Bringing Generative Artificial Intelligence (GenAI) to Education

Hang Li
Michigan State University
East Lansing, Michigan
lihang4@msu.edu

Kaiqi Yang
Michigan State University
East Lansing, Michigan
kqyang@msu.edu

Yucheng Chu
Michigan State University
East Lansing, Michigan
chuyuch2@msu.edu

Jiliang Tang
Michigan State University
East Lansing, Michigan
tangjili@msu.edu

ABSTRACT

Generative Artificial Intelligence (GenAI) has rapidly emerged as a transformative technology across multiple research communities. With its exceptional capabilities in generating images, audio, video, and natural language, GenAI has achieved unprecedented success in various domains. In this tutorial, we aim to provide a comprehensive overview of integrating GenAI into education, presenting recent advancements in the field, exploring both its opportunities and challenges, and inspiring further research in this direction. We begin by introducing fundamental concepts of GenAI to equip attendees with the necessary background knowledge for the discussions that follow. Next, we examine the impact of GenAI on education, summarizing key transformations, opportunities, and challenges to provide a broad perspective on its applications in educational settings. Building on this foundation, we present detailed case studies of representative works that implement GenAI in education. By showcasing pioneering research and applications, we illustrate the benefits, challenges, and limitations of GenAI-driven educational tools. Additionally, we have prepared a live demonstration session where participants will experience two LLM-powered educational systems: an interactive learning platform and an automated grading tool. Through hands-on interaction, attendees will gain firsthand insight into how GenAI is shaping education. Finally, we conclude with a discussion session, encouraging participants to share their thoughts, insights, and ideas on the future of GenAI in education. By the end of this tutorial, we hope attendees will gain a solid understanding of current trends in GenAI-driven education, develop deeper insights into this evolving field, and find inspiration for future research and applications.

Keywords

Generative AI, AI in Education, Education Technology

Hang Li, Kaiqi Yang, Yucheng Chu, and Jiliang Tang. Bringing Generative Artificial Intelligence (GenAI) to Education. In Caitlin Mills, Giora Alexandron, Davide Taibi, Giosuè Lo Bosco, and Luc Paquette (eds.) Proceedings of the 18th International Conference on Educational Data Mining, Palermo, Italy, July, 2025, pp. 699–704. International Educational Data Mining Society (2025).

https://doi.org/10.5281/zenodo.15870320

1. TARGET AUDIENCE

The intended audience for this tutorial mainly includes researchers, graduate students, and professionals who are new to this area or who already have some experience with GenAI and data mining in education. The audience is expected to have some basic understanding of linear algebra, data mining, machine learning and generative model. However, the tutorial will be presented at college junior/senior level and should be comfortably followed by academic researchers and practitioners from the industry.

2. WHY IS THIS TOPIC IMPORTANT?

The rise of GenAI has recently driven revolutionary advancements across various fields, including finance [39], law [30], and medicine [31]. With its exceptional performance in key areas such as natural language processing (NLP), computer vision (CV), and speech signal processing (SSP), GenAI has demonstrated significant potential in enhancing problem-solving efficiency for complex real-world tasks [2, 26]. This improvement is evident in both human-AI collaborative approaches and fully automated solutions. In the field of education, one of the most persistent challenges in delivering high-quality learning experiences is the intensive demand for human effort. The emergence of GenAI introduces new solutions to this issue by leveraging its problem-solving and collaborative capabilities. With the integration of AI-driven tools, the future of education is likely to become more efficient and scalable. Beyond improving efficiency, GenAI's generative capabilities may also open up new directions in educational research, enabling advanced studies to address previously unexplored challenges [5, 27]. Given these transformative possibilities, we believe that GenAI in education will be a critical research topic in the coming years. Understanding how to responsibly and effectively integrate GenAI into educational systems will be essential to ensuring its successful and ethical implementation in future research and practice.

3. WHAT IS THE BENEFIT TO AUDIENCE?

Through this tutorial, participants will gain a general understanding of Generative AI in education. By exploring existing applications of GenAI in educational settings, attendees will develop a comprehensive perspective on this emerging field and find inspiration for their future research. Overall, the expected learning outcomes of the tutorial are:

- Develop a foundational understanding of GenAI. They will learn how different GenAI models are developed, their unique characteristics, and their applications.
- Gain insights into the role of GenAI in education. Participants will explore key directions in integrating GenAI into educational practices, along with the opportunities and challenges associated with its development.
- Learn from prior research and real-world applications. By studying existing implementations of GenAI in education, attendees will gain valuable insights and inspiration for their own research and projects.
- Engage with interactive demo systems. Participants will experience hands-on interaction with GenAI integrated educational tools, helping them understand the potential transformations AI brings to the future of learning.

4. OUTLINE OF THE TUTORIAL

The proposed duration of this tutorial is one day (6 hours), and it consists of five key components: (1) Introduction about GenAI, (2) Overview of GenAI in Education, (3) Application Cases in Education, (4) Demo System Experience and (5) Future Directions and Q&A. In the following section, we provide a detailed introduction to the contents covered in each component.

4.1 Introduction about GenAI

In this component, we aim to provide the audience with foundational knowledge about Generative AI (GenAI) through a structured three-step approach. First, we introduce key concepts and mathematical foundations commonly used in generative AI models, including probability theory, parameter estimation, and optimization methods. Building on this foundation, we discuss the modeling objectives of generative AI models and highlight their differences from other machine learning approaches, such as discriminative models. Once the theoretical groundwork is established, we introduce widely used generative AI models, categorizing them based on their primary application domains. For natural language processing, we discuss large language models (LLMs)[40]. In the field of computer vision, we cover generative diffusion models (GDM)[7]. For video generation, we examine SORA [4], and for audio processing, we introduce Whisper [25]. For each model, we explore its design, training methodology, and inference characteristics. Finally, we explore the practical applications of GenAI models across various tasks and discuss their transformative impact on different domains.

4.2 Overview of GenAI in Education

In this component, we present an overview of promising approaches for integrating GenAI into education. Specifically, we summarize recent survey studies, such as LLMs for Education [33] and Generative Models for Adaptive Learning [17], to highlight key research directions in applying GenAI to educational settings. Beyond summarizing existing work, we analyze predominant trends, examining the strengths and limitations of current research. Additionally, we explore insights from these studies on the compatibility of different GenAI models with various educational tasks.

4.3 Application Cases in Education

In this component, we introduce key applications of GenAI in education. Specifically, we highlight five active research directions in the field, presenting representative studies for each. By examining these works in detail, we aim to familiarize the audience with recent advancements in GenAI for education and inspire future research in this evolving area.

4.3.1 Grading and Feedback

Automatic grading systems have advanced significantly with the advent of GenAI. GenAI-powered systems have demonstrated promising results across various studies and subjects, including STEM and essay writing. Specifically, pioneering studies [36, 24] explore the use of large language models (LLMs) for automatic scoring of open-ended questions and essays, leveraging prompt-tuning algorithms. By incorporating comprehensive contexts, clear rubrics, and high-quality examples, LLMs have shown satisfactory performance in both grading tasks. Subsequent research [35] investigates the logical reasoning capabilities of LLMs by integrating Chain-of-Thought (CoT) reasoning into the grading process. This approach instructs LLMs to first analyze and explain the provided materials before making final score determinations. With such modifications, LLMs not only generate scores but also provide detailed comments on students' responses, helping them understand how to improve in future attempts. Beyond these developments, many recent studies explore the integration of advanced techniques such as automatic prompt optimization [6] and multi-agent systems [37] to further enhance grading performance. In addition to assigning scores, recent research has explored the use of GenAI for generating formative feedback, producing customized explanations and suggestions that have received positive responses from students. In essay writing, AI-powered systems provide paragraph-level suggestions by cross-referencing rubric deficiencies with genre-specific writing models [29]. In coding education, models can diagnose errors based on grading results and generate line-by-line debugging advice [14]. Furthermore, some studies investigate methods to align AI-generated grading and feedback more closely with human expert evaluations [22].

4.3.2 Content Creation

Renowned for its powerful generative capabilities, GenAI masters the skills of generating human-like content in multiple modals, especially the text. GenAI can integrate information from its pre-training data and task-specific prompts, including domain knowledge, pedagogical strategies, rules for formats and tones, etc. For example, LLMs can be instructed to generate learning materials that cover knowledge and ways to memorize, such like keyword mnemonics for language courses [16], dialogue scripts and stories [11], clinical vignettes for health education [3]. Besides, there are works leveraging GenAI to generate questions or tests to assess students' understanding of knowledge. In biology teaching, open-ended questions are generated by LLMs with curated prompts and related textbooks [34]. For mathematical exams, LLMs can generate multiple-choice questions (MCQ) and their options of high quality [9]. In addition, LLMs tailored for teaching are proposed to generate guidelines or hints for students' learning. SocraticLM [19] proposed a "Thought-Provoking" dialogue system based on the LLMs tuned on a Socratic-style dataset, which can provide analysis, knowledge and insightful questions in response to students' answers. EduChat [8] is an LLM-based chat bot specialized for Chinese middle schools, which is versatile in providing multiple kinds of teaching materials. In conclusion, GenAI is capable of generating various kinds of educational content encompassing both teaching and evaluation.

4.3.3 Education Assistant

Using an LLM-based chatbot as an educational assistant is a straightforward yet effective way to introduce GenAI into education, and recent studies have highlighted its numerous advantages and opportunities [8]. In general, LLM chatbots can tailor their responses to the individual needs of learners, providing personalized feedback and support. This customization accommodates different learning styles, speeds, and preferences. One key benefit is their 24/7 availability, making learning accessible anytime, anywhere. This is particularly advantageous for students in different time zones or those with varying schedules. Additionally, many studies have found that the interactive nature of chatbots enhances engagement and makes learning more enjoyable [18]. They can simulate conversations, create interactive learning scenarios, and provide instant feedback, which can be more effective than passive learning methods. Chatbots also offer scalability, as they can handle thousands of queries simultaneously, making them an efficient solution for educational institutions to support a large number of learners without requiring additional teaching staff. By automating repetitive tasks such as grading quizzes and providing basic feedback, they allow educators to focus on more complex and creative aspects of teaching. Moreover, with the integration of visual GenAI models, GenAI-powered education can also generate visual explanations for complex topics, further enriching the learning experience [15].

4.3.4 Role-Play Simulation

With abilities of understanding and generating, GenAI can be prompted to perform role playing by showing a wide variety of human-like behaviors. Beyond the functions mentioned in the prior sections, role-playing takes LLMs as non-deterministic simulator, which illustrating beliefs, emotions and goals to imitate an infinity of characters [28]. Generally, existing works design intelligent agents powered by LLMs with or without external tools to simulate humans in social context, predicting future actions of humans or providing helpful feedbacks for them [10]. SimClass [38] proposed a system of classroom simulation, in which agents performs as teachers, assistants, or students of different characteristics. In SimClass, agents play their roles following their human-like background and psychological patterns, thus allowing human users learn in an attractive and interactive manner. There are also works focusing on simulation of specific roles as surrogates of humans. For example, Virtual TA [20] is a simulation of computer science teachers. Supported by LLMs and auxiliary tools like document retrieval and text classification, Virtual TA can automatically judge the current situation and provide personalized feedback to students. By harnessing and implementing a wide range of abilities, GenAI underpins the foundations of intelligent agents and stimulates students' interest in learning.

4.3.5 Career Development

The broad knowledge base of GenAI models makes them not only suitable for STEM education but also recognized as valuable tools in various career development programs. Additionally, they have been increasingly integrated into higher education [1]. For example, the exceptional programming capabilities of LLMs have led to numerous studies exploring their use in computer programming education [13]. Specifically, research has investigated their role as teaching assistants in introductory programming courses, with many studies reporting positive student feedback on using GenAI for learning [21]. Furthermore, some works focus on evaluating the impact of different forms of assistance. Based on these findings, collaboration between students and GenAI models during programming education has demonstrated significant advantages over traditional textbook-based learning [12]. Beyond programming courses, GenAI applications have been explored in advanced educational subjects such as physics [18], medicine [32], and even career-oriented training, including vehicle driving skills [23]. By leveraging their generative capabilities, GenAI models enable a highly personalized learning experience, keeping students engaged and improving learning outcomes compared to control groups.

4.4 Demo Systems Experience

In this session, we will showcase two GenAI-powered demo systems we have developed, providing the audience with a hands-on experience of how GenAI is transforming education. The first system, I-VIP, is a teacher career education platform designed to offer an interactive training experience for educators. Unlike traditional teacher training programs that rely on rigid, rule-based frameworks, I-VIP leverages a dynamic Q&A format between large language models (LLMs) and participants. This approach enhances pedagogical skill development by simulating real-world interactions, making training more engaging and adaptive. Through this demo, we aim to demonstrate how GenAI is reshaping conventional education practices. The second system we will present is an LLM-powered automatic grading system. By showcasing its performance across various grading tasks, we will illustrate how LLMs can effectively handle logical reasoning and instruction-following, thereby improving grading efficiency and accuracy. This demonstration will provide insight into the practical applications of GenAI in streamlining assessment processes. Attendees are encouraged to actively engage with both demo systems by testing them with their own questions. We will also be available for discussions to address any questions or concerns. The systems will be accessible via our website, and testing user accounts will be provided at the beginning of the session.

4.5 Future Directions and Q&A

By the end of this tutorial, we will briefly summarize our discussion on the applications of Generative AI (GenAI) in education. Additionally, we will highlight key open questions in the field. Finally, we will hold a Q&A session to address any related questions from the audience.

4.6 Schedule Overview

1. Introduction to GenAI (40 min)
   - Concept Introduction
   - Common GenAI Models in various domains
   - Examples of GenAI Models in applications

   Break (10 min)
2. Overview of GenAI in Education (40 min)
   - Summarization of application directions
   - Discussion on challenges and opportunities

   Break (10 min)
3. Application Cases in Education I (60 min)
   - Grading and feedback
   - Content Creation
   - Education Assistant

   Lunch Break
4. Application Cases in Education II (30 min)
   - Role-play Simulation
   - Career Development

   Break (10 min)
5. Demo Systems Presentation (90 min)
   - Teacher career education system
   - Automatic grading system

   Break (10 min)
6. Future Directions and Q&A (30 min)

5. PRESENTERS

Hang Li. is a Ph.D. student at Michigan State University. He holds an M.S. in Statistics from the University of Illinois at Urbana-Champaign and a B.S. in Information and Computing Science from Beijing Jiaotong University. His research interests include Graph Neural Networks, Generative AI, and AI for Education. He has received several accolades, including 2nd Place in the OGB-LSC @ NeurIPS Node Classification Competition and 1st Place in the ACM Ubicomp STABILO Time Series Classification Challenge 2020. His prior research has been published in top-tier AI and education conferences, including AAAI, KDD, ACL, EMNLP, AIED, and EDM, among others. He regularly serves as a external reviewer for various data mining, natural language processing and machine learning conferences including ACL, AAAI, WWW, KDD, TKDE, IJCAI, etc.

Kaiqi Yang. is a Ph.D. student of Computer Science and Engineering at Michigan State University. He received a Master's degree in Applied Statistics and a Bachelor's degree in Sociology from Fudan University. His research interests include Social Computing, AI for Social Science, and Social Networks. His work has been accepted at leading conferences on data mining and natural language processing, such as CIKM, EMNLP, and AIED. He serves as a reviewer for ACL, TKDD, TKDE, CIKM, KDD, RecSys, etc.

***Yucheng Chu.*.** is a Ph.D. student at Michigan State University. She holds a B.S. in Computer Science from Columbia University. Her research interests including Generative AI, and AI for Education. Her prior research has been published in top-tier AI and education conferences such as AIED.

Jiliang Tang. is University Foundation Professor in the computer science and engineering department at Michigan State University. He got one early promotion to Associate Professor in 2021 and then a promotion to Full Professor (designated as MSU foundation professor) at 2022. Before that, he was a research scientist in Yahoo Research. He got his Ph.D. from Arizona State University in 2015 and MS and BE from Beijing Institute of Technology in 2010 and 2008, respectively. His research interests include graph machine learning, trustworthy AI, and their applications in Education. He authored the first comprehensive book "deep learning on graphs" with Cambridge University Press and developed various well-received open-sourced tools including scikit-feature for feature selection, DeepRobust for trustworthy AI and DANCE for single-cell analysis. He was the recipient of various career awards (2022 AI's 10 to Watch, 2022 IAPR J. K. AGGARWAL, 2022 SIAM SDM, 2021 IEEE ICDM, 2021 IEEE Big Data Security, 2020 ACM SIGKDD, 2019 NSF), numerous industrial faculty awards (Meta, JP Morgan, Amazon, Cisco, Johnson&Johnson, Criteo Labs and SNAP), and 8 best paper awards (or runner-ups) including WSDM2018 and KDD2016. He serves as conference organizers (e.g., KDD, SIGIR, WSDM and SDM) and journal editors (e.g., TKDD, TKDE and TOIS). He has published his research in highly ranked journals and top conference proceedings, which have 43,000+ citations with h-index 98 and extensive media coverage.

6. AUDIENCE PARTICIPATION AND INTERACTIVITY

We plan to ask some questions to the audience during the tutorial presentation, especially inspiring our audience to critically and actively consider the weakness of the defending algorithms. In this way, we can hopefully help the audience grip the ideas of our introduced topic of adversarial learning. In addition, we plan to allocate 5-10 minutes for questions and discussions after each section of the tutorial.

7. EQUIPMENT

The standard equipment normally available at a conference venue should be sufficient for the tutorial and no additional equipment is needed. We will bring a laptop and pointer with us. Attendees are encouraged to bring their laptops and smartphones for the demo session.

8. TUTORIAL WEBSITE

`https://korbiny.github.io/tutorials/EDM2025/`

9. REFERENCES

[1] N. P. Bakas, M. Papadaki, E. Vagianou, I. Christou, and S. A. Chatzichristofis. Integrating llms in higher education, through interactive problem solving and tutoring: Algorithmic approach and use cases. In *European, Mediterranean, and Middle Eastern*

Conference on Information Systems, pages 291–307. Springer, 2023.

[2] F. Barreto, L. Moharkar, M. Shirodkar, V. Sarode, S. Gonsalves, and A. Johns. Generative artificial intelligence: Opportunities and challenges of large language models. In *International conference on intelligent computing and networking*, pages 545–553. Springer, 2023.

[3] J. R. Benoit. Chatgpt for clinical vignette generation, revision, and evaluation. *MedRxiv*, pages 2023–02, 2023.

[4] T. Brooks, B. Peebles, C. Holmes, W. DePue, Y. Guo, L. Jing, D. Schnurr, J. Taylor, T. Luhman, E. Luhman, et al. Video generation models as world simulators. 2024. *URL https://openai.com/research/video-generation-models-as-world-simulators*, 3:1, 2024.

[5] T. K. Chiu. The impact of generative ai (genai) on practices, policies and research direction in education: A case of chatgpt and midjourney. *Interactive Learning Environments*, 32(10):6187–6203, 2024.

[6] Y. Chu, H. Li, K. Yang, H. Shomer, H. Liu, Y. Copur-Gencturk, and J. Tang. A llm-powered automatic grading framework with human-level guidelines optimization. *arXiv preprint arXiv:2410.02165*, 2024.

[7] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10850–10869, 2023.

[8] Y. Dan, Z. Lei, Y. Gu, Y. Li, J. Yin, J. Lin, L. Ye, Z. Tie, Y. Zhou, Y. Wang, et al. Educhat: A large-scale language model-based chatbot system for intelligent education. *arXiv preprint arXiv:2308.02773*, 2023.

[9] W. Feng, J. Lee, H. McNichols, A. Scarlatos, D. Smith, S. Woodhead, N. O. Ornelas, and A. Lan. Exploring automated distractor generation for math multiple-choice questions via large language models. *arXiv preprint arXiv:2404.02124*, 2024.

[10] Ö. Gürcan. Llm-augmented agent-based modelling for social simulations: Challenges and opportunities. *HHAI 2024: Hybrid Human AI Systems for the Social Good*, pages 134–144, 2024.

[11] J. Jeon and S. Lee. Large language models in education: A focus on the complementary relationship between human teachers and chatgpt. *Education and Information Technologies*, 28(12):15873–15892, 2023.

[12] M. Kazemitabaar, X. Hou, A. Henley, B. J. Ericson, D. Weintrop, and T. Grossman. How novices use llm-based code generators to solve cs1 coding tasks in a self-paced learning environment. In *Proceedings of the 23rd Koli calling international conference on computing education research*, pages 1–12, 2023.

[13] M. Kazemitabaar, R. Ye, X. Wang, A. Z. Henley, P. Denny, M. Craig, and T. Grossman. Codeaid: Evaluating a classroom deployment of an llm-based programming assistant that balances student and educator needs. In *Proceedings of the 2024 chi conference on human factors in computing systems*, pages 1–20, 2024.

[14] J. Y. Khan and G. Uddin. Automatic code documentation generation using gpt-3. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, pages 1–6, 2022.

[15] N. W. Kim, H.-K. Ko, G. Myers, and B. Bach. Chatgpt in data visualization education: A student perspective. In *2024 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 109–120. IEEE, 2024.

[16] J. Lee, H. McNichols, and A. Lan. Exploring automated keyword mnemonics generation with large language models via overgenerate-and-rank. *arXiv preprint arXiv:2409.13952*, 2024.

[17] H. Li, T. Xu, C. Zhang, E. Chen, J. Liang, X. Fan, H. Li, J. Tang, and Q. Wen. Bringing generative ai to adaptive learning in education. *arXiv preprint arXiv:2402.14601*, 2024.

[18] A. Lieb and T. Goel. Student interaction with newtbot: An llm-as-tutor chatbot for secondary physics education. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2024.

[19] J. Liu, Z. Huang, T. Xiao, J. Sha, J. Wu, Q. Liu, S. Wang, and E. Chen. Socraticlm: Exploring socratic personalized teaching with large language models. *Advances in Neural Information Processing Systems*, 37:85693–85721, 2025.

[20] M. Liu and F. M'hiri. Beyond traditional teaching: Large language models as simulated teaching assistants in computer science. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, pages 743–749, 2024.

[21] W. Lyu, Y. Wang, T. Chung, Y. Sun, and Y. Zhang. Evaluating the effectiveness of llms in introductory computer science education: A semester-long field study. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale*, pages 63–74, 2024.

[22] J. K. Matelsky, F. Parodi, T. Liu, R. D. Lange, and K. P. Kording. A large language model-assisted education tool to provide feedback on open-ended responses. *arXiv preprint arXiv:2308.02439*, 2023.

[23] M. Murtaza, C.-T. Cheng, M. Fard, and J. Zeleznikow. Transforming driver education: a comparative analysis of llm-augmented training and conventional instruction for autonomous vehicle technologies. *International Journal of Artificial Intelligence in Education*, pages 1–38, 2024.

[24] G. Pinto, I. Cardoso-Pereira, D. Monteiro, D. Lucena, A. Souza, and K. Gama. Large language models for education: Grading open-ended questions using chatgpt. In *Proceedings of the XXXVII Brazilian Symposium on Software Engineering*, pages 293–302, 2023.

[25] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*, pages 28492–28518. PMLR, 2023.

[26] S. S. Sengar, A. B. Hasan, S. Kumar, and F. Carroll. Generative artificial intelligence: a systematic review and applications. *Multimedia Tools and Applications*, pages 1–40, 2024.

[27] S. Shailendra, R. Kadel, and A. Sharma. Framework

for adoption of generative artificial intelligence (genai) in education. *IEEE Transactions on Education*, 2024.

[28] M. Shanahan, K. McDonell, and L. Reynolds. Role play with large language models. *Nature*, 623(7987):493–498, 2023.

[29] Y. Song, Q. Zhu, H. Wang, and Q. Zheng. Automated essay scoring and revising based on open-source large language models. *IEEE Transactions on Learning Technologies*, 2024.

[30] J. Sun, C. Dai, Z. Luo, Y. Chang, and Y. Li. Lawluo: A chinese law firm co-run by llm agents. *arXiv preprint arXiv:2407.16252*, 2024.

[31] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting. Large language models in medicine. *Nature medicine*, 29(8):1930–1940, 2023.

[32] E. Waisberg, J. Ong, M. Masalkhi, and A. G. Lee. Large language model (llm)-driven chatbots for neuro-ophthalmic medical education. *Eye*, 38(4):639–641, 2024.

[33] S. Wang, T. Xu, H. Li, C. Zhang, J. Liang, J. Tang, P. S. Yu, and Q. Wen. Large language models for education: A survey and outlook. *arXiv preprint arXiv:2403.18105*, 2024.

[34] Z. Wang, J. Valdez, D. Basu Mallick, and R. G. Baraniuk. Towards human-like educational question generation with large language models. In *International conference on artificial intelligence in education*, pages 153–166. Springer, 2022.

[35] C. Xiao, W. Ma, S. X. Xu, K. Zhang, Y. Wang, and Q. Fu. From automation to augmentation: Large language models elevating essay scoring landscape. *arXiv preprint arXiv:2401.06431*, 2024.

[36] K. P. Yancey, G. Laflair, A. Verardi, and J. Burstein. Rating short l2 essays on the cefr scale with gpt-4. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 576–584, 2023.

[37] K. Yang, Y. Chu, T. Darwin, A. Han, H. Li, H. Wen, Y. Copur-Gencturk, J. Tang, and H. Liu. Content knowledge identification with multi-agent large language models (llms). In *International Conference on Artificial Intelligence in Education*, pages 284–292. Springer, 2024.

[38] Z. Zhang, D. Zhang-Li, J. Yu, L. Gong, J. Zhou, Z. Hao, J. Jiang, J. Cao, H. Liu, Z. Liu, et al. Simulating classroom education with llm-empowered agents. *arXiv preprint arXiv:2406.19226*, 2024.

[39] H. Zhao, Z. Liu, Z. Wu, Y. Li, T. Yang, P. Shu, S. Xu, H. Dai, L. Zhao, G. Mai, et al. Revolutionizing finance with llms: An overview of applications and insights. *arXiv preprint arXiv:2401.11641*, 2024.

[40] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 1(2), 2023.