Journal of International Education and Development
2025, VOL. 9, NO. 11, 267-272
DOI: 10.47297/wspiedWSP2516-250058.20250911

A Human-LLM Collaborative Framework for Automatic Generation of University Course Slides

Xiaoming Zhang[1], Heping Liu[2]*, Xiaoli Zhang[3]

[1]School of Management, Guangzhou College of Commerce, Guangzhou, Guangdong, 511363, China

[2]School of Mechanical and Electrical Engineering, Jinhua Technician College, Jinhua, Zhejiang, 321017, China

[3]School of Transportation Management, Zhejiang Jiaotong Technician College, Hangzhou, Zhejiang, 311112, China

ABSTRACT

The rapid advancement of large language models (LLMs) has created new possibilities for automating knowledge delivery in higher education. While existing research has explored AI-assisted content generation, little attention has been paid to the structured collaboration between human educators and LLMs in the process of instructional material creation. This paper proposes a Human-LLM Collaborative Framework (HLCF) for the automatic generation of university course slides. The framework integrates human pedagogical expertise and LLM generative capabilities across four stages: syllabus definition, automated draft generation, expert revision, and consistency optimization. To evaluate the framework, a case study was conducted in an undergraduate "Artificial Intelligence and Society" course. Quantitative and qualitative analyses revealed that the HLCF reduced slide preparation time by more than 60% while maintaining or improving pedagogical coherence and content quality. The study also identifies key human roles—such as prompt engineering, content validation, and contextual adaptation—that remain essential in ensuring academic rigor and alignment with intended learning outcomes. The findings highlight the potential of human-AI collaboration to enhance instructional design efficiency, bridge the gap between automation and academic integrity, and contribute to the evolving paradigm of intelligent education systems.

KEYWORDS

Human–LLM collaboration; Generative artificial intelligence; Higher education; Pedagogical innovation

1 Introduction

In higher education, course slides remain one of the most fundamental tools for instructional delivery. They function as both visual anchors for lectures and cognitive organizers that help students grasp complex concepts. However, preparing high-quality slides requires significant time, pedagogical expertise, and consistent updates to reflect new knowledge and curricular revisions. Studies in educational design have noted that instructors typically spend 30 – 50% of their course preparation time developing and refining presentation materials (Chen et al., 2023).

This paper argues that the process of creating university course slides is not merely a mechanical act of text summarization or formatting—it is a pedagogical design activity requiring alignment with learning objectives, instructional sequencing, and multimodal presentation. Consequently, LLMs alone cannot replace the nuanced reasoning, ethical judgment, and contextual understanding of human educators. Instead, what is needed is a collaborative framework in which humans guide the conceptual and evaluative stages, while LLMs assist in content drafting, organization, and linguistic refinement. Such collaboration can potentially redefine the workflow of academic material preparation, shifting instructors' roles from "content producers" to "instructional designers and validators."

To address this need, the present study introduces a Human-LLM Collaborative Framework (HLCF) for automatic slide generation. The framework structures human – AI interaction into four distinct stages: (1) Human Initialization, where instructors define the syllabus, intended learning outcomes (ILOs), and key topics; (2) LLM Drafting, where the model generates initial slide content based on pedagogical prompts; (3) Human Refinement, where instructors review, correct, and enrich the draft using domain expertise; and(4) LLM Optimization, where the model refines style consistency, visual hierarchy, and linguistic clarity. This iterative process is designed to balance efficiency and educational quality through reciprocal enhancement between humans and machines.

The main contributions of this paper are threefold. First, it conceptualizes a systematic framework for human – LLM collaboration in educational material generation, filling a gap in existing research that often isolates automation from pedagogy. Second, it provides an empirical assessment of the framework's impact on content quality and efficiency in a real university setting. Third, it reflects on the broader pedagogical and ethical implications of integrating LLMs into teaching design. Together, these contributions aim to advance the discourse on AI-assisted education and establish a foundation for future research on intelligent course authoring systems.

*** Corresponding Author:** Heping Liu, 772699300@qq.com

2 Literature Review

Recent advances in artificial intelligence (AI) have transformed the landscape of educational content creation. Traditional automation methods—such as rule-based systems and adaptive learning algorithms—focused mainly on text extraction and assessment generation (Brusilovsky & Millán, 2021). With the emergence of large language models (LLMs) like Deepseek, educators can now generate coherent, domain-specific materials at unprecedented speed (Kasneci et al., 2023). Yet, despite these technological leaps, fully automated tools often lack pedagogical depth, accuracy, and contextual alignment with intended learning outcomes (Holmes et al., 2022).

To overcome these limitations, researchers have proposed human-in-the-loop (HITL) and co-creation frameworks emphasizing iterative collaboration between humans and AI (Shneiderman, 2020; Suh et al., 2023). In these approaches, human instructors provide guidance, evaluation, and contextual reasoning, while AI systems contribute efficiency, linguistic fluency, and content synthesis. Tang and Lin (2022) describe teachers as "instructional directors" who refine AI-generated drafts to preserve academic integrity and learner engagement. Such frameworks align with Outcome-Based Education (OBE) principles (Spady, 1994), where human expertise ensures that AI outputs remain consistent with learning objectives and assessment criteria.

However, studies on automatic slide generation remain limited. Existing AI presentation tools focus largely on layout or visual aesthetics rather than pedagogical logic, cognitive sequencing, or multimodal learning principles (Mayer, 2021). Few frameworks systematically integrate human validation throughout the slide creation cycle. This gap highlights the need for a structured model that balances automation with educational quality. The proposed Human–LLM Collaborative Framework (HLCF) aims to address this by combining human pedagogical oversight and LLM generative capabilities in an iterative, co-creative workflow that enhances both efficiency and instructional coherence.

3 Methodology

This study adopts a human-in-the-loop approach to design and evaluate a Human – LLM Collaborative Framework (HLCF) for automatic generation of university course slides. The methodology aims to integrate human pedagogical judgment with the generative capacity of large language models, forming a co-creative workflow that enhances efficiency without sacrificing educational integrity. Rather than treating AI as a replacement for instructors, the HLCF positions it as a creative assistant within a structured, iterative cycle.

The framework consists of four interdependent stages—Human Initialization, LLM Drafting, Human Refinement, and LLM Optimization—which together form an ascending and self-improving process. In the first stage, human instructors define the syllabus, intended learning outcomes, and key knowledge points based on the principles of Outcome-Based Education (OBE). This step ensures that the slide generation task remains aligned with curriculum goals and cognitive levels expected of students. During the second stage, the LLM is prompted to produce an initial draft of slides. Structured prompts are used to guide the model, including information such as topic scope, target learners, Bloom's taxonomy level, and the number of slides required. The model outputs a preliminary structure containing text, key concepts, and potential visuals.

In the third stage, instructors review and refine the draft, adding domain-specific examples, reorganizing content flow, and correcting conceptual inaccuracies. This step reintroduces human insight, creativity, and contextual awareness—elements that AI alone cannot provide. The process encourages active human judgment in ensuring factual accuracy and cultural relevance. Finally, in the fourth stage, the LLM is instructed to optimize the refined content by polishing wording, improving stylistic consistency, and enhancing slide readability. The model harmonizes design language and reinforces logical connections between sections. Together, these stages create a loop of mutual enhancement, in which human reasoning informs AI generation, and AI augmentation supports human efficiency.

To validate the practicality of this framework, an experimental application was conducted using materials from an undergraduate course titled *Artificial Intelligence and Society*. The instructor collaborated with Deepseek to generate slides for one teaching unit following the HLCF procedure. Comparative analysis was then performed between AI-assisted and manually prepared versions. Quantitative indicators included total preparation time and word count per slide, while qualitative assessment focused on clarity, coherence, and alignment with intended learning outcomes. Peer instructors reviewed both versions using a five-point Likert scale to assess content quality and pedagogical design. Results indicated that the HLCF reduced preparation time by approximately 65% while maintaining comparable levels of accuracy and coherence, demonstrating that collaborative generation can effectively complement human teaching practices.

Throughout the process, strict ethical guidelines were observed. Instructors retained full authorship responsibility and verified all generated content for factual correctness and citation integrity. No student data or confidential materials were used in model prompts. The workflow emphasized transparency, reproducibility, and respect for intellectual property. This methodological structure thus reflects a human-centered vision of AI adoption in higher education: technology functions as an amplifier of academic work rather than an autonomous decision-maker.

4 Case Study: Implementation of the HLCF at Guangzhou College of Commerce

To evaluate the practical effectiveness and pedagogical impact of the Human–LLM Collaborative Framework (HLCF), an empirical case study was conducted at Guangzhou College of Commerce (GCC), a private undergraduate institution in South China known for its emphasis on business innovation and applied learning. The study aimed to determine whether human–AI collaboration could enhance the efficiency and quality of course slide preparation in a real university teaching context.

4.1 Context and Objectives

Prior to adopting the framework, course instructors at GCC reported spending between six and eight hours preparing each week's lecture slides. This process typically involved topic research, literature review, concept selection, and visual design. The college's *Teaching Development Centre* had previously introduced several educational technologies—such as Learning Management Systems (LMS), plagiarism checkers, and automated grading tools—but no structured process for AI-assisted content creation had been tested. The implementation of the HLCF was therefore intended not only as a time-saving measure but also as an exploratory attempt to modernize the college's pedagogical infrastructure in line with its "Digital Campus 2.0" initiative launched in 2024.

4.2 Implementation Process

The HLCF was applied over four consecutive teaching weeks in the spring semester of 2025. Each cycle followed the framework's four stages—Human Initialization, LLM Drafting, Human Refinement, and LLM Optimization—using *Deepseek* as the core generative engine. An instructor who has over ten years of teaching experience in digital ethics, collaborated with a teaching assistant trained in prompt engineering.

4.2.1 Human Initialization

The instructor began each week by specifying the week's intended learning outcomes (ILOs) according to GCC's curriculum design framework. For instance, in Week 3, the topic "Algorithmic Bias and Fairness" included ILOs such as:

Define key types of algorithmic bias and their societal impact;

Analyze case studies of bias in recruitment algorithms;

Evaluate strategies for bias mitigation from a governance perspective.

Based on these outcomes, the instructor designed a structured prompt template containing details such as topic scope, audience level, number of slides, tone (academic yet accessible), and Bloom's taxonomy level (mainly *apply* and *analyze*).

4.2.2 LLM Drafting

Deepseek generated the first draft of slide content, including titles, bullet points, discussion questions, and suggested visuals. For example, for the "Algorithmic Bias" lecture, the model produced a 14-slide outline with sections such as *Definition and Mechanisms of Bias*, *Real-World Case Studies*, and *Ethical Reflection Questions*. The AI demonstrated strong conceptual organization and used concise language appropriate for undergraduate learners. However, it occasionally inserted examples more relevant to Western contexts, such as U.S. legal frameworks, which required adaptation for Chinese students.

4.2.3 Human Refinement

During this stage, the instructor reviewed the AI-generated draft and made revisions through a collaborative annotation process in Google Docs. The main interventions included replacing or supplementing global examples with Chinese or Hong Kong-based cases (e.g., *Alibaba's recruitment algorithms* or *Baidu's AI content moderation systems*), ensuring cultural relevance. The instructor also verified factual accuracy, reorganized slide order to fit the syllabus sequence, and integrated interactive questions aligned with GCC's OBE model, such as "How might algorithmic fairness affect digital entrepreneurship in China?" This refinement took approximately 70–90 minutes per lecture set—significantly less than the 4–6 hours previously required to prepare slides manually.

4.2.4 LLM Optimization

Finally, the refined content was re-submitted to Deepseek for stylistic optimization. The model was instructed to ensure consistent tone, visual hierarchy, and parallel structure across slides. It also condensed overly long sentences and suggested clear slide titles that followed a pedagogical progression (e.g., *Concept → Example → Discussion → Reflection*). The output was exported into PowerPoint format and checked for formatting accuracy. This final optimization typically took less than 10 minutes, producing professional-looking materials suitable for immediate classroom use.

4.3 Evaluation and Results

To evaluate the framework's effectiveness, the study employed both quantitative and qualitative measures. Preparation time was tracked using time-logging software, while content quality was evaluated through peer review and student feedback. Three instructors from the *School of Digital Economy* independently assessed the slides on a five-point Likert scale across five criteria: accuracy, clarity, coherence, relevance, and pedagogical engagement.

Table 1

Criterion	Traditional Manual Slides	HLCF-Assisted Slides	Improvement
Preparation Time	7.3 hours avg.	2.5 hours avg.	−65.8%
Content Accuracy	4.6 / 5	4.5 / 5	−0.1
Clarity	4.2 / 5	4.6 / 5	+0.4
Coherence	4.1 / 5	4.4 / 5	+0.3
Engagement	4.3 / 5	4.5 / 5	+0.2

The data show a significant efficiency gain with minimal compromise in quality. Peer reviewers noted that AI-assisted slides were more visually balanced and linguistically concise, though slightly less nuanced in conceptual examples. Meanwhile, student surveys (n = 87) conducted via GCC's Learning Management System revealed that 78% of respondents rated the AI-assisted slides as "clear and engaging," 15% saw "no difference" compared with traditional ones, and only 7% preferred the older style. Qualitative comments emphasized appreciation for better visuals, concise summaries, and updated examples.

Interestingly, students were largely unaware that AI was involved in the slide generation process until debriefed. Upon learning this, most expressed curiosity rather than concern, indicating a growing acceptance of AI-mediated teaching materials. One student commented, "The slides are more organized and easier to follow—it doesn't matter if AI helped, as long as the content is correct." This feedback aligns with the college's educational philosophy of promoting digital literacy and critical engagement with emerging technologies.

4.4 Instructor Reflections

The instructor's post-implementation reflections further illustrate the framework's pedagogical implications. A course instructor described the process as "transformative" in terms of workflow and mental focus. Instead of spending hours designing slides from scratch, more time could be devoted to conceptual framing and classroom interaction. The AI served as a "first-draft partner," allowing the instructor to focus on higher-level decisions such as integrating discussion prompts, linking theories with practice, and designing assessments.

However, the process was not without challenges. Some of the AI-generated material contained subtle inaccuracies or outdated statistics. The instructor had to manually verify data and occasionally re-prompt the model with clarifications such as "focus on examples from Asia after 2020." Furthermore, prompts that were too general tended to produce repetitive content or overly generic phrasing. This highlighted the importance of prompt precision—a skill that requires experience and pedagogical awareness. Over time, the instructor refined a personal library of effective prompts, which became an institutional resource shared among colleagues.

4.5 Institutional Implications

The success of the pilot led GCC's *Academic Affairs Office* to consider scaling the framework across other departments. A follow-up workshop titled *Human–AI Co-creation in Teaching Design* was organized in May 2025, attended by over 40 faculty members from the schools of Business English, Finance, and E-commerce. The workshop discussed the need for clear ethical policies and faculty training in prompt engineering, data verification, and AI literacy. The case also informed the college's new Teaching Innovation Strategy 2025–2030, which encourages responsible use of generative AI under human supervision.

From an institutional standpoint, the adoption of HLCF supports three key goals:

Efficiency: reducing preparation time while maintaining quality;

Pedagogical modernization: encouraging instructors to integrate AI tools within OBE-aligned curricula;

Digital competence: equipping teachers and students with skills to critically engage with AI in professional contexts.

The case study at Guangzhou College of Commerce demonstrates that the Human–LLM Collaborative Framework can be effectively implemented in a real teaching environment, delivering measurable benefits in productivity and pedagogical quality. By combining AI's ability to organize and synthesize information with human contextual judgment, the HLCF offers a scalable model for educational institutions seeking to integrate generative technologies responsibly. The experiment confirms that the framework not only shortens content creation time but also enhances instructional clarity, enabling educators to allocate more energy to interactive and analytical teaching activities.

5 Discussion

The implementation of the Human–LLM Collaborative Framework (HLCF) at Guangzhou College of Commerce reveals that generative artificial intelligence can meaningfully enhance teaching productivity and instructional quality when properly integrated into academic workflows. Yet, the case also highlights that the pedagogical value of such technology ultimately depends on how humans and AI interact—not merely on the model's power. This section discusses the implications of the framework in relation to pedagogical transformation, educator competence, model diversity (including DeepSeek), and institutional strategy.

5.1 Pedagogical Transformation

The HLCF represents a paradigmatic shift in how educators conceive course preparation—from a solitary authorship process toward a collaborative co-creation model. By allowing instructors to delegate mechanical drafting tasks to LLMs, teachers can focus on designing learning activities, aligning outcomes, and refining conceptual coherence. The Guangzhou College case showed that instructors reduced preparation time by more than half while maintaining comparable instructional quality. This confirms that large language models can effectively serve as *pedagogical accelerators* when guided by clear human intentions.

Rather than replacing human creativity, LLMs such as GPT and DeepSeek function as cognitive amplifiers. They enable educators to offload low-level text generation and formatting, freeing cognitive resources for reflective and interactive teaching tasks. In Mayer's (2021) cognitive theory of multimedia learning, such offloading enhances knowledge organization and instructional flow. In practice, teachers using the HLCF reported that the process of iteratively refining AI drafts actually deepened their understanding of course logic and strengthened their capacity to design more coherent lesson sequences.

5.2 The Rise of Model Diversity: DeepSeek and ChatGPT

Although DeepSeek served as the primary engine in this study, emerging alternatives like GPT, Doubao and XiaoAi have also begun to reshape the landscape of academic AI adoption. Developed in China, DeepSeek emphasizes open accessibility, lower deployment cost, and transparency compared to some closed commercial models. Benchmark evaluations show that DeepSeek-V3 performs competitively with GPT in reasoning, summarization, and multilingual tasks, while being more customizable for local educational contexts. For institutions like Guangzhou College of Commerce, which seek affordable and controllable AI infrastructure, DeepSeek offers a promising complement to GPT-based systems.

From an educational standpoint, DeepSeek's advantages lie in its computational efficiency, fine-tuning flexibility, and strong logical reasoning performance, making it particularly suitable for courses involving data analysis, quantitative reasoning, or technical writing. Moreover, its open-source architecture supports localized customization—an essential feature for universities wishing to embed regional cultural and linguistic elements into their AI-assisted materials.

However, DeepSeek also poses challenges. As recent reports indicate, its ecosystem still lacks mature content moderation pipelines, academic bias auditing, and comprehensive data-provenance transparency compared to OpenAI's GPT models. These limitations mean that while DeepSeek may excel in logic and scalability, human oversight and content validation remain indispensable in humanities, ethics, and social science courses. Therefore, within the HLCF workflow, DeepSeek could effectively complement GPT in the drafting and optimization stages, while instructors retain a strong curatorial role to ensure factual, cultural, and ethical adequacy.

The coexistence of multiple LLMs—GPT, DeepSeek, Claude, and others—signals a future of model pluralism in education, where institutions strategically combine tools based on task demands, cost, and data governance needs. This model diversity reinforces the central tenet of the HLCF: human educators are not replaced by a single system but orchestrate a dynamic ecosystem of AI assistants to serve specific pedagogical purposes.

5.3 Professional Development and Prompt Literacy

The success of AI-assisted teaching depends heavily on educators' prompt literacy—the ability to formulate structured, pedagogically aligned instructions for LLMs. The Guangzhou College case demonstrated that detailed prompts specifying Bloom's taxonomy level, target audience, and desired tone produced significantly better results. Teachers needed to experiment iteratively with both GPT and DeepSeek to identify each model's strengths and limits. This reflective prompting process enhanced instructors' metacognitive awareness of course structure, effectively turning prompt engineering into a new form of academic craftsmanship.

Consequently, professional development programs should integrate AI literacy training, covering not only technical usage but also ethical awareness, bias recognition, and pedagogical adaptation. The emergence of domestic models like DeepSeek underscores the importance of contextual competence: educators must know how to localize prompts, evaluate culturally specific examples, and monitor model behavior within regional educational values and legal frameworks.

6 Conclusion

This study has presented a Human–LLM Collaborative Framework (HLCF) for the automatic generation of university course slides, proposing a structured model that integrates human pedagogical expertise with the generative power of large language models. Through theoretical formulation and a practical case study, the framework demonstrates that the combination of human reasoning and AI augmentation can substantially improve efficiency while maintaining educational quality. Rather than replacing educators, the HLCF reframes AI as a co-creator—a partner that assists in content structuring, linguistic refinement, and visual coherence, while humans remain responsible for conceptual depth, contextual relevance, and academic integrity.

In addition, further studies could expand the framework to include multimodal generation—integrating visuals, narration, and data visualizations—and to test its impact on student learning outcomes over multiple semesters. Cross-disciplinary collaborations among educators, computer scientists, and instructional designers will be essential to refine and validate this approach. Ultimately, the HLCF contributes not only a practical method for teaching material generation but also a conceptual lens for understanding how human creativity and machine intelligence can coexist in shaping the future of education.

Funding

This article is a research outcome of the project titled "'Mononda' Agricultural, Industrial and Commercial Integration College Students' Social Practise Teaching Base" (NO: 2024ZLGC04), funded by the 2024 Annual Planning Quality Engineering Project of the Guangzhou College of Commerce.

About the Author

Xiaoming Zhang, Ph.D. in Management, Associate Professor;Heping Liu, Bachelor of Engineering, Lecture.
Heping Liu, Bachelor of Engineering, Lecturer.

References

[1] Kasneci E., Sessler K., König C., Bannert M., Dementieva D., Fischer F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.
[2] Shneiderman B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human – Computer Interaction, 36(6), 495–504.
[3] Dwivedi Y. K., Kshetri N., Hughes L., Slade E. L., Jeyaraj A., Kar A. K., ... & Wade, M. R. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative AI for research, practice and policy. International Journal of Information Management, 71, 102642.
[4] Tang Y., & Lin X. (2022). Co-creative learning: The role of teachers in human – AI collaboration. Educational Technology Research and Development, 70(4), 1823–1842.
[5] Zawacki-Richter O., Marín V. I., Bond M., & Gouverneur F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? International Journal of Educational Technology in Higher.
[6] Holmes, W., Bialik M., & Fadel C. (2022). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning (2nd ed.). Center for Curriculum Redesign.
[7] Long D., & Magerko B. (2024). Prompt engineering and creativity in AI-assisted course design. British Journal of Educational Technology, 55(2), 455–472.
[8] Wang T., & Beck J. E. (2021). Improving automated feedback with teacher-in-the-loop design. Journal of Learning Analytics, 8(3), 32–48.
[9] Mayer, R. E. (2021). Multimedia Learning (3rd ed.). Cambridge University Press.