

LLM-RAG 系统的设计与教学实践应用

Design and Teaching Practice Application of LLM-RAG System

尹强¹, 于晓梅^{1*}, 郑向伟¹, 赵丽香¹

¹ 信息科学与工程学院, 山东师范大学

* yxm0708@126.com

【摘要】 本研究设计开发了一个基于 LLM-RAG 技术的智能问答助手系统, 该系统旨在为学生提供一个基于自建知识库功能的学习导师, 其中 RAG 技术可以有效解决 LLM 中出现的幻觉问题, 自建知识库功能可以有效解决大模型信息实时性不足的问题, 设计了 9 个 python 编程相关问题对智能问答助手进行测试, 经过三名相关领域专业教师的评估, 在智能问答助手回答问题准确性方面取得了良好的效果, 后续基于智能问答助手进行了两个班级的教学实践, 实践结果表明, 使用智能问答助手教学的班级成绩提高率明显高于未使用智能问答助手教学的班级。

【关键词】 LLM; RAG; GAI; 编程教学

Abstract: This research designs and develops an intelligent question answering assistant system based on llm-rag technology. The system aims to provide students with a learning tutor based on the function of self built knowledge base. Rag technology can effectively solve the hallucination problem in LLM, and the function of self built knowledge base can effectively solve the problem of insufficient real-time information of large models. Nine questions related to python programming are designed to test the intelligent question answering assistant. After the evaluation of three professional teachers in related fields, it has achieved good results in the accuracy of the intelligent question answering assistant. The subsequent teaching practice based on the intelligent question answering assistant in two classes shows that the improvement rate of class performance using the intelligent question answering assistant is significantly higher than that without using it. Intelligent question and answer assistant teaching class.

Keywords: LLM; RAG; GAI; Programming teaching

1. 引言

近年来, 人工智能在教育中的应用获得了巨大的吸引力, 旨在通过智能技术增强和创新学习和教学方法。像 ChatGPT 这样的生成式 AI 工具在教育领域表现出了巨大的潜力, 提供个性化的学习体验、学生支持和创新的课程内容交付 (Christos-Nikolaos, 2023)。这些工具根据学生的学习进度和需求提供实时互动反馈, 大大提高了学习效率和参与度 (Humble et al., 2024; Yu et al., 2024)。生成式人工智能可以通过对话的方式协助用户完成文本生成、检查代码等各类事项 (Lim et al., 2023), 能够及格或较为优秀的成绩通过大学预修生物学、人文、社会科学和法律考试 (Farazouli et al., 2024)。ChatGPT 在程序学习中的使用最近也变得很普遍, 因为学习编程对大多数人来说是一个具有挑战性和复杂的过程 (Rahman & Watanobe, 2023)。它可以帮助学生了解编程语言的复杂性, 例如使用代码进行调试, 并提供实时解决问题的帮助 (Biswas, 2023)。大量研究已经验证了 GPT 在解决编程问题方面的性能, 包括调试、生成代码和提供解释 (Pankiewicz & Baker, 2023; Phung et al., 2023; Tian et al., 2023)。

尽管生成式人工智能以其强大的能力受到教育工作者的追捧, 但是其编码和解决问题等基本技能的潜在负面影响, 同样引发了教育工作者的担忧 (Wise et al., 2024)。其中幻觉、过时的知识等问题是生成式人工智能的一大挑战 (Dhingra et al., 2022)。

RAG 是解决这些问题的有效途径 (Lee, 2024), RAG, 全称 Retrieval-Augmented Generation, 是一种结合了信息检索与文本生成优势的模型架构, 将检索方法与深度学习技术相结合, 通过动态集成最新的外部信息来解决大型语言模型 (LLM) 的静态局限性。其核心思想是通过“检索-生成”双重机制来提高生成模型的表现。在检索阶段, 模型通过检索系统从知识库中找到与输入相关的文档或段落; 在生成阶段, 生成模型利用检索到的信息作为上下文, 生成最终的答案或文本。

本文的工作介绍如下:

1. 以 Python 中的 Flask 为框架, 采用 MTV 开发模式, 以 LLM-RAG 为核心技术, 创建了一个基于自建知识库、自选模型以及对话历史功能的智能问答助手系统, 旨在解决大模型幻觉、过时知识问题以及提升学生使用体验。

2. 对开发的智能问答助手进行知识回答准确性测验, 设计了关于 Python 知识 3 个维度的九个问题, 邀请了三名具有丰富 Python 教学经验的教师进行答案评分, 取得了良好的结果。

3. 基于开发的智能问答助手系统应用于教学实践, 分别应用于中职学校计算机专业的两个班级, 进行了为期十课时的 python 编程教学, 实践结果表明, 应用智能问答助手的班级成绩提高明显优于未使用智能问答助手教学的班级。

2. 研究现状

2.1. GAI 在教育中的应用

ChatGPT 为教育提供了各种好处。它提供个性化的学习支持、自动化的管理流程和对在线学习的支持。ChatGPT 可以满足教育界的特定要求, 包括教职员工、学生 (Mogavi et al., 2024)。教师可以利用 ChatGPT 进行作业评估以及考试设计 (Ali et al., 2024), 学习监督 (Cowling et al., 2023), 课程设计 (Meron & Araci, 2023) 等。学生可以进行学习支持 (Rawas, 2024), 写作翻译 (Keiper et al., 2023), 个性化的学习辅导 (Bommineni et al., 2023) 等。

2.2. GAI 在编程教育中的应用

在编程教育中, ChatGPT 也显示出解决不同编码问题的潜力。作为一种可以使用自然语言进行交互的 AI 语言模型, ChatGPT 甚至可以让那些没有编程知识的人轻松解决编码问题, 大大降低了学习编程的门槛 (Surameery & Shakor, 2023)。最近的研究强调了 ChatGPT 在各种编程任务中的强大能力。例如, Tian et al. (2023) 实证分析了 ChatGPT 作为代码生成、程序修复和代码摘要的全自动编程助手的潜力。他们的结果表明, ChatGPT 有效地处理了典型的编程挑战, 例如修复错误、提供描述和生成代码。此外, 研究发现, ChatGPT 的错误修复性能与常见的深度学习方法相比具有竞争力, 并且明显优于标准的程序修复技术 (Sobania et al., 2023)。Phung et al. (2023) 系统地将 GPT 模型与使用不同的 Python 编程问题和现实世界的 buggy 程序的人类导师进行了比较, 揭示了 GPT-4 在几种情况下的表现接近于人类导师的表现。此外, ChatGPT 协助提供有关编程作业的反馈, 支持学生理论知识的实际应用。

2.3. LLM-RAG 在教育中的应用

RAG 技术致力于解决大模型回答问题时产生的幻觉问题, 教育领域的知识准确性要求使得研究者们对这一技术产生了浓厚的兴趣。其中卢宇 et al. (2024) 提出了大模型与检索增强生成技术结合的教学智能体基本框架及概念, 阐述了其应用于教学的优势。王珊 et al. (2025) 在 GLM-4 大模型的基础上结合检索增强生成等技术构建了助力家庭教育的理论模型, 取得了良好的实践效果。Lee (2024) 基于 ChatGPT 和检索增强生成技术开发了一个

计算机导师，能够为所有设计的二十个问题提供正确和相关的答案，并根据要求生成 R 代码片段。这项研究强调了基于 RAG 的导师为学生提供准确和个性化学习体验的潜力，同时降低了提供通常与 LLM 相关的虚假信息的风险。Muludi et al. (2024)评估了使用 ChatGPT 模型的生成预训练 Transformer 3.5 或 GPT-3.5-turbo 的 RAG 及其对文档数据处理的影响，并将其与其他应用程序进行了比较，结果突出了 RAG 的优越性。

3.系统构建

3.1. MTV 框架模式

本系统整体架构采用 Python 编程语言中的 Flask 这一轻量级且高度可扩展的 Web 框架进行构建。Flask 框架采用 MTV 模式，也就是通过 Model（模型）、View（视图）以及 Template（模板）这三个部分来组织和实现。MTV 模式的工作流程如图 1 所示。

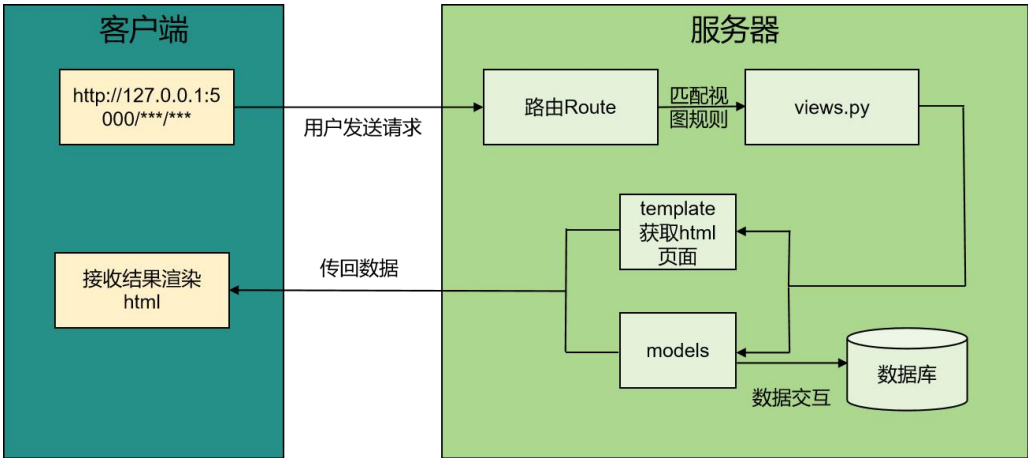


图 1 MTV 框架

3.2. LLM-RAG 技术

LLM-RAG 技术实现原理流程如图 2 所示：

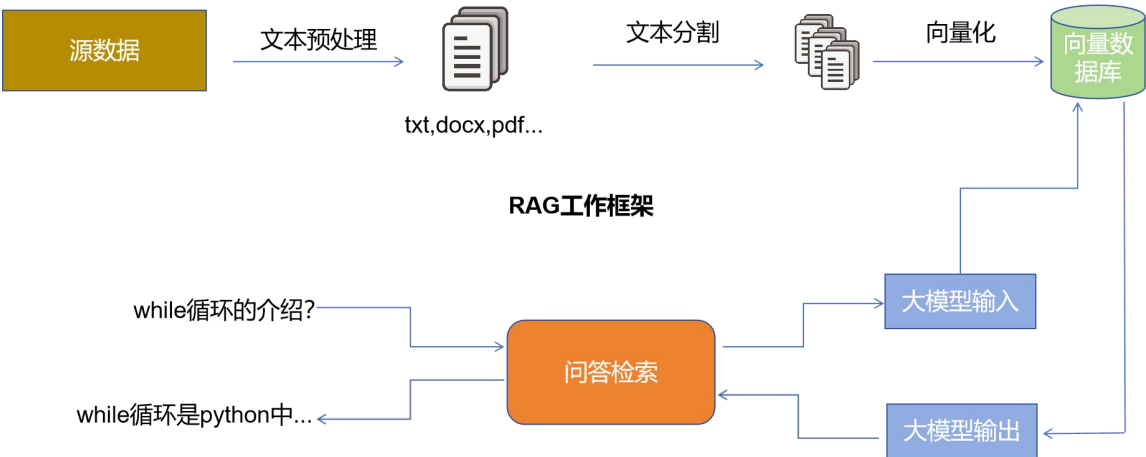


图 2 LLM-RAG 技术流程

首先进行数据的获取，为了保证数据的真实性和准确性，我们的数据获取源包括 Python 程序设计教材、菜鸟教程（python 知识网站）等，对获取的数据进行预处理，与具有多年 Python 课程教学经验的教师进行交流探讨，对重复知识以及无关知识进行剔除。

然后对处理好的数据进行文本分割，分为结构化片段和非结构化片段。针对结构化数据（csv 和 xlsx 表格文件）片段分割，直接按行进行数据读取即可。非结构化数据包括 txt，

docx, pdf 等文件, 直接按照段落长度进行切割, 这里设计片段长度为 500, 重复长度为 100, 这样做的目的是防止遗漏的信息检索。使用 openai 的 text-embedding-3-small 模型对数据进行向量化并将其存入向量数据库中供大模型根据问题进行匹配。

最后大模型对问题进行检索, 答案召回阶段按照以下流程进行: 用户提问——>大模型输入——>问题向量化——>L2 相似度匹配——>答案召回——>大模型总结输出。

采用 L2 欧几里得距离 (Euclidean distance) 进行向量化后的问题与向量数据库进行相似度匹配, L2 距离可以通过以下公式计算:

$$D = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

3.3. 基于 LLM-RAG 的智能问答助手

基于上述技术构建的智能问答助手系统如图 3 所示。

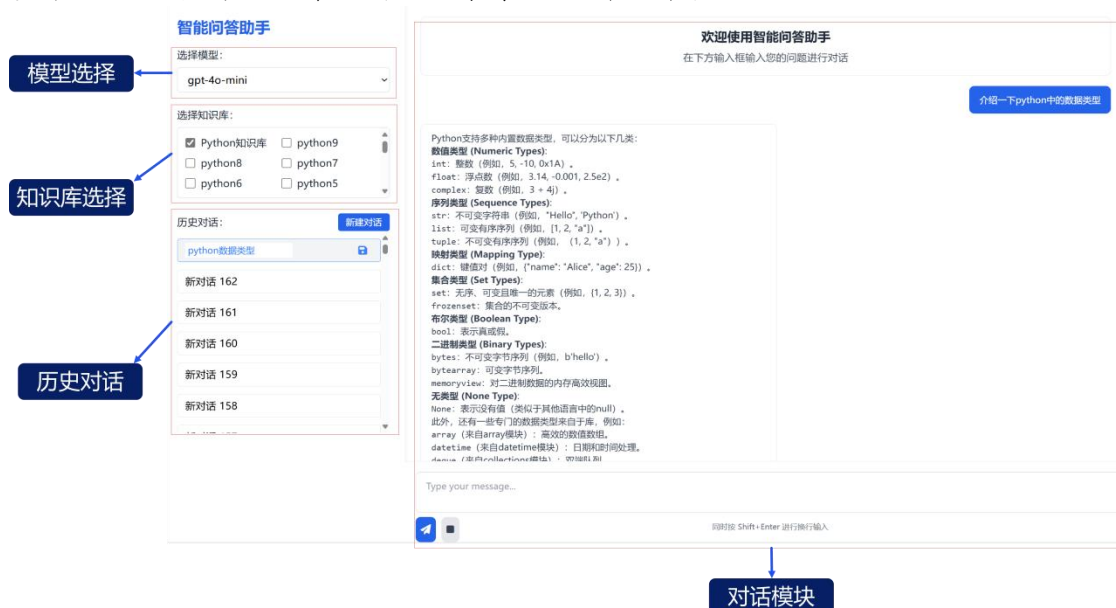


图 3 智能问答助手

模型选择功能: 考虑到 openai 的 api 需要用户订阅, 设计了多样性的模型选择功能, 这样的设计可以接入免费的大模型 api, 用户可以基于自身需求选择对话模型, 这能够消除社会中的数字鸿沟影响, 促进教育公平。

知识库选择功能: 用户可以根据自身需求上传文档, 智能问答助手根据上传的文档自动完成文档的分割和向量化任务, 从而创建知识库, 这意味着用户不仅可以针对于 Python 编程创建知识库满足编程需求, 也可以延伸到其他学科的学习中, 只需要勾选自己想要连接的知识库, 智能问答助手便可以根据知识库的内容来进行对话服务。

历史对话功能: 用户可以在历史对话中查看自己与智能问答助手的对话记录, 这有利于用户对学习的总结归纳。

对话功能: 用户可以根据自身需求选择模型、知识库, 基于此与智能问答助手进行对话, 得益于 RAG 功能, 大模型基于知识库检索进行问题回答, 避免了产生的幻觉与知识过时问题。

4. 准确性测试

为了保证智能问答助手在各方面问题的回答准确性，基于《Python 程序设计》教材，设计了 3 个维度的 9 个相关问题，问题相关介绍如表 1 所示，详细问题及答案见补充材料。

表 1 问题维度及描述

维度	问题
基础定义类问题	问题 1: Python 语言的特点?
	问题 2: Python 语言的数据类型有哪些?
	问题 3: Python 中增加列表的方法有哪些?
	问题 4: 利用 for 循环相关知识, 在屏幕上输出九九乘法表 (三角形式)。
编程类问题	问题 5: 一个 5 位数, 判断它是不是回文数。即 12321 是回文数, 个位与万位相同, 十位与千位相同。
	问题 6: 求输入数字的平方, 如果平方运算后小于 50 则退出。
	问题 7: 判断一个数是否为质数。
纠错类问题	问题 8: while 循环输出 0-10。
	问题 9: 有一分数序列: 2/1, 3/2, 5/3, ... 求出这个数列的前 20 项之和。

为保证问题回答的正确性，设计了一个评分表，评分表详细描述如表 2 所示。

表 2 评分表

指标	得分
智能问答助手回答不准确。	1 分
智能问答助手回答部分准确, 但包含不正确的信息。	2 分
智能问答助手回答准确, 但不够充分。	3 分
智能问答助手回答准确充分, 但包含不必要的信息, 不够简洁。	4 分
智能问答助手回答准确充分并且简洁易懂。	5 分

邀请了三名具有丰富 Python 编程教学经验的教师进行智能问答助手提供的答案评估打分，打分情况如表 3 所示。

表 3 教师评分情况

问题	分数 (教师 1)	分数 (教师 2)	分数 (教师 3)
1	4	4	4
2	5	5	5
3	4	5	4
4	5	5	5
5	5	5	5
6	4	4	4
7	5	5	5
8	5	5	5
9	5	5	5

Fleiss Kappa 判断多列离散无序数据，用于评估评分者之间的一致性程度。经过计算，智能问答助手答案的评分 K 值为 0.822，表明三位教师对智能问答助手回答的质量评分有很强的一致性。

如表 3 所示，教师对于智能问答助手的答案评分大多数为 5 分，说明智能问答助手对于问题的回答有较高的准确性。

5.教学实践测试

将智能问答助手应用于信息技术课程教学实践是检验其效果的有效方式(于晓梅, 2023)。本次实验在中国 J 市某中职学校开展，选择了计算机专业二年级的两个班级作为研究对象。其中，1 班被设定为实验班，在教学过程中融入了智能问答助手辅助教学；而 2 班则作为对照班，采用了传统的教学方法进行授课。

为了确保教学实验结果的准确性和可信度，在正式开展教学活动之前，采取了一项预备措施：为保证实验结果的可靠性，在进行教学实验之前选用《Python 程序设计》第一章 Python 程序基础的单元测试结果作为前测数据。

本研究使用 SPSS 数据分析软件，对 1 班（实验班）和 2 班（对照班）的前测成绩进行了统计分析，统计结果详见表 4。依据 Shapiro-Wilk 检验所得出的 p 值，可以判定两个班级的成绩均服从正态分布。如表 4 所示，实验班（1 班）的平均成绩为 52.42，标准差为 7.916；而对照班（2 班）的平均成绩为 53.02，标准差为 7.331。从这两个统计指标来看，在未进行教学前两个班级的成绩水平和离散程度都较为接近，说明两个班级在成绩表现上具有相似性。

表 4 前测成绩分析结果

名称	样本量	平均值	标准差	Kolmogorov-Smirnov 检验		Shapiro-Wilk 检验	
				统计量 D 值	p	统计量 W 值	p
1 班	45	52.42	7.916	0.077	0.200	0.966	0.207
2 班	44	53.02	7.331	0.117	0.151	0.967	0.240

分别从学生与教师两个维度出发，设计了基于智能问答助手系统的教学活动，这些活动巧妙地贯穿于课前准备、课堂实施及课后巩固这三个关键环节之中，整体设计思路如图 4 所示。

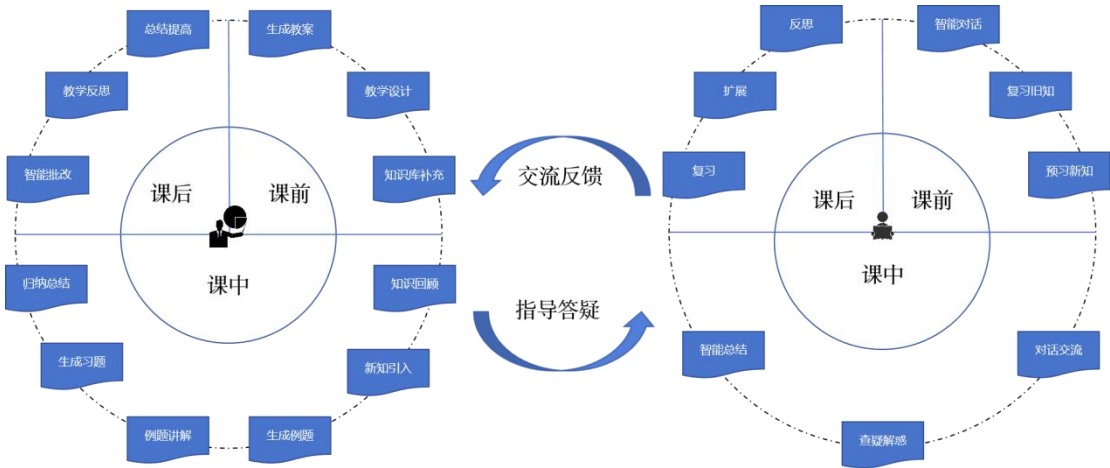


图 4 教学活动设计

在完成十课时的教学后，对实验组班级和对照组班级分别进行了 Python 程序语句相关的后测。测试内容为第二章节 Python 程序语句相关知识。利用 SPSS 数据分析工具对后

测成绩进行了统计分析，结果如表 5 所示。数据分析显示，使用了智能问答助手的班级，其成绩平均值和标准差均高于未使用智能问答助手的班级，经过计算，实验组班级学生平均成绩提高率为 23.4%，对照组班级学生平均成绩提高率为 11.6%，实验组班级成绩提高率比对照组班级平均成绩提高率高了 11.8%。这表明，在教学中引入智能问答助手有助于提升学生的学习成绩，从而验证了该助手在教学中提高学生成绩的有效性。

表 5 后测成绩分析结果

名称	样本量	最小值	最大值	平均值	标准差
1 班	45	48	85	64.69	7.951
2 班	44	44	80	59.16	7.631

6.结论

这项研究表明，基于 LLM-RAG 构建的智能问答助手系统对问题的回答有较高的准确性，解决了幻觉和知识过时的问题，不仅如此，知识库、模型的选择功能以及历史对话功能都能够很好的满足学生的学习需求。这项研究取得了良好的效果，但是也存在一些问题需要在后续的研究中进行解决。首先知识回答包含不必要的信息、答案不够简洁，这可能是由于知识库中数据过于繁杂造成的，为了解决这一问题，在后续的研究中应该对知识库中的数据进行更加严格的筛选以及预处理。其次本研究局限于智能问答助手应用在教学中学生成绩的变化探索，在后续的研究中要对学生学习过程、问题解决能力等方面的变化进行深入探讨。

致谢

本研究项目资助：山东省研究生优质教育教学资源项目（SDYKC2024146）。

参考文献

Ali, K., Barhom, N., Tamimi, F., & Duggal, M. (2024). ChatGPT—A double-edged sword for healthcare education? Implications for assessments of dental students. *European Journal of Dental Education*, 28(1), 206-211.

Biswas, S. (2023). Role of ChatGPT in Computer Programming. *Mesopotamian Journal of Computer Science*, 2023, 9-15.

Bommineni, V. L., Bhagwagar, S., Balcarcel, D., Bommineni, V., Davazitkos, C., & Boyer, D. (2023). Performance of ChatGPT on the MCAT: the road to personalized and equitable premedical learning. *MedRxiv*, 2023.2003. 2005.23286533.

Christos-Nikolaos, A. (2023). ChatGPT impacts in programming education: A recent literature overview that debates ChatGPT responses. *arXiv preprint arXiv*, 12348.

Cowling, M., Crawford, J., Allen, K.-A., & Wehmeyer, M. (2023). Using leadership to leverage ChatGPT and artificial intelligence for undergraduate and postgraduate research supervision. *Australasian Journal of Educational Technology*, 39(4), 89-103.

Dhingra, B., Cole, J. R., Eisenschlos, J. M., Gillick, D., Eisenstein, J., & Cohen, W. W. (2022). Time-aware language models as temporal knowledge bases. *Transactions of the Association for Computational Linguistics*, 10, 257-273.

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers’ assessment practices. *Assessment & Evaluation in Higher Education*, 49(3), 363-375.

- Humble, N., Boustedt, J., Holmgren, H., Milutinovic, G., Seipel, S., & Östberg, A.-S. (2024). Cheaters or AI-enhanced learners: Consequences of ChatGPT for programming education. *Electronic journal of e-Learning*, 22(2), 16-29.
- Keiper, M. C., Fried, G., Lupinek, J., & Nordstrom, H. (2023). Artificial intelligence in sport management education: Playing the AI game with ChatGPT. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 33, 100456.
- Lee, Y. (2024). Developing a computer-based tutor utilizing Generative Artificial Intelligence (GAI) and Retrieval-Augmented Generation (RAG). *Education and Information Technologies*, 1-22.
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The international journal of management education*, 21(2), 100790.
- Meron, Y., & Araci, Y. T. (2023). Artificial intelligence in design education: evaluating ChatGPT as a virtual colleague for post-graduate course development. *Design Science*, 9, e30.
- Mogavi, R. H., Deng, C., Kim, J. J., Zhou, P., Kwon, Y. D., Metwally, A. H. S.,...Gujar, S. (2024). ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans*, 2(1), 100027.
- Muludi, K., Fitria, K. M., & Triloka, J. (2024). Retrieval-Augmented Generation Approach: Document Question Answering using Large Language Model. *International Journal of Advanced Computer Science & Applications*, 15(3).
- Pankiewicz, M., & Baker, R. S. (2023). Large Language Models (GPT) for automating feedback on programming assignments. *arXiv preprint arXiv*.
- Phung, T., Pădurean, V.-A., Cambronero, J., Gulwani, S., Kohn, T., Majumdar, R.,...Soares, G. (2023). Generative AI for programming education: benchmarking ChatGPT, GPT-4, and human tutors. *Proceedings of the 2023 ACM Conference on International Computing Education Research-Volume 2*.
- Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied sciences*, 13(9), 5783.
- Rawas, S. (2024). ChatGPT: Empowering lifelong learning in the digital age of higher education. *Education and Information Technologies*, 29(6), 6895-6908.
- Sobania, D., Briesch, M., Hanna, C., & Petke, J. (2023). An analysis of the automatic bug fixing performance of chatgpt. *2023 IEEE/ACM International Workshop on Automated Program Repair (APR)*.
- Surameery, N. M. S., & Shakor, M. Y. (2023). Use chat gpt to solve programming bugs. *International Journal of Information Technology & Computer Engineering*(31), 17-22.
- Tian, H., Lu, W., Li, T. O., Tang, X., Cheung, S.-C., Klein, J., & Bissyandé, T. F. (2023). Is ChatGPT the ultimate programming assistant--how far is it? *arXiv preprint arXiv*.
- Wise, B., Emerson, L., Van Luyn, A., Dyson, B., Bjork, C., Thomas, S. E. J. H. E. R., & Development. (2024). A scholarly dialogue: writing scholarship, authorship, academic integrity and the challenges of AI. 43(3), 578-590.
- Yu, X., Mao, Q., Wang, X., Yin, Q., Che, X., & Zheng, X. J. E. S. w. A. (2024). CR-LCRP: Course recommendation based on Learner–Course Relation Prediction with data augmentation in a heterogeneous view. 249, 123777.
- 卢宇, 余京蕾, & 陈鹏鹤. (2024). 基于大模型的教学智能体构建与应用研究. *中国电化教育* (07), 99-108.

王珊, 肖义墙, 潘亦宁, & 彭唯. (2025). 生成式人工智能助力家庭教育的模型构建及系统实践. *电化教育研究*, 46(03), 64-71.

于晓梅. (2023). *软件与教育信息服务*. 山东人民出版社.