Textbook-Embedded Virtual Teaching Assistant That Answers CS1 Questions With Verifiable Citations

Yuanrong Han
University of Toronto
Toronto, Canada
rex.han@mail.utoronto.ca

Salma Emara
University of Toronto
Toronto, Canada
salma.emara@mail.utoronto.ca

Abstract

Students in CS1 often struggle to find course-appropriate answers across scattered resources and may not know how to ask questions due to hidden knowledge gaps. We present a Retrieval-Augmented Generation (RAG) Virtual Teaching Assistant(VTA) that extracts knowledge from course materials to answer exam and conceptual questions with CS1-aligned explanations and verifiable textbook citations. Experiments show reliable, complete responses with correct links to relevant textbook sections. Usage logs and surveys indicate that students ask more conceptual than problem-solving questions, and 89% report that the assistant is helpful—suggesting that grounded LLM support can encourage deeper engagement with core concepts.

CCS Concepts

• **Applied computing → Education.**

Keywords

Large Language Model, Retrieval-Augmented Generation, Education, Chatbot, Virtual Teaching Assistant, CS1

ACM Reference Format:

Yuanrong Han and Salma Emara. 2026. Textbook-Embedded Virtual Teaching Assistant That Answers CS1 Questions With Verifiable Citations. In *2026 IEEE/ACM 48th International Conference on Software Engineering (ICSE-Companion '26), April 12–18, 2026, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3774748.3795670

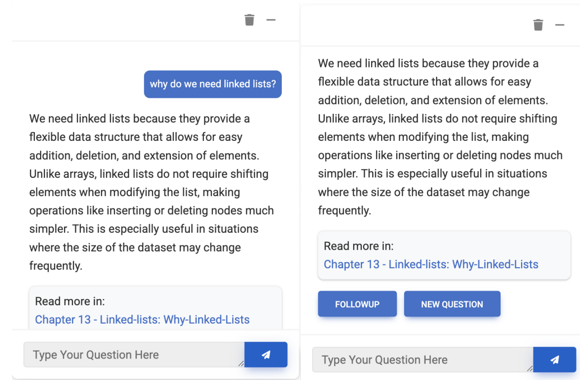

Figure 1: Example response with citation links to the textbook.

1 Problem

CS1 students frequently need *just-in-time* explanations while studying, but relevant information is spread across a textbook, lecture notes, and past exams. Students may not know how to phrase questions precisely, and web search or general-purpose GenAI can yield content that is mismatched to the course level and difficult to verify. Instructors and TAs are not always available outside of class and office hours, so students often look for alternative sources when they get stuck. However, answers without provenance make it hard to confirm correctness, especially when misconceptions can be subtle (e.g., pointers, memory models, or edge cases in C language). We target an assistant that (i) stays aligned with the course textbook and (ii) makes the evidence for each answer easy to inspect. This design also helps address trust and academic-integrity concerns by making the source material explicit rather than producing opaque solutions.

ICSE-Companion '26, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-2296-7/2026/04
https://doi.org/10.1145/3774748.3795670

2 Approach

Our VTA is embedded in the course textbook and provides an explanation, along with direct links to the textbook sections used to generate the answer, the ui is shown in Figure 1.

We preprocess the open-source textbook by segmenting it into chunks and storing each chunk with its section-level metadata (chapter and section identifiers) to support both retrieval and citation generation.

Building on prior work on LLM-based VTAs in CS education [Lau and Cooper-Stachowsky 2024; Liu et al. 2024], we build on a standard RAG pipeline and tailor retrieval for exam-related queries by combining keyword-based section search (BM25) with semantic reranking (cosine similarity), the pipeline shown in Figure 2.

To keep answers verifiable and course-aligned, we (1) retrieve textbook chunks with section metadata, (2) prompt the generator to ground the response in the retrieved context, and (3) surface the corresponding section links alongside the answer.

Because exam questions tend to be keyword-heavy (year, term, question number, topic names), we route exam-related queries through a BM25 stage to identify the most likely section before semantic reranking.

In our deployment, we retrieve the top-N contexts (we use N=10) and pass them to the generator, which produces a concise explanation and the section identifiers needed to construct hyperlinks into the textbook.

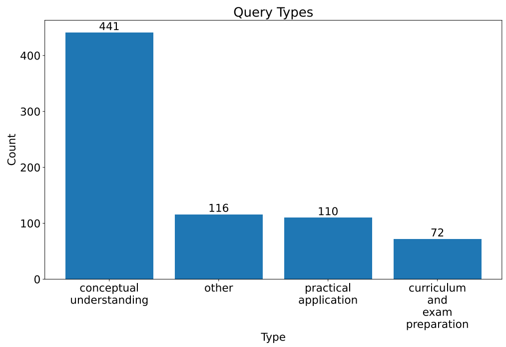

Figure 4: Distribution of query categories observed in usage logs.

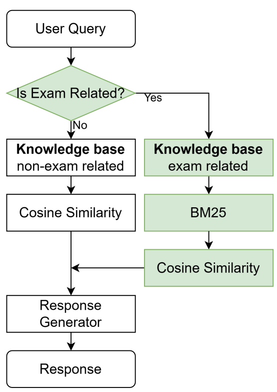

Figure 2: Deployed retrieval workflow

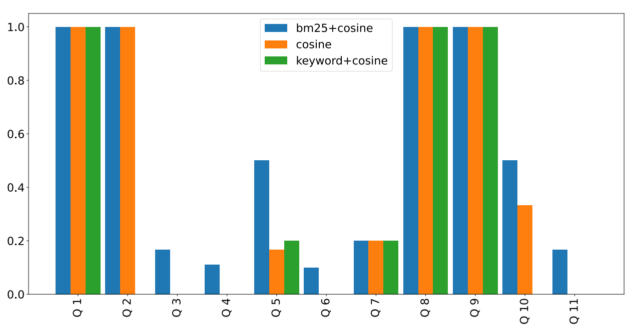

Figure 3: RR comparison on exam-related queries.

Table 1: MRR on exam-related queries.

Mechanism	MRR
BM25+Cosine Similarity	0.50
Cosine Similarity	0.45
Keyword Matching+Cosine Similarity	0.34

Concretely, we prompt the model to first decide whether to use keyword-based retrieval or cosine similarity, optionally extract query keywords for BM25, and return a structured response containing the answer text and citation fields (e.g., chapter/section names) that are rendered as clickable links in the textbook UI.

3 Results

Retrieval. We compared three retrieval mechanisms on exam-related queries: cosine similarity, BM25 for exam-related queries with cosine similarity otherwise (BM25+Cosine), and keyword matching for exam-related queries with cosine similarity otherwise (Keyword+Cosine). BM25+Cosine performs best on direct exam queries.

Answer quality. A course teaching assistant assessed generated answers across accuracy, relevance, completeness, and reference correctness. 93% of answers were accurate; 96% were relevant; 83% were complete; and references were correct in 89% of cases.

Usage and perceptions. Usage logs and survey responses indicate students asked more conceptual questions than problem-solving questions. Overall, 89% of surveyed students found the VTA helpful; 43% reported being likely to ask conceptual questions, while 78% reported being unlikely to use it for problem-solving, often citing academic-integrity concerns.

4 Limitations and Future Work

The current VTA is constrained by the textbook's scope and struggles with questions that require visual explanations (e.g., diagrams) or broader programming guidance (e.g., debugging strategies).

Future work includes improving support for visual content, refining retrieval to better summarize larger units (e.g., sections or chapters), and studying interventions that increase student trust for appropriate problem-solving use.

5 Takeaways

- **Evidence-backed responses.** Embedding the assistant in the textbook and returning section-level citations supports verification and reduces reliance on ungrounded explanations.
- **Hybrid retrieval for exam queries.** BM25 section ranking combined with cosine-similarity reranking improves retrieval for keyword-dense, exam-related questions, showing reranking effectiveness with BM25 by measuring mean reciprocal rank(MRR), shown in Figure 3 and Table 1.
- **Task-dependent use and perceived appropriateness.** Students report a higher willingness to ask conceptual questions (43%) than problem-solving questions (78% unlikely), shown in Figure 4 and qualitative feedback indicates that academic-integrity concerns can discourage problem-solving use; this motivates clearer guidance on acceptable use and onboarding that aligns the tool with course policies.

References

David Lau and Michael Cooper-Stachowsky. 2024. Architecture and Usage Information for a Course-Aligned, GPT-Powered Chatbot. *Proceedings of the Canadian Engineering Education Association (CEEA)* (Dec. 2024). doi:10.24908/pceea.2024.18547

Rongxin Liu, Carter Zenke, Charlie Liu, Andrew Holmes, Patrick Thornton, and David J. Malan. 2024. Teaching CS50 with AI: Leveraging Generative Artificial Intelligence in Computer Science Education. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) *(SIGCSE 2024)*. Association for Computing Machinery, New York, NY, USA, 750–756. doi:10.1145/3626252.3630938