Publications included in "Generative Artificial Intelligence in Secondary STEM Education in the Light of Human Flourishing: A Scoping Literature Review"

Abolnejadian, M., Alipour, S., & Taeb, K. (2024). Leveraging ChatGPT for Adaptive Learning through Personalized Prompt-based Instruction: A CS1 Education Case Study. In F. F. Mueller, P. Kyburz, J. R. Williamson, & C. Sas (Eds.), *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1–8). ACM. https://doi.org/10.1145/3613905.3637148

Abramski, K., Citraro, S., Lombardi, L., Rossetti, G., & Stella, M. (2023). Cognitive Network Science Reveals Bias in GPT-3, GPT-3.5 Turbo, and GPT-4 Mirroring Math Anxiety in High-School Students. *Big Data and Cognitive Computing*, *7*(3). https://doi.org/10.3390/bdcc7030124

Adeyele, V. O., & Ramnarain, U. (2024). Exploring the Integration of ChatGPT in Inquiry-based Learning: Teacher Perspectives. *International Journal of Technology in Education*, *7*(2), 200–217. https://doi.org/10.46328/ijte.638

Alabidi, S., Alarabi, K., Alsalhi, N. R., & Mansoori, M. A. (2023). The Dawn of ChatGPT: Transformation in Science Assessment. *Eurasian Journal of Educational Research*, *2023*(106), 321–337. https://doi.org/10.14689/ejer.2023.106.019

Alarbi, K., Halaweh, M., Tairab, H., Alsalhi, N. R., Annamalai, N., & Aldarmaki, F. (2024). Making a revolution in physics learning in high schools with ChatGPT: A case study in UAE. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(9). https://doi.org/10.29333/ejmste/14983

Alhazzani, N. S. (2024). Enhancing mathematics teachers' pedagogical skills by using ChatGPT. *International Journal of Innovative Research and Scientific Studies*, *7*(4), 1614–1627. https://doi.org/10.53894/ijirss.v7i4.3460

Ali, S., DiPaola, D., Lee, I., Hong, J [Jenna], & Breazeal, C. (2021). Exploring Generative Models with Middle School Students. In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, & S. Drucker (Eds.), *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). ACM. https://doi.org/10.1145/3411764.3445226

Anand, A., Addala, K., Baghel, K., Goel, A., Hira, M., Gupta, R., & Shah, R. R. (2023). Revolutionizing High School Physics Education: A Novel Dataset. In V. Goyal, N. Kumar, S. S. Bhowmick, P. Goyal, N. Goyal, & D. Kumar (Eds.), *Lecture Notes in Computer Science. Big Data and Artificial Intelligence* (Vol. 14418, pp. 64–79). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-49601-1_5

Anghel, E., Wang, Y [Yu], Gopalakrishnan, M., Mansukhani, P., Bergner, Y., & Benedetto L. Caines A. Duenas G. Galvan-Sosa D. Loukina A. Taslimipoor S. Zesch T. (2024). *Can LLMs evaluate items measuring collaborative problem-solving?* In *CEUR Workshop Proceedings*. https://ceur-ws.org/Vol-3772/paper2.pdf

Araujo, J. L., & Saude, I. (2024). Can ChatGPT Enhance Chemistry Laboratory Teaching? Using Prompt Engineering to Enable AI in Generating Laboratory Activities. *JOURNAL of CHEMICAL EDUCATION*, *101*(5), 1858–1864. https://doi.org/10.1021/acs.jchemed.3c00745

Avila, K. E., Steinert, S., Ruzika, S., Kuhn, J., & Küchemann, S. (2024). Using ChatGPT for teaching physics. *PHYSICS TEACHER*, *62*(6), 536–537.

Bainbridge, K., Walkington, C., Ibrahim, A., Zhong, I., Mallick, D. B., Washington, J., Baraniuk, R [Rich], & Moore S. Stamper J. Tong R. Cao C. Liu Z. Hu X. Lu Y. Liang J. Khosravi H. Denny P. Singh A. Brooks C. (2023). A Case Study Using Large Language Models to Generate Metadata for Math Questions. In *CEUR Workshop Proceedings*. https://ceur-ws.org/Vol-3487/short5.pdf

Barany, A., Nasiar, N., Porter, C., Zambrano, A. F., Andres, A. L., Bright, D., Shah, M., Liu, X., Gao, S., Zhang, J., Mehta, S., Choi, J., Giordano, C., & Baker, R. S. (2024). ChatGPT for

Education Research: Exploring the Potential of Large Language Models for Qualitative Codebook Development. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Lecture Notes in Computer Science. Artificial Intelligence in Education* (Vol. 14830, pp. 134–149). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64299-9_10

Barno, E., Albaladejo-González, M., & Reich, J. (2024). Scaling Generated Feedback for Novice Teachers by Sustaining Teacher Educators' Expertise: A Design to Train LLMs with Teacher Educator Endorsement of Generated Feedback. In D. Joyner, M. K. Kim, X. Wang, & M. Xia (Eds.), *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (pp. 412–416). ACM. https://doi.org/10.1145/3657604.3664677

Baytak, A. (2024). The Content Analysis of the Lesson Plans Created by ChatGPT and Google Gemini. *Research in Social Sciences and Technology*, *9*(1), 329–350. https://doi.org/10.46303/ressat.2024.19

Beege, M., Hug, C., & Nerb, J. (2024). AI in STEM education: The relationship between teacher perceptions and ChatGPT use. *Computers in Human Behavior Reports*, *16.* https://doi.org/10.1016/j.chbr.2024.100494

Bellettini, C., Lodi, M., Lonati, V., Monga, M., & Morpurgo, A. (2023). Davinci Goes to Bebras: A Study on the Problem Solving Ability of GPT-3. In *Proceedings of the 15th International Conference on Computer Supported Education* (pp. 59–69). SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0012007500003470

Bešlić, A., Bešlić, J., & Kamber Hamzić, D. (2024). Artificial Intelligence in Elementary Math Education: Analyzing Impact on Students Achievements. In T. Volarić, B. Crnokić, & D. Vasić (Eds.), *Communications in Computer and Information Science. Digital Transformation in Education and Artificial Intelligence Application* (Vol. 2124, pp. 27–40). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-62058-4_3

Bešlić, J., Bešlić, A., Haverić, A., & Gajević, M. (2024). Leveraging AI in the Classroom to Transform Adolescent Dietary Habits Through Innovative Nutritional Education. *SN Computer Science*, *5*(757). https://doi.org/10.1007/s42979-024-03119-2

Bessas, N., Tzanaki, E., Vavougios, D., & Plagianakos, V. P. (2023). Implementing AI in Physics Lessons in the High School. In *2023 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1775–1779). IEEE. https://doi.org/10.1109/CSCI62032.2023.00293

Bewersdorff, A., Hartmann, C., Hornberger, M., Seßler, K., Bannert, M., Kasneci, E., Kasneci, G., Zhai, X., & Nerdel, C. (2024). *Taking the Next Step with Generative Artificial Intelligence: The Transformative Role of Multimodal Large Language Models in Science Education*. https://arxiv.org/abs/2401.00832v3

Bewersdorff, A., Seßler, K., Baur, A., Kasneci, E., & Nerdel, C. (2023). Assessing student errors in experimentation using artificial intelligence and large language models: A comparative study with human raters. *Computers and Education: Artificial Intelligence*, *5.* https://doi.org/10.1016/j.caeai.2023.100177

Bitzenbauer, P. (2023). ChatGPT in Physics Education: A Pilot Study on Easy-to-Implement Activities. *Contemporary Educational Technology*, *15*(3). https://doi.org/10.30935/cedtech/13176

Blonder, R., & Feldman-Maggor, Y. (2024). AI for chemistry teaching: responsible AI and ethical considerations. *Chemistry Teacher International*, *6*(4), 385–395. https://doi.org/10.1515/cti-2024-0014

Borisova, M., Hadzhikoleva, S., & Hadzhikolev, E. (2023). Use of Artificial Intelligence technologies in studying the phenomenon of electric current in physics education. In *Proceedings of the International Conference on Virtual Learning - VIRTUAL LEARNING - VIRTUAL REALITY (18th edition)* (pp. 215–224). The National Institute for Research & Development in Informatics - ICI Bucharest (ICI Publishing House). https://doi.org/10.58503/icvl-v18y202318

Borovský, D., Hanč, J., & Hančová, M. (2024). Innovative approaches to high school physics competitions: Harnessing the power of AI and open science. *Journal of Physics: Conference Series*, *2715*(1). https://doi.org/10.1088/1742-6596/2715/1/012011

Bradford, A., Li, W [Weiying], Gerard, L., & Linn, M. C. (2024). Comparing Expert and ChatGPT-authored Guidance Prompts. In D. Joyner, M. K. Kim, X. Wang, & M. Xia (Eds.), *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (pp. 388–392). ACM. https://doi.org/10.1145/3657604.3664669

Butgereit, L., & van Staden, A. (2023). An AI Contribution to African Economic Development Through Mother Tongue Education - A Linguistic Evaluation of an African-Based AI Mathematics Tutoring Bot. In *2023 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)* (pp. 7–12). IEEE. https://doi.org/10.1109/ICT4DA59526.2023.10302241

Canonigo, A. M. (2024). Levering AI to enhance students' conceptual understanding and confidence in mathematics. *JOURNAL of COMPUTER ASSISTED LEARNING*, *40*, 3215–3229. https://doi.org/10.1111/jcal.13065

Chen, C.-H. (2023). The Role of Students' Help-seeking Profiles in the AI-assisted Game-based Learning. In J.-L. Shih, A. Kashihara, W. Chen, & H. Ogata (Chairs), *International Conference on Computers in Education*. https://doi.org/10.58459/icce.2023.1467

Chen, C.-H., & Chang, C.-L. (2024). Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior. *EDUCATION and INFORMATION TECHNOLOGIES*, *29*, 18621–18642. https://doi.org/10.1007/s10639-024-12553-x

Chen, Y [Yanxin], & He, L. (2024). *Research on the Application of Large Language Models in Automatic Question Generation: A Case Study of ChatGLM in the Context of High School Information Technology Curriculum.* https://arxiv.org/abs/2408.11539

Chen, C.-H., & Shih J.-L. Kashihara A. Chen W. Chen W. Ogata H. Baker R. Chang B. Dianati S. Madathil J. Yousef A.M.F. Yang Y. Zarzour H. (2023). Exploring the Benefits of Chatbots Game-based Learning in Science Learning Outcomes and Motivation. In J.-L. Shih, A. Kashihara, W. Chen, & H. Ogata (Chairs), *International Conference on Computers in Education*. https://doi.org/10.58459/icce.2023.1418

Chen, E., Wang, D., Xu, L., Cao, C., Fang, X., & Lin, J. (2024). *A Systematic Review on Prompt Engineering in Large Language Models for K-12 STEM Education*. https://arxiv.org/abs/2410.11123v1

Chen, Y [Yuyan], Wu, C., Yan, S., Liu, P., & Xiao, Y. (2024). Dr.Academy: A Benchmark for Evaluating Questioning Capability in Education for Large Language Models. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 3138–3167). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.173

Cheung, K. K. C., Pun, J. K. H., & Li, W [Wangyin] (2024). Students' Holistic Reading of Socio-Scientific Texts on Climate Change in a ChatGPT Scenario. *RESEARCH in SCIENCE EDUCATION*, *54*(5), 957–976. https://doi.org/10.1007/s11165-024-10177-2

Cohn, C., Hutchins, N., Le, T., Biswas, G., & Wooldridge M. Dy J. Natarajan S. (2024). A Chain-of-Thought Prompting Approach with LLMs for Evaluating Students' Formative Assessment Responses in Science. In M. Wooldridge, J. Dy, & S. Natarajan (Chairs), *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence* (pp. 23182 – 23190). Association for Computational Linguistics. https://doi.org/10.1609/aaai.v38i21.30364

Cohn, C., Snyder, C., Montenegro, J., & Biswas, G. (2024). Towards a Human-in-the-Loop LLM Approach to Collaborative Discourse Analysis. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Communications in Computer and Information Science.*

Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky (Vol. 2151, pp. 11–19). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64312-5_2

Cohn, C., Snyder, C., Fonteles, J. H., T. S., A., Montenegro, J., & Biswas, G. (2025). A multimodal approach to support teacher, researcher and AI collaboration in STEM +C learning environments. *BRITISH JOURNAL of EDUCATIONAL TECHNOLOGY*, *56*(2), 595–620. https://doi.org/10.1111/bjet.13518

Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *JOURNAL of SCIENCE EDUCATION and TECHNOLOGY*, *32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Cooper, G., & Tang, K.-S. (2024). Pixels and Pedagogy: Examining Science Education Imagery by Generative Artificial Intelligence. *JOURNAL of SCIENCE EDUCATION and TECHNOLOGY*, *33*(4), 556–568. https://doi.org/10.1007/s10956-024-10104-0

Daher, W., & Gierdien, F. (2024). Use of Language By generative AI Tools in Mathematical Problem Solving: The Case of ChatGPT. *African Journal of Research in Mathematics, Science and Technology Education*, *28*(2), 222–235. https://doi.org/10.1080/18117295.2024.2384676

Dao, X.-Q., & Le, N.-B. (2023). *Investigating the Effectiveness of ChatGPT in Mathematical Reasoning and Problem Solving: Evidence from the Vietnamese National High School Graduation Examination*. https://arxiv.org/abs/2306.06331v3

Dao, X.-Q., & Le, N.-B. (2023). LLMs Performance on Vietnamese High School Biology Examination. *International Journal of Modern Education and Computer Science*, *15*(6), 14–30. https://doi.org/10.5815/ijmecs.2023.06.02

Dao, X.-Q., Le, N.-B., Vo, T.-D., Phan, X.-D., Ngo, B.-B., Nguyen, V.-T., Nguyen, T.-M.-T., & Nguyen, H.-P. (2023). *VNHSGE: VietNamese High School Graduation Examination Dataset for Large Language Models*. https://arxiv.org/abs/2305.12199v1

Dao, X.-Q., Le, N.-B., Phan, X.-D., Ngo, B.-B., & Vo, T.-D. (2023). *Evaluation of ChatGPT and Microsoft Bing AI Chat Performances on Physics Exams of Vietnamese National High School Graduation Examination*. https://arxiv.org/abs/2306.04538v3

Dasari, D., Hendriyanto, A., Sahara, S., Suryadi, D., Muhaimin, L. H., Chao, T., & Fitriana, L. (2024). ChatGPT in didactical tetrahedron, does it make an exception? A case study in mathematics teaching and learning. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1295413

Dengel, A., Gehrlein, R., Fernes, D., Görlich, S., Maurer, J., Pham, H. H., Großmann, G., & Eisermann, N. D. g. (2023). Qualitative Research Methods for Large Language Models: Conducting Semi-Structured Interviews with ChatGPT and BARD on Computer Science Education. *Informatics*, *10*(4). https://doi.org/10.3390/informatics10040078

Dimmel, J. K., & Bayyurt, I. (2023). Rise of the Machines: Navigating the Opportunities and Challenges of AI-Assisted Research and Learning. In T. Lamberg & D. Moss (Chairs), *Proceedings of the forty-fifth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. https://files.eric.ed.gov/fulltext/ED658455.pdf

Ding, J., Cen, Y., & Wei, X [Xinyuan]. (2023). *Using Large Language Model to Solve and Explain Physics Word Problems Approaching Human Level*. https://arxiv.org/abs/2309.08182v2

dos Anjos, J. R., Souza, M. G. de, Andrade Neto, A. S. de, & Souza, B. C. de (2024). AN ANALYSIS OF THE GENERATIVE AI USE AS ANALYST IN QUALITATIVE RESEARCH IN SCIENCE EDUCATION; [UMA ANÁLISE DO USO DE IA GENERATIVA COMO INVESTIGADORA EM PESQUISAS QUALITATIVAS EM EDUCAÇÃO CIENTÍFICA]. *Revista Pesquisa Qualitativa*, *12*(30). https://doi.org/10.33361/RPQ.2024.v.12.n.30.724

Dos Santos, R. P. (2023) *Enhancing Physics Learning with ChatGPT, Bing Chat, and Bard as Agents-to-Think-With: A Comparative Case Study*. https://arxiv.org/abs/2306.00724v1

dos Santos, R. P. (2023). Enhancing Chemistry Learning with ChatGPT and Bing Chat as Agents to Think With: A Comparative Case Study. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4447416

dos Santos, R. P. (2023). *Enhancing Chemistry Learning with ChatGPT, Bing Chat, Bard, and Claude as Agents-to-Think-With: A Comparative Case Study*. https://arxiv.org/abs/2311.00709v1

Drushlyak, M., Lukashova, T., Sabadosh, Y., Melnikov, I., & Semenikhina, O. (2024). Using ChatGPT for the Development of Critical Thinking in Youth: Example of Inequality Proof. In *2024 47th MIPRO ICT and Electronics Convention (MIPRO)* (pp. 334–339). IEEE. https://doi.org/10.1109/MIPRO60963.2024.10569759

Due, S., Das, S., Andersen, M., López, B. P., Nexø, S. A., & Clemmensen, L. (2024). *Evaluation of Large Language Models: STEM education and Gender Stereotypes*. https://arxiv.org/abs/2406.10133v1

Dumitran, A. M., Badea, A. C., & Muscalu, S.-G. (2024). Evaluating the Performance of Large Language Models in Competitive Programming: A Multi-Year, Multi-Grade Analysis. In *2024 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)* (pp. 1–7). IEEE. https://doi.org/10.1109/INISTA62901.2024.10683837

Egara, F. O., & Mosimege, M. (2024). Exploring the Integration of Artificial Intelligence-Based ChatGPT into Mathematics Instruction: Perceptions, Challenges, and Implications for Educators. *Education Sciences*, *14*(7). https://doi.org/10.3390/educsci14070742

Ergene, O., & Ergene, B. C. (2025). AI ChatBots' solutions to mathematical problems in interactive e-textbooks: Affordances and constraints from the eyes of students and teachers. *EDUCATION and INFORMATION TECHNOLOGIES*, *30*(1), 509–545. https://doi.org/10.1007/s10639-024-13121-z

Exintaris, B., Karunaratne, N., & Yuriev, E. (2023). Metacognition and Critical Thinking: Using ChatGPT-Generated Responses as Prompts for Critique in a Problem-Solving Workshop (SMARTCHEMPer). *JOURNAL of CHEMICAL EDUCATION*, *100*(8), 2972–2980. https://doi.org/10.1021/acs.jchemed.3c00481

Fang, L., Lee, G.-G., & Zhai, X. (2023). *Using GPT-4 to Augment Unbalanced Data for Automatic Scoring*. https://arxiv.org/abs/2310.18365v3

Farhana, E., Sarkar, S., Knipper, R., Dey, I., Narayanan, H., Puntambekar, S., & Karmaker, S. (2024). SimPal: Towards a Meta-Conversational Framework to Understand Teacher's Instructional Goals for K-12 Physics. In D. Joyner, M. K. Kim, X. Wang, & M. Xia (Eds.), *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (pp. 461–465). ACM. https://doi.org/10.1145/3657604.3664695

Feldman-Maggor, Y., Blonder, R., & Alexandron, G. (2025). Perspectives of Generative AI in Chemistry Education Within the TPACK Framework. *JOURNAL of SCIENCE EDUCATION and TECHNOLOGY*, *34*(1), 1–12. https://doi.org/10.1007/s10956-024-10147-3

Feng, W., Lee, J., McNichols, H., Scarlatos, A., Smith, D., Woodhead, S., Ornelas, N. O., & Lan, A. (2024). *Exploring Automated Distractor Generation for Math Multiple-choice Questions via Large Language Models*. https://arxiv.org/abs/2404.02124v3

Fernandez, N., Scarlatos, A., Feng, W., Woodhead, S., & Lan, A. (2024). *DiVERT: Distractor Generation with Variational Errors Represented as Text for Math Multiple-choice Questions*. https://arxiv.org/abs/2406.19356v2

Feser, M. S. (2024). PARENTS'VIEWS ON THE USE OF AI-BASED CHATBOTS SUCH AS CHATGPT IN HIGH SCHOOL (STEM) EDUCATION. *JOURNAL of BALTIC SCIENCE EDUCATION*, *23*(1), 4–8. https://doi.org/10.33225/jbse/24.23.04

Frazier, M., Damevski, K., & Pollock, L. (2024). Customizing ChatGPT to Help Computer Science Principles Students Learn Through Conversation. In M. Monga, V. Lonati, E. Barendsen, J. Sheard, & J. Paterson (Eds.), *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1* (pp. 633–639). ACM. https://doi.org/10.1145/3649217.3653570

Garofalo, S. G., & Farenga, S. J. (2025). Science Teacher Perceptions of the State of Knowledge and Education at the Advent of Generative Artificial Intelligence Popularity. *SCIENCE & EDUCATION*, *34*(2), 893–912. https://doi.org/10.1007/s11191-024-00534-y

Gattupalli, S., Lee, W., Allessio, D., Crabtree, D., Arroyo, I., Woolf, B., & Moore S. Stamper J. Tong R. Cao C. Liu Z. Hu X. Lu Y. Liang J. Khosravi H. Denny P. Singh A. Brooks C. (2023). Exploring Pre-Service Teachers' Perceptions of Large Language Models-Generated Hints in Online Mathematics Learning. In *CEUR Workshop Proceedings*. https://ceur-ws.org/Vol-3487/paper10.pdf

Ghazali, S. A., Zaki, N., Ali, L., & Harous, S. (2024). Exploring the Potential of ChatGPT as a Substitute Teacher: A Case Study. *International Journal of Information and Education Technology*, *14*(2), 271–278. https://doi.org/10.18178/ijiet.2024.14.2.2048

Golding, Y. (2024). *Extending Interactive Science Exhibits into the Classroom using Anthropomorphized Chatbots and Bloom's Taxonomy*. https://arxiv.org/abs/2402.01770v1

Gottam, V., Dehbozorgi, N., & Lee, S. (2024). AI-Based Discourse Analysis System (ADAS) for Improved STEM Education. In *2024 IEEE Integrated STEM Education Conference (ISEC)* (pp. 1–4). IEEE. https://doi.org/10.1109/ISEC61299.2024.10665112

Govender, R. (2023). The impact of artificial intelligence and the future of ChatGPT for mathematics teaching and learning in schools and higher education. *Pythagoras*, *44*(1). https://doi.org/10.4102/PYTHAGORAS.V44I1.787

Gregorcic, B., & Pendrill, A.-M. (2023). ChatGPT and the Frustrated Socrates. *Physics Education*, *58*(3). https://doi.org/10.1088/1361-6552/acc299

Gregorcic, B., Polverini, G., & Sarlah, A. (2024). ChatGPT as a tool for honing teachers' Socratic dialogue skills. *Physics Education*, *59*(4), 45005. https://doi.org/10.1088/1361-6552/ad3d21

Guler, N. K., Dertli, Z. G., Boran, E., & Yildiz, B. (2024). An Artificial Intelligence Application in Mathematics Education: Evaluating ChatGPT's Academic Achievement in a Mathematics Exam. *Pedagogical Research*, *9*(2). https://doi.org/10.29333/pr/14145

Hadley-Hulet, A., Ellis, M., Moore, A., Lehnardt, E., & Longhurst, M. (2024). Using Lessons from History to Guide the Implementation of AI in Science Education. *Science Teacher*, *91*(2), 29–34. https://doi.org/10.1080/00368555.2024.2308315

Han, Z. F., Lin, J., Gurung, A., Thomas, D. R., Chen, E., Borchers, C., Gupta, S., Koedinger, K. R., & Ananda M. Malick D.B. Burstein J. Liu L.T. Liu Z. Sharpnack J. Wang Z. Wang S. (2024). Improving Assessment of Tutoring Practices using Retrieval-Augmented Generation. In M. Ananda, D. B. Malick, J. Burstein, L. T. Liu, Z. Liu, J. Sharpnack, . . . S. Wang (Chairs), *Proceedings of the 2024 AAAI Conference on Artificial Intelligence*. https://arxiv.org/abs/2402.14594

Hashem, R., Ali, N., Zein, F. E., Fidalgo, P., & Khurma, O. A. (2024). AI to the rescue: Exploring the potential of ChatGPT as a teacher ally for workload relief and burnout prevention. *Research and Practice in Technology Enhanced Learning*, *19*. https://doi.org/10.58459/rptel.2024.19023

Henkel, O., Horne-Robinson, H., Kozhakhmetova, N., & Lee, A. (2024). Effective and Scalable Math Support: Experimental Evidence on the Impact of an AI-Math Tutor in Ghana. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Communications in Computer and Information Science. Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky* (Vol. 2150, pp. 373–381). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64315-6_34

Henkel, O., Horne-Robinson, H., Dyshel, M., Ch, N., Moreau-Pernet, B., & Abood, R. (2024). *Learning to Love Edge Cases in Formative Math Assessment: Using the AMMORE Dataset and Chain-of-Thought Prompting to Improve Grading Accuracy*. https://arxiv.org/abs/2409.17904v1

Hess, C., & Kunz, S. (2024). The Soccer-Playing Unicorn – Mitigating Gender Bias in AI-Created STEM Teaching Materials. In C.-P. M. Ballester (Chair), *Proceedings of the 7th International Conference on Gender Research* (pp. 158–166). 10.34190/icgr.7.1.2058

Hu, B., Zheng, L., Zhu, J., Ding, L., Wang, Y [Yilei], & Gu, X. (2024). Teaching Plan Generation and Evaluation with GPT-4: Unleashing the Potential of LLM in Instructional Design. *IEEE TRANSACTIONS on LEARNING TECHNOLOGIES, 17*, 1471–1485. https://doi.org/10.1109/TLT.2024.3384765

Ji, Y., Zou, X., Li, T., & Zhan, Z. (2023). The effectiveness of ChatGPT on pre-service teachers' STEM teaching literacy, learning performance, and cognitive load in a teacher training course. In *Proceedings of the 2023 6th International Conference on Educational Technology Management* (pp. 16–22). ACM. https://doi.org/10.1145/3637907.3637948

Jia, J., Wang, T., Zhang, Y [Yuyue], & Wang, G. (2024). The Comparison of General Tips for Mathematical Problem Solving Generated by Generative AI with Those Generated by Human Teachers. *ASIA PACIFIC JOURNAL of EDUCATION*, *44*(1), 8–28. https://doi.org/10.1080/02188791.2023.2286920

Jiang, Z., & Jiang, M. (2024). *Beyond Answers: Large Language Model-Powered Tutoring System in Physics Education for Deep Learning and Precise Understanding*. https://arxiv.org/abs/2406.10934v1

Jin, H., Yoo, M., Park, J., Lee, Y., Wang, X., & Kim, J [Juho]. (2024). *TeachTune: Reviewing Pedagogical Agents Against Diverse Student Profiles with Simulated Students*. https://arxiv.org/abs/2410.04078v1

Jr, M. V., Friday, M., & Zaghi, A. (2024). *Evaluating the capability of large language models to personalize science texts for diverse middle-school-age learners*. https://arxiv.org/abs/2408.05204v1

Kakarla, S., Thomas, D. R., Lin, J., Gupta, S., Koedinger, K. R., & Ananda M. Malick D.B. Burstein J. Liu L.T. Liu Z. Sharpnack J. Wang Z. Wang S. (2024). Using Large Language Models to Assess Tutors' Performance in Reacting to Students Making Math Errors. In M. Ananda, D. B. Malick, J. Burstein, L. T. Liu, Z. Liu, J. Sharpnack, . . . S. Wang (Chairs), *Proceedings of the 2024 AAAI Conference on Artificial Intelligence*.

Kazemitabaar, M., Chow, J., Ma, C. K. T., Ericson, B. J., Weintrop, D., & Grossman, T. (2023). Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming. In A. Schmidt, K. Väänänen, T. Goyal, P. O. Kristensson, A. Peters, S. Mueller, J. R. Williamson, & M. L. Wilson (Eds.), *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–23). ACM. https://doi.org/10.1145/3544548.3580919

Kharchenko, Y. V., Babenko, O. M., & Papadakis S. University of Crete, Gallos Campus, Crete (2024). Advantages and limitations of large language models in chemistry education: A comparative analysis of ChatGPT, Gemini and Copilot. In *CEUR Workshop Proceedings*. https://ceur-ws.org/Vol-3781/paper03.pdf

Kieser, F., Tschisgale, P., Rauh, S., Bai, X., Maus, H., Petersen, S., Stede, M., Neumann, K., & Wulff, P. (2024). David vs. Goliath: comparing conventional machine learning and a large language model for assessing students' concept use in a physics problem. *Frontiers in Artificial Intelligence*, *7*. https://doi.org/10.3389/frai.2024.1408817

Kieser, F., Wulff, P., Kuhn, J., & Kuechemann, S. (2023). Educational data augmentation in physics education research using ChatGPT. *PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH*, *19*(2). https://doi.org/10.1103/PhysRevPhysEducRes.19.020150

Kruis, J., Pera, M. S., Napel, Z. ten, Landoni, M., Murgia, E., Huibers, T., & Feskens, R. (2024). Toward Personalised Learning Experiences: Beyond Prompt Engineering. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference* (pp. 644–649). ACM. https://doi.org/10.1145/3628516.3659367

Küchemann, S., Steinert, S., Revenga, N., Schweinberger, M., Dinc, Y., Avila, K. E., & Kuhn, J. (2023). Can ChatGPT Support Prospective Teachers in Physics Task Development? *PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH*, *19*(2). https://doi.org/10.1103/PhysRevPhysEducRes.19.020128

Küchemann, S., Steinert, S., Kuhn, J., Avila, K., & Ruzika, S. (2024). Large language models—Valuable tools that require a sensitive integration into teaching and learning physics. *PHYSICS TEACHER*, *62*(5), 400–402. https://doi.org/10.1119/5.0212374

Kunz, S., & Steffen, A. (2024). Criticize my Code, not me: Using AI-generated Feedback in Computer Science Teaching. International Conference on Gender Research, 7(1), 200–207. https://doi.org/10.34190/icgr.7.1.2002

Latif, E., Parasuraman, R., & Zhai, X. (2024). PhysicsAssistant: An LLM-Powered Interactive Learning Robot for Physics Lab Investigations. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)* (pp. 864–871). IEEE. https://doi.org/10.1109/RO-MAN60168.2024.10731312

Latif, E., & Zhai, X. (2024). Fine-tuning ChatGPT for automatic scoring. *Computers and Education: Artificial Intelligence*, *6*, 100210. https://doi.org/10.1016/j.caeai.2024.100210

Lee, G.-G., Latif, E., Wu, X., Liu, N [Ninghao], & Zhai, X. (2024). Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence*, *6*, 100213. https://doi.org/10.1016/j.caeai.2024.100213

Lee, J., Smith, D., Woodhead, S., & Lan, A. (2024). *Math Multiple Choice Question Generation via Human-Large Language Model Collaboration.* 17th International Conference on Educational Data Mining. https://arxiv.org/abs/2405.00864

Lee, A. V. Y., Teo, C. L., & Tan, S. C. (2024). Prompt Engineering for Knowledge Creation: Using Chain-of-Thought to Support Students' Improvable Ideas. *AI*, *5*(3), 1446–1461. https://doi.org/10.3390/ai5030069

Levonian, Z., Li, C., Zhu, W., Gade, A., Henkel, O., Postle, M.-E., & Xing, W. (2023). *Retrieval-augmented Generation to Improve Math Question-Answering: Trade-offs Between Groundedness and Human Preference.* 17th International Conference on Educational Data Mining. https://arxiv.org/abs/2310.03184

Li, M. Z. (2024). Using Prompt Engineering to Enhance STEM Education. In *2024 IEEE Integrated STEM Education Conference (ISEC)* (pp. 1–2). IEEE. https://doi.org/10.1109/ISEC61299.2024.10664775

Li, Z., & Ironsi, C. S. (2024). The Efficacy of Generative Artificial Intelligence in Developing Science Education Preservice Teachers' Writing Skills: An Experimental Approach. *JOURNAL of SCIENCE EDUCATION and TECHNOLOGY.* Advance online publication. https://doi.org/10.1007/s10956-024-10148-2

Li, T., Ji, Y., & Zhan, Z. (2024). Expert or Machine? Comparing the Effect of Pairing Student Teacher with In-Service Teacher and ChatGPT on Their Critical Thinking, Learning Performance, and Cognitive Load in an Integrated-STEM Course. *ASIA PACIFIC JOURNAL of EDUCATION*, *44*(1), 45–60. https://doi.org/10.1080/02188791.2024.2305163

Li, R., Wang, Y [Yiting], Zheng, C., Jiang, Y.-H., & Jiang, B. (2024). Generating Contextualized Mathematics Multiple-Choice Questions Utilizing Large Language Models. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Communications in Computer and Information Science. Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky* (Vol. 2150, pp. 494–501). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64315-6_48

Li, Z.-Z., Zhang, M.-L., Yin, F., Ji, Z.-L., Bai, J.-F., Pan, Z.-R., Zeng, F.-H., Xu, J., Zhang, J.-X., & Liu, C.-L. (2024). *CMMaTH: A Chinese Multi-modal Math Skill Evaluation Benchmark for Foundation Models*. https://arxiv.org/abs/2407.12023v1

Liang, Y., Di Zou, Xie, H., & Wang, F. L. (2023). Exploring the Potential of Using ChatGPT in Physics Education. *Smart Learning Environments, 10.* https://doi.org/10.1186/s40561-023-00273-7

Lieb, A., & Goel, T. (2024). Student Interaction with NewtBot: An LLM-as-tutor Chatbot for Secondary Physics Education. In F. F. Mueller, P. Kyburz, J. R. Williamson, & C. Sas (Eds.), *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1–8). ACM. https://doi.org/10.1145/3613905.3647957

Lin, T., & Riccomini, P. (2024). Inclusive Education in Rural Settings: Leveraging Technology for Improved Math Learning Outcomes Among Students With Disabilities. *JOURNAL of SPECIAL EDUCATION TECHNOLOGY*, *40*(2), 219–226. https://doi.org/10.1177/01626434241263045

Liu, N [Naiming], Sonkar, S., Le, M., & Baraniuk, R [Richard] (2024). MalAlgoQA: Pedagogical Evaluation of Counterfactual Reasoning in Large Language Models and Implications for AI in Education. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 15554–15567). Association for Computational Linguistics. 10.18653/v1/2024.findings-emnlp.913

Long, Y., Luo, H., & Zhang, Y [Yu] (2024). Evaluating large language models in analysing classroom dialogue. *NPJ SCIENCE of LEARNING*, *9*(1). https://doi.org/10.1038/s41539-024-00273-3

Lopez-Simo, V., & Rezende, M. F. (2024). Challenging ChatGPT with Different Types of Physics Education Questions. *PHYSICS TEACHER*, *62*(4), 290–294. https://doi.org/10.1119/5.0160160

Lu, J., Zheng, R., Gong, Z., & Xu, H. (2024). Supporting Teachers' Professional Development with Generative AI: The Effects on Higher Order Thinking and Self-Efficacy. *IEEE TRANSACTIONS on LEARNING TECHNOLOGIES*, *17*, 1279–1289. https://doi.org/10.1109/TLT.2024.3369690

Mahon, J., Mac Namee, B., & Becker, B. A. (2023). No More Pencils No More Books: Capabilities of Generative AI on Irish and UK Computer Science School Leaving Examinations. In T. Astarte, F. Moller, K. Quille, & S. Russell (Eds.), *The United Kingdom and Ireland Computing Education Research (UKICER) conference* (pp. 1–7). ACM. https://doi.org/10.1145/3610969.3610982

Maj, S. P. (2024). Solving the global STEM educational crisis using Cognitive Load Optimization and Artificial Intelligence–A preliminary comparative analysis. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(5). https://doi.org/10.29333/ejmste/14448

Malik, R., Abdi, D., Wang, R., & Demszky, D. (2024). Scaling High-Leverage Curriculum Scaffolding in Middle-School Mathematics. In D. Joyner, M. K. Kim, X. Wang, & M. Xia (Eds.), *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (pp. 476–480). ACM. https://doi.org/10.1145/3657604.3664698

McNichols, H., Feng, W., Lee, J., Scarlatos, A., Smith, D., Woodhead, S., & Lan, A. (2023). *Automated Distractor and Feedback Generation for Math Multiple-choice Questions via In-context Learning.* NeurIPS'23 Workshop: Generative AI for Education (GAIED). https://arxiv.org/abs/2308.03234

McNichols, H., Lee, J., Fancsali, S., Ritter, S., & Lan, A. (2024). *Can Large Language Models Replicate ITS Feedback on Open-Ended Math Questions?* Educational Data Mining 2024. https://arxiv.org/abs/2405.06414v2

Monteiro, F. F., Souza, P. V. S., Da Silva, M. C., Maia, J. R., Da Silva, W. F., & Girardi, D. (2024). ChatGPT in Brazilian K-12 science education. *Frontiers in Education, 9.* https://doi.org/10.3389/feduc.2024.1321547

Moreau-Pernet, B., Tian, Y., Sawaya, S., Foltz, P., Cao, J., Milne, B., & Christie, T. (2024). Classifying Tutor Discursive Moves at Scale in Mathematics Classrooms with Large Language Models. In D. Joyner, M. K. Kim, X. Wang, & M. Xia (Eds.), *Proceedings of the*

Eleventh ACM Conference on Learning @ Scale (pp. 361–365). ACM. https://doi.org/10.1145/3657604.3664664

Nascimento Junior, Wilton J. D., Morais, C., & Girotto Junior, G. (2024). Enhancing AI Responses in Chemistry: Integrating Text Generation, Image Creation, and Image Interpretation through Different Levels of Prompts. *JOURNAL of CHEMICAL EDUCATION*, *101*(9), 3767–3779. https://doi.org/10.1021/acs.jchemed.4c00230

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering Student Self-Regulated Learning and Science Education through ChatGPT: A Pioneering Pilot Study. *BRITISH JOURNAL of EDUCATIONAL TECHNOLOGY*, *55*(4), 1328–1353. https://doi.org/10.1111/bjet.13454

Nicula, B., Dascalu, M., Arner, T., Balyan, R., & McNamara, D. S. (2023). Automated Assessment of Comprehension Strategies from Self-Explanations Using LLMs. *Information*, *14*(10). https://doi.org/10.3390/info14100567

Norberg, K. A., Almoubayyed, H., Ley, L. de, Murphy, A., Weldon, K., & Ritter, S. (2024). Rewriting Content with GPT-4 to Support Emerging Readers in Adaptive Mathematics Software. *International Journal of Artificial Intelligence in Education.* Advance online publication. https://doi.org/10.1007/s40593-024-00420-2

Okulu, H. Z., & Muslu, N. (2024). Designing a course for pre-service science teachers using ChatGPT: what ChatGPT brings to the table. *INTERACTIVE LEARNING ENVIRONMENTS*, *32*(10), 7450–7467. https://doi.org/10.1080/10494820.2024.2322462

Parra, V., Sureda, P., Corica, A., Schiaffino, S., & Godoy, D. (2024). Can Generative AI Solve Geometry Problems? Strengths and Weaknesses of LLMs for Geometric Reasoning in Spanish. *INTERNATIONAL JOURNAL of INTERACTIVE MULTIMEDIA and ARTIFICIAL INTELLIGENCE*, *8*(5), 65–74. https://doi.org/10.9781/ijimai.2024.02.009

Pavlova, N. H. (2024). Flipped Dialogic Learning Method with ChatGPT: A Case Study. *International Electronic Journal of Mathematics Education*, *19*(1). https://doi.org/10.29333/iejme/14025

Pernaa, J., Ikävalko, T., Takala, A., Vuorio, E., Pesonen, R., & Haatainen, O. (2024). Artificial Intelligence Chatbots in Chemical Information Seeking: Narrative Educational Insights via a SWOT Analysis. *Informatics*, *11*(2). https://doi.org/10.3390/informatics11020020

Pituxcoosuvarn, M., & Murakami, Y. (2024). Transforming Children's Python Turtle Graphics Learning with LLM Technology: A Design Proposal. In *2024 9th International STEM Education Conference (iSTEM-Ed)* (pp. 1–6). IEEE. https://doi.org/10.1109/iSTEM-Ed62750.2024.10663193

Polverini, G., & Gregorcic, B. (2024). Evaluating vision-capable chatbots in interpreting kinematics graphs: a comparative study of free and subscription-based models. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1452414

Polverini, G., & Gregorcic, B. (2024). How understanding large language models can inform the use of ChatGPT in physics education. *EUROPEAN JOURNAL of PHYSICS*, *45*(2). https://doi.org/10.1088/1361-6404/ad1420

Polverini, G., & Gregorcic, B. (2024). Performance of ChatGPT on the Test of Understanding Graphs in Kinematics. *PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH*, *20*(1). https://doi.org/10.1103/PhysRevPhysEducRes.20.010109

Prihar, E., Lee, M., Hopman, M., Tauman Kalai, A., Vempala, S., Wang, A., Wickline, G., Murray, A., & Heffernan, N. (2023). Comparing Different Approaches to Generating Mathematics Explanations Using Large Language Models. In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda & O.C. Santos (Eds.), Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky (Vol. 1831, pp. 290–295). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36336-8_45

Puech, R., Macina, J., Chatain, J., Sachan, M., & Kapur, M. (2024). *Towards the Pedagogical Steering of Large Language Models for Tutoring: A Case Study with Modeling Productive Failure*. https://arxiv.org/abs/2410.03781v1

Qadir, J. (2023). Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. In *2023 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1–9). IEEE. https://doi.org/10.1109/EDUCON54358.2023.10125121

Ramnarain, U., Ogegbo, A. A., Penn, M., Ojetunde, S., & Mdlalose, N. (2024). Pre-Service Science Teachers' Intention to use Generative Artificial Intelligence in Inquiry-Based Teaching. *JOURNAL of SCIENCE EDUCATION and TECHNOLOGY.* Advance online publication. https://doi.org/10.1007/s10956-024-10159-z

Rezende Junior, M. F., & López-Simó, V. (2024). What are the perceptions of physics teachers in Brazil about ChatGPT in school activities? *Journal of Physics: Conference Series*, *2693*(1). https://doi.org/10.1088/1742-6596/2693/1/012011

Ribeiro, A. R. A., Navarro, E. R., & Kalinke, M. A. (2024). USING CHATGPT TO SOLVE MATHEMATICAL PROBLEMS ABOUT DIRECTLY AND INVERSELY PROPORTIONAL VARIABLES; [O USO DO CHATGPT PARA RESOLVER PROBLEMAS MATEMÁTICOS SOBRE GRANDEZAS DIRETA E INVERSAMENTE PROPORCIONAIS]. *Revista Pesquisa Qualitativa*, *12*(30). https://doi.org/10.33361/RPQ.2024.v.12.n.30.716

Rodrigues, L., Dwan Pereira, F., Cabral, L., Gašević, D., Ramalho, G., & Ferreira Mello, R. (2024). Assessing the quality of automatic-generated short answers using GPT-4. *Computers and Education: Artificial Intelligence*, *7*. https://doi.org/10.1016/j.caeai.2024.100248

Rouzegar, H., & Makrehchi, M. (2024). Generative AI for Enhancing Active Learning in Education: A Comparative Study of GPT-3.5 and GPT-4 in Crafting Customized Test Questions. In R. Goebel & F. Zarrinkalam (Chairs), *Proceedings of the 37th Canadian Conference on Artificial Intelligence*. https://arxiv.org/abs/2406.13903v1

Sadidi, F., & Prestel, T. (2024). *Impact of Criterion-Based Reflection on Prospective Physics Teachers' Perceptions of ChatGPT-Generated Content.* https://arxiv.org/abs/2410.01354v1

Salcedo-Lagos, P., Pinacho-Davidson, P., Pinninghoff J., M. A., Contreras A., R., Fuentes-Riffo, K., & Carrillo, M. F. (2024). Optimizing Didactic Sequences with Artificial Intelligence: Integrating Bloom's Taxonomy and Emotion in the Selection of Educational Technologies. In J. M. Ferrández Vicente, M. Val Calvo, & H. Adeli (Eds.), *Lecture Notes in Computer Science. Artificial Intelligence for Neuroscience and Emotional Systems* (Vol. 14674, pp. 509–517). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-61140-7_48

Scarlatos, A., Feng, W., Smith, D., Woodhead, S., & Lan, A. (2024). Improving Automated Distractor Generation for Math Multiple-choice Questions with Overgenerate-and-rank. In E. Kochmar, M. Bexte, J. Burstein, A. Horbach, R. Laarmann-Quante, A. Tack, . . . Z. Yuan (Chairs), *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*. https://aclanthology.org/2024.bea-1.19/

Scarlatos, A., Smith, D., Woodhead, S., & Lan, A. (2024). Improving the Validity of Automatically Generated Feedback via Reinforcement Learning. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Lecture Notes in Computer Science. Artificial Intelligence in Education* (Vol. 14829, pp. 280–294). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64302-6_20

Schorcht, S., Baumanns, L., Buchholtz, N., Huget, J., Peters, F., & Pohl, M. (2024). The tasty world of irrational numbers–explorative mathematical visualizations generated by AI image models. *Journal of Mathematics and the Arts*, *18*(3-4), 272–289. https://doi.org/10.1080/17513472.2024.2407273

Schorcht, S., Buchholtz, N., & Baumanns, L. (2024). Prompt the problem – investigating the mathematics educational quality of AI-supported problem solving by comparing prompt techniques. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1386075

Selahattin, A., & Eyup, Y. (2024). Flipped Learning: An Innovative Model for Enhancing Education through ChatGPT. *International Journal of Modern Education Studies*, *8*(1), 124–148. https://doi.org/10.51383/ijonmes.2024.328

Shafayat, S., Hasan, H. M. Q., Mahim, M. R. C., Putri, R. A., Thorne, J., & Oh, A. (2024). *BEnQA: A Question Answering and Reasoning Benchmark for Bengali and English*. https://arxiv.org/abs/2403.10900v1

Shankar, S. K., Pothancheri, G., Sasi, D., & Mishra, S. (2024). Bringing Teachers in the Loop: Exploring Perspectives on Integrating Generative AI in Technology-Enhanced Learning. *International Journal of Artificial Intelligence in Education*, *35*, 155–180. https://doi.org/10.1007/s40593-024-00428-8

Singh, A., Ehtesham, A., Kumar, S., Gupta, G. K., & Khoei, T. T. (2024). *Encouraging Responsible Use of Generative AI in Education: A Reward-Based Learning Approach*. https://arxiv.org/abs/2407.15022v1

Smart, F., Bos, N. D., & Bos, J. T. (2024). Can Large Language Models Recognize and Respond to Student Misconceptions? In R. A. Sottilare & J. Schwarz (Eds.), *Lecture Notes in Computer Science. Adaptive Instructional Systems* (Vol. 14727, pp. 288–299). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60609-0_21

Song, Y., Kim, J [Jinhee], Liu, Z [Zifeng], Li, C., & Xing, W. (2025). Students' perceived roles, opportunities, and challenges of a generative AI-powered teachable agent: a case of middle school math class. *Journal of Research on Technology in Education*, 1–19. https://doi.org/10.1080/15391523.2024.2447727

Sonkar, S., Chen, X., Le, M., Liu, N [Naiming], Basu Mallick, D., & Baraniuk, R [Richard] (2024). Code Soliloquies for Accurate Calculations in Large Language Models. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 828–835). ACM. https://doi.org/10.1145/3636555.3636889

Spreitzer, C., Straser, O., Zehetmeier, S., & Maaß, K. (2024). Mathematical Modelling Abilities of Artificial Intelligence Tools: The Case of ChatGPT. *Education Sciences*, *14*(7). https://doi.org/10.3390/educsci14070698

Taani, O., & Alabidi, S. (2024). ChatGPT in education: benefits and challenges of ChatGPT for mathematics and science teaching practices. *International Journal of Mathematical Education in Science & Technology*. Advance online publication. https://doi.org/10.1080/0020739X.2024.2357341

Tabuenca, B., Martín, S., Greller, W., Tillmann, A., Uche-Soria, M., Castro, M., Tovar, E., & Rodríguez-Artacho, M. (2024). IoT and Generative AI Technologies to Support Urban Environmental Learning. In *2024 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1–4). IEEE. https://doi.org/10.1109/EDUCON60312.2024.10578567

Tang, K.-S. (2024). Informing Research on Generative Artificial Intelligence from a Language and Literacy Perspective: A Meta-Synthesis of Studies in Science Education. *SCIENCE EDUCATION*, *108*(5), 1329–1355. https://doi.org/10.1002/sce.21875

Tang, K.-S., & Cooper, G. (2024). The Role of Materiality in an Era of Generative Artificial Intelligence. *SCIENCE & EDUCATION*, *34*, 731–746. https://doi.org/10.1007/s11191-024-00508-0

Tassoti, S. (2024). Assessment of Students Use of Generative Artificial Intelligence: Prompting Strategies and Prompt Engineering in Chemistry Education. *JOURNAL of CHEMICAL EDUCATION*, *101*(6), 2475–2482. https://doi.org/10.1021/acs.jchemed.4c00212

Teegavarapu, R. R., & Sanghvi, H. (2023). Analyzing the Competitive Mathematical Problem-Solving Skills of ChatGPT. In *2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)* (pp. 1–7). IEEE. https://doi.org/10.1109/ICDSAAI59313.2023.10452659

Tong, D., Tao, Y., Zhang, K., Dong, X., Hu, Y., Pan, S., & Liu, Q. (2023). Investigating ChatGPT-4's performance in solving physics problems and its potential implications for education. *ASIA PACIFIC EDUCATION REVIEW*, *25*, 1379–1389. https://doi.org/10.1007/s12564-023-09913-6

Trckova, K., Marsalek, R., Zacek, M., & Tepla, M. (2024). Why Do Fishes Die? Lab Assignment Created Using ChatGPT. *JOURNAL of CHEMICAL EDUCATION*, *101*(8), 3171–3178. https://doi.org/10.1021/acs.jchemed.4c00130

van Pham, P. L., Duc, A. V., Hoang, N. M., Do, X. L., & Luu, A. T. (2024). ChatGPT as a Math Questioner? Evaluating ChatGPT on Generating Pre-university Math Questions. In J. Hong, J. W. Park, & A. Przybyłek (Eds.), *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing* (pp. 65–73). ACM. https://doi.org/10.1145/3605098.3636030

Vankúš, P. (2024). Generative Artificial Intelligence on Mobile Devices in the University Preparation of Future Teachers of Mathematics. *International Journal of Interactive Mobile Technologies*, *18*(18), 19–33. https://doi.org/10.3991/ijim.v18i18.51221

Vartiainen, H., Kahila, J., Tedre, M., Lopez-Pernas, S., & Pope, N. (2024). Enhancing children's understanding of algorithmic biases in and with text-to-image generative AI. *NEW MEDIA & SOCIETY*. Advance online publication. https://doi.org/10.1177/14614448241252820

Vasconcelos, M. A. R., & dos Santos, R. P. (2023). Enhancing STEM learning with ChatGPT and Bing Chat as objects to think with: A case study. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7). https://doi.org/10.29333/ejmste/13313

Verma, G., Campbell, T., Melville, W., & Park, B.-Y. (2023). Navigating Opportunities and Challenges of Artificial Intelligence: ChatGPT and Generative Models in Science Teacher Education. *Journal of Science Teacher Education*, *34*(8), 793–798. https://doi.org/10.1080/1046560X.2023.2263251

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7). https://doi.org/10.29333/ejmste/13272

Wei, X [Xin] (2024). Evaluating chatGPT-4 and chatGPT-4o: performance insights from NAEP mathematics problem solving. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1452570

Wijaya, T. T., Su, M., Cao, Y., Weinhandl, R., & Houghton, T. (2024). Examining Chinese preservice mathematics teachers' adoption of AI chatbots for learning: Unpacking perspectives through the UTAUT2 model. *EDUCATION and INFORMATION TECHNOLOGIES*, *30*, 1387–1415. https://doi.org/10.1007/s10639-024-12837-2

Wood, Z., & Greenberg, M. (2024). Board 43: AP-CS, ChatGPT and Me: a High School Student Perspective. In *2024 ASEE Annual Conference & Exposition Proceedings*. ASEE Conferences. https://doi.org/10.18260/1-2--47020

Wu, X., Saraf, P. P., Lee, G.-G., Latif, E., Liu, N [Ninghao], & Zhai, X. (2024). *Unveiling Scoring Processes: Dissecting the Differences between LLMs and Human Graders in Automatic Scoring*. https://arxiv.org/abs/2407.18328v1

Wulff, P. (2024). Physics language and language use in physics-What do we know and how AI might enhance language-related research and instruction. *EUROPEAN JOURNAL of PHYSICS*, *45*(2). https://doi.org/10.1088/1361-6404/ad0f9c

Xie, B., Sarin, P., Wolf, J., Garcia, R. C., Delaney, V., Sieh, I., Fuloria, A., Dennison, D. V., Bywater, C., Lee, V. R., & Wooldridge M. Dy J. Natarajan S. (2024). Co-designing AI Education Curriculum with Cross-Disciplinary High School Teachers. In M. Wooldridge, J. Dy, & S. Natarajan (Chairs), *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, *38*(21), 23146-23154. https://doi.org/10.1609/aaai.v38i21.30360

Yan, Y., Wang, S [Shen], Huo, J., Li, H [Hang], Li, B., Su, J., Gao, X., Zhang, Y.-F., Xu, T., Chu, Z., Zhong, A., Wang, K., Xiong, H., Yu, P. S., Hu, X., & Wen, Q. (2024). *ErrorRadar: Benchmarking Complex Mathematical Reasoning of Multimodal Large Language Models Via Error Detection*. https://arxiv.org/abs/2410.04509v2

Yang, K., Chu, Y., Darwin, T., Han, A., Li, H [Hang], Wen, H., Copur-Gencturk, Y., Tang, J., & Liu, H. (2024). Content Knowledge Identification with Multi-agent Large Language Models (LLMs). In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Lecture Notes in Computer Science. Artificial Intelligence in Education* (Vol. 14830, pp. 284–292). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64299-9_23

Yeadon, W., & Hardy, T. (2024). The impact of AI in physics education: a comprehensive review from GCSE to university levels. *Physics Education*, *59*(2). https://doi.org/10.1088/1361-6552/ad1fa2

Yue, M., Mifdal, W., Zhang, Y [Yixuan], Suh, J., & Yao, Z. (2024). *MathVC: An LLM-Simulated Multi-Character Virtual Classroom for Mathematics Education*. https://arxiv.org/abs/2404.06711v1

Yunianto, W., Lavicza, Z., Kastner-Hauler, O., & Houghton, T. (2024). Investigating the use of ChatGPT to solve a GeoGebra based mathematics+computational thinking task in a geometry topic. *Journal of Mathematics Education*, *15*(3), 1027–1052. https://doi.org/10.22342/jme.v15i3.pp1027-1052

Zhai, X., Nyaaba, M., & Ma, W. (2024). Can Generative AI and ChatGPT Outperform Humans on Cognitive-Demanding Problem-Solving Tasks in Science? *SCIENCE & EDUCATION*, *34*, 649–670. https://doi.org/10.1007/s11191-024-00496-1

Zhang, F., Li, C., Henkel, O., Xing, W., Baral, S., Heffernan, N., & Li, H [Hai] (2024). Math-LLMs: AI Cyberinfrastructure with Pre-trained Transformers for Math Education. *International Journal of Artificial Intelligence in Education.* Advance online publication. https://doi.org/10.1007/s40593-024-00416-y