







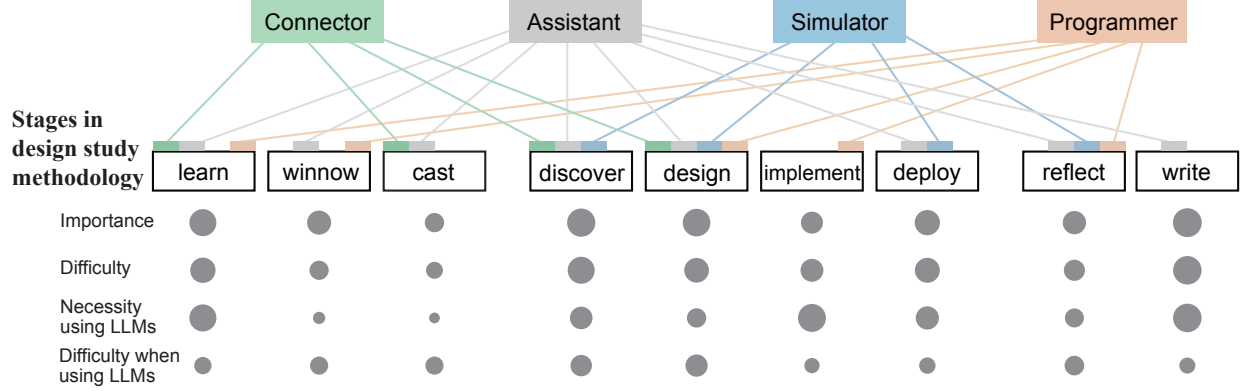


Qualitative Study for LLM-assisted Design Study Process: Strategies, Challenges, and Roles

Shaolun Ruan , Rui Sheng , Xiaolin Wen , Jiachen Wang , Tianyi Zhang ,
Yong Wang , Tim Dwyer , and Jiannan Li 

Roles for LLM-assisted design study



Ratings for researchers' attitudes toward with and without LLMs (1 point ~ ● 7 points)

Fig. 1: An overview of the summarized roles of LLMs in the LLM-assisted design study process. Following the framework proposed by Sedlmair et al. [34], we coded all nine stages of the design study by the compiled roles: 1) *Connector*: LLMs are used to bridge gaps between visualization researchers and domain experts; 3) *Simulator*: LLMs support simulating user behavior, predicting feedback, and testing workflows to uncover usability issues and hidden use cases; 4) *Programmer*: LLMs are asked to assist in coding tasks, from prototyping to debugging. The bottom shows the 30 researchers' ratings of importance, difficulty, necessity of using LLMs, and difficulty of using LLMs for each stage, where the size of each circle represents the average rating, with larger circles indicating higher scores.

Abstract—Design studies aim to create visualization solutions for real-world problems of different application domains. Recently, the emergence of large language models (LLMs) has introduced new opportunities to enhance the design study process, providing capabilities such as creative problem-solving, data handling, and insightful analysis. However, despite their growing popularity, there remains a lack of systematic understanding of how LLMs can effectively assist researchers in visualization-specific design studies. In this paper, we conducted a multi-stage qualitative study to fill this gap, involving 30 design study researchers from diverse backgrounds and expertise levels. Through in-depth interviews and carefully-designed questionnaires, we investigated strategies for utilizing LLMs, the challenges encountered, and the practices used to overcome them. We further compiled and summarized the roles that LLMs can play across the different stages of the design study process. Our findings highlight practical implications to inform visualization practitioners, and provide a framework for leveraging LLMs to enhance and innovate the design study process in visualization research.

Index Terms—Design Study, Large Language Models (LLMs), Qualitative Study, Visualization

1 INTRODUCTION

Sedlmair *et al.* introduce the term *design study* to describe an applied research methodology that focuses on creating visualizations to address

specific, real-world problems [34]. More specifically, they defined a visualization design study as “a project in which visualization researchers analyze a specific problem faced by domain experts and design a visualization system that supports solving this problem” [34]. Building upon this definition, a nine-stage methodology framework was proposed for conducting design studies, as shown in Figure 1, which has become a common guide for this category of research in the visualization community. Design studies are crucial because they bridge the gap between theoretical visualization research and practical applications, ensuring that visualization tools are both usable and useful in real-world contexts [29]. Basically, researchers conducted design studies through labor-intensive processes (*e.g.*, manually connecting with domain experts, extracting requirements via interviews, and iteratively designing and developing visualization systems), requiring significant experience and refined knowledge to gain effective solutions.

Meanwhile, Large Language Models (LLMs) have gained increasing popularity in recent years due to their impressive capabilities in natural language processing and generation [48]. LLMs have been increasingly used to enhance various aspects of the design study process, including

- S. Ruan, T. Zhang, and J. Li are with Singapore Management University. E-mail: {slruan.2021, tianyizhang.2023}@phdcs.smu.edu.sg and jiannanli@smu.edu.sg. S. Ruan is also with Monash University.
- R. Sheng is with the Hong Kong University of Science and Technology. E-mail: rshengac@connect.ust.hk.
- X. Wen and Y. Wang are with Nanyang Technological University. E-mail: {xiaolin.wen, yong-wang}@ntu.edu.sg.
- J. Wang is with Zhejiang University. E-mail: wangjiachen@zju.edu.cn.
- T. Dwyer is with Monash University. E-mail: tim.dwyer@monash.edu.
- J. Li and Y. Wang are the co-corresponding authors.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxx

assisting in creative problem-solving, handling large volumes of data, and generating insightful analysis results [18, 24, 38, 50]. Thus, by leveraging these capabilities, LLMs show the potential to significantly streamline the workflow of design studies and make the process more efficient and effective [2, 15].

Despite the growing popularity of utilizing LLMs in conducting design studies, the question of how LLMs can effectively assist researchers in this domain remains underexplored. For example, researchers with less expertise need to “deepen” their understanding of LLM-assisted design studies to reach higher levels of proficiency. Also, those with moderate-level of expertise can master the practice from peers to quickly “broaden” their knowledge in powering design studies by LLMs. However, existing research [12, 14, 36] on how LLMs can aid co-design methodologies in other fields, such as poster design [25], cannot be easily generalized to the field of visualization due to its visualization-specific workflows and methodologies. Thus, a systematic study on how to facilitate the LLM-assisted design study process is urgently needed in the visualization domain.

The requirement above points out several specific needs for achieving effective integration of LLMs into the design study process. First, design study researchers often utilize LLMs based on their own understanding and personal experience, without any systematic and rigorous stage-by-stage guidance on how LLMs can assist in each procedure of the design study methodology as references. Second, researchers are often unaware of the potential challenges they may face during the study, which can lead to significant confusion and time-consuming corrective efforts [42]. Even though these challenges have been recognized, effective practices to correct these problematic responses are also missing for individual design study researchers. Third, to easily understand and utilize LLMs in design studies, there is a lack of a high-level summary of what LLMs are able to help with, which significantly hinders the formation of a systematic framework for general design study researchers.

To fill the research gap, we conduct a multi-stage qualitative study that investigates common strategies and challenges for researchers who utilize LLMs in their design study processes’ current practices to inform the effective integration of LLMs into this methodology. To provide a comprehensive perspective, we invited 30 design study developers and researchers from five different countries with varying levels of expertise (i.e., novice, advanced beginner, intermediate, proficient, and expert) to share their insights. Specifically, we first collected their background and demographic information for situating their qualitative responses. To address the first need above, we conducted interviews with those participants to explore their *strategies* for how LLMs could assist in design studies. Then, we asked them follow-up questions to identify the *challenges* they encountered while implementing these strategies and the practice they used to address them. To further characterize the types of tasks that LLMs can effectively assist with, we derived a set of *roles* that LLMs may play during the design study stages by synthesizing all feedback and performing a post-study analysis. In addition to the interviews, we also incorporated a questionnaire to gather more nuanced insights into how developers and researchers navigate each stage of the design study methodology [34] via comparing their viewpoints with and without LLM assistance. Finally, we distilled implications from all the findings that can inform both visualization practitioners and LLM researchers of a more productive LLM-assisted design study process in the future.

2 RELATED WORK

This section provides an overview of the existing literature relevant to this study, which can be categorized into three major areas: Human-LLM collaboration in visualization, LLM-assisted co-design process, and qualitative study in visualization.

2.1 Human-LLM Collaboration in Visualization

Several studies have explored the dynamics of human-AI collaboration, particularly in the context of interactive machine learning and visual analytics. Saha et al. [32] examined the design and evolution of interactive machine-learning interfaces and pointed out usability

challenges and interaction paradigms that optimize collaboration. Similarly, Kovalerchuk et al. [20] investigated the role of AI in visual knowledge discovery and also identified key challenges for AI integration in visualization processes. Another relevant study by Schelble et al. [33] analyzed shared mental models and trust in human-agent teams. Furthermore, Chen et al. [4] presented a case study on human-AI co-innovation, showcasing how AI augments human creativity and analytical tasks to enhance problem-solving capabilities. In recent years, with the increasing prevalence of LLMs, the integration of large language models (LLMs) in visualization has been an active research area. For example, Fill et al. [7] provided an overview of how LLMs contribute to accessibility and efficiency in visualization tools. Also, researchers also developed an LLM-powered visualization system to support complex data analysis. Kim et al. [17] proposed a framework that improves LLMs’ understanding of visual data to enhance user interactions with visualization systems. Recent developments in natural language interfaces for visualization include Maddigan et al. [28], which introduced a system that translates natural language queries into data visualizations. Similarly, Narechania et al. [30] developed a toolkit that enables users to specify and translate them into visual representations to streamline analytics workflows. Similarly, Zhao et al. [49] proposed LightVA, a lightweight visual analytics system leveraging LLMs for automated task planning and execution in data analysis workflows. More recently, Kim et al. [19] studied how could ChatGPT be used in the educational scenario. Additionally, Xu et al. [45] assessed the effectiveness of LLMs in performing low-level analytic tasks on SVG visualizations, demonstrating their potential for enhancing user experience and efficiency in data interpretation. Moreover, prior work explored leveraging LLMs for visual workflow generation, regarding financial Q&A tasks [47], industrial applications [21], and business process automation [39, 44].

Despite some insightful knowledge has been yielded from the above studies, they barely investigated the high-level summary of how LLMs are utilized in the whole process of the design study for visualization, which can directly inform design study researchers of the effective usage of LLMs.

2.2 LLM-assisted Co-design Process

Apart from the visualization field, an increasing integration of Large Language Models (LLMs) across various domains has also led to extensive research for contextualizing the role of LLMs in assisting co-design workflows. As an active field of utilizing LLMs, Human-Computer Interaction (HCI), researchers have explored its potential to enhance user experience, model transparency, and human-AI collaboration. Specifically, Kim et al. [16] investigated the co-design process of LLM-powered human-robot interactions, which highlights the ability of these models to facilitate seamless communication and adaptive interactions. Similarly, Liao et al. [23] examined the need for model transparency in AI-powered user experience design and emphasized the importance of interpretability for designers working with AI-generated suggestions. Sun et al. [40] explored the role of LLMs in UI/UX design by demonstrating their capacity to advance the development of interactive systems. Also, H. Fill et al. [8] discussed how the LLMs can be used to generate and interpret various types of conceptual models, e.g., Entity-Relationship (ER) diagram. The work was followed by a study of a framework called Conceptual Model Augmented Generative Artificial Intelligence (CMAG) [9], which uses conceptual models to validate and enhance the outputs of generative AI models for improving their reliability and interpretability. Beyond HCI, LLMs have also shown great potential in facilitating co-design processes in medical applications. For example, Swanson et al. [41] introduced a virtual lab framework in which AI agents designed novel SARS-CoV-2 nanobodies. The use of LLMs in the engineering field has also been studied. Chiarello et al. [5] analyzed the role of generative LLMs in engineering design by identifying key challenges associated with automating design processes. Similarly, Gomez et al. [11] explored the application of LLMs in complex system design to facilitate knowledge transfer, automate documentation, and enhance problem-solving in engineering workflows. LLMs have also been employed in game design to support

creative and educational development. Gatti et al. [10] investigated how ChatGPT can inspire and improve serious board game design which enhances player engagement. Furthermore, previous studies also investigated the LLM-driven design authoring tool for various objectives, such as dashboard [37], programming [35, 46], and videos [22, 26, 31]

While prior work has demonstrated the capacity of LLMs to assist in creative problem-solving, knowledge synthesis, and automated design, challenges such as model transparency, reliability, and domain-specific adaptation in the area of visualization and visual analytics remain under-explored, which has yielded a strong need in the visualization community.

2.3 Qualitative Study in Visualization

The integration of qualitative methods, such as interviews, into visualization research has been explored across various domains, which provides valuable insights for informing the design and application of visualization tools. More specifically, in the domain of data representation, Hogan et al. [13] introduced the Elicitation Interview Technique to capture users’ experiences with visual data, which provides a structured method to enable researchers to better understand how users interpret and interact with visualizations. Also, Batch and Elmqvist [3] investigated the gap in interactive visualization during initial exploratory data analysis via qualitative interviews with data analysts. After that, they proposed a methodological framework for highlighting how pair interviews enhance data collection processes, particularly in collaborative visualization studies [1]. Meanwhile, Wu et al. [43] also initiated a workshop to discuss the critics of the contribution and rigor of visual analytics. In the context of public health in recent years, Çay et al. [51] explored user experiences with COVID-19 maps using remote elicitation interviews, revealing how users interpret pandemic-related visualizations and identified challenges in designing effective visual communication tools. Visualization-specific user perspectives have also been explored in statistical and technical domains. For instance, the role of visualization in inferential statistics was studied in [6]. Similarly, challenges and opportunities in distributed tracing visualization were characterized by Davidson et al. [27].

The aforementioned studies have demonstrated that qualitative research is a highly effective method for uncovering valuable insights in the field of visualization, given its inherently human-centered and subjective nature. However, the prior studies rarely explored how emerging tools, like LLMs, could help facilitate the design study process in visualization. This paper addresses this gap by investigating the potential strategies and challenges of utilizing LLMs to streamline workflows and foster more efficient and creative approaches during the design study process.

3 METHODOLOGY

To explore the strategies, challenges, and roles of the LLM-assisted design study process, we conducted a multi-stage qualitative study, involving participant profiling, an interview, and a follow-up questionnaire. This section introduced our methodological approach. We first described the participants involved, the data collection methods employed, and the procedures followed throughout the study.

3.1 Participants

We recruited 30 paid participants via direct email request with the compensation of 15 SGD in the form of a gift card. As listed in Figure 3, the participants were affiliated with 11 institutions located across Singapore, mainland China, Hong Kong, Australia, and the United States. Their academic and professional roles included professors, Ph.D. candidates, post-doctoral researchers, and undergraduate students. With this participant group, we aimed to survey a diverse range of perspectives about how LLMs are perceived and utilized across varying levels of experience.

Participants’ expertise in visualization and design study was identified by themselves into five levels: novice, advanced beginner, intermediate, proficient, and expert. Meanwhile, the participants’ design study research spanned a wide range of interdisciplinary research domains, including biomedical science, finance, smart cities, sports, etc. In terms

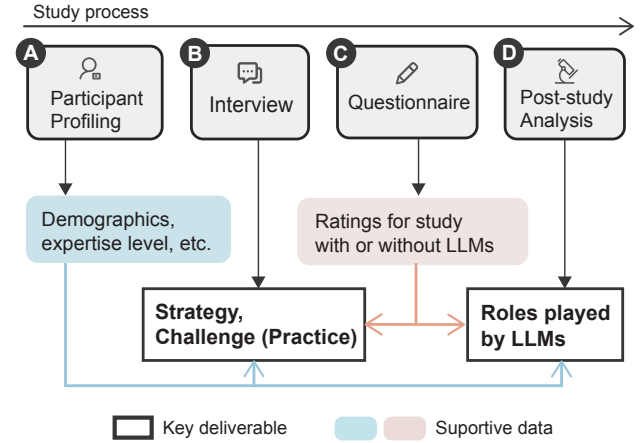


Fig. 2: Our study consists of four steps: (A) participant profiling, (B) interview process, (C) questionnaire, and (D) post-study analysis. The outcomes of supportive data from (A) and (C) are used to support the in-depth analysis of interviews (B) and post-study analysis (D).

of research productivity, our participants had contributed to 400 research papers in total, of which 250 were accepted in peer-reviewed venues. Novice participants typically had limited publication records, often having contributed to one or two design study research. Proficient participants demonstrated more substantial outputs and many of them had authored 10 to 20 papers. Experts had the most extensive publication records, with some contributing to over 40 papers.

Furthermore, we also requested participants to report the most-used LLM variants and frequency of LLM usage in their workflows. We gathered this information to understand their overall familiarity with LLMs and the extent to which LLMs are important to their research workflows.

3.2 Data Collection

We illustrate our data collection process in Figure 2. As shown in Figure 2, we use the visual representation to better illustrate the study workflow. Firstly, all interviews were conducted via online video calls. Before each interview, we asked if the participant agreed with the recording, and we recorded the whole session upon their consent. At the same time, the interviewer took notes and created sketches based on the visualizations described by the participants during each session. We shared these notes and sketches with participants to make sure they accurately reflected participants’ original intentions. Additionally, the researcher documented general impressions and recalled details immediately to capture information that may not have been noted in real-time.

For the stages of participant profiling (Figure 2(A)) and questionnaire (Figure 2(C)), we verbally inquired all interviewees during the meeting right before and after the qualitative interview. Note that all participants consented to provide their demographic information before we moved on to the stage of participant profiling. All data was collected via Google Sheets and every participants can access and edit it anytime.

3.3 Procedure

Participant Profiling. Before the interviews began, participants were asked to provide demographic information, their visualization expertise, and LLM usage. We also collected information about their research expertise in visualization, including their self-reported expertise level among five levels (novice, advanced beginner, intermediate, proficient, and expert) and the number of research papers they had worked on and published. Additionally, participants were asked to indicate their primary design study domain.

To further characterize their reliance on LLMs, we asked participants to report their most-used LLM services and the frequency of usage. Specifically, usage frequency refers to the approximate number

A				B		C		D		
	Ongoing degree	Vis expertise	Design study domain	# of research involved (accepted)	Mostly-used LLMs	Usage frequency of LLMs	Residence	Age	Self-described gender	
1	Undergraduate	Novice	Society Goods	1 (0)	GPT-4o	11-20 ■■	Singapore	21-25	Male	
2	Undergraduate	Novice	Education	1 (0)	GPT-4o	1-10 ■	Singapore	21-25	Male	
3	Undergraduate	Novice	Music	1 (0)	GPT-4o, GPT-o1	1-10 ■	Singapore	16-20	Female	
4	Undergraduate	Novice	Music	1 (0)	GPT-4o	11-20 ■■	Singapore	21-25	Female	
5	Undergraduate	Novice	Society Goods	1 (0)	GPT-4o	21-30 ■■■■	Singapore	16-20	Female	
6	Undergraduate	Novice	Society Goods	1 (0)	GPT-4o	21-30 ■■■■	Singapore	21-25	Male	
7	Undergraduate	Novice	Healthcare	1 (0)	GPT-4o, GPT-o1	11-20 ■■	Singapore	21-25	Male	
8	Undergraduate	Advanced beginner	Human-AI Collaboration	3 (1)	GPT-4o, Deepseek	21-30 ■■■■	Mainland China	21-25	Male	
9	PhD	Intermediate	Text Visualization	4-5 (3)	GPT-4o, Claude 3.5	> 30 ■■■■	Australia	26-30	Male	
10	PhD	Proficient	Finance; Software	18 (10)	GPT-4	21-30 ■■■■	Singapore	26-30	Male	
11	PhD	Proficient	Biomedical	21 (3)	GPT-3.5	21-30 ■■■■	Hong Kong	26-30	Male	
12	PhD	Proficient	Bioinformatics	16 (9)	GPT-4o	21-30 ■■■■	Singapore	26-30	Male	
13	PhD	Proficient	E-learning	16 (6)	GPT-4	21-30 ■■■■	Hong Kong	21-25	Male	
14	PhD	Proficient	Creative Design	17 (10)	GPT-4o, DeepSeek	11-20 ■■	Hong Kong	26-30	Female	
15	PhD	Proficient	Color	15 (9)	GPT-4o, DeepSeek	21-30 ■■■■	Mainland China	21-25	Female	
16	PhD	Proficient	Cultural Heritage	4 (2)	GPT-4o, DeepSeek	> 30 ■■■■	Mainland China	26-30	Male	
17	PhD	Proficient	Medical; Notebook	10 (4)	GPT-4o, GPT-4, DeepSeek	21-30 ■■■■	Mainland China	21-25	Male	
18	PhD	Proficient	XAI	15 (10)	GPT-4o, other (Kimi)	> 30 ■■■■	Mainland China	31-35	Male	
19	PhD	Advanced beginner	Human-AI Collaboration	3 (1)	DeepSeek	> 30 ■■■■	Hong Kong	26-30	Male	
20	PhD	Advanced beginner	AR/VR	6 (3)	GPT-4	21-30 ■■■■	Hong Kong	26-30	Male	
21	PostDoc	Proficient	Storytelling	15 (9)	GPT-4	1-10 ■	Australia	21-25	Female	
22	PostDoc	Proficient	Notebook	13 (8)	GPT-4o, GPT-o1	21-30 ■■■■	Hong Kong	26-30	Female	
23	Professor	Expert	Sports	20 (15)	GPT-4o	1-10 ■	Mainland China	26-30	Male	
24	Professor	Intermediate	Drug; Biomedical	21 (16)	DeepSeek	11-20 ■■	Mainland China	26-30	Female	
25	Professor	Expert	Smart City	40 (20)	GPT-4	11-20 ■■	Mainland China	31-35	Male	
26	Professor	Expert	Data Video	20 (10)	GPT-4o, DeepSeek	11-20 ■■	Hong Kong	31-35	Female	
27	Professor	Expert	Smart City	8 (5)	GPT-4o	11-20 ■■	Australia	41-45	Female	
28	Professor	Proficient	Quantum Computing	15 (8)	GPT-4o, Claude 3.5	21-30 ■■■■	United States	46-50	Male	
29	Professor	Proficient	Material Science	12 (8)	DeepSeek	21-30 ■■■■	Mainland China	41-45	Male	
30	Professor	Proficient	High-performance Comp.	11 (9)	GPT-4o	1-10 ■	United States	41-45	Female	

Fig. 3: Table for the participant profiling of 30 participants in our study. All dimensions are grouped into four categories: educational background (Area A), visualization study engagement (Area B), LLM utilization (Area C), and demographics (Area D). We sorted all participants according to their ongoing degrees. All participants' names were anonymized.

of prompts one sends per day during the design study process. We categorized the number into four ranges: 0–10 prompts/day, 11–20 prompts/day, 21–30 prompts/day, and more than 30 prompts/day. The participant profiling stage was brief, with each session lasting approximately 5 minutes.

Interview. At the beginning of each interview session, we first introduced the Sedlmair et al. [34] nine-stage framework to participants. To ensure clarity, we provided detailed explanations of each stage, supplemented by illustrative examples to help participants understand the purpose and process of each stage. This introduction served as the foundation for the subsequent discussions. After that, we conducted one-on-one, semistructured, hour-long interviews with each participant. The interviews were designed around the three phases outlined in the original paper, i.e., precondition, core design study, and analysis. This allowed us to systematically explore participants' experiences and practices at different stages of the design study process. During the interviews, we first asked participants to describe the strategies they typically employ in their common design study processes to learn about their research workflows and approaches. Subsequently, we discussed with participants about the challenges or pitfalls in applying these strategies and how they attempted to address the challenges. During the interviews, we took notes and asked follow-up questions for clarification.

Questionnaire. After completing the interviews, all participants were asked to fill in a post-study questionnaire. This questionnaire included four questions for each individual stage in the nine-stage framework, totaling 36 items (Figure 4). Specifically, firstly participants were asked to rate the *importance* and *difficulty* of each stage without LLMs. Then they also rate the *necessity* of each stage requiring the assistance of LLMs, followed by the rating for asking how *difficult* it is to utilize LLMs in this stage. These ratings were based on a 7-point Likert scale, where 1 indicated “not important/difficult at all” and 7

Q1	How significant do you believe this stage is?
Q2	How difficult do you find this stage to be?
Q3	How necessary do you believe the LLM assistance is in this stage?
Q4	How difficult is it to utilize the LLM assistance to help with this stage?
X	
learn winnow cast	discover design implement deploy reflect write
PRECONDITION	CORE ANALYSIS

Fig. 4: The questionnaire questions upon the interview. All participants were invited to rate each question regarding the nine stages of the framework of the design study.

indicated “extremely important/difficult”. The purpose of the questionnaire was twofold: 1) To understand how much participants value each stage and perceive its difficulty when performing the design study process without LLM assistance; 2) To explore the perceived necessity of using LLMs for each stage, as well as the difficulty of incorporating LLM assistance into the process. We used the responses to this questionnaire to complement the qualitative feedback gathered during the interviews and broaden our understanding of how participants with varying attitudes toward LLMs adapt their research strategies in design studies.

Post-study analysis. We conducted a post-interview analysis to identify the high-level roles that LLMs can play in design study processes. The goal was to unify the diverse strategies employed by participants with different backgrounds and experiences. Our analysis mainly focused on the qualitative feedback collected during the qualitative interview step. Furthermore, we leveraged the information from participant profiling and their questionnaire responses to more accurately characterize individual researchers according to their exper-

tise, levels of reliance on LLMs, and attitudes toward LLM-assisted workflows. By synthesizing these data sources, we were able to identify and analyze participants' varying perspectives on each stage of the design study process. To compile the roles, a team of three researchers independently reviewed the collected qualitative data, including interview transcripts, questionnaire responses, and participant profiles. Each researcher proposed an initial set of roles based on their interpretation of the data. These roles were then discussed in a series of collaborative meetings to reach a consensus. In cases where disagreements arose, the team revisited the original data and engaged in iterative discussions to refine the roles until all members reached an agreement. These identified roles are intended to serve as a reference and inspiration for the broader visualization community, offering guidance on how LLMs can be effectively incorporated into diverse workflows.

4 ROLES OF LLMs IN DESIGN STUDIES

In this section, we introduce the roles of LLMs in supporting design study processes as our study found. By analyzing the strategies collected from the 30 visualization researchers during the qualitative interview, we identified four roles, i.e., **connector**, **simulator**, **programmer**, and **assistant**. We first introduce the roles and then present the main strategies related to these roles in Section 5.

Connector: The **connector** role focuses on bridging the gap between visualization researchers and domain experts. Specifically, LLMs in this role help researchers learn domain-specific knowledge, extract design requirements, and interact with domain experts more effectively. For novice researchers, LLMs act as a learning tool, providing background knowledge about the target domain to help them quickly familiarize themselves with unfamiliar concepts. For proficient PhD students, LLMs assist in analyzing domain-specific workflows, identifying key tasks, and collecting practical design requirements. For example, a novice researcher might use an LLM to understand the basics of bioinformatics before beginning a visualization project (e.g., P6). Meanwhile, a proficient PhD student could use an LLM to record how bioinformatics researchers analyze tumor tissue slices, ensuring their visualization aligns with real-world tasks. Experts might use LLMs to update their knowledge of a certain domain or generate drafts for academic papers based on high-level summaries. Additionally, LLMs support communication by helping researchers articulate their needs to domain experts or translating domain-specific terminology into accessible language. This role is most frequently relied upon by novice researchers to acquire domain knowledge and by advanced and proficient PhD students to learn design requirements and communicate effectively with domain experts.

Simulator: In their **simulator** roles, LLMs are often used to simulate responses to questionnaires, predict potential user feedback, or test outcomes of workflows. By simulating user behavior, LLMs help researchers uncover hidden use cases, evaluate system usability, and identify cognitive load issues before presenting the system to real users. Additionally, LLMs can take on the role of a critical reviewer, challenging the reasoning and logic behind a system and suggesting potential improvements. For example, a proficient PhD student might use an LLM to simulate user responses during the evaluation process, combining these responses with real user feedback to refine their system. This role is most frequently relied upon by proficient PhD students who are experienced enough to incorporate simulated insights into their design studies effectively.

Programmer: The **programmer** role is one of the most widely used roles of LLMs, which can support researchers in software development tasks ranging from prototyping to debugging. This usage helps researchers generate, optimize, and revise code in order to facilitate the technical implementation process. Specifically, novice, intermediate, and proficient researchers all reported heavy usage of LLMs to handle various programming tasks. For beginners, LLMs provide guidance for setting up basic frameworks or generating simple scripts, such as creating mock datasets or pre-processing data. Intermediate researchers use LLMs for more complex coding tasks, such as generating JavaScript code for interactive visualizations or implementing visualization de-

signs. Proficient researchers are often able to utilize LLMs to improve the overall user interactions of their visualization system prototype. For example, an intermediate researcher could use it to create JavaScript code that translates hand-drawn drafts into functional visualizations. Meanwhile, a proficient researcher might use an LLM to simplify SVG elements or optimize a visualization's runtime performance.

Assistant: The assistant role encompasses general-purpose, miscellaneous assistance that LLMs provide to enhance researchers' productivity and efficiency by automating repetitive or time-consuming tasks. For example, LLMs are usually used for tasks such as searching academic papers, summarizing collaborator's background from CV, or generating scripts from evaluation recordings. Unlike the other roles, this role is used by researchers across all experience levels. For example, a novice researcher might use an LLM to summarize a complex academic paper into key points they can easily understand. A postdoc researcher could use it to transcribe interview recordings and distill core messages from them.

5 QUALITATIVE INTERVIEW RESULTS

In this section, we present the qualitative feedback collected from 30 participants, focusing on their experiences with leveraging LLMs to assist design studies in individual stages of the Sedlmair et al. framework [34]. We explore the *strategies* they employed, the *challenges* they encountered, and any useful *practices* they adopted in the process when working with LLMs playing the four kinds of roles as defined above. We label findings with the associated LLM roles. Our analysis aims to uncover diverse insights that emerge across participants with varying levels of visualization expertise and LLM usage frequency.

1 Precondition: Learn

"In summary, LLMs were found to be highly effective in aiding novice researchers to quickly grasp domain fundamentals, while advanced users leveraged them to stay updated with recent developments, together with a "validation-loop" between visualization researchers, domain experts, and LLMs. However, challenges such as inaccuracies in technical details and confirmation bias need users to adopt practices like consulting domain experts to ensure reliability."

The "Learn" stage highlights how researchers use LLMs to quickly acquire foundational knowledge in unfamiliar domains or stay updated with recent developments in their own fields. Most novice visualization developers found LLMs particularly helpful for quickly grasping fundamental concepts in a specific domain **connector**. For example, P6 noted that LLMs allowed them to "quickly grasp the basics of a domain that would otherwise require extensive manual searching." However, proficient and advanced users use LLMs to stay updated with the latest developments in their fields. Despite these benefits, participants faced challenges, particularly novices, who often treated LLMs as search engines without the expertise to discern inaccuracies in the responses. P5 mentioned that "LLMs occasionally provide incorrect answers, but as a beginner, it is difficult to tell if there are actually any incorrect contents." To overcome this uncertainty, novices are often forced to double-check information with people with more expertise often at a significant cost of time. Another common challenge was the accuracy of detailed explanations provided by LLMs. While their high-level summaries were generally correct, technical details were sometimes flawed, as noted by P23, who gave an example where the explanation of an object detection algorithm was conceptually accurate but the technical details about the model layer were incorrect. To address these issues, participants adopted various practices, including consulting domain experts, cross-referencing LLM outputs with literature, and generating tailored learning materials using LLMs. For instance, P9 used LLMs to create coding tutorials and background explanations for self-training in new domains. But he noted the limitations that LLMs could not effectively use many emerging but immature tools (e.g., the package 'turf.js' in the field of geospatial visualization package), even when provided with its tutorial.

Another interesting finding is from advanced users. They noted a “validation loop”, involving a dynamic process where visualization researchers, domain experts, and LLMs cross-verify with each other `connector` assistant . For example, visualization researchers used LLMs to validate domain knowledge provided by domain practitioners, while domain experts reviewed LLM-generated designing methods with visualization researchers to ensure alignment with real practice. Meanwhile, visualization users also validate the LLM-generated domain knowledge with domain practitioners as well. P21 described this triangulation process as effective for ensuring reliability but sometimes introducing pitfalls, because LLMs often aligned with the phrasing or assumptions embedded in prompts, a phenomenon referred to as “confirmation bias.” One example is that if the domain experts ask LLM “Please help me to decide if the stacked bar charts can fit my data or not,” and LLM will most likely provide you with the answers about why stacked bar charts fit your data instead of critical answers. To mitigate this, P21 recommended allowing LLMs to generate independent evidence and conclusions.

Furthermore, participants also highlighted unique strategies for leveraging LLMs during this “Learn” stage. For example, P24 used LLMs to fill gaps in domain-specific terminology `connector` , enabling them to bridge knowledge gaps more efficiently. However, P21 also warned against relying on LLMs for summarizing design requirements, as these summaries often failed to align with their domain expertise. Similarly, P9 employed LLMs to generate tailored courses and answers for self-training for a new domain and also the technical stack `connector` programmer .

2 Precondition: Winnow

“In summary, LLMs provided significant support in evaluating visualization necessity, identifying collaborators, and generating mock datasets. Despite challenges like hallucinated references and privacy concerns, participants employed strategies such as using LLM-powered platforms and refining generated content to ensure reliability and privacy.”

The “Winnow” stage is about filtering and narrowing down potential problems and collaborators. Novice students often used LLMs to evaluate whether visualization techniques could effectively solve a given problem. Specifically, they prompted LLMs to provide the most relevant non-visualization papers to check if there were existing automatic approaches or algorithms that could address the problem assistant . Additionally, they explored whether visualization methods, even from different domains, could be generalized and applied to their context assistant . However, a major challenge arose due to LLM hallucinations, as popular models frequently recommended papers that did not exist. To overcome this, participants turned to LLM-powered literature search platforms, such as Scite.com, to conduct more reliable literature reviews and verify the authenticity of suggested resources.

PhD students noted that prior to the availability of LLMs, they primarily relied on personal connections or pre-existing knowledge of researchers in the domain for collaborator recruitment. With LLMs, however, they could directly request a list of relevant researchers working in the same field assistant . Subsequently, some advanced researchers utilized LLMs to generate recruitment emails for inviting collaborators from other domains assistant . While this approach saved time, it introduced specific challenges. First, the emails were often too generic, lacking personalized details such as why the recipient was being invited or how their research aligned with the project. To address this, participants attached the recipient’s CV or homepage link to the LLM prompt, enabling the model to generate more tailored emails. Second, privacy concerns arose when describing research projects to LLMs, as the model could extract sensitive keywords or overly detailed descriptions, risking unintentional idea leakage. To mitigate this, participants stressed the need to manually verify generated content before sharing.

P11 highlighted using LLMs to address delays in obtaining datasets

by generating mock datasets or expanding historical data while awaiting real data programmer . Once real data was available, LLMs helped quickly adapt prototypes. Additionally, P11 used LLMs to draft Institutional Review Board (IRB) applications or consent forms, saving time while ensuring compliance with research protocols assistant .

3 Precondition: Cast

“Generally, LLMs facilitated role assignment and collaboration by analyzing collaborators’ expertise, and also acted as agents to bridge interdisciplinary gaps. However, limitations in assessing personal traits and achieving granularity required additional input from principal investigators (PIs) or ‘gatekeepers’ to finalize decisions.”

The “Cast” stage focuses on identifying and assembling the right team of collaborators for a design study. Advanced beginners in visualization used LLMs to assign roles within a project by providing the model with collaborators’ CVs or brief descriptions of their expertise assistant . While this approach facilitates the initial role assignment, participants noted challenges in achieving sufficient granularity. For example, it was difficult for LLMs to distinguish whether a collaborator with a visualization background would be better suited as a designer or a programmer. Furthermore, LLMs lacked the ability to assess personal traits, such as leadership or communication skills, making it challenging to identify appropriate candidates for roles like team manager or connectors, who are responsible for coordinating all members. To address these limitations, participants often relied on the Principal Investigator (PI) or a designated “gatekeeper” to make the final decisions regarding role assignments.

Fresh PhD students faced additional difficulties in distinguishing between front-line analysts and translators within their teams connector . Thus, they usually used LLMs to analyze collaborators’ published papers and professional backgrounds, allowing the model to suggest suitable candidates for these roles. Beyond role assignment, LLMs can also be directly regarded as an “agent” to facilitate collaboration among researchers from different domains connector . For example, LLMs helped by providing task abstractions and translating domain-specific terminology into language that visualization researchers could understand.

4 Core: Discover

“LLMs were useful in uncovering or validating domain-specific requirements and summarizing literature, though challenges like hallucinated references and misinterpretation of visual design contributions needed manual validation and domain expertise.”

The “Discover” stage highlights how researchers use LLMs to uncover knowledge, identify related work, and understand domain-specific requirements. Novice researchers often used LLMs to explore the literature review by asking LLMs to suggest papers or summarize research trends assistant connector , which was particularly helpful for those unfamiliar with the domain. However, this strategy often introduced hallucinated references to novices, where LLMs usually provided nonexistent papers or incorrect publication details. According to P21, who is a professor and lecturer in visualization, LLMs often fail to recognize the visual design elements but the students totally cannot identify this incorrectness. Proficient PhD students and professors used LLMs to summarize the takeaways from the academic paper assistant , e.g., key contributions, methodologies, and findings from visualization papers. However, participants noted that LLMs often misinterpreted the contributions of visualization papers, particularly when analyzing visual design elements. For instance, the models often failed to accurately understand visual encodings or the rationale behind specific visual channels. This limitation required significant additional manual effort to validate outputs provided by the LLMs.

Beyond literature discovery, participants also explored interest-

ing ways to use LLMs to understand domain-specific requirements **connector**. Instead of relying solely on interviews with domain experts, some researchers used LLMs to analyze routine tasks performed by professionals in their fields. For example, P12 described using Copilot embedded in IDE to document how bioinformatics researchers analyze tumor tissue slices, which captures the practical workflows that might otherwise be overlooked in traditional interviews. This approach significantly reduced information loss and provided a more precise understanding of domain requirements.

Participants also expressed that they often use LLMs to double-check the collected design requirements from domain experts **connector**. However, LLMs' answers are usually too generic to help for the requirement validation. This limitation was particularly evident when researchers attempted to use LLMs to generate design requirements or propose novel solutions. As a result, researchers emphasized the importance of combining LLM outputs with their own expertise, using the models as tools to augment, rather than replace, their discovery process. P14 also mentioned the usage of the generation of the questionnaire for the pilot study or formative study assistant. Also, P14 used LLMs to generate the possible answers from the provided questionnaire **simulator**, which is used to compare with the real answers from the participants to identify some unique and insightful findings in this domain.

5 Core: Design

"In general, LLMs supported visualization design by bridging gaps between designers and domain experts, generating design alternatives, and optimizing usability. However, limitations in global design understanding and integration into complex frameworks required researchers to refine and adapt LLM outputs with domain experts."

The "Design" stage focuses on leveraging LLMs to assist visualization researchers in converting domain requirements into concrete visualization designs, generating design alternatives, and optimizing usability. Three professors highlighted the use of LLMs as "translators" during the design process to bridge the gap between visualization designers and domain experts **connector**. In this case, LLMs were used to facilitate communication by summarizing iterative meetings, generating meeting notes, and translating domain-specific knowledge and terminology into practical insights for visualization designers. Some PhD students reported using LLMs to decompose domain requirements, obtained during the "Discover" stage, into sub-tasks and brainstorm design ideas based on these sub-tasks assistant. For example, they would ask LLMs to search approaches used in prior visualization studies, such as those published in IEEE TVCG papers, for similar design requirements.

An advanced usage of LLMs involved refining initial visualization designs by analyzing code or design structures **programmer**. Participants described the strategy that feeding LLMs the HTML elements of their designs (e.g., including SVG elements) along with the design requirements for understanding the structure of the visualization. The LLMs were then tasked with simplifying or optimizing the code to improve usability. Additionally, researchers used hand-drawn drafts as input for LLMs to generate JavaScript code **programmer**, which was then validated to determine if the initial design could reveal the expected data patterns.

For system-wide design tasks, such as determining the overall layout, interaction logic, and workflow between views, LLMs were used to provide high-level suggestions **simulator**. However, participants noted that LLMs often struggled with these tasks because they lacked the ability to maintain a comprehensive understanding of the entire system. This limitation required researchers to manually refine and integrate the LLM's outputs into a cohesive global design.

LLMs were also employed as sources of design inspiration **assistant**. Some researchers provided a domain and asked the LLM to suggest straightforward visualization designs to inspire their work. While this approach often generated initial ideas, participants noted that the

suggestions were often limited to popular approaches. To address this challenge, researchers came up with practices such as incorporating evaluation metrics (e.g., significance or novelty) to assess the quality of the proposed designs. Moreover, PhD designers used LLMs to search for commercial tools online, leveraging existing solutions as inspiration for their designs.

6 Core: Implement

"Basically, LLMs were widely used for data processing, refining prototypes, and code generation. Despite challenges like incompatibility with complex frameworks and difficulty in fine-tuning designs, participants adopted best practices such as providing detailed context and writing explicit code comments to improve outcomes."

The "Implement" stage in the design study methodology explores how LLMs can be utilized to support various phases of visualization system development. According to our interview, this stage is the most common scenario where visualization developers rely on LLMs. First of all, for data processing tasks, novices and PhD students can use LLMs to generate code for data cleaning or other pre-processing tasks **programmer**. Also, after the system prototype is becoming mature, LLMs can be employed to generate large-scale datasets to test the system's scalability **programmer**, enabling researchers to initially evaluate if the prototype can seamlessly fit the domain real needs.

In terms of technical stack setup, according to the findings from the professors, LLMs can help beginners to generate basic but functional architectures **programmer**, such as a simple React.JS and Flask framework that supports fundamental tasks like data requesting. This approach allows developers to focus on design and functionality while avoiding the time-consuming process of learning new frameworks.

For implementing the design, LLMs can assist in generating code for visual design implementation **programmer**. While LLMs work well for basic chart designs when provided with a sketch or bitmap, they often fail to recognize hierarchical structures. For example, in the case of a stacked bar chart, LLMs may generate separate bars instead of a cohesive stacked structure. As a workaround, novices may turn to user-friendly tools like Tableau, but it introduced new issues of flexibility and interactivity, limiting the visualization to only simple analysis tasks. Another significant challenge arises when the generated JavaScript code is only suitable for static HTML pages but fails to integrate into component-based frameworks like React.JS. This incompatibility comes from several issues: complex interaction logic makes the code difficult to adapt, conflicts such as class name duplication can cause system bugs, the generated design may not align with the overall size or color scheme of the system, and SVG elements may not be correctly grouped, leading to issues with animations. The best practices to address these challenges include providing detailed context about all components to hint LLMs and reduce integration issues.

Several challenges also exist in the process of fine-tuning prototype systems. It is difficult for LLMs to modify specific annotated parts of a design, as they struggle to recognize which part of the design is being referenced. Also, if a user requests modifying the "top left" circle, LLMs lack spatial awareness and cannot identify the target. Moreover, some descriptions, such as making a color scheme "look cool" or "minimalist but not simplistic", are often misunderstood by LLMs. A possible practice to address these issues is to write detailed comments in the code that explicitly specify the modifications required.

7 Core: Deploy

"Interestingly, LLMs were used to simulate user behavior, analyze performance, and formulate potential cases in the evaluation process. Challenges like generic outputs and hallucinations were mitigated by combining LLM insights with human expertise and testing with multiple models for diverse perspectives."

In the Deploy stage, LLMs are used to evaluate the system's usability, explore potential usage scenarios, and test its performance under

real-world constraints. Advanced students have reported using LLMs to simulate user behavior, allowing the model to “pretend” to be a user `simulator`. This approach helps uncover hidden and potential use cases for the system, as well as identify alternative workflows that differ from conventional ones proposed by human experts. However, two significant challenges arise: first, the generic background knowledge of LLMs can lead to superficial stories that lack depth; second, hallucinations by LLMs may result in incorrect stories based on non-existent system features.

To evaluate whether the system adapts to real-world constraints, such as performance bottlenecks with large datasets, LLMs can be used to analyze backend network logs `assistant`. For example, using Chrome’s network log data, LLMs can identify performance problems related to data requests between backend data port, which can detect bottlenecks that may hinder system efficiency. Also, beginners often use LLMs to test example workflows to determine whether the cognitive load is manageable before presenting the system to domain users `simulator`. This approach ensures that the system’s design is intuitive and accessible to its target audience. On the other hand, experts have explored using different types of LLMs to test the system’s performance `simulator`. For example, models like GPT-4o are suited for reasoning tasks, while Claude may excel in logic-oriented evaluations. These experiments allow developers to assess how the system performs from the perspectives of various domain users.

Moreover, LLMs have mentioned some useful in supporting user studies and feedback analysis. Specifically, developers can transcribe user study recordings into text and use LLMs to summarize key takeaways and feedback `assistant`. Similarly, LLMs can assist in generating questionnaires and designing tasks based on the provided workflow and requirements. However, challenges exist in these practices. For example, relying on LLMs to summarize feedback may result in conclusions that do not align with the developer’s domain expertise. Additionally, when generating questionnaires based on historical papers, LLMs may over-rely on the provided materials, like previous papers, leading to repetitive questions.

8 Analysis: Reflect

“To sum up, LLMs supported reflective practices by simulating reviewer feedback and analyzing usability data. However, limitations in specificity and understanding user experiences required researchers to provide domain-specific data and detailed prompts to enhance feedback quality.”

The “Reflect” stage emphasizes critical evaluation of research progress, tools, and methodologies to identify limitations and improve outcomes. Advanced researchers used LLMs to simulate the role of a reviewer, challenging their research projects based on the abstract and identifying potential limitations `simulator`. While this approach provided an initial critique, participants noted that the feedback was often too generic, focusing on broad issues such as unclear motivation, scalability, or generalizability of the proposed design. To address this, participants adopted two practices to make the feedback more specific and actionable. First, they fed the LLM their own published papers to help the model learn the typical way of limitations relevant to their research domain. Second, they provided the LLM with a more specific direction, such as asking it to consider potential solutions for model fine-tuning.

Frequent LLM users highlighted the difficulty of capturing user experiences verbally via the evaluation process, particularly when investigating the usability of tools. To address this, they used LLMs to analyze performance data and assess whether the tool was suitable for real-world users `assistant` `programmer`. For example, participants prompted LLMs to generate code that could track cursor events within a visualization system, such as the duration of cursor hover in a view, click frequency of the SVG elements, or movement trajectories, which can be used to evaluate the usability of each view, with longer hover times and higher click frequencies indicating lower usability, and move-

ment patterns providing insights into whether the view layout was intuitive and effective.

P13 specifically mentioned using LLMs to generate reflection summaries based on user interview scripts `assistant`. Specifically, researchers can ask LLM to provide the key insights and provide with a concise summary of user feedback by analyzing the transcripts.

9 Analysis: Write

“Generally, LLMs facilitated academic writing by generating drafts and refining text, especially for novices. Challenges like inconsistent domain terminology and lack of contextual awareness were addressed by training LLMs on domain-specific papers and adopting multi-step writing processes.”

The “Write” stage focuses on using LLMs to assist in drafting academic papers, reports, and summaries. Novice students often leverage LLMs to generate papers based on high-level summaries they provided for each sentence, typically written in their native (non-English) language `assistant`. While this approach lowered the language barrier and helped novices produce drafts more quickly, it introduced several challenges. First, the generated text often lacked detail, insight, or unique understanding, as the LLM tended to directly translate the student’s input. Second, when using their native language to create summaries, LLM frequently struggled to capture the key points of a sentence, resulting in the misunderstanding of the user’s intention. Participants at intermediate, proficient, and expert levels took a more structured approach to writing. More specifically, they provided high-level summaries for entire sections, allowing the LLM to draft longer passages `assistant`. These drafts were then manually fine-tuned to ensure accuracy, coherence, and alignment with the overall paper. A common challenge across all expertise levels was the lack of contextual awareness for the whole paper, which led to inconsistent terminology and style compared to the rest of the paper. To address these issues, participants adopted a practice of feeding LLMs domain-specific papers to learn the appropriate writing style and refine the generated text for consistency.

Some frequent LLM users among visualization researchers pointed out the process of combining the two steps above into one `assistant`. Specifically, they first used reasoning-capable LLMs, such as DeepSeek or GPT-o1, to first generate a detailed outline for a section based on a high-level summary. These tools then expanded the outline into more detailed summaries iteratively, down to the level of individual sentences. Finally, advanced LLMs, such as GPT-4o, were used to craft each sentence based on the detailed summaries. This multi-step process allowed for greater precision and coherence in the final output.

However, some participants, such as P26, expressed reservations about relying on LLMs for academic writing. They argued that academic writing is a generative task, requiring the creation of original content (The leap from nothing to something) based on human-distilled knowledge—an area where LLMs are less effective. Instead, they viewed LLMs as being better suited for tasks like summarizing or reporting, where the model works with existing material to produce concise summaries. This perspective highlights that novice writers may benefit most from LLM-generated drafts, because they tend to use LLMs for tasks that align with the model’s strengths, such as summarization.

6 GUIDELINE SUMMARIZATION

We proposed a set of role-based guidelines for LLMs’ usage in design studies, which are as follows:

Connector: As an connector, LLMs can assist researchers in understanding domain-specific terminology and foundational concepts (*Learn*) and support collaboration by analyzing the expertise of potential collaborators to suggest suitable roles within a project (*Cast*). Also, LLMs can assist in summarizing complex domain-specific workflows and extracting design requirements (*Discover*). In this stage, LLMs can also be employed to verify collected design requirements, ensuring

that they align with the practical challenges of the domain. Finally, in the *Design* stage, LLMs can facilitate communication between visualization designers and domain experts by translating domain-specific knowledge into actionable design insights (*Design*).

Simulator: In their simulator roles, LLMs can simulate user behavior to uncover hidden use cases, alternative workflows, and potential usage scenarios (*Deploy*). They can also generate potential responses to questionnaires, enabling researchers to compare these with real participant feedback to identify unique insights (*Discover*). Additionally, LLMs can simulate reviewer feedback to identify potential limitations in research projects, such as unclear motivation or scalability issues (*Reflect*). They may also provide high-level suggestions for system-wide design tasks like layout and interaction logic as the basis for researchers to further refine and align with domain requirements (*Design*).

Programmer: LLMs in the programmer role can assist researchers in generating tailored coding tutorials and learning new technical stacks, such as emerging tools, to support self-training in new domains (*Learn*). LLMs can generate mock datasets to address delays in obtaining real data and enable researchers to continue prototyping (*Winnow*). Additionally, LLMs can create large-scale mock datasets to test system scalability and generate code for data cleaning and pre-processing tasks (*Implement*). For visual design, LLMs can help with generating and optimizing code for prototypes and initial design (*Design, Implement*). Furthermore, LLMs refine specific parts of prototypes and generate usability tracking code, such as cursor events and click frequency, to assess system usability and layout effectiveness (*Reflect*).

Assistant: As the assistant, LLMs can aid researchers in filling gaps in domain-specific terminology and foundational concepts (*Learn*) and support collaboration by evaluating visualization techniques or generating recruitment emails for collaborators (*Winnow*). They can help summarize complex workflows, extract design requirements (*Discover*). Additionally, LLMs can assist in translating domain-specific knowledge into actionable design insights, decomposing requirements into sub-tasks, and providing design inspiration (*Design*). LLMs can analyze performance bottlenecks, and summarize user feedback to evaluate system usability (*Deploy*). Finally, they can summarize user interviews, and assist in drafting and refining academic papers (*Reflect, Write*).

7 IMPLICATIONS

In this section, we introduced the future direction of LLM-assisted design study and limitations of this work.

7.1 Future Direction

We introduce the future direction of the LLM-assisted design study according to the distilled feedback from the qualitative interview process.

7.1.1 Toward a Project-Oriented LLM Copilot

Current LLMs provide fragmented and localized assistance in design study projects, lacking the ability to capture the entire workflow holistically. As discussed in Section 5, many of the challenges associated with LLM-assisted strategies are from their inability to maintain an understanding of the project’s global context. For example, LLMs often lose track of prior information when starting a new session, which needs users to repeatedly reintroduce the same content. Additionally, the narrative-driven prompts are often imprecise to LLMs, making it difficult for the model to fully grasp the actual user intention. If possible, by integrating with various technical stacks—such as Overleaf for writing, VS Code for coding, or Figma for design—these advanced LLMs could provide contextualized assistance through textual writing, coding, or visualization design. A future direction is to develop LLM systems that can naturally integrate into the entire design study workflow.

7.1.2 LLMs for Automated Task Decomposition

Another promising direction for LLM-assisted design studies is the development of agent-based LLMs which is capable of autonomously decomposing course-grained tasks. Instead of requiring users to manually break down tasks or provide detailed prompts, these advanced LLM could operate based on high-level commands, automatically decomposing them into sub-tasks and executing them seamlessly. For

example, given a set of collected design requirements, an agent-based LLM could independently perform task abstraction, and then it can generate the outputs, such as code from global visual representations (e.g., views) to detailed visual encodings (e.g., glyphs).

Furthermore, inspired by other domains, like chemistry, where multi-agent systems can already carry out autonomous research workflows. These systems can perform tasks such as hypothesis generation, experiment design, data collection, and analysis without requiring constant human oversight. However, compared to hard sciences like chemistry, visualization is inherently a “soft science” where human interpretation, creativity, and subjective judgment play a critical role. Thus, visualization often requires humans to remain in the loop to provide feedback.

7.1.3 Are Human Researchers Indispensable in Design Study Process?

While LLMs have made significant strength in assisting with various aspects of design studies, all participants commented that it is unlikely that they will fully replace human researchers in the foreseeable future. First, the process of brainstorming or assessing ideas cannot be entirely taken over by LLMs. Even though LLMs have a vast corpus of visualization knowledge, they struggle to determine whether an idea is actually interesting or not. Human researchers, on the other hand, leverage their subjective judgment, taste, and domain expertise to evaluate ideas based on higher-level abstractions built upon their accumulated knowledge in visualization.

Second, while LLMs excel at summarizing existing knowledge, they fall short in creative thinking. Although there have been advancements in enabling LLMs to generate creative outputs, the majority of their results remain generic and lack the originality required for groundbreaking insights. Tasks that needs critical thinking or deep reflection are areas where LLMs are less capable in, while human researchers can provide profound and unique perspectives that are important to advancing the field.

Finally, experts agreed that one of the most challenging parts of visualization design for LLMs is interaction design. The success or failure of an interaction is highly dependent on the subjective experiences of domain users, which vary significantly across different systems and tasks. Evaluating and refining interactions requires iterative collaboration between visualization researchers and real users, thus it remains an open challenge for LLMs.

7.2 Limitations

While this study provides valuable insights into the roles and applications of LLMs in design studies, there are still some limitations.

7.2.1 Trust and Usability Concerns

Although most participants in our study expressed a positive attitude towards LLMs and their potential to support design studies, there are also concerns regarding trust and usability regarding LLMs. For example, some participants, such as P21, highlighted the “dark side” of LLM usage, where users may think of these tools as unreliable for specific tasks. Future work should focus on understanding the reasons behind these negative perceptions and exploring how LLMs can be improved to address these concerns. Second, although we have spent every effort involving the participants who voluntarily joined from different regions, these participants may already have a predisposition toward using LLMs that do not represent the whole population of design study researchers. Third, while the paper proposes guidelines for LLM usage in design studies, it does not empirically validate these recommendations. Future work should involve testing these guidelines in real-world scenarios.

7.2.2 Rapid Evolution of LLMs

This paper focuses on the current generation of LLMs and how they are being used in design studies. However, it is obvious that LLMs are evolving rapidly, with continuous improvements in their capabilities and applications. Thus, some of the insights presented in this work may become outdated as newer and more advanced LLMs emerge. Despite

this limitation, we have provided deep discussions and practical guidelines for the design study field. For example, we provided potential research directions for better integrating LLMs into design studies in the future. Additionally, we have discussed why humans cannot be replaced by LLMs in the foreseeable future. These contributions provide value for future studies to build upon, even as LLM technology becomes increasingly advanced.

7.2.3 Limited Focus on Domain- and LLM-Specific Usage

This paper did not focus much on the detailed comparison of how LLMs are used across different domains or how various types of LLMs differ in their applications. For example, we did not conduct an in-depth comparison of how LLM usages are different between scientific domains (e.g., biomedical research or quantum computing) and application domains (e.g., smart cities or sports). Also, although the study focuses on visualization-specific design studies and cannot be directly applied to other fields, the human-centered research, such as HCI or UX design, can also yield the relevant benefits, like how to properly boost the human-AI collaboration, from this study. Similarly, we did not explore much about the differences in usage patterns between various types of LLMs (e.g., reasoning or coding). Future research could address these gaps by conducting comparative studies in each above contexts.

8 CONCLUSION

This paper investigated the strategies, challenges, and roles associated with integrating Large Language Models (LLMs) into the visualization-specific design study process. Through a multi-stage qualitative study involving 30 participants across various expertise levels and domains, we identified four primary roles for LLMs: *Connector*, *Simulator*, *Programmer*, and *Assistant*. Each role supports researchers differently across the nine stages of the design study methodology. Also, we also conducted post-study analysis to reveal how researchers treat LLM-assisted design study regarding the importance and difficulty of using LLMs. Moreover, we proposed the useful guidelines to benefits all domain researchers with different expertise, followed by the discussion of several implications for future research.

REFERENCES

- [1] G. Akbaba and N. Elmqvist. "Two Heads are Better than One": Pair-Interviews for Visualization. In *IEEE VIS*, 2023. doi: 10.1109/VIS54172.2023.00050 3
- [2] R. C. Basole and T. Major. Generative ai for visualization: Opportunities and challenges. *IEEE Computer Graphics and Applications*, 44(2):55–64, 2024. 2
- [3] J. Batch and N. Elmqvist. The interactive visualization gap in initial exploratory data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):104–113, 2018. doi: 10.1109/TVCG.2017.2743990 3
- [4] H. Chen et al. Designing the future: A case study on human-ai co-innovation. *Creative Education*, 15:645–658, 2024. 2
- [5] A. Chiarello and A. Author2. Generative large language models in engineering design: Opportunities and challenges. *Design Studies*, 78:101073, 2024. doi: 10.1016/j.destud.2024.101073 2
- [6] N. E. I. . C. Eric Newburger. An interview study on the role of visualization for inferential statistics. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):110–120, 2024. doi: 10.1109/TVCG.2023.3326521 3
- [7] H. Fill and A. Muff. Visualization in the era of artificial intelligence. *Jusletter IT*, 2023. 2
- [8] H.-G. Fill, P. Fettke, and J. Köpke. Conceptual modeling and large language models: impressions from first experiments with chatgpt. *Enterprise Modelling and Information Systems Architectures (EMISAJ)*, 18:1–15, 2023. 2
- [9] H.-G. Fill, F. Härer, I. Vasic, D. Bocard, B. Reitemeyer, F. Muff, S. Curty, and M. Bühlmann. Cmag: A framework for conceptual model augmented generative artificial intelligence. 2024. 2
- [10] A. Gatti and A. Author2. How chatgpt can inspire and improve serious board game design. *International Journal of Game-Based Learning*, 14(1):22–35, 2024. doi: 10.4018/IJGBL.2024010102 3
- [11] A. Gomez and A. Author2. Large language models in complex system design. In *Proceedings of the Design Society*, vol. 3, pp. 123–132, 2024. doi: 10.1017/pds.2024.13 2
- [12] J.-Y. He, Z.-Q. Cheng, C. Li, J. Sun, W. Xiang, X. Lin, X. Kang, Z. Jin, Y. Hu, B. Luo, et al. Wordart designer: user-driven artistic typography synthesis using large language models. *arXiv preprint arXiv:2310.18332*, 2023. 2
- [13] J. Hogan et al. The elicitation interview technique: Capturing people's experiences of data representations. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):804–813, 2016. doi: 10.1109/TVCG.2015.2511718 3
- [14] Y. Hou, M. Yang, H. Cui, L. Wang, J. Xu, and W. Zeng. C2ideas: Supporting creative interior color design ideation with a large language model. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2024. 2
- [15] M. Hutchinson, R. Jianu, A. Slingsby, and P. Madhyastha. Llm-assisted visual analytics: Opportunities and challenges. *arXiv preprint arXiv:2409.02691*, 2024. 2
- [16] A. Kim and A. Author2. Understanding large-language model (llm)-powered human-robot interaction. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1):1–23, 2024. doi: 10.1145/1234567 2
- [17] J. Kim et al. Visually-situated natural language understanding with contrastive reading model and frozen large language models. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):1684–1695, 2023. 2
- [18] J. Kim, S. Lee, H. Jeon, K.-J. Lee, H.-J. Bae, B. Kim, and J. Seo. Phenoflow: A human-llm driven visual analytics system for exploring large and complex stroke datasets. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2
- [19] N. W. Kim, H.-K. Ko, G. Myers, and B. Bach. Chatgpt in data visualization education: A student perspective. In *2024 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pp. 109–120. IEEE, 2024. 2
- [20] B. Kovalerchuk et al. Visual knowledge discovery with artificial intelligence: Challenges and future directions. *AI & Society*, 36(4):675–688, 2021. 2
- [21] H. Koziolok, S. Grüner, R. Hark, V. Ashiwal, S. Linsbauer, and N. Eskandani. Llm-based and retrieval-augmented control code generation. In *Proceedings of the 1st International Workshop on Large Language Models for Code*, pp. 22–29, 2024. 2
- [22] Y. Li, H. Xu, and F. Tian. From shots to stories: Llm-assisted video editing with unified language representations. *arXiv preprint arXiv:2505.12237*, 2025. 3
- [23] A. Liao and A. Author2. Designerly understanding: Information needs for model transparency to support design ideation for ai-powered user experience. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2024. doi: 10.1145/1234568 2
- [24] Z. Liu, X. Xie, M. He, W. Zhao, Y. Wu, L. Cheng, H. Zhang, and Y. Wu. Smartboard: Visual exploration of team tactics with llm agent. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2
- [25] J. Ma, Y. Deng, C. Chen, N. Du, H. Lu, and Z. Yang. Glyphdraw2: Automatic generation of complex glyph posters with diffusion models and large language models. *arXiv preprint arXiv:2407.02252*, 2024. 2
- [26] Y. Ma, Y. He, H. Wang, A. Wang, L. Shen, C. Qi, J. Ying, C. Cai, Z. Li, H.-Y. Shum, et al. Follow-your-click: Open-domain regional image animation via motion prompts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 6018–6026, 2025. 3
- [27] T. D. E. W. J. Mace. A qualitative interview study of distributed tracing visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 30(7):130–140, 2024. doi: 10.1109/TVCG.2023.3241596 3
- [28] D. Maddigan and M. Sušnjak. Chat2vis: Generating data visualizations via natural language. In *Proceedings of IEEE VIS*, pp. 456–467, 2022. 2
- [29] M. Meyer, M. Sedlmair, P. Quinan, and T. Munzner. Criteria for rigor in visualization design study. *IEEE transactions on visualization and computer graphics*, 26(1):87–97, 2019. 1
- [30] N. Narechania et al. Nl4dv: A toolkit for generating analytic specifications for data visualization from natural language queries. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):530–540, 2021. 2
- [31] A. Roza-Torres, C. J. Latorre-Rojas, and W. J. Sarmiento. Prompt engineering-based video prototyping for immersive interaction design: Limits, opportunities and perspectives. In *Iberoamerican Workshop on Human-Computer Interaction*, pp. 252–266. Springer, 2025. 3
- [32] S. Saha. Human-ai collaboration: Exploring interfaces for interactive machine learning. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2264–2275, 2022. 2

- [33] D. Schelble et al. Let's think together! assessing shared mental models, performance, and trust in human-agent teams. *ACM Transactions on Interactive Intelligent Systems*, 10(4):1–29, 2020. 2
- [34] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics*, 18(12):2431–2440, 2012. 1, 2, 4, 5
- [35] P. Shanbhag. Tewn: A prompt-based system for context-aware website generation. 2025. 3
- [36] L. Shen, H. Li, Y. Wang, and H. Qu. From data to story: Towards automatic animated data video creation with llm-based multi-agent systems. In *2024 IEEE VIS Workshop on Data Storytelling in an Era of Generative AI (GEN4DS)*, pp. 20–27. IEEE, 2024. 2
- [37] S. Shen, Z. Lin, W. Liu, C. Xin, W. Dai, S. Chen, X. Wen, and X. Lan. Dashchat: Interactive authoring of industrial dashboard design prototypes through conversation with llm-powered agents. *arXiv preprint arXiv:2504.12865*, 2025. 3
- [38] S. Shin, S. Hong, and N. Elmqvist. Visualizationary: Automating design feedback for visualization designers using llms. *arXiv preprint arXiv:2409.13109*, 2024. 2
- [39] F. K. Sufi. Ai-globalevents: A software for analyzing, identifying and explaining global events with artificial intelligence. *Software Impacts*, 11:100218, 2022. 2
- [40] A. Sun and A. Author2. Llms and diffusion models in ui/ux: Advancing human-computer interaction. *International Journal of Human-Computer Studies*, 150:102630, 2024. doi: 10.1016/j.ijhcs.2024.102630 2
- [41] A. Swanson and A. Author2. The virtual lab: Ai agents design new sars-cov-2 nanobodies with experimental validation. *Nature Biotechnology*, 42:123–130, 2024. doi: 10.1038/s41587-024-01234-5 2
- [42] J. Walny, C. Frisson, M. West, D. Kosminsky, S. Knudsen, S. Carpendale, and W. Willett. Data changes everything: Challenges and opportunities in data visualization design handoff. *IEEE transactions on visualization and computer graphics*, 26(1):12–22, 2019. 2
- [43] A. Wu, D. Deng, F. Cheng, Y. Wu, S. Liu, and H. Qu. In defence of visual analytics systems: Replies to critics. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1026–1036, 2022. 3
- [44] J. Xu, W. Du, X. Liu, and X. Li. Llm4workflow: An llm-based automated workflow model generation tool. In *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering*, pp. 2394–2398, 2024. 2
- [45] L. Xu and E. Wall. Exploring the capability of llms in performing low-level visual analytic tasks on svg data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 2
- [46] J. Zamfirescu-Pereira, E. Jun, M. Terry, Q. Yang, and B. Hartmann. Beyond code generation: Llm-supported exploration of the program design space. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2025. 3
- [47] Z. Zeng, W. Watson, N. Cho, S. Rahimi, S. Reynolds, T. Balch, and M. Veloso. Flowmind: automatic workflow generation with llms. In *Proceedings of the Fourth ACM International Conference on AI in Finance*, pp. 73–81, 2023. 2
- [48] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 1(2), 2023. 1
- [49] X. Zhao et al. Lightva: Lightweight visual analytics with llm agent-based task planning and execution. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2
- [50] Y. Zhao, Y. Zhang, Y. Zhang, X. Zhao, J. Wang, Z. Shao, C. Turkay, and S. Chen. Leva: Using large language models to enhance visual analytics. *IEEE transactions on visualization and computer graphics*, 2024. 2
- [51] E. Çay et al. Understanding user experience of covid-19 maps through remote elicitation interviews. 2020. doi: 10.36227/techrxiv.12912242 3