



# Harmonizing Assistance: Moderating Visual and Textual Aids in AI-Enhanced Textbook Reading with *IRead*

Xiaoyu Zhang<sup>1,3,4</sup> · Vincent Dörig<sup>2,3</sup> · Peng Cui<sup>2</sup> · Vilém Zouhar<sup>2</sup> · Torbjørn Netland<sup>3</sup> · Mrinmaya Sachan<sup>2</sup>

Received: 19 July 2024 / Accepted: 6 September 2025 / Published online: 11 November 2025  
© The Author(s) 2025

## Abstract

Textbooks continue to be one of primary mediums of learning. Students often need additional support during the process of reading textbooks, leading to several research efforts that aim to increase student engagement and provide tailored experiences in textbook reading. However, providing excessive information beyond the textbook can also distract students from the reading task. When enhancing the reading experience, one has to strike a delicate balance between providing sufficient informational support and maintaining students' focus on textbook reading. Fusing together latest developments in large language models (LLMs) and their applications in education and several pedagogical theories, we design a textbook reading guidance mechanism. We introduce *IRead*, an interactive tool for textbook reading which uses LLMs with visualization and interaction techniques, to enhance students' reading and learning experiences. *IRead* incorporates conceptual visualizations that reflect the textbook's content and features an AI-driven question bot that generates questions in response to student reading and interaction history. We evaluate *IRead* with a between-subject user study and measure the effectiveness of our methodology in supporting the students' reading experience based on the Bloom's Taxonomy and the ARCS model. We collect feedback from participants ranging from undergraduate to doctorate students. The results highlight the effectiveness of simple yet intuitive visualizations, such as the concept tree in *IRead*. We also derive general insights for the development of tools that enhance educational reading experiences.

**Keywords** Textbook reading support · Learning engagement · Question generation · Artificial intelligence · ARCS model

---

Xiaoyu Zhang and Vincent Dörig contributed equally to this work.

---

Extended author information available on the last page of the article

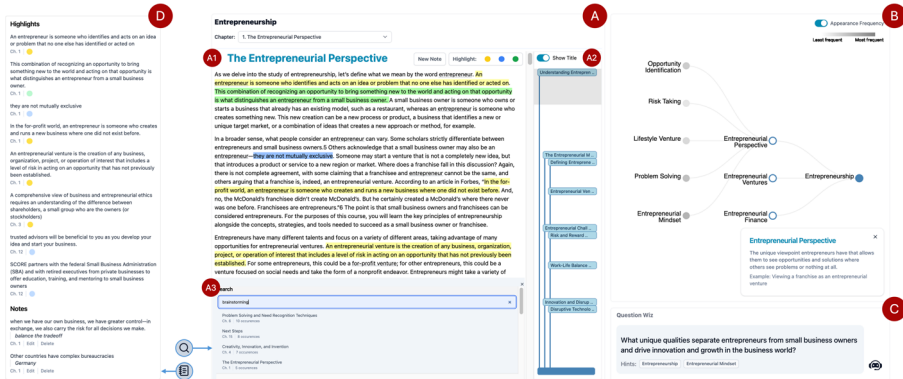
## Introduction

With the latest advances in large language models (LLMs) and the increased focus on Human-AI collaboration, traditional methods of learning and teaching are undergoing significant transformations, introducing new opportunities and challenges to educational settings. The conventional process of textbook reading may also change, given the capabilities of LLMs to provide high-quality summaries (Zhang et al., 2023, 2024; Cohan et al., 2018), retrieve information (Lewis et al., 2020; Zhu et al., 2023), and answer questions (Ko et al., 2022; Guu et al., 2020). Institutions around the world are looking for ways to integrate LLM in teaching and learning, but are also concerned about risks of counterproductive outcomes (Kasneji et al., 2023; Jeon and Lee, 2023; Tlili et al., 2023). However, they also voice concerns that these new methods might completely replace traditional textbook reading practices, which would conflict with the university's objective of fostering comprehensive student development. Motivated by these insights, we explore and design a novel textbook reading tool, *IRead*<sup>1</sup> for higher-education students grounded in pedagogical principles, aiming to enhance rather than distract them from the learning process.

The core challenge of our research lies in maintaining a balance between providing sufficient supplementary information and avoiding distracting readers with an excess of it. Many existing AI-enhanced textbook or document reading tools aim to leverage the full potential of AI to offer maximum support. In contrast, we position our work within the context of higher education, focusing on designing mechanisms to control the volume of visual and textual information and deliver it at the appropriate times for the users. For instance, consider two existing reading products: Apple Books and Kindle. Apple Books offers extensive reading materials with internal and external links, while Kindle provides a more streamlined reading environment. Interestingly, students tend to use Apple Books for leisure reading and Kindle for serious academic reading (Moore, 2009). This dichotomy underscores the challenge of balancing educational objectives with student engagement. From a pedagogical point of view, we aim to encourage students to engage in a deeper reflection and explore the topic independently, though many of them prefer to seek the quickest and easiest path to complete their reading tasks (Bharuthram, 2017; Mizrachi et al., 2018). Therefore, the primary research focus of this work is to determine the right moments and methods to provide an appropriate level of informational support to learners.

Our solutions are grounded on the guided discovery learning theory (Brown, 1992; Brown and Campione, 1994; Palincsar et al., 1989), which balances didactic teaching and unassisted discovery learning (Dewey et al., 1897) by providing learners with carefully designed examples or tasks to help them discover the knowledge themselves. This approach is learner-centered and can foster their autonomy, curiosity, and critical thinking abilities (Bustos, 2020). To implement this approach, we employ questions as a means to guide and stimulate the reading process for students. We follow Keller's ARCS model (Attention, Relevance, Confidence, Satisfaction) (Keller, 1987) to develop a question-triggering mechanism that identifies the right moments

<sup>1</sup> The project webpage and prototype system are available at <https://iread.ethz.ch>,



**Fig. 1** The *IRead* interface comprises three major components: (1) A1 - an enhanced textbook reader with A2 - a minimap for logical overview and A3 - a query function for keyword searching; (2) B - a concept visualization for conceptual overview and interactive exploration of concept definitions; and (3) C - Question Wiz, which presents user interaction-specific questions for active and engaged reading. Users can review their highlighting and note-taking records in D - a collapsible history view

to display the questions and motivate the reader. This mechanism is implemented in an AI-assisted textbook reading tool – *IRead* (Fig. 1). To equip *IRead* with high-quality questions, we train a question generation model on a textbook corpus sourced from OpenStax<sup>2</sup> Cui et al. (2024) and adapt it according to the constructivist learning theory (Matthews, 1998). In particular, the questions are individualized based on reader’s highlighting, note-taking or other interaction behaviors. Guided by these questions, the reader can further interact and read through *IRead*, as well as reviewing their interaction history through a collapsible concept tree anytime they need.

We evaluate *IRead* and the question-triggering mechanism through a between-subject comparison study. The results show that the participants prefer the novel reading experience provided by *IRead*, which also stimulated more diverse thinking compared to the baseline system. The results also indicate that participants with different levels of education, knowledge backgrounds, and reading habits demonstrated diverse preferences for the components of *IRead*. From the study, we see the power of simple yet intuitive visualizations, such as the concept tree in *IRead*, and have observed how various users use it in various ways to serve specific purposes.

## Related Work

### Guided Discovery Learning for Learning Engagement

The problem of driving learning engagement has existed in pedagogy for a long time and has been approached in various ways such as using instructions (Guthrie and Wigfield, 1997) or questions associated with the text (Guthrie et al., 1997). Among

<sup>2</sup><https://openstax.org>

the existing attempts, one widely accepted theory is discovery learning developed by Jerome Bruner, John Dewey et al., where the student is provided with the learning materials rather than the exact answer so that they could find the answer themselves (Dewey et al., 1897). This learning technique also aligns with constructivist-based education principles (Matthews, 1998), where students actively construct their understanding and knowledge by integrating new information with their prior knowledge, rather than passively receiving new information. However, cognitive load theory reveals that “*free exploration of a highly complex environment can overload working memory and hinder learning*” (Kirschner et al., 2006). To balance didactic teaching and unassisted discovery learning, researchers have developed the guided discovery learning approach (Brown, 1992; Brown and Campione, 1994; Palincsar et al., 1989). In practice, such a guided discovery learning experience can be achieved through inquiry-based learning (Bruner, 1961), which fosters a deeper understanding of the material with reduced cognitive load. In this work, we facilitate discovery learning and enhance student engagement by developing an interactive textbook exploration tool. This tool generates questions based on the reader’s reading history, but does not provide answers. We combine teacher expertise with LLMs to develop a facilitator that guides students through their learning adventures. This approach encourages students to seek answers through further exploration of the textbook, resulting in iterative rounds of discovery and constructivist-based learning.

## Reading Facilitation Techniques

Enhancing the digital reading experience has long been explored by researchers in pedagogy and computer science. Before the advent of artificial intelligence (AI), various methods were developed to individually improve the reading experience, commonly measured by reading comprehension performance. One of the simplest methods is note taking, in which readers take independent notes and summarize the texts they read, demonstrating significant benefits (Rahmani and Sadeghi, 2011; Özçakmak, 2019). Another popular technique is manual text highlighting, which has also been effectively implemented with digital tools to improve the reading experience (Leroy and Kammerer, 2023).

In recent years, the surge of AI has created more possibilities to enhance the reading experience. Natural language processing (NLP) techniques now enable automatic highlighting of extracted keywords or concepts and displaying their definitions for readers (Zhang et al., 2021, 2023). Vision-language models can automatically enhance textbooks by matching images from the Web with textbook content (Singh et al., 2023). Beyond AI solutions alone, studies have shown that instructors seek high-quality human-AI teaming approaches to improve their teaching efficiency (Lu et al., 2023; Wang et al., 2022) and support students in reading (Wang et al., 2022). In response to this, Ruan et al. developed BookBuddy, which converts reading materials into interactive conversation-based tutors (Ruan et al., 2019). VR and AR technologies have also been applied to create immersive visual interactions for children with storybooks (Wang et al., 2024). However, many of these solutions target entertainment reading for children, leaving a gap in addressing the balance between engagement and learning outcomes in serious reading context. Although some approaches

incorporate question-answer practices into reading (Johnson et al., 2024; Willis et al., 2019; Ruan et al., 2019), few explore the integration of additional visual and textual aids into an interactive textbook reading tool.

As outlined in Table 1, *IRead* stands out against other reading facilitation tools due to its textbook format that pairs visualizations with an AI-driven question bot. In particular, *IRead* employs concept tree visualizations, enabling learners to explore connections between ideas and have the ability to generate contextual questions, fostering deeper engagement. Compared to tools such as QuizBot (Ruan et al., 2019), which focuses on active recall through dialogue, or ReadingQuizMaker (Lu et al., 2023), aimed at generating quiz questions for educators, *IRead* uniquely addresses the need for a cohesive reading tool that improves comprehension without distracting from the primary text. This contributes to a more supportive learning environment, ultimately enhancing the student's ability to grasp complex materials effectively.

### NLP Techniques for Question Generation

NLP techniques have been widely used to support learning and teaching. For example, they have been used to create educational content (Sarsa et al., 2022), simulate student behaviors (Markel et al., 2023), and offer adaptive feedback (Liu et al., 2020). Among them, question generation, the technique to automatically create relevant questions in a given context, has great potential to facilitate active learning (Kurdi et al., 2020). Existing question generation approaches typically adopt language models, such as BART (Lewis et al., 2020) and T5 (Raffel et al., 2020), to create questions in an end-to-end manner. However, question generation systems in the NLP community mainly focus on the relevance and fluency of questions while overlooking their broader educational implications, such as the potential to foster deeper engagement with or critical thinking about the material (Zhang et al., 2022). As a result, recent efforts incorporate pedagogical principles into system design and demonstrate their effectiveness in various scenarios, such as math problem solving (Jiao et al., 2023) and language learning (Cui and Sachan, 2023). In addition to end-to-end models, Wang et al. (2022); Lu et al. (2023) emphasize the iterative process of question creation and collect a suite of NLP tools for Human-NLP collaborative question generation, including text summarization, paraphrasing, and negation transformation. However, the questions of most existing work are designed for assessment or practice purposes. In contrast, the question generation model in *IRead* is trained to mimic “*what an instructor would ask students during the reading*”, aiming to create an interactive reading experience.

### ARCS Model

The ARCS model is a motivational design model that provides a systematic framework to answer “*how to create instruction that would stimulate the motivation to learn*” (Keller, 1987). It extends the two motivational categories for engagement, the satisfaction of personal needs (the value aspect) and the positive expectancy for success (the expectancy aspect), in the expectancy-value theory (Tolman, 1932; Lewin, 1968), to four categories: (1) *Attention*—the extent to which students' interest and

**Table 1** Comparison of *IRead* against related tools

Tool	Target Audience	Main Purpose	User Interaction	Key Features	Educational Theory	Evaluation Methods	Key Findings
IRead	Students	Enhance the textbook reading experience by balancing supplementary info without distraction	Interactive textbook interface with visualizations and an AI-driven question bot	Concept tree visualization, AI-driven questions based interactions with the tool	Bloom's Taxonomy, ARCS model, active reading and guided discovery learning	Between-subject user study	Enhanced student engagement and learning outcomes, insights into balancing AI support and minimizing distractions
QuizBot (Lu et al., 2023)	Students	Learn factual knowledge through dialogue with an AI tutoring agent	Dialogue-based Q&A with active recall and adaptive feedback	Personalized feedback and explanations, adaptive questioning and conversational avatar for engagement	Testing effect via active recall and recognition	Two within-subject studies with students	Over 20% higher recall and recognition compared to flashcards, increased engagement
Reading QuizMaker (Lu et al., 2023)	Instructors and educators	Help design high-quality quiz questions to enhance student's comprehension	Interactive interface with tools to generate and adopt quiz questions	Authoring panel with detailed controls and question previews	Active reading and learning	Formative and evaluative study with instructors	Question generation comparable to manual creation, preference for human-AI teaming over automatic generation, users desire control and flexibility
Metabook (Wang et al., 2024)	Children and teachers	Automatically generate AR storybooks to improve reading engagement	Upload books to generate AR experiences with 3D avatars as reading companions	Story-to-3D book generation pipeline and 3D avatar companion for interactive reading	Teacher feedback, learning enthusiasm	Formative study with adults and user studies with children	Increased interest in reading and vocabulary
Story-Buddy (Zhang et al., 2022)	Parents and children	Enhance interactive storytelling experiences with AI-assisted question answering	Parent-AI co-reading in storytelling sessions	AI-generated questions, follow-up questions and child progress tracking	Dialogic reading, parental assessment	Formative user study with parents and children	Effective in engaging children, reducing parental cognitive load, supporting diverse storytelling practices and promoting language development.
BOOKR Class <sup>a</sup>	Children	Enhance children's reading skills, comprehension and foster a habit of reading from an early age	Interactive e-books that include animations, narration, and educational games	Multimedia combining narration, illustrations and educational games	Dual coding, active learning, gamification	N/A	Better reading comprehension compared to traditional print books, higher recall facilitated by animated illustrations

<sup>a</sup> <https://bookrclass.com>

curiosity is stimulated by the lesson; (2) *Relevance*—the materials’ and instructions’ relatedness with students’ goals during the learning process; (3) *Confidence*—students’ perceived competence to achieve their educational goals; (4) *Satisfaction*—the extent to which students are satisfied with the course and their achievements. The author of the ARCS model also provides detailed subcategories and sample motivational strategies for each category in Keller (1987). Guided by this model, following researchers created survey questions to measure the various dimensions of the motivation effects. It has been proved that the ARCS model is not only applicable to traditional classroom (Feng and Tuan, 2005), but also to online learning (Huett et al., 2008) and even blended learning environments (Chang and Chen, 2015). Furthermore, the model has been adapted to incorporate elements of digital literacy and professional competency (Maiti et al., 2023), leveraging AI technologies to create more personalized and motivating learning experiences (Wu and Liou, 2024). Therefore, the ARCS model is a good fit for our project to motivate the readers in the digital reading context.

## Methods

### Design Considerations

The design of *IRead* was informed by a preliminary study presented in Cui et al. (2024). In this study, participants were asked to read textbook articles with additional questions displayed in a side panel (see experiment interface in Appendix A). The results indicate that intriguing and informative questions can enhance engagement and stimulate thoughts, but participants disliked questions that were too easy or had immediate answers in the article. Moreover, participants’ criteria for good questions varied based on the reading scenario, e.g., first-time reading vs. content reviewing, learning scenarios vs. examination scenarios, and serious learning vs. casual reading. The study also identified a “cold start” phenomenon, where readers need more support at the beginning to grasp the context of an article. Based on these insights, we summarize the following requirements for a comprehensive reading assistance tool:

**R1 Overview:** The tool should provide a high-level overview of the content and structure of the textbook. This overview can take the form of a textual summary detailing the article content or a graphical diagram illustrating the logical structure, with the latter being the preferred choice for the majority of participants.

**R2 Comprehension Support:** The tool should provide enhanced reading support compared to a standard PDF reader. In addition to conventional features such as highlighting, taking notes and searching for content, the tool should also provide supplementary information on terms that may impede the reading process. This additional information may include definitions, examples, external links, and illustrative figures.

**R3 Intelligent Guidance:** The tool should aptly navigate the student’s reading process, tailoring guidance to their individual reading pace. An intelligent chatbot can be used to offer help when readers face challenges such as cold start, reading barriers, or reading fatigue. The questions posed by the chatbot should be intelligently positioned, with an appropriate frequency, and of suitable complexity, ensuring that

they enhance the reading process rather than becoming a distraction.

**R4 Customized Information Inquire:** The readers should have control over the extent of additional information they wish to access, catering to their individual reading preferences in various scenarios, such as first-time reading, in-depth reading, serious reading, and casual reading. This includes the ability to control the visibility of the information outlined in **R1** to **R3**.

To fulfill the aforementioned requirements, we turn to the guided discovery learning theory (Brown, 1992; Brown and Campione, 1994; Palincsar et al., 1989) for providing a learner-centered reading experience. This theory emphasizes a balance between didactic teaching and unassisted discovery learning (Dewey et al., 1897), which could be implemented with questions that guide learners through the reading process. In order to determine the right moments to display these questions and motivate the reading process, we turn to another theory—Keller’s ARCS (Attention, Relevance, Confidence, and Satisfaction) model (Keller, 1987)—to develop a question-triggering mechanism that could provide context-sensitive information to motivate the reading process. We connected the four dimensions of the ARCS model with five typical types of interactions in reading, designing specific question-triggering mechanisms for each interaction type. Table 2 outlines these five interaction types, their triggering mechanisms, the presentation of questions, and the corresponding ARCS principles. Moreover, we followed the constructivist learning theory (Matthews, 1998) to individualize questions for each learner based on their highlighting, note-taking, and other interaction behaviors, as well as track their interaction history. Guided by these questions, readers will further read and get more individualized questions, accumulating knowledge through their unique learning paths.

### Textbook Data Processing Workflow

The content presented in *IRead* is generated by the textbook data processing workflow illustrated in Fig. 2. This workflow is a one-time setup for every newly imported textbook. It takes the textbook chapters as input, and leverages an LLM to generate a structured overview of the main ideas (concept hierarchy) (Section “[Concept Tree Generation](#)”) and the table of contents for each chapter. It also feeds the textbook into a pre-trained question generation model to generate questions and fun facts about the book content (Section “[Question Generation Model](#)”).

### Question Generation Model

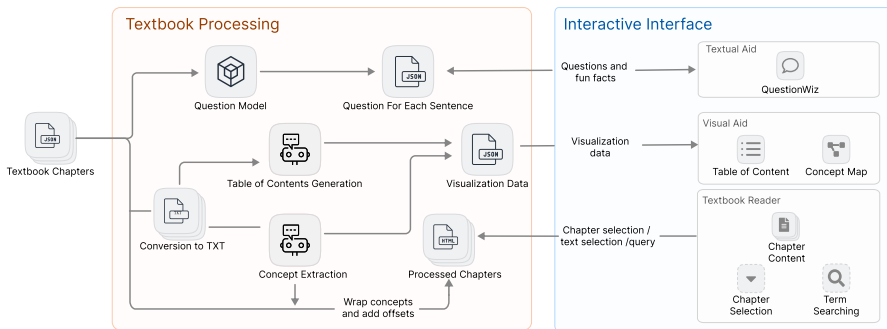
The core idea behind *IRead* is a question-driven active reading mechanism. Our goal is to use questions as a communicative and interactive tool to maintain reader interest and attention and guide them through the reading process. However, existing question generation (QG) systems are not adequate for our needs because:

- Question generation models in the NLP community mainly focus on factual questions that lack educational value (Wang et al., 2022; Zhang et al., 2022).
- Most questions are generated for evaluation purposes *after* reading, and it remains unclear how to use them in an interactive scenario *during* reading, such as

**Table 2** An overview of five primary interaction types in *IRead*, their triggering operations, and corresponding questions generated by Question Wiz

Interaction	Trigger	Question Wiz	Prompt for Customization	ARCS
Highlighting	Select and highlight a piece of text	A question related to the selected sentence with a list of relevant concepts below	<i>Paraphrase the following question such that it is intriguing, informative and stimulates thought and reflection: {original_question}</i>	Relevance
Note-Taking	Create a note for a selected piece of text	A question related to both the selected sentence and the note, with a list of relevant concepts below	<i>Generate a question regarding "{question}" based on the emotion/purpose indicated in {note}:</i>	Relevance
Concept Query	Click on an underlined concept in the reader	A question related to the concept with a list of hint concepts below	<i>Ask a question about the potential application of {concept} in the context of {chapter_name}. The question should be open-ended and thought-provoking. Return the question without any additional information.</i>	Attention Relevance
Term Searching	Search for a term, click on a search result to go to a different chapter and waits 5 seconds	A progress reminder of prior chapter reading, with a button to return to the prior chapter	-	Attention Relevance Satisfaction
Inactivity	There has been no activity on the page for two minutes	A fun fact about the current chapter that doesn't demand problem-solving or deep cognitive processing	<i>Generate a fun fact for a college student reading a textbook called "{book_title}" in a chapter titled "{chapter_title}" and containing the following concepts: {concepts}. Simply return the fact without anything else.</i>	Attention

These questions align with one or more ARCS principle dimensions and are further customized using an LLM



**Fig. 2** Textbook data processing and interaction workflow

when to present them to the reader.

To address these limitations, we have developed a question generation model considering three perspectives:

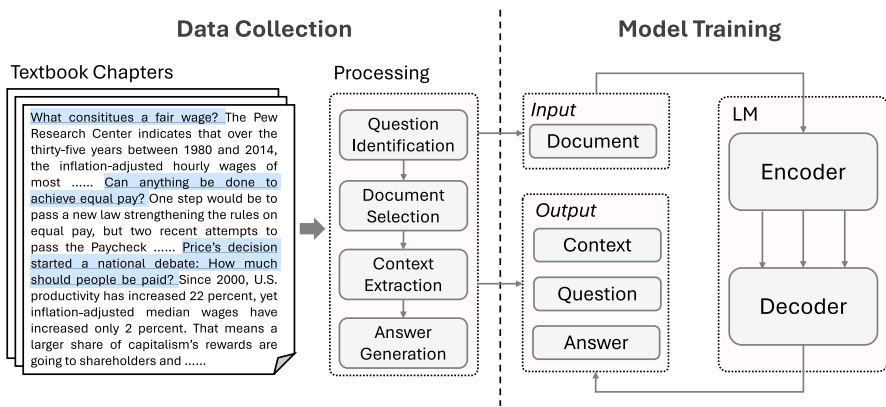
**From the data perspective**, we build our question generation model on a textbook corpus collected from OpenStax. The dataset consists of textbook chapters in various disciplines, such as business and social sciences. In particular, instead of using assessment questions, we collect questions embedded in the content of the chapter. These in-text questions are designed by expert educators to provoke readers' thoughts *during* reading and learning. Meanwhile, they perform a variety of interactional functions to engage readers (Hyland, 2002), therefore, are the ideal choice to build our model.

**From the model design perspective**, traditional QG models in NLP primarily focus on generating the content of questions, and optionally, their answers. However, effective questions should not only be relevant in content but also introduced at appropriate moments that align with the discourse structure and progression of the document (Cui et al., 2024). To address this, we incorporate question position into the learning objective, ensuring that each generated question is tied to a specific point in the source text—i.e., where it naturally fits within the flow of reading. This better supports our goal of enhancing the overall reading experience.

**From the interaction perspective**, we modify the configuration of the question generation model to adapt to our interaction mechanism (per Section “**Interaction**”). In particular, we generate a baseline question for every sentence. In doing so, we ensure that each sentence is associated with a relevant question that can be evoked upon the user's interaction. In cases where a sentence is less important, the corresponding question may stem from a more significant surrounding sentence. Building upon these base questions, we further employ ChatGPT to integrate user input and customize the questions displayed based on their specific interactions.

As illustrated in Fig. 3, we implement our question generation model through data collection and model training. Specifically, we begin by extracting questions from OpenStax textbook chapters and selecting documents that contain more than three questions, as this indicates frequent use of questions and suggests their pedagogical importance. This results in 621 documents and 3,593 questions in total. For each question, we extract the preceding sentence to serve as its positional and contextual reference. Additionally, we employ ChatGPT to generate answers to these questions based on the chapter content. Based on the collected dataset, we fine-tune Flan-T5, an encoder-decoder based language model. We use chapter content as input, and train the model to generate a concatenation of the contextual sentence, the corresponding question, and its answer as output. During inference, the contextual sentence can be specified as the prefix to the output, enabling control over the sentence on which the generation should be conditioned.

The question generation model was evaluated with both automatic and human evaluations, both of which demonstrated strong performance. For automatic evaluation, we used reference-based metrics (i.e., textual similarity between the generation and human-written question), achieving scores of 18.3 Rouge-L (Lin, 2004), 20.7 Meteor (Banerjee and Lavie, 2005), and 84.5 BertScore (Zhang et al., 2019),



**Fig. 3** An illustration of our question generation model. We construct our training dataset from a textbook chapter corpus (left), and train a Flan-T5 model to generate questions, their corresponding answers, and the contextual positions within the chapter where these questions should arise (right)

outperforming zero-shot GPT-4, which scored 15.7, 18.9, and 84.3, respectively. For human evaluation, we invited 15 annotators to evaluate 32 model-generated questions from five textbook chapters, and another 15 annotators to evaluate 32 reference questions written by the textbook authors from the same chapters. The evaluation criteria include relevance to the context, appropriate positioning without distraction, and importance to the central topic. On a scale of 1 to 5, the model-generated questions received scores of 4.2, 4.3, and 4.2 for the three criteria, respectively. These scores are comparable to those for the reference questions, which received 4.0, 4.2, and 4.4. The participants' final ratings were averaged to mitigate individual variability. We measured the consistency of annotator judgments using Krippendorff's Alpha, which yielded a moderate consistency score of 0.66. Given the subjective nature of the task, we consider this level of consistency acceptable.

### Concept Tree Generation

In addition to the questions, *IRead* includes a concept visualization feature to visually illustrate key concepts in the textbook and their hierarchical relationships. To generate the concept tree data for visualization, we leverage the capabilities of LLMs, which excel in contextual comprehension and information synthesis. These models can also efficiently handle long texts due to their large context window sizes, allowing *IRead* to process an entire textbook chapter at once.

In practice, we instruct the LLM to identify key concepts and their related subconcepts, along with their respective children, to structure the content hierarchically. We also request concept definitions and examples for inclusion in information cards and determine each concept's frequency within the chapter to illustrate its prevalence in the visualization. The resultant data are formatted as a nested JSON object.

We iterated through various versions of the prompt to identify one that consistently produces the desired data when applied to texts of varying lengths and domains. The prompt is provided in Appendix B-Listing 1, utilizing the GPT-4

(| gpt-4-turbo-preview|) model, which stands as OpenAI’s latest and most advanced model at the time of development. After autogeneration, we employ human oversight to ensure the balance of the tree structure and the quality of the concept, which are difficult to consistently manage through the prompt alone. In addition, we offer an interactive visual interface that allows authors and instructors to further edit the concepts. This method of automated concept tree generation, plus minimal post-inspection, is efficient as it eliminates the need for additional inputs beyond the text itself while maintaining the concept tree quality. It is also cost-effective as it only needs to be done once per chapter.

## Concept Mapping

To bridge concept visualization with textbook content, we implement a fuzzy concept mapping technique that aligns the concept tree with the text. Instead of restricting the mapping to exact matches, our approach aims to identify words or phrases closely related to the concepts, even with slight variations. To achieve this, we devise a method that utilizes synonyms generated by the LLM to pinpoint specific terms corresponding to each concept. We employ word embeddings that map words or sentences into a multi-dimensional space, using cosine similarity through a sentence-transformer model. This allows us to assess how well these terms correlate with the concept within a sentence. Terms exceeding a predefined similarity threshold are highlighted as clickable links and associated with a concept in the visualization within *IRead*. For a detailed implementation overview, refer to Algorithm 1.

### Algorithm 1 Naïve Concept Mapping with Semantic Similarity

---

```

1: function MAPCONCEPTS(text, threshold)
2:   concepts  $\leftarrow$  EXTRACTCONCEPTSWITHLLM(text)
3:   sentences  $\leftarrow$  SPLITTEXT(text)                                 $\triangleright$  split textbook content into sentences
4:   for all sentence  $\in$  sentences do
5:     for all word  $\in$  sentence do
6:       if word  $\in$  concepts then
7:         MARKASCONCEPT(word, concept)
8:         break
9:       end if
10:       $\vec{w} \leftarrow$  ENCODE(word, sentence)                             $\triangleright$  context-aware word embedding
11:      for all concept  $\in$  concepts do
12:         $\vec{c} \leftarrow$  ENCODE(concept, sentence)
13:        if  $\frac{\vec{w} \cdot \vec{c}}{\|\vec{w}\| \cdot \|\vec{c}\|} \geq \textit{threshold}$  then                         $\triangleright$  cosine similarity threshold
14:          MARKASCONCEPT(word, concept)
15:        end if
16:      end for
17:    end for
18:  end for
19: end function

```

---

## Interface Design

### Visual Interface

To address the challenges discussed in Section “Introduction” and meet the design requirements outlined in Section “Design Considerations”, we design a visual interface (Section “Visual Interface”) and an interaction mechanism (Section “Interaction”) grounded in multiple pedagogical theories, as illustrated in Fig. 4. The visual interface and interaction mechanisms are coordinated to create a cohesive, interactive learning environment, personalized for each user’s exploration of textbook content.

The *IRead* interface consists of three major components: an enhanced textbook reader (Fig. 1-A), a concept visualization (Fig. 1-B), and Question Wiz (Fig. 1-C). These components, along with their subcomponents, are designed to provide moderate visual and textual aids for users.

**Enhanced Textbook Reader** To provide a fundamental reading experience similar to typical PDF readers such as Acrobat, *IRead* includes annotative features to support active reading strategies. Users can highlight sentences in three different colors and add notes to selected text, which are automatically saved for future reference (R2). Furthermore, a sidebar (Fig. 1-D), which can be displayed or hidden, organizes and displays the saved highlights and notes, allowing users to review and edit them as needed. To provide a linear overview and help the reader navigate the text (R1), we include a minimap on the right side of the text, serving as a table of contents (Fig. 1-A2). Clicking on the section titles takes the user directly to that section. Additionally, a search function allows users to perform queries across different chapters (Fig. 1-A3) and see how the same term is addressed in different contexts (R4).

**Concept Visualization** Following the “overview first, details on demand” mantra (Shneiderman, 2003), this component graphically represents the concepts in the textbook and their relationships within a hierarchical tree structure (R1). Concepts on the left side of the graph have higher-level and more abstract semantic meanings compared

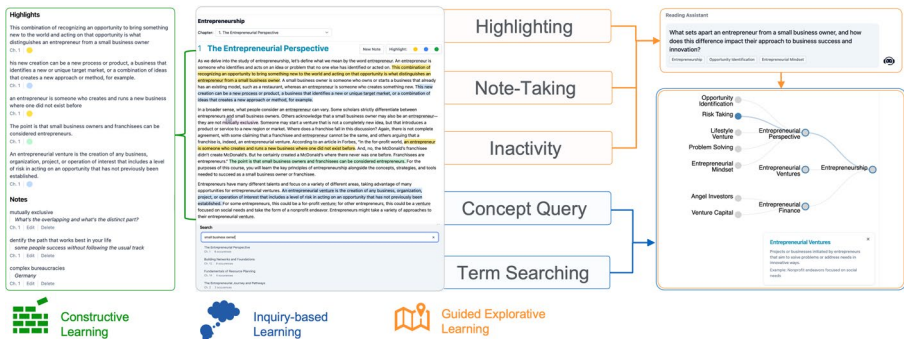


Fig. 4 The five different interaction mechanisms and corresponding implementation in *IRead*

to those on the right side. When users hover over a concept, an information card appears with a definition and examples to aid comprehension (R2). Clicking on a node dynamically adjusts the tree to reveal the direct children of the selected concept. By expanding or collapsing the concept nodes, users can navigate the tree, promoting an engaging and exploratory way of learning (R4). These concepts are also visually distinguished from other text with dotted underlines in the textbook reader. When a user clicks on such a term, the visualization of the concept updates to highlight this concept, revealing its information card and showing its family path in the tree. Meanwhile, Question Wiz will pose a relevant question, further engaging the user with the material.

**Question Wiz** Located in the bottom right corner of the interface is the “Question Wiz”, *IRead*’s reading assistant designed to show feedback from the various interaction mechanisms presented in Section “Textbook Data Processing Workflow” (R3). When a question is posed, the assistant not only presents the question but also offers basic hints in the form of related concepts to aid in formulating an answer. Users can interact with these hints by clicking on them, which highlights the respective concept in the visualization, strategically positioned right above to facilitate quick cross-referencing. Additionally, when a user moves to a new chapter via search, the assistant will offer to return them to the last reading position from the previous chapter.

**Concept Mapping in the Text** (Figure 1-A1). In *IRead*, the user will see that specific words are visually distinguished from others by being underlined with a dotted line. These denote concepts and clicking such a term triggers two responses: it highlights the concept in the visualization and prompts the question bot to pose a relevant question, further engaging the user with the material. Concept mapping aids in comprehension and allows cross-referencing between textbook content and the visualization (R2).

## Interaction

*IRead* provides an improved textbook reader tool that supports highlighting and note-taking, similar to a conventional PDF reader, and stores these user-specific records in a history view for later review (in line with constructivist learning principles). Moreover, users can query critical concepts that they find challenging or search for keywords in multiple chapters of the textbook. *IRead* provides definitions and examples related to the concepts or presents the context of the keyword in all chapters for further understanding (aligning with inquiry-based learning principles). Furthermore, when users highlight, take notes, search concepts or pause reading with *IRead*, the corresponding questions are triggered to encourage them to continue reading and explore concepts they are interested in (in accordance with the principles of guided exploration learning).

## User Study

We have designed and conducted a between-subject comparison study to explore our target users’ preferences and assess the effectiveness of *IRead* in enhancing their motivation and learning dynamics during textbook reading. Through this user study,

we aim to understand the following three research questions by gathering questionnaire responses and observing participants' behavioral patterns:

**Q1:** Can questions corresponding to various interactions effectively motivate users in reading and reflecting on the content? If so, what are the optimal content, frequency, and display timing for these questions?

**Q2:** Can the visualization we have designed aid in knowledge exploration? If so, what is the ideal level of detail and visual representation?

**Q3:** Can *IRead*, equipped with suitable prompted questions and visualization support, engage users in textbook reading and facilitate deeper content understanding? In addition to these three research questions, we also aim to assess how well *IRead* fulfills the design requirements outlined in Section “[Design Considerations](#)”. Furthermore, we seek to gather insights into users' feedback on the usability of our system and potential enhancements.

## Participants

We recruited 30 participants for the user study: 15 for the experiment group and 15 for the control group. All participants were current or past university students who read textbooks in their courses. They had good English reading and writing abilities and no mental, cognitive, or severe sight disabilities or impairments. On average, they read about 12 pages of text per day. The participants included 13 females and 17 males with educational backgrounds in computer science, management, biology, economics, and other fields. Among them were 2 undergraduate students, 16 Master's students, 11 doctoral students, and 1 postdoc researcher.

## Setup and Procedure

To involve participants globally and simulate their typical learning environments, we offered the option for them to participate either remotely or in person. Tutorials for *IRead* were shared with participants at least 24 hours before the online study session began. We set up a remote server for remote participants to access *IRead* and join the study with their own machine and external devices. To understand their cognitive process when interacting with *IRead*, participants were instructed to follow the think-aloud protocol (Lewis, 1982) to continuously express their thoughts, reflecting on their understanding, strategies, and any difficulties they encountered. Their audio and screen activities were also recorded during the task. The participants were rewarded with a voucher corresponding to ~\$27/hr.

To understand the effects of the visual and textual aids in *IRead*, we design the user study as a between-subject comparison study. Participants in the experiment group use *IRead* for the study, while those in the control group use a baseline variation with only the PDF reader and annotation functions (i.e., A1 panel in Fig. 1). Informed by two pilot studies, we structured the study with four parts: a tutorial session, two task sessions, and an experience survey and interview session. In the tutorial session, we demonstrated the usage of the reading tool and asked the participants to perform a few mini-tasks themselves. Afterwards, we ask them to perform the two tasks below involving textbook reading and content question answering:

**T1 Textbook Reading.** Participants need to read the first half of a textbook chapter titled “The Entrepreneurial Perspective”, available as an open-source resource on OpenStax. They are required to complete the reading within 25-30 minutes and were given the freedom to choose their preferred method of utilizing the assigned reading tool based on their individual reading habits.

**T2 Content Question Answering.** Participants answer six open-ended questions regarding the textbook content within a time frame of 15-20 minutes. We formulate the questions corresponding to the six levels of Bloom’s taxonomy (Bloom et al., 1956) to assess various depths of understanding of the participants. They are encouraged to refer to *IRead* as needed, simulating an open-book examination scenario. After the two tasks, we ask the participants to respond to an experience survey adapted from the standard motivation questionnaires of ARCS (Keller, 1987; Ma and Lee, 2021), which examines Attention, Relevance, Confidence, and Satisfaction individually. We also interview them promptly based on their responses to specific questions to better understand their learning experience, the reasons behind their actions, and the usability limitations of *IRead*. Additionally, participants were invited to respond to the following three questions either orally during the interview or in writing on the questionnaire.

- What is the function(s) that you found the most helpful in *IRead*?
- What are your most significant challenges when reading the textbook and answering the content questions?
- What types of additional information, visual clues, or interactions do you believe would be helpful in addressing these challenges?

## Data Collection and Analysis

We collected three types of data from the study and analyzed them using different approaches. Specifically, the data and corresponding analysis include:

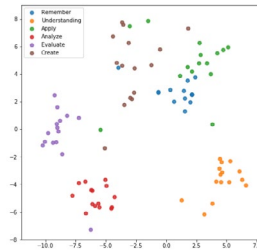
**Participants’ responses to the six open-ended content questions:** We converted them from unstructured text into word embeddings with an open-sourced sentence transformer<sup>3</sup> and calculated the average in-cluster answer distance,  $d_{ans} = \frac{1}{n} \sum_{i=1}^n \|v_i - \frac{1}{n} \sum_{j=1}^n v_j\|$ , for both the control and experimental groups (see results in Fig 5(a)). The in-cluster answer distance could reflect the variability of participants’ responses to the same question, thus indicating the level of inspiration in their thoughts when using *IRead* or the baseline system. We also applied a dimensionality reduction algorithm to reduce the embeddings to 2D and visualize them in Fig. 5(b) and (c), which facilitates a more intuitive comparison.

**Participants’ responses to the 5-point ARCS-model guided experience survey:** We conducted a pairwise t-test to assess significant differences between the control and experimental groups. We also calculated the Pearson correlation coefficient ( $r$ ) to examine the relationship between participants’ education levels and their ratings of perceived question quality and distraction levels.

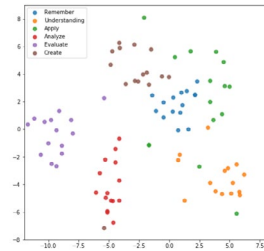
<sup>3</sup><https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>

	Control Group	Experiment Group
Q1: Remember	0.37	<b>0.49</b>
Q2: Understanding	0.60	<b>0.62</b>
Q3: Apply	0.67	<b>0.76</b>
Q4: Analyze	0.54	<b>0.59</b>
Q5: Evaluate	0.61	<b>0.66</b>
Q6: Create	<b>0.72</b>	0.67

(a) Average In-Cluster Distance



(b) Control Group



(c) Experiment Group

**Fig. 5** The average in-cluster answer distance indicates that participants using *IRead* (experiment group) were more inspired in their thoughts and thus provided more varied answers compared to those using the baseline system (control group) for 5 out of the 6 content-related questions designed according to Bloom's Taxonomy. The 2D projections of the word embeddings of participants' answers also align with this insight

***Participants' oral feedback gathered through think-aloud protocols and post-survey interviews:*** We converted the audio recordings from the think-aloud sessions and short post-survey interviews into transcripts and conducted a thematic analysis of the transcripts. One co-author annotated the data, while another verified the results. We distilled insightful observations and explanations for the participant behaviors we observed during the study.

## Results

The statistical results show that *IRead* outperforms the baseline system in 8 out of 14 measured dimensions, with 5 achieving statistical significance (see Table 3). Though no significant difference is detected, we observe a potential trend of *IRead* to reduce participants' confidence in understanding the content and can sometimes be distracting. We provide further discussion and explanations based on observations and interviews from the study as follows.

**Question is Not Just Question** Although the questions in *IRead* were originally intended to engage the reader, the study shows that participants use them for various purposes, including but not limited to previewing upcoming content, facilitating comprehension, and reviewing sections they have completed. For instance, P1 mentions that “*sometimes it takes the context of the paragraph that's coming next...so you know what's the main point of the paragraph, and so you can read it even faster*”. P12 creatively utilizes the hints of the questions to “*pull up the concepts they want to check*”. In addition, many participants find the questions regarding their highlighted content to be “*good as a recap*” (P3, P7) and “*trigger reflection*” (P4, P10, P14). As P3 commented, “*It satisfies in a way that you feel like you learn double because it's not just reading a text... I see where content is connected also with the question. It keeps me thinking about that and keeps me summarizing. It's like, always a revision. It's not just reading words, and then you forget. You continuously being in contact and remind of what you just read*”. In general, the prompt questions perform a positive role in inspiring readers' thoughts and improving learning effectiveness, which is

**Table 3** Comparison between experiment group with IRead and a control group with the baseline PDF reader in IRead

Statement	Condition	N	M	SD	P	Agreement to the Statement (1 - 5)	
Attention	S1: I feel rather disappointed with this tool.	Baseline	15	2.27	0.68	0.1413	
	IRead	15	<b>1.87</b>	0.72			
	S2: This tool has very little in it that captures my attention.	Baseline	15	2.93	0.77	<b>0.0086**</b>	
	IRead	15	<b>2.0</b>	0.97			
S3: I am often distracted while reading with this tool.	Baseline	15	2.0	0.89	0.2226		
IRead	15	2.4	0.8				
S4: I do NOT think I will benefit much from this tool.	Baseline	15	2.73	0.77	0.1081		
IRead	15	<b>2.07</b>	1.29				
Relevance	S5: The reading support provided by this tool helps me address my confusion.	Baseline	15	3.4	0.95	0.3378	
	IRead	15	3.07	0.85			
Confidence	S6: The things I read with this tool will be useful to me.	Baseline	15	3.4	0.61	<b>0.0038**</b>	
	IRead	15	<b>4.0</b>	0.37			
Satisfaction	S7: As I read with this tool, I believe I can understand all textbook content if I try hard.	Baseline	15	3.93	1.0	0.7476	
	IRead	15	3.8	1.17			
	S8: I feel confident that I will do well in the question answering section.	Baseline	15	3.87	0.5	0.4615	
	IRead	15	3.67	0.87			
Satisfaction	S9: I find the tasks of reading and answering questions to be manageable for me.	Baseline	15	3.93	0.44	0.096	
	IRead	15	<b>4.27</b>	0.57			
	S10: I enjoy reading with this tool.	Baseline	15	3.67	0.47	<b>0.0457*</b>	
	IRead	15	<b>4.0</b>	0.37			
Satisfaction	S11: I feel that reading with this tool gives me a lot of satisfaction.	Baseline	15	2.67	0.47	<b>0.0001**</b>	
	IRead	15	<b>3.67</b>	0.7			
	S12: I am curious about the textbook content I was reading.	Baseline	15	3.93	0.77	0.8247	
	IRead	15	3.87	0.81			
Satisfaction	S13: This tool presents the textbook content in a way that fulfills my expectation and goals.	Baseline	15	3.53	0.81	0.3339	
	IRead	15	3.2	0.98			
S14: This tool presents various supplementary information in an engaging way.	Baseline	15	2.67	1.01	<b>0.0004**</b>		
IRead	15	<b>4.0</b>	0.73				

also evidenced by the fact that participants in experiment group provides more variant answers to the content-related questions compared to the control group (Fig. 5).

**Simple Visualizations Can Be Powerful** *IRead* employs a simple concept tree visualization to present the semantic hierarchy of concepts. We have observed that the majority of participants react positively to this visualization and use it for various purposes in accordance with the design requirements. Their ratings for engagement and satisfaction are also significantly higher than those of the baseline group (see Table 3-S10, S11, S14). Four participants in the experiment group utilize the concept tree to gain an overview and navigate through “*the things they have to learn*” before reading (R1). Six participants use it more frequently during the reading process to check concept definitions (R3) and search for answers to questions indicated by question hints (R2). They comment that “*it’s really clear...better to understand...even better than a definition from Google or Wikipedia*”(P1). Meanwhile, we have observed participants with a deeper background knowledge of the textbook topic (P5, P7) using the concept tree as a tool to review and confirm their understanding of all related key concepts in the chapter. In contrast, participants with almost zero background knowledge (P11) treat the concept tree itself as a learning material and directly learn from it. Additionally, we have observed three “*non-visual people*” (P3, P15) choose to skip the tree and use other

**Table 4** Statistics of participants' preference towards different questions timing

Question Content	Upcoming	Current	Previous
Positive	4	4	0
Negative	5	4	1
Neutral	3	2	1
Not Trigger	3	5	13

Questions about previously read content can only be triggered by highlighting a few heading sentences of subsections, so not all participants have the opportunity to encounter them

functions to facilitate reading. This flexibility in using the same visualization in different ways demonstrates the level of customization we provide for users with varying educational backgrounds and personal reading habits (**R4**).

**Question Timing is Critical** In the study, all participants in the experiment group comment and express their preferences regarding the timing of the questions. We track the timing of these questions based on interactions including highlighting, note-taking, and concept-definition checking initiated by the participants. Table 4 illustrates their opinions regarding questions about upcoming, current, and previously read content. We observed that most participants exhibited a consistent preference about upcoming and current content. This observation is also relevant to the convergent-thinking versus divergent-thinking reading habits. When the type of questions they prefer dominates the reading process, participants tend to be more satisfied with the quality of the questions. It's also worth noting that P5's attitude towards all questions, especially early ones, changed from negative to positive after overcoming the learning curve and becoming accustomed to the questions in *IRead*.

**Education Backgrounds Influence Preferences** Given that a significant number of participants find the distracting level of *IRead* to be high, we conduct a closer investigation and discover that higher academic ranks (bachelor, master, doctoral, post-doctoral) correlate with lower perceived quality ( $\rho = 0.2$ ) and increased reports of distraction ( $\rho = 0.1$ ). This is computed as Pearson correlations (academic rank from 0 to 3 against the target variable) with a significance level of  $p < 0.05$  two-sided (Kowalski, 1972). Participants in these higher academic ranks express a preference for receiving less visual and textual aid from *IRead* or no questions at all (see P12, P15). In contrast, participants with lower academic ranks described their ideal AI tools as providing “*intelligent summarization of content*” (P2, P10) and high-quality questions (P10). This finding aligns with prior HCI studies that highlight the influence of user expertise (Sauer et al., 2010; Bubalo et al., 2016). For different questions with objective, personal, or inspiring content, participants' opinions also vary. Divergent-thinking readers with business or management backgrounds (P3, P7, P14) prefer inspiring questions and questions that connect to their personal experiences. In contrast, convergent-thinking readers with mathematical or engineering backgrounds (P2, P13, P15) prefer more objective questions.

**Reading Scenarios and States Influence Behaviors** During the study, participants exhibit three distinct modes influenced by time pressure and task requirements. In the initial 15-20 minutes of **T1**, participants are relaxed to explore and comment on functions in *IRead*.

They tend to utilize highlighting and note-taking functions more extensively and engage in question answering and visualization exploration to deepen their learning. As the reading task approached the 10-minute mark, we reminded participants about the time, prompting most of them (13 out of 15) to transition into a pressure mode and prioritize the completion of the reading task. In this mode, their usage was limited to highlighting and concept definition functions. Finally, during **T2**, simulating an open-examination scenario, participants utilize content searching functions more frequently for questions corresponding to the “memorize” and “understanding” levels of Bloom’s taxonomy. Meanwhile, participants tend to review their own highlighting and note-taking history while also exploring the concept visualization for addressing higher-level open questions that require deeper personal reflection. Some participants noted that their behavior might vary when using *IRead* with textbooks from different disciplines. For instance, P3 mentions their potential difference in behavior when using math textbooks, stating, “*I would have very specific questions... it’s very hard to generate questions exactly focusing on this one...because you cannot make sure that it answered the questions I have, because maybe I don’t understand an example.*” Moving forward, we plan to incorporate a customization component that involves instructors in contributing to the cultivation of the concept tree and the generation of questions.

## Discussion

From the user study result in Section “[User Study](#)”, we observed interesting reading patterns and varying attitudes toward the visual and textual support provided by *IRead*. In this section, we will further discuss these observations and reflect on the optimal ways to enhance the pedagogical context through appropriate AI and visualization technologies. To supplement this discussion with broader perspectives, we also presented *IRead* and gathered feedback from two textbook authors during unstructured interviews, as well as 21 potential users including professors, lecturers, scientists, and students, in an onsite Learning & Teaching Fair at a research university. Due to the diverse backgrounds of the interviewees and the limited interaction time, we were unable to establish a formal study format in advance. Nevertheless, we outline our plan to incorporate the insights gained into the next development iteration and to conduct more rigorous, large-scale studies in the future in Section “[Limitation and Future Work](#)”.

## AI-Generated Content Quality

Although AI becomes increasingly popular in assisting various educational contexts, concerns about the quality of AI-generated content continue to emerge (Jeon and Lee, 2023; Kasneci et al., 2023; Tili et al., 2023). In our user study, participants with higher academic ranks (P12, P15) exhibited less trust in AI-generated questions, preferring less visual and textual aid from *IRead* or opting for no questions at all. Our interaction with potential users revealed that, while most of them had a positive impression of *IRead* and expressed a willingness to explore the tool, they were less inclined to incorporate it into their daily teaching or learning routines. Five explicitly voiced concerns regarding the quality of AI-generated content. And when we showcased *IRead* to textbook authors with their own textbook, they can quickly identify any discrepancies between their own knowledge and the content gener-

ated by the LLM. Nevertheless, they acknowledge that AI serves as a good starting point for them to work on, even if it cannot always discern subtle differences or generate content that satisfies the authors or instructors. To alleviate this issue, implementing a human-in-the-loop approach for quality control is an effective solution. This could involve more extensive post-editing and customization of AI-generated content to ensure, as one textbook author noted, “*at least we ensure nothing is wrong (before releasing it to the students)*”. We also advocate for adherence to the transparency obligations outlined in the Artificial Intelligence Act of the European Union (EU AI Act) (Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence, 2021) and explicitly annotate which parts are generated by AI to avoid potential confusion.

### Information Support and Attention Guidance in Learning Context

Reflecting on the long learning curve and complex AI-generated content of many prior visual analytics systems, especially those with sophisticated workflow and AI-driven mechanisms (Zhang et al., 2023; Zytek et al., 2021), we make a different attempt in *IRead* to employ a simple collapsible tree for the concept visualization and a question bot rather than a chatbot for engagement. As demonstrated in Section “[Results](#)”, this design has yielded overwhelmingly positive feedback and has met nearly all requirements outlined in Section “[Design Considerations](#)”. From this practice, we learned that visualization for learning support should prioritize reducing cognitive load over showcasing elaborate functionalities, facilitating a more intuitive transfer of information from the AI agent to users. This insight addresses our research question **Q2**.

Nevertheless, statistical results in Table 3 show that participants still felt *IRead* could be more distracting (statement S3) and confusing (statement S5) compared to the baseline tool, primarily due to the questions displayed in Question Wiz, which provide a relatively negative response to research question **Q1**. We partly attribute this to the adaptation curve associated with the new reading paradigm introduced in *IRead* (reading with continuously updating questions), and specifically to the lack of direct answers to those questions. Participants’ perception that the content in *IRead* did not meet their expectations compared to the baseline system (statement S13) further supports this explanation. As a quick solution, we have implemented a “focus mode” in *IRead*, allowing readers to hide content generated by Question Wiz during reading and review them afterwards for self-checking.

Another potential cause of this issue is the fun fact triggered by inactivity, which might impose additional extraneous cognitive load (Clark et al., 2011; Abeysekera and Dawson, 2015) on readers. However, these fun facts are prompted to be simple, relevant, and without demanding problem-solving or deep cognitive processing. This design aligns with findings in multimedia learning, which suggest that well-placed “*interesting but non-demanding*” elements can enhance engagement without impairing learning outcomes if managed effectively (Mayer, 2005). Moreover, fun facts appear only in rare instances when readers have paused their reading for a significant duration, indicating that learners have already disengaged from the learning process. At such moments, fun facts can help reorient disengaged learners rather than impose extraneous cognitive load. We believe that the new “focus mode” mentioned in the last paragraph can mitigate potential side effects. This mode enables self-paced learning, which is the primary recommendation in Clark et al. (2011) for reducing extraneous cognitive load. Additionally, Clark and Mayer (2023)

suggests emotional design as an alternative to extraneous text to enhance motivation and learning. To incorporate emotional design in *IRead*, we plan to display encouraging words alongside progress reminders or utilize the robot icon displayed at the lower right corner of Question Wiz, giving it varied expressions similar to the owl in the Duolingo App (Bicknell et al., 2023). This strategy aims to boost learner motivation without adding significant cognitive load from reading extra text.

In the meantime, our user study also revealed that some participants were not satisfied with the amount of information provided in *IRead*. For instance, four participants explicitly request that they want a Google Search function integrated in the tool to support searching of any selected keyword, and three mention they would like to directly communicate with the chatbot. These features were intentionally avoided during the design phase, as we believed they could easily distract the readers from trivial topics. From their lower ratings on confidence-related dimensions (statement S7, S8) in the questionnaire and the follow-up interviews, we learned that such requests may be due to a lack of security and confidence. For instance, P6 mentions that “*if I have more information, I feel safer. And then, I would like to learn more, not just the text*”. They acknowledge that Google Search can be “*kind of distracting*”, but they still feel it is necessary. As a tradeoff, we plan to provide a deprived version of internal search, but also try to maintain the reader’s attention by avoiding excessive information or external links.

### Individual Difference in Reading Preference

During the user study and our conversation with textbook authors, we have identified some typical usage habits and preferences. We document the most common types along with the textbook authors’ opinions as a reference for future researchers in the field below. These varied reading patterns also indicate that there is no unified answer to research question Q3. Depending on the reader’s background, reading habits, and purpose, the role and effectiveness of *IRead* can vary.

**Novice vs. Professional Reading** As evidenced in Section “[Results](#)”, the functionalities provided by *IRead* can already adaptively cater to varied reading requirements. However, we still want to highlight the different reading preferences of novice and professional readers. Novice readers usually benefit from a linear reading approach as they are unfamiliar with the organization of the content in the textbook. For these readers, it is critical to keep them engaged and guide them through the entire reading process. In contrast, professionals usually possess more domain knowledge and have specific information to find when they read. Their biggest challenge is usually “*not knowing the exact name or location*”, so we need to help them quickly find the content they need and filter out any AI-generated content that might distract them.

**Linear vs. Nonlinear Reading** As mentioned above, we observed both linear and nonlinear reading styles during the user study. Contrary to our expectations, discussions with the textbook authors revealed a surprisingly open attitude toward students’ reading strategies. They did not intend to enforce the linear structure they designed for the textbook and agreed that “*students have different reading strategies*”. One of the authors, who also had experience instructing a master’s course using the textbook they wrote, mentioned that

they had tried to force students to read the textbook linearly, but “*it didn’t work*”. The enhanced PDF reader in *IRead* supports linear reading, while the visualization and searching functions facilitate nonlinear reading. We believe this flexibility could allow readers to personalize their reading experiences, maximizing their benefit from the learning process.

**Reading First vs. Question First** In *IRead*, the primary form of AI assistance is through AI-generated questions. And we observed that participants in the user study had varying preferences regarding when and how to utilize these questions. Some used the questions as a preview, while others employed them as a recap. For instance, P5 preferred to view questions before reading to “*read with questions in mind*”, whereas P7 and P15 preferred to read all questions after completing the reading to maintain uninterrupted focus. In teaching practice, educators need to be aware that students may have different understandings and approaches to use learning materials, particularly AI-generated content, which can easily be biased and misleading (Lee et al., 2024; Nguyen, 2024). While maintaining an open attitude towards diversity, educators should also be mindful of the potential risks and solutions associated with the misuse of such materials.

### Limitation and Future Work

Our primary user study was conducted in a lab setting over a relatively short period, without tracking the long-term learning performance of students. Therefore, the measures obtained from the questionnaire developed based on the principles of the ARCS model may be limited and potentially subjective. As part of our future work, we aim to conduct a larger-scale field study to assess *IRead* in real-world educational contexts and measure longitudinal learning outcomes, such as comprehension and retention. We also plan to interview textbook authors and other stakeholders to collect comprehensive feedback on *IRead* from diverse perspectives.

Due to time constraints in the development cycle of *IRead*, we were unable to implement high levels of customization or individualized feedback, such as enabling manual editing of the concept map or providing lightweight checkpoints for specific learners. Therefore, we recommend using *IRead* as one of several learning modes in a course, so that feedback mechanisms can be implemented in the wider course design (including quizzes, group work, assignments, discussions, exams). Besides, the scalability of the validation mechanisms for AI-generated questions and concept maps is also limited and still relies on manual oversight. To eliminate this problem, we plan to develop a flagging system that allows users (students or instructors) to highlight potentially incorrect or unclear content. We will also explore automated validation methods to improve scalability. For users who find the AI-generated questions and answer-feedback functionalities too distracting, we will provide an option to disable them.

### Conclusion

In this paper, we present *IRead*, an interactive textbook reading tool grounded in guided learning theory. Our system integrates a state-of-the-art question generation model and an interactive visualization to enrich students’ reading and learning experiences. It also

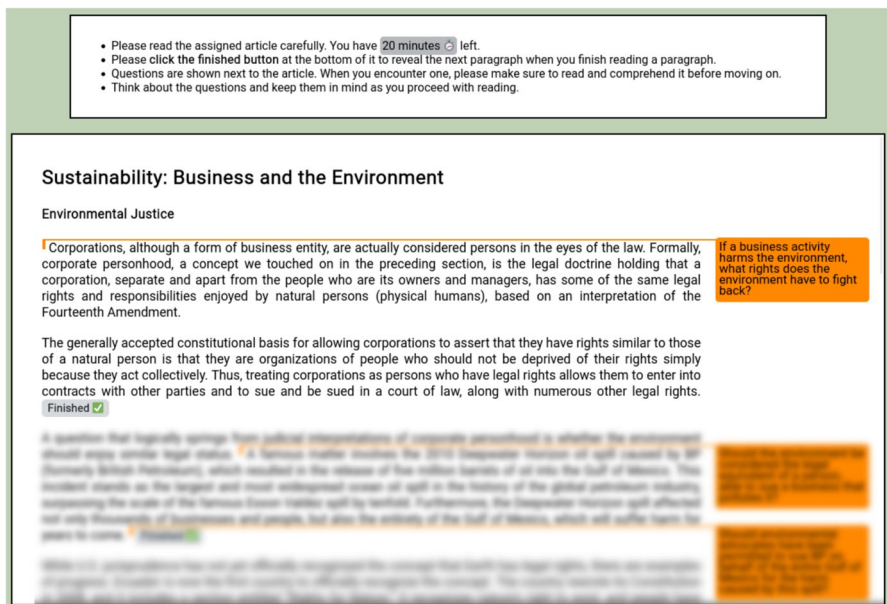
implements a question-driven learning guidance mechanism built upon the ARCS model. Through a comparative study involving current and past students involved in higher education, we demonstrate the effectiveness of our approach in engaging students with textbook reading and the potential of applying human-AI collaboration technologies for educational purposes. Two universities have already expressed interest in integrating *IRead* into their teaching and learning systems, motivating our immediate next step to deploy and further evaluate its functionality in real-world teaching environments. Our ultimate goal is to make *IRead* publicly accessible and easily generalizable, allowing students of all locations, education levels, and knowledge backgrounds to benefit from the novel learning experiences offered by *IRead* and the new generation of AI-enhanced education.

## Supplementary Information

We have provided a demo video to illustrate the interactions in *IRead* with the two use case scenarios described in Appendix C.

## Appendix A Experiment Interface for Preliminary Study

The preliminary study described in Section “[Design Considerations](#)” was conducted using a textbook reading interface, as illustrated in Fig. 6.



**Fig. 6** The document contains generated active reading questions. Paragraphs not yet reached by the reader are blurred

## Appendix B LLM Prompts

Listing 1 is the prompt used to generate the concept tree data described in Section “[Concept Tree Generation](#)”. Listing 2 is the prompt used to generate the table of content data mentioned in Section “[Visual Interface](#)”.

### Listing 1 Concept Generation Prompt

Your task is to extract the primary overarching concepts given the text below. For each primary concept, discern a maximum of five sub-concepts with their own respective sub-concepts that are directly related. Ensure that the list of concepts does not become too long. Organize these concepts hierarchically, showcasing the main concepts at the top level and their corresponding sub-concepts indented beneath them. Aim to capture the fundamental structure of the document’s concepts in a hierarchical format. The reader is a university student with good understanding in the text’s field.

Return a nested JSON object with the following keys:

- "name": the concept
- "short\_name": the abbreviated concept (if there is one or it is necessary)
- "definition": a short definition of the concept
- "example": an example of the concept (if applicable, may be a sentence or two)
- a "verbatim\_words" list: containing strings that appear in the text verbatim about the current concept (this list will be used to highlight the concept within the text to the user, be as specific as possible)
- a "children" array: containing the children (if they exist) in the same format

Additionally, for the leaf concepts, add a "relative\_frequency" key denoting the appearance frequency (between 0 and 1) and an "absolute\_frequency" that is the absolute appearance count (number) of this concept in the text. For the absolute frequency, assign each sentence to one of the leaf concepts if applicable and sum them for each leaf concept. The relative frequencies should sum to 1.

**Listing 2** Table of Contents Generation Prompt

Your task is to generate a table of content.

Return a nested JSON object with 3 levels. Every level includes following keys:

- "index": section index as string
- "title": the title of the current section (no more than 5 words)
- "summary": a short summary of the current section
- "length": the number of characters in the current section. It should equal to the total length of all its child subsections plus its own character count.
- "char\_start": the index of first character the corresponding section
- "char\_end": the index of last character of the current section. It should equal to char\_end of its last children subsection.
- "sen\_start": the first sentence of the current section
- "sen\_end": the last sentence of the current section. It should equal to sen\_end of its last children subsection.
- "subsections": a list of subsections with the same keys as the current one
- "depth": the depth number of current layer

Eliminate all empty object or list.

## Appendix C Use Case Scenarios

### C.1 Navigating New Knowledge as Novice Learner

In this use case, we describe a curious student with minimal background in psychology using *IRead* to read a psychology textbook (Spielman et al., 2020) for a single-semester introductory psychology course. The student, interested in the topics of emotion and motivation, first browses the chapter titles and selects Chapter 10 to begin reading. Once the chapter is loaded, the student immediately notices a figure from the book illustrating intrinsic and extrinsic motivation. Curious about these terms, they click on the underlined concept *intrinsic motivation* from the figure title, which updates the visualization in the upper right corner and reveals the hierarchical position of the concept. There, they learn that intrinsic and extrinsic motivation are the two main types of motivation. They also hover over the concept to reveal both the definition and examples of these terms, grasping an overview of the conceptual grounding before delving into the reading material. Meanwhile, *IRead* simultaneously prompts a related question in the reading assistant panel as shown in Fig. 7-A. This question, along with hints to related concepts, helps the student engage with the material early on.

As the student progresses with the reading, they encounter a statement that aligns with their previous findings from the visualization: “*motivations can be intrinsic (arising from internal factors) or extrinsic (arising from external factors)*”. They choose to highlight this as an important point using the yellow highlighting color. Upon creating the highlight, a new question appears: “*What drives you to seek knowledge and pursue education from within yourself?*” Reflecting on and trying to answer

The screenshot displays the IRead interface for a psychology textbook chapter on 'Emotion and Motivation'. On the left, a 'Concept Definition Checking' sidebar shows a mind map with nodes for 'Types of Motivation', 'Extrinsic Motivation', 'Overjustification Effect', 'Cultural Influences on Motivation', 'Theories About Motivation', and 'Intrinsic Motivation'. The main textbook page includes a title '10 Emotion and Motivation', a sub-section 'Psychology', and a paragraph discussing intrinsic vs. extrinsic motivation. A diagram shows 'Intrinsic Motivation (From within)' with sub-points: Autonomy, Mastery, Purpose and 'Extrinsic Motivation (From outside)' with sub-points: Compensation, Punishment, Reward. Below the text, a 'Question Highlighting' box contains a question: 'What drives you to seek knowledge and pursue education from within yourself?'. A 'Question for Note-Taking' box contains: 'Why do we sometimes prioritize financial gain over our personal preferences when making decisions or taking action?'. A 'Question for Inactivity' box contains: 'Did you know that the Over-justification Effect can sometimes lead to a decrease in intrinsic motivation for an activity you love? This happens when external rewards (like money or prizes) are given for doing something you already find rewarding, potentially making the activity feel less like a choice and more like a chore.' The interface also includes a 'Question Wizard' and a 'Feedback' button.

**Fig. 7** *IRead* with a textbook on psychology while reading chapter 10 with “intrinsic motivation” as selected concept

this question encourages a deeper personal connection with the content and enhances the student’s learning experience.

In further reading, the student notes the passage: “*physical reinforcement (such as money) and verbal reinforcement (such as praise) may affect an individual in very different ways.*” Wanting to personalize this information, the student highlights it and creates a note stating, “*I am more likely to do something I dislike if I receive financial compensation.*” They then pause to reflect on this topic and gradually drift away in thought. The system detects this inactivity and presents a fun fact about the current topic to re-engage the reader. This new information triggers the student’s curiosity and motivation, prompting them to resume reading. Eventually, the student finishes the chapter with determination and enthusiasm, in an interactive and immersive reading environment fostered by *IRead*.

## Reviewing Previous Content for Exam Preparation

The second use case illustrates how students can use *IRead* for review and exam preparation. Consider a business student who used the system throughout the semester and is now leveraging it to reinforce and self-test critical concepts related to entrepreneurship in preparation for an exam. The student begins by carefully reviewing the concepts addressed in the chapter through the concept visualization, recalling the specific contexts in which they were discussed. They also turn on the frequency toggle within the visualization to identify the key concepts to focus on. During this process, they realize that they cannot clearly recall the meaning of “*entrepreneurial perspective*”, so they hover over the concept to review its definition and click on it to explore the children concepts.

Afterwards, the student turns to the textbook content. Given that this textbook was utilized throughout the semester, it already contains numerous notes and highlights made by the user (Fig. 1-D), making it easy for them to navigate through the content. They then notice a notable new sentence: “*This combination of recognizing an opportunity to bring something new to the world and acting on that opportunity is what distinguishes an entrepreneur from a small business owner*” and highlight this sentence in green. This interaction triggers a question: “*What unique qualities sepa-*

rate entrepreneurs from small business owners and drive innovation and growth in the business world?” Recognizing the importance of this question, the student spends extra time finding the answer. When faced with a similar question on the exam, they are able to answer correctly due to their thorough preparation.

As they continue reading, they come across another concept “*brainstorming*”. Remembering that this concept is addressed in multiple chapters with different contexts, they use the search feature to locate additional information. The search bar provides a list of chapters where the concept occurs, along with the frequency of its mentions. The student navigates to the top chapter listed in the search results, “*Chapter 6 – Problem Solving and Need Recognition Techniques*”, and review their previous highlights there (Fig. 1-A3). Recognizing this brief disruption, Question Wiz suggests returning to the prior reading and finishing reviewing the entire chapter by reminding the reading progress. The student accepts the suggestion and picks up their prior reading context. Finally, the student accesses all their notes and highlights in the sidebar, double-checking all concepts in the concept tree once more. Now, they feel well-prepared and confident for the upcoming exam.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40593-025-00515-4>.

**Acknowledgements** This research is partially sponsored by CityU Startup Grant (project number 9610708). We thank the participants of our studies and the organizers and attendees of the Learning & Teaching Fair 2024 at ETH Zurich for their feedback to IRead. We are grateful to Mr. Michel Baudin for allowing us to test IRead on his textbook and for participating in our interviews as a textbook author. We extend our thanks to Prof. Dr. Menna El-Assady for her high-level suggestions on this work. We also thank Ms. Katalin Tesch for curating the demo video for this submission. Last but not least, we appreciate the valuable feedback and suggestions from the reviewers.

**Author Contributions** Xiaoyu Zhang and Vincent Dörig contributed equally to this paper. Xiaoyu Zhang and Vincent Dörig developed the IRead system, conducted the user studies and wrote the main manuscript text except for Section “*Participants*”. Peng Cui developed the question generate model and wrote Section “*Participants*”. Vilém Zouhar developed the interactive interface in Appendix A for the preliminary study and helped with the user study result analysis. Torbjorn Netland and Mrinmaya Sachan provided high-level instructions to this work. All authors reviewed the manuscript.

**Funding** Open access publishing enabled by City University of Hong Kong Library's agreement with Springer Nature. This research is partially sponsored by City University of Hong Kong Startup Grant (project number 9610708).

**Data Availability** Data analysis is provided within the manuscript.

## Declarations

**Competing Interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abeysekera, L., Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *Higher Education Research & Development* 34(1), 1–14.
- Banerjee, S., Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In: Goldstein, J., Lavie, A., Lin, C.-Y., Voss, C. (eds.) Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation And/ or Summarization, pp. 65–72. Association for Computational Linguistics, Ann Arbor, Michigan . <https://aclanthology.org/W05-0909>
- Bharuthram, S. (2017). The reading habits and practices of undergraduate students at a higher education institution in south africa: A case study. *The Independent Journal of Teaching and Learning*, 12(1), 50–62.
- Bicknell, K., Brust, C., & Settles, B. (2023). How duolingo’s ai learns what you need to learn: The language-learning app tries to emulate a great human tutor. *IEEE Spectrum*, 60(3), 28–33.
- Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D.R., et al. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain. Longman, New York, NY
- Brown, A. L., & Campione, J. C. (1994). Guided Discovery in a Community of Learners. *The MIT Press, Cambridge, MA*. <https://doi.org/10.7551/mitpress/1861.001.0001>
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141–178.
- Bruner, J.S. (1961). The act of discovery. Harvard educational review
- Bubalo, N., Honold, F., Schüssel, F., Weber, M., Huckauf, A. (2016). User expertise in multimodal HCI. In: Proceedings of the European Conference on Cognitive Ergonomics, pp. 1–6. <https://dl.acm.org/doi/abs/10.1145/2970930.2970941>
- Bustos, A. (2020). *Guided Discovery approach in CEI Remote teaching*. Buenos Aires, Argentina: British Council.
- Chang, N.-C., & Chen, H.-H. (2015). A motivational analysis of the arcs model for information literacy courses in a blended learning environment. *Libri*, 65(2), 129–142.
- Clark, R.C., Mayer, R.E. (2023). E-learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. John Wiley & sons, ???
- Clark, R.C., Nguyen, F., Sweller, J. (2011). Efficiency in Learning: Evidence-based Guidelines to Manage Cognitive Load. John Wiley & Sons, ???
- Cohan, A., Derroncourt, F., Kim, D.S., Bui, T., Kim, S., Chang, W., Goharian, N. (2018). A discourse-aware attention model for abstractive summarization of long documents. In: Walker, M., Ji, H., Stent, A. (eds.) Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), pp. 615–621. Association for Computational Linguistics, New Orleans, Louisiana. <https://doi.org/10.18653/v1/N18-2097>
- Cui, P., Sachan, M. (2023). Adaptive and personalized exercise generation for online language learning. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (eds.) Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 10184–10198. Association for Computational Linguistics, Toronto, Canada. <https://doi.org/10.18653/v1/2023.acl-long.567>
- Cui, P., Zouhar, V., Zhang, X., Sachan, M. (2024). How to engage your readers? generating guiding questions to promote active reading. In: Ku, L.-W., Martins, A., Srikumar, V. (eds.) Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 11749–11765. Association for Computational Linguistics, Bangkok, Thailand. <https://aclanthology.org/2024.acl-long.632>
- Dewey, J., et al. (1897). The university elementary school: History and character. *University Record*, 2, 72–5.

- Feng, S.-L., & Tuan, H.-L. (2005). Using arcs model to promote 11th graders' motivation and achievement in learning about acids and bases. *International Journal of Science and Mathematics Education*, 3, 463–484.
- Guthrie, J.T., Wigfield, A. (1997). Reading Engagement: Motivating Readers Through Integrated Instruction. ERIC, Washington, DC. <https://eric.ed.gov/?id=ED404627>
- Guthrie, J. T., Alao, S., & Rinehart, J. M. (1997). Literacy issues in focus: Engagement in reading for young adolescents. *Journal of Adolescent & Adult Literacy*, 40(6), 438–446.
- Guu, K., Lee, K., Tung, Z., Pasupat, P., Chang, M. (2020). Retrieval augmented language model pre-training. In: International Conference on Machine Learning, pp. 3929–3938. PMLR
- Huett, J. B., Kalinowski, K. E., Moller, L., & Huett, K. C. (2008). Improving the motivation and retention of online students through the use of arcs-based e-mails. *The American Journal of Distance Education*, 22(3), 159–176.
- Hyland, K. (2002). What do they mean? questions in academic writing. *Text & Talk*, 22(4), 529–557.
- Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and chatgpt. *Education and Information Technologies*, 28(12), 15873–15892.
- Jiao, Y., Shridhar, K., Cui, P., Zhou, W., Sachan, M. (2023). Automatic educational question generation with difficulty level controls. In: International Conference on Artificial Intelligence in Education, pp. 476–488. Springer
- Johnson, B.G., Van Campenhout, R., Jerome, B., Castro, M.F., Bistolfi, R., Dittel, J.S. (2024). Automatic question generation for spanish textbooks: Evaluating spanish questions generated with the parallel construction method. *International Journal of Artificial Intelligence in Education*, 1–20
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. (2023). Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274
- Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development*, 10(3), 2–10.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86.
- Ko, W.-J., Dalton, C., Simmons, M., Fisher, E., Durrett, G., Li, J.J. (2022). Discourse comprehension: A question answering framework to represent sentence connections. In: Goldberg, Y., Kozareva, Z., Zhang, Y. (eds.) Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pp. 11752–11764. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates. <https://doi.org/10.18653/v1/2022.emnlp-main.806>
- Kowalski, C. J. (1972). On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 21(1), 1–12.
- Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 30, 121–204.
- Lee, G. G., Goodman, D., & Chang, T. C. (2024). Racial and gender bias in artificial intelligence generated ophthalmologic educational material. *Investigative Ophthalmology & Visual Science*, 65(7), 352–352.
- Leroy, C., & Kammerer, Y. (2023). Reading multiple documents on a health-related issue: The roles of a text-highlighting tool and re-reading behaviour in integrated understanding. *Behaviour & Information Technology*, 42(14), 2331–2352.
- Lewin, K. (1968). *The Conceptual Representation and the Measurement of Psychological Forces. Contributions to psychological theory.* Duke University Press, ??? <https://books.google.com/books?id=0yd9AAAAAAAJ>
- Lewis, C. (1982). Using the 'thinking-aloud' method in cognitive interface design. Research Report RC9265, IBM T. J. Watson Research Center
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J. (eds.) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7871–7880. Association for Computational Linguistics, Online. <https://doi.org/10.18653/v1/2020.acl-main.703>

- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.
- Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In: Text Summarization Branches Out, pp. 74–81. Association for Computational Linguistics, Barcelona, Spain. <https://aclanthology.org/W04-1013>
- Liu, H., Liu, Z., Wu, Z., Tang, J. (2020). Personalized multimodal feedback generation in education. In: Scott, D., Bel, N., Zong, C. (eds.) Proceedings of the 28th International Conference on Computational Linguistics, pp. 1826–1840. International Committee on Computational Linguistics, Barcelona, Spain (Online). <https://doi.org/10.18653/v1/2020.coling-main.166>
- Lu, X., Fan, S., Houghton, J., Wang, L., Wang, X. (2023). 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, pp. 1–18 (2023). <https://dl.acm.org/doi/full/10.1145/3544548.3580957>
- Maiti, M., Priyaadharshini, M., et al. (2023). Design and evaluation of a revised arcs motivational model for online classes in higher education. *Heliyon* 9 (12)
- Ma, L., & Lee, C. S. (2021). Evaluating the effectiveness of blended learning using the ARCS model. *Journal of Computer Assisted Learning*, 37(5), 1397–1408.
- Markel, J.M., Opferman, S.G., Landay, J.A., Piech, C. (2023). GPTeach: Interactive TA training with GPT-based students. In: Proceedings of the Tenth ACM Conference on Learning @ Scale. L@S '23, pp. 226–236. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3573051.3593393>.
- Matthews, M. R. (1998). *Constructivism in Science Education: A Philosophical Examination*. Berlin, Germany: Springer.
- Mayer, R.E. (2005). *The Cambridge Handbook of Multimedia Learning*. Cambridge university press, ???
- Mizrahi, D., Salaz, A.M., Kurbanoglu, S., Boustany, J., Group, A.R. (2018). Academic reading format preferences and behaviors among university students worldwide: A comparative survey analysis. *PLoS one*, 13(5), 0197444.
- Moore, L.M. (2009). At your leisure: assessing ebook reader functionality and interactivity. Unpublished dissertation). University College London, London
- Nguyen, T. (2024). *ChatGPT in medical education: a precursor for automation bias?* Canada: JMIR Publications Toronto.
- Özçakmak, H. (2019). Impact of note taking during reading and during listening on comprehension. *Educational Research and Reviews*, 14(16), 580–589.
- Palincsar, A. S. (1989). Collaborative research and development of reciprocal teaching. *Educational Leadership*, 46(4), 37–40.
- Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). Accessed: 2025-01-12. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140), 1–67.
- Rahmani, M., & Sadeghi, K. (2011). Effects of note-taking training on reading comprehension and recall. *Reading*, 11(2), 116–128.
- Ruan, S., Jiang, L., Xu, J., Tham, B.J.-K., Qiu, Z., Zhu, Y., Murnane, E.L., Brunskill, E., Landay, J.A. (2019). Quizbot: A dialogue-based adaptive learning system for factual knowledge. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–13
- Ruan, S., Willis, A., Xu, Q., Davis, G.M., Jiang, L., Brunskill, E., Landay, J.A.: Bookbuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. In: Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale, pp. 1–4 (2019)
- Sarsa, S., Denny, P., Hellas, A., Leinonen, J. (2022). Automatic generation of programming exercises and code explanations using large language models. In: Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 1. ICER '22, pp. 27–43. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3501385.3543957> .
- Sauer, J., Seibel, K., & Rüttinger, B. (2010). The influence of user expertise and prototype fidelity in usability tests. *Applied Ergonomics*, 41(1), 130–140. <https://doi.org/10.1016/j.apergo.2009.06.003>
- Shneiderman, B. (2003). The eyes have it: A task by data type taxonomy for information visualizations. In: *The Craft of Information Visualization*, pp. 364–371. Elsevier, ???

- Singh, J., Zouhar, V., Sachan, M. (2023). Enhancing textbooks with visuals from the web for improved learning. [arXiv:2304.08931](https://arxiv.org/abs/2304.08931)
- Spielman, R. M., Jenkins, W. J., & Lovett, M. D. (2020). *Psychology 2e*. Houston, TX: OpenStax.
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: Chatgpt as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15.
- Tolman, E. (1932). *Purposive Behavior in Animals and Men*. Appleton-Century, ???
- Wang, X., Fan, S., Houghton, J., Wang, L. (2022). Towards process-oriented, modular, and versatile question generation that meets educational needs. In: Carpuat, M., Marneffe, M.-C., Meza Ruiz, I.V. (eds.) Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 291–302. Association for Computational Linguistics, Seattle, United States. <https://doi.org/10.18653/v1/2022.naacl-main.22>
- Wang, Y., Mao, Y., Ni, S.-t. (2024). Metabook: An automatically generated augmented reality storybook interaction system to improve children's engagement in storytelling. [arXiv:2405.13701](https://arxiv.org/abs/2405.13701)
- Willis, A., Davis, G., Ruan, S., Manoharan, L., Landay, J., Brunskill, E. (2019). Key phrase extraction for generating educational question-answer pairs. In: Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale, pp. 1–10
- Wu, C.-H., Liou, G.-M. (2024). Arcs model for exploring the enhancement of learning motivation and engagement through aigc technology in computer graphics courses. In: Proceedings of the 2024 8th International Conference on Education and Multimedia Technology, pp. 51–59
- Zhang, X., Chandrasegaran, S., Ma, K.-L. (2021). Conceptscope: Organizing and visualizing knowledge in documents based on domain ontology. In: Proceedings of the 2021 Chi Conference on Human Factors in Computing Systems, pp. 1–13
- Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., Artzi, Y. (2019). Bertscore: Evaluating text generation with bert. [arXiv:1904.09675](https://arxiv.org/abs/1904.09675)
- Zhang, X., Li, J., Chi, P.-W., Chandrasegaran, S., Ma, K.-L. (2023). ConceptEVA: Concept-based interactive exploration and customization of document summaries. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–16
- Zhang, Z., Xu, Y., Wang, Y., Yao, B., Ritchie, D., Wu, T., Yu, M., Wang, D., Li, T.J.-J. (2022). Storybuddy: A Human-AI collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3491102.3517479>.
- Zhang, X., Xuan, X., Dima, A., Sexton, T., Ma, K.-L. (2023). Labelvizier: Interactive validation and relabeling for technical text annotations. In: 2023 IEEE 16th Pacific Visualization Symposium (Pacific-Vis), pp. 167–176. IEEE
- Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., & Hashimoto, T. B. (2024). Benchmarking large language models for news summarization. *Transactions of the Association for Computational Linguistics*, 12, 39–57.
- Zhu, Y., Yuan, H., Wang, S., Liu, J., Liu, W., Deng, C., Chen, H., Dou, Z., Wen, J.-R. (2023). Large language models for information retrieval: A survey. *CoRR* [arXiv:2308.07107](https://arxiv.org/abs/2308.07107)
- Zytek, A., Liu, D., Vaithianathan, R., & Veeramachaneni, K. (2021). Sibyl: Understanding and addressing the usability challenges of machine learning in high-stakes decision making. *IEEE Transactions on Visualization and Computer Graphics*, 28(1), 1161–1171.

## Authors and Affiliations

**Xiaoyu Zhang<sup>1,3,4</sup> · Vincent Dörig<sup>2,3</sup> · Peng Cui<sup>2</sup> · Vilém Zouhar<sup>2</sup> ·  
Torbjørn Netland<sup>3</sup> · Mrinmaya Sachan<sup>2</sup>**

✉ Xiaoyu Zhang  
xiaoyu.zhang@cityu.edu.hk

Vincent Dörig  
vdoerig@ethz.ch

Peng Cui  
peng.cui@inf.ethz.ch

Vilém Zouhar  
vzouhar@ethz.ch

Torbjørn Netland  
tnetland@ethz.ch

Mrinmaya Sachan  
msachan@ethz.ch

<sup>1</sup> School of Creative Media, City University of Hong Kong, Hong Kong, China

<sup>2</sup> Department of Computer Science, ETH Zürich, Zürich 8092, Switzerland

<sup>3</sup> Department of Management, Technology, and Economics, ETH Zürich, Zürich 8006, Switzerland

<sup>4</sup> ETH AI Center, ETH Zürich, Zürich, Switzerland