iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: https://doi.org/10.22214/ijraset.2026.85157

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue X Oct 2026- Available at www.ijraset.com

PDF-SmartAI - An AI Study Companion for Uploaded Textbooks: A Book-Grounded Question-Answering System Using In-Browser PDF Extraction and Google Gemini

Fariya Malkani[1], Perseus Bhavnagri[2]
[1, 2]*Department of Information Technology, Wilson College (Autonomous), Mumbai, India*

Abstract: ***Students increasingly turn to general-purpose AI chatbots for help with their studies, yet these chatbots answer from what they absorbed during training rather than from the textbook a student is examined on, and they can state wrong facts with confidence. This paper presents PDF-SmartAI, a zero-cost web application in which a student uploads up to two PDF textbooks and asks questions that an assistant, AI Buddy, answers only from that book, citing the page numbers it used and saying so when the book does not cover a question.***

Text is extracted entirely inside the browser by PDFium compiled to WebAssembly, split into page-aware chunks, stored per user in Cloud Firestore under owner-only security rules, and sent to Google Gemini through Firebase AI Logic on free tiers — the whole book when it is short (up to 100,000 characters), otherwise the eight chunks chosen by a plain-Dart keyword retriever. The study is secondary research combined with the evaluation of a built artefact; no participants were recruited. A public global survey of higher-education students (analytic n = 12,909) shows a moderate positive association between how much students use ChatGPT and how useful they find it for learning (Spearman ρ = 0.376, p < .001), while a substantial share of students (44.3%) agree that ChatGPT might mislead with inaccurate information (Wilcoxon z = 21.286, p < .001). On 10,570 SQuAD v1.1 questions over 13 synthetic books, the app's own retrieval code placed the answer-bearing passage among the eight excerpts sent to the AI for 94.9% of questions, far above a random baseline (4.9%) and within 0.05 of BM25 (95.7%), although statistically slightly below it. All three hypotheses are supported. AI answer quality was checked only through live smoke tests and remains to be benchmarked.

Keywords: Large language models, retrieval-augmented generation, question answering, hallucination, page citation, PDF text extraction, WebAssembly, Flutter Web, Firebase, Google Gemini.

I. INTRODUCTION

Generative AI chatbots have quickly become part of how students in higher education study. Large language models (LLMs) can summarise a chapter, explain a difficult idea in simpler words and generate practice questions on demand [1]. In a global survey of 23,218 students from 109 countries and territories, students valued ChatGPT for simplifying complex information and summarising extensive content, but only 41% believed it could provide reliable information [2]. For a student preparing for an examination this gap matters: questions are set from prescribed textbooks and notes, and an answer that is fluent but does not match the book can cost marks.

An LLM answering from its own memory can produce hallucinations — content that is nonsensical or unfaithful to the source [3]. A general chatbot also cannot show where in the book an answer comes from. Retrieval-augmented generation (RAG) finds relevant passages first and asks the model to answer from them [4], but typical RAG systems rely on servers, vector databases and embedding models that usually cost money to run.

PDF-SmartAI addresses this gap: its assistant, AI Buddy, answers a student's questions only from the student's own uploaded PDF textbook, shows the pages each answer is drawn from, admits when the book does not cover a question and runs entirely on free tiers. Section II reviews related work, Section III describes the methodology and hypotheses, Section IV the design and implementation of the system, Section V the results, and Section VI concludes.

©IJRASET: All Rights are Reserved

A. Problem Statement and Need for the Study

Students already use general AI chatbots for study, yet a substantial share doubt the accuracy of the answers. Such chatbots do not answer from the student's own textbook, show no page references and rarely admit that the book does not cover a question, while book-grounded alternatives need paid servers or per-token API calls. A study companion is needed that answers only from the student's own book, shows where each answer comes from and runs entirely on free tiers — with evidence that such a cheap design is good enough. This study is undertaken to:

- examine, from published research and a public student-survey dataset, how students use AI chatbots for learning and how far they trust the accuracy of the answers;
- design and build a zero-cost web application that answers only from a student's uploaded book and cites the pages used;
- test whether a simple keyword retriever running in the browser reliably finds the passage that holds the answer in books too large to be sent to the AI whole;
- report honestly on the boundaries of the design — scanned books, paraphrased questions, and AI answer quality that has not yet been benchmarked.

B. Objectives

This study aims to:

- review the literature on LLMs in education, hallucination, RAG and client-side document processing;
- build a zero-cost web app that answers questions only from a student's uploaded PDF book, with page references;
- extract and chunk PDF text entirely in the browser and store it securely per user;
- stay within free tiers, with a hard stop and no paid fallback when the free AI quota runs out;
- evaluate the retrieval step on a public question-answering dataset, and the system by testing;
- test hypotheses on students' use of, and concerns about, AI chatbots using a public survey dataset.

II. LITERATURE REVIEW

A. AI and LLMs in Higher Education

Zawacki-Richter et al. [5] reviewed 146 studies of AI in higher education (2007–2018) and found almost no critical reflection on risks. Kasneci et al. [1] describe the value of LLMs for summaries, explanations and practice material, but warn that students may struggle to tell real knowledge from convincingly written but unverified output, and advise corroborating answers against books. Wang et al. [6] name hallucination as a key reliability problem in education and point to RAG as an effective way to address it.

B. Student Perceptions and Learning Outcomes

Among 399 students in Hong Kong, agreement that they understand generative AI "can generate output that is factually inaccurate" was among the highest (M = 4.10 of 5), and willingness to use it correlated with frequency of use (r = 0.326) [7]. Controlled studies show that design matters: a GPT-4 tutor whose prompts contained pre-written expert solutions, to avoid hallucination, more than doubled learning gains over in-class active learning [8], whereas unguarded GPT-4 access lowered later exam scores by 17% and answered the practice problems correctly only 51% of the time [9].

C. Hallucination, Retrieval and Citation

Ji et al. [3] define hallucination as output unfaithful to its source. RAG couples a generator with retrieved passages [4], and retrieval reduced hallucinated responses in knowledge-grounded conversation by over 60% [10]. Citations let users check claims, yet even the best models lacked complete citation support about half of the time on the long-form ELI5 dataset [11], so citation must be enforced by design.

D. Retrieval Methods and Long Contexts

BM25 remains the standard keyword-retrieval baseline [12]. Dense retrieval beat BM25 by 9–19 points in top-20 accuracy on most benchmarks, but not on SQuAD [13], whose questions share many words with their passages [14]. Gemini 1.5 reported near-perfect (> 99%) retrieval of a planted fact from contexts of up to at least 10 million tokens [15], yet models use the middle of long contexts poorly [16], and although long-context prompting beat RAG on average, RAG was far cheaper [17].

E. PDF Extraction in the Browser

Text in scanned pages must be recovered by optical character recognition (OCR); the neural OCR model Nougat took about 19.5 s per batch of six pages even on a 24 GB server graphics card, very slow compared with classical extraction [18]. For text-based PDFs, WebAssembly makes in-browser processing practical: it was designed as safe, fast, portable bytecode, with most early benchmarks running close to native speed [19], though on large programs it runs about 45–55% slower than native code [20].

F. Research Gap

Four gaps stand out. First, general chatbots answer from model memory, not from the student's own book, and give no page references. Second, the RAG and citation systems in the literature assume servers, vector indexes and embedding models, and long-context prompting is priced per input token; none targets a deployment that costs a student nothing. Third, perception studies document doubts about accuracy but do not test a book-grounded tool that could answer them. Fourth, there is little evidence that a deliberately simple keyword retriever running in the browser is accurate enough to feed the right passages to the model. PDF-SmartAI targets all four: answers only from the uploaded book, page citations limited to the pages actually sent, an honest "not in this book", in-browser extraction and a measured retrieval step, within free tiers.

III.METHODOLOGY

A. Research Design

The study is quantitative, descriptive and evaluative secondary research, combined with the design and evaluation of a working artefact. No questionnaire was administered and no participants were recruited. The survey variables are ordinal, so non-parametric tests were used (Spearman rank correlation for H1; one-sample Wilcoxon signed-rank test against the midpoint for H2) [21], with effect sizes reported alongside p-values. For H3 the app's real Dart retrieval code was run unchanged in a stand-alone harness and compared with baselines computed on identical chunks: random selection, BM25 ($k1 = 1.5$, $b = 0.75$) using the app's own tokeniser, and TF-IDF cosine similarity from scikit-learn [22].

B. Data Collection

Three secondary sources were used. (i) Higher Education Students' Early Perceptions of ChatGPT: Global Survey Data [23], published on Mendeley Data under a CC BY 4.0 licence: an anonymous online questionnaire (October 2023 – February 2024, convenience sampling) answered by 23,218 students in 109 countries and territories. (ii) The SQuAD v1.1 development set [13], licence CC BY-SA 4.0: 10,570 crowd-written questions on 48 Wikipedia articles, each with answer spans. (iii) The published literature reviewed above. The system itself supplied 33 automated unit tests, static analysis and live smoke tests run on 10 October 2026 against the real Firebase project and the real Gemini model.

C. Sample Size

The minimum sample for estimating a proportion in a large population follows Cochran's formula [24]:

$$n = (Z^2 \times p \times (1 - p)) / e^2 \quad (1)$$

where $Z = 1.96$ (95% confidence level), $p = 0.5$ (maximum variability – safest assumption) and $e = 0.05$ (5% margin of error), giving $n = 0.9604 / 0.0025 = 384.16 \approx 385$. After cleaning — keeping students who had used ChatGPT and answered all study variables, gender, level and field — the analytic survey sample is $n = 12{,}909$, about 33.5 times the minimum. For H3, all 10,570 SQuAD questions were used. Because any single article (14,706–85,305 characters) would be sent to the AI whole, articles were concatenated into 13 synthetic books of 100,115–208,340 characters, forcing the retrieval path: 2,067 pages (one paragraph per page) split by the app's chunker into 2,137 chunks.

D. Variables and Hypotheses

Table I lists the variables. H1 and H2 use variables from the survey dataset; H3 uses properties of the retrieval step measured on the built system. All hypotheses are tested at $\alpha = 0.05$.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

TABLE I
VARIABLES OF THE STUDY

Hypothesis	Independent variable	Dependent variable	Data source
H1	Extent of general ChatGPT use (Q15; 1 Rarely … 5 Extensively)	Perceived usefulness for learning, PU (mean of ten items Q26a–j; α = 0.913)	Survey dataset [23]
H2	— (one-sample test against the neutral midpoint 3)	Concern about accuracy: "ChatGPT might mislead with inaccurate information" (Q22e; 1–5)	Survey dataset
H3	Retrieval method: PDF-SmartAI retriever vs random, BM25 and TF-IDF baselines	Recall@8 (answer-bearing chunk among the 8 sent to the AI); MRR@8	SQuAD v1.1 dev set [13]

1) *H1 (Usefulness Rises With Use):* $H0_1$: There is no positive association between the extent of ChatGPT use and perceived usefulness for learning. $H1_1$: Students who use ChatGPT more rate it as more useful for learning ($\rho > 0$).
2) *H2 (Concern About Accuracy):* $H0_2$: Students are, on balance, not concerned that ChatGPT may give inaccurate information (location of Q22e ≤ 3). $H1_2$: Students lean towards agreeing that ChatGPT might mislead with inaccurate information (location > 3) — a concern that motivates answers grounded in the student's own book with page references.
3) *H3 (Zero-Cost Retrieval Is Sufficient):* $H0_3$: For books too large to be sent whole, the in-browser keyword retriever includes the answer-bearing passage among the excerpts sent to the AI for fewer than 80% of questions (Recall@8 < 0.80). $H1_3$: Recall@8 ≥ 0.80, far above a random baseline and comparable to BM25 (Recall@8 within 0.05 of BM25 at k = 8).

IV. SYSTEM DESIGN AND IMPLEMENTATION

PDF-SmartAI is a Flutter Web application [25] in which a student uploads a text-based PDF (at most 50 MB; two books per account). The pdfrx package [26], which runs the PDFium engine [27] compiled to WebAssembly, extracts the text page by page inside the browser; the PDF itself is never uploaded. Each page is cut into chunks of at most 1,500 characters that never span two pages, so every chunk carries its page number, and the chunks are stored in Cloud Firestore (Mumbai region) under owner-only, deny-by-default security rules [28]. A book of up to 100,000 characters is sent to the AI whole; for a larger book a plain-Dart keyword retriever selects the eight best-matching chunks, each labelled [Page N]. These go, with the last six chat messages, to Gemini 3.5 Flash-Lite through Firebase AI Logic on the free Gemini Developer API [29], [30], which must answer only from the excerpts in the JSON form {answer, pages, found}; cited pages are filtered to the pages actually sent. The project runs on the free Spark plan with no billing account [31], so Cloud Functions and Cloud Storage are not used, and when the free daily AI quota runs out the app stops with a clear message instead of falling back to a paid service. Fig. 1 shows the architecture.

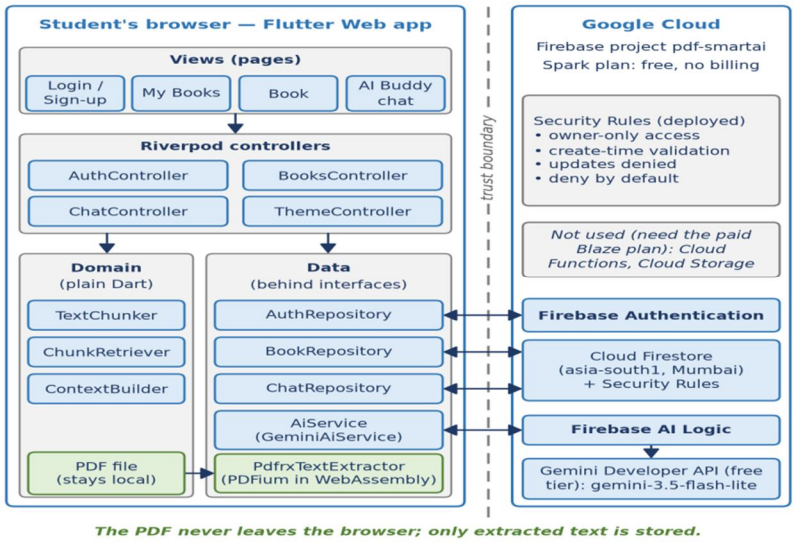


Fig. 1 Architecture of PDF-SmartAI — all processing of the PDF stays in the student's browser; only extracted text and chat messages reach Firebase.

V. RESULTS AND DISCUSSION

A. Description of the Data Sets

The analytic survey sample (n = 12,909) is 54.3% female; 83.3% are undergraduates, 12.4% master's and 4.3% doctoral students; the largest fields are Social Sciences (40.5%) and Applied Sciences including Computer Science and IT (36.5%); mean age is 22.92 years; and 88.4% used only the free ChatGPT-3.5. Students come from 94 countries and territories, but India contributes only 65 (0.5%), so the results describe students in general rather than Mumbai students. The retrieval set has 10,570 questions over 13 books of 124–266 chunks each. Table II summarises the survey variables.

TABLE II
NUMERICAL SUMMARY OF THE SURVEY STUDY VARIABLES (1–5 SCALES)

Variable	Mean	SD	Median	% agree / high (4–5)
Extent of general ChatGPT use (Q15)	2.483	1.094	2	18.0
Perceived usefulness for learning, PU index (Q26a–j)	3.502	0.733	3.500	27.6†
ChatGPT might mislead with inaccurate information (Q22e)	3.232	1.138	3	44.3
ChatGPT can provide reliable information (Q19e)	3.277	0.916	3	40.7
Satisfied with the accuracy of information (Q24g)	3.158	0.969	3	37.6
ChatGPT can simplify complex information (Q19g)	3.792	0.847	4	68.1

(† Share of students whose PU mean is 4 or higher.) Students find ChatGPT helpful — 68.1% agree it simplifies complex information — but only 40.7% agree it gives reliable information and only 37.6% are satisfied with its accuracy. Mean PU rises steadily with extent of use, from 3.081 (rarely) to 4.130 (extensively).

B. Retrieval Effectiveness

Table III compares the four retrieval methods on all 10,570 questions. A question counts as a hit at k when one of the top-k chunks comes from the answer's page and contains a gold answer string; MRR@8 is the mean reciprocal rank of the first hit.

TABLE III
RETRIEVAL EFFECTIVENESS ON 10,570 SQUAD V1.1 QUESTIONS (13 BOOKS, 2,137 CHUNKS)

Method	R@1	R@3	R@5	R@8	MRR@8
PDF-SmartAI retriever (real Dart code)	0.748	0.891	0.925	0.949	0.824
BM25 (k1 = 1.5, b = 0.75; app tokeniser)	0.791	0.912	0.939	0.957	0.855
TF-IDF cosine (scikit-learn, English stop words)	0.698	0.863	0.911	0.937	0.787
Random 8 chunks (mean of 5 seeds)	0.006	0.018	0.030	0.049	0.017

The answer-bearing passage reached the AI for 94.9% of questions and was ranked first for 74.8% (Fig. 2). Retrieval returned nothing for only 9 questions (0.1%), which then fell back to ten chunks spread evenly through the book. Retrieval took 7.8 ms per question on average (median 7.4 ms, p95 12.1 ms) on a desktop Dart VM — negligible next to the AI reply of about 4 s observed in live tests. A sample of the misses shows paraphrase, synonyms, acronyms and inflected forms (for example "weigh" against "weighed"), the known weakness of keyword matching.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

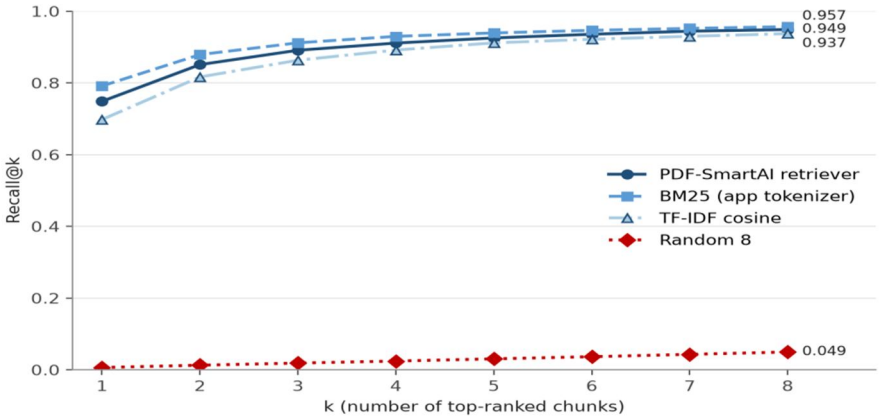


Fig. 2 Recall@k (k = 1–8) of the PDF-SmartAI retriever compared with BM25, TF-IDF cosine and a random baseline.

C. Hypothesis Testing

Table IV gives the outcome of the three hypothesis tests.

TABLE IV
RESULTS OF HYPOTHESIS TESTING (α = 0.05)

Hypothesis	Test / Evaluation	Statistic / Observed	p-value	Decision
H1	Spearman rank correlation, one-tailed (n = 12,909)	ρ = 0.376, 95% CI [0.361, 0.390] (moderate)	< .001	Reject $H0_1$ — supported
H2	One-sample Wilcoxon signed-rank test vs midpoint 3, one-tailed (n = 12,909; 8,956 non-neutral)	z = 21.286; r = 0.225 (small); 44.3% agree, 25.1% disagree	< .001	Reject $H0_2$ — supported (small effect)
H3 (criterion)	Recall@8 vs 0.80, one-sided exact binomial test; 95% Wilson CI	0.949 (10,027 of 10,570); CI 0.944–0.953	< .001	Reject $H0_3$ — criterion met
H3 vs random	Recall@8 compared with random selection of 8 chunks	0.949 vs 0.049 (about 19×)	—	Far above random
H3 vs BM25	Exact McNemar test on per-question hits at k = 8	0.949 vs 0.957; 45 vs 130 discordant questions	< .001	Within 0.05 of BM25 (gap 0.008); statistically slightly lower
H3 vs TF-IDF	Exact McNemar test on per-question hits at k = 8	0.949 vs 0.937; 271 vs 150 discordant questions	< .001	Significantly above TF-IDF

For H1, all ten usefulness items correlate positively with extent of use (ρ = 0.238 to 0.330). For H2, of students who took a side, 63.8% agreed; the median answer is still "Neutral", so the concern is held by a substantial share of students — a plurality, not a majority. For H3, BM25 is better by 0.043 at rank 1 and 0.008 at k = 8, most likely because the app's scoring has no document-length normalisation.

D. Discussion

1) H1 Is Supported: Extent of ChatGPT use and perceived usefulness for learning show a moderate positive association (ρ = 0.376, p < .001). Students who use ChatGPT more tend to rate it as more useful for learning, which points to real demand for AI study help; the result is correlational, not causal.

2) H2 Is Supported, With a Small Effect: Concern that ChatGPT might mislead with inaccurate information outweighs reassurance (z = 21.286, p < .001; 44.3% agree against 25.1% disagree). A substantial share of students therefore doubt the accuracy of chatbot answers, which supports PDF-SmartAI's design choices: answering only from the student's own book, showing the pages used and admitting when the book does not cover a question.

3) *H3 Is Supported:* All three parts of the criterion are met. The zero-cost retriever passed the answer-bearing passage to the AI for 94.9% of questions (95% CI 0.944–0.953), well above the 0.80 criterion and about 19 times the random baseline, and it is within 0.05 of BM25 (0.008 lower at k = 8), although the McNemar test shows it is statistically slightly below BM25. SQuAD questions share many words with their passages, which favours keyword methods [14], so recall on real student questions is likely lower; and a hit means the answer was available to the AI, not that the AI answered correctly. Answer quality was checked only by live smoke tests on 10 October 2026, in which AI Buddy answered specific, comparison, summary and follow-up questions and correctly reported an off-topic question as not covered; no large-scale AI accuracy benchmark was run, as the AI is deliberately paused until shortly before the examination.

VI. CONCLUSION AND FUTURE WORK

PDF-SmartAI shows that a book-grounded AI study companion can be built and run at zero cost: the PDF is read inside the browser by PDFium compiled to WebAssembly, the text is held privately under owner-only rules, a simple keyword retriever chooses the passages, and Google Gemini answers only from them, with page references and an honest "not in this book" when needed. Secondary analysis of 12,909 students' survey responses confirms both the demand (usefulness rises with use) and the problem (a substantial share doubt the accuracy of chatbot answers). On 10,570 SQuAD questions the retriever delivered the answer-bearing passage 94.9% of the time — well above the 0.80 criterion, far above chance and only slightly behind BM25. The limits are stated plainly: the survey concerns ChatGPT rather than PDF-SmartAI, the benchmark favours keyword methods, scanned books are unsupported, and the AI's final answers have so far been confirmed only through smoke tests.

Based on the study, the following future work is recommended:

1) add BM25-style document-length normalisation to the retriever — a free, plain-Dart change that, judging by BM25's lead, could add about 0.04 at rank 1 on this data — plus light stemming and typo tolerance for word-form misses;

2) once the AI is re-enabled, benchmark answer faithfulness and page-citation accuracy within the free daily quota, and run a consented user study with students on their own textbooks;

3) enforce the two-book limit in the Firestore security rules as well as in the app, since a determined user could currently bypass the client-side check;

4) support scanned books through free on-device text recognition; server-side OCR would require the paid Blaze plan and is deliberately excluded from the zero-cost scope;

5) for teachers and colleges: encourage students to check AI answers against the cited pages of the prescribed book, in line with the corroboration advice of Kasneci et al. [1].

REFERENCES

[1] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," Learning and Individual Differences, vol. 103, Art. no. 102274, 2023, doi: 10.1016/j.lindif.2023.102274.

[2] D. Ravšelj et al., "Higher education students' perceptions of ChatGPT: A global study of early reactions," PLOS ONE, vol. 20, no. 2, Art. no. e0315011, 2025, doi: 10.1371/journal.pone.0315011.

[3] Z. Ji et al., "Survey of hallucination in natural language generation," ACM Computing Surveys, vol. 55, no. 12, Art. no. 248, 2023, doi: 10.1145/3571730.

[4] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 9459–9474. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

[5] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?," International Journal of Educational Technology in Higher Education, vol. 16, no. 1, Art. no. 39, 2019, doi: 10.1186/s41239-019-0171-0.

[6] S. Wang et al., "Large language models for education: A survey and outlook," arXiv:2403.18105, 2024, doi: 10.48550/arXiv.2403.18105.

[7] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: Perceptions, benefits, and challenges in higher education," International Journal of Educational Technology in Higher Education, vol. 20, no. 1, Art. no. 43, 2023, doi: 10.1186/s41239-023-00411-8.

[8] G. Kestin, K. Miller, A. Klales, T. Milbourne, and G. Ponti, "AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting," Scientific Reports, vol. 15, no. 1, Art. no. 17458, 2025, doi: 10.1038/s41598-025-97652-6.

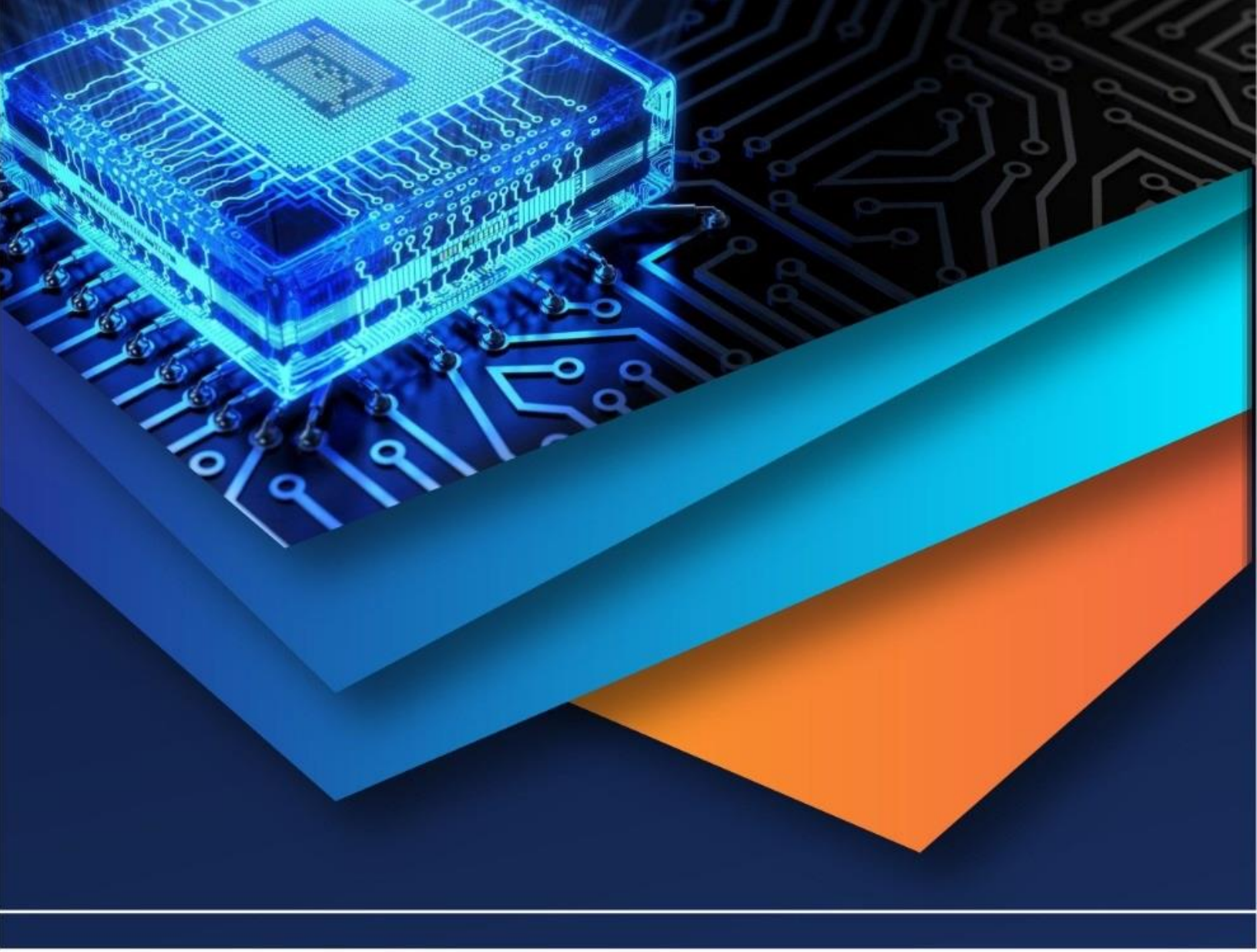
[9] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, "Generative AI without guardrails can harm learning: Evidence from high school mathematics," Proceedings of the National Academy of Sciences of the United States of America, vol. 122, no. 26, Art. no. e2422633122, 2025, doi: 10.1073/pnas.2422633122.

[10] K. Shuster, S. Poff, M. Chen, D. Kiela, and J. Weston, "Retrieval augmentation reduces hallucination in conversation," in Findings of the Association for Computational Linguistics: EMNLP 2021, 2021, pp. 3784–3803, doi: 10.18653/v1/2021.findings-emnlp.320.

[11] T. Gao, H. Yen, J. Yu, and D. Chen, "Enabling large language models to generate text with citations," in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023, pp. 6465–6488, doi: 10.18653/v1/2023.emnlp-main.398.

[12] S. Robertson and H. Zaragoza, "The probabilistic relevance framework: BM25 and beyond," Foundations and Trends in Information Retrieval, vol. 3, no. 4, pp. 333–389, 2009, doi: 10.1561/1500000019.

[13] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ questions for machine comprehension of text," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 2383–2392, doi: 10.18653/v1/D16-1264.
[14] V. Karpukhin et al., "Dense passage retrieval for open-domain question answering," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 6769–6781, doi: 10.18653/v1/2020.emnlp-main.550.
[15] Gemini Team, Google, "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," arXiv:2403.05530, 2024, doi: 10.48550/arXiv.2403.05530.
[16] N. F. Liu et al., "Lost in the middle: How language models use long contexts," Transactions of the Association for Computational Linguistics, vol. 12, pp. 157–173, 2024, doi: 10.1162/tacl_a_00638.
[17] Z. Li, C. Li, M. Zhang, Q. Mei, and M. Bendersky, "Retrieval augmented generation or long-context LLMs? A comprehensive study and hybrid approach," in Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track, 2024, pp. 881–893, doi: 10.18653/v1/2024.emnlp-industry.66.
[18] L. Blecher, G. Cucurull, T. Scialom, and R. Stojnic, "Nougat: Neural optical understanding for academic documents," in The Twelfth International Conference on Learning Representations (ICLR 2024), 2024. [Online]. Available: https://arxiv.org/abs/2308.13418
[19] A. Haas et al., "Bringing the web up to speed with WebAssembly," in Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation, 2017, pp. 185–200, doi: 10.1145/3062341.3062363.
[20] A. Jangda, B. Powers, E. D. Berger, and A. Guha, "Not so fast: Analyzing the performance of WebAssembly vs. native code," in Proceedings of the 2019 USENIX Annual Technical Conference, 2019, pp. 107–120. [Online]. Available: https://www.usenix.org/conference/atc19/presentation/jangda
[21] S. Siegel and N. J. Castellan, Jr., Nonparametric Statistics for the Behavioral Sciences, 2nd ed. New York, NY, USA: McGraw-Hill, 1988.
[22] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[23] D. Ravšelj et al., "Higher education students' early perceptions of ChatGPT: Global survey data," Mendeley Data, Version 2, Feb. 2025, doi: 10.17632/ymg9nsn6kn.2.
[24] W. G. Cochran, *Sampling Techniques*, 3rd ed. New York, NY, USA: John Wiley & Sons, 1977.
[25] Flutter, "Web support for Flutter," Flutter documentation. [Online]. Available: https://docs.flutter.dev/platform-integration/web (accessed Oct. 10, 2026).
[26] espresso3389.jp, "pdfrx | Flutter package," pub.dev. [Online]. Available: https://pub.dev/packages/pdfrx (accessed Oct. 10, 2026).
[27] PDFium Authors, "pdfium - Git at Google," Google Git. [Online]. Available: https://pdfium.googlesource.com/pdfium/ (accessed Oct. 10, 2026).
[28] Google, "Get started with Cloud Firestore Security Rules," Firebase documentation. [Online]. Available: https://firebase.google.com/docs/firestore/security/get-started (accessed Oct. 10, 2026).
[29] Google, "Gemini API using Firebase AI Logic," Firebase documentation. [Online]. Available: https://firebase.google.com/docs/ai-logic (accessed Oct. 10, 2026).
[30] Google, "Gemini Developer API pricing," Google AI for Developers. [Online]. Available: https://ai.google.dev/gemini-api/docs/pricing (accessed Oct. 10, 2026).
[31] Google, "Firebase Pricing," Firebase. [Online]. Available: https://firebase.google.com/pricing (accessed Oct. 10, 2026).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)