ARCHED: A Human-Centered Framework for Transparent, Responsible, and Collaborative AI-Assisted Instructional Design

Hongming Li[1]   HLI3@UFL.EDU

Yizirui Fang[2]   YFANG52@JHU.EDU

Shan Zhang[1]   ZHANGSHAN@UFL.EDU

Seiyon M. Lee[1]   LEESEIYON@UFL.EDU

Yiming Wang[3]   WANG.YIMING8@NORTHEASTERN.EDU

Mark Trexler[2]   MTREXLER@JHU.EDU

Anthony F. Botelho[1]   ABOTELHO@COE.UFL.EDU

[1] *University of Florida, Gainesville, FL, USA*

[2] *Johns Hopkins University, Baltimore, MD, USA*

[3] *Northeastern University, Seattle, WA, USA*

Abstract

Integrating Large Language Models (LLMs) in educational technology reveals unprecedented opportunities to improve instructional design (ID), yet current approaches often prioritize automation over pedagogical rigor and human agency. This paper introduces ARCHED (AI for Responsible, Collaborative, Human-centered Education Instructional Design), a framework that implements a structured multi-stage workflow between educators and AI. Unlike existing tools that generate complete instructional materials autonomously, ARCHED cascades the development into distinct stages, from learning objective formulation to assessment design, each guided by Bloom's taxonomy and enhanced by LLMs. This framework employs multiple specialized AI components that work in concert: one generating diverse pedagogical options, another evaluating their alignment with learning objectives while maintaining human educators as primary decision-makers. ARCHED addresses critical gaps in current AI-assisted instructional design regarding transparency, pedagogical foundation, and meaningful human agency through this approach. This research advances the responsible integration of AI in education by providing a concrete, theoretically grounded framework that prioritizes human expertise and educational accountability.

Keywords: responsible AI, artificial intelligence, educational technology, instructional design, learning objectives, large language models, human-AI collaboration, Bloom's taxonomy

1. Introduction

The rise of Large Language Models (LLMs) has significantly changed educational technology, especially in instructional design (Choi et al., 2024). As Hodges and Kirschner (2024) observe, this transformation presents both unprecedented opportunities and significant challenges for educators. While LLMs offer powerful capabilities for content generation and assessment creation, their current implementation in educational settings often lacks pedagogical foundations and meaningful human oversight.

Recent studies have revealed concerning trends in AI-assisted instructional design. Parsons and Curry (2024) demonstrate that current AI tools can generate seemingly competent

instructional materials, yet often fail to incorporate crucial pedagogical considerations. This automation-first approach risks diminishing the role of human expertise in educational design. Furthermore, Sridhar et al. (2023) highlight that many existing systems operate as black boxes, making it difficult for educators to understand and validate the pedagogical reasoning behind generated content.

These challenges manifest in three critical areas. (1) The opacity of AI decision-making processes compromises educators' ability to ensure alignment with established educational frameworks. (2) Current emphasis on full automation marginalizes human expertise, potentially undermining the quality and effectiveness of instructional design. (3) Standardization of AI-generated assessments often results in a narrow range of evaluation methods, limiting opportunities for innovative and diverse learning experiences (Cheng et al., 2024).

This paper addresses these challenges through ARCHED, which fundamentally reimagines human-AI collaboration in instructional design by leveraging the capabilities of modern LLMs to enhance, rather than replace, human expertise. Specifically, we investigate the following research questions:

RQ1: How can AI-assisted instructional design systems maintain transparency while supporting efficient content development?

RQ2: What mechanisms enable meaningful integration of human expertise in AI-supported learning objective generation?

RQ3: How can AI tools promote diversity in assessment design while maintaining alignment with learning objectives?

Our work makes several contributions to the field. First, we introduce a transparent workflow that explicitly connects AI-generated content with established pedagogical frameworks through a structured multi-stage process. Second, we demonstrate how specialized AI components working in concert can enhance rather than diminish educator expertise, as validated through our empirical evaluation. Finally, we provide concrete evidence for how structured human-AI collaboration can diversify assessment strategies while maintaining pedagogical rigor, addressing a critical gap in current educational technology research.

2. Related Work

The integration of Large Language Models (LLMs) in educational technology shows a transformative shift in instructional design practices. Our analysis of existing research reveals both significant advances and critical limitations in four interconnected domains.

Evolution of AI-Assisted Instructional Design. The development of AI-assisted instructional design evolves from basic automation to increasingly sophisticated pedagogical attempts. Choi et al. (2024)'s SWOT analysis discloses a fundamental tension: while AI tools offer unprecedented efficiency in content creation, they require significant domain expertise for quality outcomes. Recent frameworks such as GAIDE (Dickey and Bejarano, 2024) and the design-build-test-learn approach (Chan et al., 2024) demonstrate improved efficiency, but the challenges in maintaining pedagogical rigor persist. Hu et al. (2024)'s evaluation of GPT-4 demonstrates that, while systems can produce structurally sound teaching plans, they struggle with context-specific pedagogical requirements. This limitation extends to graduate-level materials, where Parsons and Curry (2024) found that AI-generated content often lacks deep pedagogical understanding despite surface-level competence. DaCosta

and Kinsell (2024a) and DaCosta and Kinsell (2024b) further reveal that standardization in media selection and assessment methods may compromise educational effectiveness.

Learning Objectives and Assessment Generation. Sridhar et al. (2023) struggled with maintaining consistent taxonomic levels and authentic course alignment. Assessment generation faces similar challenges, with Cheng et al. (2024)'s TreeQuestion system revealing how AI-generated assessments tend toward conventional formats, limiting evaluation diversity. This observation aligns with Doughty et al. (2024)'s findings of significant quality variations between AI-generated and human-crafted programming assessments. Rouzegar and Makrehchi's multi-role AI assessment generation and Lin et al. (2024)'s explanatory feedback highlighting demonstrate the performance of AI while revealing limitations in handling nuanced, domain-specific responses.

Human-AI Collaboration Models in Education. Almasre (2024)'s GPT models show a consistent superiority of integrating human expertise over fully automated approaches, supporting Lin et al. (2022)'s necessitating pedagogical agency in utilizing AI capabilities. Innovative approaches include Krushinskaia et al. (2024)'s co-instructional designer bot and Li et al. (2024)'s ChatGPT in middle school instruction. While Moussa (2024) and Madunić and Sovulj (2024) structured approaches to GPT integration, Lan and Chen (2024) argue current frameworks often fail to fully address the complex interplay between human expertise and AI capabilities.

Transparency and Accountability Challenges. Park et al. (2024) systematically reveals widespread challenges in understanding and validating AI decision-making, while do Amaral (2024) highlights the need for transparent processes to maintain academic integrity while fostering innovation. Hatmanto et al. (2024) demonstrates how lack of transparency could disengage educator engagement and compromise pedagogy. Yadav (2023)'s work on scaling evidence-based instructional design expertise through LLMs emphasizes the importance of maintaining clear connections to educational research and best practices while listing the ongoing challenge of accountability in AI-generated content.

3. The ARCHED Framework: A Human-Centered Approach to AI-Assisted Instructional Design

This section presents ARCHED (AI for Responsible, Collaborative, Human-centered Education Instructional Design), a framework that fundamentally reimagines the integration of AI in instructional design. Rather than pursuing full automation, ARCHED establishes a structured collaborative workflow that maintains human agency while leveraging AI capabilities to enhance the development of educational content.

3.1. Theoretical Foundation and Design Principles

The development of ARCHED is grounded in the recognition that effective instructional design requires both pedagogical expertise and creative insight—qualities that neither AI systems nor humans alone can fully provide. Our framework builds upon Bloom's taxonomy as its pedagogical foundation, incorporating this established framework into a novel human-AI collaborative workflow. This approach addresses a fundamental limitation in current AI-assisted instructional design systems: their tendency to marginalize human expertise in favor of automation.

Three core principles guide our framework's design: First, human agency must be maintained throughout the instructional design process. Rather than generating complete instructional materials autonomously, AI systems should serve as collaborative tools that enhance human decision-making. Second, the framework must provide transparent reasoning for all AI-generated suggestions, enabling educators to make informed decisions about content adoption or modification. Third, the system should promote diversity in both learning objectives and assessment strategies, moving beyond the standardized approaches often favored by current AI systems.

3.2. Framework Architecture and Implementation

ARCHED implements a multi-stage collaborative workflow that integrates two specialized AI components: a Learning Objective Generation System (LOGS) and an Objective Analysis Engine (OAE). As shown in Figure 1, these components work in concert with human educators through three distinct phases while maintaining pedagogical rigor.
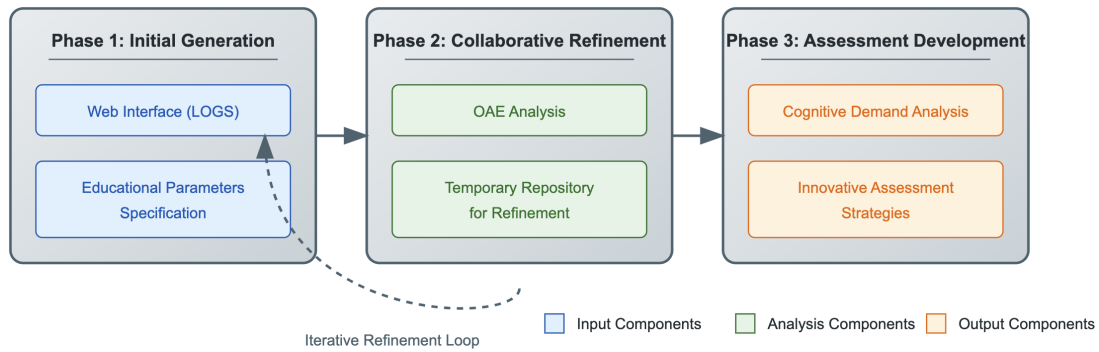

Figure 1: ARCHED Implementation Methodology: This shows the three-phase workflow: (a) Initial generation through LOGS with educator input, specifying parameters such as grade level, subject area, and desired Bloom's taxonomy levels (b) Collaborative refinement with OAE analysis, and (c) Assessment development based on finalized objectives. Human educators maintain control throughout the process.

LOGS serves as the primary interface for the development of learning objectives. Unlike existing systems that attempt to generate complete sets of learning objectives autonomously, LOGS implements a staged approach that begins with educator input. Educators specify fundamental parameters including grade level, subject area, and desired Bloom's taxonomy level. This initial human input ensures that all generated content aligns with the specific educational context and learning goals.

OAE provides analytical support by evaluating generated objectives against established pedagogical criteria. This separation of generation and evaluation capabilities represents a key innovation in our framework. Using distinct AI systems for these tasks, we maintain clearer accountability and enable more robust analysis. OAE generates detailed reports that highlight potential improvements while maintaining explicit connections to pedagogical principles, enabling educators to make informed decisions about objective refinement.

3.3. Technical Implementation

We have developed an open-access web-based tool[1] that implements ARCHED through multiple specialized AI agents. The system employs careful prompt engineering and chain-of-thought reasoning, with LOGS using structured prompts incorporating Bloom's taxonomy for objective generation and OAE utilizing step-by-step reasoning chains for evaluation.
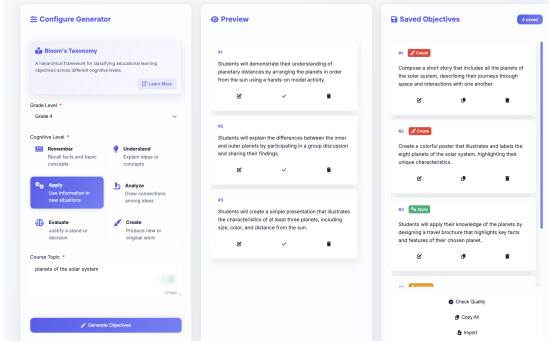

Figure 2: Main interface showing the three-phase workflow integration with continuous educator input in the loop
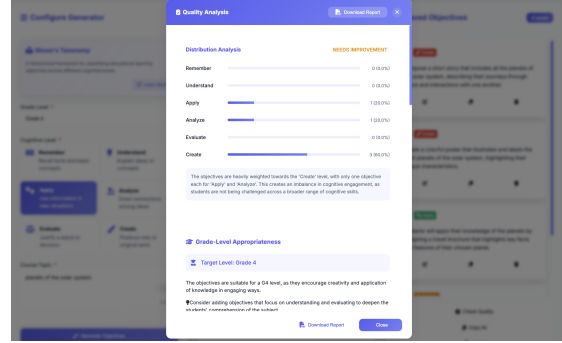

Figure 3: OAE analysis interface providing detailed pedagogical feedback and guidance

Based on OpenAI's GPT-4O series APIs, the modular architecture of the system supports integration with various language models, including local deployments. Through an intuitive interface (Figure 2), educators define course parameters and learning goals while receiving detailed feedback from OAE (Figure 3). We initially tested both `GPT-4o` and `GPT-4o mini` and found their generation quality to be largely comparable. To enhance accessibility and ensure feasibility in resource-limited educational contexts, we primarily use `gpt-4o-mini-2024-07-18`, which offers a cost-effective alternative. Additionally, we explore open-weight LLMs as potential solutions, though barriers such as hardware acquisition, server hosting, and API costs remain challenges for widespread adoption. We continue to evaluate the most sustainable deployment strategies that balance affordability, accessibility, and pedagogical effectiveness.

4. Results

We evaluated ARCHED through (1) a technical assessment of its learning objective analysis and generation capabilities and (2) preliminary expert feedback on its practical utility.

4.1. Technical Evaluation

To assess the framework's analytical capabilities, we tested it against 120 existing learning objectives from computer science education, with each objective independently classified by an educational expert according to Bloom's taxonomy. We employed weighted Cohen's Kappa (κ_w) for evaluation, using an ordinal weighting scheme to account for the hierarchical nature of taxonomic levels. The weighting scheme assigned 1.0 for perfect agreement, 0.8 for one-level difference, 0.6 for two-level difference, 0.4 for three-level difference, 0.2 for

---

1. A publicly available preview version of the tool can be accessed at https://logen.viablelab.org/.

four-level difference, and 0.0 for five-level difference. For example, confusing adjacent levels (e.g., "Create" with "Evaluate") was penalized less than distant levels (e.g., "Create" with "Remember").

The framework achieved $\kappa_w = 0.834$ (95% CI: [0.771, 0.891]), indicating strong agreement with expert classification. The confusion matrix analysis (Figure 4) reveals several key patterns: strongest performance at the extremes of the taxonomy (Remember and Create levels), high accuracy in distinguishing higher-order thinking skills (Analysis and Evaluation), and expected confusion patterns between adjacent cognitive levels. This confusion primarily occurs in the middle range of the taxonomy, where cognitive processes naturally overlap, particularly between Understanding and its adjacent levels. These patterns align with known challenges in cognitive-level discrimination, even among human experts.

Figure 4: Confusion matrix showing agreement between AI and expert classifications across Bloom's taxonomy levels
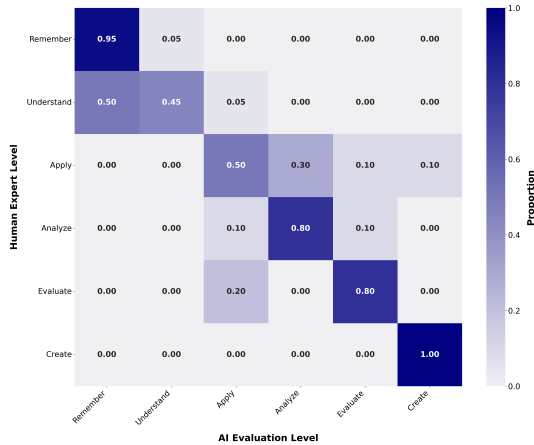
Criterion	ARCHED	Human
Structural	4.1±0.4	4.2±0.4
Taxonomic	4.0±0.5	4.1±0.4
Measurable	3.9±0.5	4.0±0.5
Clarity	4.0±0.4	4.1±0.4
Technical	3.8±0.6	4.0±0.5

Table 1: Quality scores (mean±SD), with a maximum score of 5

For generation quality assessment, we conducted a blind comparative analysis of 30 ARCHED-generated and 30 human-created learning objectives in computer science, mathematics, and physics. An expert evaluator assessed all objectives using five criteria on a scale of 1-5: structural completeness (ABCD and SMART model) (Heinich, 1996; Doran, 1981), taxonomic alignment, measurability, content clarity, and technical accuracy. Mann-Whitney U tests showed no significant differences between ARCHED and human-created objectives across all criteria ($p > 0.05$, Bonferroni-corrected), demonstrating the framework's ability to generate pedagogically sound objectives comparable to those created by experienced educators.

4.2. Pilot User Feedback

Two educational experts provided preliminary feedback on ARCHED's implementation, offering insights into how the framework's core principles manifest in practical instructional design workflows.

The experts' observations validated the framework's fundamental design principle of

maintaining human agency through iterative refinement. The three-panel interface design, which implements ARCHED's staged workflow, was noted for effectively supporting this principle. As reflected in one observation, *"The ability to adjust parameters and regenerate objectives based on specific needs is particularly useful. The interface makes it easy to fine-tune grade levels and cognitive levels until the objectives match course requirements."* This feedback demonstrates how the framework's multi-stage workflow successfully enables meaningful human-AI collaboration.

The selective curation feature, a key implementation of ARCHED's human-centered design principle, was highlighted as essential for maintaining educator autonomy. *"Having the freedom to select which objectives align with student needs and course goals, rather than accepting all generated content, helps maintain instructional control while leveraging AI assistance."* noted an expert. This aligns with the framework's core goal of positioning AI as a supportive tool rather than an autonomous solution.

The quality analysis functionality, which implements ARCHED's emphasis on pedagogical rigor and transparency, received positive attention. The integration of Bloom's taxonomy into the analysis process, as specified in the framework design, was noted as particularly valuable. As an expert observed, *"The visualization of the distribution analysis and the downloadable reports make it easy to identify gaps in cognitive coverage and document the development process."* This feedback validates the framework approach to maintain the pedagogical alignment through explicit evaluation criteria.

The import feature for existing learning objectives demonstrated the flexibility of the framework in supporting various instructional design approaches. An observation highlighted that *"The ability to upload and analyze existing objectives streamlines the review process, though integration with learning management systems would make it even more valuable."* This capability exemplifies ARCHED's design principle of enhancing rather than replacing existing educational practices.

This initial feedback suggests that ARCHED's implementation successfully realizes its core design principles of human agency, pedagogical rigor, and transparent AI assistance. The practical application of the framework demonstrates its potential to improve instructional design processes while maintaining educator control and pedagogical quality.

5. Discussion

The development and implementation of ARCHED represents a significant advancement in AI-assisted instructional design by introducing a structured, human-centered framework that maintains pedagogical rigor while leveraging AI capabilities. Our evaluation results demonstrate how the framework's core design principles and components effectively address the key challenges identified in this work.

Regarding RQ1 on maintaining transparency in AI-assisted instructional design, ARCHED's multi-agent architecture fundamentally transforms the typical "black box" approach. By separating objective generation (LOGS) and evaluation (OAE) into distinct components, each with explicit reasoning processes, the framework provides educators with clear insight into the AI's decision-making process. The high rates of agreement ($\kappa_w = 0.834$) between OAE analysis and expert evaluations validate this approach, demonstrating that transparent AI systems can maintain high performance while supporting informed decision-making.

This aligns with Park et al. (2024)'s vision for explicable AI systems in education while addressing Sridhar et al. (2023)'s concerns about opacity in AI-assisted instruction.

The collaborative workflow staged within the framework directly addresses RQ2 by reimagining how human expertise integrates with AI capabilities. Rather than attempting to automate the entire process, ARCHED implements specific interaction points where educator input guides AI generation and refinement. This design choice is validated by our comparative analysis, where ARCHED-generated objectives achieved quality scores comparable to human-created ones on all evaluation criteria. The framework's ability to maintain pedagogical sophistication while preserving educator agency demonstrates the effectiveness of our approach, supporting Lin et al. (2022)'s emphasis on meaningful human involvement in AI-assisted education.

For RQ3, concerning assessment diversity and objective alignment, our analysis focused primarily on the framework's ability to maintain pedagogical alignment in learning objective generation. The high-quality scores across structural completeness (4.1 ± 0.4), taxonomic alignment (4.0 ± 0.5), and measurability (3.9 ± 0.5) demonstrate ARCHED's capability to generate well-aligned learning objectives that support diverse assessment possibilities. Although our current evaluation concentrated on the generation and alignment of objectives, the framework lays the foundation for the future development of assessment strategies that maintain this demonstrated pedagogical rigor. This addresses part of Cheng et al. (2024)'s concerns about assessment alignment, though further work is needed to fully explore the framework's potential indirect assessment generation.

Although our evaluation demonstrates ARCHED's effectiveness, particularly in computer science education contexts, several limitations warrant consideration. Following Hu et al. (2024), the ability of the framework to generate higher-order thinking objectives requires further validation across diverse disciplines. Additionally, as noted by Krushinskaia et al. (2024), system adoption may face challenges in resource-constrained settings due to familiarization requirements.

Future research will focus on increasing the capacity of ARCHED in several domains and improving interaction with current learning management systems (Moussa, 2024). These developments aim to create a more accessible and comprehensive tool for AI-assisted instructional design while maintaining our commitment to transparency, human agency, and pedagogical quality.

Acknowledgments

References

Miada Almasre. Development and Evaluation of a Custom GPT for the Assessment of Students' Designs in a Typography Course. *Education Sciences*, 14(2):148, February 2024. ISSN 2227-7102. doi: 10.3390/educsci14020148.

Ka Ching Chan, Sachithra Lokuge, Mahdi Fahmideh, and Michael Steven Lane. AI-Assisted

Educational Design: Academic-GPT Collaboration for Assessment Creation, October 2024.

Zirui Cheng, Jingfei Xu, and Haojian Jin. TreeQuestion: Assessing Conceptual Learning Outcomes with LLM-Generated Multiple-Choice Questions. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW2):431:1–431:29, November 2024. doi: 10.1145/3686970.

Gi Woong Choi, Soo Hyeon Kim, Daeyeoul Lee, and Jewoong Moon. Utilizing Generative AI for Instructional Design: Exploring Strengths, Weaknesses, Opportunities, and Threats. *TechTrends*, 68(4):832–844, July 2024. ISSN 1559-7075. doi: 10.1007/s11528-024-00967-w.

Boaventura DaCosta and Carolyn Kinsell. Enhancing Instructional Design with Generative AI: A Comparative Analysis of Media Selection Using ChatGPT. In *Society for Information Technology & Teacher Education International Conference*, pages 68–73. Association for the Advancement of Computing in Education (AACE), March 2024a. ISBN 978-1-939797-76-6.

Boaventura DaCosta and Carolyn Kinsell. Investigating Media Selection through ChatGPT: An Exploratory Study on Generative Artificial Intelligence in the Aid of Instructional Design. *Open Journal of Social Sciences*, 12(4):187–227, April 2024b. doi: 10.4236/jss.2024.124014.

Ethan Dickey and Andres Bejarano. GAIDE: A Framework for Using Generative AI to Assist in Course Content Development, September 2024.

Inês do Amaral. Reflection on the use of Generative Language Models as a Tool for Teaching Design. In *2024 IEEE World Engineering Education Conference (EDUNINE)*, pages 1–4, March 2024. doi: 10.1109/EDUNINE60625.2024.10500634.

George T Doran. There's a smart way to write managements's goals and objectives. *Management review*, 70(11), 1981.

Jacob Doughty, Zipiao Wan, Anishka Bompelli, Jubahed Qayum, Taozhi Wang, Juran Zhang, Yujia Zheng, Aidan Doyle, Pragnya Sridhar, Arav Agarwal, Christopher Bogart, Eric Keylor, Can Kultur, Jaromir Savelka, and Majd Sakr. A Comparative Study of AI-Generated (GPT-4) and Human-crafted MCQs in Programming Education. In *Proceedings of the 26th Australasian Computing Education Conference*, ACE '24, pages 114–123, New York, NY, USA, January 2024. Association for Computing Machinery. ISBN 9798400716195. doi: 10.1145/3636243.3636256.

Endro Dwi Hatmanto, Sittie Noffaisah B. Pasandalan, Fitria Rahmawati, and Maryam Sorohiti. Empowering Creative Education: Applying Chat GPT for Enhancing Student Engagement in Senior Teacher-Driven Instructional Design in the Philippines. *E3S Web of Conferences*, 570:03007, 2024. ISSN 2267-1242. doi: 10.1051/e3sconf/202457003007.

R Heinich. Instructional media and technologies for learning. *Nj: Merrill*, 1996.

Charles B. Hodges and Paul A. Kirschner. Innovation of Instructional Design and Assess-

ment in the Age of Generative Artificial Intelligence. *TechTrends*, 68(1):195–199, January 2024. ISSN 1559-7075. doi: 10.1007/s11528-023-00926-x.

Bihao Hu, Longwei Zheng, Jiayi Zhu, Lishan Ding, Yilei Wang, and Xiaoqing Gu. Teaching Plan Generation and Evaluation With GPT-4: Unleashing the Potential of LLM in Instructional Design. *IEEE Transactions on Learning Technologies*, 17:1471–1485, 2024. ISSN 1939-1382. doi: 10.1109/TLT.2024.3384765.

Kristina Krushinskaia, Jan Elen, and Annelies Raes. Design and Development of a Co-instructional Designer Bot Using GPT-4 to Support Teachers in Designing Instruction. 2024.

Yu-Ju Lan and Nian-Shing Chen. Teachers' agency in the era of LLM and generative AI: Designing pedagogical AI agents. *Educational Technology & Society*, 27(1):I–XVIII, 2024. ISSN 1176-3647.

Yingzhe Li, Jianling Liu, and Shichen Yang. Is ChatGPT a Good Middle School Teacher? An Exploration of its Role in Instructional Design. In *Proceedings of the 3rd International Conference on New Media Development and Modernized Education, NMDME 2023, October 13–15, 2023, Xi'an, China*, January 2024. ISBN 978-1-63190-445-5.

Jionghao Lin, Eason Chen, Zeifei Han, Ashish Gurung, Danielle R. Thomas, Wei Tan, Ngoc Dang Nguyen, and Kenneth R. Koedinger. How Can I Improve? Using GPT to Highlight the Desired and Undesired Parts of Open-ended Responses, May 2024.

Xiao-Fan Lin, Lu Chen, Kan Kan Chan, Shiqing Peng, Xifan Chen, Siqi Xie, Jiachun Liu, and Qintai Hu. Teachers' Perceptions of Teaching Sustainable Artificial Intelligence: A Design Frame Perspective. *Sustainability*, 14(13):7811, January 2022. ISSN 2071-1050. doi: 10.3390/su14137811.

Jelena Madunić and Matija Sovulj. Application of ChatGPT in Information Literacy Instructional Design. *Publications*, 12(2):11, June 2024. ISSN 2304-6775. doi: 10.3390/publications12020011.

Moustafa Kamal Moussa. Towards Reliable Utilization: An Instructional Design Model for Integrating Generative Pre-trained Transformer (GPT) in Education. In Amina Al-Marzouqi, Said A. Salloum, Mohammed Al-Saidat, Ahmed Aburayya, and Babeet Gupta, editors, *Artificial Intelligence in Education: The Power and Dangers of ChatGPT in the Classroom*, pages 481–496. Springer Nature Switzerland, Cham, 2024. ISBN 978-3-031-52280-2. doi: 10.1007/978-3-031-52280-2_30.

Sungkyung Park, Sumin Hong, Eunseon Lim, and Cheolil Lim. Analysis of the state of instructional research using AI and exploration key lesson design stages through a systematic literature review. *Korean Association for Educational Information and Media*, 30:1135–1158, August 2024. doi: 10.15833/KAFEIAM.30.4.1135.

Bruce Parsons and John H. Curry. Can ChatGPT Pass Graduate-Level Instructional Design assignments? Potential Implications of Artificial Intelligence in Education and a Call to Action. *TechTrends*, 68(1):67–78, January 2024. ISSN 1559-7075. doi: 10.1007/

s11528-023-00912-3.

Hamidreza Rouzegar and Masoud Makrehchi. Generative AI for Enhancing Active Learning in Education: A Comparative Study of GPT-3.5 and GPT-4 in Crafting Customized Test Questions.

Pragnya Sridhar, Aidan Doyle, Arav Agarwal, Christopher Bogart, Jaromir Savelka, and Majd Sakr. Harnessing LLMs in Curricular Design: Using GPT-4 to Support Authoring of Learning Objectives, June 2023.

Gautam Yadav. Scaling Evidence-based Instructional Design Expertise through Large Language Models, June 2023.