InkSpire: An LLM-Powered System for Designing and Generating Disciplinarily Aligned and Context-Aware Reading Scaffolds

Yuwei Liang[1], Ruiqian Wang[1] and Bodong Chen[1]

[1]*University of Pennsylvania, Philadelphia, PA 19104, USA*

Abstract

Reading scaffolds aligned with disciplinary literacy practices and instructional contexts are essential for supporting engagement with disciplinary texts in higher education, yet remain difficult to design effectively at scale. We present InkSpire, an LLM-powered system for creating disciplinarily aligned and context-aware reading scaffolds for digital textbooks. InkSpire introduces a human-in-the-loop workflow for structured integration of instructional context and pedagogical considerations, along with a theory-grounded generation pipeline based on the disciplinary literacy framework [1] to produce scaffolds aligned with disciplinary epistemologies and reading practices. A preliminary pilot across 11 readings generated 56 scaffolds, with the majority accepted with minimal revision, suggesting the system's potential for supporting disciplinarily aligned and context-aware scaffold authoring at scale.

Keywords

Disciplinary literacy, Reading scaffolds, Human-in-the-loop, Large language models, Digital textbooks

1. Introduction

Reading is essential for academic success in higher education [2, 3]. Across disciplines, students are expected to engage extensively with textbooks as a core component of their learning [4]. However, texts in higher education are highly complex and disciplinary in nature, requiring students to to draw on epistemological understandings of disciplinary traditions to construct meaning and apply strategies aligned with discipline-specific standards of knowing [5, 6]—demands that general reading strategies often prove insufficient to address [7]. These discipline-specific literacy learning needs have inspired the development of disciplinarily aligned reading scaffolds as effective instructional support [8, 9].

However, designing high-quality disciplinary reading scaffolds for reading materials such as digital textbooks is a time-consuming, knowledge-intensive process [10]. Doing so reliably at scale is even harder, motivating research on automated question and scaffold generation for digital textbooks. While these systems have made meaningful progress, they exhibit critical limitations in producing disciplinarily aligned scaffolds sensitive to context. Earlier systems primarily relied on surface-level textual features for scaffold question generation [11, 12], making it difficult to capture the higher-order, discipline-specific epistemological differences and generate truly disciplinarily aligned scaffolds. Such generation approach also struggles

iTextbooks'26: Seventh Workshop on Intelligent Textbooks, June 28, 2026, Seoul, Republic of Korea

ywliang@upenn.edu (Y. Liang); wrq@upenn.edu (R. Wang); cbd@upenn.edu (B. Chen)

0009-0002-3561-3213 (Y. Liang); 0009-0000-8120-5320 (R. Wang); 0000-0003-4616-4353 (B. Chen)

to incorporate higher-order instructional contexts and pedagogical goals. While more recent LLM-based approaches offer greater generative flexibility [10, 13], existing research has largely overlooked whether generated scaffolds align with and support disciplinary reading practices, and has not sufficiently addressed the structured integration of instructional context and pedagogical considerations into the generation process.

To address these limitations, we present InkSpire, an LLM-powered scaffold design and generation system for higher education digital readings. By describing InkSpire's theory-aligned design and preliminary pilot findings, this paper makes the following contributions: (1) the design and implementation of InkSpire, a human-AI co-design system for context-aware and disciplinarily aligned scaffold generation; (2) a human-in-the-loop workflow that integrates instructional context and pedagogical considerations as foundational inputs to scaffold generation and positions instructors as active reviewers who validate and refine output scaffolds against their instructional goals; and (3) a theory-informed generation pipeline grounded in disciplinary literacy framework [1].

2. Related Work

2.1. Disciplinary Literacy and Textbook Reading in Higher Education

Traditional content-area reading instruction has emphasized general reading strategies, such as prediction, think-aloud, and activating prior knowledge, which are well established as effective for supporting basic comprehension [14, 15]. However, as literacy demands become more disciplinary and technical beyond secondary education, these strategies grow increasingly insufficient, as each discipline constructs and evaluates knowledge differently and thus requires distinct reading practices [16]. To address the disciplinary literacy needs, disciplinary literacy instruction advocates fostering discipline-appropriate literacy practices within content areas [17]. Goldman et al. [1] propose a conceptual framework that helps operationalize discipline-specific literacy learning goals through five core constructs: (a) epistemology—beliefs about the nature of knowledge and knowing; (b) inquiry practices/strategies of reasoning—ways in which claims and evidence are established and validated; (c) overarching concepts, themes, and frameworks—the core ideas and principles for warranting claims; (d) forms of information representation/types of texts—types of texts and media through which information is communicated; and (e) discourse and language structures—the oral and written language forms that express disciplinary knowledge.

In higher education, academic texts, such as textbooks, are often highly complex and disciplinary in nature [18, 19], and general reading strategies have proven insufficient when applied to such materials [7]. To support students' engagement with disciplinary texts, empirical research has explored and highlighted the value of discipline-specific reading scaffolds. Van Camp and Van Camp [9] found that scaffolded questions in weekly assignments significantly improved students' critical reading of psychology texts, while the READ project showed that discipline-specific tasks (such as discipline-relevant vocabulary building) improved pass rates and higher-order competencies across biology, engineering, and marketing courses [8].

These findings suggest that discipline-specific reading scaffolds are effective and that embedding them during reading holds significant promise. However, how to systematically design

and implement them in interactive digital textbooks remains underexplored, motivating our review of existing automated scaffold and question generation systems.

2.2. Current systems designed for scaffold generation and reading support

Research on automated scaffold and question generation for academic texts has attracted growing research attention. One research strand focuses on fully automated generation pipelines. Earlier automatic question generation systems largely follow a pre-LLM paradigm spanning rule-based and linguistically driven feature extraction methods and neural network–based representation learning architectures [12, 20]. More recent work has leveraged LLMs for educational question generation based on textbooks, demonstrating greater cross-domain flexibility through prompting [10, 13]. However, LLM-generated questions still fall short in engaging with discipline-specific reading practices and structured integration with pedagogical goals.

A different approach to the problem is to involve instructors in the scaffold generation process. ReadingQuizMaker, for instance, introduces a human-NLP collaborative paradigm, providing instructors with question templates and an NLP toolbox for text summarization and negation [11]. Though the co-design workflow affords instructors greater involvement, the tool's support remains primarily linguistic: it neither models the discipline-specific epistemological characteristics of texts nor encodes pedagogical goals [21], affording instructors limited agency to act as disciplinary designers or to align scaffolds with broader learning objectives.

Across both strands, reading scaffolds generally lack attention to disciplinary literacy practices and structured mechanisms for integrating instructional context and pedagogical goals. Earlier work largely anchors scaffold generation in surface textual features, which limits meaningful engagement with disciplinary reading practices and constrains the integration of instructional context. More recent LLM-based approaches offer greater generative flexibility, yet continue to overlook disciplinary grounding and structured pedagogical integration.

3. System design

To address these gaps, we designed InkSpire, an LLM-powered system that supports the design and generation of context-aware and disciplinarily aligned reading scaffolds (including questions and prompts that provoke thought and guide deeper comprehension and discussion).

To better account for instructional context, InkSpire adopts a human-in-the-loop workflow that actively engages instructors throughout the process. The workflow begins with a Class Profile co-design process, in which instructors contribute their understanding of the learning context and share pedagogical considerations. The verified Class Profile then guides the LLM-driven scaffold generation pipeline, and the workflow concludes with an interactive review interface for instructors to evaluate and refine the generated scaffolds.

To ground the generated scaffolds in discipline-specific reading practices, Goldman et al.'s [1] disciplinary core constructs framework is embedded throughout the workflow and guides each stage of the generation pipeline. During Class Profile construction, the LLM performs a theory-informed analysis to identify the core constructs shaping disciplinary literacy practices, subject to instructor verification. In the scaffold generation pipeline, these constructs play a central

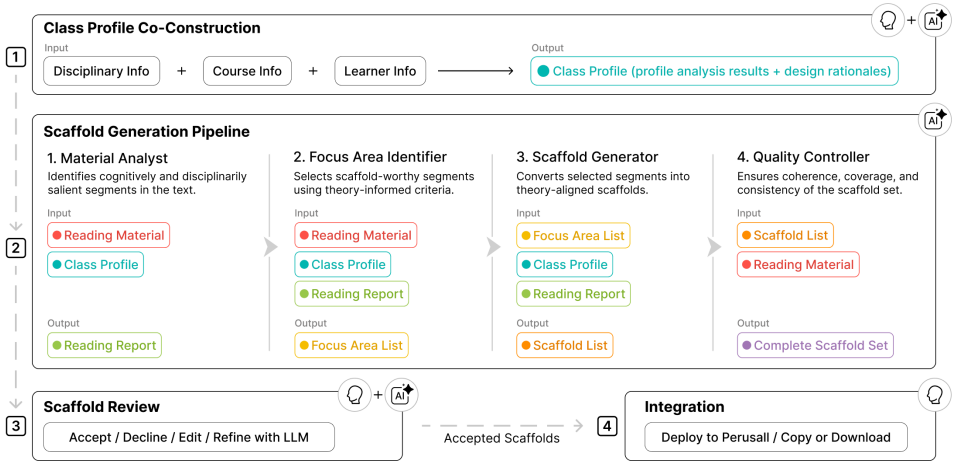

Figure 1: InkSpire System Workflow

role at each stage—guiding the identification of scaffold-worthy text segments and informing the generation of scaffold types responsive to each disciplinary literacy construct.

InkSpire comprises three core components: (1) Class Profile Co-Construction, (2) a Theory-Informed Scaffold Generation Pipeline, and (3) an Interactive Review Layer with integration into collaborative annotation technologies.

3.1. Component 1: Class Profile Co-Construction

The first core component of InkSpire focuses on the co-construction of a comprehensive class profile of the instructional context between the instructor and the system, which serves as the foundation for subsequent scaffold design and generation. Instructors provide inputs at the disciplinary, course, and class levels, which an LLM-assisted analysis processes into a structured interpretation of the instructional context across four levels—disciplinary (how knowledge is read, reasoned about, and validated in the field [1]), course (learning goals, key concepts and terms), learner (prior knowledge, challenges, and suitable scaffold formats), and design considerations (high-priority constraints on what to emphasize or avoid).

To support interpretability and instructor oversight, the system generates an explicit design rationale alongside the profile analysis, enabling instructors to review, validate, and refine the system's interpretation before it is passed to the subsequent scaffold generation pipeline.

3.2. Component 2: Theory-Informed Scaffold Generation Pipeline

The second core component is a theory-informed generation pipeline that takes the verified Class Profile as input and produces disciplinarily aligned and context-sensitive scaffolds.

The pipeline uses LangChain and LangGraph for structured, multi-step workflow management, which offers greater controllability than single-pass prompting. It runs on Gemini 2.5 Flash at temperature 0.3 to prioritize output stability, reproducibility, and faithful preservation of pedagogical theories and intent. The pipeline proceeds through four sequential stages.

The first stage, the **Material Analyst**, is prompted to "analyze the reading material section by section", annotating each section for textual features (content type, cognitive load, and reading patterns) and identifying sections aligned with the instructional context and pedagogical considerations—an identification grounded in a prior step that first analyzes "key ideas, potential student challenges, and instructional opportunities" based on the class profile. It also flags texts related to the core disciplinary literacy constructs [1], including markers of *disciplinary epistemology*, *inquiry practices and strategies of reasoning*, *overarching disciplinary concepts and frameworks*, critical *representation transitions*, and discipline-specific *discourse and language structures*. This stage produces a structured characterization report of the reading, indicating which sections contain scaffold-worthy content.

The second stage, the **Focus Area Identifier**, takes the class profile and the Material Analyst's reading report as input and applies a principled rule set—specified explicitly and rigorously in the agent's system prompt—to identify scaffold-worthy segments. Informed by cognitive load theory[22] and disciplinary literacy framework [1], the rules prioritize segments characterized by high cognitive load, containing or embodying core disciplinary constructs (like discipline-specific *epistemological practices* or *inquiry strategies*), and potential misconceptions, etc. The Key Course Alignment rule further ensures that selected segments align with course and session goals, preserving context-awareness. Each selected segment is assigned a rule-grounded rationale and priority ranking to guide subsequent generation.

The third stage, the **Scaffold Generator**, then maps each flagged scaffold segment and its associated rationale to theoretically grounded scaffold types. Segments with high cognitive load, for instance, are addressed through simplification and clarification prompts or questions, while segments involving *inquiry practices* are translated into prompts that encourage discipline-specific ways of thinking or inquiry, such as interpreting evidence or constructing explanations in the manner of a discipline expert. The generator is further prompted to calibrate language difficulty to the class's proficiency level and, where possible, connect scaffolds to course learning goals and key terms.

The fourth stage, the **Quality Controller**, reviews the complete scaffold set for coherence and coverage before surfacing the outputs for instructor review.

The pipeline ultimately transforms reading materials into pedagogically purposeful scaffolds—e.g., "How does this approach (Productive Failure) challenge traditional notions of 'success' in learning, and what makes this initial failure 'productive' rather than discouraging?"

3.3. Component 3: Interactive Scaffold Review and Integration with Collaborative Annotation Technologies

Finally, the scaffold review and deployment component concludes the system's human-in-the-loop design workflow. The system provides a review interface where instructors can actively engage with generated scaffolds—accepting, declining, editing, or refining them with the LLM.

To reduce the cognitive load associated with the scaffold validation, each scaffold is displayed alongside a reading in which the corresponding source texts are highlighted. Clicking a scaffold card scrolls the reading directly to the highlighted source, enabling efficient cross-referencing between scaffolds and the source texts throughout the review process.

Beyond individual review, the system supports downstream scaffold dissemination through

export functionalities such as copy and download, as well as built-in integration with external collaborative reading and discussion platforms (i.e., Perusall) [23]. The highlighted text segments and their associated scaffolds accepted in the review interface are automatically transformed into instructor-authored annotations anchored to the corresponding texts in Perusall. This integration streamlines scaffold deployment while enabling the scaffolds to serve as interactive entry points for collaborative disciplinary discourse through social annotation [24].

4. Preliminary Pilot Findings

To examine instructor engagement with InkSpire, one of the researchers piloted the system in his own course over 6 weeks across 6 design sessions. 56 scaffolds were generated across 11 readings. Of these, 36 scaffolds were accepted without modification, 15 were accepted after one manual revision, and 3 were accepted after two rounds of revision. Two generated scaffolds were initially rejected but were ultimately accepted. One case appears to be a misclick, as it was immediately followed by an accept on identical content. The other case was rejected because the scaffold misattributed a concept to the wrong paper—two sources appeared within the same source fragment, which likely contributed to the confusion. A subsequent "refine with LLM" attempt produced a similarly flawed result. The instructor then manually corrected the citation and accepted it.

5. Conclusion

In this paper, we presented InkSpire, an LLM-powered system for designing and generating disciplinarily aligned and context-aware, reading scaffolds for digital textbooks. Designing high-quality disciplinary reading scaffolds is traditionally a time-intensive, expert-dependent process, and existing automated systems lack the disciplinary grounding and pedagogical context integration necessary to produce scaffolds that both align with disciplinary reading practices and are responsive to instructional context [10]. In its current form, InkSpire offers a step toward more disciplinarily aligned and pedagogically grounded scaffold authoring.

InkSpire advances the field in two key respects. First, it addresses the limited incorporation of instructional context in prior systems [10, 12] by adopting a structured human-in-the-loop workflow that enables instructors to co-construct a class profile as foundational input to scaffold generation, and to actively review, validate, and iteratively refine generated outputs to ensure alignment with their pedagogical goals. Second, by grounding the generation pipeline in a disciplinary literacy framework [1] rather than structural or syntactic heuristics [11, 12], InkSpire produces scaffolds more genuinely aligned with discipline-specific epistemological and reading practices, better positioning it to foster students' disciplinary literacy.

Pilot findings suggest the system's potential for supporting disciplinarily aligned and context-aware scaffold authoring at scale. Future work will include a more detailed instructor evaluation of the generated scaffolds, testing across different disciplines, alongside a closer look into hallucination rates and efficiency gains (e.g., time saved relative to manual scaffold design). In addition, Perusall interaction data provides opportunities for understanding how disciplinarily aligned and context-aware scaffolds influence students' engagement with disciplinary texts.

Acknowledgments

We thank Yue Zhao for contributing to the early-stage backend development of InkSpire.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-5 to identify and correct grammatical errors, typos, and other writing mistakes. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] S. R. Goldman, M. A. Britt, W. Brown, G. Cribb, M. George, C. Greenleaf, C. D. Lee, C. Shanahan, Project Readi, Disciplinary Literacies and Learning to Read for Understanding: A Conceptual Framework for Disciplinary Literacy, Educational Psychologist 51 (2016) 219–246. URL: http://www.tandfonline.com/doi/full/10.1080/00461520.2016.1168741. doi:10.1080/00461520.2016.1168741.

[2] E. P. Bettinger, B. T. Long, Addressing the needs of underprepared students in higher education: Does college remediation work?, Journal of Human resources 44 (2009) 736–771.

[3] S. Bharuthram, Making a case for the teaching of reading across the curriculum in higher education, South African Journal of Education 32 (2012) 205–214.

[4] H. St Clair-Thompson, A. Graham, S. Marsham, Exploring the Reading Practices of Undergraduate Students, Education Inquiry 9 (2018) 284–298. URL: https://doi.org/10.1080/20004508.2017.1380487. doi:10.1080/20004508.2017.1380487, _eprint: https://doi.org/10.1080/20004508.2017.1380487.

[5] S. L. Pugh, F. Pawan, C. Antommarchi, Academic literacy and the new college learner, in: R. F. Flippo, D. C. Caverly (Eds.), Handbook of college reading and study strategy research, Lawrence Erlbaum Associates Publishers, 2000, pp. 25–42.

[6] C. Shanahan, Disciplinary comprehension, in: S. E. Israel, G. G. Duffy (Eds.), Handbook of research on reading comprehension, Routledge, New York, NY, 2009, pp. 240–260.

[7] M. D. Shepherd, A. Selden, J. Selden, University Students' Reading of Their First-Year Mathematics Textbooks, Mathematical Thinking and Learning 14 (2012) 226–256. URL: https://doi.org/10.1080/10986065.2012.682959. doi:10.1080/10986065.2012.682959, _eprint: https://doi.org/10.1080/10986065.2012.682959.

[8] J. But, P. Brown, D. Smyth, Reading Effectively Across the Disciplines (READ): A Strategy to Improve Student Success, InSight: A Journal of Scholarly Teaching 12 (2017) 30–50. URL: https://insightjournal.park.edu/wp-content/uploads/2020/01/3-But-et-al.pdf. doi:10.46504/12201702bu.

[9] D. Van Camp, W. Van Camp, Using content reading assignments in a psychology course to teach critical reading skills., Journal of the Scholarship of Teaching and Learning 13 (2013) 86–99.

[10] A. Durg, C. Kultur, A. Zhang, J. Savelka, Llm-powered framework for automatic generation of metacognitive scaffolding cues for introductory programming in higher education, in: Proceedings of the Sixth Workshop on Intelligent Textbooks (iTextbooks'25), Palermo, Italy, 2025. ITextbooks'25 Workshop.

[11] X. Lu, S. Fan, J. Houghton, L. Wang, X. Wang, ReadingQuizMaker: A Human-NLP Collaborative System that Supports Instructors to Design High-Quality Reading Quiz Questions, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, ACM, Hamburg Germany, 2023, pp. 1–18. URL: https://dl.acm.org/doi/10.1145/3544548.3580957. doi:10.1145/3544548.3580957.

[12] T. Steuer, Automatic Question Generation to Support Reading Comprehension of Learners - Content Selection, Neural Question Generation, and Educational Evaluation, PhD Thesis, 2023. URL: https://tuprints.ulb.tu-darmstadt.de/handle/tuda/9867. doi:https://doi.org/10.26083/tuprints-00023032.

[13] S. Maity, A. Deroy, S. Sarkar, Exploring the Capabilities of Prompted Large Language Models in Educational and Assessment Applications (2024). URL: https://zenodo.org/doi/10.5281/zenodo.12730013. doi:10.5281/ZENODO.12730013.

[14] P. D. Pearson, N. K. Duke, Comprehension instruction in the primary grades, Comprehension instruction: Research-based best practices (2002) 247–258.

[15] N. NICHD, Report of the national reading panel. teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction, Reports of the subgroups (2000).

[16] T. Shanahan, C. Shanahan, Teaching Disciplinary Literacy to Adolescents: Rethinking Content-Area Literacy, Harvard Educational Review 78 (2008) 40–59,279. URL: https://www.proquest.com/docview/212267873/abstract/6ED7131E82FB42DAPQ/1, num Pages: 21.

[17] T. Shanahan, C. Shanahan, What is disciplinary literacy and why does it matter?, Topics in language disorders 32 (2012) 7–18.

[18] Z. Fang, Scientific literacy: A systemic functional linguistics perspective, Science education 89 (2005) 335–347.

[19] K. Hubbard, Disciplinary literacies in stem: what do undergraduates read, how do they read it, and can we teach scientific reading more effectively?, Higher Education Pedagogies 6 (2021) 41–65.

[20] Y. Song, J. Du, Q. Zheng, Automatic item generation for educational assessments: A systematic literature review, Interactive Learning Environments 33 (2025) 5386–5405. doi:10.1080/10494820.2025.2482588.

[21] C. Hynd-Shanahan, What does it take? the challenge of disciplinary literacy, Journal of Adolescent & Adult Literacy 57 (2013) 93–98.

[22] J. Sweller, J. J. G. van Merrienboer, F. G. W. C. Paas, Cognitive Architecture and Instructional Design, Educational Psychology Review 10 (1998) 251–296. URL: https://doi.org/10.1023/A:1022193728205. doi:10.1023/A:1022193728205.

[23] A. J. Clarke, Perusall: Social learning platform for reading and annotating (perusall LLC, perusall.com), Journal of Political Science Education 17 (2021) 149–154. URL: https://doi.org/10.1080/15512169.2019.1649151. doi:10.1080/15512169.2019.1649151, _eprint: https://doi.org/10.1080/15512169.2019.1649151.

[24] X. Zhu, B. Chen, R. M. Avadhanam, H. Shui, R. Z. Zhang, Reading and connecting: using social annotation in online classes, Information and Learning Sciences 121 (2020) 261–271. URL: https://www.emerald.com/insight/content/doi/10.1108/ILS-04-2020-0117/full/html. doi:10.1108/ILS-04-2020-0117.