The Pennsylvania State University

The J. Jeffrey and Ann Marie Fox Graduate School

UNDERSTANDING RESEARCHERS' BEHAVIORS AND DESIGN CONSIDERATIONS FOR AI-ASSISTED SCIENTIFIC CAPTION WRITING

A Thesis in

Informatics

by

Ho Yin Ng

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

December 2024

The thesis of Ho Yin Ng was reviewed and approved by the following:

Ting-Hao 'Kenneth' Huang
Associate Professor of Information Sciences and Technology
Thesis Advisor

Saeed Abdullah
Associate Professor of Information Sciences and Technology

Xiaolong (Luke) Zhang
Associate Professor of Information Sciences and Technology

Dongwon Lee
Professor of Information Sciences and Technology
Professor in Charge for IST Graduate Program

ABSTRACT

This study investigates the potential of AI, specifically Large Language Models (LLMs), in assisting researchers with figure caption writing—a crucial yet often tedious aspect of academic publishing. While previous research has focused on caption generation for readers, our study uniquely addresses the writer's perspective. We conducted a mixed-methods study with 18 participants, involving a writing task using AI-generated captions and semi-structured interviews. The study examined participants' caption writing practices, challenges, and views on AI assistance in scientific publishing. Quantitative analysis compared preferences among AI-generated caption and self-reported improvements, while qualitative analysis revealed insights across Task Characteristics, User Capabilities and Perceptions, Ecosystem Constraints, and Desired Interaction Features. Our findings provide a framework for understanding effective caption creation, inform researchers' writing processes, and identify design considerations to guide future AI-assisted academic writing tools in enhancing scientific communication.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Scientific publications rely heavily on figures and their accompanying captions to convey complex information concisely. Captions play a critical role in making figures more interpretable and accessible, yet writing effective captions remains a challenging and tedious task for many researchers.

A fundamental misalignment exists between how scientific content is written and how it is consumed. While writers typically focus their efforts on crafting detailed body text, readers often begin by examining figures and their captions before delving into the main text. This disconnect is particularly problematic because readers expect to find specific information in particular places, and when these expectations are violated, they must divert energy from understanding the content to unraveling its structure. Despite captions serving as crucial bridges between figures and detailed textual explanations, researchers often dedicate minimal time to caption writing compared to the extensive effort invested in the main text.

With the advancement of AI technologies, particularly Large Language Models (LLMs), there is growing potential to assist researchers in this process. However, the application of AI in scientific writing raises important questions about how researchers interact with and perceive such assistance from AI technologies.

Although previous research has explored AI-generated captions from the reader's perspective [22], our study uniquely addresses the writer's viewpoint. We investigate how researchers engage with AI-generated caption suggestions and how this interaction might inform the design of future AI writing assistants for scientific publications.

To explore this space, we conducted a mixed-method user study with 18 participants, primarily Ph.D. students from diverse fields. Participants engaged in a writing task in which they rewrote captions for figures from their own published work using AI-generated suggestions. This was complemented by semi-structured interviews to understand their existing practices, challenges, and perspectives on AI assistance in caption writing. Our study address two main research questions:

RQ1: How do researchers interact with and utilize multiple AI-generated captions in the context of writing captions for scientific publications?

RQ2: What are the possible design considerations for AI writing assistants in scientific figure caption writing from the researchers' perspective?

Our analysis leverages an existing design space for AI-assisted writing [16], adapting it to focus on the needs of researchers in scientific caption writing. While the original framework included the theme of 'Technology', our study emphasizes user-centered aspects, particularly the themes of 'Task', 'User', 'Ecosystem', and 'Interaction'. This adaptation allows us to explore the unique challenges and requirements of caption writing in scientific publications from the perspective of the researchers themselves.

Our findings reveal a complex connection between the researcher's existing practices, their interactions with AI-generated suggestions, and the broader ecosystem of scientific publishing. We adapted the original design space to our specific context to create four themes: 'Task Characteristics', 'User Capabilities and Perceptions', 'Ecosystem Constraints', and 'Desired Interaction Features'. These themes serve as the basis for identifying insights from participants to form design considerations. These insights contribute to our understanding of how AI can be effectively integrated into the scientific writing process, particularly for figure captions. This paper makes the following contributions:

1. An empirical understanding of how researchers interact with and perceive AI-generated captions in the context of scientific writing.

2. Insights into the diverse needs, challenges, and evaluation criteria researchers apply when considering AI assistance for caption writing.

3. Design implications for future AI writing assistants tailored to the specific requirements of scientific figure captions.

Our work lays the ground for developing more effective and user-centered AI writing assistants that can enhance the quality and efficiency of scientific communication while addressing the specific needs and concerns of researchers.

Chapter 2

Related Work

2.1 Figure Caption Generation and Evaluation

Despite the growing focus on evaluating AI models for generating captions of scientific figures, these assessments often overlook the writer's perspective. Previous studies have widely used automatic metrics to evaluate the quality of AI-generated figure descriptions. For instance, BLEU measures word overlap between the generated and reference texts [1, 4–6, 17, 21] and, in a similar vein, ROUGE focuses on how much content from the reference text is included in the generated text [4–6, 25]. METEOR evaluates the generated output by measuring its semantic similarity to reference translations [1, 6], considering factors such as synonyms and morphological variations. Additionally, previous works [21] have used accuracy metrics including SP-Acc and NLI-Acc for evaluating the logical fidelity of the generated text by comparing it against the gold standard, along with Adv-Acc assesses robustness by evaluating the model's performance on adversarial examples designed to challenge its accuracy. While automatic evaluation metrics can provide a basic measure of performance, they have a limited capacity to understand a writer's perspective for several reasons. First, human-written captions in scientific papers are often of low quality, making the comparison between machine output and manual captions unreliable. Another problem is that automatic evaluations do not always align well with human judgments, revealing a gap between machine scoring and actual human comprehension. Finally, such comparisons focusing solely on the written artifact miss what was not explicitly stated including the author's intentions and considerations. To address these limitations, several studies have incorporated human evaluation alongside automatic metrics. However, these human evaluations often fail to adequately capture the writer's intent behind the captions [14, 15, 18, 19]. For example, annotators were asked to evaluate the matching degree, which reflects the extent to which the data in the generated summary matched the chart, and reasoning correctness, which assesses whether the generated summary accurately reflected the intended message [18]. Similarly, another study [19] conducted human evaluations with three respondents per chart, assessing informativeness, conciseness, coherence, and fluency, but also faced limitations related to the lack of consideration for the writer's perspective. These methods are generally confined to evaluating how faithfully the generated text describes the given figures, or how effectively it is understood by a reader, rather than the deeper communicative goals of the caption's author.

2.2 AI Writing Assistants

The development of AI writing assistants has been widely discussed to explore their potential in various writing contexts. These studies have investigated the application of AI in creative writing [23], legal writing [27], and medical practice [9], demonstrating the broad applicability of Human-AI collaboration in specialized writing tasks.

Focusing in the area of academic writing, researchers have examined AI's role in supporting various aspects of the scholarly process. For example, Sparks [11] explored how language models can generate sentence-level suggestions to aid science writing, while others have investigated AI support for literature review writing [2, 7]. These studies highlight the potential of AI to address specific challenges in academic writing, such as summarizing complex information and synthesizing existing literature. The more specialized application of AI in academic writing has also been studied through more previous works. MetaWriter [26] focused on assisting in the writing of academia meta-reviews, demonstrating how AI can support critical evaluation and synthesis in peer review processes. While FigurA11y [24] explored AI assistance in creating alt text for scientific figures, it's important to note the distinction between alt text and captions. Alt text primarily serves accessibility purposes for visually impaired readers, whereas captions cater to a broader audience and often contain more detailed interpretations or analyses of figures. This distinction emphasizes the need for specialized AI support in caption writing with broader communicative goals.

The application of AI writing assistants raises important questions about accuracy and trust in specialized academic tasks. Previous studies [3] have evaluated the efficacy and ethical implications of using AI in academic writing, highlighting both potential benefits and challenges of integration AI into academic workflows.

While many studies have explored AI in academic writing, many specific tasks within academic writing remain understudied in the context of AI assistance. As mentioned earlier, while some work has touched on related areas, such as generating descriptions for charts or figures, these studies often prioritize on the reader's perspective rather than the writer's experience. Our work aims to address this gap by specifically examining how researchers interact with and perceive AI-generated captions in scientific writing.

2.3 Human-AI Collaboration

Human-AI collaboration has gained significant attention in HCI research, especially as AI systems become more advanced and capable of integration into complex knowledge work. This trend is especially relevant in academic contexts, where AI tools are beginning to assist with various aspects of the research and writing process.

A recent systematic review by Shen et al. [20] proposes a framework for 'bidirectional human-AI alignment', which encompasses both the process of aligning AI systems with human values and the adaptation of humans to AI capabilities. This dual perspective is particularly relevant to our study, as we examine both how AI can be tailored to assist researchers in caption writing and how researchers adapt their practices when working with AI-generated captions. In the context of our work, we explore how AI systems can be designed to generate captions that meet the specific needs and expectations of scientific authors (‘Aligning AI to Humans’), and we investigate how researchers adapt their caption writing processes and evaluate the quality of AI-generated content (‘Aligning Humans to AI’). Furthermore, Lee et al. [16] explored the design space for intelligent writing assistants, highlighting the importance of considering both the AI’s capabilities and the user’s writing process. Their work emphasizes the interconnection between users and AI systems, providing a foundation for our in-depth investigation of researchers’ perspectives on caption-writing tasks for scientific publications. This approach emphasize the need for writing tools that not only generate content but also support the user’s cognitive processes and decision-making. These works provide a foundation for our investigation into researchers’ perspectives on AI-assisted caption writing for scientific publications.

Chapter 3

User Study

3.1 Study Design

To address our research questions, we conducted a user study with 18 participants, consisting of semi-structured interviews and a writing task. The study was conducted remotely via Zoom and lasted approximately 60 minutes. It was organized into three main phases: 1) Pre-task interview, 2) Writing task, and 3) Post-task interview.

3.2 Participant Recruitment and Demographics

Participants were recruited through a questionnaire assessing their academic writing experience. The questionnaire collected data on research areas, years of experience in academic writing, English language proficiency, and publication history. Participants also provided up to three recent published papers (Details in 3.3), which were used to produced AI-generated captions using three different configurations of the GPT-4o model:

1. GPT-4o (img+text) with 30-word limit (*30words*)

2. GPT-4o (text only) with unlimited length (*text_only*)

3. GPT-4o (img+text) with unlimited length (*unlimited*)

This approach was based on a previous study on caption generation using GPT-4V [12].

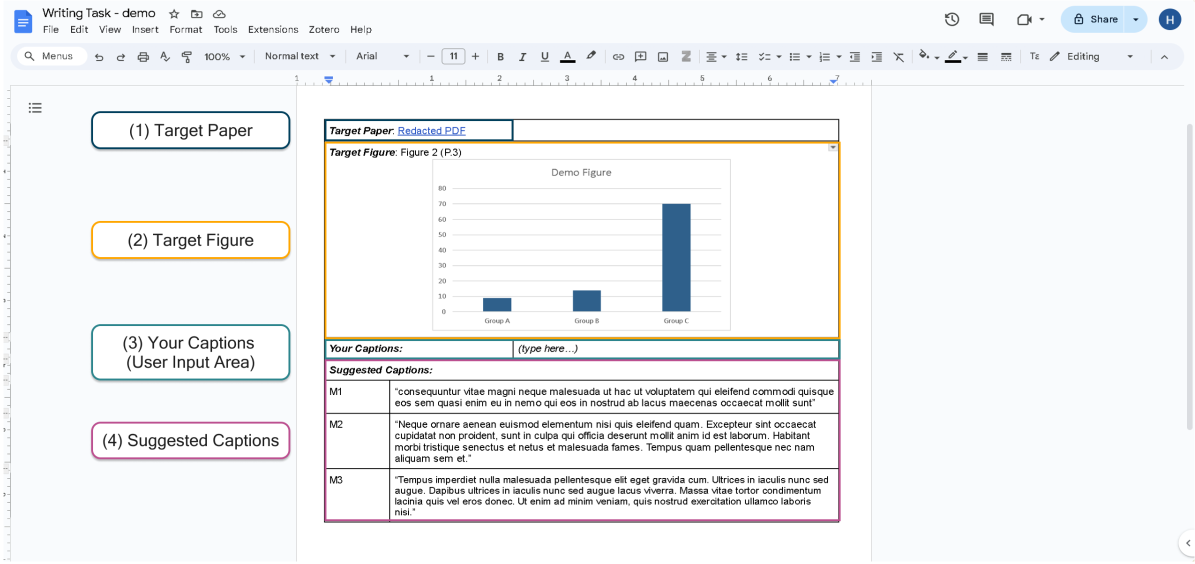
The 18 participants represented diverse research areas, with the majority in Computer Science/Informatics(28%) and Human-Computer Interaction(22%), other also includes

Artificial Intelligence/Robotics(17%), Energy and Minerals Engineering(6%), Mechanical Engineering(6%), Environmental Engineering(6%), Chemistry/Biochemistry(6%), Materials Science(6%), Cybersecurity(6%). Participants' age ranged from 22 to 44, with 78% between 26-29 years old. 72% (13) of participants reported that English is not their first language.

3.3 Task Design

We designed a writing task required participants to write captions for two figures from their recently published papers. 'Recent' was defined as paper published within the last 3 years but not less than 1 month ago. This timeframe was chosen to ensure participants were still familiar with the context of their work while reflecting realistic writing scenarios. By focusing on participants' own published work, we aimed to simulate an authentic writing environment where researchers possess the necessary domain knowledge and motivation for high-quality output. This approach allowed us to study how researchers interact with AI-generated captions in a context closely mirroring their actual writing practices, providing insights into the potential integration of AI assistance in scientific writing workflows.

We aimed to select diverse figure types to cover a range of caption writing requirements, focusing primarily on statistical and conceptual figures. While we recognize that scientific publications contain a much wider variety of figure types, we limited our selection to these two broad categories for practical reasons. This simplification allowed us to maintain a reasonable timeframe for the rewriting task and subsequent interview. We defined 'statistical' figures'= as those presenting quantitative data through graphs or charts, while 'conceptual' figures were those illustrating theoretical models or processes. Such distinction provided a basis for comparing caption writing approaches across different types of visual information. However, it is important to note that the actual selection



Target Paper: Redacted PDF

Target Figure: Figure 2 (P.3)

Your Captions:	*(type here…)*

Suggested Captions:

M1	"consequuntur vitae magni neque malesuada ut hac ut voluptatem qui eleifend commodi quisque eos sem quasi enim eu in nemo qui eos in nostrud ab lacus maecenas occaecat mollit sunt"
M2	"Neque ornare aenean euismod elementum nisi quis eleifend quam. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Habitant morbi tristique senectus et netus et malesuada fames. Tempus quam pellentesque nec nam aliquam sem et."
M3	"Tempus imperdiet nulla malesuada pellentesque elit eget gravida cum. Ultrices in iaculis nunc sed augue. Dapibus ultrices in iaculis nunc sed augue lacus viverra. Massa vitae tortor condimentum lacinia quis vel eros donec. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi."

Figure 3-1. User interface for figure caption writing task. The layout includes (from top to bottom): (1) 'Target Paper' - the hyperlink to the redacted PDF, (2) 'Target Figure' - displaying the figure image for the writing task, with the figure number and the page number in the redacted PDF, (3) 'Your Captions' - User input area for caption writing, and (4) 'Suggested Captions' - AI-generated caption suggestions from 3 different configuration(*unlimited*, *text_only*, *30words*), presented in randomized order for each caption item.

of figures was sometimes constrained by the nature of participants' research fields, and we could not always achieve an equal distribution between these two types for every participant.

The writing task was conducted in a prepared Google Doc (Fig. 3-**1**), ensuring a consistent writing environment across all participants. For each figure, participants were provided with the following to works on the caption writing:

1. Three AI-generated captions presented in randomized order

2. The corresponding PDF file with the original caption redacted

3.4 Procedure

After obtaining informed consent, the study proceeded as follows:

3.4.1 Pre-task interview

The pre-task interview served as an important foundation for our study, allowing us to frame the study and capture preliminary information. This semi-structured interview focused on several key areas:

1. Exploring current workflows: Participants described their usual process for writing figure captions, including any tools or resources they typically use, time and effort spent on caption writing, and how caption writing fits into their overall paper writing process.

2. Defining 'good' captions: We prompted participants to articulate their perspective on what constitutes a high-quality figure caption in scientific writing. This included discussions on essential elements, common pitfalls, and how caption quality might vary across different scientific disciplines.

3.4.2 Writing task

Participants engaged in the writing task using a think-aloud protocol [10]. While we estimated approximately 10 minutes per caption, no strict time limit was imposed to ensure a comfortable writing scenario. Upon completing each caption, participants:

1. Ranked the three AI-generated captions based on usefulness

2. Provided rationale for their ranking and comments on each AI-generated captions

3. Compared their resulting caption to the original (revealed after writing) by evaluating the statement 'My rewritten caption is better than the original caption' using a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree)

3.4.3 Post-task interview

Following the writing task, we conducted a semi-structured post-task interview. This interview aimed to:

1. Explore perception of AI-generated captions: We asked participants about their overall experience using AI-generated captions for caption writing. Questions explored their perceived benefits, challenges, and potential concerns regarding the use of AI in this context.

2. Compare with usual writing process: Participants were asked to compare this AI-assisted writing experience with their usual caption writing process, highlighting any differences in efficiency, quality, or cognitive load.

3. Gather future design insights: We solicited participants' input on potential future designs for caption-writing tools in scientific writing. This included discussions about desired features, integration with existing workflows, and ways to address any limitations they experienced during the task.

4. Collect general reflections: Participants were given the opportunity to share any additional thoughts or reflections on the experience that weren't covered by our specific questions.

3.5 Data Analysis Methods

Our study employed a mixed-methods approach, combining qualitative and quantitative analyses to provide a comprehensive understanding of participants' interactions with AI-generated captions, their writing processes, and their perspectives on using AI technology for caption writing in scientific publications.

3.5.1 Interaction Analysis

We conducted interaction analysis [13] to examine participants' writing behaviors and their engagement with AI-generated captions. We developed two coding schemes: one for the initial adoption behavior of AI-generated captions and another for the ongoing editing process. These coding schemes allowed us to capture how participants incorporated, modified, or rejected the AI-generated captions throughout their writing process.

The initial adaptation behavior coding focused on how participants began their writing task in relation to the AI-generated captions, ranging from completely ignoring the suggestions to copying them entirely(Table 3-**1**). The editing process coding captured the various strategies participants employed as they continued to work with the AI-generated captions, such as copying text, making corrections, adding their own information, and combining multiple suggestions(Table 3-**2**). This analysis complemented our qualitative interview data, providing a more comprehensive understanding of how participants interact with AI-generated captions in scientific writing.

Table 3-1. Coding Scheme for Initial Adoption Behavior in AI-Assisted Writing: Ideation and Drafting Stage

Code	**Description**
Init-Complete	Copying the entire AI suggestion verbatim("as is") into the draft
Init-Full	Copying or referring to a full sentence (defined by punctuation) from the AI suggestion to start working
Init-Part	Copying or referring to a partial sentence segment (e.g., phrase) from the AI suggestion
Init-Keyword	Copying or referring to a key term or single word from the AI suggestion to start working
Init-Ignore	Ignoring the AI suggestion and starting with manual typing and drafting

Table 3-2. Coding Scheme for Ongoing Editing Process Behavior in AI-Assisted Caption Writing

Code	Description
AI-Copy	Participant makes a new attempt to copy or refer to additional text from AI suggestions.
AI-Combine	Participant copies or refers to text from different AI suggestions (beyond the first suggestion).
AI-Adapt	Participant follows the AI suggestion but makes slight modifications, like rearranging text or deleting articles, while retaining the main idea.
AI-Delete	Participant removes redundant text from the AI suggestions (e.g., incorrect or unnecessary information).
AI-Correct	Participant attempts to correct the text based on their understanding or requirements.
Human-Add	Participant adds new information not generated by the AI.

3.5.2 Quantitative Data Analysis

We performed quantitative analyses on two key aspects of the study:

1. Ranking preference for AI-generated captions: We analyzed the frequency of each caption being ranked as most useful. Weighted ranking analysis is employed to consider the relative importance of each rank position. Weights were assigned to each rank (3 for 1st, 2 for 2nd, 1 for 3rd), and a weighted score was calculated for each AI-generated captions. These weighted scores were then converted to percentages to determine the final preference ranking.

2. Perceived improvement ratings: We calculated descriptive statistics for the 5-point Likert scale ratings comparing resulting captions to original ones.

3.5.3 Thematic Analysis

We employed thematic analysis [8] using a hybrid approach that combined deductive and inductive coding methods. As a starting point, we utilized the existing design space for AI-assistive writing developed by Lee et al. [16] as our initial coding framework. It provides a structured set of themes and codes highly relevant to our study of AI-generated captions for figure in scientific writing. It helps to address our research questions about researchers' interaction with and perceptions of AI-generated captions.

To ensure our analysis captured the unique aspects of our research context, we also investigated emerging themes and patterns in our data that might not have been fully captured by the existing design space. Our coding process involved reviewing the interview transcripts and video recordings, applying codes from the design space where appropriate, and developing new codes when necessary to capture novel insights specific to our study. This allows us to leverage existing theoretical frameworks while maintaining the flexibility to identify and explore unanticipated themes in our data.

Chapter 4

Findings: Quantitative analysis for Participants' Experience With Multiple AI-Generated Captions

Our finding revealed diverse patterns in the participants' interactions with AI-generated captions, their writing processes, and their perceptions of the resulting captions. We present our findings across four key areas: initial adoption behaviors, ongoing writing processes, perceived usefulness of AI-generated captions, and overall improvement ratings.

4.1 Initial Adoption Behaviors at Ideation and Drafting Stage

Analysis of participants' initial interactions with AI-generated captions revealed a tendency towards higher levels of adoption, with many using AI-generated captions as a starting point for their drafting process. Across 36 caption items from 18 participants, we observed the following distribution of initial behaviors(Fig. 4-**1**): 12 occurences in 'Init-Complete' (33.3%), 10 occurences in 'Init-Full'(27.8%), 6 occurences in 'Init-Ignore'(16.7%), 5 occurences in 'Init-Part'(13.9%), 3 occurences in 'Init-Keyword'(8.3%). Seven participants (38.9%) demonstrated consistent behavior across both caption tasks. The most common consistent behavior was'Init-Complete' (3 participants, 16.7%), followed by 'Init-Ignore'(2 participants, 11.1%), 'Init-Full'(1 participant, 5.6%) and 'Init-Part'(1 participant, 5.6%).

4.2 Ongoing Writing Process Behaviors

Our analysis of the ongoing writing process revealed various integration strategies employed by participants:

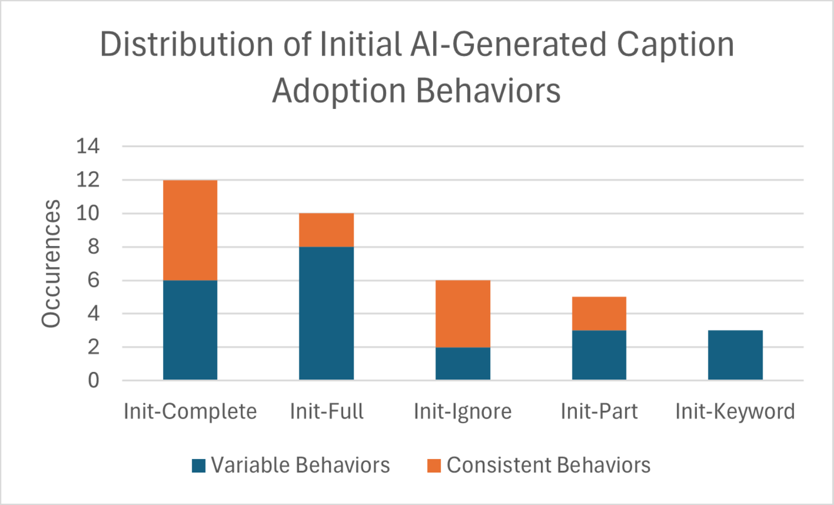


Figure 4-1. Distribution of Initial AI-Generated Caption Adaption Behaviors Among Participants (N=36 caption items)

- **Human Integration:** In 35 out of 36 caption items(97.2%), participants engaged in some degree of human integration with AI-generated captions. This included correcting suggestions (AI-Correct), deleting redundant or incorrect words (AI-Delete), and adding new information that is not included in the AI-generated captions (Human-Add). This high rate of human integration highlights the collaborative nature of the AI-assisted caption writing process, where researchers actively refine and supplement AI-generated content.

- **Direct Adoption:** In 1 out of 36 writing tasks (2.8%), a participant used an AI-generated caption verbatim without any modifications or additional writing process behaviors after the initial adoption.

- **Reversion to AI-generated Caption:** In 2 out of 36 caption items (5.6%), participants initially made integration efforts but ultimately reverted to using the AI-generated caption verbatim for the resulting caption. This behavior suggests a complex decision-making process where participants initially attempted to customize the AI output but eventually decided the original suggestion was sufficient.

- **Combination of Suggestions:** In 11 out of 36 caption items (30.6%), participants combined content from multiple AI-generated captions to form their resulting captions. This behavior, observed in 10 out of 18 participants (55.6%), suggests that many researchers found value in synthesizing ideas from various AI-generated captions rather than relying on a single AI-generated caption.

4.3 Perceived Usefulness of AI-Generated Captions

The result of the weighted ranking analysis (Table 4-**1**) shows that suggestion generated by '*unlimited*' is the most preferred, with a weighted percentage of 39.35%, while '*text_only*' and '*30words*' had very similar weighted percentages of 30.56% and 30.09% respectively, indicating a nearly equal preference for these two suggestions.

Table 4-1. Ranking Preference for AI-Generated Captions Suggestion from Different Configurations

Configuration	Rank	W%	W.Score	1st	2nd	3rd
unlimited	**1**	39.35%	85	18	13	5
text_only	**2**	30.56%	66	8	14	14
30words	**3**	30.09%	65	10	9	17

4.4 Perceived Improvement of Resulting Captions

Participants generally perceived improvements in their resulting captions compared to the original captions:

- 69% of items received ratings of 4 or 5, indicating substantial perceived improvement

- 14% (5 items) were rated as 3 ('Neutral')

- 17% received negative ratings: 11% (4 items) rated as 2, and 6% (2 items) rated as 1

The mean rating across all caption items was 3.83 (SD = 1.21), suggesting that participants generally found the resulting captions, assisted by AI-generated captions, to be better than their original captions.

4.5 Qualitative Insights on AI-Generated Captions

In addition to quantitative data, participants provided verbal comments on the AI-generated captions and their resulting captions. These comments offer valuable insights into participants' rationales for their ratings and their perceptions of the AI-generated captions. These qualitative data were incorporated into our thematic analysis, which will be discussed in the following section.

Chapter 5

Findings: Design Considerations for AI-Assisted Scientific Caption-Writing

Our thematic analysis revealed several key insights into researchers' experiences with AI-generated captions for scientific publications. We organize our findings around four key considerations: 'Task Characteristics', 'User Capabilities and Perception', 'Ecosystem Constraints', and 'Desired Interaction Features'.

5.1 Task Characteristics

5.1.1 Purpose

5.1.1.1 Diverse Purposes

Captions served multiple functions, including analytical interpretation, narrative storytelling, and ensuring accessibility. For example, P17 emphasized the analytical purpose: "I think in order to effectively communicate the idea of the paper, that interpretation is necessary." In contrast, P12 highlighted the narrative aspect: "It really helps the researchers to get a clear flow in terms of how we tell this whole story." Some participants aimed to address multiple purposes simultaneously. P11 exemplified this multifaceted approach:

> I would try to give more context, have larger captions. (**Expository**) [...] You can see in the picture we have like some variables, right? So you might be tempted to say like we need to also describe these. (**Descriptive**) [...] To some extend, you want to sell your paper, like a good scientist as a salesman that you are selling your paper. (**Persuasive**)

This diversity in purpose highlights the complexity of caption writing and suggests that

AI writing assistants need to support multiple, sometimes conflicting, goals.

5.1.2 Specificity

5.1.2.1 Adapting to Figure Location

Participants reported adjusting their caption writing approach based on the figure's position within the paper. As P17 noted:

> It depends on which section this figure is going to be. If it's a teaser image or a figure in the system section, then maybe the caption could be more about describing what's on the image. Whereas if you have figures on the latter part of the finding section or the discussion section, maybe that could lead to more of the interpretation side.

This context sensitivity suggests that AI tools should consider the figure's location when generating caption suggestions, potentially offering different styles or levels of detail based on the section.

5.1.2.2 Balancing Caption and Main Text Content

Participants demonstrated varying approaches to integrating caption content with the main text:

1. **Avoiding Redundancy:** Some participants, like P11, preferred to minimize repetition between captions and the main text: "The figure that you are already did mentioned in the text of our paper, there is no need, you know, to repeat the information over and over again in the caption."

2. **Complementary Information:** Others, such as P17, saw value in some overlap to enhance reader understanding: "I have the figures there because I think it nicely complements what I have in the body text. There might be some overlap but I think it helps better describe or help the reader better understand the body text."

3. **Self-Contained Captions:** Some participants aimed for comprehensive, standalone captions. P19 explained: "Because not everyone understands the trend described in the figure and the claim I want to make. So sometimes I will make some connections to make sure that the figure and the captions together are self-explanatory. So people can get a general idea without referring to the main text."

These diverse approaches suggest that AI writing assistants should offer flexibility in generating captions that can either complement or stand independently from the main text, based on the author's preference and the paper's structure.

5.1.2.3 Adapting to Figure Types

Participants reported varying caption writing strategies based on figure types, extending beyond our initial categorization of statistical and conceptual figures:

1. **Experimental Results:** For figures presenting experimental data, participants like P18 preferred detailed descriptions: "The figure content is all experimental results. So currently, I think if the caption can describe the experiment to this extent, I will be very satisfied."

2. **Abstract Figures:** P18 also noted the challenge with more abstract figures:

   > But for other like for HCI if they have some figures that's like the pipeline or the design or the interface, I think if they try to make a caption, it will be more difficulty because the thing itself is not very objective, it needs people's understand, but it's hard to describe that in one sentence in the caption.

3. **Teaser Figures:** For introductory figures, participants like P08 emphasized high-level descriptions: "So I think this figure is supposed to be teaser, right? So they should give us like a very high level perspective."

4. **Complex, Multi-Part Figures:** Some participants, like P19, dealt with highly complex figures requiring detailed, step-by-step caption writing processes:

> I would do some drafting in my notebook or just on a piece of paper and describe a general idea of what I would like to convey through the figure. Then I will move on to my actual data. And then convert the data into figures or schematic illustrations after which I will have the first draft of my figure legend or in your words, it's the captions and then I will start writing the main text. After having the main text at hand, I will make some changes to my captions to make it more coherent with the text in my main text.

These findings highlight the need for AI writing assistants to recognize and adapt to various figure types, offering tailored suggestions that align with the specific requirements of each figure category.

5.1.3 Audience

Our analysis revealed that participants' approach to caption writing was significantly influenced by their perception of the intended audience, reflecting diverse needs and expectations of different readers in scientific publications.

5.1.3.1 Diverse Group of Readers

Participants mentioned a wide range of specific Participants identified a wide range of potential readers for their captions, demonstrating the complex ecosystem of scientific communication. Reviewers were the most frequently mentioned audience (P02, P09, P12, P16, P20), highlighting the critical role of captions in the peer review process. Other specific audiences included advisors (P12, P15), co-authors (P12), and peer researchers (P11). Some participants (P06, P07, P10, P13, P18, P19) used the general term 'reader', suggesting a broader, less specialized audience.

5.1.3.2 Adapting Detail Level

The diversity of potential readers led participants to consider flexibility in the level of detail provided in captions. This adaptability was seen as crucial for catering to audiences with varying levels of expertise. P19 articulated this challenge:

> I am not sure how I can improve the caption writing for someone new to the field just to how can I make it accessible to a general audience? If someone's totally new to this topic or this field, it might be difficult for them to really understand what I have in the figure caption.

This reflection highlights the tension between providing sufficient detail for novices while maintaining relevance for expert readers, a common challenge in scientific communication.

5.1.3.3 Anticipating Reader Behavior

Interestingly, participants often based their caption writing approach on assumptions about reader behavior. These assumptions guided decisions about content, structure, and detail level. For example, P20 anticipated that readers might only briefly engage with captions: "People might not read the whole thing. So maybe if they just glance at the caption." Other participants made assumptions about the sequence in which readers would engage with different elements of the paper. P17 noted: "When readers come to the paper, I think the first thing they look at is usually the figure and the caption." Similarly, P19 considered both the order of engagement and the potential for reader fatigue with lengthy captions: "Most of the time they will go through the figure first and as they go through the figures, they will read the captions. But if the captions are too long, then people are less likely to go through captions" These assumptions about reader behavior influenced participants' strategies for caption writing, balancing the need for comprehensive information with the reality of how readers might interact with the academic paper.

5.2 User Capabilities and Perceptions

5.2.1 User capabilities

Our analysis revealed several key aspects of users' capabilities and challenges in caption writing, which can be categorized into efficiency concerns, varying confidence levels, and cognitive challenges.

5.2.1.1 Time and Effort Constraints

Participants reported significant time constraints and difficulties in structuring captions efficiently. This challenge was exemplified by P17, who stated: "I don't allocate a lot of time of caption writing."

This sentiment reflects a common struggle in balancing the demands of caption writing with other aspects of scientific paper preparation.

5.2.1.2 Confidence Disparities in Language and Domain Knowledge

A notable contrast emerged in participants' confidence levels between language proficiency and technical knowledge:

1. **Language Proficiency Concerns:** Many participants, particularly non-native English speakers, expressed lower confidence in writing captions due to concerns about English proficiency. P15 articulated this challenge: "I'm not good at writing the English captions right now because I'm not good at English."

2. **Technical Knowledge Confidence:** In contrast, participants generally reported high confidence in their technical and domain knowledge. Some, like P01, even expressed greater trust in their own expertise compared to AI capabilities: "I feel like I know best. And so I would do what I think is best. I feel like I might know better than AI on how to caption the figure on my paper."

This disparity suggests a potential area where AI assistance could be particularly beneficial, complementing users' strong domain knowledge with language support.

5.2.1.3 Cognitive Challenges

Two primary cognitive challenges emerged from our analysis:

1. **Attention Allocation:** Some participants admitted to prioritizing the main text over captions, potentially undervaluing the importance of figure captions. P20 noted: "This is an under focused area because it kind of like, doesn't matter but it does. They're probably gonna focus more on the content of the paper."

2. **Information Overload:** Participants often felt overwhelmed by the amount of information in figures, leading to difficulties in writing concise yet comprehensive captions. P15 described this challenge: "I cannot explain all of the details inside the figure. So I need to omit some details. But sometimes it is very difficulty to distinguish whether the part is omitted or not."

These cognitive challenges highlight the complex decision-making process involved in caption writing and the potential for AI tools to assist in information prioritization and concise summarization.

5.2.2 Relationship to System

Our analysis revealed complex dynamics in participants' relationships with the AI system, particularly concerning integrity and trust. These factors significantly influenced participants' perceptions and use of AI-generated captions.

5.2.2.1 Prioritizing Integrity in Academic Writing

Participants consistently emphasized the critical importance of accuracy in AI-generated captions, viewing it as fundamental to maintaining the credibility of their academic work. This sentiment was succinctly captured by P11: “I think in academic captioning like in papers being right and correct is the bottom line like that is necessary.” This strong emphasis on integrity highlights the high stakes involved in academic writing and the potential risks associated with incorporating AI-generated content without careful verification.

5.2.2.2 Mixed Trust in AI Capabilities

Participants exhibited varied levels of trust in AI-generated captions, revealing a mixed perspective that acknowledged both the strengths and limitations of AI systems. This variability in trust manifested in three key areas:

1. **Recognizing AI’s Summarization Strengths:** Some participants appreciated AI’s ability to distill complex information, as noted by P20: “I think it helps me, this kind of provided a summary in a way, you know, so that I can kind of, I could see concretely what it is that I’m working with in another, in another way.” This suggests potential value in using AI tools to help researchers create concise, informative captions.

2. **Skepticism About Deep Research Understanding:** Despite acknowledging AI’s summarization capabilities, participants expressed doubts about its ability to grasp varying research intentions. P19 observed: “I would say they’re good at summarizing but not necessarily deriving the claim or to make connection between the summary and the author’s claim.” This highlights a perceived gap between AI’s information processing abilities and its capacity to understand the deeper implications of research findings.

3. **Concerns About Increased Cognitive Load:** Some participants worried that relying on AI-generated captions might actually increase their workload due to the need for careful verification. P18 explained:

> I feel like the fear of having and including incorrect information might create an unnecessary stress within my writing process. So I feel like if I start from beginning with my own draft, it might be less stressful of the process and I can trust the model way more instead of having some pre-made suggestions from the model that I need to double check

This perspective suggests that for some researchers, the perceived benefits of AI assistance may be outweighed by the additional mental effort required to ensure accuracy.

5.2.3 System Output Preference

Our analysis of participants' interactions with AI-generated captions revealed specific preferences regarding system outputs. These preferences primarily centered around textual coherence and diversity, highlighting the complex connection between AI-generated captions and researchers' expectations in scientific writing.

5.2.3.1 Contextual Relevance and Domain-Specific Language.

Participants expressed a strong preference for suggestions that were both grammatically correct and contextually relevant. However, they noted concerns about the contextual understanding demonstrated in the existing suggestions, particularly regarding word choice. P12 articulated this concern: "There might be some differences in terms of the choice of words. For example, we would probably not use the word 'decline' here when writing research results." This observation highlights the critical importance of domain-specific language understanding in AI writing assistants. Inappropriate word choices can significantly detract from the credibility of the caption, potentially undermining the researcher's work.

5.2.3.2 Value of Multiple Suggestions.

The provision of multiple AI-generated captions was generally well-received by participants. They appreciated the ability to compare and combine different captions to create more comprehensive and refined final versions. P09 articulated this benefit: "I sometimes found that the mix of two would be better choices because I think sometimes one generated caption will miss some point." This preference for diverse suggestions highlights the value of generating multiple options in AI writing assistance tools. It allows users to leverage the strengths of different outputs, potentially leading to more comprehensive and accurate captions. This approach also aligns with the complex nature of scientific writing, where multiple perspectives or ways of describing a figure can be valuable.

5.2.3.3 Recurring Information as Reliability Indicator.

An unexpected finding emerged regarding the impact of consistency across multiple suggestions. Some participants reported that seeing the same piece of information across multiple AI-generated captions increased their confidence in including that information in their final writing. P01 explained: "So if there were certain information inside of all of this or the majority of the suggested captions, then I assumed I needed to include that information in my caption as well." This observation suggests that consistency across multiple AI-generated captions may serve as an indicator for importance or relevance, influencing users' decision-making processes in caption writing. This finding has interesting implications for the design of AI writing assistants, suggesting that providing information about the frequency or consistency of certain elements across multiple suggestions could be a valuable feature.

5.3 Ecosystem Constraints

Our analysis revealed that caption writing in scientific publications is significantly influenced by the broader ecosystem of academic publishing, particularly the norms and rules imposed by different publication venues. These considerations play a crucial role in shaping researchers' approaches to caption writing.

5.3.1 Norms and Rules

5.3.1.1 Venue-Specific Guidelines and Constraints

Participants reported that their caption writing strategies are often dictated by the specific guidelines of their target publication venues, such as conferences and journals. These guidelines can vary significantly and impact caption content and structure in several ways:

1. **Page and Word Limits:** Many venues impose strict page or word count limits on submissions. As P06 noted:

   > In many conferences, there is explicit page that I can't go above this number of pages. So at the time I need to make it short, I don't have any option even in journals, like as far as you can remember, International Journal of Medical of else here. They have a number or limitation but maybe figure caption is not included there.

   This constraint can lead researchers to make strategic decisions about caption length and content. Interestingly, some venues exclude captions from word count limits, potentially encouraging more detailed captions as a way to conserve word count in the main text.

2. **Field-Specific Norms:** Participants also highlighted the existence of unwritten norms within their specific research fields regarding caption content. P16 observed: "Such description should not be included in the caption and I have never seen

this style of writing in our field." This suggests that researchers must navigate not only explicit guidelines but also implicit expectations within their academic communities.

3. **Visual Layout Considerations:** The visual presentation of the paper emerged as another factor influencing caption writing. Participants reported adjusting caption lengths to optimize the overall layout and formatting of their submissions. P10 explained:

   > So actually if you want to like to shorten that I can remove the like the numbers here, just use the model one, just remove that. These are the options if I consider the limit. You can see that it's fully it even the reference is not at the end of the page. So I try to fit the paper, the pages I see.

   This highlights how caption writing is not just about content, but also about managing the visual appeal and space efficiency of the entire paper.

5.4 Desired Interaction Features

5.4.1 UI - Layout

Our analysis revealed strong participant preferences for seamlessly integrated user interfaces that support in-situ interactions between writing processes and AI-generated outputs. These preferences centered around two main aspects: integration with existing writing environments and support for the broader research workflow.

5.4.1.1 Integration with Writing Environments

Participants expressed a desire for AI caption tools to be directly integrated into their existing writing platforms. This preference was exemplified by P5, who suggested:

> I would like to integrate this into Overleaf because we use Overleaf to write research paper, right? It's like a Copilot. Integrate directly to the Overleaf

> and, every time you put in a figure, the tool will generate a caption for you first.

This sentiment reflects a desire for seamless workflow integration, where AI assistance becomes an organic part of the writing process rather than a separate tool.

5.4.1.2 Support for Research Workflow

Some participants envisioned AI caption tools extending beyond writing to support the broader research process, including data analysis and figure generation. P10 proposed: “Maybe another way is to use as a Visual Studio Code extension. Like because I do analyze using Python that I feel that kind of tool integrating with the existing tool that you’re familiar with using.” This suggestion highlights the potential for AI writing assistants to bridge the gap between data analysis and manuscript preparation, offering a more holistic approach to research documentation.

5.4.2 UI - Visual Differentiation

Our analysis revealed a strong preference among participants for clear visual differentiation of various elements in the caption writing interface. This desire for visual clarity centered around three key aspects: distinguishing user input from AI-generated content, tracking information sources, and managing multiple AI-generated captions.

5.4.2.1 Distinguishing User and AI Content

Participants emphasized the importance of clearly delineating between their original input and AI-generated content. This distinction was seen as crucial for maintaining authorial control and understanding the contribution of AI to the writing process. As P18 noted: “So that I know which part was generated and which part was my original content.” This preference highlights the need for AI writing assistants to provide clear visual cues that allow users to easily identify and differentiate between human-authored and AI-generated text.

5.4.2.2 Source Tracking for Explainability

Participants expressed a desire for features that would allow them to track the sources of information used in generating AI-generated captions. This explainability feature was seen as valuable for understanding the rationale behind generated content and assessing its relevance and accuracy. P18 suggested: "Maybe it can highlight something like this to show that this part was originally or it's exactly a copy paste from your text or your material into caption, something like that maybe might be useful." This insight underscores the importance of transparency in AI-assisted writing tools, particularly in academic contexts where source attribution is critical.

5.4.2.3 Managing Multiple AI-Generated Captions

In cases where multiple AI-generated captions were integrated, participants wanted the ability to track the origin of different parts of the final caption. This feature was seen as essential for effective management and refinement of captions. P15 proposed: "Maybe it can highlight something like this to show that this part was originally or it's exactly a copy paste from your text or your material into caption, something like that maybe might be useful." This suggestion highlights the need for AI writing assistants to support complex writing processes where users may combine and refine multiple AI suggestions.

5.4.3 User - Integrating System Output

Our analysis revealed participants' preferences for integrating AI-generated content into their writing process, highlighting the need for flexible and user-centric design in AI writing assistants.

5.4.3.1 Customizable Template Structures

Participants expressed a desire for AI systems to provide adaptable sentence structures that could be easily customized with context-specific information. This preference for a template-based approach was articulated by P17: “Maybe if the tool could give me like a format and I could just fill in the blanks ... then I would be able to just fill them in.” This suggestion indicates a potential for AI writing assistants to offer scaffolding for caption writing, allowing researchers to maintain control over the content while benefiting from AI-generated structural guidance.

5.4.4 System - User Data Access

Our analysis revealed participants’ preferences for providing diverse and relevant data to the AI system to enhance caption generation. These preferences centered around four key areas: research intentions, contextual materials, reference styles, and domain knowledge.

5.4.4.1 Communicating Research Intentions

Participants expressed a desire for the system to understand their research intentions, akin to how they frame research questions. They suggested various methods to achieve this, including prompts and additional documentation. P12 highlighted the potential of supplementary information:

> I can see like the AI isn’t able to put those here because uh those are not included in the main text or something. But like AI if you like put the supplementary information which like, we have like descriptions of these models and stuff. So that, I guess that that might be helpful.

This suggests that AI writing assistants could benefit from access to a broader range of research materials beyond the main text.

5.4.4.2 Contextual Highlighting for AI Focus

Participants emphasized the importance of relevant materials from the same paper for generating accurate and detailed captions. They suggested features such as highlighting specific areas of figures to guide the AI's focus. P15 proposed:

> can just select the part of the image that I want to add in the caption. Like first if I select all those image, whole image, then they can make the caption for the whole image. OK. And then if I select the part A in the image, then they can make the caption for part A.

Additionally, participants stressed the need for real-time updates to the main body text to ensure accurate generation. P20 noted: "I wouldn't want it to look at my old draft because, you know, that's outdated" These insights suggest that AI writing assistants should be designed with features that allow for dynamic input and real-time updates.

5.4.4.3 Reference Styles

Some participants suggested incorporating exemplar papers or style guides to help the AI learn preferred caption styles. P09 proposed: "Do you think we can like input other's paper with caption? [...] if I felt one paper has a very good caption and can it learn the style of this kind of caption writing." This indicates a desire for AI systems that can adapt to specific stylistic preferences or disciplinary norms.

5.4.4.4 Domain Knowledge Integration

Participants highlighted the importance of integrating domain-specific knowledge to improve the accuracy and relevance of AI-generated captions. P19 suggested: "I think it would be ideal if the authors or users can also feed the model with some Wikipedia page or just encyclopedia, it does not have to be super specific, just the basic knowledge of the field." This emphasis on domain knowledge integration reflects the need for AI writing assistants that can understand and incorporate field-specific context.

5.4.5 System - Output Type

While our study primarily focused on AI-generated captions for direct incorporation into papers, participants expressed a desire for a broader range of system outputs to support their caption writing process. These additional output types fall into three main categories: analytical feedback, explanatory content, and consistency management.

5.4.5.1 Analytical Feedback on Caption Quality

Participants suggested that the system could provide evaluative feedback on their captions. P17 proposed: "Maybe like something that evaluates my caption if they look good enough. But I think that might be challenging because I don't know what would be the metrics to evaluate those questions" This suggestion highlights a desire for objective assessment of caption quality, while also acknowledging the complexity of defining appropriate evaluation metrics for scientific writing.

5.4.5.2 Explanatory Content for AI Suggestions

Participants expressed interest in understanding the rationale behind AI-generated captions. P09 articulated this need:

> not only making suggestions, it also tell me why it is making this modification like I input the original caption and it says 'I don't think it's good because it is too short. I suggest that you adding another sentence describing the results.

This desire for explanatory content reflects a need for transparency in AI decision-making, potentially enhancing user trust and facilitating more informed use of AI suggestions.

5.4.5.3 Automated Consistency Management

Participants highlighted the challenge of maintaining consistency across various elements of their papers, including figures, captions, and main text. P17 described this issue:

> I had multiple versions of my teaser image just because I had to change like the 'summary bar' used to be like 'preference space', but then become 'summary bar'. So if those texts are updated, I need to go back to caption and recheck if like the texts are properly inserted. And if that could be done automatically, I think I would definitely use it

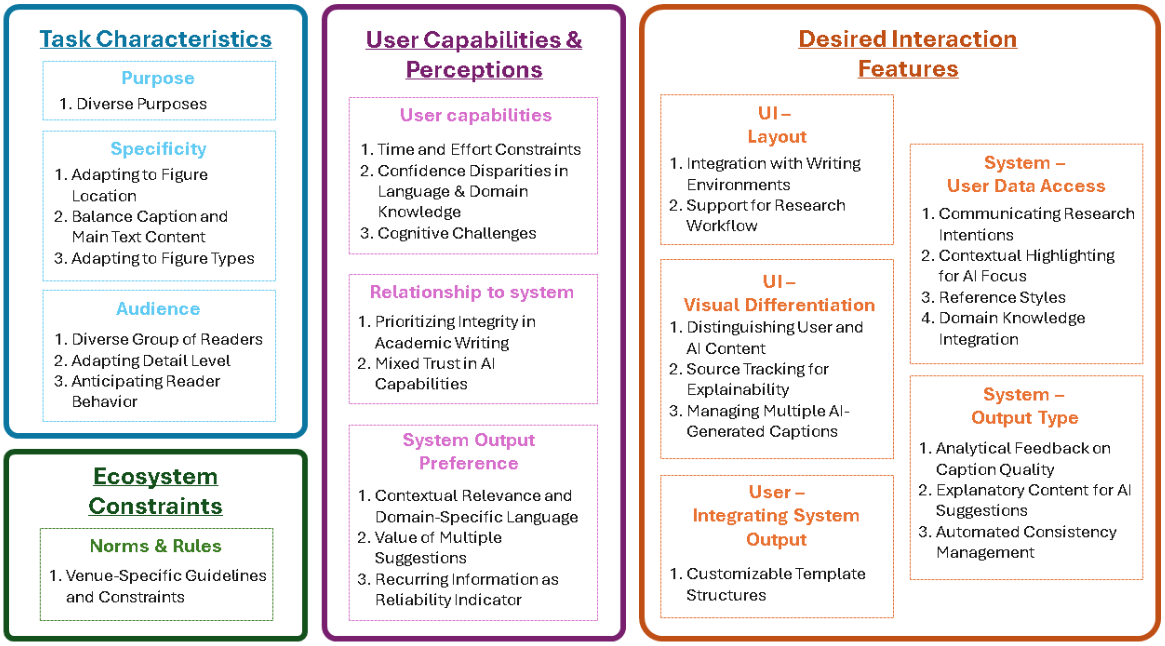

Task Characteristics
Purpose
1. Diverse Purposes
Specificity
1. Adapting to Figure Location
2. Balance Caption and Main Text Content
3. Adapting to Figure Types
Audience
1. Diverse Group of Readers
2. Adapting Detail Level
3. Anticipating Reader Behavior
Ecosystem Constraints
Norms & Rules
1. Venue-Specific Guidelines and Constraints
User Capabilities & Perceptions
User capabilities
1. Time and Effort Constraints
2. Confidence Disparities in Language & Domain Knowledge
3. Cognitive Challenges
Relationship to system
1. Prioritizing Integrity in Academic Writing
2. Mixed Trust in AI Capabilities
System Output Preference
1. Contextual Relevance and Domain-Specific Language
2. Value of Multiple Suggestions
3. Recurring Information as Reliability Indicator
Desired Interaction Features
UI – Layout
1. Integration with Writing Environments
2. Support for Research Workflow
UI – Visual Differentiation
1. Distinguishing User and AI Content
2. Source Tracking for Explainability
3. Managing Multiple AI-Generated Captions
User – Integrating System Output
1. Customizable Template Structures
System – User Data Access
1. Communicating Research Intentions
2. Contextual Highlighting for AI Focus
3. Reference Styles
4. Domain Knowledge Integration
System – Output Type
1. Analytical Feedback on Caption Quality
2. Explanatory Content for AI Suggestions
3. Automated Consistency Management

Figure 5-1. Framework of Design Considerations for AI-Assisted Scientific Caption-Writing: Adapting existing design space [16] to qualitative insights on caption-writing tasks, encompassing four main themes: (1) **Task Characteristics**, addressing the purpose, specificity, and audience of captions; (2) **User Capabilities & Perceptions**, covering user skills, relationship to the AI system, and output preferences; (3) **Ecosystem Constraints**, focusing on venue-specific guidelines; and (4) **Desired Interaction Features**, detailing UI requirements, data access needs, and system output types. This framework aims to guide the development of more effective and user-centered AI-assisted caption writing tools for scientific publications.

This suggests a need for AI systems that can track and manage changes across different components of a scientific paper, ensuring consistency and reducing manual cross-checking.

5.5 Summary

Our analysis of the design space for caption writing in scientific publications leveraged existing frameworks to derive thematic insights based on our observations and interviews (Fig. 5-**1**). This analysis revealed numerous in-depth considerations and feedback from participants across various themes. These insights are crucial for understanding the challenges in caption writing for scientific publications and should be considered when designing intelligent caption writing tools to help researchers create better captions.

Chapter 6

Discussion

Our study reveals several key insights into the design of AI-assisted caption writing tools for scientific publications. We organize our discussion around three main themes: (1) Holistic Approach to Caption Writing, (2) Adaptability for Diverse Fields and Readers, and (3) Human-Centered Design with Feedback and Input. We also address the limitations of our study and suggest directions for future research.

6.1 Holistic Approach to Caption Writing

Our findings emphasize the interconnected nature of caption writing within the broader context of scientific paper preparation. Participants consistently expressed a preference for AI-generated captions that integrated both figure content and body text, as evidenced by the higher ranking of the '*unlimited*' configuration in our quantitative analysis. This preference aligns with participants' emphasis on maintaining coherence between captions, main text, and figures.

These insights suggest that future AI writing assistants should adopt a more comprehensive approach, considering not just the figure in isolation, but also its relationship to the surrounding text and overall paper structure. Such tools could potentially:

1. Analyze the entire paper draft to ensure consistency in terminology and style between captions and main text.

2. Suggest ways to reference figures in the main text that complement the caption content.

3. Offer options for different levels of detail in captions based on the figure's position and role in the paper (e.g., more conceptual for introductory figures, more detailed for results).

6.2 Adaptability for Diverse Fields and Readers

The diversity of research fields and potential readers emerged as a crucial factor in caption writing. Despite our relatively small sample size, we observed significant variations in caption writing approaches across disciplines and target audiences. This diversity underscores the need for highly adaptable AI writing tools. Future AI writing assistants should consider:

1. Discipline-specific models trained on field-relevant corpora to better capture the distinctive features of different research areas.

2. User-configurable settings to tailor output for different audiences (e.g., reviewers, general readers, experts in the field).

3. Flexibility in caption length and detail to accommodate varying publication venue requirements and reader expectations.

Moreover, the tool should be capable of adapting to different figure types and their specific captioning needs, as highlighted by our participants' varied approaches to different figure types.

6.3 Human-Centered Design with Feedback and Input

While our quantitative analysis showed generally positive ratings for AI-generated captions, qualitative feedback revealed important gaps between AI suggestions and researchers' intentions. This highlights the critical need for human input and feedback in the caption writing process. Design considerations for future tools should include:

1. Interactive interfaces that allow researchers to highlight specific parts of figures or text for focused AI suggestions.

2. Mechanisms for researchers to input their research intentions, hypotheses, or key messages to guide AI caption generation.

3. Customization options for preferred writing styles, terminology, or level of detail.

4. Explainable AI features that provide rationales for suggestions, allowing researchers to better understand and trust the AI's output.

5. Iterative feedback loops where researchers can refine AI suggestions, helping the system learn and improve over time.

6.4 Limitations and Future Work

Our study setup involved participants rewriting existing captions from their own academic papers rather than writing new captions from scratch. This approach allowed us to simulate interactions with AI-generated captions in a controlled environment, focusing on the specific needs of academic writing. However, it may not fully replicate the complexities and dynamics of real-world scientific writing scenarios. For example, researchers might not have had the same level of engagement or urgency for caption writing as they would in a fresh, ongoing research project. Furthermore, contextual information from past research papers might have been less salient compared to ongoing projects, potentially affecting how researchers interacted with AI suggestions.

Another limitation of our study lies in the nature of the AI-generated caption presentation. In our experimental setup, we provided participants with multiple AI suggestions without an interactive prompting process. While this approach allowed us to investigate how various AI suggestions affect the writer's perspective, it may not fully align with real-world processes where researchers might engage in iterative prompting, similar to interactions with conversational AI like ChatGPT. Our focus on the ’writing process’ rather than the ’prompting process’ was intentional, as it allowed us to examine how writers incorporate and adapt AI suggestions into their work. However, this methodology doesn’t capture the potential benefits or challenges of an interactive AI writing assistant. Future research could explore a more dynamic interaction between researchers and AI tools, perhaps incorporating a dialogue-based system that allows for refinement and clarification of caption suggestions. This would provide insights into how researchers might leverage AI assistance in a more collaborative and iterative manner, potentially leading to the development of more sophisticated and user-friendly AI writing tools for scientific publications.

Despite these limitations, our study provided valuable insights into the design space for AI-generated captions in academic writing and laid a foundation for future research. To address these limitations, we can design future studies to more accurately capture empirical findings from real-world academic writing tasks and interactions with AI-generated content. This may include prototype development and testing of AI writing tools specifically tailored for researchers, incorporating the design considerations identified in this study. Such field studies in more realistic academic environments would provide a deeper understanding of how researchers interact with AI-assisted caption writing tools in their actual scientific writing processes.

Chapter 7

Conclusion

This study explored the potential of AI-assisted caption writing for scientific publications by examining researchers' interactions with multiple AI-generated captions during a writing task. Through semi-structured interviews and analysis of the caption writing process, we mapped the design space for AI-assisted caption writing in scientific contexts.

Our findings highlight key considerations for developing AI writing assistants tailored to scientific caption writing, including the need for contextual adaptability, user-centered design, and seamless integration with existing workflows. These insights aim to support researchers in producing higher-quality captions with reduced effort, ultimately enhancing the communication of research findings to both specialized and general audiences.

This study lays the foundation for future investigations and scholarly discourse in the domain, contributing to the iterative development of AI-driven systems that can enhance scientific communication processes. As AI continues to evolve, there is significant potential to create writing assistants that not only assist with mechanical aspects but also improve the overall quality and accessibility of scientific publications. Future work should build upon these insights to develop AI writing tools that are more attuned to the specific needs and challenges of scientific caption writing, potentially transforming how researchers disseminate their knowledge to diverse audiences.

Bibliography

[1] Vibhav Agarwal, Sourav Ghosh, Harichandana BSS, Himanshu Arora, and Barath Raj Kandur Raja. Tricy: Trigger-guided data-to-text generation with intent aware attention-copy. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 32:1173–1184, jan 2024.

[2] Ömer Aydın and Enis Karaarslan. Openai chatgpt generated literature review: Digital twin in healthcare. *Aydın, Ö., Karaarslan, E.(2022). OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. In Ö. Aydın (Ed.), Emerging Computer Technologies*, 2, 2022.

[3] Oğuz'Oz' Buruk. Academic writing with gpt-3.5 (chatgpt): reflections on practices, efficacy and transparency. In *Proceedings of the 26th International Academic Mindtrek Conference*, pages 144–153, 2023.

[4] Juan Cao, Junpeng Gong, and Pengzhou Zhang. Open-domain table-to-text generation based on seq2seq. In *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence*, ACAI '18, New York, NY, USA, 2018. Association for Computing Machinery.

[5] Juan Cao, Junpeng Gong, and Pengzhou Zhang. Two-level model for table-to-text generation. In *Proceedings of the 2019 International Symposium on Signal Processing Systems*, SSPS '19, page 121–124, New York, NY, USA, 2019. Association for Computing Machinery.

[6] Charles Chen, Ruiyi Zhang, Sungchul Kim, Scott Cohen, Tong Yu, Ryan Rossi, and Razvan Bunescu. Neural caption generation over figures. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '19 Adjunct, page 482–485, New York, NY, USA, 2019. Association for Computing Machinery.

[7] Kiroong Choe, Seokhyeon Park, Seokweon Jung, Hyeok Kim, Ji Won Yang, Hwajung Hong, and Jinwook Seo. Supporting novice researchers to write literature review using language models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–9, 2024.

[8] Victoria Clarke and Virginia Braun. Thematic analysis. *The journal of positive psychology*, 12(3):297–298, 2017.

[9] Carla Parra Escartín, Sharon O'Brien, Marie-Josée Goulet, and Michel Simard. Machine translation as an academic writing aid for medical practitioners. In *Proceedings of Machine Translation Summit XVI: Research Track*, pages 254–267, 2017.

[10] Marsha E Fonteyn, Benjamin Kuipers, and Susan J Grobe. A description of think aloud method and protocol analysis. *Qualitative health research*, 3(4):430–441, 1993.

[11] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. Sparks: Inspiration for science writing using language models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, pages 1002–1019, 2022.

[12] Ting-Yao Hsu, Chieh-Yang Huang, Shih-Hong Huang, Ryan Rossi, Sungchul Kim, Tong Yu, C Lee Giles, and Ting-Hao Kenneth Huang. Scicapenter: Supporting caption composition for scientific figures with machine-generated captions and ratings. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–9, 2024.

[13] Brigitte Jordan and Austin Henderson. Interaction analysis: Foundations and practice. *The journal of the learning sciences*, 4(1):39–103, 1995.

[14] Shankar Kantharaj, Rixie Tiffany Leong, Xiang Lin, Ahmed Masry, Megh Thakkar, Enamul Hoque, and Shafiq Joty. Chart-to-text: A large-scale benchmark for chart summarization. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4005–4023, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[15] Anirban Laha, Parag Jain, Abhijit Mishra, and Karthik Sankaranarayanan. Scalable micro-planned generation of discourse from structured data. *Comput. Linguist.*, 45(4):737–763, jan 2020.

[16] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A Alghamdi, et al. A design space for intelligent and interactive writing assistants. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–35, 2024.

[17] Cencen Liu, Yi Xu, Wen Yin, and Dezhang Zheng. Structure-aware table-to-text generation with prefix-tuning. In *Proceedings of the 2023 4th International Conference on Control, Robotics and Intelligent System*, CCRIS '23, page 135–140, New York, NY, USA, 2023. Association for Computing Machinery.

[18] Mengsha Liu, Daoyuan Chen, Yaliang Li, Guian Fang, and Ying Shen. ChartThinker: A contextual chain-of-thought approach to optimized chart summarization. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 3057–3074, Torino, Italia, May 2024. ELRA and ICCL.

[19] Jason Obeid and Enamul Hoque. Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model. In Brian Davis, Yvette

Graham, John Kelleher, and Yaji Sripada, editors, *Proceedings of the 13th International Conference on Natural Language Generation*, pages 138–147, Dublin, Ireland, December 2020. Association for Computational Linguistics.

[20] Hua Shen, Tiffany Knearem, Reshmi Ghosh, Kenan Alkiek, Kundan Krishna, Yachuan Liu, Ziqiao Ma, Savvas Petridis, Yi-Hao Peng, Li Qiwei, et al. Towards bidirectional human-ai alignment: A systematic review for clarifications, framework, and future directions. *arXiv preprint arXiv:2406.09264*, 2024.

[21] Weiwei Shi, Yubo Liu, Jie Wu, and Jianming Liao. Three-stage logical table-to-text generation based on type control. In *Proceedings of the 2022 5th International Conference on Algorithms, Computing and Artificial Intelligence*, ACAI '22, New York, NY, USA, 2023. Association for Computing Machinery.

[22] Ashish Singh, Prateek Agarwal, Zixuan Huang, Arpita Singh, Tong Yu, Sungchul Kim, Victor Bursztyn, Nikos Vlassis, and Ryan A Rossi. Figcaps-hf: A figure-to-caption generative framework and benchmark with human feedback. *arXiv preprint arXiv:2307.10867*, 2023.

[23] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L Glassman. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Transactions on Computer-Human Interaction*, 30(5):1–57, 2023.

[24] Nikhil Singh, Lucy Lu Wang, and Jonathan Bragg. Figura11y: Ai assistance for writing scientific alt text. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*, pages 886–906, 2024.

[25] Andrea Spreafico and Giuseppe Carenini. Neural data-driven captioning of time-series line charts. In *Proceedings of the 2020 International Conference on Advanced Visual Interfaces*, AVI '20, New York, NY, USA, 2020. Association for Computing Machinery.

[26] Lu Sun, Stone Tao, Junjie Hu, and Steven P Dow. Metawriter: Exploring the potential and perils of ai writing support in scientific peer review. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1):1–32, 2024.

[27] Florian Weber, Thiemo Wambsganss, Seyed Parsa Neshaei, and Matthias Soellner. Legalwriter: An intelligent writing support system for structured and persuasive legal case writing for novice law students. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–23, 2024.