EduPresenta: A Conversational AI Agent for Pedagogically Sound Presentation Generation for Instructors

Saha Kuljit Shantanu[*†1,2][0009−0000−6175−0970], Saad Mohammad Rafid Pial[*1][0009−0008−5121−8348], and Sadia Sharmin[1][0000−0002−6381−1308]

[1] Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh

[2] Department of Computer Science and Engineering, International University of Business Agriculture and Technology, Dhaka 1230, Bangladesh

sahakuljitshantanubuet@gmail.com, saadmrp@gmail.com, sadia@teacher.cse.buet.ac.bd

[*]These authors contributed equally to this work.

[†]Corresponding author. Email: sahakuljitshantanubuet@gmail.com

Abstract. Recent researches are widely focusing on applying AI-based system to generate academic presentations, as reflected through efforts like AutoPresent, SlideSpawn, Presentify and PPTAgent. This progress has paved research directions to work on summarizing research works into a presentable format. However, the reliability of such systems in contexts of teaching-learning remains unexplored, where considering the pedagogical matrix is highly significant compared to the content organization and visual quality of the presentation slides. Additionally, capabilities of AI to summarize and organize contents may not necessarily map to effective learning support for students. In educational contexts, this approach fails to address a fundamental requirement: slides must align with specific pedagogical goals including learning objectives, cognitive complexity levels (Bloom's Taxonomy), depths of knowledge and outcome-based education frameworks. In this work, we introduce *EduPresenta*, an AI-based system that generates presentation to fulfil the stated requirement with support for conversational editing. We employ human educators, specifically Science, Technology, Engineering and Math (STEM) instructors of undergraduate studies, who use this system to generate presentation slides for the courses they are teaching or once taught. Later on, we assess the correctness of the system to maintain pedagogical goals and alignment through the utilization of the resources provided by the instructors and necessary conversation regarding edits. We conclude that, though *EduPresenta* has the potential to generate an appealing presentation, it is not intelligent to function without human guidance in ensuring pedagogical goals and alignment.

Keywords: Educational slides · Pedagogical goals · STEM · Human-in-the-Loop.

1 Introduction

Generating Presentations with the help of a conversational AI has been a very popular topic to work with in recent days. Existing AI-based systems have exhibited moderate proficiency in generating academic presentation slides. Although this practice has solved a set of purposes with a set of different approaches, the work has often been overlooked in the field of education. Educational slides often require contents based on profound understanding that will achieve some organized learning goals. Modern AI tools provide no structured mechanisms to capture the educator's pedagogical goals and teaching objectives [5]. The engagement of instructor driven perspective to modern studies will be a paradigm shift in generating presentations that overcome this limitations.
Alternatives such as Gamma AI and SlidesAI can generate visually appealing educational presentations with promising contents. However, significant challenges remain to ensure reliability, validate factual correctness and achieve pedagogical objectives. Prior studies claim that AI models may produce misleading information [9], underscoring the need for educator validation. In educational slides, the factual accuracy and pedagogical paradigms are more crucial than content organization and visual layout. Failure of system in following instructor guidance may lead to the worst case: Generation of a presentation lacking reliability and alignment with teaching goals.
Current systems relegate pedagogical considerations to unstructured prompts, making educational intent implicit and easily violated during generation or refinement. Furthermore, the trustworthiness of AI-sourced content remains a critical concern for educators who are accountable for accuracy and pedagogical appropriateness of teaching materials [18]. Motivated by these concerns, we introduce **EduPresenta**, that prepares and updates educational slides whilst maintaining a simultaneous communication with an educator. The conversation is initiated through plain text and resources by an instructor of a particular course. In response to the prompt, *EduPresenta* generates a teaching presentation. All the slides of the generated presentation are subjected to instructor driven editing. The instructor can iteratively edit the slides until the presentation satisfies their pedagogical goals. This study evaluates the potential of the proposed system in preparing a teaching effective presentation and focuses on investigating the following question:

- **RQ1:** Can presentations generated using AI agents reduce cognitive load of instructors?
- **RQ2:** Do the generated presentations align with the particular teaching objectives of the instructors?
- **RQ3:** Are the generated presentations reliable and trustworthy?

We employ a particular number of instructors from STEM fields to assess the credibility of the system. The findings of this study contribute to the field of intelligent slide generation systems and the role of AI in teaching. Particularly, our work makes the following contributions:

- We introduce a **Human-in-the-Loop AI model** that leverages continuous improvements to generate an appealing presentation.
- We introduce a **pedagogically interactive mechanism** that adapts to pedagogical requirements across multiple STEM fields.
- We performed an empirical analysis of **the new system** based on the feedback from experts.

2 Related Work

The automated generation of presentation slides is a long-standing challenge that intersects natural language processing, computer vision, and human-computer interaction. Our work, *EduPresenta*, builds upon two primary streams of research: (1) systems for the automated generation of presentations from various source materials, and (2) the integration of pedagogical principles into educational technology. While significant progress has been made in the former, the latter remains a critical but underexplored dimension in the context of AI-driven content creation.

2.1 Automated Presentation and Content Generation

Early research into automated slide generation primarily focused on extracting and summarizing textual content from structured documents. These systems often employed rule-based and extractive summarization techniques, relying on discourse structure analysis [16], semantic annotations [19], or query-specific summarization [17] to identify and organize key points. Some works leveraged the inherent structure of technical papers [13] or LaTeX documents [21] to generate basic slides, demonstrating the feasibility of automating this tedious task. However, these early methods were often limited in their semantic understanding, produced text-heavy slides, and offered minimal control over visual layout and design. The advent of Large Language Models (LLMs) has catalyzed a new generation of more sophisticated systems. Researchers have explored multi-staged, end-to-end pipelines that divide the complex generation task into manageable sub-problems, such as outline generation, content mapping, and image extraction [1]. A significant advancement has been the shift towards agent-based frameworks that mimic human workflows. For instance, recent work introduces an edit-based approach where an agent modifies reference slides [23], while other systems develop language-driven agents for performing precise edits on existing presentations [11]. These systems offer greater control over multimodal content and layout. Additional recent work has focused on generating slides from scratch based on natural language instructions [6] or creating complex multimodal posters from full scientific papers [14], pushing the boundaries of what AI can achieve in visual design. Furthermore, a specialized sub-field has emerged focusing on generating presentations from computational notebooks. Recent systems [20, 22] aim to help data scientists communicate their findings by transforming code and analyses into coherent narratives. These tools highlight the importance of a human-in-the-loop paradigm and the utility of structured inputs, such as user-provided outlines, to guide the generation process. Concurrently, research on enhancing the accessibility and consumption of presentations, for instance by adding synchronized highlights [8] or extracting structure for non-visual access [15], underscores the field's growing focus on the end-user's experience. Despite these significant technical advancements, a common limitation persists: these systems are primarily optimized for content fidelity, structural coherence, and visual appeal, but they generally lack an explicit framework for incorporating pedagogical principles. They do not formally consider the learning objectives, the target audience's prior knowledge, or the specific educational context, which is the critical gap our work aims to address.

2.2 Pedagogical Principles in Educational Technology

The effectiveness of presentations as educational tools has been a subject of study long before the rise of AI. Foundational work established that the pedagogical value of slides depends heavily on their design [2], noting that the inclusion of multimedia elements like images or sounds can sometimes be detrimental to learning if not directly relevant to the content. This highlights a core principle: effective educational content is not merely about information delivery but about facilitating comprehension and retention. With the integration of AI into education, this principle has become even more critical. A growing body of research is exploring the design and impact of AI-powered pedagogical agents and educational tools. For example, research investigates the design of interactive pedagogical agents to assist higher education instructors [4], emphasizing the need for systems that align with teachers' existing workflows and pedagogical goals. Similarly, work on generating personalized educational content, such as transforming presentation slides into customized textbooks [24], demonstrates the potential of LLMs to adapt content to diverse learning needs. A crucial aspect of this research is the evaluation of AI-generated educational content not just for factual accuracy but for its pedagogical utility. Recent work proposes frameworks for evaluating the pedagogical value of LLM-generated explanations [10], arguing that a good explanation must be tailored to the learner's context and knowledge level. This shift from technical correctness to pedagogical effectiveness is a hallmark of mature educational technology. Our work, **EduPresenta**, is situated at the intersection of these two research areas. While the systems reviewed in the first subsection have made remarkable progress in the *technical* automation of slide generation, they have largely operated without the explicit pedagogical awareness explored in the second. EduPresenta addresses this gap by introducing a novel framework where pedagogical parameters such as learning objectives, audience knowledge level, and content focus are treated as first-class inputs to the generation process. By allowing educators to guide the AI with specific educational goals and to iteratively refine the output through an interactive interface, our system moves beyond simple content automation to facilitate the creation of truly educationally-grounded presentations.

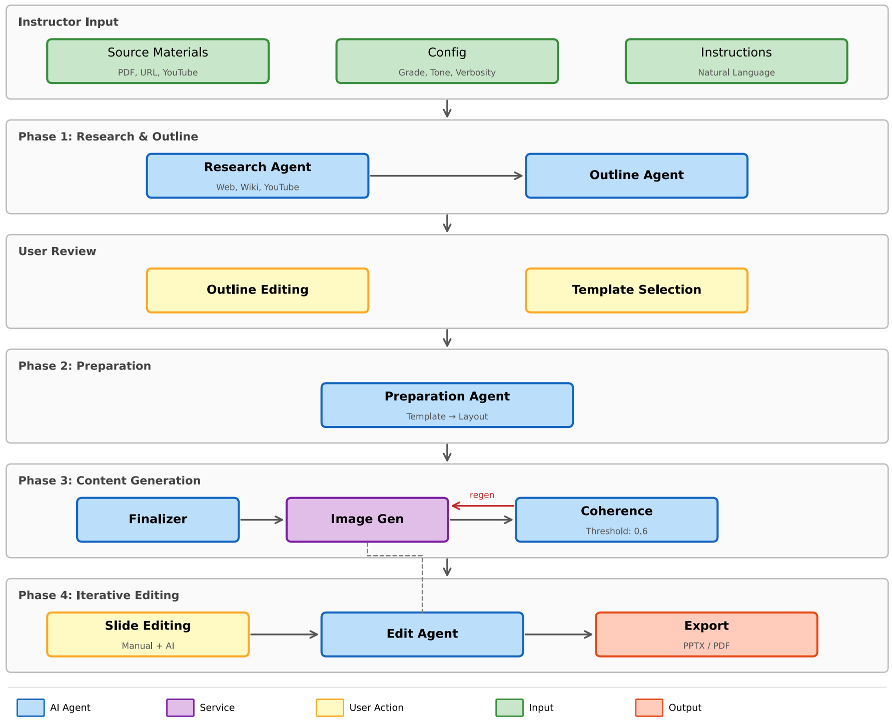

Fig. 1. Architecture of EduPresenta showing the multi-agent pipeline from input processing through slide generation and editing.

3 System Description

This section describes *EduPresenta*, presenting the user workflow, multi-agent architecture, and design decisions supporting instructor control over pedagogical outcomes.

3.1 System Overview

EduPresenta accepts instructor-provided materials and specifications, generates a presentation through a multi-agent pipeline, and supports iterative refinement. The workflow proceeds through three stages: input configuration, outline review with template selection, and slide editing. Figure 1 illustrates the system architecture.

3.2 User Workflow

Figure 2 shows the interface across the three workflow stages.

Input Configuration Instructors provide source materials and configuration parameters (Figure 2A). The system accepts multiple input types: PDF documents, web URLs, YouTube videos, and plain text. These sources ground the presentation in instructor-validated content.

Configuration parameters include grade level, tone (academic, professional, or casual), verbosity, output language, and number of slides. A free-text instruction field captures additional pedagogical requirements, such as "emphasize practical applications," "align with Bloom's taxonomy comprehension level," or "include formative assessment questions." These instructions allow instructors to specify learning objectives, cognitive targets, and content emphases that persist through the generation pipeline.

Outline Review and Template Selection The system generates an outline with slide titles and content points (Figure 2B). Instructors can reorder slides, edit content, add new slides, or remove unwanted ones. On the same page, a separate tab provides template selection (Figure 2C). Available templates span multiple categories with layouts suited for different content types, including content-image splits, data tables, formula highlights, and more. Each template defines the visual structure and content slots for generated slides. After reviewing both outline and template, instructors trigger slide generation.

Slide Editing and Export Generated slides appear in an editor view (Figure 2D). Instructors can modify content directly or use per-slide AI editing by clicking an edit button and describing the desired change (e.g., "add a concrete example," "simplify this explanation," or "use a different image"). The system updates content and regenerates images as needed while preserving the slide layout. Completed presentations export to PPTX or PDF.

3.3 Multi-Agent Architecture

The generation pipeline has six agents. A *Research Agent* processes input sources using web search, Wikipedia, and YouTube retrieval tools, producing a summary. A *Slide Outline Agent* converts this into a presentation structure based on the instructor's parameters.

After instructor approval of the outline, a *Slide Preparation Agent* analyzes the selected template to determine content requirements for each layout type. A *Slide Finalizer Agent* then generates content for each slide, embedding image prompts where visual elements are needed.

Images are sourced through AI generation or web search. Generated images pass through an *Image Coherence Agent* that scores text-image alignment from 0 to 1. Images below 0.6 are regenerated up to three times or flagged for review. This catches a common failure mode: images that look plausible but misrepresent the content.

When instructors request modifications, a *Slide Edit Agent* processes the natural language instructions with full presentation context, determining whether to update text, regenerate images, or both.

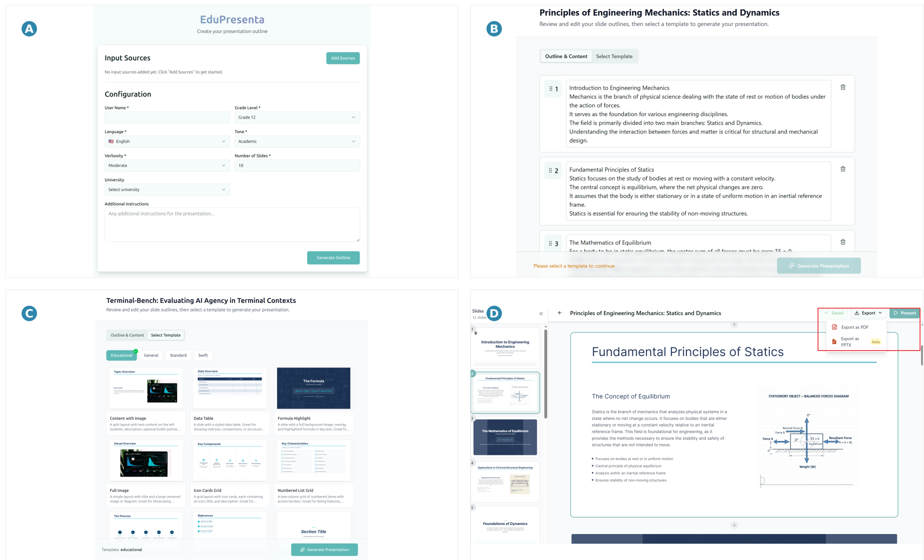

Fig. 2. EduPresenta interface: (A) source upload and parameters; (B) editable presentation outline; (C) template selection with layout previews; (D) slide editor supporting direct and AI-assisted modifications.

3.4 Design Rationale

The workflow places human checkpoints before expensive operations. Instructors review the outline before slide generation runs, so unwanted content gets cut early. Per-slide editing preserves context across the presentation, so fixing one slide does not require regenerating the rest.

4 Evaluation

We conducted a user study with university educators to evaluate whether *EduPresenta* produces pedagogically aligned presentations, matches instructor objectives, and earns user trust.

4.1 Study Design

Participants. We recruited 12 educators from four universities in Bangladesh, covering nine departments across science, engineering, health, and agricultural disciplines. Qualifications ranged from BSc to Post-doctoral, with teaching experience from under one year to eight years. Table 1 lists participant profiles.

Table 1. Participant demographics (n=12).

ID	Department	Qualification	Experience
T01	Nursing	MSc	4 years
T02	Mathematics	MSc	7 years
T03	Computer Science	BSc	<1 year
T04	Software Eng.	BSc	1 year
T05	Mechanical Eng.	MSc	4 years
T06	Computer Science	MSc	1 year
T07	Electrical Eng.	BSc	<1 year
T08	Computer Science	BSc	1 year
T09	Civil Eng.	BSc	1 year
T10	Agricultural Sci.	Post-doc	1 year
T11	Environmental Eng.	Post-doc	8 years
T12	Electrical Eng.	BSc	1 year

Procedure. Each participant received a brief walkthrough of *EduPresenta* and then selected a topic from a course they teach. They used the system to generate a full presentation: providing source materials, setting pedagogical parameters, reviewing and editing the outline, choosing a template, and refining the final slides. Afterward, participants completed a structured survey with 5-point Likert scales and multi-select items. We have taken their consent to publish their responses and all the comments they make.

Measures. Survey items were grouped into four constructs, each mapped to one or more research questions:

- *Pedagogical alignment* (RQ2): There are four traditional Depth of Knowledge (DoK) levels by Norman Webb [3] and six cognitive domains classified by Bloom [12]. Three DOK levels and five cognitive domains can be attained in presentations. We classified the three DOK levels into eight sublevels. Participants selected which DOK sublevels and Bloom's Taxonomy domains the presentation covered, and which it should have covered.
- *Cognitive load* (RQ1, NASA-TLX): Five items adapted from the Raw TLX (RTLX) approach on a 5-point Likert scale [7], covering mental demand, temporal demand, effort, frustration, and performance. Physical demand was omitted as the task is purely cognitive. Rated for both EduPresenta and manual preparation.
- *Trust and alignment* (RQ3): Three items on content accuracy trust, AI intent matching, and content retention. Two additional items on teaching confidence and objective alignment.
- *Usability*: Three items on learnability, adoption intent, and recommendation likelihood. Rated for both EduPresenta and participants' usual tools.

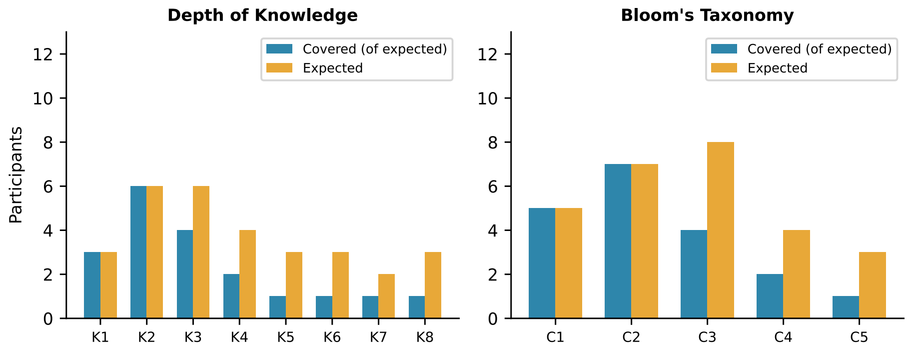

Fig. 3. Pedagogical attainment (n=12): number of participants who expected each learning objective to be attained versus how many of those had it attained by the generated presentation.

4.2 Results

Pedagogical Alignment. Figure 3 shows, for each DOK sublevel and Bloom's cognitive domain, how many participants expected that learning objective to be achieved and how many of those participants found it attained in the generated presentation.

At the participant level, about half reported full alignment between their expected and attained DoK levels, with the remainder showing partial or no overlap. Bloom's Taxonomy alignment was similar, with roughly 40% reporting full attainment. Lower cognitive levels (K1–K3, C1–C2) were consistently covered when expected, while higher-order levels (K5–K8, C3–C5) were frequently expected but not attained.

Cognitive Load (NASA-TLX). We adapted the Raw TLX (RTLX) approach to measure cognitive workload on a 5-point Likert scale across five dimensions: mental demand, temporal demand, effort, frustration, and performance satisfaction. All scores were oriented so that lower values indicate less workload. Each participant's RTLX is the unweighted mean of these five dimensions. Figure 4 compares EduPresenta with manual preparation.

The overall RTLX was 2.38 (SD=0.65) for EduPresenta and 3.35 (SD=0.92) for manual preparation, a 28.9% reduction in perceived workload. Temporal demand and frustration showed the largest gaps (2.17 vs. 3.83 and 2.17 vs. 3.92, respectively). Mental demand was closer (3.33 vs. 3.58). Reviewing and refining AI output still requires cognitive effort, but the procedural burden of building slides from scratch is reduced.

Trust, Alignment, and Usability. Figure 5 shows trust, teaching confidence, and usability ratings.

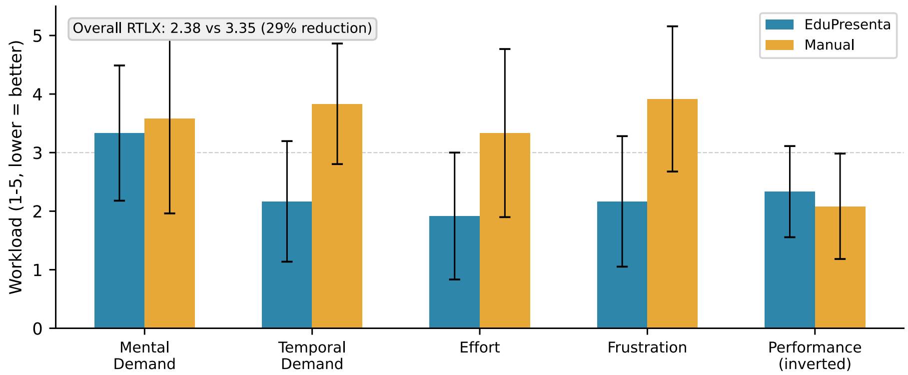

Fig. 4. NASA-TLX workload comparison (n=12, 1–5 scale, lower = less workload). EduPresenta reduces overall cognitive workload by 28.9% compared to manual preparation.

Trust in accuracy scored 4.00 (SD=0.74), AI intent matching 3.75 (SD=0.45), and content retention 3.75 (SD=0.62). Teaching confidence scored 3.75 (SD=0.87) and teaching objective alignment 4.08 (SD=0.67), suggesting the generated output works as a usable starting point. On usability, EduPresenta scored higher than participants' usual tools on all three items: learnability (4.42 vs. 3.83), adoption intent (3.75 vs. 3.33), and recommendation (3.92 vs. 3.42). Learnability at 4.42 (SD=0.51) was the highest-rated item in the survey.

4.3 Findings from instructors

All the instructors were suggested to comment upon system issues and their preferences. They got to comment on the both merits and limitations of the system. Most instructors(n=9, 75%) agreed that content accuracy is one of the reasons that contributed to minimizing their cognitive workload. However, a few instructors contested this claim. One of our participants (T04) noted some observations on our system: *"The topics were clearly structured. The system, however, can not extract relevant information from double hop links and drive links"(T04).* Another participant (T06) added: *"It would be better if the system accepted all types of documents (sheets, ppt etc)"(T06)* Apart from content accuracy, few instructors(n=5, 41.7%) identified visual clarity as a significant feature of our work. One participant (T07) appreciated the visual clarity: *"Visual clarity and explanations were much good in the slide." (T07).* Nevertheless, instructors also expressed some of their concerns regarding visual appearance. Four of our participants(33.3%) were concerned about the size of content font being insufficient for readability in presentation. One participant (T04) suggested that the system should have flexible theme selection for instructors: *"Some more default presentation templates would have been nice too. Overall it is a great system but has*

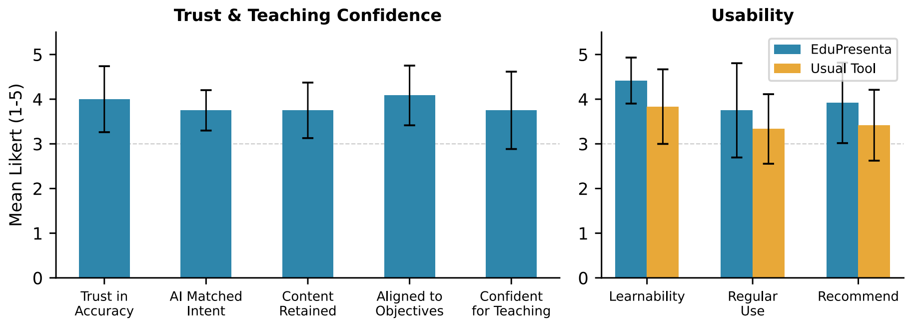

Fig. 5. Trust, teaching confidence (n=12, EduPresenta only), and usability comparison (EduPresenta vs. usual tool). Error bars show standard deviation.

scopes of improvement." (T04) On the other hand, one of our participants(T08) appreciated the existing visual layouts: *"The slide layout worked really well. I selected the swift version and it was very interesting." (T08)* Despite all minor issues, our system did not compromise with pedagogical quality of the presentations. Most instructors (n=7, 58.3%) found the presentations generated from our system suitable for student level adaptation.

5 Discussion

In this section we answer all our research Questions corresponding to the findings from our study.

RQ1: Can presentations generated using AI agents reduce cognitive load of instructors?

We have assessed the mental demand, temporal demand, effort, performance and frustration for both scenario (a) an instructor prepares their own presentations on a certain topic under a certain course and (b) the same instructor prepares their presentations on that same topic using our system. In case(a), the load might appear from an effort to prepare everything from scratch. In contrast, load in case(b) might come from selecting resource, thinking critically while writing a prompt, understanding and editing outline or manipulating slides. Our findings portray that the performance comparison between the manual tool and our proposed system is context dependent. Our system, however, happens to reduce effort and frustration of instructors while preparing the presentation as the contents generate automatically following their guidance.

RQ2: Do the generated presentations align with the particular teaching objectives of the instructors?

In our work, we have taken pedagogical alignment as the primary criteria for evaluation. Each of the presentations were evaluated by instructors with a provided description on Bloom's Taxonomy in Five Cognitive Domains. For a fair assessment, we have categorized three coverable depths of knowledge levels into eight knowledge sublevels to ensure a better clarity for instructors. The instructors expect their presentations to attain the maximum of their teaching objectives. As a result, it is not feasible for a presentation generated using AI agents to entirely satisfy the expectations of instructors. Our findings, however, suggest that our proposed system generates presentations which are able to meet the expectations of the instructors substantially.

RQ3: Are the generated presentations reliable and trustworthy?

The instructors have generated presentations from our system guided by the manual. Influenced by their responses, we have evaluated trust in accuracy, AI generated content and teaching confidence for the generated presentations. Additionally, we compare the overall response on the usability of our system with the tools instructors generally use for preparing their presentation slides. The substantial pedagogical alignment of the generated presentations make them reliable for the instructors . Analysis of the resources validated by instructors influences their trust and confidence on the system. In our findings, our proposed system outperforms the baseline presentation tools in learnability, regular usage and recommendations.

6 Limitations & Future Work

Our work mostly focuses on generating presentation slides that retain teaching qualities and achieve pedagogical goals for educators. However, our work is not yet evaluated from the learners' perspectives. Moreover, the generated presentation still lacks robustness in proper analysis of resources provided. While we specialized our work from educational perspective, we compromised with the user flexibility on visual layout of the presentation slides. Sometimes, the system hallucinates and performance degrades after Human-in-the-Loop modification. But our system lacks a version control approach to stash versions, raising inability in reverting to previous state. Additionally, we envisioned the system to function for overall STEM education. This resulted program-specific flaws in the system, like designing codes, graphs and circuits. We aim to extend our work to enclave validation from both teaching and learning perspectives. Our future direction of this work also includes enhancement of robustness in resource analysis and flexibility in manipulating contents. The flexibility will also ensure in removing program-specific flaws, optimizing export quality and reinstating previous versions.

7 Conclusion

Our work explores a system that generates presentation with instructor guided scaffolding of courses. We contribute to perform an empirical evaluation grounded in cognitive load theory and attainment of pedagogical goals. Our findings state that our work outperforms the baseline in reduction of cognitive load. While all the criteria are assessed by undergraduate instructors across STEM disciplines, the evaluation of the quality of this work from learners' perspective still remains as a future direction to the work.

References

1. Bandyopadhyay, S., Maheshwari, H., Natarajan, A., Saxena, A.: Enhancing presentation slide generation by llms with a multi-staged end-to-end approach. arXiv preprint arXiv:2406.06556 (2024)
2. Bartsch, R.A., Cobern, K.M.: Effectiveness of powerpoint presentations in lectures. Computers & education **41**(1), 77–86 (2003)
3. Bennet, D., Bennet, A.: The depth of knowledge: surface, shallow or deep? Vine **38**(4), 405–420 (2008)
4. Chen, S., Metoyer, R., Le, K., Acunin, A., Molnar, I., Ambrose, A., Lang, J., Chawla, N., Metoyer, R.: Bridging the ai adoption gap: Designing an interactive pedagogical agent for higher education instructors. arXiv preprint arXiv:2503.05039 (2025)
5. Díaz, B., Nussbaum, M.: Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence. Computers & Education **217**, 105071 (2024)
6. Ge, J., Wang, Z.Z., Zhou, X., Peng, Y.H., Subramanian, S., Tan, Q., Sap, M., Suhr, A., Fried, D., Neubig, G., et al.: Autopresent: Designing structured visuals from scratch. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 2902–2911 (2025)
7. Georgsson, M.: Nasa rtlx as a novel assessment for determining cognitive load and user acceptance of expert and user-based evaluation methods exemplified through a mhealth diabetes self-management application evaluation. In: pHealth 2019, pp. 185–190. IOS Press (2019)
8. Holmberg, A.: Generating narrated lecture videos from slides with synchronized highlights. arXiv preprint arXiv:2505.02966 (2025)
9. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., Fung, P.: Survey of hallucination in natural language generation. ACM computing surveys **55**(12), 1–38 (2023)
10. Joshi, B., He, K., Ramnath, S., Sabouri, S., Zhou, K., Chattopadhyay, S., Swayamdipta, S., Ren, X.: Eli-why: Evaluating the pedagogical utility of language model explanations. arXiv preprint arXiv:2506.14200 (2025)
11. Jung, K., Cho, H., Yun, J., Yang, S., Jang, J., Choo, J.: Talk to your slides: Language-driven agents for efficient slide editing. arXiv preprint arXiv:2505.11604 (2025)
12. Krathwohl, D.R.: A revision of bloom's taxonomy: An overview. Theory into practice **41**(4), 212–218 (2002)
13. Kumar, K., Chowdary, R.: Slidespawn: An automatic slides generation system for research publications. arXiv preprint arXiv:2411.17719 (2024)
14. Pang, W., Lin, K.Q., Jian, X., He, X., Torr, P.: Paper2poster: Towards multimodal poster automation from scientific papers. arXiv preprint arXiv:2505.21497 (2025)
15. Peng, Y.H., Chi, P., Kannan, A., Morris, M.R., Essa, I.: Slide gestalt: Automatic structure extraction in slide decks for non-visual access. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–14 (2023)
16. Shibata, T., Kurohashi, S.: Automatic slide generation based on discourse structure analysis. In: International Conference on Natural Language Processing. pp. 754–766. Springer (2005)

17. Sravanthi, M., Chowdary, C.R., Kumar, P.S.: Slidesgen: Automatic generation of presentation slides for a technical paper using summarization. In: FLAIRS (2009)
18. Taheri, R., Nazemi, N., Pennington, S.E., Clark, J.A., Dadgostari, F.: Factors influencing educators' ai adoption: A grounded meta-analysis review. Computers and Education: Artificial Intelligence p. 100464 (2025)
19. Utiyama, M., Hasida, K.: Automatic slide presentation from semantically annotated documents. In: Coreference and Its Applications (1999)
20. Wang, F., Lin, Y., Yang, L., Li, H., Gu, M., Zhu, M., Qu, H.: Outlinespark: Igniting ai-powered presentation slides creation from computational notebooks through outlines. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. pp. 1–16 (2024)
21. Yasumura, Y., Takeichi, M., Nitta, K.: A support system for making presentation slides. Transactions of the Japanese Society for Artificial Intelligence **18**(4), 212–220 (2003)
22. Zheng, C., Wang, D., Wang, A.Y., Ma, X.: Telling stories from computational notebooks: Ai-assisted presentation slides creation for presenting data science work. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. pp. 1–20 (2022)
23. Zheng, H., Guan, X., Kong, H., Zheng, J., Zhou, W., Lin, H., Lu, Y., He, B., Han, X., Sun, L.: Pptagent: Generating and evaluating presentations beyond text-to-slides. arXiv preprint arXiv:2501.03936 (2025)
24. Zhou, Y.: Slide2text: Leveraging llms for personalized textbook generation from powerpoint presentations. arXiv preprint arXiv:2503.17710 (2025)

A Depth of Knowledge Levels

The depths of knowledge(DOK) from level 1 to 3 have been further classified into the following eight sublevels, which were used in the pedagogical alignment assessment. Participants selected which sublevels the generated presentation covered and which it should have covered.

K1: Recall basic facts, symbols, and terminology
K2: Understand and explain basic concepts
K3: Interpret representations such as graphs, diagrams, or equations
K4: Apply standard procedures, formulas, or algorithms
K5: Independently apply knowledge to solve routine problems
K6: Analyze relationships, patterns, and underlying mechanisms
K7: Integrate multiple concepts across contexts or domains
K8: Evaluate methods or solutions using evidence and criteria

B Bloom's Taxonomy Domains

Five cognitive domains from Bloom's Taxonomy were used, each with a short description provided to participants:

C1: *Remember* — Definitions and facts
C2: *Understand* — Interpretation of concepts
C3: *Apply* — Solve real-life problems
C4: *Analyze* — Dig deep into the solutions to a particular problem
C5: *Evaluate* — Compare different solutions of an existing problem

C Evaluation Questionnaire

Table 2 shows how survey questions map to evaluation constructs. The full list of questions follows. All Likert items used a 5-point scale (Strongly Disagree to Strongly Agree).

Table 2. Construct-to-question mapping.

Construct	Questions
Pedagogical Alignment	Q1–Q4 (multi-select)
Cognitive Load (NASA-TLX)	Q5–Q9 (rated for EduPresenta and manual)
Trust & Teaching Confidence	Q10–Q14
Usability	Q15–Q17 (rated for EduPresenta and usual tool)

Q1. Which sublevels of knowledge does the presentation cover? (see Appendix A)
Q2. Which sublevels of knowledge should the presentation have covered?

Q3. Which Bloom's Taxonomy domains does the presentation cover? (see Appendix B)
Q4. Which Bloom's Taxonomy domains should the presentation have covered?
Q5. I had to put a little mental effort
Q6. I had trouble preparing it under time pressure
Q7. I had to put unnecessary efforts
Q8. I was frustrated for the process being lengthy
Q9. I was satisfied with my performance
Q10. I trusted the AI-generated content to be accurate
Q11. The AI suggestions matched what I had in mind
Q12. How much of the AI-generated content did you keep
Q13. The generated presentation can be confidently used for teaching
Q14. The generated presentation is aligned to my teaching objectives
Q15. It was easy to learn how to use the system
Q16. I would use this system regularly
Q17. I would recommend the tool to a colleague