Scim: Intelligent Faceted Highlights for Interactive, Multi-Pass Skimming of Scientific Papers

Raymond Fok
rayfok@cs.washington.edu
University of Washington
Seattle, Washington, USA

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

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

Kyle Lo
kylel@allenai.org
Allen Institute for AI
Seattle, Washington, USA

Marti A. Hearst
hearst@berkeley.edu
University of Berkeley
Berkeley, California, USA

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

ABSTRACT

Researchers are expected to keep up with an immense literature, yet often find it prohibitively time-consuming to do so. This paper explores how intelligent agents can help scaffold in-situ information seeking across scientific papers. Specifically, we present Scim, an AI-augmented reading interface designed to help researchers skim papers by automatically identifying, classifying, and highlighting salient sentences, organized into rhetorical facets rooted in common information needs. Using Scim as a design probe, we explore the benefits and drawbacks of imperfect AI assistance within an augmented reading interface. We found researchers used Scim in several different ways: from reading primarily in the 'highlight browser' (side panel) to making multiple passes through the paper with different facets activated (e.g., focusing solely on OBJECTIVE and NOVELTY in their first pass). From our study, we identify six key design recommendations and avenues for future research in augmented reading interfaces.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

augmented reading interfaces, scientific papers, skimming, faceted highlights

1 INTRODUCTION

The rise of knowledge work and contemporaneous information explosion demand the ability to quickly sift through rapidly evolving information. For example, scientific researchers spend a tremendous amount of effort staying up to date with literature in their field. The process typically involves researchers foraging for a set of potentially relevant work, skimming or reading the selected papers, and organizing the relevant aspects contextually. AI has been used to assist several aspects of this process: scholarly search engines, such as Google Scholar and Semantic Scholar [2, 7], paper recommendation systems [5, 8], and in-depth reading support tools [29], but information overload remains a huge challenge for researchers.

Before deciding to read a paper in depth, researchers often skim it selectively [58]. Skimming is a reading technique that involves quickly glancing over text to gain a general idea of its content, extracting only the most important information [49]. Since skimming a paper takes a fraction of the time required for a deep read, researchers have adopted skimming as one strategy to keep pace with the growing literature, a practice intensified by a shift to online scholarly reading [43, 66].

Experienced researchers rely on years of practice skimming papers, accumulating repetitions in key techniques such as scanning for headings, selectively identifying key words that indicate areas of the text to devote more attention, and focusing on visual content. However, skimming can still be a challenging skill to learn and effectively harness [20, 49, 73]. Skimming imposes a cognitive burden on our limited working memory, requiring readers to navigate with deliberate saccades through a text to identify important information and understand the text that is fixated on. Novice researchers with less developed skimming know-how, and even experienced researchers may find themselves struggling to effectively identify and navigate between important parts of the text. Effective skimming can then be considered a rapid decision-making process under time pressure, balancing a determination of parts of the text that should be read in more depth and an assimilation of the processed content into their working understanding of the paper.

Prior research has explored ways to reduce the cognitive effort required for various aspects of reading and skimming papers [29, 42, 56]. Experimental tools have been developed to specifically help skimming readers, for instance by spotlighting visual content within a paper [39] and automatically enlarging section headers within paper thumbnails [11]. Within the natural language processing community, automated techniques for document understanding and summarization of scientific papers [63, 71] have seen a recent growth in research interest and performance, and could be possibly leveraged for developing novel tools to support skimming, but have yet to be meaningfully integrated and studied within user-facing interactive systems.

We seek to address these issues by exploring the strengths, limitations, and design implications of reading interfaces for skimming augmented by the output of these potentially imprecise AI systems. We present Scim, a prototype tool designed to support readers in skimming scientific papers (Figure 1). Scim's key features involve automatically identifying and highlighting salient sentences within a paper, classifying these sentences into four rhetorical facets that

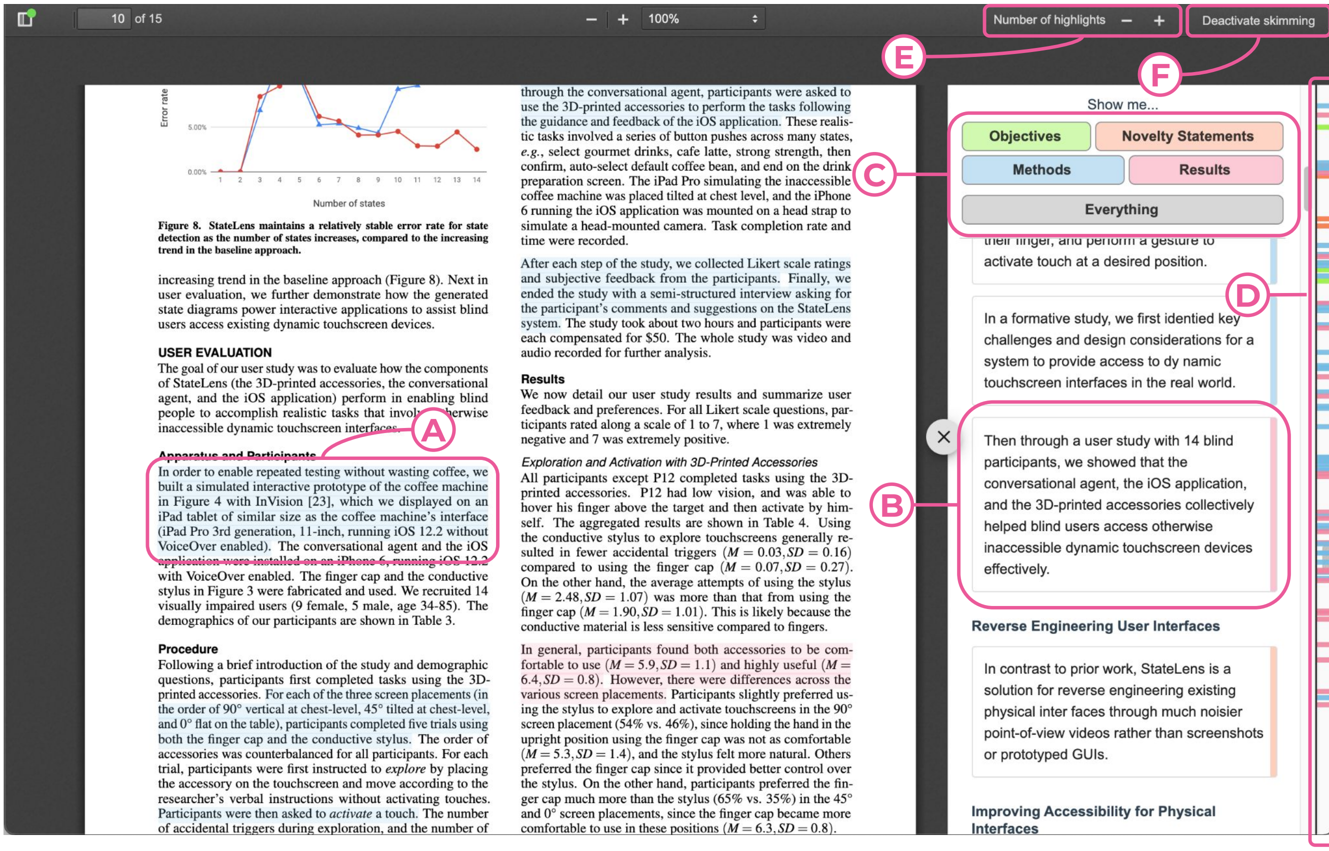

Figure 1: Scim's user interface. Scim marks salient sentences with a layer of colored highlights atop the original document (A). The highlighted sentences are organized within a *highlight browser*, and clicking on a highlight scrolls to the corresponding location in the document (B). A *facet palette* enables filtering of highlights within the document and highlight browser according to a specific rhetorical facet (C). A spatial distribution of highlights throughout the paper is visualized with *scrollbar annotations* (D). Controls allow readers to adjust the quantity of presented highlights (E) and toggle Scim on and off (F).

correspond to common information needs of readers, providing an integrated list view of these highlights organized and linked to its corresponding context in the paper, and offering readers interactive controls to support their skimming.

We used Scim as a design probe to (1) explore the opportunities for an AI-powered reading interface for supporting researchers as they skim scientific papers and (2) characterize the benefits and risks of integrating imprecise AI support into augmented reading interfaces. In a study with 13 participants, we observed researchers using Scim in various ways. As might be expected, researchers used the highlights as suggestions of importance from an opaque intelligent agent, treating them as cues of attention to complement linear skimming processes. We also saw participants making use of Scim in novel and unexpected ways, highlighting important design considerations for future AI-assisted reading interfaces. For instance, we observed readers using the features to support multiple, complementary skimming passes over a paper. Some readers tended to filter highlights to two of the rhetorical facets—Objective and Novelty—and jumped to read these highlights within the paper first, before continuing to skim the paper from the beginning. Others used a summary of highlights, provided in an adjacent highlight browser, as a rereading mechanism to efficiently verify their comprehension.

Participants largely expressed interest in using Scim's core features in future augmented reading interfaces, and were excited for the potential of AI assistance in reading interfaces, despite their potential for errors. Related to prior studies exploring mental model formation in human-AI systems [55], we also observed researchers forming a first impression for Scim's reliability in identifying important sentences by evaluating the quality of the first highlights they saw. While they made further adjustments to their impressions of reliability based on subsequent highlights, researchers felt their initial assessment played a vital role in determining their desire to continue using or abandon the system. Based on qualitative insights from our design probe, we distilled six key design recommendations for the development of AI-infused reading interfaces. Our recommendations highlight how these systems may support more interactive, need-driven reading experiences, and how their designs may consider errors and biases made by the underlying intelligent agents. Finally, we discuss future research opportunities in collaborative highlights and personalization within intelligent reading interfaces.

2 RELATED WORK

2.1 Prior Studies on Skimming

Skimming is widely considered to be a form of rapid reading in which the goal is to get a general idea of the text or visual content, typically accomplished by focusing on information relevant to one's goals and skipping over irrelevant information [48, 58]. Skimming is a particularly necessary and useful skill for scholars who read scientific papers. As the number of published papers continue to increase year over year and technology has caused a gradual move from print towards a digital medium, scholars have adapted by reading more papers while spending less time on each [43, 66].

Prior results from the psychology literature have found that skim readers are not generally very accurate at selecting goal-relevant information for processing within text, and that physical limitations in the oculomotor system responsible for controlling eye movements largely preclude rapid, accurate placements of eye gaze for extended periods such as when skimming a long document [47, 48]. Beyond limitations in visual acuity, skimming can also be a cognitively demanding task as readers are continually building an ongoing mental model of the text and integrating information across sentences as they read [57, 58, 65].

Other studies suggest that skim readers may be able to effectively direct attention to more important content, for instance by reading in a satisficing manner [20, 21, 59]. Satisficing is a skim reading strategy in which readers are inherently sensitive to a proxy for information gain. Under this strategy, readers set a information threshold, and if while reading a unit of text they determine that the information gain falls below a designated information threshold, they proceed on to the next unit of text. These studies have found that as a result, people tend to spend more time at the beginning of paragraphs, toward the top of pages, and at the beginning of documents [20]. We use Scim to study how automated assistance may support skimming by cueing readers towards salient sentences, suggested by an AI system, thereby shifting the initial locus of attention for readers under the satisficing strategy.

One study on skimming for scientific document triage found that readers were hasty and incomplete, and documents were scrolled through quickly with attention paid to highly visual content and section headers [44]. Since information-dense content may be buried within pages of plain text, we see an opportunity for automated assistance in facilitating the discovery of these relevant information units that may otherwise be skipped. Scientific documents are also laden with visual content, typographical cues (e.g., italicized, bold, or colored text), and structural information. Studies have found that readers draw on document features to support rapid comprehension via these macro- and micro-structures [13, 38, 45] and visual content [34, 73]. Scim's design as an AI-augmented reading interface enables readers to leverage AI assistance while still retaining access to a paper's visual and structural information.

2.2 Tools for Reading and Skimming

Researchers have long sought to equip readers with tools that support and augment their cognition while reading documents. The nascent days of human-computer interaction saw the introduction of augmented reading interfaces to support the reading process, including fluid documents that provided contextual access to supplemental information between lines of text [15], fluid hypertext [74], visualizations for social annotations within papers [30], and affordances for annotating papers and jumping readers to passages of interest [25, 62]. Since then, several classes of approaches have been proposed to support the various aspects of reading, such as document navigation and comprehension.

2.2.1 Modified Scrolling Interactions. One line of research sought to facilitate the rapid exploration of long documents by modifying the behavior of reading interfaces during scrolling. Applications of content-aware scrolling were used to redefine the presentation order of content within a document [33], provide pseudo-haptic feedback when scrolling past relevant information [36], and dynamically resize document headings within paper thumbnails in a document viewer [11]. Spotlights implemented an attention allocation technique that pinned headings and figures as static overlays to a document as it was continuously scrolled [39].

2.2.2 Typographical Cueing. Another approach involved augmenting reading interfaces with typographical cues (e.g., highlighting). The Semantize system used highlights to visualize sentiment within a document, and underlined words with positive or negative sentiment in different colors [70]. The ScentHighlights system used highlights to identify conceptually relevant text based on a user's query [17]. The HiText technique introduced dynamic graded highlighting of sentences within a document in accordance with their salience [72]. Modern reading interfaces also commonly support readers in marking regions of interest with a document with highlights or free-text annotations. The pervasiveness of highlighting as a technique for drawing readers' attention can be attributed to the von Restorff isolation effect, which states that an item isolated against a homogenous background will be more likely to be attended to and remembered [69]. Studies have since found evidence of this effect on the visual foraging behavior of readers on highlighted documents, finding that highlights attract about half of the total number of fixations within a document, and are often drawn to by readers' eyes [16].

2.2.3 Document Augmentations. Beyond typographical cues, other reading interface augmentations exist to specifically support the reading of scientific papers. For instance, online paper providers like ScienceDirect, PubMed, and Semantic Scholar provide readers with in-context citation information. Experimental systems have linked document text to marks within charts [37] and cells within tables [35], generated on-demand visualizations based on text within the paper [4], augmented static visualizations with animated [26] or interactive [46] overlays, and provided in-context definitions for nonce words [29]. We design Scim with inspiration from many of these prior augmented reading interfaces, augmenting scientific papers with interactive highlights that guide reader attention. Extending prior systems, Scim not only extracts salient sentences, but also classifies each highlight into common classes of information needs for readers.

2.2.4 Summarization. An alternative method to skimming a full paper is to read a shortened representation of the paper's content in the form of a summary. An author-provided summary is de facto included with each paper as an abstract, which researchers often read before continuing to the rest of the paper. Automated summarization has garnered significant interest from the natural language processing community, and extractive and abstractive methods for generating summaries from long-form documents have been developed over the years [3, 53, 63]. Some methods have even been proposed for generating extreme (single sentence) summaries, called TLDRs, from full papers [12].

However, providing only a summary to readers is often unsatisfactory. Despite recent improvements in the quality of generated summaries, they remain error-prone, susceptible to hallucination [76], and are not reliable enough to be used as a standalone replacement for reading the paper itself. Furthermore, summaries do not provide readers with the ability to interact with the full paper. For instance, as readers' goals and interests change while reading a paper, they may wish to explore certain sections in further detail. While traditional summaries cannot support this interaction, augmented reading interfaces naturally retain the context of the paper. We leverage natural language processing techniques to identify salient sentences and classify sentences into rhetorical facets using a pretrained language model, and present the output within a carefully-designed augmented reading interface to support the interactivity and context lacking in standalone summaries.

3 DESIGN MOTIVATIONS

3.1 Formative Study

To better understand how an intelligent reading interface might support readers in skimming scientific papers, we conducted a small formative study with eight researchers. In the study, researchers were asked to describe their typical approaches and goals for skimming. We found researchers were largely predisposed to established strategies—what they considered as conventional wisdom—for skimming scientific papers. Most focused on reading the Abstract and Introduction sections of a paper, before searching for a list of the paper's contributions, a summary of results, or the paper's conclusions.

Strategies for skimming beyond this conventional wisdom tended to diverge between individual researchers, influenced by factors such as their goal for skimming a particular paper and their experience in that paper's research area. Researchers mentioned using heuristics for identifying regions that were likely to provide the most information gain—typographical cues (e.g., bold or italicized text), structural cues (e.g., section headers), the presence of visual media (e.g., figures and tables), and rhetoric structure (e.g., first and last sentences of a paragraph). They also expressed a variety of goals that would require different levels of detail in skimming—these included learning about specific techniques introduced in a paper, determining a paper's relationship with prior work or their own research, discovering new research directions, or gaining a general understanding to discuss a paper with colleagues.

3.2 Design Motivations

Our formative study suggested that researchers skimming papers may have a diverse set of goals and strategies, but often start by searching for areas indicating a paper's significance before diving into further detail. We see an opportunity to leverage automated tools to accelerate the discovery of significance-defining cues within the paper. Based on our formative study and a review of related literature, we identified four design motivations that guide the design of Scim:

DM1. Scaffold information discovery throughout an entire paper. Though conventional wisdom guided readers toward a handful of common sections at the beginning and end of papers, these is little support for the discovery of relevant content within the middle, often text-heavy sections of a paper. We aimed to explore opportunities in leveraging natural language processing techniques for document-level understanding to provide support for skimming these sections of papers. By supporting content discovery and classification, we believe intelligent agents may reduce the cognitive processing readers need to perform during skimming.

DM2. Connect readers to salient content in context. One method to support skimming of long texts is through a condensed view of salient text content. These views aim to provide a representation of the text that requires less effort to process but conveys the same level of understanding. However, this condensed representation may lack sufficient context if presented statically. Our design should therefore provide readers with the ability to engage with and gain more context on-demand for any AI-suggested content within these condensed views.

DM3. Direct reader attention while minimizing distractions. Studies in cognitive psychology have found that visual cueing mechanisms can be effective in focusing reader attention [16] and improving retention of material [24, 61]. In developing Scim, we evaluated three common cueing mechanisms: underlining, highlighting, and masking. User evaluations of these mechanisms within prototypes of Scim found that highlighting was the most familiar due to its use for annotation in existing PDF viewers, underlining was too subtle to consistently attract attention, and masking required additional cognitive effort to shift one's eyes between masked and unmasked text. Overall, highlighting was recommended by users as the most effective cueing mechanism, which we incorporated into the current design of Scim.

DM4. Support error recovery. Automated approaches for identifying and classifying salient content within papers are susceptible to errors. In addition to systematic errors, these methods may also make relevance errors when integrated into user-facing systems like an augmented reading interface. For example, suggested content that is important to one reader may be deemed irrelevant by another reader under a particular context. Reflecting guidelines in developing imperfect AI-infused user applications [1], we aimed to design Scim to retain the agency of users in efficiently recovering from and dismissing fallible AI assistance.

4 THE SCIM SYSTEM

Scim provides an augmented reading interface that organizes and filters salient content within a scientific paper (Figure 2). The interface is enabled by an end-to-end document processing pipeline that localizes, classifies, and ranks important sentences in a paper.

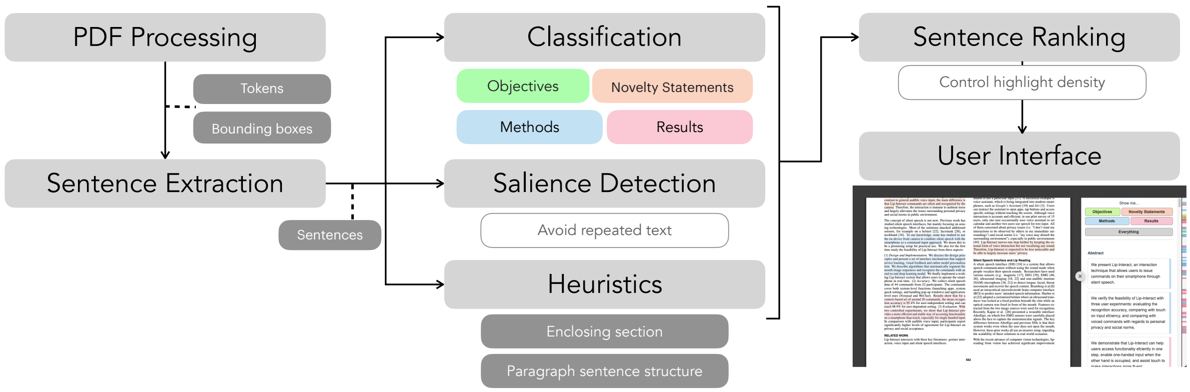

Figure 2: Overview of Scim, a reading interface leveraging faceted highlights to support skimming of scientific papers.

4.1 User Interface

Scim's user interface was developed through an iterative design process, guided by our formative study and informal usability studies of initial prototypes. We indicate whenever a feature or design choice was informed by one of the four design motivations introduced in Section 3.2.

We implemented the user interface as a web application, with features built atop the PDF rendering platform `pdf.js` [51]. Scim's user interface consists of a suite of features that augment the original text document through overlays and side panels, largely allowing readers to peruse the document in its original structure. One goal was to enable readers to effortlessly switch between standard skimming processes and guided reading through augmented interface features. Scim accomplishes this by presenting readers with the full document and incorporating guidance via faceted highlighting throughout the text. Scim also retains text markup that may aid readers in skimming, such as hyperlinks, interactive citations, bold and italicized text, and other visual cues provided by the authors. To minimize the friction in adopting Scim into readers' existing skimming processes, we used common design patterns found in current PDF viewers and integrated development environments.

4.1.1 Faceted Highlights. To support readers in identifying salient sentences within the context of a paper, Scim uses *colored highlighting* as visual cues throughout the document (Figure 1.A). The opacity of highlights ensure visual salience while remaining unobtrusive to readers attending to both the highlighted text and its surrounding context (*DM3*). In-context highlights also give readers an intuition for the structure of the paper relative to the highlighted content, and enabling them to maintain a sense of how much content is skipped when navigating between consecutive highlights.

Our formative study revealed that despite variation in goals and experience, readers had common high-level information needs while skimming. To support the information foraging process, we organized highlights into *rhetorical facets* based on the type of information they provided. Numerous schemes exist for sentence-based classification of scientific literature into rhetorical facets. Coarse-grained schemes classify sentences according to typical section names found in scientific literature [18, 31], and are composed of a small number of rhetorical facets such as Objective, Method, Result, and Conclusions. Other schemes are more fine-grained and classify based on argumentative zones and conceptual structure [40, 41, 67, 68].

We aimed to select a minimal set of rhetorical facets for Scim that most closely corresponded to the information needs of readers while skimming. We combined aspects of a coarse-grained schema for classifying scientific abstracts [18] and the NOV_ADV category (i.e., sentences presenting novelty or advantage) from Argumentative Zoning [68] to create a final taxonomy of four rhetorical facets (hereafter referred to as *facets*). Scim uses four colors to identify highlights associated with each of the four facets: Objective (green), Novelty (orange), Method (blue), and Result (red). With repeated usage, we envision these colors could become a simple association device for each facet, enabling readers to quickly scroll through a paper and identify highlights of interest.

4.1.2 Highlight Browser. Rather than scrolling through an entire paper, readers may want an efficient means to view all AI-suggested highlights. Scim provides a *highlight browser* (or browser for short) that displays an condensed list of all highlights distributed throughout the paper. The browser is implemented as side panel anchored to the right of the reading interface, and can be hidden away when not needed. Highlights in the browser are ordered by their location within the paper, and demarcated by which section of the paper they are contained within (Figure 1.B). A colored indicator is also displayed to the right of each highlight in the browser, offering a subtle cue its classified facet directly within the browser (*DM3*). Readers desiring more context for a particular highlight can click on it within the browser to scroll to its corresponding position within the paper—we refer to this interaction throughout the paper as *context linking* (*DM2*).

4.1.3 Scrollbar Marks. Inspired by edit wear affordances [30] and scrollbar maps within integrated development environments [50], Scim provides colored *scrollbar annotations* in the vertical document scrollbar (Figure 1.D). These marks intend to convey at a glance the quantity and positional distribution of highlights throughout the document (*DM3*).

4.1.4 Highlight Controls. Scim also provides several features that enable readers to tailor the highlighting experience to their own skimming needs. At the top of the browser, a *facet palette* (Figure 1.C) contains buttons that can be clicked to select a single facet, displaying only highlights of that facet throughout the document, within the browser, and in the scrollbar annotations. Highlights for all facets are enabled by default, and a single button labeled "Everything" allows readers to show all highlights again after filtering to a single facet. For more or less automated guidance while skimming, readers can increase or decrease the overall number of displayed highlights via buttons in the menu bar above the document (Figure 1.E). Finally, another button allows readers to toggle between Scim and their default document viewer (Figure 1.F) (*DM4*).

4.2 Document Processing Pipeline

4.2.1 PDF Component Extraction. We used the open-source Multimodal Document Analysis (MMDA) library [23] to process all elements within a PDF document, including textual tokens, mathematical symbols, section headers, and metadata. By default, MMDA only provides PDF bounding box detection for pages, tokens, and rows, so we segmented a paper's text into sentences and merged token and row bounding boxes to form sentence bounding boxes. For each sentence we stored its corresponding section headers, which we later use in heuristics for sentence prioritization (described in Section 4.2.4).

4.2.2 Salient Sentence Extraction. To identify sentences of general importance to a reader, we defined an initial salience score for each sentence within a document. We used the Universal Sentence Encoder [14] to retrieve an embedding v for each sentence s, and calculated the cosine similarity between embeddings for all pairs of sentences.

$$s_c = \sum_{v'} \frac{v \cdot v'}{||v||||v'||}$$

The sum represents the global similarity of each sentence to other sentences in the document. The intuition behind this salience measure is that sentences which are most similar to many other sentences in the document are likely to contain information central to the ideas of the document. We clipped negative similarity scores to 0, and normalized all scores to $[0, 1]$.

4.2.3 Facet Classification. To classify sentences into the previously described facets, we implemented the BERT-based sentence classifier with SciBERT pretrained weights introduced in [18]. We trained the model on three Nvidia Titan X GPUs, keeping the same dropout rate, optimizer, number of epochs, and learning rate noted in the original paper. To create a larger training set, we merged the three splits of the CSAbstruct dataset [18], a corpora of computer science abstracts with sentences annotated according to their rhetorical role (one of Background, Method, Objective, Result, or Other). We used this trained model to classify sentences within papers for three of the facets within our taxonomy: Objective, Method, and Result. For the remaining Novelty facet, since collecting a separate annotated dataset necessary for training was outside the scope of this work, we instead used a small set of lexical heuristics for classification. Specifically, we use a rule-based approach based on discourse and lexical constraints, similar to [28]. We classify sentences as a Novelty statement by matching against a lexical set containing words such as *inconsistent* and *however*, and then verifying the presence of nearby indicators of author intent (i.e., aliases of {we, our, this study}) and comparison to previous work (i.e., aliases of {previous, recent}). While we found these methods were sufficient for developing Scim as a design probe, future iterations of the system could investigate more sophisticated techniques for salient sentence extraction and classification.

4.2.4 Additional Heuristics. Several heuristics based on common structural patterns within scientific papers were used to refine the ranking of salient sentences. These heuristics included favoring sentences that appeared at the beginning and ends of paragraphs, and sentences that appeared within a predefined set of "expected sections" (e.g., Objective within an Introduction section and Novelty within a Related Work section). These heuristics were combined with the probability assigned to each classified sentence by the previous sentence classifier and the computed salience scores to produce a sentence prioritization schema, which was used to select which sentences to highlight within the reading interface. Pilot studies with earlier Scim prototypes found that readers generally expected AI highlights distributed throughout the document rather than concentrated only in particular sections. To distribute highlights throughout a paper, a final heuristic was included to prioritize sentences within paragraphs not containing other highlighted sentences.

4.2.5 Final Output. For each paper, the pipeline outputs a ranked list of sentence objects. Each sentence object contains the sentence's original text, its classified facet, its bounding box within the PDF, and its enclosing section headers. These sentences were saved to a single JSON file, which was then stored in a remote server and processed by the previously described user interface.

5 USABILITY STUDY

We used Scim as a technological design probe [9] to explore the impact of an augmented reading interface with faceted highlights on skimming scientific papers. We aimed to observe the realistic use of the system within two skimming tasks, and to collect feedback from participants on the current design and potential improvements for the system. We focused our observations from the user study to answer two research questions:

RQ1. How do readers use Scim and its faceted highlights for skimming papers?

RQ2. How does AI assistance influence readers while skimming papers?

5.1 Participants

We recruited 13 participants (8 male, 5 female) via purposive and snowball sampling, primarily through university mailing lists. Participants were required to have some prior experience in reading or writing scientific papers. Eleven participants were graduate students with between one and six years of experience in a computer science PhD program, one was an undergraduate computer science student, and one was a senior research programmer. Six participants were between the ages of 18-25, and seven participants were between the ages of 26-40. Nine participants described themselves as somewhat experienced or very experienced with reading papers in computer science, and one participant described themself as somewhat inexperienced. Participants self-reported a median of four scientific papers authored, suggesting most were experienced in skimming and reading scientific papers. None of the participants were affiliated with the organization conducting this research. Participants were compensated with $20 (USD) for their time.

5.2 Procedure

We conducted the study remotely through a video conferencing platform. As a result, we could not account for factors such as the technological setup of each participant (e.g., hardware specifications, screen size) and the presence of external distractions. Rather than a limitation, we see these factors as enabling a more diverse evaluation of the system within realistic reading environments. We recorded each participant's screen and audio, and logged all interactions with Scim during the study session. Each study was between 45 minutes and one hour long, and consisted of three stages: tutorial, skimming tasks, and exit interview.

5.2.1 Tutorial. Participants first completed a self-guided tutorial that introduced the features of Scim within a sample paper (selected to be similar in style and length to those used in the skimming tasks). The tutorial directed participants to explicitly perform actions involving usage of each of Scim's features. Participants shared their screen during this process and the experimenters answered any usability questions that arose.

5.2.2 Skimming Tasks. To reflect the diversity of scenarios in which readers skim scientific papers, we used three tasks to prompt readers into a skimming mindset. Given the cognitive effort and time required to skim multiple scientific papers in a single study session, participants were split into two groups and assigned separate tasks. All participants were provided with the same three papers [27, 60, 64] selected from recent proceedings of the ACM Symposium on User Interface Software and Technology. We prompted participants with the following scenario: *Imagine you are a member of a paper reading group, and you need to select a paper to present. Please skim the selected paper to get a high-level gist of its content.*

In the first group, participants were given six minutes to skim a paper, and another six minutes to answer ten true or false questions on high-level concepts from the paper (Task 1). In the second group, participants were given six minutes to skim a paper and another six minutes to list three strengths and weaknesses of the paper (Task 2). After a short break, they were given another six minutes to skim a second paper and fifteen minutes to create an outline of a short presentation for the paper (Task 3). All participants were allowed to revisit the paper and encouraged to talk aloud while completing their tasks. All reading was done using Scim.

5.2.3 Exit Interview. After completing the reading tasks, participants completed a System Usability Survey [10], a ten-question industry-standard survey assessing a system's perceived usability. Participants then answered several questions regarding the anticipated usefulness of each of Scim's features and quality of the automatically generated highlights, and miscellaneous demographic questions. We concluded with a 10-15 minute semi-structured interview with participants about their experience using Scim, probing into features or interactions they liked, areas for improvement, and opportunities for future AI-assisted reading interfaces.

5.3 Analysis

Two authors analyzed the interview data, following a qualitative approach described in [19]. Throughout the analysis process, transcripts were created from audio recordings, themes developed and refined, and relevant utterances from participants extracted. In our results, we refer to participants with the pseudonyms P1-P13. The utterances presented below were edited to elide identifying information, while preserving their meaning.

6 RESULTS

6.1 RQ1. How do readers use Scim and its faceted highlights for skimming papers?

Most participants were positive about Scim and saw the potential for AI-infused reading interfaces to assist in facilitating the process of knowledge distillation while skimming. For instance, P4 mentioned that Scim could help facilitate currently manual processes with its automatically generated faceted highlights:

> *I'm working on a lit review project and I realize that your tool will be super helpful given that the information I need to extract from the papers are mostly method and significance/contribution, which are already highlighted by the tool.*

Scim was also well-regarded by less seasoned readers. For instance, P10 praised Scim's potential for guidance as they continued to develop their own skimming strategies.

> *I think it was quite amazing, especially as a first-year, just getting into this. So I'm always looking for those keywords, and if I don't find those I feel really lost. For papers that didn't have those key words, it was very helpful for something to at least give a clear area of where to look.*

By design, Scim offered participants an augmented skimming experience that resembled closely a traditional document reader. Participants found the system "easy to use" (P8) and "very understandable" (P9), and skimmed by scrolling linearly through a paper, using the faceted highlights as cues for visual attention. Many appeared to adopt a satisficing skimming strategy, relying on the faceted highlights and more conventional cues (e.g., lead sentences of paragraphs, presence of jargon, visual content) to approximate the perceived importance of a particular section of a paper. This dominant usage pattern is to be expected since participants were given a short amount of time to skim a long research paper, participants were relatively unfamiliar with the system (other than interactions from the tutorial), and paper reading processes are deeply ingrained within researcher practices.

Some found that faceted highlights helped them attend to areas of the paper they might have skipped otherwise. For example, one

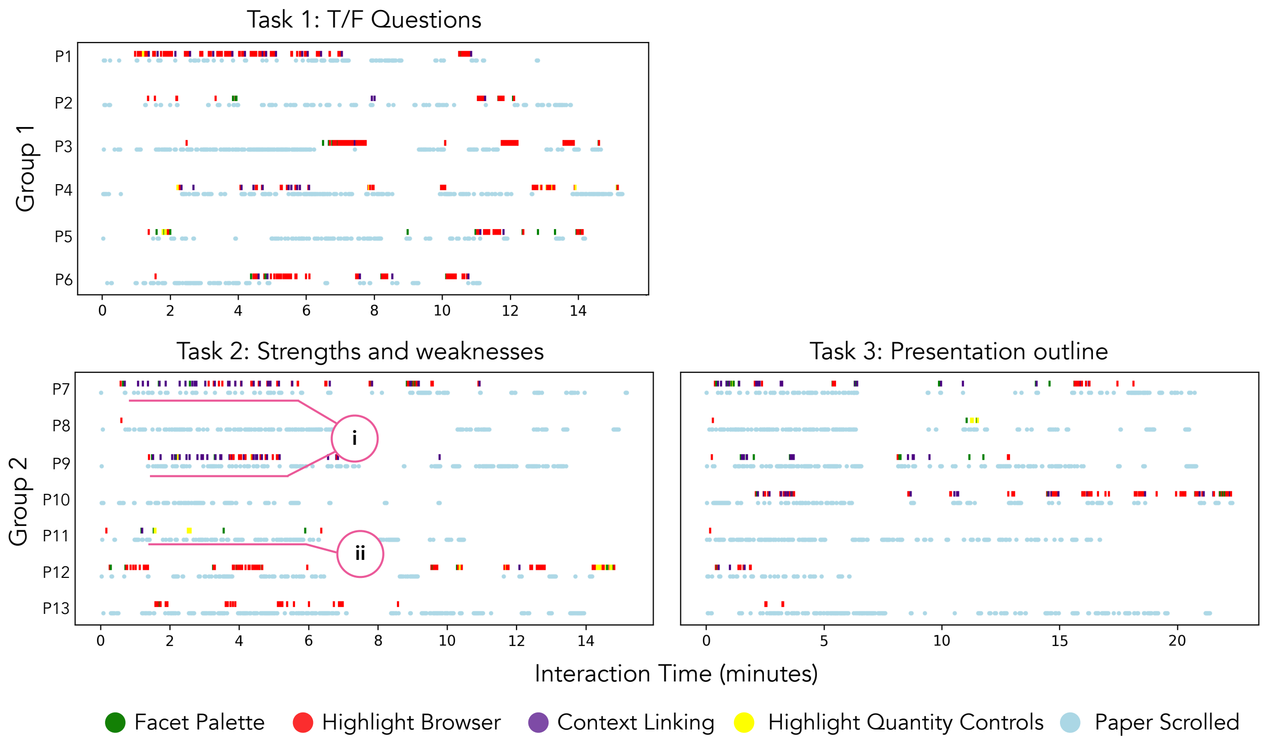

Figure 3: Overview of participant interactions with the Scim reading interface. General scrolling of the paper is shown in light blue, and interactions with various features are marked. Subjects are split by group, with tasks as described in Section 5.2.2. We see instances of participants (i) clicking on sentences within the *highlight browser* to rapidly navigate through a paper (indicated by spans of red and purple marks), and (ii) making multiple high-level passes through paper after filtering highlights to a single facet (indicated by green marks which denote interactions to select a particular facet within the *facet palette*).

participant noted that the Novelty highlights in the Related Work section—a section they typically skip during skimming—prompted them to slow down and read highlighted sentences with relevant information (P3). Participants also mentioned how AI support could be particularly helpful while skimming under time pressure, since they believed the AI-powered highlights could surface sentences of greater importance than readers could identify manually given limited time (P1, P5, P7, P8, P9, P10).

In addition to the inline faceted highlights, participants also used Scim's highlight browser to support navigation and scaffold comprehension while skimming a paper. Rather than scrolling through a paper itself, participants would instead scroll through the list of faceted highlights in the browser (P1, P3, P7, P9). Figure 3 illustrates this pattern, specifically in rows where scrolling interactions through the paper (shown in light blue) are punctuated by scrolling interactions in the highlight browser (red markers) and context linking interactions from highlights in the browser to the paper context (purple markers). The highlight browser offered readers a sentence-level index into the paper, allowing efficient contextual exploration of a particular highlight of interest.

Participants used Scim's facet palette to filter highlights to reflect specific reading goals. For instance, they often began a skim by searching for information that conveyed the paper's significance, using the facet palette to filter highlights to two specific facets—Objective and Novelty—before reading through the filtered highlights in the browser. Some would further click on each highlight to view it in the context of the paper, after which they would resume skimming linearly through the paper. Other participants opted for a multi-pass approach: they first skimmed linearly through the paper, then used the facet palette to filter the highlights to one or more facets of interest (e.g., Method or Result), and performed one or more additional passes over the paper with the filtered highlights.

Participants also saw Scim as potentially useful in assimilating the knowledge gained while skimming. For instance, P3 mentioned how they used Scim's highlight browser as a reminder of what they had previously skimmed in the paper and to verify their understanding:

> *It was more useful as a post-reading exercise. After I skimmed the entire thing, I'd go into the sidebar, click on the different colored tabs and just skim through the sentences corresponding to that filter. Just to summarize again for myself and get a gist of the paper.*

Usability. Participants gave Scim an average SUS score of 86.8, indicating excellent overall usability (based on the canonical interpretation of SUS scores [6]). They lauded Scim for its visual and interactive simplicity, and its ability to nest AI-powered features within an intuitive layout which paralleling software for other reading contexts. Those reporting lower usability scores noted the potential for inconsistency within the AI highlights, resulting in an additional cognitive load in occasionally needing to verify the accuracy of the AI highlights.

Additional Features. Participants suggested additional features they believed would improve the usability of Scim. Some mentioned annotating papers during their usual skimming sessions, and asked for analogous features to engage with the faceted highlights within Scim, such as the ability to add, update, and annotate highlights. The increased interactivity may also help Scim better support both skimming and deep reading processes. To persist the augmented reading experience, some suggested Scim should support exporting automated highlights and any user-generated annotations as a post-reading summary.

6.2 RQ2. How does AI assistance influence readers while skimming papers?

Numerous guidelines have emerged over the past two decades from the human-computer interaction literature outlining design paradigms for effective human interaction with AI-infused systems [1, 32, 54]. We present our findings from Scim to expand upon these guidelines within the specific context of AI-infused reading interfaces.

6.2.1 Errors in AI-Powered Highlighting. Participants identified several types of errors that they encountered while using Scim: (i) salience detection errors—highlighting of an unimportant sentence (False Positive) or lack of highlighting of an important sentence (False Negative), (ii) facet classification errors, and (iii) PDF processing errors. Most participants were not affected by unimportant highlights while scrolling through a paper since they could be easily ignored, and mentioned simply spending more time during a skim in sections of interest when Scim provided less highlights than expected. While individual errors did not appear to strongly influence the usability of Scim, participants appeared to develop a judgment on the perceived quality of the AI highlights while skimming a paper. They were quick to form initial impressions, focusing on the precision of the first few highlights in the paper or highlight browser. Those who gained a positive impression continued to use Scim for skimming and completing the tasks:

> *This was the first sentence that I clicked. This actually pointed out a very important sentence in this paper. At that time, I felt like the tool was very powerful. (P7)*

Others who deemed the quality of the highlights too low based on their first impressions disabled the features of Scim to complete their skim of subsequent papers (P11, P12). Our results echo prior studies on trust formation and automation reliance that found trust declines rapidly and slowly increases over periods of appropriate behavior [22, 52].

While some participants felt their experience with the system was strongly affected by errors made by the AI models (P5, P11, P12), others were more tolerant to potential errors, since they felt the benefits of having automated reading guidance outweighed the potential for errors (P10). Our results do not indicate any tolerable level of error within AI-infused reading interfaces, but rather that expectations of individual researchers will drive the eventual decision to adopt these intelligent tools. That said, we found that errors in Scim served to violate an assumption of the system's trustworthiness, and were therefore better remembered than "correct" highlights made by the system.

To reduce the impact of errors, some participants suggested Scim could include affordances that offered readers better transparency into the AI model. For instance, the system could convey the AI's confidence in highlighting particular sentences (e.g., by varying the opacity of highlights or with a confidence score), or how the AI selected the prioritization of sentences to be shown when readers modified the density of highlights with the highlight controls.

6.2.2 Trade-offs in Precision And Recall for AI-Powered Highlights. An alternative approach to reduce the impact of errors in AI-infused systems is to focus on calibrating the system to a high precision, and hence avoid false positive errors at the expense of recall. We found that participants' opinions varied regarding the importance of recall and precision within the AI-powered highlights. Pilot studies with Scim suggested that readers typically wanted to see a greater number of highlights distributed throughout the paper, a reassurance of the system's ability to recall all important information. While in our usability study some participants echoed this preference for high recall, others preferred fewer highlights with greater precision:

> *If it began with a place where the highlights were restricted to where the AI has very high confidence, and it's up to me to populate with more noise, that might be more effective for doing a first pass through deciding if I want to read. (P12)*

Within AI-infused reading interfaces, a lack of AI-powered highlights requires a reader to revert to a typical skimming and reading flow (appearing to provide insufficient support). In contrast, an overabundance of irrelevant AI-powered highlights requires a reader to recover from errors by ignoring those highlighted sentences. While the cognitive effort required to ignore individual highlights may not be significant, these loads may accumulate over time, resulting in lowering impressions of AI competence or abandonment of the reading interface entirely.

6.2.3 Potentially Biased Narratives from AI-Powered Highlighting. While automated highlights may identify important content for readers, the resultant story that is crafted by the selected highlights could unintentionally offer a biased skim of a paper. For instance, P8 noted how they relied heavily on Scim's highlights:

> *Having the highlights in the paper and then having the different colors was a really helpful system as I'm skimming through. My eyes definitely focused on the highlights, and maybe read 10% of what wasn't highlighted..which is good and bad.*

They further described their typical skimming technique involved a penchant for critique, but felt that skimming only highlights provided by the AI caused them to feel unnaturally positive about a paper:

> *I felt that I got a very positive impression of the work based on the highlights...which for me is surprising because I tend to pick out tiny details...and really complain about the small details like a wizard of oz method. But*

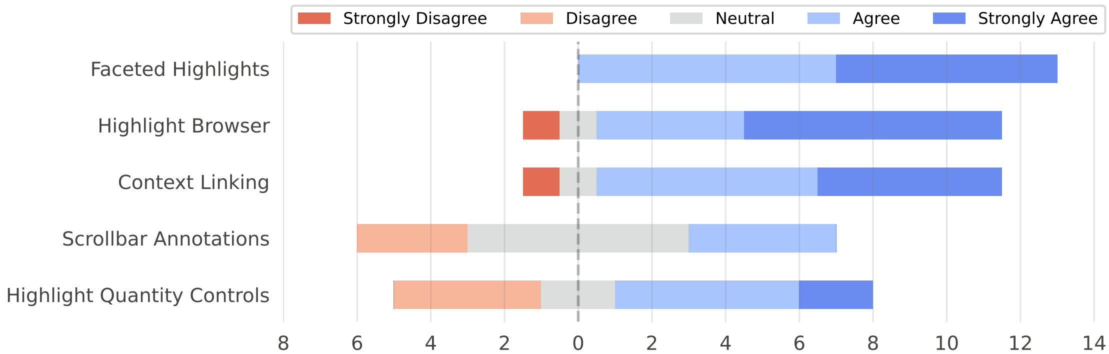

Figure 4: Distribution of responses regarding desire to use each feature in a future reading interface for scientific papers.

> *when I'm hopping from highlight to highlight there was a bit of skew towards the good, important things are highlighted.*

Though we had designed Scim's facets based on common information needs of skimming readers, facets such as OBJECTIVE and NOVELTY may by nature tend to favor spotlighting sentences that reinforce the strengths of the paper rather than its limitations. The other facets—METHOD and Result—also tend to prescribe objective sentences to readers, and therefore do not typically expose the limitations of a paper that may be surfaced in a longer, more critical read.

6.2.4 Toward Adopting AI-Infused Reading Interfaces. Overall, participants appeared optimistic about the potential for intelligent reading interfaces to support the paper skimming experience despite their potential for errors, but suggested improvements to enable future adoption. Participants desired more control and personalization over highlights suggested by the system, as many believed there should be some fine-tuning into what was highlighted for each individual reader. One participant remarked how an interpretation of a paper represented by AI-powered highlights may resonate across different readers, though perhaps highlights pertaining to more technical aspects may be more universally useful.

> *Reading these research articles is like art, music. It's based on interpretation, it's like a very abstract concept. Different people reading papers will have different points...Though I think there are definitely some common concepts that's similar across different papers and very technical things that can be extrapolated by the AI, like methods. (P11)*

Participants also believed that continued usage of Scim would be useful to familiarize themselves with the various features, facet color associations, and reliability of the AI-powered highlights. One participant mentioned how their "main issue is getting used to seeing highlights that aren't my own" (P13), suggesting that readers may need some time to acclimate to AI assistance in traditionally manual processes.

7 DISCUSSION

7.1 Design Recommendations

R1. Support details-on-demand. Reading interfaces should offer readers the ability to view details-on-demand. Scim provided a context linking feature which connected readers from highlights in the browser to its context in the paper. Participants also suggested alternative designs for details-on-demand. One suggested providing a summary of a paper that could be iteratively "zoomed in" on, perhaps visualized as a tree-like structure. Readers could then interact with sentences in the tree, expanding it on-demand to display additional detail from the paper, akin to Wikum, a tool for recursive summarization of discussion forums [75].

R2. Prefer a consistent distribution of generated highlights. Through initial usability testing in the iterative design process of Scim, we found that readers generally preferred highlights distributed throughout the entire paper, since they expected salient content within most sections of the paper. When skimming past extended regions of text without highlights, readers believed the AI system was missing relevant content, and devolved their skim into a detailed reading of those sections. Developers of future AI-infused reading interfaces should carefully consider the impact of probabilistic features (e.g., the distribution of generated highlights) on reader perceptions, accounting for potentially adverse effects.

R3.Support passive and active reading processes. While some participants preferred to passively skim papers, others engaged in more active reading such as highlighting and note-taking in an external tool while skimming. Reading interfaces should consider providing affordances that support both passive and active reading processes. For instance, they could present nested features—colored highlights support readers in passive skimming, and clicking on a highlight could reveal tools for on-demand annotation. Their implementation should be uncluttered and intuitive, supporting readers while minimizing any introduced distractions from reading.

R4. Persist augmented reading beyond a single reading session. Participants in our study wanted to export paper annotations after an initial skim. Augmented reading interfaces provide an additional layer of information atop the paper itself that readers spend cognitive effort in consuming. These systems should therefore give users the ability to retain the reading process beyond the lifespan of a single reading session. The exported information represents a snapshot of the human-AI collaborative reading process, including both AI-provided augmentations such as faceted highlights and annotations created by the reader. Persistence can also manifest as saving user preferences or interactions with the paper (e.g., likes or dislikes).

R5. Support readers in recovering from errors in AI features. Instantiating guidelines for designing interactions with imperfect AI-infused systems [1], AI-infused reading interfaces should offer readers paths to recovery when the AI models err. In Scim, visualizing AI suggestions as highlights atop a paper naturally lends itself to a simple recovery mechanism when the highlights are irrelevant or inadequate—ignore the highlight, or read the surrounding text to gain additional context. However, AI errors may not always be easily recoverable from by design, for instance if paper context is inaccessible in a particular reading interface. Reading interfaces should also enable readers to stay in control of their reading experience—provide readers with the ability to efficiently disable and reactivate the AI augmentations within the reading interface. They should also assist readers in developing an intuition for the system's reliability—offer transparency into the system's reasoning or confidence in highlighting a particular sentence. This recommendation hearkens back to our fourth design motivation for Scim (Section 3.2).

R6. Reduce biases induced by the AI-infused reading interface. Some participants in our study relied on skimming mostly content highlighted by the AI, reading less of the non-highlighted text than they may have in a traditional skim. In our results, we discussed potential consequences of this behavior—one participant believed Scim inadvertently induced a positive bias on their skim of a paper through its selective highlighting. Future AI-infused reading interfaces should carefully consider how the design and visualization of their AI-powered features may impact readers, particularly in offering a biased or selectively incomplete narrative of a paper.

7.2 Limitations

One limitation of our study is its focus on only one category of papers within the human-computer interaction field of computer science. Papers across computer science (let alone other scholarly disciplines) can vary greatly in structure, clarity, or length. Despite grounding the design of Scim in established natural language processing techniques and user interface paradigms, future studies should evaluate the system's generalizability on a variety of papers.

Participants had limited time to interact with Scim during the user study, and interacted with a small set of predetermined papers. Some participants expected to benefit more from Scim with continued usage, particularly as they gained familiarity with the faceted highlights, facets' color associations, and types of reading tasks the system could support. Though, given longitudinal access to Scim, participants may also discover limitations that could preclude them from further adoption. While observations from our study suggest readers would like to use Scim within numerous reading scenarios, we believe a longitudinal deployment with in-situ feedback would be valuable in assessing how readers might truly use the tool "in the wild".

7.3 Future Work

7.3.1 Highlights via collaborative reading. One reason readers may be hesitant to adopt an augmented reading interface like Scim is distrust of the system's ability to provide the most relevant highlights. Some were concerned whether AI-powered highlights could ever instill enough confidence to allow readers to navigate between only the automated highlights. Instead, some mentioned potentially greater trust in highlights created by other people (e.g., fellow researchers). We note that platforms such as Medium show "popular highlights," which appear popular. Combining social and AI-powered highlights raises interesting design challenges.

7.3.2 Eliciting user interactions for improved highlights. As readers continue to interact with augmented papers, we envision an opportunity for AI models to adaptively learn from these user interactions. Future systems could elicit and learn from passive (e.g., dwell patterns on certain parts of a paper) or active (e.g., highlighting a sentence or removing an AI-generated highlight) user interaction signals. They could also explore user interactions with features designed for conveying the uncertainty of individual highlights and recovery from AI errors as trainable instances for improving the underlying models.

7.3.3 Tailoring the reading experience through personalization. In addition to improving the overall performance of the models, future reading interfaces could be tailored to individual readers' characteristics. One could explore approaches for modeling long-term user attributes—experience generally reading scientific papers, experience with the research area of the current paper, types of information a reader typically looks for in a paper, and their expectations for AI assistance within a reading interface—or short-term states for a particular paper, such as their goals for reading the paper or the number of times the paper had been read previously.

8 CONCLUSION

We presented Scim, an augmented reading interface that provides faceted highlights to support researchers when skimming scientific papers. Using Scim as a design probe with 13 participants, we found that colored highlights placed within the paper and organized into rhetorical facets within a highlight browser complemented participants' existing skimming processes by facilitating behaviors such as rereading and guided reading. We found that participants were eager to use Scim's contextually-linked faceted highlights and highlight browser in their future reading interfaces, and were more broadly enthusiastic about the potential of AI-augmented reading interfaces. Based on qualitative insights from our study, we identified several design recommendations that we hope can inform and inspire designers of future augmented reading interfaces.

ACKNOWLEDGMENTS

This project is supported in part by NSF Grant OIA-2033558, NSF RAPID award 2040196, ONR grant N00014-21-1-2707, and the Allen Institute for Artificial Intelligence (AI2).

REFERENCES

[1] Saleema Amershi, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, Eric Horvitz, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, and Paul N. Bennett. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–13.

[2] Waleed Ammar, Dirk Groeneveld, Chandra Bhagavatula, Iz Beltagy, Miles Crawford, Doug Downey, Jason Dunkelberger, Ahmed Elgohary, Sergey Feldman, Vu Ha, Rodney Kinney, Sebastian Kohlmeier, Kyle Lo, Tyler Murray, Hsu-Han Ooi, Matthew Peters, Joanna Power, Sam Skjonsberg, Lucy Wang, Chris Wilhelm, Zheng Yuan, Madeleine van Zuylen, and Oren Etzioni. 2018. Construction of the Literature Graph in Semantic Scholar. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)*. Association for Computational Linguistics, New Orleans - Louisiana, 84–91.

[3] Abdelkrime Aries, Djamel Eddine Zegour, and Walid-Khaled Hidouci. 2019. Automatic text summarization: What has been done and what has to be done. *CoRR* abs/1904.00688 (2019).

[4] Sriram Karthik Badam, Zhicheng Liu, and Niklas Elmqvist. 2019. Elastic Documents: Coupling Text and Tables through Contextual Visualizations for Enhanced Document Reading. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 661–671.

[5] Xiaomei Bai, Mengyang Wang, Ivan Lee, Zhuo Yang, Xiangjie Kong, and Feng Xia. 2019. Scientific Paper Recommendation: A Survey. *IEEE Access* 7 (2019), 9324–9339. Conference Name: IEEE Access.

[6] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies* 4, 3 (may 2009), 114–123.

[7] Joeran Beel and Bela Gipp. 2009. Google Scholar's Ranking Algorithm : An Introductory Overview. In *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, Vol. 1. "International Society for Scientometrics and Informetrics", "Rio de Janeiro, Brazil", 230–2041.

[8] Joeran Beel, Bela Gipp, Stefan Langer, and Corinna Breitinger. 2016. Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries* 17, 4 (Nov. 2016), 305–338.

[9] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 1077–1086.

[10] John Brooke. 2013. SUS: a retrospective. *Journal of Usability Studies* 8, 2 (Feb. 2013), 29–40.

[11] George Buchanan and Tom Owen. 2008. Improving skim reading for document triage. In *Proceedings of the second international symposium on Information interaction in context - IIiX '08*. ACM Press, London, United Kingdom, 83.

[12] Isabel Cachola, Kyle Lo, Arman Cohan, and Daniel Weld. 2020. TLDR: Extreme Summarization of Scientific Documents. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, 4766–4777.

[13] Patricia L. Carrell. 1985. Facilitating ESL Reading by Teaching Text Structure. *TESOL Quarterly* 19, 4 (1985), 727–752.

[14] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. 2018. Universal Sentence Encoder. *CoRR* abs/1803.11175 (2018).

[15] 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* (San Francisco, California, USA) *(UIST '98)*. Association for Computing Machinery, New York, NY, USA, 123–132.

[16] Ed Chi, Michelle Gumbrecht, and Lichan Hong. 2007. *Visual Foraging of Highlighted Text: An Eye-Tracking Study*. Vol. 4552. Springer Berlin Heidelberg, Berlin, Heidelberg.

[17] Ed H. Chi, Lichan Hong, Michelle Gumbrecht, and Stuart K. Card. 2005. ScentHighlights: highlighting conceptually-related sentences during reading. In *Proceedings of the 10th international conference on Intelligent user interfaces - IUI '05*. ACM Press, San Diego, California, USA, 272.

[18] Arman Cohan, Iz Beltagy, Daniel King, Bhavana Dalvi, and Daniel S. Weld. 2019. Pretrained Language Models for Sequential Sentence Classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Online, 3691–3697.

[19] John W Creswell and Cheryl N Poth. 2016. *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications, United States of America.

[20] Geoffrey Duggan and Stephen Payne. 2009. Text Skimming: The Process and Effectiveness of Foraging Through Text Under Time Pressure. *Journal of experimental psychology. Applied* 15 (09 2009), 228–42.

[21] Geoffrey B. Duggan and Stephen J. Payne. 2011. Skim reading by satisficing: evidence from eye tracking. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 1141.

[22] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce, and Hall P. Beck. 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies* 58, 6 (June 2003), 697–718.

[23] Allen Institute for Artifical Intelligence. 2022. MMDA - Multimodal Document Analysis. https://github.com/allenai/mmda

[24] Robert L. Fowler and Anne S. Barker. 1974. Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology* 59, 3 (1974), 358–364.

[25] Jamey Graham. 1999. The reader's helper: a personalized document reading environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*. ACM Press, Pittsburgh, Pennsylvania, United States, 481–488.

[26] 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*. ACM, Seoul Republic of Korea, 461–475.

[27] Anhong Guo, Junhan Kong, Michael Rivera, Frank F. Xu, and Jeffrey P. Bigham. 2019. StateLens: A Reverse Engineering Solution for Making Existing Dynamic Touchscreens Accessible. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, New Orleans LA USA, 371–385.

[28] Yufan Guo, Roi Reichart, and Anna Korhonen. 2013. Improved Information Structure Analysis of Scientific Documents Through Discourse and Lexical Constraints. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Atlanta, Georgia, 928–937.

[29] 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 413, 18 pages.

[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* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 3–9.

[31] Kenji Hirohata, Naoaki Okazaki, Sophia Ananiadou, and Mitsuru Ishizuka. 2008. Identifying Sections in Scientific Abstracts using Conditional Random Fields. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*. Association for Computational Linguistics, Hyderabad, India.

[32] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*. ACM Press, Pittsburgh, Pennsylvania, United States, 159–166.

[33] Edward W. Ishak and Steven K. Feiner. 2006. Content-aware scrolling. In *Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST '06*. ACM Press, Montreux, Switzerland, 155.

[34] S. Keshav. 2007. How to read a paper. *Comput. Commun. Rev.* 37 (2007), 83–84.

[35] 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*. ACM, Berlin Germany, 423–434.

[36] 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*. ACM, Honolulu Hawaii USA, 123–127.

[37] Nicholas Kong, Marti A. Hearst, and Maneesh Agrawala. 2014. Extracting references between text and charts via crowdsourcing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 31–40.

[38] Natasha Lacroix. 1999. Macrostructure construction and organization in the processing of multiple text passages. *Instructional Science* 27, 3/4 (1999), 221–233.

[39] 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*. ACM, San Jose California USA, 5203–5214.

[40] Maria Liakata, Shyamasree Saha, Simon Dobnik, Colin Batchelor, and Dietrich Rebholz-Schuhmann. 2012. Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics* 28, 7 (April 2012), 991–1000.

[41] Maria Liakata, Simone Teufel, Advaith Siddharthan, and Colin Batchelor. 2010. Corpora for the Conceptualisation and Zoning of Scientific Papers. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), Valletta, Malta.

[42] Xiaozhong Liu, Zhuoren Jiang, and Liangcai Gao. 2015. Scientific Information Understanding via Open Educational Resources (OER). In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago, Chile) *(SIGIR '15)*. Association for Computing Machinery, New York, NY, USA, 645–654.

[43] Ziming Liu. 2005. Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation* 61, 6 (Jan.

2005), 700–712. Publisher: Emerald Group Publishing Limited.
[44] Fernando Loizides and George Buchanan. 2009. An Empirical Study of User Navigation during Document Triage. In *Research and Advanced Technology for Digital Libraries*. Vol. 5714. Springer Berlin Heidelberg, Berlin, Heidelberg, 138–149.
[45] Tonja Machulla, Mauro Avila, Pawel Wozniak, Dillon Montag, and Albrecht Schmidt. 2018. Skim-Reading Strategies in Sighted and Visually-Impaired Individuals: A Comparative Study. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference* (Corfu, Greece) *(PETRA '18)*. Association for Computing Machinery, New York, NY, USA, 170–177.
[46] 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*. ACM, Honolulu HI USA, 1–13.
[47] Michael E. Masson. 1982. Cognitive processes in skimming stories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8, 5 (1982), 400–417.
[48] Michael E. J. Masson. 1983. Conceptual processing of text during skimming and rapid sequential reading. *Memory & Cognition* 11, 3 (May 1983), 262–274.
[49] Martha J. Maxwell. 1972. Skimming and Scanning Improvement: The Needs, Assumptions and Knowledge Base. *Journal of Reading Behavior* 5, 1 (March 1972), 47–59. Publisher: SAGE Publications.
[50] Microsoft. 2022. Scroll bar map mode and bar mode - Visual Studio (Windows). https://docs.microsoft.com/en-us/visualstudio/ide/how-to-track-your-code-by-customizing-the-scrollbar
[51] Mozilla. 2022. PDF.js. https://github.com/mozilla/pdf.js
[52] Bonnie M. Muir. 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* 37, 11 (1994), 1905–1922.
[53] Ani Nenkova and Kathleen McKeown. 2012. A Survey of Text Summarization Techniques. In *Mining Text Data*, Charu C. Aggarwal and ChengXiang Zhai (Eds.). Springer US, Boston, MA, 43–76.
[54] Donald A. Norman. 1994. How Might People Interact with Agents. *Commun. ACM* 37, 7 (jul 1994), 68–71.
[55] Mahsan Nourani, Chiradeep Roy, Jeremy E Block, Donald R Honeycutt, Tahrima Rahman, Eric Ragan, and Vibhav Gogate. 2021. Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. In *26th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, USA, 340–350.
[56] Brett Powley, Robert Dale, and Ilya Anisimoff. 2009. Enriching a document collection by integrating information extraction and PDF annotation. In *Document Recognition and Retrieval XVI*, Kathrin Berkner and Laurence Likforman-Sulem (Eds.), Vol. 7247. International Society for Optics and Photonics, SPIE, San Jose, CA, USA, 47 – 56.
[57] David N. Rapp and Paul van den Broek. 2005. Dynamic text comprehension: An integrative view of reading. *Current Directions in Psychological Science* 14, 5 (2005), 276–279.
[58] Keith Rayner, Elizabeth R. Schotter, Michael E. J. Masson, Mary C. Potter, and Rebecca Treiman. 2016. So Much to Read, So Little Time: How Do We Read, and Can Speed Reading Help? *Psychological Science in the Public Interest* 17, 1 (May 2016), 4–34.
[59] William R. Reader and Stephen J. Payne. 2007. Allocating Time across Multiple Texts: Sampling and Satisficing. *Hum.-Comput. Interact.* 22, 3 (aug 2007), 263–298.
[60] Mark Richardson, Matt Durasoff, and Robert Wang. 2020. Decoding Surface Touch Typing from Hand-Tracking. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. ACM, Virtual Event USA, 686–696.
[61] John P. Rickards. 1980. Notetaking, Underlining, Inserted Questions, and Organizers in Text: Research Conclusions and Educational Implications. *Educational Technology* 20, 6 (1980), 5–11.
[62] 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* (Los Angeles, California, USA) *(CHI '98)*. ACM Press/Addison-Wesley Publishing Co., USA, 249–256.
[63] Athar Sefid and C. Lee Giles. 2022. SciBERTSUM: Extractive Summarization for Scientific Documents. *CoRR* abs/2201.08495 (2022).
[64] Ke Sun, Chun Yu, Weinan Shi, Lan Liu, and Yuanchun Shi. 2018. Lip-Interact: Improving Mobile Device Interaction with Silent Speech Commands. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, Berlin Germany, 581–593.
[65] Craig S. Tashman and W. Keith Edwards. 2011. Active Reading and Its Discontents: The Situations, Problems and Ideas of Readers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2927–2936.
[66] Carol Tenopir, Donald King, Sheri Edwards, and Lei Wu. 2009. Electronic Journals and Changes in Scholarly Article Seeking and Reading Patterns. *Carol Tenopir* 61 (01 2009).
[67] Simone Teufel and Marc Moens. 2002. Articles Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status. *Computational Linguistics* 28, 4 (2002), 409–445.
[68] Simone Teufel, Advaith Siddharthan, and Colin Batchelor. 2009. Towards Domain-Independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Singapore, 1493–1502.
[69] Hedwig von Restorff. 1933. Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung* 18, 1 (Dec. 1933), 299–342.
[70] Alan J. Wecker, Joel Lanir, Osnat Mokryn, Einat Minkov, and Tsvi Kuflik. 2014. Semantize: visualizing the sentiment of individual document. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces - AVI '14*. ACM Press, Como, Italy, 385–386.
[71] Wen Xiao and Giuseppe Carenini. 2019. Extractive Summarization of Long Documents by Combining Global and Local Context. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 3011–3021.
[72] 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 - WWW '17 Companion*. ACM Press, Perth, Australia, 311–319.
[73] Ji Soo Yi. 2014. QnDReview: read 100 CHI papers in 7 hours. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 805–814.
[74] 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* (Pittsburgh, Pennsylvania, USA) *(HYPERTEXT '98)*. Association for Computing Machinery, New York, NY, USA, 50–57.
[75] Amy X. Zhang, Lea Verou, and David Karger. 2017. Wikum: Bridging Discussion Forums and Wikis Using Recursive Summarization. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2082–2096.
[76] Zheng Zhao, Shay B. Cohen, and Bonnie Webber. 2020. Reducing Quantity Hallucinations in Abstractive Summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, 2237–2249.