



# More AI Assistance Reduces Cognitive Engagement: Examining the AI Assistance Dilemma in AI-Supported Note-Taking

XINYUE CHEN, University of Michigan, USA

KUNLIN RUAN, University of Michigan, USA

KEXIN PHYLLIS JU, University of Michigan, USA

NATHAN YAP, University of Michigan, USA

XU WANG, University of Michigan, USA

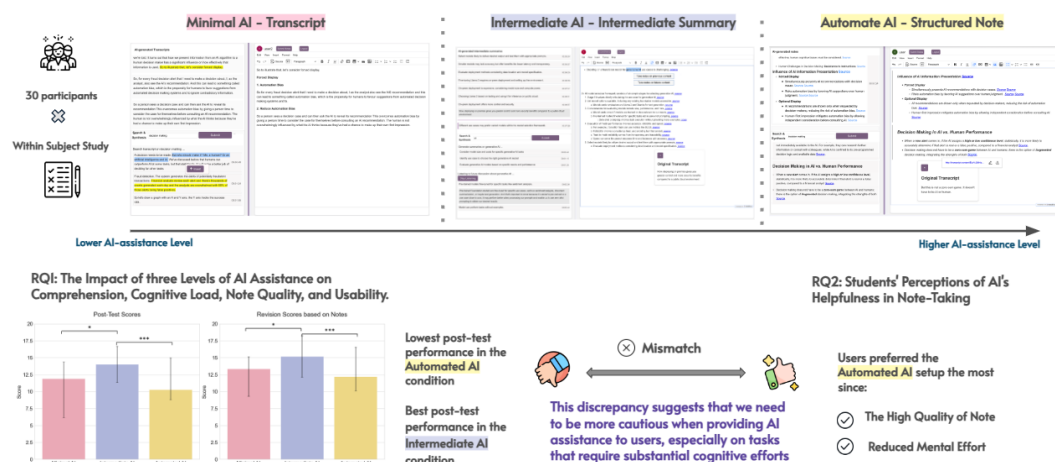


Fig. 1. In a within-subject experiment with 30 participants, we tested three conditions: **Automated AI** (high assistance), **Intermediate AI** (moderate assistance with summaries), and **Minimal AI** (transcript-only). Results show **Intermediate AI** yielded the highest learning outcomes, while **Automated AI** scored lowest despite being the most preferred for ease of use. These findings highlight the gap between preferred convenience and cognitive benefit, offering guidance for designing AI that enhances cognitive engagement in learning tasks.

As AI tools become increasingly integrated into cognitively demanding tasks, like note-taking, questions remain about whether they enhance or compromise cognitive engagement. This paper investigates the "AI Assistance Dilemma" in note-taking, examining how varying levels of AI support impact user engagement and comprehension. In a within-subject experiment, we asked participants ( $N=30$ ) to take notes during lecture videos under three conditions: **Automated AI** (high assistance with structured notes), **Intermediate AI** (moderate assistance with real-time summary, and **Minimal AI** (low assistance with transcript). Results

Authors' Contact Information: Xinyue Chen, xinyuech@umich.edu, University of Michigan, Ann Arbor, Michigan, USA; Kunlin Ruan, kunlinr@umich.edu, University of Michigan, Ann Arbor, Michigan, USA; Kexin Phyllis Ju, kexinju@umich.edu, University of Michigan, Ann Arbor, Michigan, USA; Nathan Yap, nyap@umich.edu, University of Michigan, Ann Arbor, Michigan, USA; Xu Wang, xwangh@umich.edu, University of Michigan, Ann Arbor, Michigan, USA.



This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2025 Copyright held by the owner/author(s).

ACM 2573-0142/2025/11-ARTCSCW451

<https://doi.org/10.1145/3757632>

reveal that Intermediate AI yields the highest post-test scores and Automated AI the lowest. Participants, however, preferred the automated setup for its perceived ease of use and perceived lower cognitive effort, suggesting a discrepancy between preferred convenience and cognitive benefit. Our study provides insights on designing AI assistance that preserves cognitive engagement, offering implications for designing moderate AI support in cognitive tasks.

CCS Concepts: • **Human-centered computing** → *Empirical studies in HCI*; **Empirical studies in interaction design**; • **Applied computing** → **Interactive learning environments**.

Additional Key Words and Phrases: AI Assistance Dilemma; Note-taking; Cognitive Load; Desirable AI Assistance

### ACM Reference Format:

Xinyue Chen, Kunlin Ruan, Kexin Phyllis Ju, Nathan Yap, and Xu Wang. 2025. More AI Assistance Reduces Cognitive Engagement: Examining the AI Assistance Dilemma in AI-Supported Note-Taking. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW451 (November 2025), 29 pages. <https://doi.org/10.1145/3757632>

## 1 Introduction

AI is increasingly used to facilitate cognitively challenging tasks, from providing next-sentence suggestions in writing [15] to generating meeting summaries [2, 9, 64]. The rise of AI productivity tools [22, 54, 57] raises the question: is AI assistance always desired and beneficial? When AI is leveraged to enhance productivity, could it compromise people's cognitive engagement in the task, leading to suboptimal outcomes? In this paper, we examine the AI assistance dilemma in the context of a crucial and prevalent task: note-taking. We examine to what extent AI-powered automatic note-taking services might reduce valuable cognitive engagement of the note-taker and how we could design human-AI collaborative mechanisms to provide the necessary AI assistance to the note-taker while preserving the cognitive effort required for meaningful content engagement.

Note-taking is a common activity during real-time information consumption, such as in meetings [60, 62], lectures [23, 52], and presentations, where participants must quickly process information without the ability to revisit what was said [9, 39]. Note-taking has proven to be a helpful tool to assist real-time information consumption by externalizing information and reducing the load on participants' working memory [39]. Note-taking serves two key functions: it helps people internalize knowledge through active engagement (encoding function) and provides a lasting reference for future recall (storage function) [34, 35].

Note-taking offers a unique and valuable context to study the AI assistance dilemma due to its dual demands on cognitive processing and external documentation [20, 65]. Effective note-taking requires users to engage deeply with information, selecting and organizing key points in real-time, which promotes active learning and understanding [17, 44]. However, when AI tools are introduced to automate or aid in this process, there is potential for reduced cognitive engagement, as users may rely on the AI rather than actively process information themselves [29].

It thus raises important questions about AI's role: Does AI assistance merely alleviate cognitive load, or does it undermine the mental engagement crucial for comprehension and retention? This concern reflects a longstanding issue known as the "Assistance Dilemma" in learning sciences, originally identified in the context of cognitive tutors deployed in classrooms [37]. This dilemma underscores that while AI-generated feedback can enhance learning, excessive assistance risks undermining learners' capacity for independent learning and critical thinking, whereas insufficient assistance may make tasks overly challenging [37, 50]. With the growing prevalence of AI assistance, we aim to investigate to what extent the AI assistance dilemma exists and how we might design human-AI collaborative mechanisms to address this challenge in a note-taking context.

To this end, we designed three levels of AI assistance to help people take notes during live lectures, namely 1) **Automated AI**, where participants received auto-generated AI notes around every

90 seconds, similar to common note-taking services; 2) **Intermediate AI**, where participants received real-time AI summary blocks; and 3) **Minimal AI**, where participants received real-time transcript blocks. In all three conditions, participants take notes in a text editor. Participants can drag and drop the AI-generated information (structured notes, summary blocks, and transcript blocks) into the text editor to compose their own notes. The **Automated AI** setup resembles popular note-taking services that provide a summary after a time interval<sup>1</sup>. With this setup, we aim to answer whether automated note-taking support would compromise people's cognitive engagement and negatively impact their understanding of the lecture content. By introducing the **Intermediate AI** and the **Minimal AI** setups, we aim to address whether giving users more control and agency during the note-taking process would enhance their understanding of the content compared to the **Automated AI** setup.

We performed a within-subject experiment with 30 participants, where they watched lecture videos and took notes in the three setups. Our findings suggest that 1) the AI assistance dilemma does exist. People had the lowest post-test performance in the **Automated AI** condition and the best post-test performance in the **Intermediate AI** condition. The difference is statistically significant (Figure 6). This finding suggests that although AI may enhance productivity, it risks compromising people's cognitive engagement and may lead to lower comprehension and retention. 2) Providing intermediate summary blocks is a useful way to address this challenge, as users in the **Intermediate AI** condition had the best post-test performance. Users had more control and more valuable cognitive engagement during the note-taking process when they were provided with the building blocks and had the agency to compose notes by themselves. 3) However, users preferred the **Automated AI** setup the most. They found that the auto-generated notes were of high quality and that they reduced their effort significantly. These findings highlight a discrepancy between preferred convenience and cognitive benefit. This discrepancy suggests that we need to be more cautious when providing AI assistance to users, especially on tasks that require substantial cognitive efforts. As users naturally prefer to minimize effort, the risk of over-reliance on AI becomes even more pronounced.

## 2 Related Work

### 2.1 Cognitive Costs and Benefits of Note-taking

Lecture note-taking has long been shown to be effective for learning. In-class note-taking helps students actively process information [44], improving memory retention [19], while the notes can also serve as an external representation for post-class review, which reinforces understanding and supports long-term knowledge retention [44, 48]. The underlying mechanism for these effects, known as the encoding-storage paradigm [34], suggests two key functions: **ENCODING**, where recording and taking notes itself promotes learning through increased attention and deeper processing, and **STORAGE**, where notes act as external memory for later review, reinforcing understanding by consolidating information and helping to slow forgetting.

Many studies have shown detrimental effects of taking notes while learning, as note-taking costs attentional resources that are needed for processing rapid and dense lecture presentations [7]. Research connecting note-taking with cognitive load theory suggests that when students attempt to write down information verbatim, it can lead to an increase in extraneous cognitive load, which negatively impacts learning effectiveness [66]. To optimize note-taking, researchers have explored structured techniques, finding that formats like pre-organized sheets [17] and outlines [35] reduce extraneous load and promote germane processes such as comprehension and synthesis [13, 30].

<sup>1</sup><http://otter.ai/>

These structured approaches can also enhance metacognition, encouraging students to focus on understanding knowledge structures and how to prioritize, organize, and learn effectively [17].

In summary, past research underscores the importance of reducing unnecessary cognitive load while supporting note-taking's encoding and storage functions. Building on these theories, our study explores how varying levels of AI assistance impact encoding, storage, and cognitive load optimization in note-taking.

## 2.2 Tools to Enhance Note-taking

HCI researchers have developed numerous tools to support note-taking. One line of the research focuses on note-taking modalities and input methods, offering flexibility for capturing notes via voice [33], digital ink [26], sketching [71], and mobile inputs [49]. Studies comparing note-taking modalities reveal that while typing enables faster capture, handwriting better supports encoding due to its hands-on nature, though results vary across contexts [4, 20]. While these tools emphasize input affordances, they didn't address real-time, content-specific support within classroom settings. Researchers also seek ways to scaffold real-time classroom note-taking, including guiding students to take notes in structured formats or with predefined structures, such as concept maps [8, 58] or designing collaborative tools such as NotePals [14], LiveNotes [32], or NoteCoStruct [18] that reduce individual cognitive load by distributing note-taking responsibilities among group members. While collaborative notes are proven to be of higher quality [43], they may compromise individual agency since students lean towards relying on contributions made by their peers instead of personal involvement [12, 13, 19]. This dependence can weaken the personal encoding benefits of note-taking, thus reducing comprehension and retention [13].

Recently, AI has increasingly supported automated note generation, especially in complex contexts like clinical documentation [40, 61, 63] and online meetings [9]. While AI automation is advantageous in environments where efficiency is paramount, classroom note-taking differs in that it serves not only as information storage but as a critical encoding process [34]. Fully automated notes in classrooms may inhibit the encoding process, yet research on the ideal level of AI assistance in note-taking remains scarce. While a few recent studies started exploring using AI to generate an intermediate level of notes that users can further work on, providing insights into balancing user engagement and AI assistance [11], the impact of AI on real-time encoding and post-lecture storage or decoding processes in educational settings has yet to be systematically quantified and understood.

## 2.3 AI Assistance Dilemma in Cognitive Tasks

The assistance dilemma describes the challenge of balancing assistance with autonomy in learning, highlighting the need to provide optimal levels of support without fostering dependency [37]. In educational research, solutions to the assistance dilemma emphasize carefully calibrated scaffolding strategies to guide students' meta-cognitive skills [51] and adjust the level of help based on learners' skill and engagement levels [31]. For example, in early problem-solving stages, worked examples are frequently paired with minimal guidance, transitioning to increased feedback or step-by-step hints as learners advance, encouraging their independent reasoning [45].

In cognitive tutoring, studies indicate that varying levels of scaffolding are effective, yet these studies primarily pertain to predefined problem domains and structured worked examples, where they can effectively trace students' knowledge progression [45, 67]. This work extends the understanding of the AI assistance dilemma beyond structured learning environments by examining how different levels of AI support affect comprehension and engagement in the context of real-time note-taking, which involves open-ended, temporally constrained sense-making.

## 2.4 Design to Balancing AI assistance in Sense-making Activities

Recent research on using large language models (LLMs) for sense-making tasks in diverse scenarios provides insights into how to balance AI assistance and human engagement in cognitive and sense-making tasks [11, 15, 42, 46, 69].

Prior work has shown that AI can support user engagement more effectively when it provides partial assistance rather than completing the task entirely [15, 46]. For example, in writing tasks, offering full paragraph suggestions may reduce user agency, while sentence-level assistance effectively reduces the difficulty of phrasing without limiting creative freedom [15].

People also explored interactive mechanisms to help people engage more in sense-making activities when receiving AI assistance. For instance, in the MeetMap system, researchers introduced a temporary AI-generated content holding area, allowing users to decide which suggestions to use or ignore [11]. This encourages users to reflect on the content before accepting it, which can help maintain cognitive engagement [11, 69].

Additionally, the level of AI abstraction and its potential uncertainty are important design considerations [22, 24]. For instance, a resilient interface can continuously highlight critical information, increasing AI's abstraction of original content as needed [24].

The design of the three levels of AI assistance in this study draws insights from these designs, balancing the granularity of AI-provided content and the agency provided to users, using a holding area where users can further interact with the AI-generated content, and adjusting abstraction levels to support effective note-taking.

## 3 Study

As demonstrated in previous research, note-taking is vital for real-time understanding and sense-making during lectures, yet students often lack effective techniques, reducing their cognitive benefits. AI tools have been introduced to aid note-taking, but fully automated methods might compromise the advantageous cognitive processes. This challenge, often framed as the assistance dilemma, has been extensively studied in structured learning environments such as cognitive tutoring. However, its implications for real-time, open-ended sense-making activities like lecture note-taking remain underexplored.

To bridge this gap, we aim to investigate to what extent the AI assistance dilemma exists and how we might design human-AI collaborative mechanisms to address this challenge in a note-taking context with the growing prevalence of AI assistance. We address these questions through the design and evaluation of a system that supports three distinct levels of AI assistance, namely **Automated AI** - which provides slightly delayed AI-generated structured notes every 1-2 minutes **Intermediate AI** - which presents real-time summary blocks, and **Minimal AI** - which provides real-time transcript blocks.

We performed a within-subject experiment with 30 participants, where they watched lecture videos and took notes in the three setups. We introduce a post-test of the lecture to measure people's understanding, which is a signal of their cognitive engagement level during the lecture.

### 3.1 Experiment design

**3.1.1 Apparatus.** We designed the note-taking tasks with three distinct levels of AI assistance, drawing from prior research on human-AI collaboration tools for sense-making activities, as discussed in section 2.4. In those studies, the assistance level was differentiated based on the degree of abstraction and scaffolding provided by AI, as well as the level of user agency retained, emphasizing how manual effort is distributed between humans and AI [16, 24]. In the context of lecture note-taking in this research, the assistance levels reflect how much of the encoding process

AI automates and the degree of user autonomy in managing and engaging with content. Besides, the level of abstraction in such real-time tasks may also be related to the synchronicity users receive the AI assistance [42]. In this study, the assistance level apparatus is designed as follows:

- **Automated AI** : AI generates a structured note block around every two minutes. Users can incorporate these structured blocks into their own notes - the notes are shown in a granularity that is similar to the business usual AI note-taking tools <sup>23</sup>.
- **Intermediate AI** : AI produces turn-level summary blocks for each speaking turn. Users can select and incorporate these intermediate summary blocks into their notes.
- **Minimal AI** : AI provides a real-time transcript divided into transcript blocks after each completed speaking turn. Users can select and incorporate these transcript blocks into their notes.

The detailed system features and how the AI assistance differs in the three setups will be discussed in section 3.2

**3.1.2 Participants and Procedure.** We recruited 30 participants through mailing lists at the University of Michigan. We selected participants who indicated that they faced challenges in note-taking. The demographic information of participants is shown in Table 1.

Gender		Age	
Male	16	0 - 20 years old	9
Female	9	21 - 30 years old	21
Genderqueer	1		
Prefer not to say	4		
Ethnicity		Year in School	
Caucasian	7	Freshman	0
Asian	19	Sophomore	9
Other/Unknown	4	Junior	4
		Senior	12
		Master	1
		PhD	4

Table 1. Demographic Information of Participants

Each participant experienced all three AI assistance conditions in person, with the conditions—**Automated AI**, **Intermediate AI**, **Minimal AI**. The order of the conditions is counterbalanced across participants to mitigate order effects. Participants were randomly assigned to one of six possible sequences to ensure balanced exposure, as shown in Figure 2.

The experiment began with an introduction to the study, providing participants with a thorough explanation of the tasks. A demo was played for each system variant, after which participants engaged in a toy task designed to familiarize them with the interface and functionalities of each system.

For each task, participants were seated in a study room with a large display simulating a classroom setting. The display played a 10-minute lecture video, mimicking a live classroom environment where the instructor delivered the lecture. During this time, participants used their own computers to take notes in real time through the customized AI-assisted system. To replicate the demands of in-situ note-taking, the video was designed to be non-replayable.

<sup>2</sup><https://www.zoom.com/en/ai-assistant/>

<sup>3</sup><https://support.microsoft.com/en-us/office/use-copilot-in-microsoft-teams-meetings>



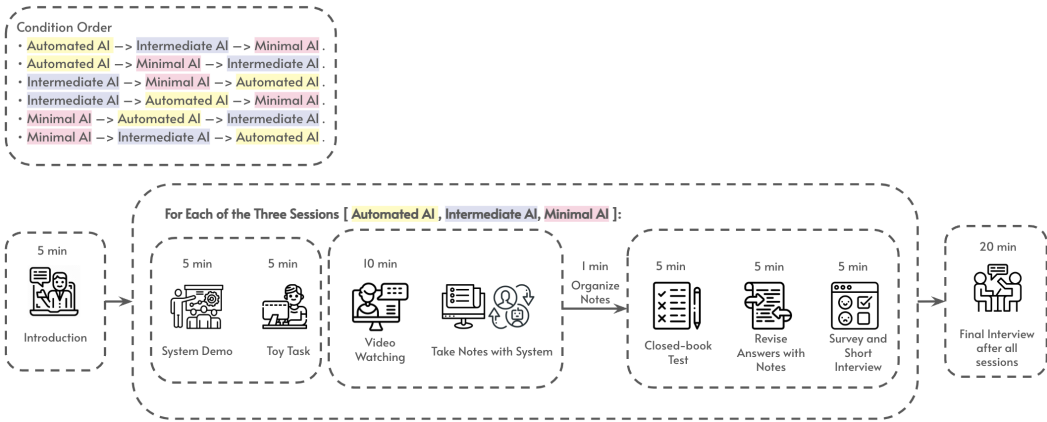


Fig. 2. **Study Design.** The study started with a 5-minute introduction outlining its objectives. Participants then experienced all three AI assistance conditions— **Automated AI** , **Intermediate AI** , **Minimal AI** , in a specified order. Each session began with a demo instruction and a toy task. Participants watched a 10-minute video while taking notes with the system, after which they had 1 minute to organize their notes. This was followed by a closed-book test lasting 5 minutes, after which participants revised their answers with their notes for an additional 5 minutes. Each session concluded with a survey and a brief interview lasting 5 minutes. A final interview was conducted after all sessions, lasting 20 minutes.

At the end of each video, participants were given one minute to finalize their notes before proceeding to a post-test, which included two multiple-choice questions (MCQs) and three open-ended questions (OEQs) <sup>4</sup>. Initially, participants completed the test without referring to their notes to assess the influence of the encoding process on learning outcomes [34]. They then submitted their responses and were asked to revisit and revise their answers with access to their notes, aligning with prior research on the storage paradigm, where notes serve as external storage for review and reinforcement [34]. Following the post-test, participants completed a post-task survey assessing their cognitive load and user experience with the system. Additionally, each task concluded with a 10-minute interview to capture immediate feedback and insights on the specific condition they had just completed.

Upon completion of all three tasks, participants participated in a final 20-minute comprehensive interview to discuss their comparative perceptions of the three AI assistance levels. Questions were tailored to explore participants' interactions with intermediate and structured AI-generated notes, their cognitive demands and agency during the process, and the trustworthiness of the AI.

In total, the study took approximately two hours per participant.

**3.1.3 Material.** We selected three lecture videos from YouTube, all presented by the same instructor and centered on an engaging, timely topic: large language models (LLMs). Each video spans approximately 8 minutes and 50 seconds to 9 minutes and 30 seconds. The topics covered are "Introduction to AI-Augmented Decision-Making," "Introduction to LLM Hallucination," and "Introduction to LLM model selection framework for your task," with the order fixed as listed. Importantly, the videos required no prior knowledge. Careful consideration was given to the selection and balance of video length and topic, minimizing potential differences in difficulty or familiarity and reducing these variables' impact on learning outcomes.

<sup>4</sup> Links to the study materials: <https://drive.google.com/file/d/17mWKqU3ydLI-dtcm2R-iwwrN7wq9rxdw>

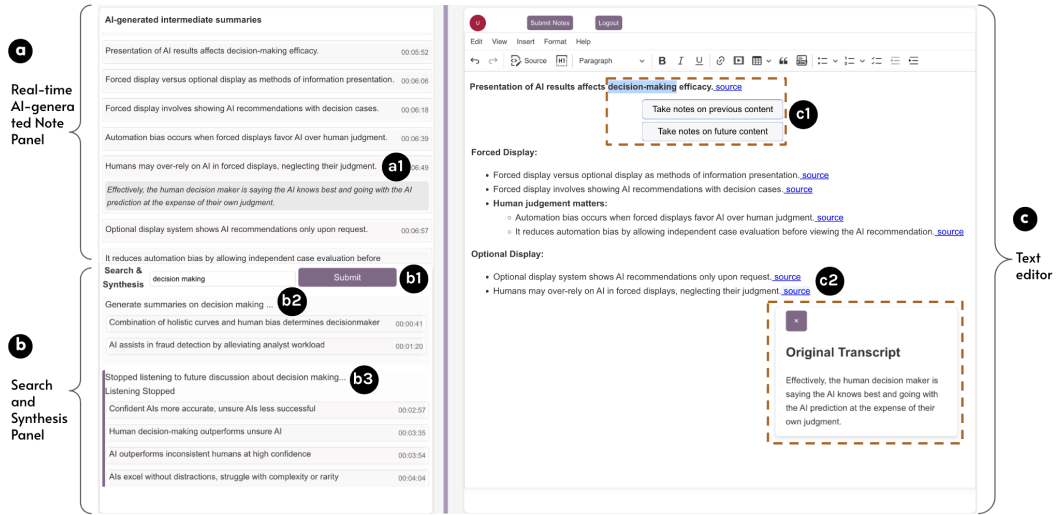


Fig. 3. **System Overview.** On the upper left is the REAL-TIME AI-GENERATED NOTE panel (a), where users can see AI-generated notes and click on them to see related transcripts (a1). On the lower left is the SEARCH AND SYNTHESIS PANEL (b), where users can interact with AI by requesting notes centered around a topic (b1) based on previous course content and receive notes as (b2). On the right is the rich Text-Editor (c), where users can create their own notes or, with the help of AI-generated notes, drag and drop the AI content from (a) or (b) to the editor. They can use in-context selection (c1) to request topic-related notes based on previous course content as (b1) does and also ask AI to continuously generate topic-related notes by listening to future course content and receive results as (b3). In addition, users can check original related transcripts for each AI-generated note by clicking on the source button (c2). This is the interface for **Intermediate AI**.

For each video, we designed a post-test comprising two multiple-choice questions (MCQs) and three open-ended questions (OEQs). An expert instructor in AI developed the tests, and their quality was validated in a pilot study involving four students who watched the videos, answered the questions, and provided feedback on any points of confusion. This feedback ensured that the questions were clear and answerable based solely on the lecture content without needing outside resources. Full details on video sources and test questions can be found in the supplementary materials <sup>4</sup>.

We opted not to include a pre-test in this study because the selected lecture videos are closely matched in length, complexity, and topic, minimizing video difficulty as a factor in learning outcomes. We excluded it to avoid lengthening the study since students had already experienced three conditions.

### 3.2 NoteCopilot: The Custom-Built AI-assisted Note-taking Tools

We developed NoteCopilot, an AI-assisted note-taking system offering adjustable levels of AI assistance - **Minimal AI**, **Intermediate AI**, **Automated AI**, each tailored to examine different degrees of cognitive assistance in note-taking.

**3.2.1 Common interaction across the three system variants.** The interface includes a REAL-TIME AI-GENERATED NOTE PANEL for proactive assistance, a SEARCH AND SYNTHESIZE PANEL for expressing specific user intents and retrieving personalized results, and a RICH TEXT EDITOR. The general interface and user interaction are shown in Figure 3.



The REAL-TIME AI-GENERATED NOTE PANEL (Figure 3(a)) displays AI-generated notes in real-time. Depending on the condition, the content in this panel varies from full transcripts (in **Minimal AI**) to intermediate summaries (in **Intermediate AI**) and structured notes (in **Automated AI**). This feature enables users to view AI-generated content as the lecture progresses, serving as the system's core AI assistance. Here, AI proactively provides content in a temporary middleware layer, allowing users to opt to integrate it into their notes or not. This design aligns with prior research that recommends a temporary holding space for AI content rather than direct insertion into the editor [11] as discussed in §2.4.

The TEXT EDITOR (Figure 3(c)) offers full note-taking functionality, supporting typing, selection, editing, and multimedia formatting. Users can drag and drop AI-generated content from panels (a) and (b) into the editor (c). When AI-generated content is dragged into the editor, a source button (c2) allows users to view the original transcripts associated with AI summaries, ensuring transparency and traceability of the information.

The above two panels form the system's core mechanism, where AI generates notes in real-time with varying granularity and timing, allowing students the flexibility to selectively incorporate AI-generated content into their notes.

In addition to proactive AI assistance, we offer users the flexibility to search for specific information as needed. The SEARCH AND SYNTHESIS PANEL (Figure 3(b)) enables users to interact with the AI by requesting targeted information on lecture topics. Users can submit a query (b1), and the AI provides responses that are aligned with the AI assistance levels based on their request (b2). Additionally, users can initiate searches directly within the editor (c). By selecting typed content, users can access a drop-down menu with two AI options: "Take notes on previous content" and "Take notes on ongoing content" (c1). The former generates AI notes based on previous lecture content, while the latter activates a listening mode, capturing any mentions of the selected topic in future content. Importantly, the granularity and abstraction of the AI responses in this panel correspond to the assigned condition: in the **Minimal AI** condition, search results return transcript blocks; in the **Intermediate AI** condition, they return summary blocks; and in the **Automated AI** condition, they provide structured notes.

The interaction design remains consistent across the three assistance levels, enabling users to interact with the AI panels similarly regardless of the specific condition. This consistency allows us to focus on examining how varying levels of AI processing and automation impact user engagement and learning.

**3.2.2 Three AI-assistance Levels.** While users interact with AI in a similar way, each system variant—**Minimal AI**, **Intermediate AI**, and **Automated AI**—offers unique levels of AI assistance. The three assistance levels reflect different extents to which AI automates the encoding process—the cognitive effort of interpreting, structuring, and recording information, and to what extent AI abstracts the information and adds its own interpretation to the content.

In the **Automated AI** condition, AI generates structured note blocks approximately every 1–2 minutes (Figure 4(n1)). This interval is chosen to provide users with timely content while allowing AI sufficient time to process and organize content into coherent structures. This design was assumed to reduce the users' note-taking encoding load and may enhance understanding by synthesizing information for them [17, 44].

In the **Intermediate AI** condition, AI generates intermediate summary blocks right after each speaking turn, as shown in Figure 3(a). These summaries offer a brief interpretation of lecture segments without a fully structured format, providing users with essential information that they can selectively integrate into their notes. The interval between summary blocks is around 15 seconds, where a natural sentence ends. Users can expand to check the original transcript of the

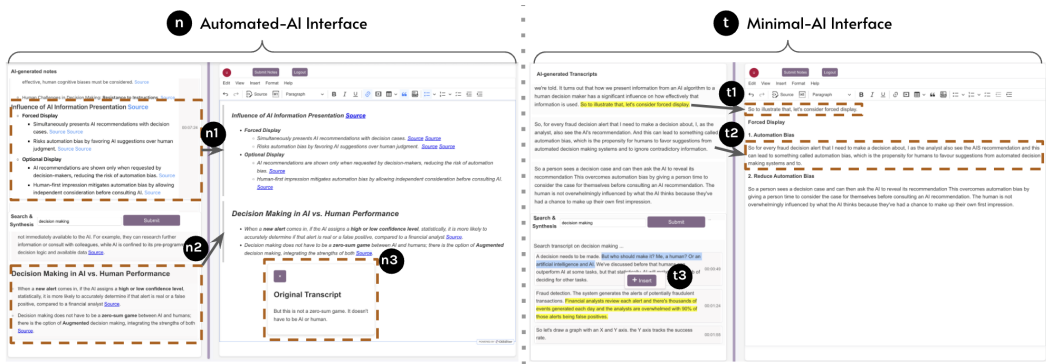


Fig. 4. **Automated AI and Minimal AI Interface.** On the left is the **Automated AI** INTERFACE (n), where users can drag and drop AI-generated notes alongside related transcripts from the **REAL-TIME AI-GENERATED NOTE PANEL** (n1) and **SEARCH AND SYNTHESIS PANEL** (n2) into the **TEXT-EDITOR PANEL**. Users can also expand related transcripts by clicking the source button (n3). On the right is the **Minimal AI** INTERFACE (t), allowing users to directly select text from the **REAL-TIME AI-GENERATED NOTE PANEL** (t1) and **SEARCH AND SYNTHESIS PANEL** (t3) for highlighting and insertion into the **TEXT-EDITOR PANEL**. Additionally, users can drag and drop entire transcript blocks into the **TEXT-EDITOR PANEL** (t2).

summary block by clicking the block Figure 3(a1). This condition leverages the cognitive benefits of “building blocks” [27]. By using intermediate summaries, this level of support is designed to encourage students to engage with the content actively, reducing the cognitive load associated with verbatim note-taking while still fostering their encoding process through selective integration and organization.

In the **Minimal AI** condition, AI provides real-time transcripts of the lecture, divided into “transcript blocks” based on turns, as shown in Figure 4(t). Users can incorporate these transcript blocks into their notes through drag-and-drop. They can also select the specific content on one transcript block and click an “insert” button to put it into the editor, shown in Figure 4(t3). This condition represents low AI assistance, where the AI merely acts as a passive transcription tool without interpretation or structuring, thereby requiring the user to handle all aspects of encoding and organization. This design mitigates the potential cognitive overload associated with fully manual note-taking [48], offering light-touch assistance that students can choose to engage with as needed.

While the **Real-Time AI-GENERATED NOTE PANEL** delivers AI-generated content proactively at predefined intervals, the **SEARCH AND SYNTHESIZE PANEL** empowers users to actively navigate the same AI assistance levels according to their own needs and points of interest. In this way, it complements the proactive panel by helping participants flexibly apply the AI assistance level to match their evolving information needs throughout the lecture. As shown in Figure 4, the content retrieved via the search panel adopts the same granularity and abstraction as the content shown in the AI-generated notes Panel. Specifically:

- **Minimal AI** : retrieves and surfaces raw transcript segments (Figure 4-t3).
- **Intermediate AI** : generates summaries based on relevant transcript chunks (Figure 3-a1).
- **Automated AI** : synthesizes structured notes from multiple segments (Figure 4-n2).

We decided against a no-AI control condition because participants can fully control and opt to disregard AI-generated content in any of the three conditions; additionally, the system permits

minimizing the AI panel to lessen visual distraction. Thus, users can maintain self-directed note-taking if desired.

**3.2.3 Design Considerations.** Our system design is intentionally constructed to reflect varying degrees of AI assistance and user cognitive engagement, aiming to empirically examine the AI assistance dilemma — the tension between automation benefits and user cognitive engagement risks.

We focus on proactive AI assistance as our primary design axis to simulate real-time note-taking support scenarios. Traditionally, learners actively engage in encoding by either transcribing content verbatim or constructing more abstractive summaries. Prior research has shown that higher degrees of information processing, such as synthesizing rather than simply copying, correlate with deeper cognitive engagement during note-taking [4, 58]. In designing our system, we intentionally align AI assistance with this natural human note-taking progression: by varying the degree of AI processing, the system can assist with this encoding process, specifically, by offering raw transcripts to support verbatim recording, concise summaries to promote selective abstraction, and structured notes to offload synthesis. However, we could hypothesize that AI-generated notes reduce the effort in note-taking, which may potentially reduce the depth of cognitive engagement of students in the *encoding process* of note-taking and influence comprehension. This AI assistance in the encoding stage may also shift the nature of users' effort in the *storage function* of note-taking. On the one hand, because AI generates structured notes, it may enhance review and understanding by synthesizing information for them. On the other hand, as AI generates more processed and synthesized notes, users may need to invest greater cognitive effort in reading, understanding, and critically evaluating these AI-generated outputs before using them. These are the nuanced influences introduced by varying levels of AI assistance during note-taking—both in encoding and storage—that we aim to empirically investigate in our study.

We deliberately balance several critical factors in designing the varied levels of AI assistance: (1) the need to provide timely, real-time AI support during the note-taking process; and (2) the necessity of maintaining comprehensive coverage of lecture content to ensure that AI-generated notes remain complete and reliable, regardless of their abstraction level. To achieve this, we control the granularity and completeness of content delivery across all conditions. Specifically, transcript and summary blocks are consistently segmented based on natural turns—a complete sentence or a set of related short sentences—ensuring comparability in chunk size and information coverage. The **Minimal AI** variant provides a real-time transcript. In the **Intermediate AI** variant, AI generates summaries immediately after each turn, averaging around 15 seconds per summary block. The difference between the two is that the summary blocks are easier to read, whereas transcript blocks are more transparent since they are true representations of the conversation. The **Automated AI** variant combines several turns into topic-based structured note blocks, with each note block between 1-2 minutes and traceable to the original transcript. Although the abstraction level varies, all AI-generated content offers a full representation of the lecture, ensuring comparability across the three system variants, as shown in Figure 5

**3.2.4 System Implementation.** We built NoteCopilot with a React.js frontend and Django backend. For experimental consistency, all AI-generated notes in **Intermediate AI** and **Automated AI** were pre-generated and timestamped to ensure identical content for all participants, avoiding variability in live AI output.

In **Minimal AI**, Azure Speech-to-Text transcribed lectures in real-time, displaying text word-by-word. We use GPT-4o to predefine sentence-level turn segmentation (Figure 5), controlling word count and timing. At each segmentation point, the system wrapped prior text into a transcript block and started a new one, ensuring consistent chunking across participants. In **Intermediate AI**,

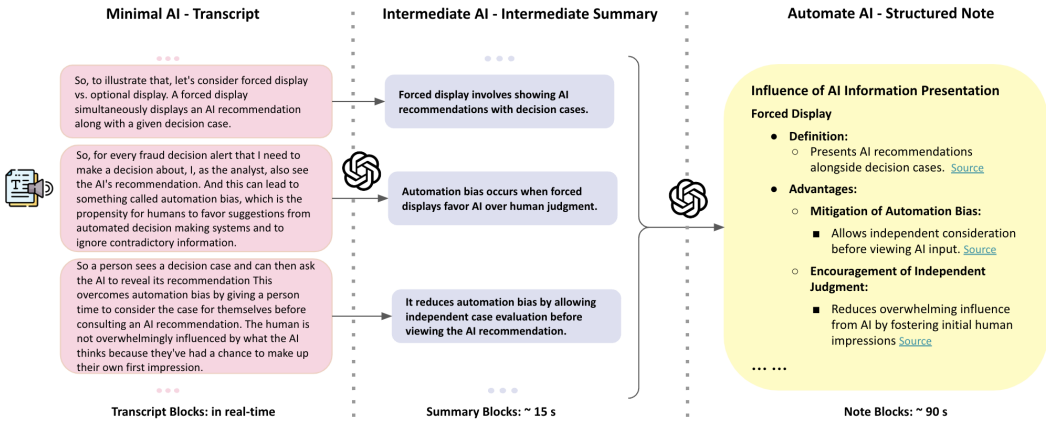


Fig. 5. **The granularity of AI notes in the three conditions.** From left to right, the AI-generated results of three systems—**Minimal AI**, **Intermediate AI**, and **Automated AI**—illustrate the topic of FORCED DISPLAY with increasing levels of AI assistance. The **Minimal AI** system presents users with raw AI-chunked transcripts, providing unfiltered information. In contrast, the **Intermediate AI** system offers concise summaries of each transcript chunk, enhancing clarity and facilitating user comprehension. Finally, the **Automated AI** system consolidates all short summaries into structured notes that include definitions and advantages of FORCED DISPLAY. This organized presentation provides detailed information for each point, allowing users to reference related source transcripts for further context.

these segments were passed to the LLM model to produce concise summaries, which were shown immediately after each lecture turn, simulating real-time generation. In **Automated AI**, GPT-4o was used to group multiple transcript segments into thematic chunks and generated structured notes synthesizing these higher-level units, pre-loaded and “replayed” along the lecture timeline.

For the SEARCH AND SYNTHESIZE PANEL, all conditions used the same prompting approach: the full transcript and user query were input to GPT-4o, with instructions tailored to the assistance level. **Minimal AI** returned relevant raw transcript segments; **Intermediate AI** returned synthesized summaries; **Automated AI** returned higher-level structured notes.

To produce a well-structured, high-quality summary and notes, researchers devoted significant effort to crafting and iterating the prompt to ensure the resulting notes appear reasonable and organized, as illustrated by the AI-generated transcript, summary, and notes shown in Figure 5. To ensure content accuracy, all generated summaries in **Intermediate AI** and notes in **Automated AI** were manually reviewed by the researchers to confirm that they captured the intended lecture and to verify its alignment with the lecture to rule out factual errors.

#### 4 Data Analysis and Measurement

After collecting the data from N=30 participants, we proceeded to analyze the data. We performed a series of quantitative analyses using several statistical models, as discussed below, to answer the following two questions:

- How do different levels of AI assistance affect learning outcomes, cognitive load, note-taking behaviors, and usability?

We use our interview data to provide further explanations on the following:

- Why do these effects occur? How do students perceive the usefulness of AI, and how do they prefer to interact with different AI assistance levels?

## 4.1 Quantitative Outcome Measures

**4.1.1 Measurement of Understanding.** To assess users' understanding of the lecture, we graded students' post-test answers across two rounds of submissions. The first round, completed without reference to notes, measured students' real-time understanding of the lecture content, referred to as the POST-TEST SCORE. The second round allowed students to revise their initial answers using their notes, called the REVISION SCORE. This score represents the understanding achieved after reviewing and refining their responses with their notes.

We developed a rubric to evaluate answers to open-ended questions (OEQ). One author initially created the rubric, with each question worth up to 4 points. Two authors then independently graded a sample (25%) of the students' answers using this rubric, blind to the experimental condition. The initial inter-coder reliability was 86.7%. After discussion and consensus on the grading standards, they divided the remaining answers for individual grading. The rubric and example students' answers are appended in the supplementary materials <sup>4</sup>.

**4.1.2 Measurement of Cognitive load.** We assessed cognitive load by employing questions tailored to evaluate overall mental effort [47], and we adapted scales from earlier studies [36, 41] to measure intrinsic (IL), extraneous (EL), and germane (GL) cognitive loads. We opted not to use a general task load measurement, such as NASA TLX, and instead chose a scale specifically designed to measure the three types of cognitive load. This approach allows us to examine how each type of cognitive load is influenced by the three conditions, providing insights into how varying levels of AI assistance can enhance beneficial aspects of cognitive load, like germane load, while mitigating the more negative aspects, such as extraneous load.

*Overall Mental Effort:* Please rate your overall mental effort (from 1 - 9). Scale 1 = very, very low mental effort, and Scale 9 = very, very high mental effort [47]

*Intrinsic Load:* This type of cognitive load refers to the cognitive effort required to understand the inherent complexity of the learning content [36]. It depends on the difficulty of the material itself, independent of the instructional methods. Questions include:

- Q1: "The content explained in this lecture is very complex."
- Q2: "This lecture contains many terms and concepts that are unfamiliar to me."

The score ranged from 1 (strongly disagree) to 7 (strongly agree).

*Extraneous Load:* Extraneous load is the cognitive burden imposed by the instructional methods or tasks, which arise from unnecessary or confusing elements in the learning environment [36]. Questions related to extraneous load include:

- Q3: "The instructor failed to explain concepts clearly."
- Q4: "Manual note-taking required a lot of my effort."
- Q5: "Interacting with the system (e.g., using AI-generated notes) required significant effort."
- Q5: "Creating notes with AI assistance also required significant effort."
- Q6: "Reading AI-generated notes required significant effort."

The score ranged from 1 (strongly disagree) to 7 (strongly agree).

*Germane Load:* Germane load is the cognitive effort devoted to processes that directly support learning, understanding, and knowledge construction [36]. This is a type of beneficial load that helps deepen comprehension and consolidate knowledge. Questions include:

- Q7: "I was fully engaged in this lecture."
- Q8: "Note-taking with AI assistance enhanced my understanding of the lecture."
- Q9: "Reading AI-generated notes enhanced my understanding of the lecture."

The score ranged from 1 (strongly disagree) to 7 (strongly agree).

**4.1.3 Note-taking Behaviors and Outcomes.** Throughout the experiment, we systematically logged users' interactions with the interface. This included actions on the REAL-TIME AI-GENERATED NOTE PANEL, SEARCH AND SYNTHESIZE PANEL, and TEXT-EDITOR. We also recorded the content created in the Text Editor, distinguishing between notes generated by AI and those created by users. By analyzing this log data, we aim to gain a detailed understanding of how participants utilized the AI assistance in their note-taking process and how this influenced their sense-making process.

*Note-Taking Behaviors.* The study tracked several note-taking behaviors, which included:

- **CLICKING ON THE TEXT EDITOR:** Frequency of checking the content in the text editor
- **TYPING IN THE EDITOR:** Frequency of typing actions in the text editor
- **DROPPING AI-GENERATED CONTENT:** Inserting AI-generated content into the editor through drag and drop or the 'insert' button
- **QUERY SEARCH:** Searching via direct queries in the SEARCH AND SYNTHESIZE PANEL
- **IN-CONTEXT SEARCH FOR PREVIOUS CONTENT:** Searching for previously discussed content around the selected texts in the text editor
- **IN-CONTEXT SEARCH FOR UPCOMING CONTENT:** Searching for future content around the selected texts in the text editor

*Note Outcome Measures.* The outcome of the note-taking process was measured based on NOTE QUANTITY:

- **TOTAL WORDS IN THE EDITOR:** The sum of all words, including both AI-generated content used by participants and manually typed content.
- **MANUALLY TYPED WORDS:** The count of words manually typed by participants in the editor.

This study only measured the QUANTITY OF NOTES and did not assess NOTE QUALITY. We acknowledge this as a limitation of our research and will address it in the discussion.

**4.1.4 Usability.** Usability was measured through a series of subjective usability questionnaire items, including scores for system ease of use, satisfaction, future usage intention, etc.

## 4.2 Interview Analysis

To understand students' perceptions of the usefulness and limitations of the three levels of AI assistance in note-taking, we conducted a thematic analysis of interview transcripts [6]. Initially, two researchers independently reviewed, commented on, and coded the transcripts. They developed 203 initial codes in this phase. They then discussed their findings to reach a consensus, after which one researcher refined the codes and identified key themes across the transcripts. All disagreements between coders were resolved through discussion. We uncovered 8 themes related to the benefits and limitations of each condition, 3 themes on how students used different AI assistance levels in note-taking, and 7 themes covering students' perceptions of agency, intention expression, needs, and trustworthiness. From these themes, we derived 5 high-level findings, which focused on AI's impact on students' learning, when and how students expressed their intentions toward the AI, and students' perceptions of agency and cognitive load.

## 5 Finding

### 5.1 RQ1: How do different levels of AI assistance affect comprehension, cognitive load, note quality, and usability?

In RQ1, we evaluated how the different levels of AI assistance (Automated AI, Intermediate AI, Minimal AI) influence students' learning and understanding, cognitive load, and usability.



**5.1.1 Understanding Levels.** To evaluate the impact of different levels of AI assistance (Automated AI, Intermediate AI, Minimal AI) on students' understanding of the lecture, we conducted mixed-effects linear regression analyses on both post-test and revision scores. In these models, we included CONDITION as a fixed factor, controlled for VIDEO as a covariate, and incorporated interaction terms between CONDITION and VIDEO. Random intercepts for INDIVIDUAL PARTICIPANTS were added to account for within-subject variability.

Coefficient (Std. Err)	Post-Test Score	Revision Score
Intercept (Automated AI)	9.550 (0.980)	12.033 (0.922)
Condition: Intermediate AI	4.172 (1.315) **	2.848 (1.190) *
Condition: Minimal AI	2.702 (1.339) *	0.367 (1.216)
Video: Video 2	2.326 (1.338)	0.565 (1.216)
Video: Video 3	-0.298 (1.368)	-0.278 (1.190)
Condition: Intermediate AI × Video 2	-1.146 (1.982)	-0.382 (1.806)
Condition: Minimal AI × Video 2	-2.197 (1.974)	2.528 (1.813)
Condition: Intermediate AI × Video 3	-0.136 (1.933)	1.339 (1.779)
Condition: Minimal AI × Video 3	-0.993 (1.997)	-0.538 (1.805)
Group Var	2.216 (0.603)	4.096 (0.894)

Table 2. Mixed-effects linear regression model results for POST-TEST SCORES and REVISION SCORES, examining the fixed effects of CONDITION and VIDEO with random intercepts for USER. The model controlled for the influence of VIDEO and included interaction terms between CONDITION and VIDEO. Results indicate that the Intermediate AI condition led to significantly higher POST-TEST SCORES and REVISION SCORES compared to Automated AI, with Minimal AI showing a higher POST-TEST SCORE than Automated AI, and no significant interactions between CONDITION and VIDEO. Significance codes: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

The analysis of POST-TEST SCORES revealed a significant main effect of Intermediate AI on learning outcomes. Specifically, students in the Intermediate AI condition (Mean = 13.88, SD = 2.34) scored significantly higher than those in the Automated AI condition (Mean = 10.28, SD = 3.75), with a coefficient of 4.17 ( $p = 0.002$ ). The Minimal AI condition also showed a significant improvement over Automated AI, with a coefficient of 2.70 ( $p = 0.044$ ). Post-hoc Tukey HSD analysis further confirmed these differences, showing that Intermediate AI outperformed Automated AI with a mean difference of 3.61 ( $p < 0.001$ ), and significantly outperformed Minimal AI with a mean difference of 2.03 ( $p = 0.034$ ).

The analysis of REVISION SCORES revealed a significant main effect of Intermediate AI on post-review learning outcomes. Specifically, students in the Intermediate AI condition (Mean = 15.29, SD = 2.99) scored significantly higher than those in the Automated AI condition (Mean = 12.12, SD = 2.76), with a coefficient of 2.85 ( $p = 0.017$ ). The Minimal AI condition showed no significant improvement over Automated AI (Coef = 0.37,  $p = 0.763$ ). Post-hoc Tukey HSD analysis supported these findings, showing that Intermediate AI outperformed Automated AI with a mean difference of 3.17 ( $p = 0.0003$ ), and significantly surpassed Minimal AI with a mean difference of 2.00 ( $p = 0.031$ ).

These findings reveal that Intermediate AI better supports real-time learning (POST-TEST SCORES) compared to Automated AI. However, contrary to our initial expectations, Intermediate AI also outperforms Automated AI in post-review tasks (REVISION SCORES). Interestingly, despite the opportunity for students to review with notes Automated AI, this condition did not yield better learning outcomes.

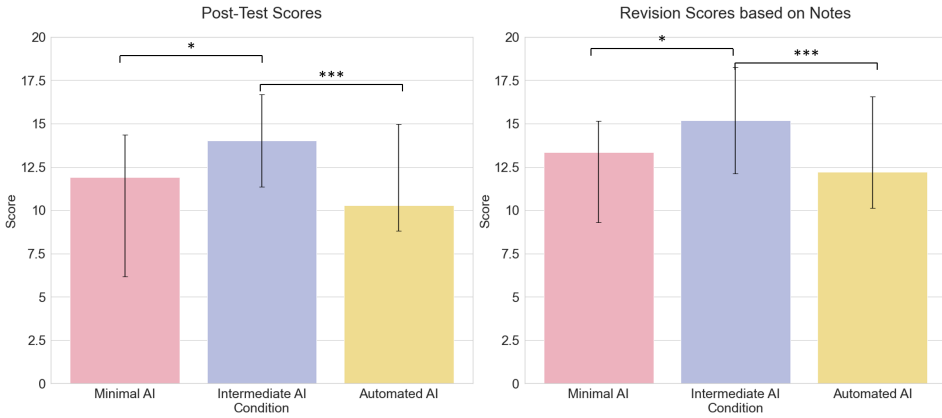


Fig. 6. **Post-hoc Tukey HSD analysis comparing post-test and revision scores across conditions.** The left bar graph shows the POST-TEST SCORES, where the Intermediate AI condition significantly outperformed both the Automated AI and Minimal AI. The right bar graph displays the REVISION SCORES, with similar trends observed.

No significant difference between VIDEO and the learning outcome was found. Hence, we argue that the difficulty levels of the videos are comparable, so we excluded video differences from further analyses in this paper.

**5.1.2 Cognitive Load.** We analyzed cognitive load across different AI-assisted note-taking conditions (Automated AI, Intermediate AI, and Minimal AI) using repeated-measures ANOVA, focusing on dimensions of mental effort, intrinsic, extraneous, and germane cognitive load.

**Overall Mental Effort.** The repeated-measures ANOVA for overall mental effort showed no significant main effect of condition ( $F(2, 58) = 1.67, p = 0.197$ ), indicating similar levels of mental effort across conditions.

**Intrinsic Cognitive Load.** For intrinsic cognitive load, measured by perceived complexity and unfamiliarity with lecture content, no significant main effect of condition was observed. These findings suggest that intrinsic load was not significantly impacted by the AI assistance levels, aligning with cognitive load theory, which posits that intrinsic load is driven primarily by the inherent complexity of the content rather than instructional conditions.

**Extraneous Cognitive Load.** Significant effects were observed for extraneous cognitive load across conditions, particularly in manual effort, AI engagement, and reading AI notes. For the effort in manual note-taking, Minimal AI (Mean = 3.77, SD = 1.43) required significantly more effort than Automated AI (Mean = 2.50, SD = 1.68,  $F = 4.39, p = 0.017$ ), indicating that manual note-taking without AI support increases extraneous load. Similarly, for engaging with the system, Minimal AI (Mean = 3.90, SD = 1.65) required significantly more engagement effort than Automated AI (Mean = 2.87, SD = 1.50,  $F = 4.35, p = 0.017$ ). For reading AI-generated notes, Minimal AI (Mean = 4.57, SD = 1.61) required significantly higher reading effort than both Intermediate AI (Mean = 3.07, SD = 1.70,  $p = 0.002$ ) and Automated AI (Mean = 2.93, SD = 1.66,  $p = 0.0007$ ) ( $F = 9.30, p < 0.001$ ).

Cognitive Load Question	Automated AI	Intermediate AI	Minimal AI	F (Pr > F)
OVERALL: Please evaluate your mental effort in the task	4.60 (2.08)	4.53 (1.74)	5.17 (1.39)	1.6716
INTRINSIC: The learning content covered in this lecture was very complex	3.03 (1.59)	2.53 (1.17)	3.00 (1.14)	1.9618
INTRINSIC: The lecture included many terms and concepts that were unfamiliar to me	3.13 (1.61)	3.13 (1.63)	3.57 (1.65)	1.1086
EXTRANEIOUS: The instructor did not explain the concepts clearly in this lecture	2.40 (1.28)	2.30 (1.21)	2.10 (1.03)	0.7346
EXTRANEIOUS: It requires a lot of effort for me to write notes by myself.	2.50 (1.68)	2.90 (1.69)	3.77 (1.43)	<b>4.3899*</b>
EXTRANEIOUS: Engaging with the system (e.g., using the AI-generated notes) took a lot of effort from me.	2.87 (1.50)	3.17 (1.18)	3.90 (1.65)	<b>4.3497*</b>
EXTRANEIOUS: It requires a lot of effort for me to create notes with the AI's help.	2.87 (1.53)	2.83 (1.23)	3.93 (1.76)	<b>4.7137*</b>
EXTRANEIOUS: It requires a lot of effort for me to read the AI-generated notes.	2.93 (1.66)	3.07 (1.70)	4.57 (1.61)	<b>9.2956***</b>
GERMANE: I am fully engaged during the lecture.	4.33 (1.53)	4.77 (1.52)	4.50 (1.61)	1.0331
GERMANE: Taking notes (with the assistance of AI) enhanced my understanding of the lecture.	4.93 (1.11)	5.23 (1.04)	4.43 (1.52)	<b>4.1359*</b>
GERMANE: Reading the AI-generated notes enhanced my understanding of the lecture.	5.23 (1.10)	4.83 (1.18)	4.23 (1.50)	<b>4.4918*</b>

Table 3. Repeated Measures ANOVA Results for Cognitive Load Questions across Conditions with Mean (SD) for Each Condition

*Germane Cognitive Load.* For germane cognitive load, significant effects were observed for both questions related to the interaction with AI. The question, "Taking notes (with the assistance of AI) enhanced my understanding of the lecture," showed a significant effect of condition ( $F = 4.14$ ,  $p = 0.021$ ), with **Intermediate AI** (Mean = 5.23, SD = 1.04) significantly outperforming **Minimal AI** (Mean = 4.43, SD = 1.52,  $p = 0.038$ ). Similarly, "Reading the AI-generated notes enhanced my understanding of the lecture," yielded a significant effect ( $F = 4.49$ ,  $p = 0.015$ ), with **Automated AI** (Mean = 5.23, SD = 1.10) significantly higher than **Minimal AI** (Mean = 4.23, SD = 1.50,  $p = 0.009$ ).

In summary, **Automated AI** is optimal for lowering extraneous load, while **Intermediate AI** supports germane cognitive load by using AI-generated Notes. Conversely, **Minimal AI** demands greater manual effort, showing higher extraneous load and challenging cognitive resources.

**5.1.3 Note-taking behavior.** We analyzed the note outcomes and note-taking behaviors across different AI-assisted note-taking conditions using repeated-measures ANOVA.

There are notable differences in average total note counts across different AI conditions, as shown in Figure 7. The **Automated AI** condition yielded a significantly higher total note count (Mean=300.68, SD=160.52) compared to both the **Intermediate AI** condition (Mean=170.46 SD=70.21) and the **Minimal AI** condition (Mean=143.69, SD=86.93). ANOVA results confirmed a significant effect of AI condition on note volume ( $F = 16.66$ ,  $p < 0.001$ ).

To further understand these differences, we classified users into three general note-taking patterns based on their approach to AI assistance: Fully AI Mode, where users only used AI-generated notes

without adding their own; Mixed AI+Human, where users combined AI notes with their manual entries; and Fully Manual, where users relied solely on self-generated notes. The distribution of these strategies across conditions reveals distinct behavioral tendencies. In the **Intermediate AI** condition, over half of the users adopted the Mixed Mode, integrating AI-generated notes with their manual inputs. A clear divergence emerged in the **Automated AI** condition: 10 users chose to take their notes manually, while 8 relied solely on AI. In the **Minimal AI** condition, users predominantly engaged in fully manual note-taking, with 19 users relying exclusively on human-generated notes and 10 employing a mixed approach.

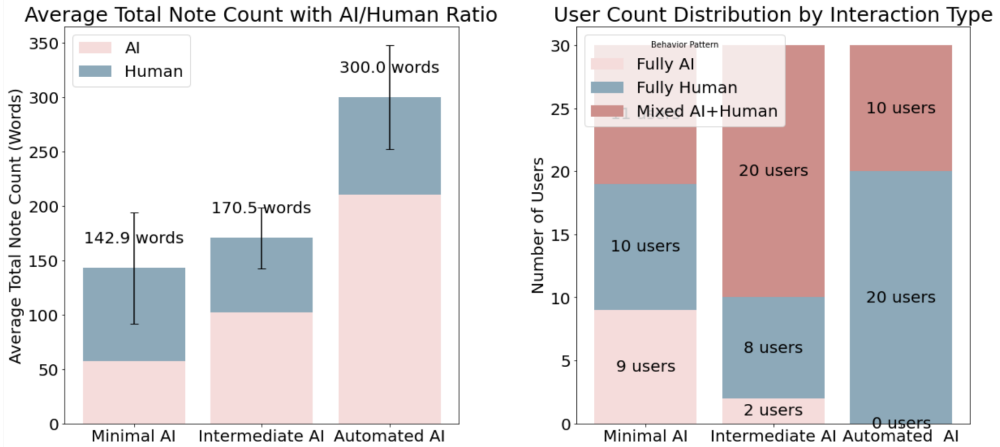


Fig. 7. **Comparison of note-taking behaviors across conditions with varying levels of AI assistance.** The left chart shows the average total note count for each condition, divided into AI-generated content and manually typed content, with error bars indicating standard deviation. On the right, a stacked bar chart illustrates the distribution of users by interaction type (fully AI-generated, mixed AI-human, or fully manual) across conditions, with labels indicating the number of users per type.

Further analysis of interaction behaviors in each condition offered insights into users' note-taking engagement. Users in the **Intermediate AI** condition frequently interacted with AI-generated notes, as indicated by the highest click counts in the text editor (Mean = 37.41, SD = 24.86) compared to the **Minimal AI** condition (Mean = 28.14, SD = 18.47) ( $p = 0.049$ ). Typing frequency was also higher in the **Intermediate AI** condition (Mean = 37.59, SD = 13.33), significantly surpassing **Minimal AI** ( $p = 0.008$ ). AI content "drops" (when AI-generated text was selectively added to the editor) were also most frequent in the **Intermediate AI** condition (Mean = 8.03, SD = 7.51), demonstrating a more active and curated approach than **Automated AI** ( $p < 0.001$ ) and the **Minimal AI** condition ( $p < 0.001$ ). Interaction with initiating a query and constantly listening to the upcoming content was notably higher in the **Intermediate AI** condition (Mean = 7.38, SD = 30.44), compared to that in **Automated AI** (Mean = 1.93, SD = 4.96) ( $p < 0.001$ ).

Are there more AI drop behaviors in **Intermediate AI** than in **Automated AI** due to different frequencies of AI content generation? To further unpack the differences in user engagement with AI-generated content across conditions, we conducted an additional analysis focusing on AI content use behavior. Specifically, we examined (1) the total number of knowledge units dropped into the editor, and (2) the number and proportion of knowledge units that were subsequently modified by users. Here, we define a "knowledge unit" as a semantic segment of AI-generated content: in **Minimal AI**, this corresponds to a transcript block; in **Intermediate AI**, a summary block;

Behavior	Minimal AI	Intermediate AI	Automated AI	F (Pr>F)
<b>Manual type words</b>	110 (83.58)	86.58 (56.01)	94.46 (61.62)	F = 1.42
<b>Click Text Editor</b>	28.14 (18.47)	37.41 (24.86)	32.07 (24.46)	F = 3.18 *
<b>Times for Typing words</b>	30.21 (15.59)	37.59 (13.33)	28.97 (12.12)	F = 0.86
<b>Drop AI Content</b>	1.14 (1.94)	8.03 (7.51)	3.00 (2.93)	F = 4.58 *
<b>Search through query</b>	0.66 (1.26)	0.55 (1.59)	0.52 (1.30)	F = 0.66
<b>In-context search</b>	0.24 (0.64)	0.14 (0.44)	0.00 (0.00)	F = 1.42
<b>Listen upcoming content</b>	N/A	7.38 (30.44)	1.93 (4.96)	N/A

Table 4. Summary of Note-Taking Behaviors During the Lecture Stage Across Conditions, with ANOVA Results for Each Behavior

and in **Automated AI**, a source-linked knowledge point rather than the full blocks, as illustrated in Figure 5. We further analyzed the extent to which users modified these AI-generated knowledge units after dropping them into the editor. Modification counts were computed by comparing the AI-generated content dropped into the editor with the final version of the user's notes.

Metric	Minimal AI	Intermediate AI	Automated AI	F(Pr>F)
Total AI content drops (normalized by knowledge units)	1.14 (1.94)	8.03 (7.51)	10.03 (4.89)	F=57.19 ***
Modified knowledge units (count)	0.30 (0.46)	3.53 (2.04)	0.76 (0.93)	F=66.37 ***
Modified knowledge units (% of dropped)	26.3%	40.0%	8.0%	F=23.60 ***

Table 5. Normalized analysis of AI content use across conditions. To account for the varying frequency and granularity of AI-generated content, we normalize the total AI content drops by the number of knowledge units. We further report the number and proportion of knowledge units that were subsequently modified by users, estimated by comparing the dropped AI content with the final editor content.

As shown in Table 5, we found that the normalized number of AI knowledge units dropped into the editor was comparable between the **Intermediate AI** (Mean = 8.03, SD = 7.51) and **Automated AI** (Mean = 9.41, SD = 4.89) conditions ( $p = 0.07 > 0.05$ ). However, users in the **Intermediate AI** condition demonstrated substantially higher modification behaviors. On average, they modified 3.53 units compared to 0.77 in **Automated AI** ( $p < 0.001$ ), and 0.30 in **Minimal AI** ( $p < 0.001$ ). These findings show some evidence that **Intermediate AI** encouraged more active refinement of AI-generated content, whereas **Automated AI** led users to more passively accept AI outputs with minimal modification. We acknowledge the limitations of our current analysis approach. While we normalized the number of AI knowledge units and captured the count of modified units, our method only reflects the end-state differences in user edits. It does not capture the dynamic editing processes that may occur during note-taking, such as additions, reorganization, or deletions of AI-generated content.

**5.1.4 Usability.** We used repeated measures ANOVA to assess usability across different levels of AI assistance, with post-hoc Tukey HSD tests to identify specific pairwise differences between conditions.

In the **Automated AI** condition, participants rated the perceived quality of AI-generated notes significantly higher (Mean = 5.20, SD = 1.27) compared to the **Minimal AI** condition (Mean = 4.20, SD = 1.61), with  $F = 3.95, p = 0.025$ . Additionally, for coverage of important topics, **Automated AI** (Mean = 5.50, SD = 1.11) was rated significantly higher than **Minimal AI** (Mean = 4.17, SD = 1.66)

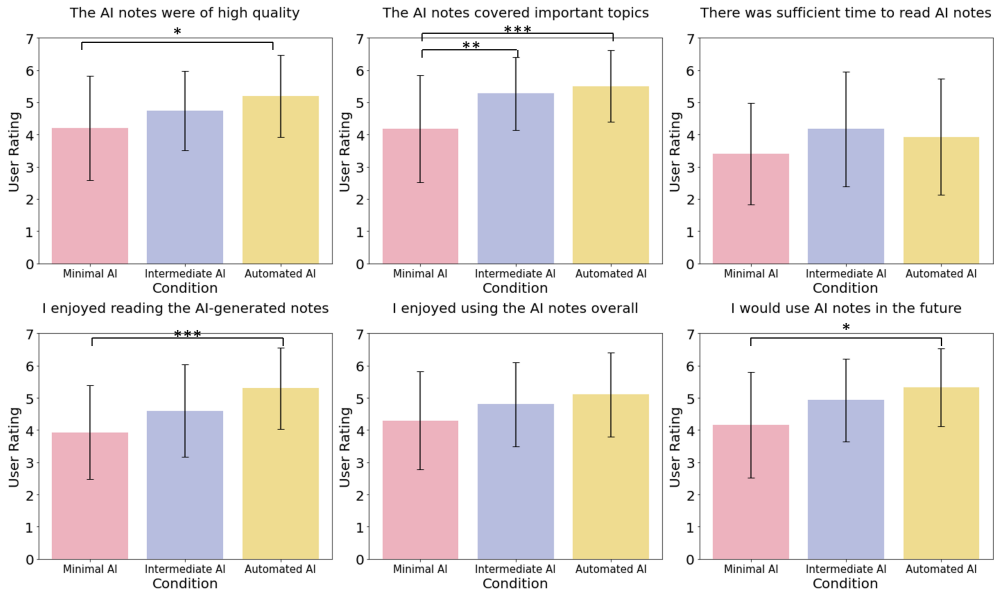


Fig. 8. Usability Rating. This figure presents the mean usability ratings across three conditions—Minimal AI, Intermediate AI, and Automated AI—for six questions assessing different aspects of the AI-generated notes. Tukey's HSD pairwise comparisons were conducted after the repeated-measure ANOVA.  $p$ -values are marked as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . The non-significant difference was annotated as NS.

( $F = 9.50 < 0.001$ ). These results suggest that users found Automated AI's output of notes more effective in capturing key points. Enjoyment in reading AI-generated notes was also substantially higher in the Automated AI condition (Mean = 5.30, SD = 1.26) compared to Minimal AI (Mean = 3.93, SD = 1.46) ( $F = 8.18, p < 0.001$ ). Furthermore, the intention for future use was notably higher in the Automated AI condition (Mean = 5.33, SD = 1.21) compared to Minimal AI (Mean = 4.17, SD = 1.64) ( $F = 5.62 = 0.006$ ).

In the Intermediate AI condition, participants rated coverage of important topics significantly higher (Mean = 5.27, SD = 1.14) than in Minimal AI (Mean = 4.16, SD = 1.66) ( $p = 0.005$ ). Besides, Minimal AI (Transcript) condition shows a higher preference for taking manual notes (Mean = 3.80, SD = 1.56) than those in the Automated AI (Auto-Note) condition (Mean = 2.63, SD = 1.45), with  $p = 0.01$ , as shown by post-hoc Tukey HSD tests. This suggests that users prefer manual note-taking more when AI is seen as unusable. We didn't see other difference in the usability rating between Intermediate AI and Minimal AI or between Intermediate AI and Automated AI.

In summary, Automated AI demonstrates optimal usability, significantly enhancing perceived quality, topic relevance, and enjoyment in reading AI-generated notes. Intermediate AI offers a balanced usability experience, showing improved coverage of important topics over Minimal AI, though it does not reach the high preference and engagement levels of Automated AI. Conversely, Minimal AI has lower usability ratings, with a higher preference for manual note-taking, suggesting that increased manual effort may detract from the perceived utility and enjoyment of AI-assisted notes.



## 5.2 RQ2. How Do Students Perceive AI's Helpfulness in Note-Taking?

For RQ2, we conducted a qualitative analysis to explore how students perceive the helpfulness of AI in the note-taking process, focusing on when specific AI conditions were considered useful or not and why certain conditions were ultimately less utilized despite initial perceptions.

**5.2.1 Intermediate AI** *enhanced students' learning by fostering active engagement with synchronous and digestible information.* Students found the summaries valuable for two main reasons. First, the summaries scaffold their cognitive engagement. Beyond serving as a "constant flow" to help them stay focused on lecture content in real-time (P13, P15), they felt that interacting with summary blocks actively deepened their understanding. This involved *"quickly glance over the summaries to check if key points were there"*(P5), *"which summaries are relevant and which ones they need"*(P3), and *"trying to put it under my own structure"*(P1). As P3 said, *"I'm just gonna drag the summaries of the last few things that you talked about onto here, and then I can kind of piece that together, and then that catches me up at least a little bit more than I normally would, where I would just miss that whole chunk of information."* As a result, students reported substantial learning gains in the **Intermediate AI** condition (P1, P13, P15).

The concise, digestible format of the **Intermediate AI** summaries makes engagement possible in real-time lectures. P2 remarked that the summaries helped them focus on the lecture by breaking down information into *"digestible chunks that were easy to put into the notes."* The granularity of the summaries matched many students' natural note-taking styles, allowing seamless integration into their notes. As students highlighted, the summaries consisted of short sentences they would typically write, enabling them to *"catch up by just dragging and dropping the summaries"* (P15) and do so *"without needing to add further details"* (P20).

**5.2.2 The Automated AI** *was perceived as beneficial for its well-structured content, but it also led to feelings of disengagement and inflexibility.* Many students perceived **Automated AI** as valuable during lectures due to their clear and logical structure. For instance, P3 and P13 noted that the structured format *"organized and contextualized the information better than other conditions."* The structure allowed P4 to *"work at their own pace and reference it later."* Students also highlighted the value of structured notes for post-lecture review and information retrieval. For example, P12 used the notes to confirm their understanding of concepts, while P15 used them to recall connections between topics. Due to the clear structure, students *"knew exactly where they could find a piece of information"* (P11) and could *"easily find a relevant bullet point by scrolling through the notes"* (P4).

However, structured content and the seemingly complete nature of notes are perceived to lead to less cognitive engagement. P3 admitted to *"zoning out"* at times, despite recognizing that the topic required their full attention. Some students felt they retained only a general impression of the course content due to over-relying on the notes, as noted by P5, *"I was struggling with answering the questions because I thought notes would help, but then I couldn't remember where it is. I felt like some sort of disconnection."* Lastly, the granularity of the notes presented challenges; P2 found it difficult to work with, commenting, *"everything was displayed in a giant chunk,"* which hindered their ability to integrate information into their own notes.

**5.2.3 Minimal AI** *has a higher cognitive load but also retains the highest level of trust.* Many students experienced the highest cognitive load during the **Minimal AI** condition. Firstly, the verbal text made it challenging to parse and transfer information (P13). Secondly, students felt distracted by the constant need to switch between interpreting the transcript, editing, and listening to the lecture. For example, P12 found it *"time-consuming to decipher the transcript while trying to keep up with the lecture"*.

However, in this condition, the AI was often seen as a *"shadow companion for real-time understanding,"* helping students capture specific terminology or follow along with the lecture content as closely as possible. Since the AI only performed basic grammatical and punctuation corrections, students felt that it *"reflected the lecture content as closely as possible"* (P29) and *"did not add the cognitive burden of learning additional AI-generated content"* (P17). Compared to the other two conditions, although the cognitive load was highest here, the AI rarely introduced misunderstandings (P9). In contrast, the other two conditions (Automated AI, Intermediate AI) occasionally led to uncertainty. Users noted that they felt *"forced to adopt a new logic to interpret the AI content, which was impossible when I hadn't fully understood the lecture in the first place"* (P19).

**5.2.4 Students valued note-taking as a learning opportunity and preferred maintaining control over their own notes.** Students value the agency they have over their notes. Most students viewed AI assistance as a *"backup"* (P20) or *"augmentation"* (P8) rather than a replacement, using it selectively for tasks like verifying information (P5), catching up on missed points (P10), and enhancing note completeness (P20). Although students found AI support helpful, they emphasized the importance of manual note-taking as a critical learning opportunity. Writing original notes allowed them to *"memorize better"* (P11, P12), *"avoid distraction and keep up with the lecture"* (P13), *"engage actively in the lecture instead of passively copying information"* (P15), and *"gain a deeper understanding of concepts"* (P20). In the Automated AI condition, even with AI-generated structured notes, some students (10/30) still preferred to maintain their own notes rather than use the AI-generated content and found it *"easier to retrieve information"* when notes followed their own structure (P4).

**5.2.5 Students actively seek real-time clarification and AI content to align with their goals.** Participants actively seek AI assistance through searching to help *"capture missed points when the lecture went fast"* (P6) and to *"efficiently access specific information"* when the sheer volume of notes felt overwhelming (P4). After the lecture, students sought more information from the AI to clarify concepts they didn't fully understand or recall. For instance, P1 noted, *"I think the notes didn't go deep into its pros and cons, so I asked it to generate more notes on this."* Additionally, students valued the ability to perform *"in-contextual searches"*; as P28 noted, *"listening to the future makes sense since I can selectively get some results related to that topic constantly, which saves my time reading."*

Interestingly, some students used the search feature not just to fill gaps but also to *"verify the accuracy of AI-generated notes"*—even for concepts they felt they already understood. In the Automated AI condition, P5 searched for a specific concept to assess the AI's accuracy, but when the result didn't meet expectations, it reduced their trust in the AI. While the search feature enabled active engagement, students felt uncertain when searching for notes on future topics. P1 described this uncertainty: *"I'm not sure whether the teacher will talk about it."*

**5.2.6 Users anticipate more contextual and intent-aware AI assistance.** Students also reported that it is important for AI assistance to align with their note-taking preferences and the nature of the task. For instance, some students felt that Intermediate AI would be particularly helpful in conceptually dense lectures, where they wanted to stay engaged and needed scaffolding to keep up with the content while still thinking actively, like P4 noted, *"If I want to understand something, I like having the summary—it gives me something to build from."* When the goal was to retain detailed information for later study, some students preferred more complete outputs like those in the Automated AI condition. *"For a review-heavy class,"* P8 explained, *"I just want something I can look back on—structured notes are better for that."*

In addition to this contextual variation, several participants expressed a desire for AI support that could better adapt to their evolving needs. They imagined that the system could gradually learn their preferences based on how they reorganized or edited the AI-generated content. As P19

shared, *"If I'm always rewriting its definition style, it should know I like it shorter and with examples."* Students hoped that the system could infer their focus or confusion based on interaction patterns. As P23 noted, *"If I spend a while hovering or editing a part, maybe it could suggest related points."*

## 6 Discussion

Our research centers around the questions of whether high levels of AI assistance harm students' cognitive engagement and comprehension and whether providing intermediate outputs can mitigate some of these negative effects. Our findings indicate that while **Automated AI** offers the highest perceived usability and lowest cognitive load, it also results in the poorest comprehension. In contrast, AI assistance in the **Intermediate AI** condition effectively mitigates the potential negative impact of AI on comprehension, with students benefiting from interacting with AI and perceiving this interaction as helpful for learning. We discuss our results with note-taking, cognitive load, and learning theories to explain the mechanisms behind this trend. We also discuss the implications of designing intermediate AI to support human cognition.

### 6.1 Unpacking the Paradox: Reduced Effort v.s. Enhanced Comprehensions

This discussion seeks to understand why **Intermediate AI** can promote comprehension and support cognitive processes to a certain extent, what mechanisms make it useful, and why **Automated AI**, despite seeming to ease the burden on students, does not truly aid comprehension and cognitive processes and, in some cases, even hinders them.

Our findings indicate that students in the **Intermediate AI** condition scored the highest on both post-lecture and review-modified comprehension tests (Table 2). The lower performance of the **Automated AI** condition may be explained by a lack of encoding processes during note-taking [34]. On the other hand, **Intermediate AI** preserved the need for learners to interpret, organize, and adapt information in real-time. This aligns with the encoding-storage paradigm [34], which posits that learning is strengthened when individuals actively process content during note-taking. Besides, **Intermediate AI** presents appropriately sized information chunks [59, 70], enabling students to efficiently scan AI-generated content, evaluate its relevance, and consider how to integrate it into their structure.

A counterintuitive finding was that even after students used the notes to adjust their responses in the post-test, **Automated AI** still performed worse. This contrasts with students' intuitive expectations, as most interviewees reported feeling that structured notes provided the most complete and organized information during the review phase. We believe that this finding aligns with our design rationale, as shown in §3.2.3. When AI takes over much of the encoding process, it may effectively reduce the burden of encoding but also increase the challenge of interpreting and making sense of this information during storage. This also ties in with collaborative sense-making theories suggesting that when collaborators largely take over the process of instantiating external representations, it can make consuming encodings more difficult [53].

Our data show that while overall cognitive load did not differ significantly, the perceived extraneous cognitive load was lowest in the **Automated AI** condition (Table 3). However, as cognitive load theory suggests, cognitive load is not always detrimental [36, 48]. Although students reported that interacting with AI in the **Intermediate AI** condition required effort, they perceived this effort as enhancing their understanding. **Intermediate AI** appears to direct user effort toward meaningful processing rather than mechanical transcription. Our study is aligned with the cognitive load theory in that, in cognitively intensive tasks such as learning and note-taking, the goal of AI assistance should not be minimizing all user effort but rather supporting users in offloading mechanical effort while maintaining space for active reasoning, reflection, and self-monitoring.

[10, 55, 59]. This perspective aligns with prior work discussing cognitive engagement with AI assistance cognitive engagement, which distinguishes cognitive engagement—where users interpret and apply information—from passive engagement, where users simply follow AI suggestions without critical processing [21].

The implications of these findings suggest that we need to be more cautious when providing AI assistance to users, especially on tasks that require substantial cognitive effort. As users naturally prefer to minimize effort, the risk of over-reliance on AI becomes even more pronounced. This insight also calls for reconsidering how we measure the effectiveness of AI tools for cognitive tasks. Many studies use task load surveys to measure the amount of load and assume that a lower load is better [25]. However, if these tools are designed for cognitive tasks, such as learning, brainstorming, or meetings, a lower load does not necessarily indicate better outcomes. More appropriate measurements should be considered.

## 6.2 Designing Intermediate AI for Real-Time Learning and Note-Taking

Our study shows that *Intermediate AI* significantly promotes student understanding and maintains cognitive engagement during note-taking. The success of *Intermediate AI* can be attributed to its design as a “middle ground,” where AI-generated “building blocks” allow students to integrate relevant information into their notes in real-time. Unlike *Automated AI*, which presents finalized content, *IntermediateAI* invites learners to evaluate, modify, and reorganize the information. Rather than reducing effort, *Intermediate AI* redistributes it toward meaningful engagement, aligning with theories of desirable difficulty [5] and cognitive scaffolding [3], which suggest that moderate challenge, when well-supported, can enhance comprehension and retention.

However, *Intermediate AI* should not be viewed as a one-size-fits-all solution. As our findings show (§5.2.6), students imagined using different types of AI assistance depending on lecture pace, topic complexity, and their own learning goals. These various assistance needs align with the Knowledge-Learning-Instruction (KLI) framework [38], which emphasizes that instructional strategies should match the type of knowledge being acquired. While our current implementation of *Intermediate AI* supports factual knowledge particularly well—each summary block typically maps to a discrete fact or definition—we observe that students often go beyond simply collecting facts. In our study, learners actively reorganized, grouped, and rephrased building blocks to construct broader conceptual insights or procedural understanding. This reveals both a strength and a design opportunity: *Intermediate AI*, by surfacing atomic knowledge units, leaves room for users to engage in deeper knowledge construction through structuring and synthesis. However, to better support learning tasks that involve procedural or conceptual knowledge from the outset, future systems could extend *Intermediate AI* to present more connected chunks of content or scaffold cross-turn integration to support comprehension and application better.

In light of these shifting demands, we argue that intermediate AI should be adaptive, not merely in the traditional sense of tailoring scaffolding based on performance assessment [3], but in a more goal-oriented and engagement-sensitive way. Students expressed interest in systems that could infer preferences from their interaction patterns, such as how they restructure summaries, what types of content they drop or emphasize, and how long they hover over certain blocks. While prior models primarily adapt support based on observable learning performance in well-defined problem spaces [31], our findings highlight the need for AI assistance to adapt in more open-ended learning environments by dynamically adjusting content granularity based on topic complexity, inferring learner state from behavioral signals (e.g., hesitations, deletions), or offering customizable output formats (e.g., bullet points, outlines, concept maps).

### 6.3 Intermediate AI as a Human-AI Collaboration Paradigm

While our study focuses on the learning context of note-taking, its design implications extend to broader cognitive tasks where AI assists human sensemaking and decision-making. We propose that *Intermediate AI* embodies a distinct paradigm—intermediate AI as a deliberate middleware layer—that supports cognitively intensive work without over-automation.

A key design principle behind *Intermediate AI* is the intentional preservation of the most creative and intellectually demanding aspects of a task for the human user. In our case, *Intermediate AI* alleviates the burden of low-level actions, such as continuous typing, while preserving higher-order sensemaking, allowing learners to evaluate, select, and integrate relevant information [34]. Similar human-led constructs appear in other cognitively valuable contexts, such as refining one’s argument during writing [15] or sustaining intentionality in meetings [10, 56]. In contrast, repetitive or mechanical subtasks—recording, formatting, or retrieval—can be offloaded to AI.

Importantly, *Intermediate AI* is not limited to the format of summarization shown in this paper. It represents a broader design space with multiple representations: building blocks [28], embedded questions [10], or interactive templates [68]. Our exploration of intermediate AI design in real-time note-taking scenarios further contributes design insights that may generalize to other intermediate AI systems. These include (1) output granularity—how abstract the AI-generated content should be and when to surface it; and (2) control dynamics—balancing pull-based user queries with push-based AI suggestions.

This paradigm is particularly promising in real-time, collaborative settings. For example, in group meetings, managing the cognitive demands of listening, reasoning, and contributing simultaneously can be challenging [64]. *Intermediate AI* enables proactive AI-generated summaries and user-initiated, in-context search, allowing participants to fluidly engage with support without interrupting their cognitive flow. MeetMap similarly adopts this middleware layer by providing a transient holding area for live AI summaries, preventing overload and allowing for timely, contextual interpretation [11].

Besides providing real-time AI help, intermediate AI design also opens up opportunities for implicit intent modeling. Users imagine that *Intermediate AI* could progressively learn from their behaviors and use these signals to infer intent and generate more tailored, personalized outputs over time. This idea is echoed in recent system GRANOLA, which observes light-touch user inputs during meetings and generates personalized post-meeting summaries aligned with the input [1].

In sum, much of today’s AI development centers on task automation—optimizing speed, output, and scale. While effective, this also reduces opportunities for learning and critical thinking. We advocate for *Intermediate AI* as a design paradigm that re-centers human agency, reducing unnecessary effort while preserving the right kind of effort—effort that enables engagement, reflection, and ownership. We hope to elicit further discussion on designing AI not only for productivity but also for thoughtfulness and augmenting human cognition.

### 6.4 Limitation

We acknowledge the limitations of this work. 1) Our sample size was limited, preventing us from using more complex models to explore relationships between variables. 2) The videos used in this study were only ten minutes long, which may not fully capture the impact of AI assistance over extended lectures or in real, long-term courses. Future work could address this with a larger sample, deploying a real classroom setting. Although we controlled for video difficulty and ensured the content was accessible without prior knowledge, our participants were from different academic years with varying knowledge levels. This may have affected their final scores; however, we mitigated this by treating individual differences as a random effect in our model, which likely



minimized its influence. Future research could incorporate pre- and post-tests to measure learning gains more accurately. Our current analysis of user interactions with the dropped AI notes only compares the final state of the notes to the original AI-generated content. Although we logged each user interaction with the editor (e.g., typing events during the session), fully capturing the dynamic editing process — such as incremental additions, deletions, or reorganizations of AI content — would require more detailed data logging and semantic analysis of intermediate note states. We will consider logging this process data and further analyzing how users interact with the AI-generated content in real-time in future studies.

## 7 Conclusion

As AI tools become integral to tasks requiring high cognitive effort, such as note-taking, questions arise about their effects on cognitive engagement. This paper explores the "AI Assistance Dilemma" in note-taking by studying the impact of different AI support levels on user engagement and comprehension through a within-subject experiment where participants (N=30) took notes during lecture videos under three conditions: **Automated AI** (high assistance with structured notes), **Intermediate AI** (moderate assistance with real-time summary), and **Minimal AI** (low assistance with transcript); results indicate that Intermediate AI led to the highest post-test scores, while Automated AI resulted in the lowest, although participants favored Automated AI for its perceived ease and lower cognitive effort, highlighting a potential mismatch between preferred convenience and cognitive benefits. Our study provides insights into designing AI systems that maintain cognitive engagement, with implications for developing moderate AI support in cognitive tasks.

## Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number IIS-2302564.

## References

- [1] 2025. <https://www.granola.ai/>
- [2] Sumit Asthana, Sagih Hilleli, Pengcheng He, and Aaron Halfaker. 2023. Summaries, Highlights, and Action items: Design, implementation and evaluation of an LLM-powered meeting recap system. *arXiv preprint arXiv:2307.15793* (2023).
- [3] Roger Azevedo, Jennifer G Cromley, and Diane Seibert. 2004. Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary educational psychology* 29, 3 (2004), 344–370.
- [4] Aaron Bauer and Kenneth R Koedinger. 2007. Selection-based note-taking applications. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 981–990.
- [5] Elizabeth L Bjork, Robert A Bjork, et al. 2011. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society* 2, 59-68 (2011).
- [6] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.
- [7] Dung C. Bui, Joel Myerson, and Sandra Hale. 2013. Note-taking with computers: Exploring alternative strategies for improved recall. *Journal of Educational Psychology* 105, 2 (2013), 299–309. <https://doi.org/10.1037/a0030367>
- [8] Ting-Ju Chen. 2021. *Association, Reflection, Stimulation: Problem Exploration in Early Design through AI-Augmented Mind-Mapping*. Thesis. <https://oaktrust.library.tamu.edu/handle/1969.1/195385> Accepted: 2022-01-27T22:18:02Z.
- [9] Xinyue Chen, Shuo Li, Shipeng Liu, Robin Fowler, and Xu Wang. 2023. Meetscript: designing transcript-based interactions to support active participation in group video meetings. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–32.
- [10] Xinyue Chen, Lev Tankelevitch, Rishi Vanukuru, Ava Elizabeth Scott, Payod Panda, and Sean Rintel. 2025. Are We On Track? AI-Assisted Active and Passive Goal Reflection During Meetings. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.
- [11] Xinyue Chen, Nathan Yap, Xinyi Lu, Aylin Gunal, and Xu Wang. 2025. MeetMap: Real-Time Collaborative Dialogue Mapping with LLMs in Online Meetings. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–35.



- [12] Jamie Costley, Matthew Courtney, and Mik Fanguy. 2022. The interaction of collaboration, note-taking completeness, and performance over 10 weeks of an online course. *The Internet and Higher Education* 52 (2022), 100831.
- [13] Jamie Costley and Mik Fanguy. 2021. Collaborative note-taking affects cognitive load: the interplay of completeness and interaction. *Educational Technology Research and Development* 69 (2021), 655–671.
- [14] Richard C Davis, James A Landay, Victor Chen, Jonathan Huang, Rebecca B Lee, Frances C Li, James Lin, Charles B Morrey III, Ben Schleimer, Morgan N Price, et al. 1999. NotePals: Lightweight note sharing by the group, for the group. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 338–345.
- [15] Paramveer S Dhillon, Somayeh Molaei, Jiaqi Li, Maximilian Golub, Shaochun Zheng, and Lionel Peter Robert. 2024. Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-writing with Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.
- [16] Paramveer S. Dhillon, Somayeh Molaei, Jiaqi Li, Maximilian Golub, Shaochun Zheng, and Lionel P. Robert. 2024. Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-writing with Language Models. <https://arxiv.org/abs/2402.11723> arXiv:2402.11723 [cs].
- [17] Jingchao Fang, Yanhao Wang, Chi-Lan Yang, Ching Liu, and Hao-Chuan Wang. 2022. Understanding the Effects of Structured Note-taking Systems for Video-based Learners in Individual and Social Learning Contexts. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (Jan. 2022), 1–21. <https://doi.org/10.1145/3492840>
- [18] Jingchao Fang, Yanhao Wang, Chi-Lan Yang, and Hao-Chuan Wang. 2021. NoteCoStruct: Powering online learners with socially scaffolded note taking and sharing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–5.
- [19] Mik Fanguy, Matthew Baldwin, Evgeniia Shmeleva, Kyungmee Lee, and Jamie Costley. 2023. How collaboration influences the effect of note-taking on writing performance and recall of contents. *Interactive Learning Environments* 31, 7 (Oct. 2023), 4057–4071. <https://doi.org/10.1080/10494820.2021.1950772>
- [20] Abraham E. Flanigan and Scott Titsworth. 2020. The impact of digital distraction on lecture note taking and student learning. *Instructional Science* 48, 5 (Oct. 2020), 495–524. <https://doi.org/10.1007/s11251-020-09517-2>
- [21] Krzysztof Z Gajos and Lena Mamykina. 2022. Do people engage cognitively with AI? Impact of AI assistance on incidental learning. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. 794–806.
- [22] Jie Gao, Kenny Tsu Wei Choo, Junming Cao, Roy Ka-Wei Lee, and Simon Perrault. 2023. CoAlcoder: Examining the effectiveness of AI-assisted human-to-human collaboration in qualitative analysis. *ACM Transactions on Computer-Human Interaction* 31, 1 (2023), 1–38.
- [23] Nitesh Goyal, Gilly Leshed, and Susan R. Fussell. 2013. Effects of visualization and note-taking on sensemaking and analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 2721–2724. <https://doi.org/10.1145/2470654.2481376>
- [24] Ziwei Gu, Ian Arawjo, Kenneth Li, Jonathan K Kummerfeld, and Elena L Glassman. 2024. An AI-Resilient Text Rendering Technique for Reading and Skimming Documents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.
- [25] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.
- [26] Ken Hinckley, Shengdong Zhao, Raman Sarin, Patrick Baudisch, Edward Cutrell, Michael Shilman, and Desney Tan. 2007. InkSeine: In Situ search for active note taking. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 251–260.
- [27] Xinying Hou, Barbara Jane Ericson, and Xu Wang. 2022. Using Adaptive Parsons Problems to Scaffold Write-Code Problems. In *Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 1*. ACM, Lugano and Virtual Event Switzerland, 15–26. <https://doi.org/10.1145/3501385.3543977>
- [28] Xinying Hou, Zihan Wu, Xu Wang, and Barbara J Ericson. 2025. Personalized Parsons Puzzles as Scaffolding Enhance Practice Engagement Over Just Showing LLM-Powered Solutions. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 2*. 1483–1484.
- [29] Faria Huq, Abdus Samee, David Chuan-en Lin, Xiaodi Alice Tang, and Jeffrey P Bigham. 2024. NoTeeline: Supporting Real-Time Notetaking from Keypoints with Large Language Models. *arXiv preprint arXiv:2409.16493* (2024).
- [30] Renée S Jansen, Daniel Lakens, and Wijnand A IJsselstein. 2017. An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review* 22 (2017), 223–233.
- [31] Slava Kalyuga. 2009. Adapting levels of instructional support to optimize learning complex cognitive skills. In *Managing Cognitive Load in Adaptive Multimedia Learning*. IGI Global, 246–271.
- [32] Matthew Kam, Jingtao Wang, Alastair Iles, Eric Tse, Jane Chiu, Daniel Glaser, Orna Tarshish, and John Canny. 2005. Livenotes: a system for cooperative and augmented note-taking in lectures. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 531–540.
- [33] Anam Ahmad Khan, Sadia Nawaz, Joshua Newn, Ryan M. Kelly, Jason M. Lodge, James Bailey, and Eduardo Velloso. 2022. To type or to speak? The effect of input modality on text understanding during note-taking. In *CHI Conference*

- on *Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–15. <https://doi.org/10.1145/3491102.3501974>
- [34] Kenneth A. Kiewra. 1989. A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review* 1, 2 (June 1989), 147–172. <https://doi.org/10.1007/BF01326640>
- [35] Kenneth A. Kiewra, Nelson F. DuBois, David Christian, Anne McShane, Michelle Meyerhoffer, and David Roskelley. 1991. Note-taking functions and techniques. *Journal of Educational Psychology* 83, 2 (June 1991), 240–245. <https://doi.org/10.1037/0022-0663.83.2.240> Publisher: American Psychological Association.
- [36] Melina Klepsch, Florian Schmitz, and Tina Seufert. 2017. Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in psychology* 8 (2017), 294028.
- [37] Kenneth R. Koedinger and Vincent Aleven. 2007. Exploring the Assistance Dilemma in Experiments with Cognitive Tutors. *Educational Psychology Review* 19, 3 (Sept. 2007), 239–264. <https://doi.org/10.1007/s10648-007-9049-0>
- [38] Kenneth R Koedinger, Albert T Corbett, and Charles Perfetti. 2012. The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive science* 36, 5 (2012), 757–798.
- [39] Anastasia Kuzminykh and Sean Rintel. 2020. Classification of Functional Attention in Video Meetings. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. <https://doi.org/10.1145/3313831.3376546>
- [40] Hui Yi Leong, Yi Fan Gao, Shuai Ji, Bora Kalaycioglu, and Uktu Pamuksuz. 2024. A GEN AI Framework for Medical Note Generation. *arXiv preprint arXiv:2410.01841* (2024).
- [41] Jimmie Leppink, Fred Paas, Cees PM Van der Vleuten, Tamara Van Gog, and Jeroen JG Van Merriënboer. 2013. Development of an instrument for measuring different types of cognitive load. *Behavior research methods* 45 (2013), 1058–1072.
- [42] Susan Lin, Jeremy Warner, JD Zamfirescu-Pereira, Matthew G Lee, Sauhard Jain, Shanqing Cai, Piyawat Lertvit-tayakumjorn, Michael Xuelin Huang, Shumin Zhai, Björn Hartmann, et al. 2024. Rambler: Supporting Writing With Speech via LLM-Assisted Gist Manipulation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.
- [43] Ching Liu, Chi-Lan Yang, Joseph Jay Williams, and Hao-Chuan Wang. 2019. Notestruct: Scaffolding note-taking while learning from online videos. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–6.
- [44] Tamas Makany, Jonathan Kemp, and Itiel E. Dror. 2009. Optimising the use of note-taking as an external cognitive aid for increasing learning. *British Journal of Educational Technology* 40, 4 (July 2009), 619–635. <https://doi.org/10.1111/j.1467-8535.2008.00906.x>
- [45] Bruce M McLaren, S Lim, and Kenneth R Koedinger. 2008. When and how often should worked examples be given to students? New results and a summary of the current state of research. In *Proceedings of the 30th annual conference of the cognitive science society*. 2176–2181.
- [46] Andrew M McNutt, Chenglong Wang, Robert A Deline, and Steven M. Drucker. 2023. On the Design of AI-powered Code Assistants for Notebooks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–16. <https://doi.org/10.1145/3544548.3580940>
- [47] Fred GWC Paas and Jeroen JG Van Merriënboer. 1993. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors* 35, 4 (1993), 737–743.
- [48] Annie Piolat, Thierry Olive, and Ronald Kellogg. 2005. Cognitive effort during note taking. *Applied Cognitive Psychology* 19 (April 2005), 291–312. <https://doi.org/10.1002/acp.1086> 270 citations (Crossref) [2024-01-03].
- [49] Yi Ren, Yang Li, and Edward Lank. 2014. InkAnchor: enhancing informal ink-based note taking on touchscreen mobile phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1123–1132.
- [50] Ido Roll, Vincent Aleven, Bruce M McLaren, and Kenneth R Koedinger. 2007. Designing for metacognition—applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning* 2 (2007), 125–140.
- [51] Ido Roll, Vincent Aleven, Bruce M McLaren, and Kenneth R Koedinger. 2011. Improving students’ help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and instruction* 21, 2 (2011), 267–280.
- [52] Ethan Z. Rong, Mo Morgana Zhou, Ge Gao, and Zhicong Lu. 2023. Understanding Personal Data Tracking and Sensemaking Practices for Self-Directed Learning in Non-classroom and Non-computer-based Contexts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–16. <https://doi.org/10.1145/3544548.3581364>
- [53] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. 1993. The cost structure of sensemaking. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93*. ACM Press, Amsterdam, The Netherlands, 269–276. <https://doi.org/10.1145/169059.169209>
- [54] Martin Schroder. 2023. AutoScrum: Automating Project Planning Using Large Language Models. <https://doi.org/10.48550/arXiv.2306.03197> arXiv:2306.03197 [cs].
- [55] Ava Elizabeth Scott, Lev Tankelevitch, Payod Panda, Rishi Vanukuru, Xinyue Chen, and Sean Rintel. 2025. What Does Success Look Like? Catalyzing Meeting Intentionality with AI-Assisted Prospective Reflection. In *Proceedings of the*

- 4th Annual Symposium on Human-Computer Interaction for Work. 1–32.
- [56] Ava Elizabeth Scott, Lev Tankelevitch, and Sean Rintel. 2024. Mental Models of Meeting Goals: Supporting Intentionality in Meeting Technologies. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.
  - [57] Orit Shaer, Angelora Cooper, Osnat Mokryn, Andrew L Kun, and Hagit Ben Shoshan. 2024. AI-Augmented Brainwriting: Investigating the use of LLMs in group ideation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–17. <https://doi.org/10.1145/3613904.3642414>
  - [58] Meng Sun, Minhong Wang, Rupert Wegerif, and Jun Peng. 2022. How do students generate ideas together in scientific creativity tasks through computer-based mind mapping? *Computers & Education* 176 (Jan. 2022), 104359. <https://doi.org/10.1016/j.compedu.2021.104359> 20 citations (Crossref) [2024-01-03].
  - [59] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The metacognitive demands and opportunities of generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–24.
  - [60] Johanna Telenius. 2016. *Sensemaking in Meetings - Collaborative Construction of Meaning and Decisions through Epistemic Authority*. Aalto University. <https://aaltodoc.aalto.fi/handle/123456789/23391> ISSN: 1799-4942 (electronic).
  - [61] Hsin-Ruey Tsai, Shih-Kang Chiu, and Bryan Wang. 2024. GazeNoter: Co-Piloted AR Note-Taking via Gaze Selection of LLM Suggestions to Match Users' Intentions. *arXiv preprint arXiv:2407.01161* (2024).
  - [62] Karthikeyan Umapathy. [n. d.]. Requirements to support Collaborative Sensemaking. ([n. d.]).
  - [63] Jixuan Wang, Jingbo Yang, Haochi Zhang, Helen Lu, Marta Skreta, Mia Husić, Aryan Arbabi, Nicole Sultanum, and Michael Brudno. 2022. PhenoPad: building AI enabled note-taking interfaces for patient encounters. *NPJ digital medicine* 5, 1 (2022), 12.
  - [64] Ruotong Wang, Lin Qiu, Justin Cranshaw, and Amy X. Zhang. 2024. Meeting Bridges: Designing Information Artifacts that Bridge from Synchronous Meetings to Asynchronous Collaboration. <http://arxiv.org/abs/2402.03259> arXiv:2402.03259 [cs].
  - [65] Amber E. Witherby and Sarah K. Tauber. 2019. The Current Status of Students' Note-Taking: Why and How Do Students Take Notes? *Journal of Applied Research in Memory and Cognition* 8, 2 (June 2019), 139–153. <https://doi.org/10.1016/j.jarmac.2019.04.002>
  - [66] Sarah Shi Hui Wong and Stephen Wee Hun Lim. 2023. Take notes, not photos: Mind-wandering mediates the impact of note-taking strategies on video-recorded lecture learning performance. *Journal of Experimental Psychology: Applied* 29, 1 (March 2023), 124–135. <https://doi.org/10.1037/xap0000375>
  - [67] David Wood. 2001. Scaffolding, contingent tutoring, and computer-supported learning. *International Journal of Artificial Intelligence in Education* 12, 3 (2001), 280–293.
  - [68] Xiaotong Xu, Jiayu Yin, Catherine Gu, Jenny Mar, Sydney Zhang, Jane L E, and Steven P Dow. 2024. Jamplate: exploring llm-enhanced templates for idea reflection. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 907–921.
  - [69] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. Visar: A human-ai argumentative writing assistant with visual programming and rapid draft prototyping. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–30.
  - [70] Zheng Zhang, Weirui Peng, Xinyue Chen, Luke Cao, and Toby Jia-Jun Li. 2025. LADICA: a large shared display interface for generative AI cognitive assistance in co-located team collaboration. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.
  - [71] Rebecca Zheng, Marina Fernández Camporro, Hugo Romat, Nathalie Henry Riche, Benjamin Bach, Fanny Chevalier, Ken Hinckley, and Nicolai Marquardt. 2021. Sketchnote components, design space dimensions, and strategies for effective visual note taking. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

Received October 2024; revised April 2025; accepted August 2025