Assessing AI tools for academic workflows: A standardized prompt evaluation

Scott A. Kerth. Department of Organizational Leadership, Lewis University, Romeoville (IL), USA. kerthsc@lewisu.edu
Sheila Boysen. Department of Organizational Leadership, Lewis University, Romeoville (IL), USA. boysensh@lewisu.edu
Maja Grgurovic. Faculty Center for the Advancement of Teaching and Learning, Lewis University, Romeoville (IL), USA. mgrgurovic@lewisu.edu
Xavier Sales. Department of Finance and Management Conttrol, EADA Business School, Spain. xsales@eada.edu
(ORCID: 0000-0002-2248-3889)

Resum
L'article presenta un estudi exploratori en curs per examinar la possible integració d'eines d'intel·ligència artificial (IA) en els fluxos de treball del professorat i dels dissenyadors a les institucions d'educació superior. El projecte de recerca desenvolupa una metodologia sistemàtica per avaluar diverses plataformes d'IA en tres fluxos de treball acadèmics principals: el disseny i la planificació curricular, el disseny de tasques i avaluacions, i la participació i interacció amb els estudiants. Mitjançant un sistema d'avaluació basat en rúbriques amb mesures de fiabilitat entre avaluadors, analitza l'eficàcia comparativa de sis eines d'IA seleccionades segons la seva prominència al mercat, diferències arquitectòniques i models de llicència. Aquesta recerca té com a objectiu reduir la resistència institucional a l'adopció de la IA mitjançant la demostració de beneficis i limitacions concretes d'aquestes tecnologies en contextos acadèmics, alhora que treballa per desenvolupar polítiques universitàries integrals per a la incorporació de la IA tant en la tasca del professorat com en l'aprenentatge dels estudiants. Els resultats preliminars suggereixen que, tot i que les eines d'IA mostren potencial per donar suport a diverses tasques acadèmiques, la seva implementació requereix una consideració acurada tant dels seus punts forts com de les seves limitacions.

Paraules clau
intel·ligència artificial, educació superior, flux de treball del professorat, disseny d'instruccions, tecnologia educativa, IA generativa.

Abstract

The paper presents an ongoing exploratory study for examining the potential integration of artificial intelligence (AI) tools into faculty and instructional designer workflows at higher education institutions. The research project develops a systematic methodology for evaluating different AI platforms across three primary academic workflows: curriculum design and planning, assignment and assessment design, and student engagement and interaction. Using a rubric-based evaluation system with inter-rater reliability measures, it will analyze the comparative effectiveness of six AI tools selected based on market prominence, architectural differences, and licensing models. This research aims to reduce institutional resistance to AI adoption by demonstrating concrete benefits and limitations of these technologies in academic contexts, while simultaneously working toward developing comprehensive university policies for AI integration in both faculty work and student learning. Preliminary findings suggest that while AI tools show promise in supporting various academic tasks, their implementation requires careful consideration of both their strengths and limitations.

Keywords

artificial intelligence, higher education, faculty workflow, instructional design, educational technology, generative AI.

1. Introduction

The emergence of sophisticated generative AI tools, particularly with ChatGPT's public release in 2022, has sparked substantial discussion across higher education institutions worldwide. This technology's unprecedented adoption rate and impact on academic practices have created both opportunities and challenges for university faculty and administrators (Bobula, 2024). At our institution, initial observations revealed improvements in student submissions, suggesting early adoption of AI tools among the student population despite varied faculty perspectives on their use.

In response to this rapidly evolving technological landscape, our research team—comprising faculty members and instructional design specialists—secured institutional funding to investigate how AI technologies could be effectively integrated into faculty workflows. This research emerges against a backdrop of significant resistance from portions of the academic community who oppose student use of AI and question its place in education more broadly (Eke, 2023; Stokel-Walker, 2022).

The present study aims to address these tensions by systematically evaluating AI tools' effectiveness across key faculty work processes, thereby providing evidence-based insights to inform institutional policies and practices. Rather than focusing exclusively on student use cases, which have dominated much of the discourse, we deliberately center our investigation on faculty applications to demonstrate AI's potential value to the skeptical members of our academic community.

University professors are increasingly exploring and adopting AI tools to streamline their workflows, enhance pedagogical practices, and augment research capabilities. From automating routine tasks to generating novel insights, AI presents transformative potential for academic life. However, the diverse landscape of available AI tools, coupled with varying levels of faculty adoption, necessitates a systematic and objective evaluation of their efficacy within specific academic contexts. This research addresses this need by conducting a standardized, prompt-based evaluation of prominent AI tools, focusing on their performance within three pivotal categories of the academic

workflow: Curriculum Design and Planning, Assignment and Assessment Design, and Student Engagement and Interaction. These categories represent core responsibilities of university professors, and their optimization through AI integration holds the promise of significant improvements in educational processes.

This paper presents our methodology and preliminary findings from this ongoing investigation, contributing to the growing body of literature examining AI integration in higher education contexts.

2. Literature Review

2.1 AI in Higher Education: Current Research Landscape

Recent research highlights the growing impact of AI for student engagement and assessment in higher education. AI tools have been shown to enhance personalized learning, improve communication, and foster collaborative environments (Msambwa et al., 2025). AI applications in education include assessment/evaluation, prediction, AI assistance, intelligent tutoring systems, and learning management (Crompton & Burke, 2023). Specifically, AI-driven chatbots like ChatGPT have demonstrated potential for increasing student engagement and supporting inquiry-based learning, though concerns about empathy and accuracy persist (Abu Khurma et al., 2024). Research also focuses on AI's role in assessing student behaviors, sentiments, and achievements (Sánchez Prieto et al., 2020). Despite the benefits, ethical considerations such as privacy and bias remain important areas for ongoing investigation and improvement (Msambwa et al., 2025).

The introduction of ChatGPT in late 2022 marked a watershed moment in educational technology, with adoption rates exceeding those of previous technological innovations. Initial institutional responses ranged from outright bans to cautious exploration, reflecting uncertainty about implications for academic integrity, student learning outcomes, and pedagogical practices (Stokel-Walker, 2022; Bender et al., 2021).

2.2 AI in Assignment and Assessment Design

Literature from 2024 and 2025 offers insightful perspectives on the integration of AI tools within assignment and assessment design in higher education. Cranfield et al. (2024) advocate for a fundamental shift in assessment practices, leveraging conversational AI, such as ChatGPT, to prioritize the development of critical thinking, communication, and adaptability skills. This aligns with Gupta et al. (2024), who demonstrate that tools like Coursera Coach, AI-assisted course building, and Turnitin significantly enhance student engagement and facilitate personalized learning experiences, ultimately optimizing assessment processes.

Thambi (2024) further contributes to this discourse by proposing a redesign of assignment tasks that explicitly incorporates AI platforms, while ensuring assessment validity through rigorous structured interviews. Zhao (2025), in a comprehensive review of AI applications including ChatGPT, GenAI, Midjourney, and automated scoring tools, highlights the potential for these systems to provide high-quality, real-time, personalized feedback, thereby fostering improved cognitive and metacognitive skills. However, Zhao (2025) also underscores the critical need to address concerns related to algorithmic bias, privacy, and academic integrity, which remain paramount considerations in the responsible deployment of AI in educational assessment.

2.3 AI in Curriculum Design and Planning

Several studies explore the multifaceted role of artificial intelligence in curriculum design and planning within higher education. Crompton and Burke (2023) highlight AI's capacity to organize curriculum sequences, design instructional frameworks, and provide structural stability, thereby streamlining the development process. Building on this, Deng (2024) demonstrates that machine learning enhances curriculum design by refining needs assessments, improving evaluation methodologies, and facilitating the personalization of learning experiences to cater to individual student needs.

Complementarily, Pusporini and Nurdiyanto (2024) suggest that AI's ability to analyze educational data, identify learning gaps, predict student outcomes, and recommend personalized learning paths supports outcome-based evaluation through machine learning and natural language processing. Finally, Wen et al. (2024) employ clustering algorithms to delineate critical dimensions—including student attitudes, expectations, adaptability, usage, safety, and perceived desirable functions—that mediate the impact of AI on learning efficacy. Collectively, these papers indicate that AI tools contribute significantly to curriculum organization, personalized learning, data-driven evaluation, and alignment with industry trends. However, the strength of evidence across these areas varies, ranging from moderate to limited, necessitating further empirical validation and exploration.

2.4 Research Gap and Study Rationale

While existing literature extensively explores the broad applications of AI in education, the benefits of workflow analysis, and the ethical considerations surrounding AI implementation, a critical gap remains in the systematic, comparative evaluation of specific AI tools within the nuanced context of core academic workflows. While there are studies that examine the use of AI for specific tasks, such as automated grading or personalized learning, there is a lack of research that holistically assesses the performance of a range of AI tools across the diverse set of responsibilities that fall under the purview of university professors.

This study aims to fill this gap by providing a comprehensive evaluation of leading AI tools, including ChatGPT, Claude, Perplexity, Gemini, Llama, and Coplit, across three key areas of academic workflow: Curriculum Design and Planning, Assignment and Assessment Design, and Student Engagement. By employing a standardized prompt-based methodology and a consistent evaluation rubric, this research will move beyond general observations of AI potential to provide empirical, comparative data, thereby offering actionable insights for educators seeking to integrate AI effectively into their daily practices.

3. Research Methodology

3.1 Academic Workflow Categories

For this research, we have identified three pivotal categories within the academic workflow: Curriculum Design and Planning, Assignment and Assessment Design, and Student Engagement and Interaction. These categories were selected due to their foundational role in educational processes and their potential for substantial enhancement through systematic analysis.

Curriculum Design and Planning forms the cornerstone of educational activities, encompassing the development and organization of course content, learning objectives, and instructional strategies. This category holds critical importance as it shapes the entire educational experience and directly influences student outcomes.

Assignment and Assessment Design is essential for evaluating student progress and comprehension. It involves the creation of meaningful tasks and evaluation methodologies that align with specified learning objectives. This category was chosen because of its direct impact on student learning and the significant time investment required from educators.

Student Engagement and Interaction represents the dynamic core of the learning process. It encompasses various activities and strategies aimed at promoting active participation, collaboration, and substantive dialogue between students and instructors. This area is particularly salient in contemporary education, characterized by diverse learning modalities and technologies.

These three categories are interconnected and represent key areas where workflow analysis can yield significant improvements in educational processes. By focusing on these aspects, our research seeks to provide comprehensive insights into optimizing academic activities, enhancing both the teaching and learning experience, and improving overall institutional effectiveness.

3.2 Tool Selection Criteria

The selection of tools for analysis was guided by a methodological framework designed to ensure comprehensive coverage, accessibility, and educational relevance. This approach prioritizes a multifaceted perspective, incorporating a diverse range of software solutions. Six AI tools were selected: ChatGPT, Claude, Perplexity, Gemini, Llama, and Coplit.

1. **Comprehensive Coverage:** The selection encompasses a broad spectrum of tools, deliberately spanning both commercial and open-source solutions. This strategy ensures a balanced representation of the current technological landscape. The selection includes both general-purpose and specialized tools, catering to a wide array of potential applications within the targeted domain. This deliberate heterogeneity extends to the inclusion of tools developed using various development approaches and architectural paradigms, providing a holistic view of existing methodologies. This selection provides a broad view of the current AI landscape. It includes both commercial (ChatGPT, Claude, Gemini, Coplit) and open-source (Llama) models, representing diverse development approaches. General-purpose LLMs (ChatGPT, Claude, Perplexity, Gemini) are included alongside a specialized tool for code generation (Coplit), and a foundational model for fine-tuning (Llama). This variety ensures a thorough analysis of different functionalities and architectures.
2. **Accessibility Considerations:** Recognizing the diverse needs of potential users, the selection process incorporated accessibility as a critical criterion. Tools were chosen to represent a range of access models, including free, paid, and institutionally licensed software. This diversity addresses budgetary constraints and ensures wider applicability. In addition, the selection accounts for varying levels of technical requirements, acknowledging the diverse skill sets of potential users. Finally, the chosen tools reflect different levels of implementation complexity, allowing for analysis of both user-friendly and more advanced solutions. The tools offer diverse access models, including free and paid options (ChatGPT, Claude, Perplexity, Gemini), addressing varying budgetary constraints. Technical requirements range from user-friendly interfaces (ChatGPT, Perplexity) to more complex deployment (Llama), accommodating diverse skill sets.

3. **Educational Relevance:** A key objective of this analysis is to explore the potential of these tools within higher education settings. Therefore, the selection prioritized tools with a demonstrated application in academic contexts. The analysis examines tools offering various levels of pedagogical support, including features designed to facilitate teaching and learning. Moreover, the selection includes tools that employ different approaches to academic content generation, enabling a comparative analysis of their effectiveness in supporting diverse educational methodologies. These tools have demonstrated applicability in higher education. LLMs like ChatGPT, Claude, Gemini, and Perplexity are used for content creation and research. Llama is employed in academic research. Coplit offers specialized code-generation capabilities relevant to computer science education. The selection also reflects diverse pedagogical support and approaches to academic content generation, enabling a comparative analysis of their educational effectiveness.

3.3 Evaluation Rubric

The development of a comprehensive rubric for evaluating artificial intelligence tools in higher education necessitates careful consideration of assessment dimensions that align with pedagogical objectives while maintaining methodological rigor. The proposed tripartite framework—comprising Clarity of Output (33.3%), Detail and Depth (33.3%), and Professional Relevance (33.3%)—draws upon established assessment principles while addressing the unique challenges of evaluating AI-generated content for academic applications.

In our research, the evaluation of AI tool performance will be conducted using the three-dimension rubric, each assessed on a 1-5 Likert scale:

Clarity of Output:

5) Output is clear, precise, and actionable
3) Output is clear but occasionally requires clarification
1) Output is ambiguous, confusing, or inappropriate

Detail and Depth:

5) Provides thorough, thoughtful analysis with relevant examples and applications
3) Offers adequate detail at a basic level of analysis
1) Provides superficial or incomplete information that cannot be effectively used

Professional Relevance:

5) Matches or exceeds university-level expectations
3) Generally appropriate for academic use
1) Below academic standards

We posit that this structured evaluation framework will facilitate a systematic assessment of AI tool performance across various academic workflows. It will provide a consistent basis for comparing different AI tools and prompting strategies, allowing for a clear understanding of their respective strengths and weaknesses.

3.4 Prompt Development and Testing

For each identified workflow task, we are developing specific prompts designed to elicit responses that would demonstrate an AI tool's capabilities within that domain. The prompts will be created to apply across the three workflow categories, with each prompt carefully constructed to represent authentic faculty needs and use cases.

These identical prompts will be then systematically applied to each of the six selected AI tools, with the three researchers dividing the work of interaction and response collection. All AI outputs will be documented verbatim for subsequent evaluation.

4. Preliminary Findings

While this research remains in progress, several preliminary observations have emerged:

1. AI tools demonstrate variable effectiveness across different workflow categories, with certain platforms exhibiting strengths in specific domains while underperforming in others.
2. The quality of AI outputs appears contingent not only on the underlying technology but also on prompt design and specificity, suggesting faculty will require development in effective prompt engineering to maximize utility.
3. Certain academic tasks appear more amenable to AI assistance than others, with clear patterns emerging that could guide institutional recommendations for appropriate AI integration points.

These early observations will be refined through completion of the evaluation process and subsequent detailed analysis.

5. Discussion

Our preliminary findings suggest several potential avenues for promoting AI adoption among faculty. By demonstrating concrete benefits in specific workflow areas, resistance may be reduced, particularly if faculty can witness efficiency gains in time-intensive tasks. Simultaneously, identifying limitations helps establish realistic expectations that may mitigate disappointment and abandonment following initial implementation attempts.

This research carries significant implications for university-wide policy development regarding AI use. Rather than implementing blanket policies governing all AI applications, our findings suggest the need for nuanced approaches that differentiate between various use cases and applications. Faculty-facing policies may require different considerations than student-facing guidelines, with appropriate scaffolding for both constituencies.

While the current study focuses primarily on existing faculty workflows, our preliminary analysis suggests several promising directions for expanded AI integration:

1. Research support applications
2. Enhanced student feedback mechanisms
3. Development of personalized learning pathways
4. Cultivation of critical thinking skills through guided AI interaction
5. Student-directed inquiry facilitated by AI tools

Each of these directions warrants dedicated investigation to determine optimal implementation strategies and potential impacts on learning outcomes.

6. Conclusion

This ongoing study represents an important step toward evidence-based integration of AI technologies in higher education environments. By systematically evaluating AI tools across authentic faculty workflows, we provide a foundation for both practical implementation guidance and policy development at institutional levels.

This research has established a rigorous methodology for evaluating the performance of various AI tools within key academic workflows, employing standardized prompts and a comprehensive rubric. By focusing on Curriculum Design and Planning, Assignment and Assessment Design, and Student Engagement and Interaction, we have targeted areas critical to the modern university professor. The systematic testing of ChatGPT, Claude, Perplexity, Gemini, Llama, and Coplit, utilizing a rubric measuring Clarity of Output, Detail and Depth, and Professional Relevance, provides a quantitative and objective assessment of each tool's capabilities.

The findings from this study will yield valuable insights into the strengths and limitations of these AI tools in supporting academic tasks. The comparative analysis will offer educators a clear understanding of which tools are best suited

for specific workflows, enabling informed decisions regarding AI integration. By identifying the tools that excel in clarity, depth, and professional relevance, this research will contribute to optimizing academic processes and enhancing overall institutional effectiveness.

Furthermore, the standardized prompt-based evaluation framework developed in this study can serve as a model for future research in this rapidly evolving field. As AI technologies continue to advance, the need for systematic and objective evaluations becomes increasingly crucial. This methodology provides a foundation for ongoing assessments and comparisons, ensuring that educators and institutions can adapt to the changing landscape of AI in higher education.

The rapid evolution of AI capabilities necessitates continued research in this domain, with particular attention to how these technologies might transform not only existing workflows but potentially redefine aspects of academic work altogether. Future research should expand beyond faculty workflows to examine more deeply how AI-assisted teaching may impact student learning outcomes and the development of critical thinking skills in the emerging technological landscape.

Ultimately, this research aims to bridge the gap between AI potential and practical application in academia. By providing a clear and comparable measure of AI tool performance, we seek to empower professors to leverage these technologies effectively, thereby improving teaching, research, and student outcomes. The insights gained from this study will contribute to the ongoing discourse on the ethical and responsible integration of AI in higher education, fostering a more informed and strategic approach to its implementation.

References

Abu Khurma, O., Albahti, F., Ali, N., & Bustanji, A. (2024). AI ChatGPT and student engagement: Unraveling dimensions through PRISMA analysis for enhanced learning experiences. *Contemporary Educational Technology, 16*(2). https://doi.org/10.30935/cedtech/14334

Belwal, R. (2025). Generative AI in Academics: Transforming Workflows, Enhancing

Efficiency, and Redefining Ethics in the Workplace. *Management Dynamics, 25*(1), Article 2. https://doi.org/10.57198/2583-4932.1340

Bobula, M. (2024). Generative artificial intelligence (AI) in higher education: A comprehensive review of challenges, opportunities, and implications. *Journal of Learning Development in Higher Education, 30*. https://doi.org/10.47408/jldhe.vi30.1137

Conklin, S., Dorgan, T., & Barreto, D. (2024). Is AI the new course creator. *Discover Education, 3*, 285. https://doi.org/10.1007/s44217-024-00386-2

Cranfield, D., Venter, I. M., & Mulyata, J. (2024). Conversational Artificial Intelligence: A Catalyst for Rethinking Assessment in Higher Education. P*roceedings of the 25th European Conference on Knowledge Management, 25*(1). https://doi.org/10.34190/eckm.25.1.2935

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education, 20*, 22. https://doi.org/10.1186/s41239-023-00392-8

Deng, Y. (2024). A Systematic Review of Application of Machine Learning in Curriculum Design Among Higher Education. *Journal of Emerging Computer Technologies, 4*(1). https://doi.org/10.57020/ject.1475566

Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology, 13*, 100060. https://doi.org/10.1016/j.jrt.2023.100060

Gupta, S., Dharamshi, R. R., & Kakde, V. (2024). An Impactful and Revolutionized Educational Ecosystem using Generative AI to Assist and Assess the Teaching and Learning benefits, Fostering the Post-Pandemic Requirements. In *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)* (pp. 1-4). https://doi.org/10.1109/ic-ETITE58242.2024.10493370

Johnson, C. W., & Paulus, T. M. (2024). Generating a Reflexive AI-Assisted Workflow for Academic Writing. *Qualitative Report, 29*(10), 2772-2792. https://doi.org/10.46743/2160-3715/2024.7634

Msambwa, M. M., Wen, Z., & Daniel, K. (2025). The Impact of AI on the Personal and Collaborative Learning Environments in Higher Education. *European Journal of Education, 60*(1). https://doi.org/10.1111/ejed.12909

Pusporini, W., & Nurdiyanto, H. (2024). Utilization of artificial intelligence in Outcome-Based Curriculum Evaluation and Development. *Journal of Research in Social Science And Humanities, 4*(1). https://doi.org/10.47679/jrssh.v4i1.109

Sánchez Prieto, J. C., Gamazo, A., Cruz, J., Therón, R., & García, F. J. (2020). AI-Driven Assessment of Students: Current Uses and Research Trends. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences. HCII 2020. Lecture Notes in Computer Science, vol 12205.* Springer, Cham. https://doi.org/10.1007/978-3-030-50513-4_22

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays — Should professors worry? *Nature*. https://doi.org/10.1038/d41586-022-04397-7

Thambi, M. (2024). Responding to AI: Redesigning an assessment in a final year ITE curriculum specialization topic. ASCILITE 2024 Conference. https://doi.org/10.14742/apubs.2024.1107

Wen, Z., Bai, E., & Li, M. (2024). An Evaluation of the Impact of Artificial Intelligence on university Students' Learning. *Journal of Innovation and Development, 6*(2). https://doi.org/10.54097/f3vvmp82

Zhao, C. (2024). AI-assisted assessment in higher education: A systematic review. *Journal of Educational Technology and Innovation, 6*(4). https://doi.org/10.61414/jeti.v6i4.209