and integer division.
10	Ripple Counter Design	100	100	Both systems correctly derived the condition ($2n \geq 5$).

Figure 3 shows the expert-scored response-quality scores of the KBRA-based chatbot and ChatGPT. The supplementary comparison results indicate that both the KBRA-based chatbot and ChatGPT demonstrated strong performance across the 10 evaluation items, suggesting that both systems were capable of providing generally accurate responses to foundational topics in computer hardware and electronics within this bounded certification domain. Based on the scores presented in the table, the KBRA-based chatbot achieved a total score of 950, corresponding to an average score of 95, whereas ChatGPT obtained a total score of 960, corresponding to an average score of 96. The difference between the two systems was therefore minimal, indicating that the KBRA-based chatbot performed at a level highly comparable to that of ChatGPT in this limited expert-scored response-quality comparison.

However, this comparison should be interpreted as a supplementary expert-scored response-quality analysis rather than as a general benchmark of chatbot performance. Because the comparison was based on 10 prompts and one domain expert rater, it does not provide evidence for broader claims about chatbot reliability across different domains, tasks, or system conditions.

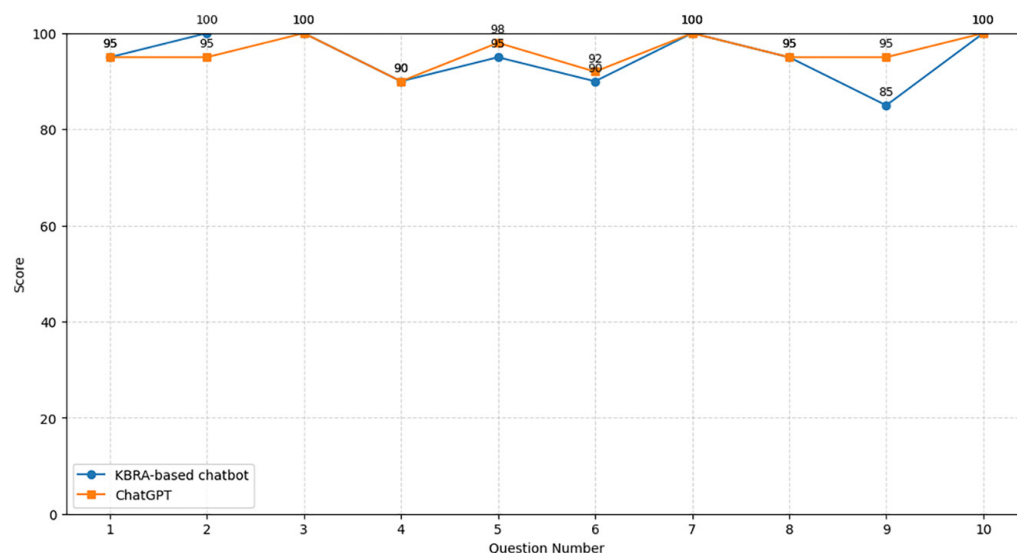


Fig. 3. Expert-scored response-quality scores of the KBRA-based chatbot and ChatGPT

A closer examination of the item-level results shows that the two systems produced identical scores on several questions, including Items 1, 3, 4, 7, 8, and 10. In particular, both systems received full marks on Items 3, 7, and 10, which suggests that they were equally effective in handling questions involving well-defined concepts, straightforward rules, or fixed computational logic. This pattern implies that, for structured knowledge-based questions with relatively unambiguous answers, the KBRA-based chatbot can achieve a level of accuracy comparable to that of an LLM such as ChatGPT.

The differences between the two systems were mainly observed in questions requiring more elaborated conceptual explanation. For example, in Item 5 (odd parity checking), ChatGPT received a slightly higher score because its explanation of the conditions for assigning 0s and 1s was more detailed. A similar pattern was observed in Item 6 (one's complement), where both systems correctly described bit inversion, but ChatGPT further extended the explanation by referring to the concept of two's complement. The largest score gap appeared in Item 9 (BASIC operator precedence), in which ChatGPT provided a more comprehensive account of the hierarchical structure of operators, including modulus and integer division. These results suggest that ChatGPT showed a relative advantage in questions that benefited from broader contextual explanation or additional instructional elaboration.

At the same time, the KBRA-based chatbot also showed competitive performance and even outperformed ChatGPT on Item 2 (optical disc data transfer rate), where it received a higher score due to its inclusion of a concrete calculation example. This result suggests that when a question corresponds closely to a well-defined knowledge entry or procedural example, the KBRA-based chatbot can provide responses that are not only accurate but also closely aligned with expected evaluation criteria.

Overall, the findings suggest that the KBRA-based chatbot has substantial potential as a domain-specific instructional support tool. Although ChatGPT showed

a slight advantage on several items requiring more detailed or extended explanations, the overall difference between the two systems was very small. The results therefore support a cautious interpretation: within this limited expert-scored comparison and bounded certification-learning domain, the KBRA-based chatbot produced response-quality scores close to those of ChatGPT. This finding suggests practical comparability under the specific conditions examined, particularly when the domain knowledge can be explicitly organized and systematically retrieved, but it does not imply general equivalence across broader domains or open-ended learning tasks.

2.3 Methods

The research framework of this study is illustrated in Figure 4. A nonequivalent pre- and post-test quasi-experimental design was adopted, comprising an experimental group and a control group. Because the study was conducted in an authentic classroom teaching context, random assignment at the individual level was not feasible. Therefore, participants were assigned to the two groups using convenience sampling based on intact class arrangements. The experimental group consisted of 26 participants (23 males and 3 females), while the control group included 25 participants (22 males and three females). The participants were students enrolled in the 2025 academic year and participated in technical certification-related learning activities. In terms of academic background, approximately four-fifths of the participants came from information/electrical-related fields, whereas approximately one-fifth came from business-related fields. Most participants had basic prior exposure to computer hardware concepts through coursework or certification preparation, but they had not previously received systematic learning support from the proposed KBRA-based chatbot. All participants were fully informed of the experimental procedures and provided written informed consent prior to participation. This study was conducted as part of the MOE Teaching Practice Research Program (Project No. PED1143504) in an authentic educational setting and involved minimal risk to the participants. No personally identifiable information was collected, and all data, including pre-test scores, post-test scores, and system usage logs, were analyzed anonymously and reported only in aggregated form. According to the institutional guidelines of National Taitung Junior College, formal institutional ethics approval was not required for this type of minimal-risk educational research. The study was conducted in accordance with relevant institutional guidelines and regulations.

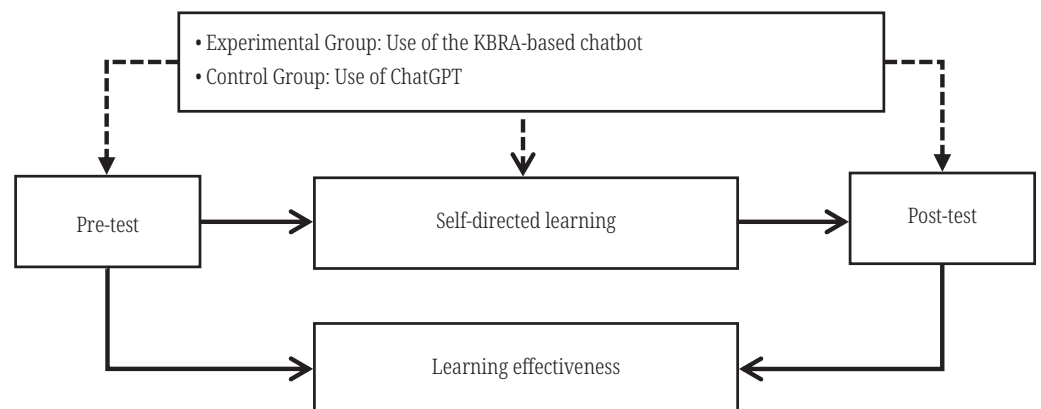


Fig. 4. Research framework of the study

To improve assessment transparency, the full pre-test and post-test questions are provided in Appendix A. The pre- and post-test consisted of 10 open-ended questions designed to assess learners' conceptual understanding of the Level B Computer Hardware Repair certification content. The two tests used different but content-equivalent items. Specifically, the items were constructed to cover the same major knowledge categories, including Karnaugh maps, optical disc data transfer rate, Boolean algebra, soldering knowledge, parity checking, one's and two's complement, IDE/SATA cable standards, BASIC operator precedence, and ripple counter design. Although the wording and numerical values differed between the pre- and post-test, the two instruments were designed to assess comparable conceptual knowledge and procedural reasoning within the same certification question-bank scope.

Each item was scored on a 10-point scale, yielding a maximum possible score of 100. A predefined scoring rubric was used for all open-ended responses. For each item, 4 points were assigned to correctness of key concepts or procedures, 3 points to completeness of the explanation or calculation process, and 3 points to relevance and clarity in relation to the expected certification content. Full credit was awarded when a response accurately addressed the required concept or procedure, provided a sufficiently complete explanation or calculation, and remained directly relevant to the question. Partial credit was awarded when a response was conceptually relevant but incomplete or only partially correct. Little or no credit was awarded when a response was incorrect, irrelevant, or failed to address the expected concept. To maintain scoring consistency, all pre- and post-test responses were evaluated by the same instructor using the same rubric. Because only one domain expert was involved in the scoring process, inter-rater reliability could not be calculated; this issue is acknowledged as a limitation of the study.

Participants in the experimental group engaged in self-directed learning using the KBRA-based chatbot proposed in this study, whereas those in the control group used ChatGPT as the baseline learning tool. Both groups followed the same learning goal, studied the same certification question-bank scope, and completed the same self-directed learning period. Participants in the control group were instructed to use ChatGPT for self-directed learning related to the Level B Computer Hardware Repair certification content during the intervention period. No fixed prompt template was imposed, as the study aimed to reflect authentic self-directed learning behavior in a natural educational setting. Participants were allowed to interact with ChatGPT using natural-language questions in a manner consistent with ordinary self-directed learning practice. No additional external tools, uploaded documents, or customized knowledge sources were integrated into the ChatGPT condition. For the experimental group, system usage logs were collected through the Google Apps Script server-side dataset, including user query records and chatbot response records generated during the use of the LINE-based KBRA chatbot. These logs were used primarily to verify system use during the intervention; detailed behavioral log analysis was beyond the scope of the present study. Prior to the intervention, all participants completed a pre-test. This was followed by an eight-week period of self-directed learning, after which a post-test was administered. Subsequently, learning effectiveness was analyzed.

To examine differences in subject-matter understanding of the Level B Computer Hardware Repair certification between participants using different chatbot systems, the instructor did not intervene in the self-directed learning process during the instructional experiment. All participants were allowed to study across the full scope of the subject question bank. Statistical analyses were conducted using SPSS version 21. In addition to analysis of covariance (ANCOVA), an equivalence analysis was conducted

using the two one-sided tests (TOST) procedure on learning gain scores ($\Delta = \text{Post} - \text{Pre}$). Given that each test consisted of 10 open-ended items scored on a 10-point scale, an equivalence margin of ± 10 points was adopted as a practically interpretable criterion, corresponding to approximately one item difference on the 100-point scale.

3 RESULTS

After the completion of the instructional experiment, statistical analyses were conducted using ANCOVA. The results are reported as follows.

1. Descriptive Statistics

This study adopted a nonequivalent pretest–posttest quasi-experimental design, consisting of an experimental group (KBRA-based chatbot) and a control group (ChatGPT). Both groups completed a pre- and a post-test. To present overall learning performance, descriptive statistics—including the mean and standard deviation—were calculated for the pre-test, post-test, and learning gain ($\Delta = \text{Post} - \text{Pre}$).

The descriptive statistics (Mean $\pm$ SD) are summarized in Table 2. As shown in Figure 5, both groups demonstrated substantial improvement from pre- to post-test, with similar mean post-test performance levels, providing an overall descriptive overview of learning gains prior to covariate adjustment. In addition, the variability of learning gains was somewhat higher in the ChatGPT group (SD = 18.35) than in the KBRA group (SD = 12.84).

Table 2. Descriptive statistics of pre-test, post-test, and learning gain scores

Group	n	Pre-Test	Post-Test	Gain (Δ)
Experimental (KBRA)	26	25.58 $\pm$ 10.71	75.77 $\pm$ 15.15	50.19 $\pm$ 12.84
Control (ChatGPT)	25	26.80 $\pm$ 12.15	76.00 $\pm$ 19.04	49.20 $\pm$ 18.35

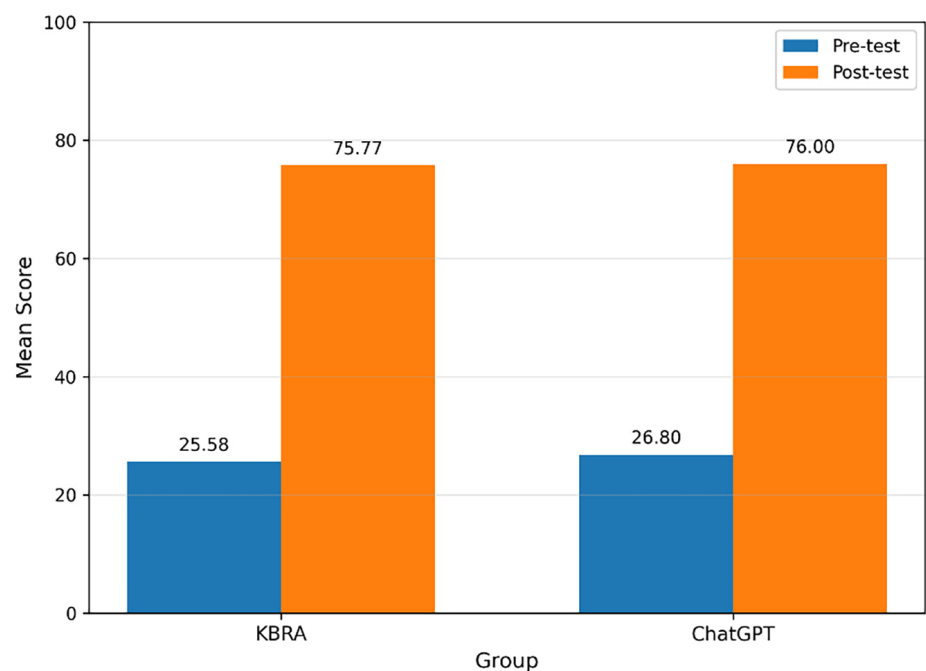


Fig. 5. Mean pre- and post-test scores for KBRA and ChatGPT

2. Baseline Group Comparison (Pre-test)

To examine whether the two groups were similar in terms of initial ability prior to the intervention, a between-group comparison was conducted using pre-test scores. Given the sample size and the potential inequality of variances between groups, Welch's independent-samples t-test was employed.

The hypotheses were defined as follows:

$$H_0: \mu p_{re,exp} = \mu p_{re,ctrl}$$

The results of Welch's t-test were as follows:

$$t = -0.381, df \cong 47.693, p = 0.705$$

The results indicate that the difference in pre-test scores between the two groups did not reach statistical significance at the $\alpha = 0.05$ level.

To further quantify the magnitude of the baseline difference, Cohen's d was calculated using the pooled standard deviation:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_p} \quad (6)$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (7)$$

Substituting the observed values in this study yields:

$$s_p = 11.44, d = \frac{25.58 - 26.80}{11.44} = -0.11$$

This result indicates a small effect size for the baseline difference between groups. Taken together, the baseline comparison suggests that the two groups were broadly similar in initial ability prior to the intervention.

3. Analysis of Covariance

This study employed a nonequivalent pretest-posttest quasi-experimental design to compare learning effectiveness between the experimental group (KBRA-based chatbot) and the control group (ChatGPT). Learning effectiveness was measured using post-test scores, with pre-test scores included as a covariate to control for the influence of learners' initial ability on post-test performance.

a) Preliminary Assumption Check: Homogeneity of Regression Slopes

Prior to conducting ANCOVA, the assumption of homogeneity of regression slopes was examined to determine whether the relationship between pre- and post-test scores was consistent across groups. To this end, an interaction term between group and pre-test scores was included in the model, as specified below:

$$Post = \beta_0 + \beta_1 Pre + \beta_2 Group + \beta_3 (Pre \times Group) + \epsilon \quad (8)$$

In this model, β_3 was used to test whether the effect of pre-test scores on post-test performance differed by group. The interaction effect was not statistically significant, with the following result:

$$F(1, 47) = 0.245, p = 0.623$$

This finding indicates that the relationship between pre- and post-test scores did not differ significantly between the two groups. As illustrated in Figure 6, the fitted regression lines for the two groups also show a broadly similar relationship between pre- and post-test scores. Therefore, the assumption of homogeneity of regression slopes was satisfied, and the conventional ANCOVA model assuming equal slopes across groups was adopted for the subsequent analysis.

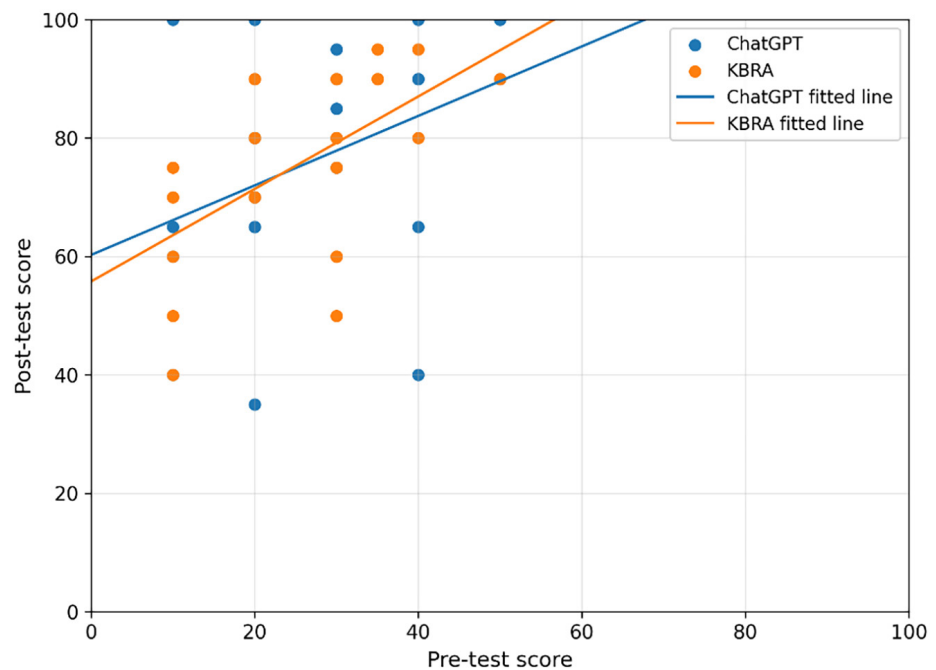


Fig. 6. Pre- and post-test scores for KBRA and ChatGPT

b) Conventional Analysis of Covariance

Because the assumption of homogeneity of regression slopes was satisfied, a conventional ANCOVA model was employed to examine differences in post-test performance between the two groups while controlling for pre-test scores.

The ANCOVA results showed that the covariate effect of pre-test score was statistically significant:

$$F(1, 48) = 12.116, p = 0.001$$

This finding indicates that learners' initial ability had a significant influence on post-test performance. However, after controlling for pre-test scores, the main effect of group was not statistically significant:

$$F(1, 48) = 0.019, p = 0.892$$

These results suggest that, after adjustment for initial ability, no statistically significant difference in post-test performance was observed between the KBRA and ChatGPT groups.

c) Effect Size Analysis of Group Effects

To quantify the practical magnitude of the group effect within the ANCOVA model, an effect size analysis was conducted. The effect size was expressed using partial eta squared (partial η^2), which represents the proportion of

variance in post-test scores attributable to group membership after controlling for other factors.

The formula for calculating partial η^2 is given as follows:

$$\eta_p^2 = \frac{SS_{group}}{SS_{group} + SS_{error}} \quad (9)$$

Based on the sums of squares obtained from the ANCOVA results, $SS_{group} = 4.471$ and $SS_{error} = 11525.395$. Substituting these values into the formula yields:

$$\eta_p^2 \approx 0.0004$$

The results indicate that the partial η^2 for the group effect was approximately 0.0004, which is considered a negligible-to-small effect size. This finding suggests that, after controlling for pre-test scores, only about 0.04% of the variance in post-test performance can be attributed to the type of chatbot used. Overall, the effect size results indicate that the magnitude of the difference in learning effectiveness between the two groups was extremely limited within the scope of this study.

4. Additional Equivalence Analysis Using TOST

To further examine whether the learning effectiveness of the two groups could be regarded as practically equivalent, an additional equivalence analysis was conducted using the TOST procedure on learning gain scores ($\Delta = \text{Post} - \text{Pre}$). Given that each test consisted of 10 open-ended items scored on a 10-point scale, an equivalence margin of ± 10 points was adopted as a practically interpretable criterion, corresponding to approximately one item difference on the 100-point scale.

The mean gain score was 49.20 ± 18.35 for the ChatGPT group and 50.19 ± 12.84 for the KBRA group, yielding a mean difference of 0.99 points (KBRA – ChatGPT). The 90% confidence interval for the mean difference was $[-6.49, 8.48]$.

When the equivalence bounds were set at ± 10 points on the 100-point scale, both one-sided tests were significant, supporting the conclusion that the two groups were equivalent under this predefined margin. Accordingly, under the ± 10 -point criterion, the learning gains of the KBRA and ChatGPT groups may be regarded as practically equivalent.

4 DISCUSSION

The discussion is organized into three parts: (1) Construction Procedure Comparison: KBRA vs. RAG, (2) Comparison of Learning Effectiveness Between the KBRA-Based Chatbot and ChatGPT, and (3) Expert-Scored Response-Quality Comparison Between the KBRA-Based Chatbot and ChatGPT.

1. Construction Procedure Comparison: KBRA vs RAG

In RAG-based systems, constructing a usable knowledge source typically requires building a vector database through a multi-stage pipeline. This pipeline begins with knowledge ingestion, where domain documents (e.g., textbooks, question banks, manuals, or web pages) are collected, followed by preprocessing

to parse, clean, and normalize the extracted text. The processed content is then segmented using a chunking strategy (e.g., fixed-size or overlapping chunks) to preserve contextual coherence. Next, each chunk is transformed into a dense vector representation during the embedding stage, which often constitutes the most computationally intensive part of the process and requires substantial CPU/GPU resources. Finally, the resulting embeddings must be organized through indexing and storage, where approximate nearest neighbor (ANN) indexes (e.g., HNSW/IVF) are created and stored in a vector database to enable efficient retrieval. As a result, RAG introduces not only online retrieval and generation components but also a non-trivial offline construction cost and maintenance overhead, particularly when knowledge sources are frequently updated or when embedding/index configurations are modified.

In contrast, the proposed KBRA-based chatbot adopts a knowledge-base-centered construction procedure that avoids the embedding-and-indexing pipeline entirely. Instead of converting documents into high-dimensional vectors, KBRA relies on a structured knowledge base in which each entry is explicitly organized to include a small, fixed set of keyword fields (keyword1–keyword5) and an associated instructional response. Similar to the initial knowledge preparation stage in RAG-based systems, the KBRA construction procedure also requires domain content collection, organization, and normalization. However, rather than parsing documents into chunks and transforming them into vector embeddings, KBRA structures the prepared content directly into knowledge base entries for subsequent keyword matching and fuzzy search. The construction procedure primarily involves preparing domain content, structuring it into knowledge base entries, and performing lightweight normalization (e.g., lowercasing, removing empty fields, and optionally consolidating synonymous terms). Once registered in the system (e.g., via a table or database), the knowledge base can be immediately used for query handling through keyword matching and fuzzy search. Because KBRA does not require vectorization or ANN index construction, it substantially simplifies system setup and updating: adding or revising knowledge can be achieved by editing or appending entries without triggering time-consuming re-embedding or re-indexing processes. Consequently, KBRA provides a deployment-friendly alternative that maintains real-time responsiveness through lightweight matching while reducing dependence on high-performance hardware required by LLM–RAG pipelines.

The contribution of the proposed system should be interpreted primarily from an implementation and evaluation perspective. The KBRA mechanism is intentionally simple, relying on structured keyword fields and fuzzy matching rather than embeddings, vector indexing, or open-ended generation. Therefore, the value of the system does not lie in replacing existing rule-based or retrieval-based chatbot approaches but in demonstrating how a lightweight knowledge-base-centered mechanism can be integrated into a mobile messaging environment and empirically evaluated in a bounded technical certification-learning context. This positioning is particularly relevant for educational settings in which ease of deployment, maintainability, and low computational requirements are important practical considerations.

2. Comparison of Learning Effectiveness Between the KBRA-Based Chatbot and ChatGPT

This study compared the learning effectiveness of a KBRA-based chatbot and ChatGPT in the context of the Level B Computer Hardware Repair certification subject. The results indicate that, after controlling for learners' initial ability,

no statistically significant difference in post-test performance was observed between the two groups. In addition, the group-related effect size was negligible, suggesting that the difference in learning effectiveness between the KBRA and ChatGPT groups was minimal in this study. These findings suggest that, under the conditions examined in this study, the KBRA-based chatbot provided learning support that was practically comparable to ChatGPT within the bounded Level B Computer Hardware Repair certification learning context. This interpretation should be made cautiously because the study involved a relatively small sample, non-random group assignment, and a single certification domain.

From a methodological perspective, these results carry important implications. ChatGPT relies on LLMs for inference and offers highly flexible natural language generation capabilities. In contrast, the KBRA-based chatbot is built upon a structured knowledge base and a keyword-based matching mechanism, responding to learners' queries through explicitly defined domain knowledge and matching rules. Given the similar learning outcomes observed in the two groups, the findings suggest that highly complex and resource-intensive generative models are not the only viable solution for supporting subject-matter understanding. For technical certification subjects with well-defined and bounded knowledge domains, the present findings indicate that a carefully structured knowledge-based chatbot may provide useful instructional support under specific implementation conditions. However, the results should not be interpreted as evidence that KBRA-based chatbots are generally equivalent to, or can broadly replace, LLM-based systems such as ChatGPT.

The additional TOST analysis further showed that, when the equivalence bounds were set at ± 10 points, the learning gains of the KBRA and ChatGPT groups could be regarded as practically equivalent. Because each test consisted of 10 open-ended items scored on a 10-point scale, this margin corresponded to approximately one item difference on the 100-point scale. This result provides further evidence that the difference in learning gains between the two groups was limited within a practically interpretable range.

Furthermore, the ANCOVA results revealed that pre-test performance had a significant effect on post-test scores, indicating that learners' initial ability remained an important factor in subsequent learning outcomes. This finding implies that, in educational practice, the selection and implementation of AI-assisted learning tools should take learners' prior knowledge and learning needs into account, rather than relying solely on whether a system is based on LLMs. By emphasizing structured domain knowledge and the alignment of key concepts, the KBRA-based chatbot may be particularly beneficial for learners who require clear guidance and focused instructional support.

3. Expert-Scored Response-Quality Comparison Between the KBRA-Based Chatbot and ChatGPT

The comparison between the KBRA-based chatbot and ChatGPT suggests that the difference between the two systems should be interpreted not only in terms of expert-scored response quality but also in terms of their response patterns and response boundary control. This analysis should be understood as a supplementary expert-scored response-quality comparison based on a limited set of prompts and one domain expert rater, rather than as a general benchmark of chatbot reliability or response accuracy. Within the scope of the Level B Computer Hardware Repair certification domain, both systems demonstrated generally strong performance across the 10 evaluation items, suggesting that both systems were able to provide largely appropriate responses when the questions

were confined to clearly structured subject knowledge. The KBRA-based chatbot achieved a total score of 950, while ChatGPT obtained 960, showing that the overall performance gap between the two systems was minimal. This result indicates that, in a well-bounded knowledge domain, a rule-based or knowledge-base-centered chatbot can perform at a level broadly comparable to that of an LLM in terms of expert-scored response quality.

A closer examination of the item-level results, however, reveals a meaningful difference in the nature of the responses produced by the two systems. ChatGPT showed a relative advantage on items that required broader conceptual explanation or supplementary contextual elaboration, such as odd parity, one's complement, and BASIC operator precedence. In these cases, ChatGPT tended to generate answers that were more expansive and instructionally elaborated, which may have contributed to its slightly higher scores. Nevertheless, this strength should not necessarily be interpreted as evidence of superior response reliability. Rather, it reflects the generative flexibility of LLMs, which enables richer explanations but may also introduce content that extends beyond the most direct or domain-constrained answer.

By contrast, the KBRA-based chatbot exhibited greater stability and tighter control in questions that closely matched predefined knowledge entries. Its response mechanism, which is grounded in structured knowledge retrieval rather than open-ended generation, appears to reduce the likelihood of unsupported elaboration. This characteristic is particularly relevant to response-quality interpretation because educational question answering may be affected not only by factual error but also by unnecessary expansion, overgeneralization, or the inclusion of plausible yet non-essential information. From this perspective, the KBRA-based chatbot may offer an advantage in maintaining response boundaries and preserving alignment with expected domain knowledge. This interpretation is also supported by the result for the optical disc data transfer rate item, in which the KBRA-based chatbot slightly outperformed ChatGPT by providing a concrete calculation example that more closely matched the evaluation expectation.

Collectively, these findings suggest that the distinction between the two systems lies less in absolute accuracy than in the balance between explanatory breadth and response control. ChatGPT showed a stronger capacity for contextual expansion and more detailed explanation, whereas the KBRA-based chatbot appeared to provide greater consistency and tighter control over unsupported content in a narrowly defined subject domain. For educational applications, this distinction is important. In instructional settings where factual precision, answer stability, and response boundary control are prioritized, the KBRA-based chatbot may represent a feasible learning support option for bounded domain-specific contexts. Nevertheless, because this comparison was based on a limited set of expert-scored items and one domain expert rater, the findings should be interpreted as supplementary evidence of response-quality differences rather than as a general evaluation of chatbot reliability or response accuracy across broader domains.

5 CONCLUSION

This study proposed a lightweight LINE-based mobile knowledge chatbot based on a KBRA for supporting learning in the Level B Computer Hardware Repair certification domain. In contrast to LLM-RAG architectures that require embedding generation, vector indexing, and substantial computational resources, the proposed

approach employed a smartphone-accessible LINE interface, a Google Apps Script backend, and a structured knowledge base to provide domain-specific instructional responses in natural language. This design substantially simplified system construction and updating while maintaining real-time responsiveness in a mobile self-directed learning context.

The empirical findings indicate that, within the bounded certification-learning context examined in this study, the KBRA-based chatbot may serve as a feasible instructional support tool for technical certification learning. Both the KBRA and ChatGPT groups showed clear improvement from pre-test to post-test. After controlling for learners' initial ability, no statistically significant difference in post-test performance was found between the two groups, and the effect size associated with chatbot type was negligible. Moreover, the TOST equivalence analysis demonstrated that the learning gains of the two groups were practically equivalent within the predefined ± 10 -point margin. These results suggest that, in the bounded and well-structured certification-learning context examined in this study, the LINE-based KBRA chatbot may support learners' subject-matter understanding at a level practically comparable to ChatGPT under the specific conditions examined. This conclusion should be interpreted cautiously and should not be extended to broader domains or general chatbot applications without further empirical validation.

The comparative analysis of response characteristics further revealed an important distinction between the two systems. ChatGPT showed a relative advantage in providing broader and more elaborated conceptual explanations, whereas the KBRA-based chatbot appeared to demonstrate greater response stability and tighter alignment with predefined domain knowledge. From the perspective of response boundary control, this difference is educationally meaningful in bounded certification-learning contexts. In instructional settings similar to the one examined in this study, where the learning content is bounded, structured, and explicitly organized in a knowledge base, the KBRA-based chatbot may represent a feasible learning support option when factual precision, answer consistency, and response boundary control are prioritized.

Overall, this study contributes to the design and evaluation of interactive mobile learning technologies for bounded technical certification domains. Rather than claiming a broadly novel chatbot algorithm, the study positions its contribution as a lightweight implementation and empirical evaluation of a LINE-based mobile knowledge chatbot in a structured certification-learning context. The findings suggest that, for the specific bounded certification-learning task examined here, learning support may be achievable through a lightweight and structured knowledge-based chatbot when the expected domain knowledge can be explicitly organized in advance. However, this result should be viewed as preliminary and context-dependent rather than as evidence against the broader value of resource-intensive generative architectures. From a practical perspective, a simple, transparent, and maintainable LINE-based mobile knowledge chatbot may offer a feasible and deployment-friendly learning support option for schools or training environments with limited computational resources.

Nevertheless, the conclusions of this study should be interpreted as preliminary and context-dependent. The relatively small sample size and the use of intact-class convenience sampling rather than random assignment limit the generalizability of the learning-effectiveness results. In addition, the study focused on a single bounded certification domain, namely the Level B Computer Hardware Repair certification, and therefore the findings may not apply to other technical subjects, broader academic domains, or more open-ended learning tasks. The use of a single domain

expert rater for both the open-ended assessments and the response-quality comparison also limited the reliability evidence, as inter-rater reliability could not be calculated. Furthermore, the ChatGPT backend model version was not controlled by the researchers, which may affect the reproducibility of the comparison across different time periods or system conditions. Finally, although system usage logs were collected for the LINE-based KBRA chatbot, detailed behavioral usage-log analysis was beyond the scope of the present study. Taken together, these limitations indicate that the observed practical comparability between the LINE-based KBRA chatbot and ChatGPT should be understood only under the specific conditions examined in this study.

In addition, the response-quality comparison should be regarded as a supplementary expert-scored analysis based on a limited set of prompts and should not be interpreted as a general evaluation of chatbot reliability or response accuracy across broader domains. Future research should further examine the scalability, long-term instructional value, and transferability of lightweight mobile chatbot systems using larger and more diverse learner samples, multiple raters, richer mobile usage-log analyses, and other bounded domain-specific learning contexts.

6 ACKNOWLEDGMENTS

This work was supported by the Ministry of Education Teaching Practice Research Program (Taiwan) under grant no. MOE-114-TPREE-0222-002Y1.

7 REFERENCES

- [1] R. L. Burke, *CAI Sourcebook*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [2] B. F. Skinner, *The Technology of Teaching*. New York: Appleton-Century-Crofts, 1968.
- [3] W. C. Hsu and C. H. Li, "A competency-based guided-learning algorithm applied on adaptively guiding e-learning," *Interactive Learning Environments*, vol. 23, no. 1, pp. 106–125, 2015. <https://doi.org/10.1080/10494820.2012.745432>
- [4] C. H. Li, "A proficiency-based learning algorithm applied to e-learning system for knowledge-oriented course," *MATEC Web of Conferences*, vol. 169, 2018. <https://doi.org/10.1051/mateconf/201816901002>
- [5] C. H. Li, "Integrating gamification elements into a personalized cognitive mobile-learning LINE Bot," *Emerging Science Innovation*, vol. 3, pp. 27–42, 2024. <https://doi.org/10.46604/emsi.2024.12980>
- [6] C. H. Li, "Enhancing vocational certification learning through a gamified chatbot: Evidence from a quasi-experimental study," *Advances in Mobile Learning Educational Research*, vol. 6, no. 1, pp. 1653–1667, 2026. <https://doi.org/10.25082/AMLER.2026.01.001>
- [7] H. B. Essel, D. Vlachopoulos, A. Tachie-Menson, E. Johnson, and P. Baah, "The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education," *International Journal of Educational Technology in Higher Education*, vol. 19, 2022. <https://doi.org/10.1186/s41239-022-00362-6>
- [8] L. Chambers and W. Owen, "The efficacy of GenAI tools in postsecondary education," *Brock Education Journal*, vol. 33, no. 3, pp. 57–74, 2024. <https://doi.org/10.26522/brocked.v33i3.1178>
- [9] I. Pesovski, R. Santos, R. Henriques, and V. Trajkovik, "Generative AI for customizable learning experiences," *Sustainability*, vol. 16, no. 7, 2024. <https://doi.org/10.3390/su16073034>

- [10] A. O'Cain, B. D. Fedoruk, Z. Masri, R. Frost, and A. Alahmar, "A system for the improvement of educational assessment using intelligent conversational agents," *SSRN Electronic Journal*, 2023. <https://doi.org/10.2139/ssrn.4393234>
- [11] D. H. Chang, M. P.-C. Lin, S. Hajian, and Q. Q. Wang, "Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization," *Sustainability*, vol. 15, no. 17, 2023. <https://doi.org/10.3390/su151712921>
- [12] F. Ali, D. Choy, S. Divaharan, H. Tay, and W. Chen, "Supporting self-directed learning and self-assessment using TeacherGAIA, a generative AI chatbot application: Learning approaches and prompt engineering," *Learning: Research and Practice*, vol. 9, no. 2, pp. 135–147, 2023. <https://doi.org/10.1080/23735082.2023.2258886>
- [13] F. Kieser, P. Wulff, J. Kuhn, and S. Küchemann, "Educational data augmentation in physics education research using ChatGPT," *Physical Review Physics Education Research*, vol. 19, arti no. 020150, 2023. <https://doi.org/10.1103/PhysRevPhysEducRes.19.020150>
- [14] Q. Xia, X. Weng, O. Fan, T. Lin, and T. Chiu, "A scoping review on how generative artificial intelligence transforms assessment in higher education," *International Journal of Educational Technology in Higher Education*, vol. 21, pp. 1–22, 2024. <https://doi.org/10.1186/s41239-024-00468-z>
- [15] D. Baidoo-Anu and L. O. Ansah, "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning," *SSRN Electronic Journal*, 2023. <https://doi.org/10.2139/ssrn.4337484>
- [16] M. Giannakos *et al.*, "The promise and challenges of generative AI in education," *Behaviour & Information Technology*, vol. 44, no. 11, pp. 2518–2544, 2024. <https://doi.org/10.1080/0144929X.2024.2394886>
- [17] U. Mittal, S. Sai, V. Chamola, and D. Sangwan, "A comprehensive review on generative AI for education," *IEEE Access*, vol. 12, pp. 142733–142759, 2024. <https://doi.org/10.1109/ACCESS.2024.3468368>
- [18] M. Allen, U. Naeem, and S. Gill, "Q-Module-Bot: A generative AI-based question and answer bot for module teaching support," *IEEE Transactions on Education*, vol. 67, pp. 793–802, 2024. <https://doi.org/10.1109/TE.2024.3435427>
- [19] W. Ma, W. Ma, Y. Hu, and X. Bi, "The who, why, and how of AI-based chatbots for learning and teaching in higher education: A systematic review," *Education and Information Technologies*, vol. 30, pp. 7781–7805, 2024. <https://doi.org/10.1007/s10639-024-13128-6>
- [20] Y. A. Malkov and D. A. Yashunin, "Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 824–836, 2020. <https://doi.org/10.1109/TPAMI.2018.2889473>
- [21] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547, 2021. <https://doi.org/10.1109/TBDATA.2019.2921572>
- [22] J. Wang *et al.*, "Milvus: A purpose-built vector data management system," in *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21)*, 2021, pp. 2614–2627. <https://doi.org/10.1145/3448016.3457550>
- [23] M. Aumüller, E. Bernhardsson, and A. Faithfull, "ANN-Benchmarks: A benchmarking tool for approximate nearest neighbor algorithms," *Information Systems*, vol. 87, 2020. <https://doi.org/10.1016/j.is.2019.02.006>
- [24] M. A. Kuhail, N. Alturki, S. Alramlawi, and K. Alhejori, "Interacting with educational chatbots: A systematic review," *Education and Information Technologies*, vol. 28, pp. 973–1018, 2023. <https://doi.org/10.1007/s10639-022-11177-3>

- [25] I. El mourabit, S. Jai Andaloussi, O. Ouchetto, and M. Miyara, "AI chatbots in higher education: Opportunities and challenges for personalized and mobile learning," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 12, pp. 19–37, 2025. <https://doi.org/10.3991/ijim.v19i12.54163>
- [26] C. Papakostas, C. Troussas, A. Krouska, and C. Sgouropoulou, "A rule-based chatbot offering personalized guidance in computer programming education," in *Generative Intelligence and Intelligent Tutoring Systems*, 2024, pp. 253–264. https://doi.org/10.1007/978-3-031-63031-6_22
- [27] E. Kasneci *et al.*, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, 2023. <https://doi.org/10.1016/j.lindif.2023.102274>
- [28] A. D. Samala, S. Papadakis, and S. Rawas, "Global insights into mobile learning in higher education: A PRISMA-guided bibliometric analysis from 2007 to 2023," *International Journal of Educational Reform*, 2025. <https://doi.org/10.1177/10567879251341869>
- [29] H. Uğraş, M. Uğraş, S. Papadakis, and M. Kalogiannakis, "ChatGPT-supported education in primary schools: The potential of ChatGPT for sustainable practices," *Sustainability*, vol. 16, no. 22, 2024. <https://doi.org/10.3390/su16229855>
- [30] N. A. Dahri, W. M. Al-Rahmi, K. A. Alhashmi, and F. Bashir, "Enhancing mobile learning with AI-powered chatbots: Investigating ChatGPT's impact on student engagement and academic performance," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 11, pp. 17–38, 2025. <https://doi.org/10.3991/ijim.v19i11.54643>
- [31] K. Baba, N.-E. El Faddouli, and N. Cheimanoff, "Mobile-optimized AI-driven personalized learning: A case study at Mohammed VI Polytechnic University," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 4, pp. 81–96, 2024. <https://doi.org/10.3991/ijim.v18i04.46547>

8 APPENDIX A. PRE- AND POST-TEST QUESTIONS AND SCORING RUBRIC

8.1 Pre-test questions

1. What is a Karnaugh map?
2. What is the maximum theoretical data transfer rate, in B/sec, of a 50× optical disc drive?
3. In Boolean algebra, according to the idempotent law, what is $A + A$?
4. When soldering general electronic active components, such as transistors, ICs, capacitors, and resistors, a soldering iron of approximately 30 W is suitable because it can provide a stable soldering temperature of approximately how many degrees Celsius?
5. What is odd parity checking?
6. What is two's complement?
7. Convert "001101001" into its two's complement form.
8. Explain the difference between IDE and SATA cables.
9. Explain the priority order of arithmetic operators in the BASIC programming language from highest to lowest.
10. How many JK flip-flops are required, at minimum, to design a divide-by-5 ripple counter?

8.2 Post-test questions

1. What is the difference between one's complement and two's complement?
2. Convert "011100011" into its two's complement form.
3. Explain the similarity between IDE and SATA cables.
4. Explain the purpose of a Karnaugh map.
5. What is the maximum theoretical data transfer rate, in B/sec, of a 120× optical disc drive?
6. In Boolean algebra, according to the idempotent law, what is $A + A + A + A$?
7. When soldering general electronic active components, such as transistors, ICs, capacitors, and resistors, approximately how many watts should the soldering iron have?
8. What is the difference between odd parity checking and even parity checking?
9. How many JK flip-flops are required, at minimum, to design a divide-by-7 ripple counter?
10. Which arithmetic operator in the BASIC programming language has the highest processing priority?

8.3 Scoring rubric

Each open-ended item was scored on a 10-point scale. The same scoring rubric was applied to both the pre-test and post-test.

Correctness of key concepts or procedures: 4 points. Responses received full credit in this category when the main concept, formula, rule, or procedure was accurate. Partial credit was awarded when the response demonstrated partial conceptual understanding but contained minor errors.

Completeness of explanation or calculation process: 3 points. Responses received full credit in this category when the answer included sufficient explanation, reasoning, or calculation steps. Partial credit was awarded when the response provided only a brief or incomplete explanation.

Relevance and clarity: 3 points. Responses received full credit in this category when the answer was directly related to the question and clearly expressed. Partial credit was awarded when the response was relevant but unclear, incomplete, or included unnecessary information.

The maximum score for each item was 10 points, and the total score for each test was 100 points. All responses were scored by the same instructor using the same rubric to maintain consistency across groups and assessment stages.

9 AUTHOR

Cheng-Hsiu Li is an Associate Professor in the Department of Information Management at National Taitung Junior College, Taiwan (R.O.C.). He received his Ph.D. degree in 2014. His research interests include e-learning, the Internet of Things (IoT), virtual reality (VR), mixed reality (MR), and computer vision (E-mail: lcs@ntc.edu.tw).