Papeos: Augmenting Research Papers with Talk Videos[*]

Tae Soo Kim[†]
School of Computing, KAIST
Daejeon, Republic of Korea
taesoo.kim@kaist.ac.kr

Matt Latzke
Allen Institute for AI
Seattle, WA, USA
mattl@allenai.org

Jonathan Bragg
Allen Institute for AI
Seattle, WA, USA
jbragg@allenai.org

Amy X. Zhang
University of Washington
Seattle, WA, USA
axz@cs.uw.edu

Joseph Chee Chang
Allen Institute for AI
Seattle, WA, USA
josephc@allenai.org

ABSTRACT

Research consumption has been traditionally limited to the reading of academic papers—a static, dense, and formally written format. Alternatively, pre-recorded conference presentation videos, which are more dynamic, concise, and colloquial, have recently become more widely available but potentially under-utilized. In this work, we explore the design space and benefits for combining academic papers and talk videos to leverage their complementary nature to provide a rich and fluid research consumption experience. Based on formative and co-design studies, we present **Papeos**, a novel reading and authoring interface that allow authors to augment their **pap**ers by segmenting and localizing talk vid**eos** alongside relevant paper passages with automatically generated suggestions. With Papeos, readers can visually skim a paper through clip thumbnails, and fluidly switch between consuming dense text in the paper or visual summaries in the video. In a comparative lab study (n=16), Papeos reduced mental load, scaffolded navigation, and facilitated more comprehensive reading of papers.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; **Empirical studies in HCI**.

KEYWORDS

Interactive Documents; Reading Interfaces; Scientific Papers; Videos

ACM Reference Format:
Tae Soo Kim, Matt Latzke, Jonathan Bragg, Amy X. Zhang, and Joseph Chee Chang. 2023. Papeos: Augmenting Research Papers with Talk Videos. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), October 29–November 01, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3586183.3606770

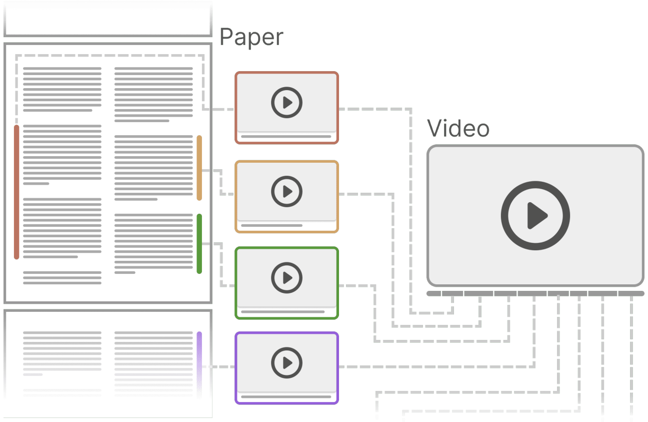


Figure 1: Papeos augment academic papers by linking relevant passages and segments of authors' talk videos. Video segments are presented as margin notes that are localized and color-coded next to relevant passages. In Papeos, users can fluidly switch between consuming the dense and detailed paper text, and the typically more concise and easier to understand talk video—providing a new scholarly reading experience. See system screenshots in Figures 3 and 4.

1 INTRODUCTION

Research progress is driven by the consumption of prior research. This is most commonly performed through reading published academic papers that are written in a rigid, dense, and formal fashion to ensure clarity, reproducibility, and to fit within a tight page limit constraint. Alternatively, another common way researchers learn about prior research is through conference presentations. Presentations or talks are typically more concise and colloquial, and can contain rich and dynamic content that cannot be included in the paper, such as screencasts of user interfaces, animated figures, and progressive diagrams.

Traditionally, author presentations were typically only available to attendees and only during conferences. However, for archival purposes [4] or to accommodate remote participants (e.g., during the COVID-19 pandemic, to be more inclusive [1], and to reduce carbon emissions [29]), recordings and pre-recordings of conference talks (i.e., talk videos) have become more widely available

[*] Click here to open the Papeo version of this document: https://papeo.app/demo
[†] Work completed during a researcher internship at Semantic Scholar Research, Allen Institute for AI.

UIST '23, October 29–November 01, 2023, San Francisco, CA, USA
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0132-0/23/10.
https://doi.org/10.1145/3586183.3606770

across different fields in recent years. Prior work, such as in psychology and education, has found various benefits in a personal and multimedia communication style (i.e., videos and dialogues) over formal and technical text, including positive social context and experiences [38], lowered cognitive load and increased interest [57], and improved comprehension when multiple alternative explanations were available [2]. However, prior HCI research has also showed that carefully designed interfaces are crucial for users to consume multiple formats without being overwhelmed [25]. In this work, we build on prior theoretical and HCI research to explore the design space for combining research papers and talk videos into a cohesive reading experience by investigating the perspectives of both paper authors and readers.

Talk videos differ from papers in format and content, and this can serve to address various challenges in research consumption. Specifically, while reading papers allows scholars to dig deep into all the details of a prior work, the process can be cognitively demanding as scholars must disentangle meaning from complex written explanations [8]. This process is further complicated as researchers may lack the background knowledge required to understand the explanations or due to variability in the quality of the writing [61, 62]. Even further, to keep pace with the rapidly expanding literature, researchers are increasingly pressured to skim papers, and they attempt to gain a high-level understanding from scattered fragments of writing [31, 54]. In contrast, a talk video may present visuals that can help illustrate complex explanations [17, 23, 68] and, due to their wider audience, focus less on specialized concepts or background knowledge while using simpler language [20, 69]. Furthermore, as talk videos typically do not contain all the details, they can present scholars with a concise and easy-to-understand overview of the corresponding papers [9, 49].

Despite the various ways in which talk videos can complement paper reading, these two formats remain largely disconnected. Readers have to choose between using either the talk video or the paper as their primary way to consume prior work, and cognitive costs to switch between the two formats could be prohibitively high. For example, if a scholar watching a talk video wants to find a specific implementation detail for a machine learning model that was omitted in the video, they must search through pages in the paper to find the corresponding passage. Similarly, when reading a paper about an interactive user interface, it can also be costly for a scholar to scrub through its talk video to search for a screencast of the system to see it in action. This disconnect prohibits readers from fluidly transitioning between papers and talk videos because context switching can be disruptive [11] and incurs significant cognitive load [6]. As a result, while the research community has recently made significant efforts in creating presentation talk videos and making them widely available even after conferences, researchers are unable to fully capitalize on their benefits.

In a formative study with researchers (n=14), we investigated opportunities and challenges in consuming papers and videos together, and the design space for combining these two formats. Instead of augmenting one format with the other, our findings revealed that researchers alternated their focus between the paper and video to control the level of detail in which they consumed the paper. Additionally, researchers observed how linking video segments to relevant paper passages (e.g., paragraphs, figures) could facilitate navigation, as the video could act as a visual map for the paper. Finally, researchers were against replacing or overlaying content in one format with content from the other as this could obscure information and the effort they dedicated in authoring both formats.

Based on these findings, we designed a novel paper reading experience, *Papeos* (**pap**er and vid**eo**), that integrates segments of the talk videos as localized *video notes* alongside corresponding sections of the paper. As a user scrolls through a Papeo, they can see color-coded *highlight bars* in the paper that hint at meaningful passages that have been covered in the video and, next to the paper, correspondingly color-coded video notes with thumbnails of the relevant video segments. When the user struggles to understand a portion of the paper, they can click on the highlight bar or video note to play the segment and get a summarized, alternative explanation. Instead of scrolling through the paper, the user can also choose to focus on the video by navigating between video notes or "autoplaying" through them. To avoid disturbing the user's watching, the system fixes the video note's position in the viewport and scrolls the paper to the relevant passage. To grant authors control on how Papeos are created for their papers and facilitate the creation process, we also present an authoring interface where authors can link their papers and talk videos with the help of AI suggestions.

To evaluate Papeos, we conducted a within-subjects study (n=16) where participants read and wrote a summary for the systems section of three papers using only the paper, the paper and talk video, or a Papeo. Our study revealed that Papeos could help researchers understand papers and decrease their mental demand during reading. Additionally, through Papeos, each format became a guide for the other which facilitated participants' navigation in the two formats and encouraged them to interact with both formats more. As a consequence of the reduced cognitive demand and improved navigation support, participants composed summaries that more comprehensively covered details from the papers. In addition, we conducted a field deployment of Papeos during an HCI conference where we had over 250 unique visitors to our reading interface.

This paper presents the following contributions:

1. A formative study using a design probe (Fig. 2) with 14 participants that revealed user needs and potential benefits of combining talk videos and research papers for readers.
2. Co-design sessions with 14 paper authors that focused on understanding how authors would like to combine their papers and talk videos, to explore the design space for combining scholarly papers with talk videos.
3. Papeos: A novel reading experience that augments research papers with margin notes that present segments from a talk video alongside relevant passages in the paper (Fig. 3).
4. A mixed-initiative authoring interface that facilitates the creation of Papeos through AI-based suggestions, to explore the costs and feasibility of creating Papeos (Fig. 6).
5. A within-subjects study with 16 participants that revealed how integrating talk videos into papers enables readers to leverage both formats for improved understanding and navigation.

2 RELATED WORK

The goal of this work is to explore the design space for augmenting scientific paper reading with corresponding presentation talk videos. To better understand this space, we first review literature around these formats: tools that support general reading, scholarly reading, and knowledge consumption using videos. Finally, we also review prior techniques in other domains for linking between text documents and videos.

2.1 Augmented Reading Interfaces

The advent of computers has enabled the creation of reading environments that transcend the limitations of static print media and, instead, allow knowledge workers to interact with and explore text dynamically [78, 79]. Hypertext [19] interconnected scattered text and documents, and this concept has been widely adopted in many reading tools today (e.g., Amazon Kindle's in-situ definitions [3], and Wikipedia's page previews [60]). Expanding on hypertext, fluid documents [11] and fluid links [85] restructure documents to incorporate this linked content within the document, and various interfaces provide links between text and other document objects, such as tables [42] or visualizations [7]. To support active reading, various interfaces allow readers to annotate documents with multiple modalities, such as ink or voice [70, 84], to manipulate the document's structure [75], or to ask questions and find answers during reading [16, 27, 34]. As documents are frequently dense in content, researchers have investigated how to scaffold navigation by providing overviews [24, 70], highlighting or fading out content to direct readers' attention [40, 83], or guiding readers based on the activity of other readers [30, 45]. Extending on this rich body of work, we investigate how to augment the dynamism of academic papers by leveraging and integrating existing talk videos.

2.2 Tools for Reading Scientific Papers

A variety of tools have been designed to address the challenges in reading papers [52]. As a crucial component of reading a paper is to contextualize it within the broader literature, CiteRead [66] augments a paper with commentary from citing papers, CiteSee [13] contextualizes inline citations to a reader's previous reading and publishing activities with visual augmentations, and Threddy [36] and Synergi [37] allow users to clip citing sentences and references to explore related themes and papers in the literature. More closely related to our work, there is a line of research that focused on enhancing both efficiency and comprehension during paper reading. Specifically, to help readers traverse the complex language and notation used in scientific papers, Paper Plain [5] provides definitions for unfamiliar terms and in-situ summaries of sections, and ScholarPhi [28] surfaces position-sensitive definitions for unique terms and symbols. Also, to facilitate skimming of papers, Scim [21] highlights salient passages of the paper to direct readers' focus, and Spotlights [48] surfaces important objects as temporary overlays to help readers identify them even as they quickly scroll through the paper. Finally, since most scholarly papers are available as PDFs, various approaches have aimed at overcoming the limitations of this format to increase accessibility [63, 81] and dynamism (e.g., embedding animations [25] or interactive elements [53]). While prior work have focused on designs that can support specific user needs such as skimming [21] or simplification [5], in this work, we explore how incorporating talk videos has the potential to embody multiple user needs when reading a paper. Specifically, a talk video can present an author-curated summary for the paper, highlight significant aspects of the work. Linking video segments back to their corresponding passages in the papers also has the potential of allowing readers to skim the paper based on the passages that the authors selected to include in their talk videos. Furthermore, talk videos include additional commentary, audibly narrate the content which can supplement screen readers, and dynamically illustrate aspects of the work such as animations and screen recordings.

2.3 Video-based Knowledge Consumption

Videos are increasingly becoming a predominant channel through which people consume and learn knowledge. According to Mayer and Moreno's principles [58], videos can be cognitively beneficial as verbal and visual explanations allow viewers to build two mental representations [55, 56] without mental overload as audio and visual channels can be processed simultaneously [59]. As support to these principles, various studies have demonstrated that videos can benefit learners in various domains [32, 39, 73]. While effective for consumption of knowledge, videos represent a continuous stream of frames, and it can be inherently difficult to skim through or locate information in videos, which prior work had shown to be a common need for scholars [21]. To overcome this limitation and harness the potential of videos, various tools have been designed to facilitate video navigation in learning contexts [43, 44, 51]. In this work, we investigate the benefits of talk videos for consumption of research, and how to combine these with papers to support both video and paper navigation—allowing scholars to fluidly switch between the two formats.

2.4 Bridging Text Documents and Videos

To overcome the difficulty in skimming and efficiently navigating videos, prior work has investigated various approaches to bridge videos with relevant text documents in a variety of domains. In education, Video Digests [65] and VideoDoc [47] segment lecture videos into sections so that students can navigate between different parts of a lecture with transcript summaries, and Shin et al. [72] further combined transcripts with extracted blackboard notes. Beyond lecture videos, Truong et al. [77] transform transcripts into hierarchical tutorials for instructional makeup videos, and Sceneskim [64] facilitates searching and browsing by temporally aligning movies with their captions, scripts and summaries. Further, Codemotion [41] automatically extracts code shown in programming tutorials to allow the user to navigate tutorials based on code-related steps. While existing research above focused on improving video navigation with text extracted from the same videos (e.g., audio transcripts or blackboard notes extracted from the frames), in this work, we explore how to bridge talk videos with research papers, which are separate entities and different media, and investigate how combining them can facilitate navigation for both media and help scholars better comprehend prior research.

3 FORMATIVE AND CO-DESIGN STUDY

To explore the design space for combining research papers and talk videos, we conducted a formative study where participants explored the opportunities and challenges in combining the two formats from the perspectives of both readers and authors.

3.1 Participants

We invited 14 researchers who had previously published at least one paper and created accompanying talk videos. 10 were doctoral students, 2 were Master's students, and the remaining 2 were a postdoc and an undergraduate student. 10 of the 14 participants identified their discipline as human-computer interaction (HCI) or related sub-fields (e.g., visualizations, AI fairness), 3 as natural language processing (NLP), 2 as machine learning (ML), and 1 as computer vision (CV).[1]

3.2 Apparatus

Consuming scholarly papers and talk videos at the same time is a new experience that may be hard for participants to imagine. In a preliminary version of this formative study, we gave participants (n=4) a paper and talk video pair side-by-side and instructed them to *"understand the content of the paper based on your real-life habits"*. Although participants could freely choose how they wished to consume the paper and video, they all watched the whole video first and then delved into the paper. Participants expressed how this was not due to a lack of desire to jump to the paper while watching the video, but due to the prohibitively high cost of cross-referencing between formats. This preliminary study revealed that unaugmented papers and videos were inadequate to explore how readers wanted to leverage both formats together.

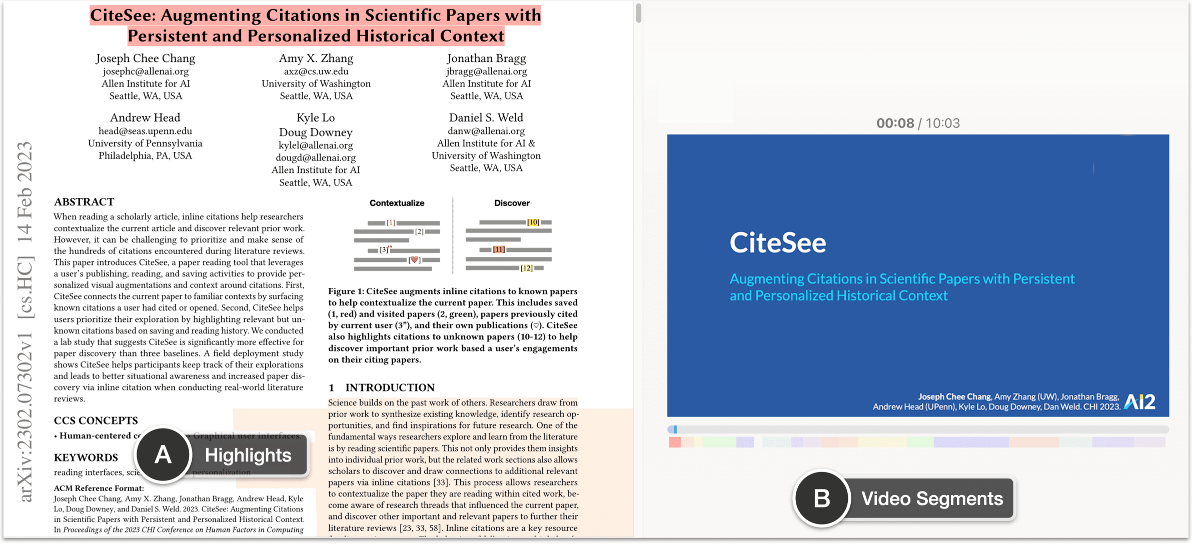
Thus, we developed a technology probe [35] (Fig. 2) where we could pre-link segments of a talk video to relevant passages in the paper (e.g., paragraphs, figures) and color-code them so that participants could switch between the two formats with lower cost. Before the study, one of the authors manually created the links between the papers and videos for three papers in each of the recruited participants' research fields (e.g., empirical HCI, systems HCI, NLP, CV). To create these links, the author followed criteria that were based on insights from the preliminary study: segment the video on slide transitions, and link segments to paragraphs based on content similarity (e.g., phrases, figures) while following the paper's reading order.

3.3 Study Procedure

The study consisted of two consecutive sessions. First, there was a formative session where participants took the perspective of paper readers and used the technology probe (Fig. 2) to read a paper where several passages were pre-linked to relevant segments of the talk video. Then, in a co-design session, participants took the perspective of paper authors and considered designs for combining their own research papers and talk videos.

For the formative session, participants chose their preferred paper from the set of pre-linked paper-video pairs and, while thinking aloud, read the paper using the technology probe for 20 minutes. In the probe, linked passages in the paper were highlighted, and participants could click on a linked passage to automatically navigate to the corresponding segment in the video. The video segments were also displayed under the video timeline, and participants could click on a video segment to scroll to the corresponding passage in the paper. After the reading period, participants were asked about the benefits and drawbacks of using the probe and the talk video during paper reading.

[1] Several participants identified with multiple disciplines.


arXiv:2302.07302v1 [cs.HC] 14 Feb 2023

CiteSee: Augmenting Citations in Scientific Papers with Persistent and Personalized Historical Context

Joseph Chee Chang
josephc@allenai.org
Allen Institute for AI
Seattle, WA, USA

Amy X. Zhang
axz@cs.uw.edu
University of Washington
Seattle, WA, USA

Jonathan Bragg
jbragg@allenai.org
Allen Institute for AI
Seattle, WA, USA

Andrew Head
head@seas.upenn.edu
University of Pennsylvania
Philadelphia, PA, USA

Kyle Lo
Doug Downey
kylel@allenai.org
dougd@allenai.org
Allen Institute for AI
Seattle, WA, USA

Daniel S. Weld
danw@allenai.org
Allen Institute for AI &
University of Washington
Seattle, WA, USA

ABSTRACT

When reading a scholarly article, inline citations help researchers contextualize the current article and discover relevant prior work. However, it can be challenging to prioritize and make sense of the hundreds of citations encountered during literature reviews. This paper introduces CiteSee, a paper reading tool that leverages a user's publishing, reading, and saving activities to provide personalized visual augmentations and context around citations. First, CiteSee connects the current paper to familiar contexts by surfacing known citations a user had cited or opened. Second, CiteSee helps users prioritize their exploration by highlighting relevant but unknown citations based on saving and reading history. We conducted a lab study that suggests CiteSee is significantly more effective for paper discovery than three baselines. A field deployment study shows CiteSee helps participants keep track of their explorations and leads to better situational awareness and increased paper discovery via inline citation when conducting real-world literature reviews.

CCS CONCEPTS

• **Human-centered c** Graphical user interfaces.

KEYWORDS

reading interfaces, scie personalization

ACM Reference Format:
Joseph Chee Chang, Amy X. Zhang, Jonathan Bragg, Andrew Head, Kyle Lo, Doug Downey, and Daniel S. Weld. 2023. CiteSee: Augmenting Citations in Scientific Papers with Persistent and Personalized Historical Context. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing*

Figure 1: CiteSee augments inline citations to known papers to help contextualize the current paper. This includes saved (1, red) and visited papers (2, green), papers previously cited by current user (3"), and their own publications (♡). CiteSee also highlights citations to unknown papers (10-12) to help discover important prior work based a user's engagements on their citing papers.

1 INTRODUCTION

Science builds on the past work of others. Researchers draw from prior work to synthesize existing knowledge, identify research opportunities, and find inspirations for future research. One of the fundamental ways researchers explore and learn from the literature is by reading scientific papers. This not only provides them insights into individual prior work, but the related work sections also allows scholars to discover and draw connections to additional relevant papers via inline citations [33]. This process allows researchers to contextualize the paper they are reading within cited work, become aware of research threads that influenced the current paper, and discover other important and relevant papers to further their literature reviews [23, 33, 58]. Inline citations are a key resource

CiteSee
Augmenting Citations in Scientific Papers with Persistent and Personalized Historical Context
Joseph Chee Chang, Amy Zhang (UW), Jonathan Bragg, Andrew Head (UPenn), Kyle Lo, Doug Downey, Dan Weld. CHI 2023.

Figure 2: Technology probe used during the formative studies. On the left, a PDF reader for the paper where passages linked to video segments are highlighted (A). On the right, a video player for the talk video accompanied by an interactive timeline and a bar displaying the location and length of segments linked to the paper (B). Linked passage-segment pairs are color-coded.

Then, participants took the perspective of authors and participated in a co-design session where they considered designs for combining their own research paper and talk video. To stimulate the participants and illustrate how to sketch designs, participants were provided with a slide deck that showed three example designs for interfaces that combined papers and videos. Participants were asked to think aloud and sketch designs in the slide deck, which was pre-populated with screenshots of the pages and key frames of the participant's paper and talk video that they provided prior to the study. To sketch out their designs, participants could resize and crop the screenshots, draw shapes, and use text boxes to describe the designs. During the session, one or two of the authors helped with the sketching by making edits based on participants' descriptions, and asked questions to encourage participants to elaborate further on their ideas or to consider alternative designs.

Aside from one in-person participant, all participants joined remotely through Google Meet.[2] This study was approved by our internal review board, and each participant was paid 45 USD for their time.

3.4 Findings

During the study sessions, we recorded participants' screens and the audio, which were then manually transcribed. Through a thematic analysis, the transcripts were coded and these codes were grouped into themes to identify the main insights from the study. Additionally, a thematic analysis was also conducted on the various designs for the co-design sessions to typify these designs based on their similarities. Based on insights from the reader and author sessions, we distilled design goals for augmenting research papers with talk videos.

3.4.1 As Readers. In contrast to participants in the preliminary study, participants in this study followed different consumption patterns with the probe: five mainly read the paper and occasionally switched to the video, and nine followed the video while intermittently pausing to dive into the paper. Based on their experiences with the technology probe, participants noted various ways in which talk videos enriched the paper. Specifically, most participants (11/14) mentioned that the video provided summaries that were easier to consume than *"dense parts of the paper"* (P5). Asides from summarizing, participants (7/14) also mentioned that videos explain details differently and that these alternative explanations were useful when they struggled to understand the paper. Participants also noted the significance of the audio-visual nature of videos. Several participants liked authors' narrations in videos (4/14) as listening could be less demanding or more *"passive"* than reading (P10), and since they could have the *"author narrate [figures] for [them]"* (P2). In terms of the visuals, various participants (5/14) described how illustrations, animations, or clips in the talk videos could better illustrate certain aspects of the paper. For example, P14 mentioned how a clip showing a demo of an interface helped them *"get a more clear idea of what the interaction would look like"*. Finally, a majority of participants (11/14) mentioned how watching the videos or skimming the video-based highlights in the paper gave them an overview of the papers and allowed them to *"make note"* (P1) of details they wanted to dive deeper into—serving as a *"launching pad"* into the paper (P3).

Despite these benefits, however, there were various interaction challenges that limited participants' use of talk videos even with the support of our technology probe (Fig 2). For example, as a paper automatically scrolled to the relevant passage when the video progressed to the next segment, participants mentioned that the probe could disrupt their reading (3/14) or cause them to get lost (6/14). Additionally, participants (4/14) mentioned how they could not predict what information would be contained in a video segment before actually watching the segment and, therefore, could not anticipate when a segment would be useful or not. Finally, as video segments were linked to relatively lengthy passages in papers, various participants (8/14) mentioned how it was difficult to locate a detail mentioned in a video segment in the paper, or to distinguish what in the paper passages had been covered or not by the segment.

3.4.2 As Authors. During the co-design sessions, participants produced a variety of designs for paper and video combinations. As seen in Table 1, several of the participants' designs shared structural similarities, but differed in terms of specific details or features. Participants considered both designs where the video supported paper reading and where the paper enhanced video watching, and some participants envisioned new formats where neither format was the main one.

Based on participants' designs and their comments during the sessions, we distilled the main goals that participants considered when designing the combinations. One of the main goals that participants (12/14) mentioned was to enable users to flexibly switch the level of detail at which they consume the content. Specifically, the user can switch from the video to the paper to *"expand to see more details"* (P1) or switch from the paper to the video to *"skip"* (P5) sections that are less interesting. Beyond consumption, several participants (7/14) considered combinations that visually represented paper passages with the video to support navigation in the paper. For example, P2's design presented slides from the video as a visual outline that the user can use to navigate the paper. Finally, due to their difficulties in locating details from the video in the paper and vice-versa during the reading session, several participants (5/14) designed interfaces that supported more fine-grained links (e.g., highlighting passages in the paper that were mentioned in the video).

Beyond revealing what authors wanted from the combinations, the co-design sessions also revealed constraints to possible designs. While several participants created designs that replaced paper passages with video elements, most participants (7/14) advocated against replacing content. Some participants mentioned that *"videos are rarely a one-to-one representation of a paper"* (P3) and that replacing could *"delete information"* (P10), while others noted how one format provided *"supplementary information"* for the other (P6) and it could be more beneficial to consume them together. Additionally, P2 mentioned how they dedicated *"significant effort"* in authoring their paper and video, and that they would want users to look at both artifacts. Another constraint was that, despite considering designs where the user mainly interacted with the video, most participants (7/14) considered the video as *"a way to advertise"* their paper (P4) and that *"ultimately"* (P11) they wanted to direct

[2] https://meet.google.com

Primary Format	Type of Design	Feature Differences
Paper	**Linked video popups**: display popup with video segment when user interacts with a linked paper passages.	Link popups on text (P4, P8, P11, P13), figures or tables (P2, P6, P8), or definitions and sections headers (P14).
		Display popup based on user's selected text (P3).
		Display thumbnail instead of video segment (P6, P7).
	Overlaid videos: overlaying video segments on relevant passages of the paper.	Overlay on videos on figures (P4, P8, P13).
		Overlay visual guides from video on tables or figures (P2, P6, P8), or mathematical equations (P8).
	Video-based outline: an outline or table of contents for the paper based on the links between video segments and paper passages.	Panel that displays a list of the slides extracted from the video as a navigational map (P2).
		Table of content for the paper but containing the titles of video sections (P11), and transcript summaries or video thumbnails (P12).
Video	**Position-sensitive details**: hovering over elements in a video frame to reveal a tooltip with related details from the paper.	Hovering over keywords to see definitions (P7, P10), summarized tables to reveal the detailed tables from the paper (P10), or elements of a system to reveal related explanations from the paper (P13).
	Guiding tooltips: tooltips that appear as the video plays to encourage the viewer to check related sections of the paper.	Tooltip is accessible through an icon that is overlaid on the video (P1), tooltip text is overlaid on the video (P10, P14) or text is shown next to the video (P1).
	Side commentary: panel next to the video that displays relevant passages from the paper as the video plays.	Commentary can include the full passages from the paper (P13), only information from the passages that is not included in the video (P5), or a summary of the passages (P5, P13)
Combined	**Interweaved paper and video**: new format that interweaves elements from the paper with those from the video.	Embedding images, animated GIFs, and clips from the video inbetween passages of text (P6), inbetween summarized passages of text (P3), or replace text with the video elements (P2, P9).
	Adaptive side-by-side: paper and video displayed side-by-side but adaptively changes the size of each format.	User can manually change the amount of space taken by each format or the interfaces automatically changes them by inferring the user's needs (P4).

Table 1: Overview of the co-design session that captured how authors envisioned combining their papers and talk videos. The table describes the types of designs that authors produced and the features that authors proposed for the different design types. Additionally, the design types were categorized based on their primary consumption formats.

the user to their paper. This was reflected through their *"guiding tooltip"* designs (Table 1).

Finally, we asked participants about whether they would be willing to create links between their papers and talk videos to enable the combinations they designed. All participants mentioned that they would create these links as it could increase the visibility of their work and *"help as many people as possible to read and understand [my paper]"* (P9). Although several participants mentioned that they would want the process of linking to be as easy as possible, all participants also mentioned that they would not want the process to be completely automatic. Instead, they would need to be *"involved in the process"* (P3) to check and edit links made by an automatic pipeline. Interestingly, some participants even expressed how they would be willing to change how they author videos to make this semi-automatic linking process easier and more accurate: *"I might start baking this stuff into the slide deck"* (P10) and *"It might have a positive influence on [...] how I design the the slides like making them more correlated to the paper"* (P6).

3.4.3 Design Goals. Based on the insights from the reader and author sessions, we distilled the following design goals for combining research papers and talk videos:

- DG1: Allow readers to both focus on either the paper or video, but also enable them to fluidly switch between the two formats when needed.
- DG2: Surface visuals from the video to help readers anticipate its content and to visually outline the paper.
- DG3: Present fine-grained links that aid in the association of related details across formats.
- DG4: Avoid occluding or replacing the content in a format with content from the other.
- DG5: Aid in the creation of links between papers and videos but grant authors control over how they want to present their work.

4 PAPEOS

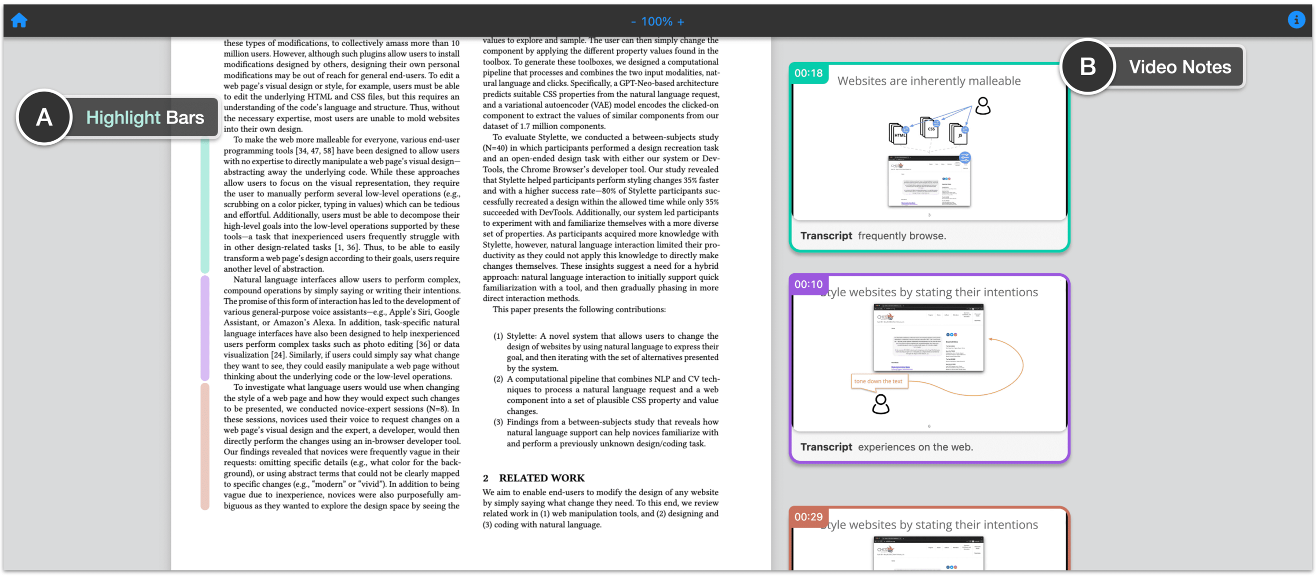
Based on the design goals, we developed *Papeos* (Figure 3), a novel reading experience that augments research papers with localized clips from the corresponding talk videos. In this section, we first illustrate the reading interface for Papeos. Then, we describe a mixed-initiative interface that allows paper authors to create Papeos for their papers and talk videos with lowered effort.

4.1 Papeo Reading Interface

The Papeo reader is designed to support a variety of use cases, such as leveraging linked video segments to guide users when text skimming (§4.1.1), support users in fluidly switching between reading text passages and watching video segments to adjust the level of details they wish to consume (§4.1.2), and allow users to continuously watch a talk video while having access to additional details in corresponding text passages (§4.1.3). For this, the Papeo reader presents video segments as *video notes* placed on the right side of paper pages and localized approximately next to their linked passages (Fig. 3). Since each page of a paper could contain multiple linked passages and video notes, Papeo renders color-coded *highlight bars* next to passages and video notes alongside a paper for linked paper passages and video segments (DG4).

4.1.1 Video-Supported Skimming. Researchers often skim read to get a high level understanding of research papers [21]. By scrolling through the Papeo reader, the user can skim the paper by looking through the highlight bars and accompanying video notes. The highlight bars (Fig. 3a) reveal the portions of the paper that the author considered important when creating their video. The video notes (Fig. 3b) reveal the content of the video segment through the thumbnail (i.e., the first frame of the video segment) and the first line from the transcript which can, respectively, visually represent and summarize these passages of importance. By skimming based on these features, for example, a reader could prioritize reading high-level descriptions of a user interface and a few important quotes from the user study that were included in the conference presentation, instead of reading all implementation details and quotes that were not included. By remembering the thumbnails and their relevant locations in the paper, the user can also develop a "spatial mental map" [66] of the paper to help them return to desired content in the paper (DG2). If the thumbnail or transcript line surfaces insufficient information about the video segment, the user can also hover and scrub over the highlight bar to peek into different moments in the segment (Fig. 4a).

4.1.2 Fluid Switching between Paper and Video. As the user is reading through the paper, they may struggle to understand certain passages or may be less interested in particular sections. For example, an expert user might need to learn the implementation details of a machine learning paper but was already familiar with the background and related work. In these cases, if a video note is linked, the user can watch an alternative and/or summarized explanation of the passage by clicking on the highlight bar or video note itself (DG1). Clicking on the bar or note "activates" the video note (Fig. 4b): the thumbnail switches into a video player that starts playing the segment, the full transcript for the segment is shown, and the note increases in size. If it is only approximately aligned with the highlight bar, the note also moves to be exactly aligned—pushing away other notes if they would overlap. As the video plays, lines of the transcript are highlighted so that the user can discern what has already been spoken.


- 100% +
A
Highlight Bars
B
Video Notes
00:18
Websites are inherently malleable
Transcript frequently browse.
00:10
Style websites by stating their intentions
tone down the text
Transcript experiences on the web.
00:29
Style websites by stating their intentions

these types of modifications, to collectively amass more than 10 million users. However, although such plugins allow users to install modifications designed by others, designing their own personal modifications may be out of reach for general end-users. To edit a web page's visual design or style, for example, users must be able to edit the underlying HTML and CSS files, but this requires an understanding of the code's language and structure. Thus, without the necessary expertise, most users are unable to mold websites into their own design.

To make the web more malleable for everyone, various end-user programming tools [34, 47, 58] have been designed to allow users with no expertise to directly manipulate a web page's visual design—abstracting away the underlying code. While these approaches allow users to focus on the visual representation, they require the user to manually perform several low-level operations (e.g., scrubbing on a color picker, typing in values) which can be tedious and effortful. Additionally, users must be able to decompose their high-level goals into the low-level operations supported by these tools—a task that inexperienced users frequently struggle with in other design-related tasks [1, 36]. Thus, to be able to easily transform a web page's design according to their goals, users require another level of abstraction.

Natural language interfaces allow users to perform complex, compound operations by simply saying or writing their intentions. The promise of this form of interaction has led to the development of various general-purpose voice assistants—e.g., Apple's Siri, Google Assistant, or Amazon's Alexa. In addition, task-specific natural language interfaces have also been designed to help inexperienced users perform complex tasks such as photo editing [36] or data visualization [24]. Similarly, if users could simply say what change they want to see, they could easily manipulate a web page without thinking about the underlying code or the low-level operations.

To investigate what language users would use when changing the style of a web page and how they would expect such changes to be presented, we conducted novice-expert sessions (N=8). In these sessions, novices used their voice to request changes on a web page's visual design and the expert, a developer, would then directly perform the changes using an in-browser developer tool. Our findings revealed that novices were frequently vague in their requests: omitting specific details (e.g., what color for the background), or using abstract terms that could not be clearly mapped to specific changes (e.g., "modern" or "vivid"). In addition to being vague due to inexperience, novices were also purposefully ambiguous as they wanted to explore the design space by seeing the

values to explore and sample. The user can then simply change the component by applying the different property values found in the toolbox. To generate these toolboxes, we designed a computational pipeline that processes and combines the two input modalities, natural language and clicks. Specifically, a GPT-Neo-based architecture predicts suitable CSS properties from the natural language request, and a variational autoencoder (VAE) model encodes the clicked-on component to extract the values of similar components from our dataset of 1.7 million components.

To evaluate Stylette, we conducted a between-subjects study (N=40) in which participants performed a design recreation task and an open-ended design task with either our system or DevTools, the Chrome Browser's developer tool. Our study revealed that Stylette helped participants perform styling changes 35% faster and with a higher success rate—80% of Stylette participants successfully recreated a design within the allowed time while only 35% succeeded with DevTools. Additionally, our system led participants to experiment with and familiarize themselves with a more diverse set of properties. As participants acquired more knowledge with Stylette, however, natural language interaction limited their productivity as they could not apply this knowledge to directly make changes themselves. These insights suggest a need for a hybrid approach: natural language interaction to initially support quick familiarization with a tool, and then gradually phasing in more direct interaction methods.

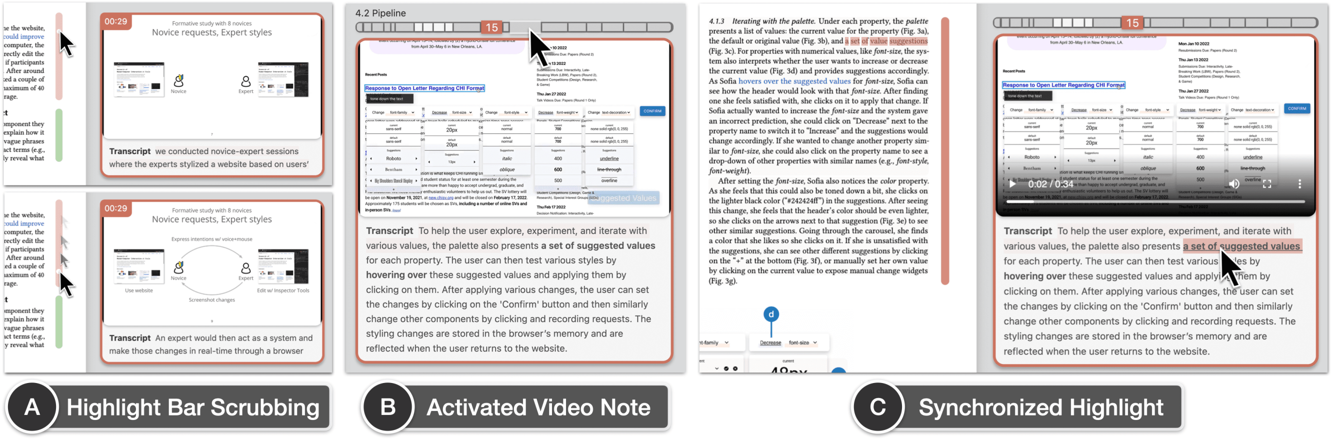
This paper presents the following contributions:

(1) Stylette: A novel system that allows users to change the design of websites by using natural language to express their goal, and then iterating with the set of alternatives presented by the system.
(2) A computational pipeline that combines NLP and CV techniques to process a natural language request and a web component into a set of plausible CSS property and value changes.
(3) Findings from a between-subjects study that reveals how natural language support can help novices familiarize with and perform a previously unknown design/coding task.

2 RELATED WORK

We aim to enable end-users to modify the design of any website by simply saying what change they need. To this end, we review related work in (1) web manipulation tools, and (2) designing and (3) coding with natural language.

Figure 3: The Papeo reader extends a PDF reader by incorporating highlight bars (A) alongside passages in a research paper that are linked to segments in the corresponding talk video. These video segments are displayed as video notes (B) that are localized next to the linked passages and present a thumbnail, a line from the transcript, and the total duration of the segment.



Transcript we conducted novice-expert sessions where the experts stylized a website based on users'

Transcript An expert would then act as a system and make those changes in real-time through a browser

Transcript To help the user explore, experiment, and iterate with various values, the palette also presents **a set of suggested values** for each property. The user can then test various styles by **hovering over** these suggested values and applying them by clicking on them. After applying various changes, the user can set the changes by clicking on the 'Confirm' button and then similarly change other components by clicking and recording requests. The styling changes are stored in the browser's memory and are reflected when the user returns to the website.

4.1.3 Iterating with the palette. Under each property, the *palette* presents a list of values: the current value for the property (Fig. 3a), the default or original value (Fig. 3b), and a set of value suggestions (Fig. 3c). For properties with numerical values, like *font-size*, the system also interprets whether the user wants to increase or decrease the current value (Fig. 3d) and provides suggestions accordingly. As Sofia hovers over the suggested values for *font-size*, Sofia can see how the header would look with that *font-size*. After finding one she feels satisfied with, she clicks on it to apply that change. If Sofia actually wanted to increase the *font-size* and the system gave an incorrect prediction, she could click on "Decrease" next to the property name to switch it to "Increase" and the suggestions would change accordingly. If she wanted to change another property similar to *font-size*, she could also click on the property name to see a drop-down of other properties with similar names (e.g., *font-style*, *font-weight*).

After setting the *font-size*, Sofia also notices the *color* property. As she feels that this could also be toned down a bit, she clicks on the lighter black color ("#242424ff") in the suggestions. After seeing this change, she feels that the header's color should be even lighter, so she clicks on the arrows next to that suggestion (Fig. 3e) to see other similar suggestions. Going through the carousel, she finds a color that she likes so she clicks on it. If she is unsatisfied with the suggestions, she can see other different suggestions by clicking on the "+" at the bottom (Fig. 3f), or manually set her own value by clicking on the current value to expose manual change widgets (Fig. 3g).

Figure 4: Illustration of features supported by the Papeo reader: (A) hovering and scrubbing over highlight bars allows users to quickly scrub through the linked video segments; (B) activated video notes present the users with player controls, the full transcript for the segment, and a segmented timeline for the whole video that presents the paper section where a note is located when the user hovers a segment; and (C) synchronized highlights are shown as blue text in the paper and bold text in the video transcript, and, when the user hovers over them, they become highlighted in sync.

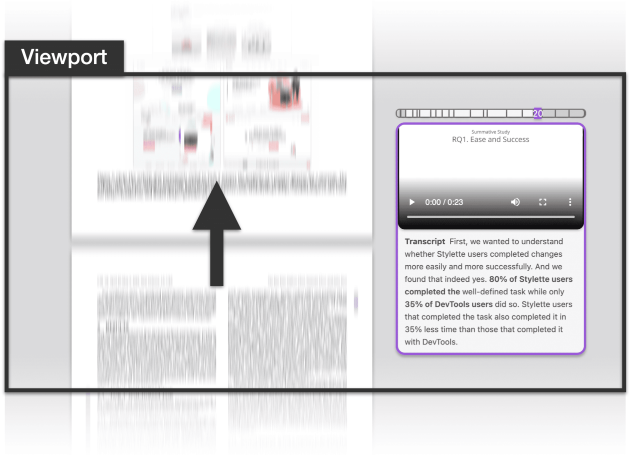
While watching the video note, the user may want to read up on the same information in the paper to acquire more details or to take in a more formalized explanation. To focus back on the reading, the user can pause the video note through the player controls or by clicking anywhere outside the note to "deactivate" it. As users may start reading while the video note plays and forget to deactivate it, each video note only streams one video segment to minimize disruption. Thus, by default, once the video note reaches the end of the current segment, the player stops instead of progressing to the next segment in the video—unlike the preliminary research probe (DG1).

Finally, when switching between the two formats mid-segment, the user may struggle to identify a detail in one format in the other due to the wording differences or the amount of text they have to traverse through. For example, if a reader watches a progressive animation explaining the architecture of a machine learning model and becomes curious about a specific hyper-parameter, it can be difficult for them to find the value of the hyper-parameter in the paper. To address this challenge, the Papeo reader provides *synchronized highlights* (Fig. 4c). Based on how the paper author created the Papeo, certain words or phrases in the video transcript and paper are bold and underlined. When the user hovers over these words or phrases, they are highlighted and the related words or phrases in the other format are also highlighted to help the user discern and match details across the formats (DG3).

4.1.3 *Video-Centric Consumption.*

Besides skimming the text of the paper and switching between text and video segments, Papeo also support users if they wish to watch multiple segments or even the entire video continuously. While each video note only streams one segment from the video, the Papeo reader also allows the user to focus on and watch the video notes in order (DG1). When a video note ends, the user is provided with the option to re-watch the video segment or to jump to the next. To watch the whole video with no interruptions, the user can activate the "autoplay" setting to automatically navigate and watch through all video segments.

Whenever the user navigates between video notes, the paper scrolls automatically to the location of the next video note to allow the user to also check and read the linked paper passages (DG1).



Transcript First, we wanted to understand whether Stylette users completed changes more easily and more successfully. And we found that indeed yes. **80% of Stylette users completed the** well-defined task while only **35% of DevTools users** did so. Stylette users that completed the task also completed it in 35% less time than those that completed it with DevTools.

Figure 5: During video note-centric scrolling, the user can navigate to the video note for the next video segment, which takes over the viewport position of the current video note. With the video note fixed in position, the paper scrolls to the passages linked to the next video segment. This allows the user to continuously watch video segments without interruption while always having access to the linked passages next to the current video playback.

To minimize disruption during autoplay, the Papeo reader employs *video note-centric scrolling* (Fig. 5). In this type of scrolling, the different video notes stay fixed in same position while the paper scrolls to corresponding linked passages as the videos play. Thus, while the user is technically navigating between video notes and scrolling through the paper, they can continue to watch the video by fixing their gaze on the same part of their screen (DG1).

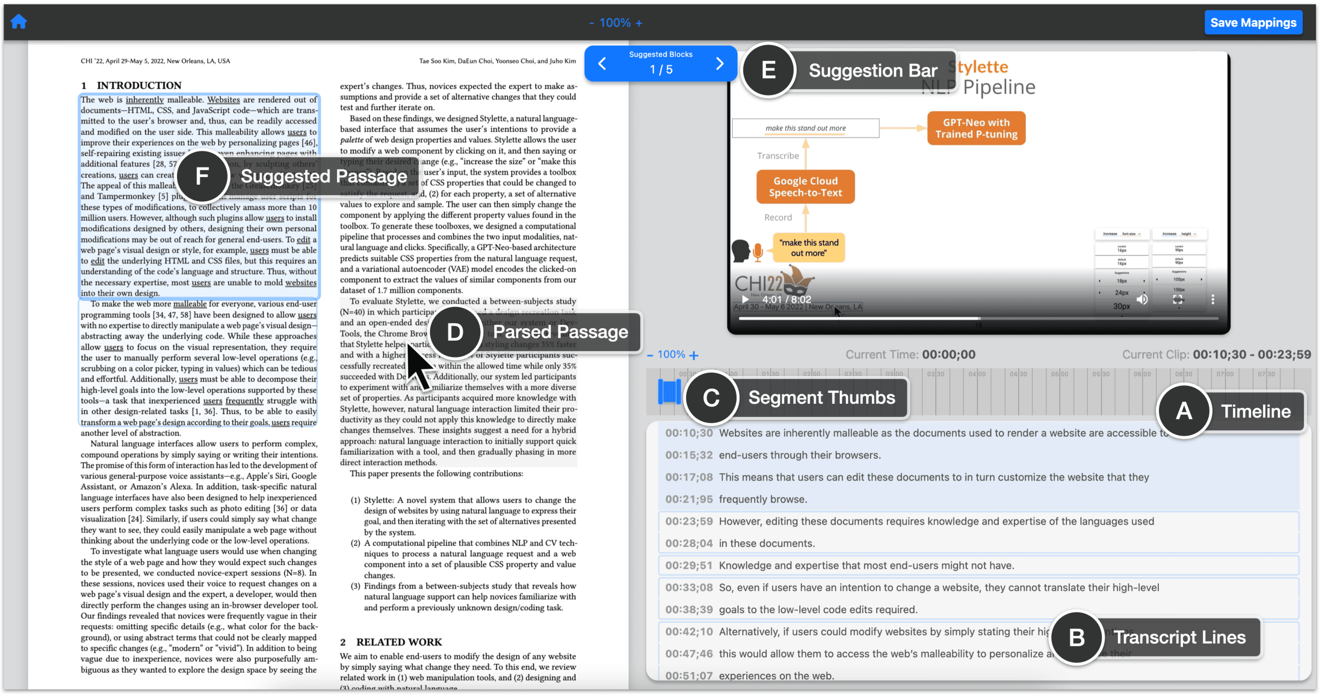
Above activated video notes, the Papeo reader also provides a timeline (top in Fig. 4b) to allow the user to navigate between video notes and, consequently, navigate the paper based on these (DG2). The timeline is fragmented where each block represents a video note and the user can navigate to these notes by clicking on the blocks—navigation occurs through *note-centric scrolling*. To help the user track where they are in the video and what they have already watched, the block for the current video note is color-coded and blocks for notes that have been watched are opaque. Before navigating to a note, the user can hover over a block to see the title of the section or sub-section where the note is located (*“4.2 Pipeline”* in Fig. 4b)—allowing them to check where they will navigate to and what type of content may be contained in the video note (DG2).

4.2 Papeo Authoring Interface

To create Papeos, we propose an authoring interface (Figure 6) through which paper authors can link their papers and talk videos—granting them control over how these formats are linked (DG5). We developed this interface through an iterative design process.

With early versions of the interface, we observed that authors dedicated significant effort to segment their videos and to search for paper passages that were relevant to these segments. To address this challenge, we adopted a mixed-initiative design for the authoring interface by providing automatic suggestions for segmenting videos and for linking papers and videos. To start authoring, the author first uploads a PDF of their paper and the talk video with transcript. Then, they access the authoring interface that consists of two panels: a video segmenter where the author can select segments of the video, and a paper annotator where they can then choose the passages to link to the segment (Fig. 6).

To start linking their paper and video, the author first needs to create a video segment. To do so, they can watch the video, click on the timeline to create an initial segment, and drag the start and end thumbs to select a time range (Fig. 6a). Alternatively, authors can read the transcript and directly select a group of transcript lines (Fig. 6b). To improve efficiency, the interface also automatically groups transcript lines at the sentence-level to act as segment suggestions. When the author clicks on a group, the interface selects a segment that contains all of the lines in the group. When creating a segment from the transcript, the author can select or de-select lines to correct errors in the segment suggestions, or further fine-tune the start and end times by using the thumbs in the timeline since transcript lines do not always align with sentence boundaries (Fig. 6c).



CHI ’22, April 29-May 5, 2022, New Orleans, LA, USA

Tae Soo Kim, DaEun Choi, Yoonseo Choi, and Juho Kim

1 INTRODUCTION

The web is inherently malleable. Websites are rendered out of documents—HTML, CSS, and JavaScript code—which are transmitted to the user’s browser and, thus, can be readily accessed and modified on the user side. This malleability allows users to improve their experiences on the web by personalizing pages [46], self-repairing existing issues [illegible] even enhancing pages with additional features [28, 57 [illegible] ion, by sculpting others’ creations, users can creat[illegible] The appeal of this malleab[illegible] the Greasemonkey [25] and Tampermonkey [5] plu[illegible] manage user scripts for these types of modifications, to collectively amass more than 10 million users. However, although such plugins allow users to install modifications designed by others, designing their own personal modifications may be out of reach for general end-users. To edit a web page’s visual design or style, for example, users must be able to edit the underlying HTML and CSS files, but this requires an understanding of the code’s language and structure. Thus, without the necessary expertise, most users are unable to mold websites into their own design.

To make the web more malleable for everyone, various end-user programming tools [34, 47, 58] have been designed to allow users with no expertise to directly manipulate a web page’s visual design—abstracting away the underlying code. While these approaches allow users to focus on the visual representation, they require the user to manually perform several low-level operations (e.g., scrubbing on a color picker, typing in values) which can be tedious and effortful. Additionally, users must be able to decompose their high-level goals into the low-level operations supported by these tools—a task that inexperienced users frequently struggle with in other design-related tasks [1, 36]. Thus, to be able to easily transform a web page’s design according to their goals, users require another level of abstraction.

Natural language interfaces allow users to perform complex, compound operations by simply saying or writing their intentions. The promise of this form of interaction has led to the development of various general-purpose voice assistants—e.g., Apple’s Siri, Google Assistant, or Amazon’s Alexa. In addition, task-specific natural language interfaces have also been designed to help inexperienced users perform complex tasks such as photo editing [36] or data visualization [24]. Similarly, if users could simply say what change they want to see, they could easily manipulate a web page without thinking about the underlying code or the low-level operations.

To investigate what language users would use when changing the style of a web page and how they would expect such changes to be presented, we conducted novice-expert sessions (N=8). In these sessions, novices used their voice to request changes on a web page’s visual design and the expert, a developer, would then directly perform the changes using an in-browser developer tool. Our findings revealed that novices were frequently vague in their requests: omitting specific details (e.g., what color for the background), or using abstract terms that could not be clearly mapped to specific changes (e.g., “modern” or “vivid”). In addition to being vague due to inexperience, novices were also purposefully ambiguous as they wanted to explore the design space by seeing the expert’s changes. Thus, novices expected the expert to make assumptions and provide a set of alternative changes that they could test and further iterate on.

Based on these findings, we designed Stylette, a natural language-based interface that assumes the user’s intentions to provide a *palette* of web design properties and values. Stylette allows the user to modify a web component by clicking on it, and then saying or typing their desired change (e.g., “increase the size” or “make this [illegible]”). Based on the user’s input, the system provides a toolbox that contains (1) a set of CSS properties that could be changed to satisfy the request, and, (2) for each property, a set of alternative values to explore and sample. The user can then simply change the component by applying the different property values found in the toolbox. To generate these toolboxes, we designed a computational pipeline that processes and combines the two input modalities, natural language and clicks. Specifically, a GPT-Neo-based architecture predicts suitable CSS properties from the natural language request, and a variational autoencoder (VAE) model encodes the clicked-on component to extract the values of similar components from our dataset of 1.7 million components.

To evaluate Stylette, we conducted a between-subjects study (N=40) in which participa[illegible]ned a design recreation task and an open-ended desi[illegible] either our system or DevTools, the Chrome Brow[illegible] tool. [illegible] that Stylette helpe[illegible] partic[illegible] styling changes 35% faster and with a higher [illegible] cess [illegible] of Stylette participants successfully recreated [illegible] within the allowed time while only 35% succeeded with De[illegible]s. Additionally, our system led participants to experiment with an[illegible] miliarize themselves with a more diverse set of properties. As participants acquired more knowledge with Stylette, however, natural language interaction limited their productivity as they could not apply this knowledge to directly make changes themselves. These insights suggest a need for a hybrid approach: natural language interaction to initially support quick familiarization with a tool, and then gradually phasing in more direct interaction methods.

This paper presents the following contributions:

(1) Stylette: A novel system that allows users to change the design of websites by using natural language to express their goal, and then iterating with the set of alternatives presented by the system.
(2) A computational pipeline that combines NLP and CV techniques to process a natural language request and a web component into a set of plausible CSS property and value changes.
(3) Findings from a between-subjects study that reveals how natural language support can help novices familiarize with and perform a previously unknown design/coding task.

2 RELATED WORK

We aim to enable end-users to modify the design of any website by simply saying what change they need. To this end, we review related work in (1) web manipulation tools, and (2) designing and (3) coding with natural language.

Current Time: 00:00;00

Current Clip: 00:10;30 - 00:23;59

00:10;30 Websites are inherently malleable as the documents used to render a website are accessible to
00:15;32 end-users through their browsers.
00:17;08 This means that users can edit these documents to in turn customize the website that they
00:21;95 frequently browse.
00:23;59 However, editing these documents requires knowledge and expertise of the languages used
00:28;04 in these documents.
00:29;51 Knowledge and expertise that most end-users might not have.
00:33;08 So, even if users have an intention to change a website, they cannot translate their high-level
00:38;39 goals to the low-level code edits required.
00:42;10 Alternatively, if users could modify websites by simply stating their hig[illegible]ntions,
00:47;46 this would allow them to access the web’s malleability to personalize a[illegible]e their
00:51;07 experiences on the web.

Figure 6: The Papeo authoring interface consists of a parsed PDF and a video segmenter. The segmenter timeline (A) displays the segments that have been created so far. The user can create a segment by clicking on the timeline or selecting lines in the transcript (B), and then dragging on the thumbs to fine-tune its length (C). Then, the user can manually click on relevant passages (D) to link them to a video segment, or review and adopt the automatically generated linking suggestions (E,F).

After creating a video segment, the author can then link it to relevant passages (e.g., paragraphs, figures) in the paper. Instead of requiring authors to manually select paragraphs or figures, the interface presents these as clickable targets so that authors can select entire paragraphs with single clicks (Fig. 6d). This is made possible by leveraging the pre-trained VILA model to automatically parse the paper PDF and identify paragraph, figure, and table boundaries [71]. Since AI models can occasionally make mistakes, the author can also click-and-drag over an area of the paper to manually select a passage to recover from errors. One remaining challenge here is that it can be time consuming to search through the paper for relevant passages. For this, the interface suggests the five most likely passages based on the current video segment (i.e., link suggestions). After a video segment is created, the paper automatically scrolls to the highlighted top-1 suggestion for the author to review (Fig. 6f). The author can further review the top 2–5 suggestions using the suggestion navigation bar (Fig. 6e).

Beyond the coarse-grained links between paper passages and video segments, the Papeo reader interface also supports fine-grained links (i.e., synchronized highlights) to help readers identify specific details. To create these fine-grained links, authors can select a paper passage or video segment that has been linked and click on the "Create Sync Highlight" button at the top of the interface. Then, the author can select words or phrases in the passages and the transcript of the video segment that they wish to link. After selecting the words, the author stores the synchronized highlight by clicking on the "Save Sync Highlight" button, and can proceed to create more synchronized highlights for the linked segment and passages.

4.2.1 Automatic Suggestions. To make authoring Papeos more efficient, Papeos' authoring interface generates automatic suggestions for video segmentation and for paper-video linking. During development, to evaluate multiple algorithms and AI models for generating suggestions, we collected a small ground-truth dataset by having three of the authors and five recruited researchers link their papers and talk videos (total of 8 pairs) using an initial version of the authoring interface without automatic suggestions. For techniques with no tunable hyperparameters, we evaluated the technique on the whole ground-truth dataset. For techniques with hyperparameters, we performed 4-fold cross-validation where 25% of the data was used to identify the best hyperparameter values and the remaining 75% was used to evaluate the technique with the best identified hyperparameter values. For each technique, we specify the hyperparamters, if any.

Segment Suggestions: We tested three different techniques for automatically segmenting videos (i.e., shot detection): (1) calculating pixel changes in the HSV (i.e., Hue, Saturation, and Value) colorspace between adjacent frames [10], (2) template matching which calculates the spatial similarity between a frame and the previous key frame [74], and (3) segmenting the video at every transcript line containing a punctuation—as authors are likely to transition between scenes at the end of sentences. Both the HSV change and template matching techniques had two hyperparameters: minimum length of a segment, and threshold (i.e., HSV change or spatial similarity value that needs to be exceeded to predict a segment boundary). For evaluation, we calculated the number of predicted segment boundaries that were within 3 seconds of ground-truth boundaries to calculate precision, recall and F1-score. Based on the interaction we designed, we expected that it would be easier (i.e., fewer clicks) for authors to merge segment suggestions compared to splitting them, so we used the F3-score, which gives more weight to favor over-segmenting (i.e., more segments) and decided to adopt the punctuation-based auto-segmenter (Table 2).

Algorithm	Precision	Recall	F1	F2	F3
Punctuation	0.405	**0.906**	0.541	0.701	**0.786**
HSV Change	0.499	0.805	0.605	**0.706**	0.751
Template Match	**0.577**	0.758	**0.635**	0.698	0.725

Table 2: Recall, precision, and F1-, F2- and F3-scores for the algorithms tested for video segmentation. Highest values for each metric are shown in bold, and the technique used in the authoring interface is shown in blue.

Linking Suggestions: Currently, the authoring interface provides *text* passage linking suggestions that appear immediately *after* a video segment was created. We initially aimed to also automatically identify video frames similar to figures and tables in the papers since authors often reuse figures and tables in their presentations. However, it became clear in early design iterations that mapping figures between papers and videos was a relatively simple task for test users. In contrast, they spent much greater effort when trying to find the right passages when mapping to text.

Therefore, we focused on providing text passage linking suggestions, and used the ground-truth video segments from our dataset to test the following measures for matching text from the segments' transcripts to text in paper passages: (1) cosine similarity based on two text embedding models (i.e., Specter [18] and MiniLM [82]), (2) ROUGE-L score [50], and (3) a baseline that chooses the first paragraph of a random section in the paper. We designed the baseline based on the assumption that talk videos provide an overview of the paper and, as a result, might state information included in the overviews of each section (i.e., the first paragraphs). As seen from the results (Table 3), ROUGE-L had the highest top-1 accuracy while MiniLM embeddings had the highest top-5 accuracy. We then combined these two measures by simply adding the two scores which achieved higher top-1 and top-5 accuracies.

Finally, we noticed how videos typically present information content in the same order as the paper—i.e., after linking a segment and passage, the next video segment would likely link to passages that appear later in the paper. Based on this, we developed an additional technique that adapts the Viterbi algorithm [22]. Using this technique, we can consider, simultaneously, the semantic similarity between paper text and video transcripts, and how information might be presented in similar ordering in the two formats (e.g., background, methods, and then evaluation). More specifically, the potential links between segments and passages are considered to be states, and an observation is whether the segment and passage are actually linked. In this context, we first normalized the combined measure of MiniLM + ROUGE to use as the emission probability (i.e., probability of linking a segment to each passage). Then, we modeled the transition probability as a hyperparameter of the likelihood of

Algorithm	Top-1	Top-5
Random first paragraph of a section	0.029	0.080
SPECTER Embeddings	0.399	0.623
MiniLM Embeddings	0.464	0.768
ROUGE-L Score	0.493	0.739
Combined (MiniLM + ROUGE-L)	0.572	0.797
Viterbi with Combined	**0.626**	**0.863**

Table 3: Top-1 and top-5 accuracy for the algorithms tested for linking paper passages and video segments. Highest values for each metric are shown in bold, and the technique used in the authoring interface is shown in blue.

linking a video segment to a passage in-order and the remaining probability becomes the likelihood of linking in reverse order.[3] This technique improved on both the top-1 and top-5 accuracies and was used to provide suggestions in the authoring interface.

4.2.2 Preliminary User Evaluation. To test the feasibility and costs of authors creating Papeos for their readers, we conducted a preliminary evaluation with 6 researchers (3 systems HCI, 3 empirical HCI, and 1 computer vision) to author a Papeo using their own research paper and talk videos. In general, participants mentioned that it was easy to use the authoring interface to link their papers and videos, and that they were enthusiastic to author Papeos for future papers through the interface. This evaluation demonstrated that participants spent an average of 25 minutes and 22 seconds (SD=5:31, max=30:17, min=15:19) to fully link their paper and video[4]. Additionally, we measured how frequently the authors used at least one of the top-5 suggestions when linking a segment to passages, and saw that suggestions were used for 71.3% (SD=11.6%, max=82.1%, min=57.1%) of all linked segments. In sum, we showed that authors can use our current authoring interface to create Papeos for their own papers with reasonable effort, and leave further automation and evaluation for future work.

4.3 Implementation Details

We implemented the reading and authoring interfaces for Papeos in around 6,500 lines of TypeScript, ReactJS, and CSS. For the PDF reader, we adapted our own open-source PDF reader library[5] and, for the video player, we used the ReactPlayer package.[6] The backend and AI-based suggestions were implemented using around 1,600 lines of Python code. We used a Flask server, the HuggingFace Transformer library[7] for the SPECTER [18] and MiniLM [82] models, and the PySceneDetect[8] and OpenCV[9] packages for shot detection based on the HSV colorspace and template matching, respectively.

5 USER STUDY

Through our formative study, we observed that talk videos and papers provided different benefits to users. Specifically, we found evidence that talk videos have the potential of complementing paper reading so that the reader can quickly get an overview but also selectively dive deeper into details. However, the interaction cost of fluidly consuming the two formats together can be prohibitively high, which led to a set of design goals that drove the development of Papeo. Thus, we conducted a within-subjects study to investigate whether Papeos can help readers to both acquire a comprehensive understanding of the paper and efficiently identify relevant details. We compared three conditions: Papeos with linked papers and videos, separated papers and talk videos, and papers only. With each condition, participants were asked to read the systems section of an assigned paper and to write a summary for the section that was *comprehensive* and *detailed.* Through this task, we investigated the following research questions:

- RQ1. Can Papeos reduce the cognitive load involved in reading and understanding research papers?
- RQ2. How do Papeos affect researchers' navigation of research papers and talk videos?
- RQ3. Can Papeos help researchers to both comprehensively cover significant aspects of papers and read these in detail?

5.1 Study Design

5.1.1 Participants. We recruited 16 early-stage researchers in HCI for the study through the authors' social media (Twitter) and snowball sampling. 12 of the participants were first to third year doctoral students, and 4 were Master's students. Our study focused on early-stage researchers as they may receive the greatest benefit from augmenting paper reading with talk videos—e.g., simplify and visually represent complex explanations, highlight important aspects of a paper. All participants reported reading research papers at least once a week to several times a day. The study lasted a total of 90 minutes, and participants were compensated 45 USD for their time. The study was approved by our internal review broad.

5.1.2 Conditions. During the study, participants read and wrote summaries for three different papers. For each paper, they used one of the following conditions: Papeo, Paper+Video, and only Paper. The ordering of the conditions was counterbalanced to mitigate the influence of ordering effects. In the Papeo condition, the participants used the Papeo reader. In the Paper+Video condition, participants used a basic PDF reader and a basic video player in separate tabs or windows, and, in the Paper condition, they only used the PDF reader. The basic PDF reader and video player were developed using the same base libraries and packages as the Papeo reader, and provided all basic functionalities available in other similar readers and players (e.g., zoom in, zoom out, playback speed controls).

5.1.3 Reading Materials. All of the participants read the same three papers [12, 33, 46] in the same order. We chose the papers from the initial dataset of linked papers and video used to evaluate the automatic suggestions (§4.2.2). Specifically, we chose HCI papers that presented systems that incorporated AI or algorithmic pipelines, and whose "System" sections were of relatively similar length. We focused on systems papers as they present interfaces that may be

[3] Based on the 4-fold cross-validation and grid-search, the transition probability was set to 0.7, 0.5, 0.6, and 0.6 in each fold, respectively.

[4] In contrast, the five researchers recruited to create the Papeo test set, who used the authoring tool without suggestions, took an average of 44 minutes (SD=10.9) to author one Papeo.

[5] https://github.com/allenai/pdf-component-library

[6] https://github.com/cookpete/react-player

[7] https://huggingface.co/docs/transformers/index

[8] https://scenedetect.com/en/latest/

[9] https://docs.opencv.org/4.x/index.html

easier to understand with videos. Additionally, as our goal was to evaluate whether Papeos can help readers identify details, we narrowed down to systems that incorporated pipelines as they may include a substantial amount of design and implementation details. To match these criteria, we chose two papers written by authors of this paper. In Appendix A, we provide a quantitative analysis of these Papeos to illustrate how they did not differ significantly from those authored by other researchers.

5.1.4 Procedure. The study was conducted through a popular video conferencing software. After a brief introduction to the overall study, participants performed the task for each paper in order. For each paper, participants were first provided with a short tutorial to the interface(s) that they would be using and, using a example paper and video, were allowed to use and test the interfaces for 5 minutes. Then, participants proceeded to the assigned paper and were instructed to first fully read the paper's abstract. After they read the abstract, participants were given 15 minutes to read the systems section of the paper and simultaneously write a summary that maximized the following criteria:

- *Comprehensive*: how well the summary provides an overview of the entire section.
- *Detailed*: how many specific details on the interactions and underlying models are included in the summary.
- *Coherent*: how well the summary flows or, in other words, how well the sentences connect logically. (This criteria was included to prevent summaries that simply listed details.)

To focus on capturing what they learned during the sessions, participants were informed that they could write a maximum of 14 sentences, were not allowed to copy-paste, and that the quality of their writing (e.g., spelling, grammar) would not be evaluated. Once the given time passed, participants were asked to complete a survey about the task and, then, proceeded to the next task. After all the tasks, we conducted a semi-structured interview about participants overall experience.

5.1.5 Measures. To evaluate the summaries, we developed a rubric for each paper where we listed all of the details contained in the system section of the paper, and we grouped these details according to the aspect of the system that they described (e.g., feature, pipeline component). Then, for each summary, we annotated whether the summary presents each of these details and rated its coherency on a 7-point Likert scale. To measure detail, we counted the number of details included in the summary, and, to measure comprehensiveness, we calculated the proportion of system aspects that were covered by the included details. Two of the authors who did not observe the studies performed the annotations while being blind to the conditions that generated the summaries. To verify reliability, the two authors first independently annotated the summaries for one paper, compared annotations and discussed to reach a consensus on the annotation process, and then independently annotated the paper again. This resulted on a Cohen's kappa of 0.712 for annotating the details and Krippendorff's alpha of 0.744 for coherency ratings. As the agreement was substantial, each of the authors was assigned with one of the remaining papers, and they independently annotated the summaries for that paper.

Additionally, we collected participants ratings, on a 7-point Likert scale, to the following five questions from the survey: *"I found it easy to write the summary"*, *"I found it easy to orient myself (i.e., know where information is) in the paper/video"*, and *"I found it easy to navigate to different parts of the paper/video"*. The survey also contained five questions from the NASA-TLX questionnaire [26] to measure perceived workload—excluding the question on physical demand. Finally, we analyzed interaction logs to measure how frequently participants (1) switched between the formats, (2) scrolled in the paper, and (3) scrubbed in the video. For switches, we counted every instance where the user interacted with one format after interacting with the other format, for scrolling and scrubbing, all consecutive actions within one second and in the same direction were counted as one action.

5.2 Results

Our results revealed that Papeos helped reduce participants' mental load during reading, facilitated and promoted navigation of both the paper and video, and led to more comprehensive summaries. For the statistic analysis of each measure, we first conducted a Shapiro-Wilk test to determine if the data was parametric or non-parametric. Then, when comparing between all three conditions, we used a one-way, repeated measures ANOVA when parametric and a Friedman test when non-parametric When comparing between the Paper+Video and Papeo conditions, we used a paired t-test when parametric and Wilcoxon signed-rank test when non-parametric.

5.2.1 ***Enhance Understanding and Decrease Mental Load.*** As seen in Figure 7, participants perceived the reading and summarizing task to be easiest with Papeos. The ANOVA analysis showed a significant effect of the condition on participants' perceived ease (Q=6.982, p=0.030) and a gradual increase between conditions, with the task perceived to be easiest in the Papeo condition. This indicates that talk videos could facilitate the task for participants, but the support was not perceived as significant until they were integrated into the reading experience in Papeos. This is also reflected

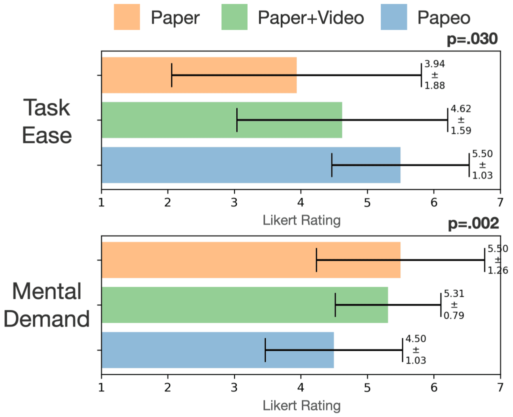


Figure 7: Perceived ease and mental demand were significantly affected by the condition used by participants. With Papeos, task ease was perceived to be highest and mental demand the lowest.

Condition	Mental	Temp.	Effort	Perf.	Frus.
Paper	5.50 (1.27)	5.13 (1.63)	5.25 (1.48)	4.44 (1.46)	4.00 (1.79)
Paper + Video	5.31 (0.79)	5.19 (1.38)	5.25 (0.93)	4.63 (1.15)	3.88 (1.15)
Papeo	**4.50** (1.03)	**4.38** (1.63)	**4.63** (1.31)	**4.94** (1.39)	**3.63** (1.59)
p-value	**0.002**	0.230	0.249	0.355	0.715

Table 4: Mean and standard deviation (in parentheses) of NASA-TLX questionnaire responses on mental demand, temporal demand, effort, performance, and frustration. (n=48, p-value based on Friedman tests.)

by responses to the NASA-TLX questionnaire as there was significant effect of the conditions on mental demand (Q=12.182, p=0.002) and demand was perceived to be lowest when participants used Papeos. Furthermore, although these results were not significant, perceived temporal demand, effort and frustration were lowest and perceived performance was highest with Papeos (Table 4).

According to participants' comments, these results could be attributed to the various ways (i.e., summaries, modalities, alternative explanations) in which talk videos supported understanding and how Papeos made these benefits available on demand. For example, P14 mentioned how Papeos summarized dense technical details but granted access to these details if needed: *"The video is high-level summary. It was easier to understand and, if I need to understand technical details, I can look the highlighted section."* Additionally, P12 mentioned how Papeos allowed them to combine and consume multiple modalities simultaneously: *"Absolutely [preferred Papeos] because I was visualizing and hearing the voice and reading the text. It was like three senses were active."* Finally, beyond helping them understand, P8 described how Papeos allowed them to check their understanding by listening to alternative explanations: *"English is not my first language so sometimes I will have a concern whether I understand the author's intention correctly. But, with the video, usually they will discuss their research in more informal way."*

5.2.2 *One Format as a Guide for the Other.*

As Papeos linked papers and videos, participants were able to use one format to guide their exploration of the other (Fig. 8). Specifically, we observed that the condition had a significant effect on participants' perceived navigation ease within the paper (Q=6.704, p=0.035), where participants perceived it to be easiest with Papeos and similar in the Paper and Paper+Video conditions. According to participants, the links between the paper and video in Papeos allowed them to navigate at a more fine-grained level than it was possible through the typical features of a paper. P11 mentioned, *"It breaks down the structure of the paper even more than the subsection headings. It also allows me to easily look for further details in the paper."* Additionally, P16 described how they were able to *"move through the paper seamlessly"* by navigating according to the video notes through the autoplay feature.

In the opposite direction, participants perceived that it was significantly easier to orient themselves within the video in the Papeo condition when compared to the Paper+Video condition (W=27.000, p=0.034). This signifies that it was easier for participants to know and remember where specific information was found within the talk video when using Papeos. P2 described that, with Papeos, it was *"clear which part of video [was] linked to"* to a specific passage of the paper, making it easier for them to find information they needed from the video. Through the localized video notes, participants could immediately access video segments that they needed when they needed them—without searching for them through the video. Thus, in Papeos, the video supported navigation in the paper, and the paper supported orientation in the video.

5.2.3 *Explore Easily, Engage More.*

Our analysis of the interaction logs (Fig. 8) revealed that participants engaged more with both formats when using Papeos. Participants switched between formats significantly more frequently in the Papeo condition compared to the Paper+Video condition (W=0.000, p<0.000). During the study, we observed that, in the Paper+Video condition, most participants watched the whole video first and then focused only on the paper during the remaining duration of the task. However, in the Papeo condition, participants continuously switched back-and-forth between the formats. Our analysis also revealed that the condition had a significant effect on how much participants scrolled in the

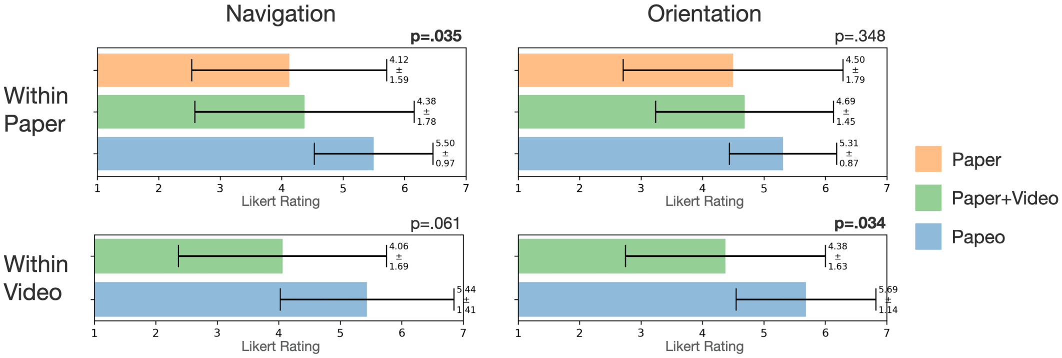


Figure 8: Results for perceived ease of navigation and orientation within the paper and the video. The condition had a significant effect on navigation within the paper and orientation within the video, with both perceived to be easiest with Papeos.

paper (F=7.065, p=0.003) with participants scrolling to a similar degree in the Papeo and Paper conditions, and scrolling less in the Paper+Video condition. Additionally, participants scrubbed in the video to a similar degree in both the Papeo and Paper+Video conditions (t=-1.810, p=0.090). Considering how participants considered that it was easier to navigate in the paper and orient oneself in the video with Papeos, these results suggest that Papeos encouraged participants to engage with both formats, and to seek for and leverage their content during the task.

5.2.4 *More Comprehensive Coverage*.

The analysis of participants' summaries (Fig. 10) revealed that there was significant effect of the condition on the comprehensiveness of participants' summaries (F=3.497, p=0.043). Summaries in the Papeo condition were rated to be the most comprehensive while those in the Paper and Paper+Video condition were relatively similar. A plausible reason for this result is that, as Papeos facilitated exploration of the content, participants were able to delve into details throughout the section and were thus able to include these in their summaries. In terms of the other measures, there was no observed effect of the condition on the detail (Q=2.000, p=0.368) or coherency (Q=2.772, p=0.250) of participants' summaries. This demonstrates that, despite participants interacting with both formats more with Papeos and writing more comprehensive summaries, this was not at the expense of other qualities in participants' summaries.

6 FIELD DEPLOYMENT

To further investigate how researchers would engage with Papeos in the wild, we deployed this new format during CSCW 2022. During the duration of the conference, we promoted our interface through social media channels and a daily newsletter sent to conference attendees. Through a portal website, conference attendees could access our reading interface and consume Papeos for specific papers that were being presented during the conference. To prepopulate this set of Papeos, we contacted several authors that were presenting in the conference and asked if they would like to use our authoring interface to create Papeos to promote their papers. Through this, we collected a set of 12 Papeos (or around six hours of volunteered authoring). The portal website also provided tutorials for using and creating Papeos and described what data is collected by the interfaces.

During the two weeks of the conference, our reading interface was visited by 288 unique users and, on average, each user visited a total of 1.20 different Papeos (min=1, max=5). To analyze the interaction logs, we identified user sessions (i.e., sequence of actions between entering and leaving the interface) and removed anomalous sessions (e.g., user left the interface immediately after entering, or user entered the interface but only interacted with it hours later). We observed that readers were actively engaged with Papeos. The average number of actions per session (e.g., scroll, play video, scrub) was 32.02 (min=2, max=255) and the average session length was 5.74 minutes (min=1.02, max=40.21). In addition to these statistics, various researchers expressed positive comments about Papeos on social media. One researcher expressed how Papeos were *"easily scannable and digestible"*, which reflected our study findings, and another researcher noted how the format can *"humanize"* papers by letting the reader *"hear the author's voice saying words that are often part of the fabric of the paper."* Beyond these benefits, a researcher noted how Papeos can *"do more than just replicate the print experience"* and *"help so all the effort we put into presentation videos doesn't get completely buried after a conference"*. In sum, through a field deployment, researchers found value in Papeo for real-world use cases, and wider adoption may require further lowering the cost of authoring Papeos.

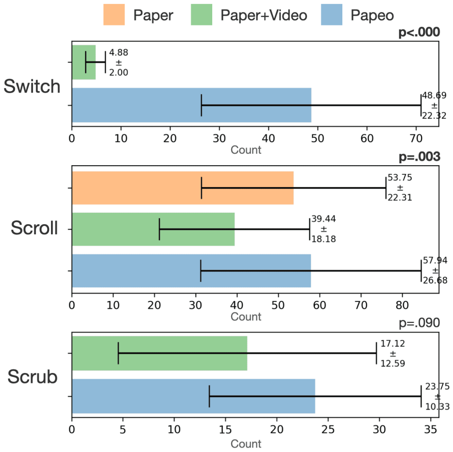


Figure 9: Analysis of the frequency of switching between formats, scrolling in the paper, and scrubbing in the video showed that the condition had a significant effect on switching and scrolling.

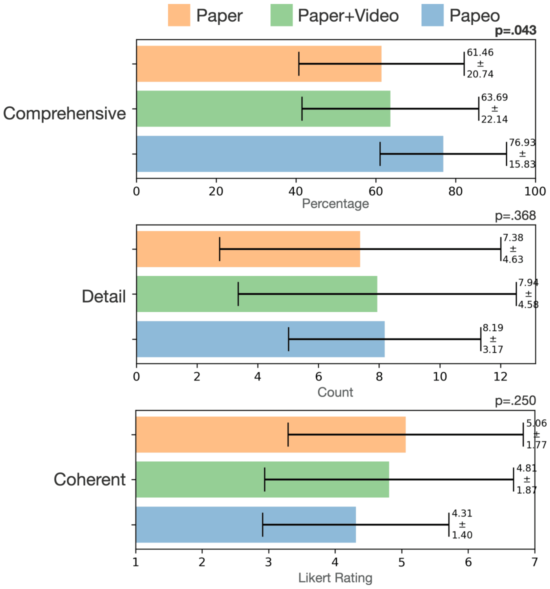


Figure 10: Results for the evaluation of participants' summaries according to comprehensiveness, detail, and coherency. The condition had a significant effect on comprehensiveness of the summaries, with summaries evaluated to be the most comprehensive in the Papeo condition.

7 DISCUSSION

In this paper, we propose Papeos, a novel reading experience that augments research papers with localized segments from talk videos to support skimming, navigation, and comprehension. While designers and researchers have taken various steps to reach the vision of *dynamic reading*, as discussed by Victor [78, 79], the experiences they proposed required a prohibitive amount of effort to realize (e.g., authoring animations or demos [25, 53]). In fact, *Distill*, a peer-reviewed journal that published interactive articles, cited authoring effort as a reason for their discontinuation [76]. In our work, we instead recognize that researchers have already dedicated significant effort in authoring talk videos that may already possess features that can enhance academic papers, like progressive animations and demo walkthroughs. Our Papeo experience leverages these talk videos to, with relatively minimal additional effort, augment the experience of reading academic papers—taking a step towards the vision of *dynamic reading*.

To extend on this vision, we identify various directions for enhancing and expanding on Papeos: automating the creation of Papeos to expand their availability, extending to other types of videos or content (i.e., blog posts), and leveraging paper-video links to generate talk videos from papers. In this section, we elaborate on the potential of Papeos and on these directions for future work.

7.1 Papeos Everywhere

Through our user study, we identified that Papeos can support understanding and navigation of papers—lowering various barriers of research consumption. Although they can aid early-stage researchers to access a larger body of knowledge, the coverage of papers that are supported with Papeos is limited by the paper authors' willingness to create Papeos. In our work, we focused on providing authors control over how their Papeos are created due to their concerns regarding automation errors. While this decision respects their preferences as authors, researchers as readers may desire a fully automatic approach as, despite possible errors, this enables them to leverage talk videos in more papers—a conflicting sentiment shared by various participants in the formative study. To increase the coverage of Papeos, future work could further develop the AI-based pipeline used for suggestions in the authoring interface. Specifically, the talk video segmentation algorithm can be enhanced by combining both visual and textual features. Additionally, while our work used general-purpose, state-of-the-art text embedding models, a small-scale dataset of paper-video links could be collected to fine-tune a sentence transformer [67] for this specific setting. However, as an improved pipeline may still present errors, the reading interface should be enhanced to provide users with error-recovery mechanisms—e.g., present multiple passages that could link to a video segment and allow the user to override erroneous links. With these improvements, future work can widen the availability of Papeos and lower the floor for early-stage researchers.

7.2 Beyond Talks and Videos

While our work focused on augmenting papers with talk videos, researchers employ an assortment of varying formats to communicate their research, and these could also be adopted to augment papers. For videos, there are various formats that exist aside from recordings of conference talks: video figures, demo videos, recordings of invited talks or thesis defenses, and, more recently, paper "explainers" on platforms like YouTube.[10] These video formats may differ from talk videos and can therefore provide different benefits when employed in Papeos. For example, demo videos can present systems and their features in more detail, invited talks or thesis defenses can contextualize a paper within a extended thread of work, and "explainer" videos can simplify the content further as their target audience can include non-researchers. Instead of depending on existing videos, authors could also create custom video clips to augment their papers with Papeos in different forms. For example, while talk videos are constrained in length and were thus useful to summarize and skim the paper, custom video clips would not be constrained and may allow authors to augment their papers with extensive, additional commentary or comprehensive walkthroughs of interfaces.

Aside from the visual aspect of videos, study participants and users from the deployment noted the significance of incorporating audio into the papers: enabling consumption with various modalities and "humanizing" papers. Future iterations of Papeo could support authors in creating additional audio-based notes to weave their voices into their papers. As an additional benefit, these audio notes could supplement screen readers and help increase the accessibility of papers by providing authors with a lightweight mechanism for creating alternative descriptions. Beyond videos, researchers frequently promote their research through other channels such as blog posts and social media (e.g., Twitter threads), and Papeos could integrate content from these formats as text-based notes. As research is increasingly distributed through a greater number of formats, Papeos can serve as a first step to connect these forms into one cohesive experience.

7.3 Generating Videos for Papers

As talk videos only cover a subset of the paper, Papeos can surface the important passages of the paper but, due to the same reason, they cannot provide video notes for the other passages. In our user study, several participants expressed how they could struggle to understand a passage, but were disappointed to not find any video notes to assist them. To remedy this limitation, future work could extend on existing work on document-to-video generation [15] to automatically generate video segments from paper passages. Specifically, with passages as input, a pipeline could generate summaries for the video's transcript [49] and slides for the frames [23, 80]. Then, the pipeline could produce video segments by combining these and incorporating audio with text-to-speech models—or even add an artificial talking head [14]. To train and tune the AI models involved in this generative pipeline, future work could use our authoring interface to collect a larger dataset of paper passage and video segment pairs. By presenting these generated video segments when requested by the reader, future Papeos can more comprehensively support the paper reading experience.

[10] Example channels include *Two Minute Papers* and *AI Coffee Break.*

7.4 Limitations

Our studies revealed various benefits of Papeos that we believe can be generalize beyond the set of papers we have tested. At the same time, we acknowledge several factors could effect the usefulness of Papeos:

- Type of work: Formative study participants noted that videos were more useful for work involving interactive and/or dynamic artifacts (e.g., HCI systems).
- Paper sections covered: User study participants expressed how Papeos were especially helpful for summarizing information dense sections.
- Visuals: Formative and user study participants noted that supplemental visuals in videos, especially those animated or presented gradually, were effective illustrating information in the paper.
- Communication style: Formative and user study participants appreciated videos that communicated paper content in a different style (e.g., informal language).

Future work should investigate the effectiveness of our approach according to these factors. Additionally, to fit the user study within 90 minutes, our user study focused on HCI papers with system contributions and only investigated the benefits of Papeos when reading one section in the paper. However, we argue that our various studies together demonstrated benefits of our approach that can generalize across papers, types of work, and domains: highlights, summaries, and audio narrations. For example, even for a qualitative paper, our approach can highlight important paper fragments (e.g., author selected themes and quotes), and provide the authors' audio narrations and summaries. Future work can conduct additional studies to investigate the significance of these benefits with papers of diverse domains and contributions.

8 CONCLUSION

This paper presents Papeos, a novel reading experience that integrates segments from talk videos as localized margin notes in academic papers. To facilitate the creation of Papeos, we introduce an authoring interface that aids paper authors in linking video segment and paper passages through algorithmic and AI-based suggestions. Through a within-subjects user study (n=16), we found that Papeos could enhance understanding of papers by providing summaries of complex passages and allowing readers to consume information in multiple modalities. With Papeos, participants leveraged each format (i.e., paper and video) to guide their navigation in the other format, which in turn facilitated navigation in both formats and encouraged more comprehensive reading of the paper. These findings and responses from researchers in a field deployment suggest the potential for leveraging existing, alternative forms of research communication to augment research papers and enable more dynamic reading experiences.

ACKNOWLEDGMENTS

The authors would like to thank Doug Downey, Shannon Zejiang Shen, Evie Cheng, and Juho Kim for their insightful discussions and feedback. We also thank the anonymous reviewers for their constructive feedback. Finally, we would like to thank the various researchers that participated in our various studies and the deployment of our system.

REFERENCES

[1] CHI 2023. 2023. The CHI2023 approach to a hybrid conference format. Retrieved February 16, 2023 from https://chi2023.acm.org/2023/01/27/the-chi2023-approach-to-a-hybrid-conference-format/

[2] Shaaron Elizabeth Ainsworth. 2006. DeFT: A Conceptual Framework for Considering Learning with Multiple Representations. *Learning and Instruction* 16 (2006), 183–198.

[3] Amazon. 2014. What is Word Wise? - Amazon Customer Service. Retrieved March 24, 2023 from https://amazon.com/gp/help/customer/display.html?nodeId=201645250

[4] Arnon Amir, Gal Ashour, and Savitha Srinivasan. 2001. Towards automatic real time preparation of on-line video proceedings for conference talks and presentations. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. IEEE, 8–pp.

[5] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A Hearst, Andrew Head, and Kyle Lo. 2022. Paper plain: Making medical research papers approachable to healthcare consumers with natural language processing. *arXiv preprint arXiv:2203.00130* (2022).

[6] Paul Ayres and Gabriele Cierniak. 2012. Split-attention effect. *Encyclopedia of the Sciences of Learning* (2012), 3172–3175.

[7] Sriram Karthik Badam, Zhicheng Liu, and Niklas Elmqvist. 2018. Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 661–671.

[8] Charles Bazerman. 1985. Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written communication* 2, 1 (1985), 3–23.

[9] Shirley Carter-Thomas and Elizabeth Rowley-Jolivet. 2003. Analysing the scientific conference presentation (CP), A methodological overview of a multimodal genre. *ASp. la revue du GERAS* 39-40 (2003), 59–72.

[10] Brandon Castellano. 2022. Scene Detection Algorithms - PySceneDetect. Retrieved March 16, 2023 from https://pyscenedetect.readthedocs.io/en/latest/reference/detection-methods/

[11] Bay-Wei Chang, Jock D Mackinlay, Polle T Zellweger, and Takeo Igarashi. 1998. A negotiation architecture for fluid documents. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*. 123–132.

[12] Joseph Chee Chang, Yongsung Kim, Victor Miller, Michael Xieyang Liu, Brad A Myers, and Aniket Kittur. 2021. Tabs.Do: Task-Centric Browser Tab Management. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 663–676. https://doi.org/10.1145/3472749.3474777

[13] Joseph Chee Chang, Amy X Zhang, Jonathan Bragg, Andrew Head, Kyle Lo, Doug Downey, and Daniel S Weld. 2023. CiteSee: Augmenting Citations in Scientific Papers with Persistent and Personalized Historical Context. *arXiv preprint arXiv:2302.07302* (2023).

[14] Peggy Chi, Tao Dong, Christian Frueh, Brian Colonna, Vivek Kwatra, and Irfan Essa. 2022. Synthesis-Assisted Video Prototyping From a Document. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–10.

[15] Peggy Chi, Nathan Frey, Katrina Panovich, and Irfan Essa. 2021. Automatic Instructional Video Creation from a Markdown-Formatted Tutorial. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 677–690.

[16] Parmit K Chilana, Amy J Ko, and Jacob O Wobbrock. 2012. LemonAid: selection-based crowdsourced contextual help for web applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1549–1558.

[17] Dorothy M Chun and Jan L Plass. 1996. Facilitating reading comprehension with multimedia. *System* 24, 4 (1996), 503–519.

[18] Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S Weld. 2020. Specter: Document-level representation learning using citation-informed transformers. *arXiv preprint arXiv:2004.07180* (2020).

[19] Jeff Conklin. 1987. Hypertext: An introduction and survey. *computer* 20, 09 (1987), 17–41.

[20] Thi Ngoc Yen Dang. 2022. A corpus-based study of vocabulary in conference presentations. *Journal of English for Academic Purposes* 59 (2022), 101144.

[21] Raymond Fok, Andrew Head, Jonathan Bragg, Kyle Lo, Marti A Hearst, and Daniel S Weld. 2022. Scim: Intelligent Faceted Highlights for Interactive, Multi-Pass Skimming of Scientific Papers. *arXiv preprint arXiv:2205.04561* (2022).

[22] G David Forney. 1973. The viterbi algorithm. *Proc. IEEE* 61, 3 (1973), 268–278.

[23] Tsu-Jui Fu, William Yang Wang, Daniel McDuff, and Yale Song. 2022. Doc2ppt: Automatic presentation slides generation from scientific documents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 634–642.

[24] Jamey Graham. 1999. The reader's helper: a personalized document reading environment. In *Proceedings of the SIGCHI conference on human factors in computing*

systems. 481–488.

[25] Tovi Grossman, Fanny Chevalier, and Rubaiat Habib Kazi. 2015. Your paper is dead! bringing life to research articles with animated figures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems.* 461–475.

[26] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology.* Vol. 52. Elsevier, 139–183.

[27] Andrew Head, Codanda Appachu, Marti A Hearst, and Björn Hartmann. 2015. Tutorons: Generating context-relevant, on-demand explanations and demonstrations of online code. In *2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC).* IEEE, 3–12.

[28] Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S Weld, and Marti A Hearst. 2021. Augmenting scientific papers with just-in-time, position-sensitive definitions of terms and symbols. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–18.

[29] James Higham and Xavier Font. 2020. Decarbonising academia: Confronting our climate hypocrisy. *Journal of Sustainable Tourism* 28, 1 (2020), 1–9.

[30] William C Hill, James D Hollan, Dave Wroblewski, and Tim McCandless. 1992. Edit wear and read wear. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 3–9.

[31] Terje Hillesund. 2010. Digital reading spaces: How expert readers handle books, the Web and electronic paper. (2010).

[32] Wen-Jung Hsin and John Cigas. 2013. Short videos improve student learning in online education. *Journal of Computing Sciences in Colleges* 28, 5 (2013), 253–259.

[33] Mina Huh, YunJung Lee, Dasom Choi, Haesoo Kim, Uran Oh, and Juho Kim. 2022. Cocomix: Utilizing Comments to Improve Non-Visual Webtoon Accessibility. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22).* Association for Computing Machinery, New York, NY, USA, Article 607, 18 pages. https://doi.org/10.1145/3491102.3502081

[34] Jessica Hullman, Yea-Seul Kim, Francis Nguyen, Lauren Speers, and Maneesh Agrawala. 2018. Improving comprehension of measurements using concrete re-expression strategies. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–12.

[35] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology Probes: Inspiring Design for and with Families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03).* Association for Computing Machinery, New York, NY, USA, 17–24. https://doi.org/10.1145/642611.642616

[36] Hyeonsu Kang, Joseph Chee Chang, Yongsung Kim, and Aniket Kittur. 2022. Threddy: An Interactive System for Personalized Thread-based Exploration and Organization of Scientific Literature. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology.* 1–15.

[37] Hyeonsu Kang, Tongshuang Wu, Joseph Chee Chang, and Aniket Kittur. 2023. Synergi: A Mixed-Initiative System for Scholarly Synthesis and Sensemaking. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology.*

[38] Myunghwa Kang and Ulrike Gretzel. 2012. Effects of podcast tours on tourist experiences in a national park. *Tourism Management* 33, 2 (2012), 440–455.

[39] Robin H Kay. 2012. Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior* 28, 3 (2012), 820–831.

[40] Caitlin Kelleher and Randy Pausch. 2005. Stencils-based tutorials: design and evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 541–550.

[41] Kandarp Khandwala and Philip J Guo. 2018. Codemotion: expanding the design space of learner interactions with computer programming tutorial videos. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale.* 1–10.

[42] Dae Hyun Kim, Enamul Hoque, Juho Kim, and Maneesh Agrawala. 2018. Facilitating document reading by linking text and tables. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology.* 423–434.

[43] Juho Kim, Philip J Guo, Carrie J Cai, Shang-Wen Li, Krzysztof Z Gajos, and Robert C Miller. 2014. Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th annual ACM symposium on User interface software and technology.* 563–572.

[44] Juho Kim, Phu Tran Nguyen, Sarah Weir, Philip J Guo, Robert C Miller, and Krzysztof Z Gajos. 2014. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the SIGCHI conference on human factors in computing systems.* 4017–4026.

[45] Juho Kim, Amy X Zhang, Jihee Kim, Robert C Miller, and Krzysztof Z Gajos. 2014. Content-aware kinetic scrolling for supporting web page navigation. In *Proceedings of the 27th annual ACM symposium on User interface software and technology.* 123–127.

[46] Tae Soo Kim, DaEun Choi, Yoonseo Choi, and Juho Kim. 2022. Stylette: Styling the Web with Natural Language. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22).* Association for Computing Machinery, New York, NY, USA, Article 5, 17 pages. https://doi.org/10.1145/3491102.3501931

[47] Rebecca Paige Krosnick. 2015. *Videodoc: Combining videos and lecture notes for a better learning experience.* Ph. D. Dissertation. Massachusetts Institute of Technology.

[48] Byungjoo Lee, Olli Savisaari, and Antti Oulasvirta. 2016. Spotlights: Attention-optimized highlights for skim reading. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* 5203–5214.

[49] Guy Lev, Michal Shmueli-Scheuer, Jonathan Herzig, Achiya Jerbi, and David Konopnicki. 2019. Talksumm: A dataset and scalable annotation method for scientific paper summarization based on conference talks. *arXiv preprint arXiv:1906.01351* (2019).

[50] Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text summarization branches out.* 74–81.

[51] Ching Liu, Juho Kim, and Hao-Chuan Wang. 2018. ConceptScape: Collaborative concept mapping for video learning. In *Proceedings of the 2018 CHI conference on human factors in computing systems.* 1–12.

[52] Kyle Lo, Joseph Chee Chang, Andrew Head, Jonathan Bragg, Amy X. Zhang, Cassidy Trier, Chloe Anastasiades, Tal August, Russell Authur, Danielle Bragg, Erin Bransom, Isabel Cachola, Stefan Candra, Yoganand Chandrasekhar, Yen-Sung Chen, Evie Yu-Yen Cheng, Yvonne Chou, Doug Downey, Rob Evans, Raymond Fok, Fangzhou Hu, Regan Huff, Dongyeop Kang, Tae Soo Kim, Rodney Kinney, Aniket Kittur, Hyeonsu Kang, Egor Klevak, Bailey Kuehl, Michael Langan, Matt Latzke, Jaron Lochner, Kelsey MacMillan, Eric Marsh, Tyler Murray, Aakanksha Naik, Ngoc-Uyen Nguyen, Srishti Palani, Soya Park, Caroline Paulic, Napol Rachatasumrit, Smita Rao, Paul Sayre, Zejiang Shen, Pao Siangliulue, Luca Soldaini, Huy Tran, Madeleine van Zuylen, Lucy Lu Wang, Christopher Wilhelm, Caroline Wu, Jiangjiang Yang, Angele Zamarron, Marti A. Hearst, and Daniel S. Weld. 2023. The Semantic Reader Project: Augmenting Scholarly Documents through AI-Powered Interactive Reading Interfaces. arXiv:2303.14334 [cs.HC]

[53] Damien Masson, Sylvain Malacria, Edward Lank, and Géry Casiez. 2020. Chameleon: Bringing Interactivity to Static Digital Documents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–13.

[54] Martha J Maxwell. 1972. Skimming and scanning improvement: The needs, assumptions and knowledge base. *Journal of Reading Behavior* 5, 1 (1972), 47–59.

[55] Richard E Mayer and Richard B Anderson. 1991. Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of educational psychology* 83, 4 (1991), 484.

[56] Richard E Mayer and Richard B Anderson. 1992. The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of educational Psychology* 84, 4 (1992), 444.

[57] Richard E. Mayer, Sherry Fennell, Lindsay Farmer, and Julie Campbell. 2004. A Personalization Effect in Multimedia Learning: Students Learn Better When Words Are in Conversational Style Rather Than Formal Style. *Journal of Educational Psychology* 96 (2004), 389–395.

[58] Richard E Mayer and Roxana Moreno. 1998. A cognitive theory of multimedia learning: Implications for design principles. *Journal of educational psychology* 91, 2 (1998), 358–368.

[59] Richard E Mayer and Roxana Moreno. 1998. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of educational psychology* 90, 2 (1998), 312.

[60] MediaWiki. 2023. Page Previews - MediaWiki. Retrieved March 24, 2023 from https://www.mediawiki.org/wiki/Page_Previews

[61] José Otero, Arthur C Graesser, et al. 2014. *The psychology of science text comprehension.* Routledge.

[62] Yasuhiro Ozuru, Kyle Dempsey, and Danielle S. McNamara. 2009. Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction* 19, 3 (2009), 228–242. https://doi.org/10.1016/j.learninstruc.2008.04.003

[63] Soya Park, Jonathan Bragg, Michael Chang, Kevin Larson, and Danielle Bragg. 2022. Exploring Team-Sourced Hyperlinks to Address Navigation Challenges for Low-Vision Readers of Scientific Papers. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–23.

[64] Amy Pavel, Dan B Goldman, Björn Hartmann, and Maneesh Agrawala. 2015. Sceneskim: Searching and browsing movies using synchronized captions, scripts and plot summaries. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology.* 181–190.

[65] Amy Pavel, Colorado Reed, Björn Hartmann, and Maneesh Agrawala. 2014. Video digests: a browsable, skimmable format for informational lecture videos.. In *UIST*, Vol. 10. Citeseer, 2642918–2647400.

[66] Napol Rachatasumrit, Jonathan Bragg, Amy X Zhang, and Daniel S Weld. 2022. CiteRead: Integrating Localized Citation Contexts into Scientific Paper Reading. In *27th International Conference on Intelligent User Interfaces.* 707–719.

[67] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[68] Elizabeth Rowley-Jolivet. 2004. Different visions, different visuals: A social semiotic analysis of field-specific visual composition in scientific conference presentations. *Visual communication* 3, 2 (2004), 145–175.

[69] Elizabeth Rowley-Jolivet and Shirley Carter-Thomas. 2005. The rhetoric of conference presentation introductions: Context, argument and interaction. *International Journal of Applied Linguistics* 15, 1 (2005), 45–70.

[70] Bill N Schilit, Gene Golovchinsky, and Morgan N Price. 1998. Beyond paper: supporting active reading with free form digital ink annotations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 249–256.

[71] Zejiang Shen, Kyle Lo, Lucy Lu Wang, Bailey Kuehl, Daniel S Weld, and Doug Downey. 2022. VILA: Improving structured content extraction from scientific PDFs using visual layout groups. *Transactions of the Association for Computational Linguistics* 10 (2022), 376–392.

[72] Hijung Valentina Shin, Floraine Berthouzoz, Wilmot Li, and Frédo Durand. 2015. Visual transcripts: lecture notes from blackboard-style lecture videos. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–10.

[73] Brent R Stockwell, Melissa S Stockwell, Michael Cennamo, and Elise Jiang. 2015. Blended learning improves science education. *Cell* 162, 5 (2015), 933–936.

[74] Deborah Swanberg, Chiao-Fe Shu, and Ramesh C Jain. 1993. Knowledge-guided parsing in video databases. In *Storage and retrieval for Image and Video Databases*, Vol. 1908. Spie, 13–24.

[75] Craig S Tashman and W Keith Edwards. 2011. LiquidText: A flexible, multitouch environment to support active reading. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3285–3294.

[76] Editorial Team. 2021. Distill Hiatus. *Distill* (2021). https://doi.org/10.23915/distill.00031 https://distill.pub/2021/distill-hiatus.

[77] Anh Truong, Peggy Chi, David Salesin, Irfan Essa, and Maneesh Agrawala. 2021. Automatic generation of two-level hierarchical tutorials from instructional makeup videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[78] Bret Victor. 2011. Explorable Explanations. Retrieved March 23, 2023 from http://worrydream.com/ExplorableExplanations/

[79] Bret Victor. 2014. Humane representation of thought: a trail map for the 21st century. In *Proceedings of the companion publication of the 2014 ACM SIGPLAN conference on Systems, Programming, and Applications: Software for Humanity*. 5–5.

[80] Fengjie Wang, Xuye Liu, Oujing Liu, Ali Neshati, Tengfei Ma, Min Zhu, and Jian Zhao. 2023. Slide4N: Creating Presentation Slides from Computational Notebooks with Human-AI Collaboration. (2023).

[81] Lucy Lu Wang, Isabel Cachola, Jonathan Bragg, Evie Yu-Yen Cheng, Chelsea Haupt, Matt Latzke, Bailey Kuehl, Madeleine van Zuylen, Linda Wagner, and Daniel S Weld. 2021. Improving the accessibility of scientific documents: Current state, user needs, and a system solution to enhance scientific PDF accessibility for blind and low vision users. *arXiv preprint arXiv:2105.00076* (2021).

[82] Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Advances in Neural Information Processing Systems* 33 (2020), 5776–5788.

[83] Qian Yang, Gerard de Melo, Yong Cheng, and Sen Wang. 2017. HiText: Text reading with dynamic salience marking. In *Proceedings of the 26th International Conference on World Wide Web Companion*. 311–319.

[84] Dongwook Yoon, Nicholas Chen, François Guimbretière, and Abigail Sellen. 2014. RichReview: blending ink, speech, and gesture to support collaborative document review. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 481–490.

[85] Polle T Zellweger, Bay-Wei Chang, and Jock D Mackinlay. 1998. Fluid links for informed and incremental link transitions. In *Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space—structure in hypermedia systems: links, objects, time and space—structure in hypermedia systems*. 50–57.

A QUANTITATIVE ANALYSIS OF PAPEOS

We quantitatively analyzed various characteristics of all the Papeos authored throughout our work (N=23). These include 8 Papeos from the test set, 6 from the preliminary user evaluation of the authoring tool, and 9 additional ones from the deployment study. Although there was a total of 12 Papeos in the deployment, 3 of them were authored during the preliminary evaluation of the authoring tool. Table 5 shows that the chosen Papeos fall within one standard deviation for all of the characteristics, which suggests that they did not deviate significantly from those authored by other researchers.

Characteristics	**Mean**	**SD**	**[46]**	**[12]**	**[33]**
Number of Linked Paper Fragments and Video Segments	20.6	7.3	20	15	24
Average Number of Paper Fragments per Link	3.2	1.6	3.9	2.8	1.8
Average Length of Linked Video Segment	24.3	8.7	24.0	29.2	19.0
Total Number of Synchronized Highlights	2.8	4.2	7	7	3

Table 5: Analysis of various characteristics of the Papeos collected during this work. The analysis shows the overall statistics (i.e., mean and standard deviation) for all the Papeos, and the statistics for the three that were selected for the user study.