Prompting Generative AI with Interaction-Augmented Instructions

Leixian Shen
The Hong Kong University of Science
and Technology
Hong Kong SAR, China
lshenaj@connect.ust.hk

Haotian Li
Microsoft Research Asia
Beijing, China
haotian.li@microsoft.com

Yifang Wang
Northwestern University
Evanston, United States
yifang.wang@kellogg.northwestern.edu

Xing Xie
Microsoft Research Asia
Beijing, China
xingx@microsoft.com

Huamin Qu
The Hong Kong University of Science
and Technology
Hong Kong SAR, China
huamin@cse.ust.hk

ABSTRACT

The emergence of generative AI (GenAI) models, including large language models and text-to-image models, has significantly advanced the synergy between humans and AI with not only their outstanding capability but more importantly, the intuitive communication method with text prompts. Though intuitive, text-based instructions suffer from natural languages' ambiguous and redundant nature. To address the issue, researchers have explored augmenting text-based instructions with interactions that facilitate precise and effective human intent expression, such as direct manipulation. However, the design strategy of interaction-augmented instructions lacks systematic investigation, hindering our understanding and application. To provide a panorama of interaction-augmented instructions, we propose a framework to analyze related tools from why, when, who, what, and how interactions are applied to augment text-based instructions. Notably, we identify four purposes for applying interactions, including restricting, expanding, organizing, and refining text instructions. The design paradigms for each purpose are also summarized to benefit future researchers and practitioners.

CCS CONCEPTS

• **Human-centered computing** → **Interaction paradigms**.

KEYWORDS

Generative AI, Text Prompt, Interaction, Human-GenAI Interaction

1 INTRODUCTION

Empowering humans with AI collaborators has been widely investigated and applied in various fields, such as decision making [60], mass communication [54], creative design [81], and scientific discovery [38]. In human-AI collaboration, one critical challenge is to convey human intent, such as the goal and mode of collaboration, effectively and efficiently to AI [31]. Only by ensuring AI collaborators align with humans' intent can the collaboration be synergistic and productive [34]. As a result, methods for conveying human intent to AI attract broad research interest, from direct interactions [33, 41] to implicit understanding of human behavior [23, 52].

The recent emerging large-scale generative AI (GenAI) models, such as large language models (LLMs, *e.g.,* GPT-4o), and image or video generation models (*e.g.,* Sora), have considerably advanced this line of research. Humans can instruct these models with natural language-based text prompts. Though these instructions are intuitive and flexible, they also inherit the intrinsic disadvantages from natural languages, such as redundancy and ambiguity [55], which lowers the effectiveness and efficiency of human intent communication to AI collaborators. To compensate for the disadvantage, researchers and practitioners have widely explored how to leverage novel interactions to enhance text-only prompts. For example, one notable application of the design strategy is OpenAI Canvas [46], which lets users specify the functional scope of text prompts to LLMs through brushing. With such **interaction-augmented instructions**, human intent can be expressed in a more precise manner and thus AI collaborators can take more targeted actions. Other examples of interaction-augmented instructions include visual brush selection on images, flowchart manipulation, inline text annotations or highlighting, *etc.* Though the strategy of enhancing text prompts with interactions has been widely applied, we notice that there still lacks systematic examination and summary of its design paradigms, hindering the re-use and innovation of related techniques.

To fill the gap, we conducted a preliminary review of existing tools that support interaction-augmented instructions. Built on the well-recognized instrumental interaction model [8], we first formalize the role of interaction-augmented instructions in the communication between humans and GenAI with a GenAI-based instrumental interaction model. The model helps us to understand
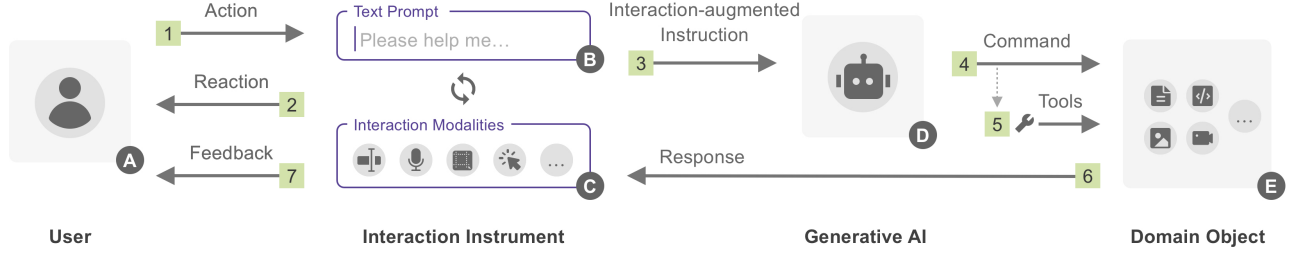
Figure 1: GenAI-based instrumental interaction model.

when interaction-augmented instructions are issued in the communication process, *who* drives the formation of these instructions, and *how* these instructions are applied on domain objects. Next, we further investigate the motivation (*why*) of applying interaction-augmented instructions, including restricting, expanding, organizing, and refining text instructions, and the detailed interaction types (*what*) to convey human intent. With the 4W1H framework, we analyze the existing tools that support interaction-augmented instructions and summarize the prominent design patterns (Appendix Tab. 1). Based on the insights, we identify a series of research opportunities and hope to further enrich our knowledge of enhancing intent communication through interaction-augmented interaction from multiple aspects, such as further understanding human agency and technical implementation behind these paradigms.

2 RELATED WORK

The rapid advancement of AI has not only facilitated the automation of numerous tasks but has also catalyzed human-AI communication and collaboration [12, 17, 45, 68, 76]. For example, Capel *et al.* [14] presented a research landscape of human-centered AI. Shen *et al.* [52] reviewed over 400 papers across multiple domains and proposed the bidirectional human-AI alignment framework. Rezwana and Maher [50] developed the Co-Creative Framework for Interaction design (COFI) and used the framework to categorize the interaction models in 92 existing human-AI co-creative systems. Moruzzi and Margarido [44] introduced a user-centered framework for human-AI co-creativity, defining key dimensions to understand the balance between user and system agency. However, these frameworks are coarse-grained and primarily emphasize high-level theoretical human-AI collaboration.

To achieve seamless human-AI collaboration, effective communication between humans and AI is essential. When communicating with GenAI, text prompts are the most commonly used approach. Due to the ambiguous and redundant nature of natural languages, prior studies have shown that crafting precise and effective prompts impose high requirements on humans' cognitive capabilities [7, 58, 63, 77]. To address the challenge, many studies incorporate interactions for creating effective GenAI prompts with minimum efforts (*e.g.,* [46, 65]). To summarize potential interaction designs, Lehmann *et al.* [33] investigated how different UIs influence users' access to the functional capabilities of GenAI models. Luera *et al.* [41] surveyed interface designs and patterns in GenAI applications, focusing on user-guided input modalities. Gao *et al.* [23] proposed a high-level taxonomy of human-LLM communication modes. However, these frameworks focus primarily

on summarizing what interactions are applied to enhance human-GenAI communication. Our work aims to dig deeper to characterize the interplay between those interactions and text prompts with a framework that focuses on *why* and *what* interactions are applied to enhance text prompts, and *who* drives the formation of interaction-augmented instructions. The framework also includes *when* these instructions were issued and *how* they function. Compared to prior research, our fine-grained framework can have stronger descriptive power [9] to summarize the design paradigms of existing tools and more generative power to point out research opportunities (Sec. 6).

3 GENAI-BASED INSTRUMENTAL INTERACTION MODEL

To understand the design paradigms of interaction-augmented instructions, we need to formalize the communication process between humans and GenAI. During the process, GenAI interprets humans' intention expressed in natural language and responds accordingly, either by taking direct actions or invoking external tools to operate on specific instruction objects when necessary.

Inspired by the report edited by Butler *et al.* [12], we realize that the interplay between humans, interaction instruments, GenAI, and target domain objects forms a GenAI-based version of the well-established *Instrumental Interaction Model* [8], where interaction instruments serve as mediators between users and domain objects. However, the original model relies on specialized instruments (*e.g.,* scrolling bar) which often provide a clear mapping from user actions to operations on target objects. The model is hard to fully capture the usage of GenAI for flexible interactions between humans and domain objects. With the participation of GenAI, users can leverage free-form text prompts as primary instruments to issue instructions with diverse intent, and GenAI will translate these instructions to executable operations on domain objects. Though flexible, these text prompts also suffer from the inherent ambiguity of natural language, To address the issue, researchers further explore enhancing text prompts with additional interactions, such as highlighting text segment with text brushing selection [46], to achieve more precise operations on domain objects. To describe the process, we propose a new *GenAI-based Instrumental Interaction Model* in Fig. 1.

In this model, interaction instruments serve as the method for human (A)-GenAI (D) communication to achieve operations on domain objects (E) based on user intent. Users act with instruments (1) via text prompts (B) and other interaction modalities (C), such as through text brushing, canvas clicking, widget selection, and textual inline annotation (all summarized interaction modalities are
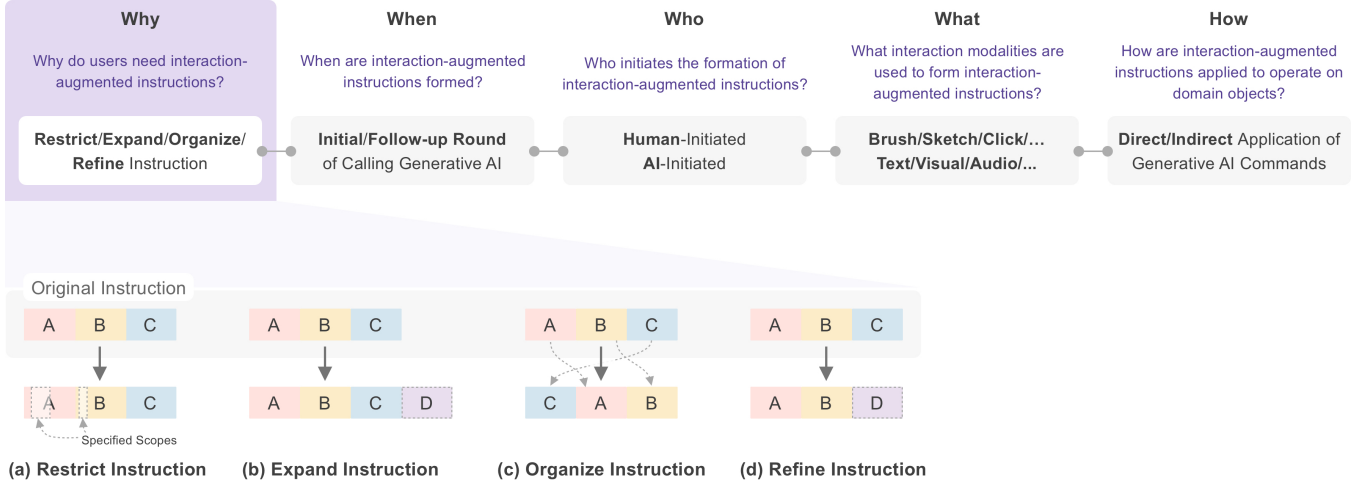
Figure 2: The 4W1H analytical framework of existing GenAI tools featuring interaction-augmented instructions. The "*Why*" dimension describes four impact types of interactions on instructions.

shown in Appendix Tab. 1), which collectively transform users' intents into interaction-augmented instructions for GenAI (3). GenAI, in turn, directly operates on domain objects (4), *e.g.,* generated text or images within GenAI's ability or produces commands to trigger other tools to operate on domain objects (5). Instruments also provide reactions (2) that allow users to refine their actions and provide feedback (7) as the instructions are executed by GenAI and applied to domain objects (6). In addition, the process is iterative.

4 FRAMEWORK

With a model to characterize the role of interaction-augmented instructions in human-GenAI communication, we propose a framework based on 4W1H—Why, When, Who, What, and How—to analyze existing interaction-augmented instruction paradigms (Fig. 2).

Why do users need interaction-augmented instructions? The impact of interaction-augmented instructions on domain objects varies based on user intents [50, 58, 68]. We categorized the impact of user interactions on original text prompts into four types (Fig. 2): (a) *Restrict Instruction*: interactions are applied to restrict the scope of AI actions (*e.g.,* a specific part of domain objects) when text prompts are vague; (b) *Expand Instruction*: interactions are applied to offer more guidance when text prompts provide limited information; (c) *Organize Instruction*: interactions are applied to structure information for AI's better understanding of free-form text prompts; (d) *Refine Instruction*: interactions are applied to improve the clarity of ambiguous text prompts.

When are interaction-augmented instructions formed? Interaction-augmented instructions can be formulated either in the *initial* or *follow-up* round of calling GenAI [7, 52]. In the initial round, users carefully prepare instructions based on application requirements before passing them to GenAI, *i.e.,* Fig. 1 (1 and 3). In the follow-up rounds, users make further decisions or improvements after receiving GenAI execution feedback, *i.e.,* Fig. 1 (6, 2 and 7).

Who initiates the formation of interaction-augmented instructions? The formulation of interactive prompts can be either *human-initiated* or *AI-initiated* [14, 58, 63]. In human-initiated mode (Fig. 1-A), users craft high-quality instructions to ensure GenAI understand the intent or refine the instructions based on GenAI decisions. In AI-initiated mode (Fig. 1-D), AI assists users by suggesting ways to specify better instructions or providing recommendations for improvement, followed by humans' final decision based on these suggestions, thus improving AI's explainability.

What interaction modalities are used to form interaction-augmented instructions? Domain objects (Fig. 1-E) in applications vary widely (*e.g.,* text, image, code, canvas, *etc.*) and can all be used to specify instructions [41]. Users can employ various actions (*e.g.,* clicking, selecting, brushing, sketching, typing, *etc.*) to generate additional interactions that augment the original text prompts (Fig. 1-B).

How are interaction-augmented instructions applied to operate on domain objects? The commands provided by GenAI based on interaction-augmented instructions can be *directly* or *indirectly* applied on domain objects. In direct mode, GenAI directly understands and executes commands to operate on domain objects (Fig. 1-4), *e.g.,* in writing and image editing tasks. In indirect mode, GenAI understands the instruction and calls other tools to operate on domain objects (Fig. 1-D), *e.g.,* utilizing a tool to generate animation effects on specific graphic elements.

5 CORPUS COLLECTION AND ANALYSIS

We applied the framework in Sec. 4 to analyze the design of existing tools with interaction-augmented instructions. The tools under analysis must adhere to the following criteria: (1) the system must be GenAI-based; (2) interaction with GenAI must involve text prompts; (3) apart from text prompts, at least one supplementary form of interaction is used to augment the text prompts. Our data collection process commenced by sourcing papers from key HCI conferences: CHI, UIST, CSCW, and IUI, using targeted keyword searches (*e.g.,* "GenAI", "Generative AI", "LLM", "Large Language Model", *etc.*) starting from November 2022 (following the release

of OpenAI's ChatGPT). Additionally, we curated a set of recent survey or position papers on human-GenAI interaction as seed papers [23, 35, 41, 52, 58, 78]. Subsequently, we analyzed the corpus of the seed papers along with their references to amplify our dataset. Finally, we collected 52 tools (see Appendix Tab. 1) with interaction-augmented instructions, covering various tasks like coding, writing, visualization authoring, image generation and editing, *etc.*

6 DESIGN PARADIGMS

We will discuss the identified patterns using our framework. These patterns can guide future human-GenAI interaction design and inspire the exploration of alternative paradigms. The discussion is organized primarily by the "Why" dimension, with further subdivision into four categories based on "When" and "Who": human-initiated instruction preparation, AI-initiated instruction preparation, human-initiated post-invocation enhancement, and AI-initiated post-invocation enhancement. Each category is analyzed with respect to corresponding "What" and "How".

6.1 Restrict Instruction

User text prompts are frequently vague when defining domain-specific objects of interest. Interactions allow users to restrict the scope of AI actions, aligning outputs more closely with human intents (29/52)[1].

Human-Initiated Instruction Preparation (14/29). These paradigms focus on instruction setup before invoking GenAI within a single interaction round. Users actively leverage diverse interaction modalities to restrict the scope of AI actions, typically interacting on user-uploaded domain objects. Interactive visual click selection is a widely adopted approach, enabling users to directly select target objects via mouse clicks (Fig. 3-a1), such as SVG elements [43], graph structure nodes [74, 79], and segmented image regions [39], which allows explicit processing by GenAI, and UI components on mobile canvas [24], which implicitly execute GenAI' decisions on the canvas. Similarly, several tools augment text prompts with preconfigured widget options, enabling users to restrict action scopes [10, 67]. Furthermore, DirectGPT [43] supports dragging visual elements and dropping them into text prompts. Visual brush selection is another common interaction method, used for controllable image generation [19], region-specific image editing [40], visual question answering [18], and implicit video editing [64]. In textual domains (*e.g.,* writing and coding), inline textual highlighting [28] and text brush selection [22, 49] are effective ways to restrict AI action scope. The selection of the text brush often triggers further widget configurations [22, 49].

Human-Initiated Post-Invocation Enhancement (4/29). In contrast to human-initiated instruction preparation paradigms, human-initiated post-invocation enhancement paradigms focus on multi-turn interactions, refining instructions after GenAI invocation while leveraging similar interaction modalities. In this context, the domain objects are outputs from a previous AI generation round, which users may wish to fine-tune or regenerate. Interactions can help restrict the scope for subsequent instructions [1, 5, 46, 70]. For example, OpenAI Canvas [46] (Fig. 3-a3) allows users to select portions of generated text or code for further querying, while

PromptCharm [70] enables users to select the image of interest from the generated set for subsequent exploration.

AI-Initiated Post-Invocation Enhancement (12/29). In these paradigms, users usually provide coarse-grained instructions, to further confirm users' intent, AI can generate fine-grained options for users' selection. In this way, the action space of AI can be restricted. The primary interaction modality is widget selection, which could be categorized into two types. One type is *accepting user feedback to guide subsequent human-GenAI interactions*. For example, in the binary selection space, InkSync [32] (Fig. 3-a4) uses inline highlighting to show all LLM-proposed modifications to user writing and provides an accept/dismiss dialogue menu. In the multi-class selection space, ConstitutionMaker [48] guides users through iterative dialogue with multiple-choice questions to customize personal chatbots, this paradigm is also common in other conversational tools [6, 11, 15, 25]. In the unknown selection space, Luminate [59] uses an initial LLM to generate task-related dimensions and widget configurations, enabling users to specify their intents, followed by another LLM to fulfill the requirements. Unlike the above explicit human-GenAI interaction, LangAware [16] offers fine-grained options for natural language interaction on mobile devices. Users confirm these options and implicitly execute them on connected sensors after LLM decision-making. The other type is *providing customization options for final outputs without additional GenAI invocation* [3, 62, 65]. For example, DynaVis [65] dynamically generates rich widgets for users to configure visualizations. Alternatively, AI performs most tasks, leaving users with simple review and confirmation, such as PDFChatAnnotator [62], which extracts user-specified PDF content as cards for review.

6.2 Expand Instruction

For complex tasks, user text prompts may lack sufficient detail or breadth. User interactions can expand instructions to guide GenAI more effectively (19/52).

Human-Initiated Instruction Preparation (9/19). In these paradigms, users expand initial text prompts through interactions to provide GenAI with richer context or clearer instructions for complex tasks. The interaction modalities in this category are often more complex and task-specific. For instance, Tableau AI [61] integrates a chatbot with Tableau's traditional WIMP interface, where all real-time interactions with visualizations provide contextual input for the chatbot. G-VOILA [71] and XR-Objects [20] enrich text queries with environmental context through viewpoint navigation using augmented reality glasses. Tools like ChatScratch [15] and DrawTalking [51] use sketches to augment text prompts for creative drawing tasks. FigurA11y [57] incorporates a set of selectable guidelines for generating figure alt text. For more complex outputs (*e.g.,* video), users often rely on extended instructions with implicit integration of auxiliary tools. For example, NotePlayer [47] enhances the context for tutorial video generation by allowing inline annotations in notebooks. LAVE [66] enables users to select a series of relevant videos as input material for text-driven video editing. Data Playwright [53] (Fig. 3-b1) encodes data video authoring commands as inline annotations within narrations, enriching the context of both narrative and authoring intents.

---

[1]This format indicates that 29 tools out of 52 are under this category.

Figure 3: Examples of interaction-augmented instructions, displayed by four impact types of interaction on instructions ("Why") and four sub-categories based on "When" and "Who" in Sec. 6. In each example, 💡 denotes the core interaction design, while 🔍 indicates the absence of typical paradigm instances identified in our paper, leaving for future exploration. The examples are from (a1) [43], (a3) [46], (a4) [32], (b1) [53], (b2) [72], (b3) [38], (b4) [80], (c1) [75], (d1) [70], and (d2) [4].

AI-Initiated Instruction Preparation (1/19). When users struggle to extend instructions—especially due to unfamiliarity with task-specific options or the overwhelming number of input candidates—AI-initiated instruction preparation paradigms can effectively recommend instruction candidates (*e.g.,* examples in text prompts), streamlining the preparation process. For instance, ScatterShot [72] (Fig. 3-b2) builds high-quality demonstration sets for LLMs by iteratively segmenting unlabeled user data into task-specific patterns for human review and selection.

Human-Initiated Post-Invocation Enhancement (5/19). Exploratory tasks often require multiple rounds of interaction with GenAI to achieve the desired outcome. Flowchart manipulation is a common interaction modality that helps users organize their thoughts and iteratively refine instructions [4, 29, 38, 60]. For example, CoQuest [38] (Fig. 3-b3) facilitates LLM-based research question exploration by generating a series of idea nodes from high-level text prompts. Users can explore each node in depth, providing feedback to guide subsequent iterations. Sensecape [60] enables users to extract key points from LLM-generated text outputs and automatically organize them into multi-level representations, facilitating navigation and exploration across complex information spaces.

AI-Initiated Post-Invocation Enhancement (4/19). In exploratory tasks, some tools utilize AI to guide users in expanding

subsequent instructions. They leverage AI to break down tasks, provide structure, and suggest further actions for enhanced task execution. For instance, ExploreLLM [42] automatically decomposes user text prompts into sub-tasks, creating nodes that represent each sub-task, and allowing users to specify personal contexts and preferences for further exploration. Similarly, VISAR [80] (Fig. 3-b4) automatically recommends possible key aspects of writing for users to select and further explore, while GAM [37] recommends data analysis utterance in spreadsheets. Graphologue [27] generates inline annotated LLM responses and parses them into diagrams to facilitate user exploration and decision-making.

6.3 Organize Instruction

User-provided text prompts are often numerous and unstructured, hindering effective task and instruction interpretation. Structuring these prompts into organized formats can enhance task comprehension and execution (6/52).

Human-Initiated Instruction Preparation (6/6). Paradigms in this group focus on user-driven efforts to organize instructions before GenAI invocation. Among these formats, tree structures are widely used to represent hierarchical relationships and logical flows [21, 73, 75, 80]. For example, CoPrompt [21] introduces a tree-based representation to organize shared coding prompts among

collaborators, supporting referencing, requesting, sharing, and linking prompts seamlessly. Similarly, CoLadder [75] (Fig. 3-c1) applies a tree structure to organize hierarchical coding tasks. VISAR [80] adopts a tree structure to aid writers in brainstorming and revising hierarchical writing goals, streamlining the creative process. Low-code LLM [13] utilizes flowcharts to help users plan complex tasks step by step, ensuring clarity in task execution. Additionally, Kim et al. [30] propose a modular framework comprising cells, generators, and lenses, allowing users to organize LLM applications flexibly. It supports various widgets for configuring parameters.

6.4 Refine Instruction

User text prompts are often imprecise, resulting in suboptimal model outputs and necessitating additional rounds of interaction. Users can clarify their instructions by incorporating interactive refinement mechanisms, improving GenAI accuracy and reducing unnecessary iterations (3/52).

Human-Initiated Instruction Preparation (2/3). In this category, users actively explore and refine their prompts to ensure better alignment between user intent and model understanding. For instance, PromptCharm [70] (Fig. 3-d1) enables users to refine text-to-image prompts by exploring a large database of styles. Users can search, compare, and modify styles through interactive lists. DreamSheets [2] leverages spreadsheet-based manipulation to quickly enumerate different prompt variations for text-to-image generation, enabling users to generate all variations simultaneously and efficiently explore the generated space.

AI-Initiated Instruction Preparation (2/3). Paradigms in this group involve tools proactively guiding users by suggesting optimal prompts. For instance, autocomplete features during text input [4] (Fig. 3-d2) assist users by completing partial phrases or suggesting contextually relevant continuations. Additionally, candidate recommendations [70] present users with optimized options after initial input, helping them construct more precise and effective prompts.

7 CONCLUSION AND FUTURE WORK

This paper provides a preliminary analysis and summary of the design paradigms for interaction-augmented instructions in existing GenAI tools, offering guidance for future human-GenAI interaction design and inspiring the exploration of additional paradigms. In the future, we plan to further investigate the following directions.
Explore More Paradigms for Interaction-Augmented Instructions. Our framework offers a starting point to explore interaction-augmented instructions. Fig. 3 presents an initial two-dimensional design space and corresponding examples. Several subspaces remain unexplored, offering promising directions for future investigation. However, it is also worth considering whether some subspaces are better suited for purely text-based prompts or traditional AI models. Theoretically, any interaction-enhanced instruction could be replicated by optimizing text prompts. Therefore, carefully evaluating the trade-offs between text-only prompts and interaction-augmented instructions is essential [63]. For example, when specifying domain objects, interactive selection methods like clicking, brushing, or lassoing on a cluttered canvas can be more intuitive, accurate, and efficient than natural language [36, 69]. However, in

selections with semantic features or between overlapping elements, natural language can offer greater efficiency and clarity [53, 56].
Build GenAI Applications with Interaction-Augmented Instructions. Our "Why" dimension highlights four key challenges users often face when crafting text prompts: overly broad scopes, limited guidance, disorganized structures, and imprecise expressions. Correspondingly, we have summarized how interaction can help restrict, expand, organize, and refine original text prompts. When designing GenAI applications, our framework and identified patterns can guide decisions about whether to incorporate interaction-augmented instructions and select appropriate interaction paradigms [63]. Moreover, effective GenAI applications can integrate multiple paradigms across different stages of interaction to address diverse user needs.
Understand Human Agency and Technical Implementation Behind Interaction-Augmented Instructions. We observe that different paradigms of interaction-augmented instructions involve different technical implementation to achieve varying levels of human agency and AI automation [26]. Through analysis with our framework, we identify different interaction granularities and mediums, reflecting variations in human involvement and influencing the underlying AI capabilities required. For example, indirect operations on domain objects require AI to select and apply additional tools. It elevates the automation level but may diminish users' agency. To make these decisions, advanced planning and tool application capabilities can be necessary for these AI models. While our current study focuses on interactions, future research is necessary to explore the interplay among interaction, human agency, and technical implementation, offering systematic guidance and further directions for enhancing human-GenAI interaction.

REFERENCES

[1] Emre Aksan, Fabrizio Pece, and Otmar Hilliges. 2018. DeepWriting: Making Digital Ink Editable via Deep Generative Modeling. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 1–14. https://doi.org/10.1145/3173574.3173779
[2] Shm Garanganao Almeda, J.D. Zamfirescu-Pereira, Kyu Won Kim, Pradeep Mani Rathnam, and Bjoern Hartmann. 2024. Prompting for Discovery: Flexible Sense-Making for AI Art-Making with Dreamsheets. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–17. https://doi.org/10.1145/3613904.3642858
[3] Peter Andrews, Oda Elise Nordberg, Stephanie Zubicueta Portales, Njål Borch, Frode Guribye, Kazuyuki Fujita, and Morten Fjeld. 2024. AiCommentator: A Multimodal Conversational Agent for Embedded Visualization in Football Viewing. In *Proceedings of the 29th International Conference on Intelligent User Interfaces, IUI'24*. ACM, 14–34. https://doi.org/10.1145/3640543.3645197
[4] Tyler Angert, Miroslav Suzara, Jenny Han, Christopher Pondoc, and Hariharan Subramonyam. 2023. Spellburst: A Node-based Interface for Exploratory Creative Coding with Natural Language Prompts. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–22. https://doi.org/10.1145/3586183.3606719
[5] Anysphere. 2023. CURSOR. https://www.cursor.com/
[6] Seungho Baek, Hyerin Im, Jiseung Ryu, Juhyeong Park, and Takyeon Lee. 2023. PromptCrafter: Crafting Text-to-Image Prompt through Mixed-Initiative Dialogue with LLM. In *Proceedings of the ICML Workshop on Artificial Intelligence & Human Computer Interaction*. 1–5. arXiv:2307.08985
[7] Gagan Bansal, Jennifer Wortman Vaughan, Saleema Amershi, Eric Horvitz, Adam Fourney, Hussein Mozannar, Victor Dibia, and Daniel S Weld. 2024. Challenges in Human-Agent Communication. *arXiv preprint arXiv:2412.10380* (2024), 1–27. https://doi.org/10.48550/arXiv.2412.10380
[8] Michel Beaudouin-Lafon. 2000. Instrumental Interaction: An Interaction Model for Designing Post-WIMP User Interfaces. In *Proceedings of the CHI conference on Human Factors in Computing Systems*. ACM, 446–453. https://doi.org/10.1145/332040.332473
[9] Michel Beaudouin-Lafon. 2004. Designing Interaction, Not Interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. 15–22. https:

//doi.org/10.1145/989863.989865

[10] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–14. https://doi.org/10.1145/3586183.3606725

[11] Victor S. Bursztyn, Jennifer Healey, Eunyee Koh, Nedim Lipka, and Larry Birnbaum. 2021. Developing a Conversational Recommendation Systemfor Navigating Limited Options. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–6. https://doi.org/10.1145/3411763.3451596

[12] Jenna Butler, Mihaela Vorvoreanu, Rebecca Janßen, Abigail Sellen, Nicole Immorlica, Brent Hecht, and Jaime Teevan. 2024. *Microsoft New Future of Work Report 2024.* Technical Report. 1–52 pages.

[13] Yuzhe Cai, Shaoguang Mao, Wenshan Wu, Zehua Wang, Yaobo Liang, Tao Ge, Chenfei Wu, WangYou WangYou, Ting Song, Yan Xia, Nan Duan, and Furu Wei. 2024. Low-code LLM: Graphical User Interface over Large Language Models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 3: System Demonstrations)*. ACL, 12–25. https://doi.org/10.18653/v1/2024.naacl-demo.2

[14] Tara Capel and Margot Brereton. 2023. What is Human-Centered about Human-Centered AI? A Map of the Research Landscape. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 1–23. https://doi.org/10.1145/3544548.3580959

[15] Liuqing Chen, Shuhong Xiao, Yunnong Chen, Yaxuan Song, Ruoyu Wu, and Lingyun Sun. 2024. ChatScratch: An AI-Augmented System Toward Autonomous Visual Programming Learning for Children Aged 6-12. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–19. https://doi.org/10.1145/3613904.3642229

[16] Weihao Chen, Chun Yu, Huadong Wang, Zheng Wang, Lichen Yang, Yukun Wang, Weinan Shi, and Yuanchun Shi. 2023. From Gap to Synergy: Enhancing Contextual Understanding through Human-Machine Collaboration in Personalized Systems. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–15. https://doi.org/10.1145/3586183.3606741

[17] Xiang Anthony Chen, Jeff Burke, Ruofei Du, Matthew K. Hong, Jennifer Jacobs, Philippe Laban, Dingzeyu Li, Nanyun Peng, Karl D. D. Willis, Chien-Sheng Wu, and Bolei Zhou. 2023. Next Steps for Human-Centered Generative AI: A Technical Perspective. *arXiv preprint arXiv:2306.15774* (2023). https://doi.org/10.48550/arXiv.2306.15774

[18] Kiroong Choe, Chaerin Lee, Soohyun Lee, Jiwon Song, Aeri Cho, Nam Wook Kim, and Jinwook Seo. 2024. Enhancing Data Literacy On-demand: LLMs as Guides for Novices in Chart Interpretation. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–17. https://doi.org/10.1109/TVCG.2024.3413195

[19] John Joon Young Chung and Eytan Adar. 2023. PromptPaint: Steering Text-to-Image Generation Through Paint Medium-like Interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–17. https://doi.org/10.1145/3586183.3606777

[20] Mustafa Doga Dogan, Eric J. Gonzalez, Karan Ahuja, Ruofei Du, Andrea Colaço, Johnny Lee, Mar Gonzalez-Franco, and David Kim. 2024. Augmented Object Intelligence with XR-Objects. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–15. https://doi.org/10.1145/3654777.3676379

[21] Li Feng, Ryan Yen, Yuzhe You, Mingming Fan, Jian Zhao, and Zhicong Lu. 2024. CoPrompt: Supporting Prompt Sharing and Referring in Collaborative Natural Language Programming. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–21. https://doi.org/10.1145/3613904.3642212

[22] Raymond Fok, Joseph Chee Chang, Tal August, Amy X Zhang, and Daniel S Weld. 2024. Qlarify: Recursively Expandable Abstracts for Dynamic Information Retrieval over Scientific Papers. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–21. https://doi.org/10.1145/3654777.3676397

[23] Jie Gao, Simret Araya Gebreegziabher, Kenny Tsu Wei Choo, Toby Jia-Jun Li, Simon Tangi Perrault, and Thomas W. Malone. 2024. A Taxonomy for Human-LLM Interaction Modes: An Initial Exploration. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–11. https://doi.org/10.1145/3613905.3650780

[24] Weiwei Gao, Kexin Du, Yujia Luo, Weinan Shi, Chun Yu, and Yuanchun Shi. 2024. EasyAsk: An In-App Contextual Tutorial Search Assistant for Older Adults with Voice and Touch Inputs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–27. https://doi.org/10.1145/3678516

[25] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing using Language Models. In *Designing Interactive Systems Conference*. ACM, 1002–1019. https://doi.org/10.1145/3532106.3533533

[26] Jeffrey Heer. 2019. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences* 116, 6 (2019), 1844–1850. https://doi.org/10.1073/pnas.1807184115

[27] Peiling Jiang, Jude Rayan, Steven P. Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and*

Technology. ACM, 1–20. https://doi.org/10.1145/3586183.3606737

[28] Jeesu Jung, Hyein Seo, Sangkeun Jung, Riwoo Chung, Hwijung Ryu, and Du-Seong Chang. 2023. Interactive User Interface for Dialogue Summarization. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. ACM, 934–957. https://doi.org/10.1145/3581641.3584057

[29] Jeongyeon Kim, Sangho Suh, Lydia B Chilton, and Haijun Xia. 2023. Metaphorian: Leveraging Large Language Models to Support Extended Metaphor Creation for Science Writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, 115–135. https://doi.org/10.1145/3563657.3595996

[30] Tae Soo Kim, Yoonjoo Lee, Minsuk Chang, and Juho Kim. 2023. Cells, Generators, and Lenses: Design Framework for Object-Oriented Interaction with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–18. https://doi.org/10.1145/3586183.3606833

[31] Yoonsu Kim, Kihoon Son, Seoyoung Kim, and Juho Kim. 2024. Beyond Prompts: Learning from Human Communication for Enhanced AI Intent Alignment. In *Proceedings of the CHI 2024 Getting Back Together Workshop*. 1–4. arXiv:2405.05678

[32] Philippe Laban, Jesse Vig, Marti Hearst, Caiming Xiong, and Chien-Sheng Wu. 2024. Beyond the Chat: Executable and Verifiable Text-Editing with LLMs. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–23. https://doi.org/10.1145/3654777.3676419

[33] Florian Lehmann and Daniel Buschek. 2024. Functional Flexibility in Generative AI Interfaces: Text Editing with LLMs through Conversations, Toolbars, and Prompts. *arXiv preprint arXiv:2410.10644* (2024), 1–54. https://doi.org/10.48550/arXiv.2410.10644

[34] Haotian Li, Yun Wang, Q Vera Liao, and Huamin Qu. 2023. Why is AI not a Panacea for Data Workers? An Interview Study on Human-AI Collaboration in Data Storytelling. *arXiv preprint arXiv:2304.08366* (2023). https://doi.org/10.48550/arXiv.2304.08366

[35] Haotian Li, Yun Wang, and Huamin Qu. 2024. Where Are We So Far? Understanding Data Storytelling Tools from the Perspective of Human-AI Collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–19. https://doi.org/10.1145/3613904.3642726

[36] Yanna Lin, Haotian Li, Leni Yang, Aoyu Wu, and Huamin Qu. 2023. InkSight: Leveraging Sketch Interaction for Documenting Chart Findings in Computational Notebooks. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2023), 944–954. https://doi.org/10.1109/TVCG.2023.3327170

[37] Michael Xieyang Liu, Advait Sarkar, Carina Negreanu, Benjamin Zorn, Jack Williams, Neil Toronto, and Andrew D. Gordon. 2023. "What It Wants Me To Say": Bridging the Abstraction Gap Between End-User Programmers and Code-Generating Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 1–31. https://doi.org/10.1145/3544548.3580817

[38] Yiren Liu, Si Chen, Haocong Cheng, Mengxia Yu, Xiao Ran, Andrew Mo, Yiliu Tang, and Yun Huang. 2024. How AI Processing Delays Foster Creativity: Exploring Research Question Co-Creation with an LLM-based Agent. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–25. https://doi.org/10.1145/3613904.3642698

[39] Zhaoyang Liu, Yinan He, Wenhai Wang, Weiyun Wang, Yi Wang, Shoufa Chen, Qinglong Zhang, Zeqiang Lai, Yang Yang, Qingyun Li, Jiashuo Yu, Kunchang Li, Zhe Chen, Xue Yang, Xizhou Zhu, Yali Wang, Limin Wang, Ping Luo, Jifeng Dai, and Yu Qiao. 2023. InternGPT: Solving Vision-Centric Tasks by Interacting with ChatGPT Beyond Language. *arXiv preprint arXiv:2305.05662* (2023), 1–17. https://doi.org/10.48550/arXiv.2305.05662

[40] Zichen Liu, Yue Yu, Hao Ouyang, Qiuyu Wang, Ka Leong Cheng, Wen Wang, Zhiheng Liu, Qifeng Chen, and Yujun Shen. 2024. MagicQuill: An Intelligent Interactive Image Editing System. *arXiv preprint arXiv:2411.09703* (2024), 1–15. https://doi.org/10.48550/arXiv.2411.09703

[41] Reuben Luera, Ryan A Rossi, Alexa Siu, Sungchul Kim, and Xiang Chen. 2024. Survey of User Interface Design and Interaction Techniques in Generative AI Applications. *arXiv preprint arXiv:2410.22370* (2024), 1–42. https://doi.org/10.48550/arXiv.2410.22370

[42] Xiao Ma, Swaroop Mishra, Ariel Liu, Sophie Ying Su, Jilin Chen, Chinmay Kulkarni, Heng-Tze Cheng, Quoc Le, and Ed Chi. 2024. Beyond ChatBots: ExploreLLM for Structured Thoughts and Personalized Model Responses. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–12. https://doi.org/10.1145/3613905.3651093

[43] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. DirectGPT: A Direct Manipulation Interface to Interact with Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–16. https://doi.org/10.1145/3613904.3642462

[44] Caterina Moruzzi and Solange Margarido. 2024. Customizing the balance between user and system agency in human-AI co-creative processes. In *Proceedings of the 15th International Conference on Computational Creativity, ICCC'24*. 1–10.

[45] Florian Floyd Mueller, Nathan Semertzidis, Josh Andres, Joe Marshall, Steve Benford, Xiang Li, Louise Matjeka, and Yash Mehta. 2023. Toward Understanding the Design of Intertwined Human–Computer Integrations. *ACM Transactions on Computer-Human Interaction* 30, 5 (2023), 1–45. https://doi.org/10.1145/3590766

[46] OpenAI. 2024. Canvas. https://openai.com/index/introducing-canvas/

[47] Yang Ouyang, Leixian Shen, Yun Wang, and Quan Li. 2024. NotePlayer: Engaging Computational Notebooks for Dynamic Presentation of Analytical Processes. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–20. https://doi.org/10.1145/3654777.3676410

[48] Savvas Petridis, Benjamin D. Wedin, James Wexler, Mahima Pushkarna, Aaron Donsbach, Nitesh Goyal, Carrie J. Cai, and Michael Terry. 2024. Constitution-Maker: Interactively Critiquing Large Language Models by Converting Feedback into Principles. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. ACM, 853–868. https://doi.org/10.1145/3640543.3645144

[49] Mohi Reza, Nathan M Laundry, Ilya Musabirov, Peter Dushniku, Zhi Yuan "Michael" Yu, Kashish Mittal, Tovi Grossman, Michael Liut, Anastasia Kuzminykh, and Joseph Jay Williams. 2024. ABScribe: Rapid Exploration & Organization of Multiple Writing Variations in Human-AI Co-Writing Tasks using Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–18. https://doi.org/10.1145/3613904.3641899

[50] Jeba Rezwana and Mary Lou Maher. 2023. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *ACM Transactions on Computer-Human Interaction* 30, 5 (2023), 1–28. https://doi.org/10.1145/3519026

[51] Karl Toby Rosenberg, Rubaiat Habib Kazi, Li-Yi Wei, Haijun Xia, and Ken Perlin. 2024. DrawTalking: Building Interactive Worlds by Sketching and Speaking. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–25. https://doi.org/10.1145/3654777.3676334

[52] Hua Shen, Tiffany Knearem, Reshmi Ghosh, Kenan Alkiek, Kundan Krishna, Yachuan Liu, Ziqiao Ma, Savvas Petridis, Yi-Hao Peng, Li Qiwei, Sushrita Rakshit, Chenglei Si, Yutong Xie, Jeffrey P. Bigham, Frank Bentley, Joyce Chai, Zachary Lipton, Qiaozhu Mei, Rada Mihalcea, Michael Terry, Diyi Yang, Meredith Ringel Morris, Paul Resnick, and David Jurgens. 2024. Towards Bidirectional Human-AI Alignment: A Systematic Review for Clarifications, Framework, and Future Directions. *arXiv preprint arXiv:2406.09264* (2024). https://doi.org/10.48550/arXiv.2406.09264

[53] Leixian Shen, Haotian Li, Yun Wang, Tianqi Luo, Yuyu Luo, and Huamin Qu. 2024. Data Playwright: Authoring Data Videos With Annotated Narration. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–14. https://doi.org/10.1109/TVCG.2024.3477926

[54] Leixian Shen, Haotian Li, Yun Wang, and Huamin Qu. 2025. Reflecting on Design Paradigms of Animated Data Video Tools. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, 1–21. https://doi.org/10.1145/3706598.3713449

[55] Leixian Shen, Enya Shen, Yuyu Luo, Xiaocong Yang, Xuming Hu, Xiongshuai Zhang, Zhiwei Tai, and Jianmin Wang. 2023. Towards Natural Language Interfaces for Data Visualization: A Survey. *IEEE Transactions on Visualization and Computer Graphics* 29, 6 (2023), 3121–3144. https://doi.org/10.1109/TVCG.2022.3148007

[56] Leixian Shen, Yizhi Zhang, Haidong Zhang, and Yun Wang. 2024. Data Player: Automatic Generation of Data Videos with Narration-Animation Interplay. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 109–119. https://doi.org/10.1109/TVCG.2023.3327197

[57] Nikhil Singh, Lucy Lu Wang, and Jonathan Bragg. 2024. FigurA11y: AI Assistance for Writing Scientific Alt Text. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. ACM, 886–906. https://doi.org/10.1145/3640543.3645212

[58] Hariharan Subramonyam, Roy Pea, Christopher Pondoc, Maneesh Agrawala, and Colleen Seifert. 2024. Bridging the Gulf of Envisioning: Cognitive Challenges in Prompt Based Interactions with LLMs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI'24*. ACM, 1–19. https://doi.org/10.1145/3613904.3642754

[59] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured Generation and Exploration of Design Space with Large Language Models for Human-AI Co-Creation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–26. https://doi.org/10.1145/3613904.3642400

[60] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–18. https://doi.org/10.1145/3586183.3606756

[61] Tableau. 2024. Tableau AI. https://www.tableau.com/products/artificial-intelligence{#}content-530266

[62] Yi Tang, Chia-Ming Chang, and Xi Yang. 2024. PDFChatAnnotator: A Human-LLM Collaborative Multi-Modal Data Annotation Tool for PDF-Format Catalogs. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. ACM, 419–430. https://doi.org/10.1145/3640543.3645174

[63] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–24. https://doi.org/10.1145/3613904.3642902

[64] Bekzat Tilekbay, Saelyne Yang, Michal Adam Lewkowicz, Alex Suryapranata, and Juho Kim. 2024. ExpressEdit: Video Editing with Natural Language and Sketching. In *Proceedings of the 29th International Conference on Intelligent User Interfaces, IUI'24*. ACM, 515–536. https://doi.org/10.1145/3640543.3645164

[65] Priyan Vaithilingam, Elena L. Glassman, Jeevana Priya Inala, and Chenglong Wang. 2024. DynaVis: Dynamically Synthesized UI Widgets for Visualization Editing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–17. https://doi.org/10.1145/3613904.3642639

[66] Bryan Wang, Yuliang Li, Zhaoyang Lv, Haijun Xia, Yan Xu, and Raj Sodhi. 2024. LAVE: LLM-Powered Agent Assistance and Language Augmentation for Video Editing. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. ACM, 699–714. https://doi.org/10.1145/3640543.3645143

[67] Chenglong Wang, John Thompson, and Bongshin Lee. 2024. Data Formulator: AI-powered Concept-driven Visualization Authoring. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 1128 – 1138. https://doi.org/10.1109/TVCG.2023.3326585

[68] Jiayin Wang, Weizhi Ma, Peijie Sun, Min Zhang, and Jian-Yun Nie. 2024. Understanding User Experience in Large Language Model Interactions. *arXiv preprint arXiv:2401.08329* (2024), 1–20. https://doi.org/10.48550/arXiv.2401.08329

[69] Yun Wang, Leixian Shen, Zhengxin You, Xinhuan Shu, Bongshin Lee, John Thompson, Haidong Zhang, and Dongmei Zhang. 2024. WonderFlow: Narration-Centric Design of Animated Data Videos. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–17. https://doi.org/10.1109/TVCG.2024.3411575

[70] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–21. https://doi.org/10.1145/3613904.3642803

[71] Zeyu Wang, Yuanchun Shi, Yuntao Wang, Yuchen Yao, Kun Yan, Yuhan Wang, Lei Ji, Xuhai Xu, and Chun Yu. 2024. G-VOILA: Gaze-Facilitated Information Querying in Daily Scenarios. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–33. https://doi.org/10.1145/3659623

[72] Sherry Wu, Hua Shen, Daniel S. Weld, Jeffrey Heer, and Marco Tulio Ribeiro. 2023. ScatterShot: Interactive In-context Example Curation for Text Transformation. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. ACM, 353–367. https://doi.org/10.1145/3581641.3584059

[73] Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J. Cai. 2022. PromptChainer: Chaining Large Language Model Prompts through Visual Programming. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–10. https://doi.org/10.1145/3491101.3519729

[74] Liwenhan Xie, Chengbo Zheng, Haijun Xia, Huamin Qu, and Chen Zhu-Tian. 2024. WaitGPT: Monitoring and Steering Conversational LLM Agent in Data Analysis with On-the-Fly Code Visualization. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–14. https://doi.org/10.1145/3654777.3676374

[75] Ryan Yen, Jiawen Stefanie Zhu, Sangho Suh, Haijun Xia, and Jian Zhao. 2024. CoLadder: Manipulating Code Generation via Multi-Level Blocks. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–20. https://doi.org/10.1145/3654777.3676357

[76] Nur Yildirim, Mahima Pushkarna, Nitesh Goyal, Martin Wattenberg, and Fernanda Viégas. 2023. Investigating How Practitioners Use Human-AI Guidelines: A Case Study on the People + AI Guidebook. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 1–13. https://doi.org/10.1145/3544548.3580900

[77] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 1–21. https://doi.org/10.1145/3544548.3581388

[78] Chaoyun Zhang, Shilin He, Jiaxu Qian, Bowen Li, Liqun Li, Si Qin, Yu Kang, Minghua Ma, Guyue Liu, Qingwei Lin, Saravan Rajmohan, Dongmei Zhang, and Qi Zhang. 2024. Large Language Model-Brained GUI Agents: A Survey. *arXiv preprint arXiv:2411.18279* (2024), 1–80. https://doi.org/10.48550/arXiv.2411.18279

[79] Rui Zhang, Ziyao Zhang, Fengliang Zhu, Jiajie Zhou, and Anyi Rao. 2024. Mindalogue: LLM-Powered Nonlinear Interaction for Effective Learning and Task Exploration. *arXiv preprint arXiv:2410.10570* (2024), 1–17. https://doi.org/10.48550/arXiv.2410.10570

[80] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. VISAR: A Human-AI Argumentative Writing Assistant with Visual Programming and Rapid Draft Prototyping. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 1–30. https://doi.org/10.1145/3586183.3606800

[81] Jiayi Zhou, Renzhong Li, Junxiu Tang, Tan Tang, Haotian Li, Weiwei Cui, and Yingcai Wu. 2024. Understanding Nonlinear Collaboration between Human and AI Agents: A Co-design Framework for Creative Design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–16. https://doi.org/10.1145/3613904.3642812

Table 1: Overview of existing tools featuring interaction-augmented instructions, labeled within our analysis framework (Sec. 4). The tools are sorted by "Why", "When", "Who", "How", and "What".

Name	Source	Year	Why	When	Who	What	How
Choe et al. [18]	TVCG	2024	Restrict	Initial	Human	Visual Brush Selection	Direct
MagicQuill [40]	arXiv	2024	Restrict	Initial	Human	Visual Brush Selection	Direct
PromptPaint [19]	UIST	2023	Restrict	Initial	Human	Visual Brush Selection	Direct
Data Formulator [67]	TVCG	2024	Restrict	Initial	Human	Widget Selection	Direct
CML [28]	IUI	2023	Restrict	Initial	Human	Inline Highlighting	Direct
DirectGPT [43]	CHI	2024	Restrict	Initial	Human	Interactive Visual Click / Drag and Drop	Direct
Mindalogue [79]	arXiv	2024	Restrict	Initial	Human	Interactive Visual Click	Direct
WaitGPT [74]	UIST	2024	Restrict	Initial	Human	Interactive Visual Click	Direct
InternGPT [39]	arXiv	2023	Restrict	Initial	Human	Interactive Visual Click	Direct
Promptify [10]	UIST	2023	Restrict / Expand	Initial / Follow-up	Human	Interactive Visual Click / Widget Selection	Direct
ABScribe [49]	CHI	2024	Restrict	Initial	Human	Text Brush Selection / Widget Selection	Direct
Qlarify [22]	UIST	2024	Restrict	Initial	Human	Text Brush Selection / Widget Selection	Direct
ExpressEdit [64]	IUI	2024	Restrict	Initial	Human	Visual Brush Selection	Indirect
EasyAsk [24]	IMWUT	2024	Restrict	Initial	Human	Interactive Visual Click	Indirect
DeepWriting [1]	CHI	2018	Restrict	Follow-up	Human	Widget Selection	Direct
CURSOR [5]	Industry	2023	Restrict	Follow-up	Human	Text Brush Selection	Direct
OpenAI Canvas [46]	Industry	2024	Restrict	Follow-up	Human / AI	Text Brush Selection / Widget Selection	Direct
DynaVis [65]	CHI	2024	Restrict	Follow-up	AI	Widget Selection	Direct
ConstitutionMaker [48]	IUI	2024	Restrict	Follow-up	AI	Widget Selection	Direct
Luminate [59]	CHI	2024	Restrict	Follow-up	AI	Widget Selection	Direct
PDFChatAnnotator [62]	IUI	2024	Restrict	Follow-up	AI	Widget Selection	Direct
PromptCrafter [6]	ICML WS	2023	Restrict	Follow-up	AI	Widget Selection	Direct
Sparks [25]	DIS	2022	Restrict	Follow-up	AI	Widget Selection	Direct
Bursztyn et al. [11]	CHI EA	2021	Restrict	Follow-up	AI	Widget Selection	Direct
InkSync [32]	UIST	2024	Restrict	Follow-up	AI	Widget Selection / Inline Highlighting	Direct
AiCommentator [3]	IUI	2024	Restrict	Follow-up	AI	Widget Selection	Indirect
LangAware [16]	UIST	2023	Restrict	Follow-up	AI	Widget Selection	Indirect
Tableau AI [61]	Industry	2024	Expand	Initial	Human	Software Manipulation	Direct
G-VOILA [71]	IMWUT	2024	Expand	Initial	Human	Viewpoint Navigation	Direct
XR-Objects [20]	UIST	2024	Expand	Initial	Human	Viewpoint Navigation / Widget Selection	Direct
ChatScratch [15]	CHI	2024	Expand / Restrict	Initial / Follow-up	Human / AI	Sketch / Widget Selection	Direct
FigurA11y [57]	IUI	2024	Expand	Initial	Human	Widget Selection	Direct
DrawTalking [51]	UIST	2024	Expand	Initial	Human	Sketch / Speech	Indirect
LAVE [66]	IUI	2024	Expand	Initial	Human	Widget Selection	Indirect
Data Playwright [53]	TVCG	2024	Expand	Initial	Human	Inline Annotation	Indirect
NotePlayer [47]	UIST	2024	Expand	Initial	Human	Inline Annotation	Indirect
ScatterShot [72]	IUI	2023	Expand	Initial	AI	Widget Selection	Direct
Spellburst [4]	UIST	2023	Expand / Refine	Follow-up / Initial	Human / AI	Flowchart Manipulation / Widget Selection	Direct
Metaphorian [29]	DIS	2023	Expand	Follow-up	Human	Flowchart Manipulation	Direct
CoQuest [38]	CHI	2024	Expand	Follow-up	Human	Flowchart Manipulation / Text Input or Editing	Indirect
Sensecape [60]	UIST	2023	Expand	Follow-up	Human	Text Brush Selection / Flowchart Manipulation	Indirect
ExploreLLM [42]	CHI EA	2024	Expand	Follow-up	AI	Widget Selection	Direct
GAM [37]	CHI	2023	Expand	Follow-up	AI	Text Input or Editing	Indirect
Graphologue [27]	UIST	2023	Expand	Follow-up	AI	Inline Annotation / Flowchart Manipulation	Indirect
CoPrompt [21]	CHI	2024	Organize	Initial	Human	Tree Manipulation	Direct
VISAR [80]	UIST	2023	Organize / Expand	Initial / Follow-up	Human / AI	Tree Manipulation / Widget Selection	Direct
CoLadder [75]	UIST	2024	Organize	Initial	Human	Tree Manipulation	Direct
Low-code LLM [13]	NAACL	2024	Organize	Initial	Human	Flowchart Manipulation	Direct
Kim et al. [30]	UIST	2023	Organize	Initial	Human	Flowchart Manipulation / Widget Selection	Direct
PromptChainer [73]	CHI	2022	Organize	Initial	Human	Flowchart Manipulation	Direct
DreamSheets [2]	CHI	2024	Refine	Initial	Human	Spreadsheet Manipulation	Direct
PromptCharm [70]	CHI	2024	Refine / Restrict	Initial / Follow-up	AI / Human	Widget Selection / Visual Brush Selection	Direct