Survey of User Interface Design and Interaction Techniques in Generative AI Applications

Reuben Luera *raluera@ucsd.edu*
University of California – San Diego

Ryan A. Rossi *ryrossi@adobe.com*
Adobe Research

Alexa Siu *asiu@adobe.com*
Adobe Research

Franck Dernoncourt *dernonco@adobe.com*
Adobe Research

Tong Yu *tyu@adobe.com*
Adobe Research

Sungchul Kim *sukim@adobe.com*
Adobe Research

Ruiyi Zhang *ruizhang@adobe.com*
Adobe Research

Xiang Chen *xiangche@adobe.com*
Adobe Research

Hanieh Salehy *deilamsa@adobe.com*
Adobe Research

Jian Zhao *jianzhao@uwaterloo.ca*
University of Waterloo

Samyadeep Basu *sbasu12@cs.umd.edu*
University of Maryland, College Park

Puneet Mathur *puneetm@adobe.com*
Adobe Research

Nedim Lipka *lipka@adobe.com*
Adobe Research

Abstract

The applications of generative AI have become extremely impressive, and the interplay between users and AI is even more so. Current human-AI interaction literature has taken a broad look at how humans interact with generative AI, but it lacks specificity regarding the user interface designs and patterns used to create these applications. Therefore, we present a survey that comprehensively presents taxonomies of how a human interacts with AI and the user interaction patterns designed to meet the needs of a variety of relevant use cases. We focus primarily on user-guided interactions, surveying interactions that are initiated by the user and do not include any implicit signals given by the user. With this survey, we aim to create a compendium of different user-interaction patterns that can be used as a reference

for designers and developers alike. In doing so, we also strive to lower the entry barrier for those attempting to learn more about the design of generative AI applications.

1 Introduction

The academic and general population have become increasingly enamored with generative artificial intelligence (AI) as it continues to revolutionize just about every field of study it is involved in. So much so, in fact, that the field of Human-Computer Interaction (HCI) has shifted much of its focus to study a sub-field of HCI called Human-AI Interaction (IXDF, 2024). While current Human-AI Interaction literature provides a broad view of the field (Shi et al., 2023), this survey aims explicitly to capture the current state of user interfaces and the respective user interactions being utilized within generative AI applications. Specifically, this survey takes a snapshot of current trends and design techniques that involve user-guided interactions (Sec. 2.1).

In doing so, we aim to create a design compendium that generative AI designers, researchers, and developers can reference to understand the current state of the user experience (UX) and user interface (UI) designs of generative AI. The overall goal is to lower the barrier to entry for those interested in the UX and UI of generative AI by giving them a foundation upon which to build. For designers and developers specifically, this paper can serve as a design library to inspire their designs for generative AI applications. This paper prevents designers and developers from needing to partake in large competitive analyses and allows them to learn from design patterns currently utilized by other generative systems. For researchers, this survey can guide further explorations of human-AI interaction. This paper lists dozens of different ways that humans and generative AI applications are currently interacting and can serve as a foundation for researchers to dive into a specific area of human-AI interaction or to dive into the area in a general sense.

1.1 Summary of Main Contributions

This survey contains several major contributions to the human-AI interaction field based on our survey of more than a hundred relevant generative AI articles. The key contributions include key definitions and disambiguation, relevant taxonomies, and research-based design principles. Specifically, the key contributions of this work are as follows:

1. **A formalization of the key notions and definitions and a disambiguation and expansion of key terms relating to UI & interactions for generative AI applications (Section 2).** We formalize vital definitions that are relevant to understanding how a user interacts with a generative artificial intelligence system. We also disambiguate user interactions by proposing a novel concept of user-guided interactions, which are interactions that a user engages in willingly and deliberately. These do not include implicit interactions or interactions that the generative system detects without the user's knowledge. Additionally, we outline the different modality types that users can utilize to interact with generative systems. In doing so, we aim to enhance the general understanding of user interaction terms and their pertinence to generative AI.
2. **A survey and taxonomy of UI interaction techniques for generative AI systems (Section 3).** We highlight and categorize common user-guided interaction design patterns and the context in which they are used. We focus on user-guided interactions, as these are examples of how users deliberately and intentionally communicate with the generative systems. We organize these interactions into a taxonomy highlighting prompting, selecting, system and parameter manipulation, and object manipulation. We aim to create a compendium of user-guided interaction techniques that designers can refer to as they plan the designs for their own generative AI applications.
3. **A survey and taxonomy of user interface layouts for generative AI systems (Section 4).** We survey key user interface design patterns utilized in various generative applications. In doing so, we present a taxonomy that categorizes and highlights common user interface structures we found through the survey. By generalizing standard UI layouts, we prevent designers from starting from scratch and encourage using these common generative AI design patterns.
4. **A survey and taxonomy of human-AI engagement levels for generative AI systems (Section 5).** We survey human-AI engagement levels, which consist of the intensity of deliberate

interaction and collaboration between a system and the user. We aim to expand on existing literature that has characterized human-AI engagement levels by presenting more recent generative AI examples and by including additional human-AI engagement levels. In doing so, we aim to present a holistic survey of the current state of human-AI interaction levels.

5. **A survey and taxonomy of AI applications and use cases for generative AI systems (Section 6)**. We highlight and categorize the different applications of generative AI systems to survey how they are used in various domains. We aim to survey these different use case areas to discover which user interfaces and interactions are most appropriate for use in various situations. With this knowledge, designers can discern which interaction patterns and user-guided techniques are best used in their respective domains.
6. **An overview of key open problems and challenges that future work should address (Section 7).** We outline problems and open issues that future research can focus on to address generative AI accessibility, growth, and ethics. In doing so, we can continue to ensure that designers continue to create designs that incorporate solutions to larger issues.

1.2 Scope of this Article

This survey focuses on user-guided interaction techniques for generative AI applications. Such user-guided interactions are sometimes referred to as controllability techniques. One simple example of such techniques is when a user prompts a generative model via text and/or images; similarly, another example is when a user selects text or a specific part of an image to control the generation. We do not attempt to survey interaction or controllability techniques that are not user-guided, nor do we survey techniques that leverage user/system feedback and the like. Given this, the purpose of this paper is to provide a recommendation on which interaction techniques are most effective when applied to specific use cases.

2 Background & Preliminaries

We begin with basic definitions and notations to formalize the terms that will be used in subsequent sections as we discuss user-guided interactions and connected concepts. We formally define and discuss the notion of user-guided interactions and explain concepts such as the interplay and differences between prompting and inputs. Then, finally, we discuss different modalities that users can utilize to interact with the generative AI systems.

2.1 User-Guided Interactions

We begin by defining user-guided interactions in order to distinguish them as a specific type of user interaction. While user interactions are *any* interactions, whether explicit or implicit, made by a user to affect a system, user-guided interactions focus solely on interactions that are explicitly made by users to affect a system in a pre-desired way. Given this, this paper will only focus on user-guided interactions and how they are used in the context of generative AI.

Definition 1 (User Guided Interaction)**.** *User-guided interactions are defined as explicit user-initiated actions that a user deliberately makes that affect the respective computer system.*

In terms of generative AI, user-guided interactions consist of any actions that a user explicitly takes to affect the generative system. This can be anything from prompting the system to complete a certain task, to selecting and manipulating objects within a system, to adjusting a system's parameters to create a specific output. For example, a user might write a prompt that generates an image. Then they might adjust sliders or select specific parts of the image to manipulate it further. All of these are examples of user-directed interactions. This definition does not, however, include implicit user interactions, which are implicit or indirect actions made by the user that the system acts upon without being explicitly tasked to do so. Implicit user interactions include implicit behavioral interactions that the system uses to create a user profile or implicit feedback where the system infers user satisfaction based on word cues or interaction delays from the user. For instance, a system might alter its answers based on a user's chat history or surface different news

stories based on a user's implied political beliefs. We acknowledge that these interactions are a crucial part of the generative process, but these interactions fall outside the scope of user-guided interactions.

All in all, we focus solely on user-guided interactions in order to provide a holistic, but focused survey of these types of interactions. Doing so affords us the ability to delve deeper into the complexities of user-guided interactions and offer a wide palette of interaction solutions. In addition, we constrain our scope solely to user-guided interactions as these are the interactions, between the user and system, that are explicitly visible. For example, one can visibly see the interaction of a user selecting a UI element or prompting a system, but they cannot, however, see a system collect implicit feedback or user data. Given that user-guided interactions are those with visible interactions, they are also the interactions that can actually be visibly designed. Given this, this survey's goal is to be especially helpful to those product, visual, user experience (UX), and any other type of designer that has been tasked with designing a generative AI user interface. To conclude, we deliberately highlight and focus on user-guided interactions as their implementation can greatly enhance the overall generative AI user experience. Utilizing this scope ensures that this survey is focused and understandable for designers of all levels.

2.2 Prompts and Inputs

Definition 2 (Inputs)**.** *An input is a piece of data, information, or content that the user uploads to the system. An input, if available, is what the prompt acts upon.*

Definition 3 (Prompts)**.** *A prompt is a type of user-guided interaction in which the user asks the generative system to complete a certain job.*

Prompting a generative system is often the most commonly thought of user-guided interaction associated with generative AI. As mentioned, it consists of a user asking a generative system to complete a specific task. While text-prompting is the most commonly thought of prompting modality, other prompting modalities include visual, audio, and multi-modal prompting. In essence, prompting is an essential way that users interact with and guide the systems that they are working with. Meanwhile, inputs are data, information, or content that is uploaded to the generative system. Like prompts, inputs can be text-based, visual, or audio. In tandem, or sometimes on their own, these are two important aspects of the generative AI user flow.

An important distinction is that this section does not consider "prompts" as an input. Instead, we view user prompts and inputs as often two separate entities, where a prompt is used to query the system, and the input is what is being acted upon. As seen in Fig. 1, if there is an audio editing generative system, then the *input* would be an audio file, but the user *prompt* can be text such as "Can you edit this audio clip so that it is only one minute long?". Thus, the input and prompting are distinct since the prompt is a user-guided interaction that acts on the input video.

We separate "prompting" and "inputting" as two different terms because, from a user interaction point of view, they are distinctly different. Whereas an input interaction is essentially some data that the generative system is acting upon, prompting is the user-guided interaction that consists of actually instructing the system to complete a specific task. Furthermore, distinguishing between the two components simplifies and focuses the upcoming discussion sections: Input Modalities (Sec. 2.3) and Prompting (Sec. 3.1).

2.3 Input Modalities

This subsection will focus solely on the input modalities that a user can use when interacting with generative AI systems. As they pertain to generative AI user interactions, we define inputs as data or information that the generative AI system is capable of analyzing or processing to generate a different output. After surveying the modalities of different generative AI systems (Fig. 3), we present the different input modalities currently in use: text-based inputs (Sec. 2.3.1), visual inputs (Fig. (Sec. 2.3.2)), and sound inputs (Sec. 2.3.3).

2.3.1 Text-based

Natural Language: Text-based natural language is a modality commonly used to interact with generative AI systems (Achiam et al., 2023; Padiyath & Magerko, 2021; Suh et al., 2023a; Wang et al., 2024c; Petridis

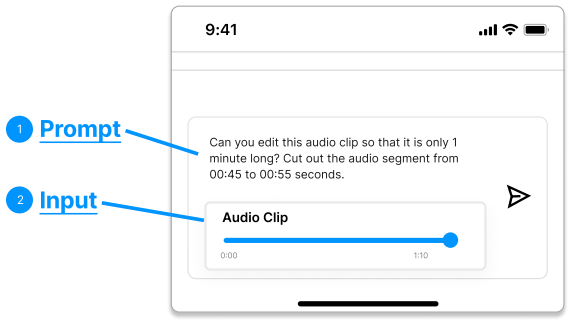


Figure 1: **Prompt vs Inputs (Sec. 2.3): A visual summary of the distinction between prompts and inputs. A prompt is a user-guided interaction where the user asks the system to complete a task. Whereas the input is the piece of data, information, or content that the prompt is acting upon.**

et al., 2024; Zhao et al., 2023; Kim et al., 2023a; Cho et al., 2024; Wang et al., 2024b). As noted, the text inputs in questions are *not* the same as prompts. Text inputs can be anything from PDF files, structured texts, and unstructured texts, but they are unique as they do not explicitly ask the system to perform an action on its own. It is the prompts that do that, as illustrated in figure 2.2. Given this point, Wang et al. (2024b)'s AesopAgent is a prime example of inputting natural language text into a generative system to create something novel and unique. With AesopAgent, a user can input a short text story and separately prompt the system to create a full script for the given short text story. The system obliges and outputs a full script with storyboard images and can even create a matching video. In doing so, the system essentially can create a short animated video from just an inputted text story.

Data: Generative AI systems are often used as a tool to synthesize, clean, or gain insights from inputted data (Achiam et al., 2023; Setlur et al., 2016; Singh et al.; Schneider et al., 2019; Swanson et al., 2024; Doe et al., 2019). In relation to generative AI, inputted data can be thought of as either raw or structured information or data that can be in the form of text files, structured data files, system log files, etc. Datasets, in general, can be large, untenable obstacles when it comes to completing both academic and industrial goals. Having generative AI help synthesize and digest data can make existing tasks easier and make new tasks possible. Take Synthemol (Swanson et al., 2024) for example. This system takes in chemical data to synthesize and design new chemical compounds that are novel and synthesizable. In essence, using data in the generative process, whether structured or unstructured, can help users complete tasks, like chemical synthesis, that would be extremely difficult to do without generative AI.

Code: Code is a common type of text that is both inputted and outputted from generative systems (Ross et al., 2023b; Achiam et al., 2023; Barke et al., 2023b; Chen et al., 2021; Finnie-Ansley et al., 2022; Yen et al., 2023; Li et al., 2023a; Okuda & Amarasinghe, 2024). When talking about code as an input or output, this can take the form of programming languages, structured scripts, markup languages, query languages, etc., and users could input it into a system to get help completing it (Achiam et al., 2023), debug it (Achiam et al., 2023), or ask questions to understand it (Ross et al., 2023b). For example, in Ross et al. (2023b)'s Programmer's Assistant, users can input raw code and ask the generative system natural language questions about either the inputted code or how to create new code. By inputting already established code into the system, the system can interact with the input and present new potential interactions that the user can partake in.

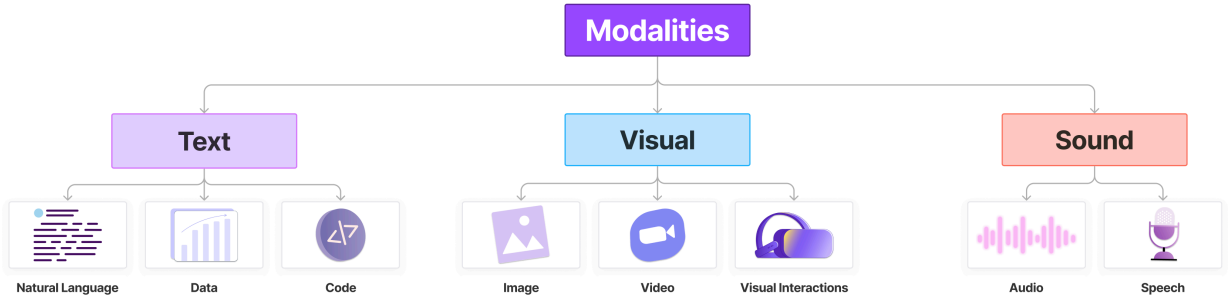


Figure 2: **Modalities: A high-level visual summary of the different modalities that generative AIs use (Sec. 2.3).**

2.3.2 Visual

Images: Using images as an input to interact with generative systems has become common, as users attempt to do everything from generating new images (Padiyath & Magerko, 2021; Jeon et al., 2021; Betker et al., 2023), to captioning images (Singh et al.; Alayrac et al., 2022), to creating infographics (Setlur et al., 2016). In essence, the image modality consists of the user interacting with images of any type, whether they are infographics, photographs, illustrations, etc., and many systems have differing and unique uses for image inputs. For example, one aspect of Singh et al.'s FigurA11y is creating alt text for figures in research papers. A user inputs figures and the system will create captions and alt text to increase accessibility or even help the reader get a deeper understanding of the paper. Overall, image inputs have a wide range of uses and are an extremely versatile modality as it pertains to generative AI systems.

Videos: Videos are also a common input and output modality for generative systems and are especially common in multi-modal LLMs (MMLLMs) (Liu et al., 2023b; Cho et al., 2024; Gao et al., 2023; Wu et al., 2023; Goyal et al., 2023). Users can use these systems to do anything from generating or finishing videos (Cho et al., 2024; Liu et al., 2023b) to highlighting or annotating them (Liu et al., 2023b). For example, in InternGPT Liu et al. (2023b), a user can input a video and prompt the system to complete a task pertaining to that video. So in the system, a user could prompt it to edit an inputted video so it matched a TikTok format. Generative AI interactions with video modalities, such as this, have a wide range of uses for both professional and amateur users.

Visual Interactions: Visual interactions are an input modality that consists of any visual interaction or gestural movement that the system records as an input. This includes visual movements in virtual or augmented reality spaces (Giunchi et al., 2024; Konenkov et al., 2024; Doe et al., 2019) or directly manipulating UI elements in a way the system takes as an input (Lin & Martelaro, 2024; Jiang et al., 2023; Suh et al., 2023a; Masson et al., 2023; Kim et al., 2023b). For example, Konenkov et al. (2024)'s VR GPT allows users to gesture at certain items to ask the system to interact with it. So if a user is trying to learn how to correctly pack a medical bag, they may gesture at a first-aid item and ask, "What is that?" In doing so, their pointing gesture is recorded in the system as an input and is used in tandem with the spoken prompt to interact with the system in a unique way.

2.3.3 Sound

Speech: Speech is a growing medium that more and more generative systems can interact with. In most use cases, speech is often used or generated by the system to help the user complete a speech-related task. Take Borsos et al. (2023)'s AudioLM, which is capable of taking a recorded spoken input and generating an "end" to the recording. So if a user generates the first half of a speech, AudioLM can generate the rest of the speech. Interactions that use speech as a modality, such as this one, often create novel and relevant use cases for their respective users.

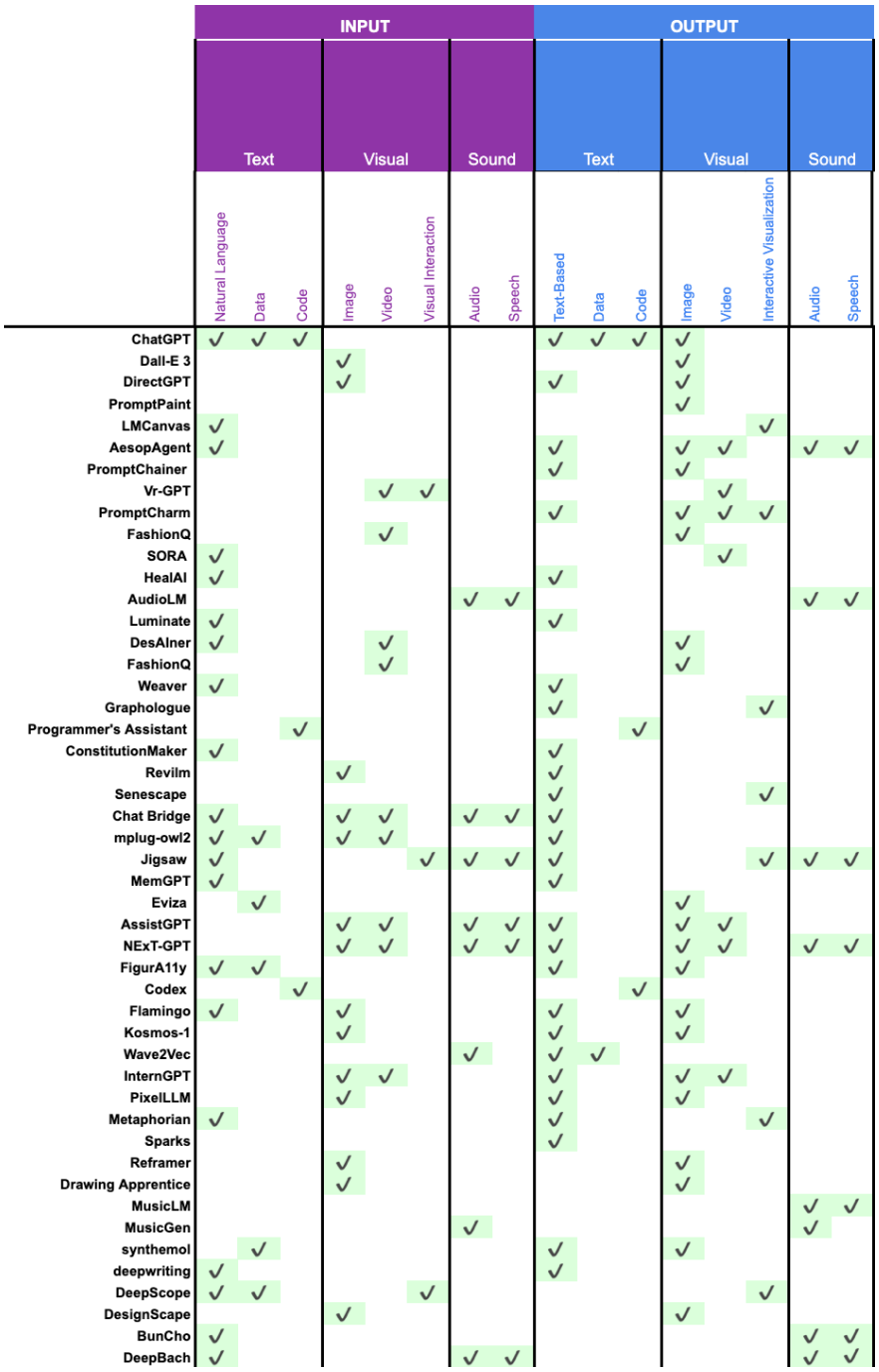


Figure 3: **Taxonomy of works by their input/output modalities.**

Audio: Similar to the video modality, audio inputs and outputs have become a versatile way that the user can interact with generative AIs (Borsos et al., 2023; Wang et al., 2024b; Zhao et al., 2023; Gao et al., 2023; Wu et al., 2023; Copet et al., 2023; Agostinelli et al., 2023). In Wu et al. (2023)'s NExT-GPT, a user can input a sound recording of something like a plane taking off. From there, the user can prompt the system to do something along the lines of "create a video that this sound would come from." The system will then use the inputed audio recording to create an accompanying video of a plane taking off. Inputting audio recordings offers another dimension that can be utilized to prompt the system to complete a unique set of tasks.

3 User-Guided Interactions

Definitions and Scope: In the case of generative AI, user-guided interactions are defined as explicit and deliberate engagements between the user and the system that the user initiates or guides in the generative process. Therefore, the scope of this section will focus primarily on the actual user interactions explicitly performed by the user, and not implicit interactions such as how a user's interactions can implicitly inform an AI system over time about their preferences. Many of these interactions are performed by interacting with UI elements, which, in this case, are defined as the visual components that users interact with to manipulate the generative process.

In the context of generative AI, we propose the following taxonomy that categorizes user-guided interactions into the following categories: prompting (Section 3.1), selection techniques (Section 3.2), system and parameter manipulation (Section 3.3), and object manipulation and transformation (Section 3.4).

1. **Prompting (Sec. 3.1)**: Prompting is a user-guided interaction in which a user asks or "prompts" the system to complete a certain task or job. Such user-guided prompting interactions include text-based prompts (Sec. 3.1.1), audio prompts (Sec. 3.1.3), visual prompts (Sec. 3.1.2), and multi-modal prompts (Sec. 3.1.4).
2. **Selection Techniques (Sec. 3.2)**: Selection techniques use various tools and methods to highlight or choose a specific UI element (*e.g.*, prompting blocks, image or text previews, parts of an image, etc.) within the generative AI system to be further interacted with during the generative process. Such selection interactions include single selection (Sec. 3.2.1), multi-selection (Sec. 3.2.2), lasso and brush selection (Sec. 3.2.3), and multi-modal selection (Sec. 3.2.2).
3. **System and Parameter Manipulation (Sec. 3.3)**: System and parameter manipulation consist of user interaction techniques that allow the user to adjust the parameters, settings, or functions of an overall generative AI system. These interactions are often used to personalize generated outputs to meet user needs. Such user-guided system and parameter manipulation interactions include menus (Sec. 3.3.1), sliders (Sec. 3.3.2), and explicit feedback(Sec. 3.3.3).
4. **Object Manipulation and Transformation (Sec. 3.4)**: Object manipulation and transformation interactions occur in situations where the user directly modifies, adjusts, and/or transforms a specific UI element, like a building block, puzzle piece, or similar entity. Doing so gives the user deeper control over the system and allows them to interact with the UI elements in a unique and novel way. Such user-guided object manipulation and transformation interactions include drag and drop interactions (Sec. 3.4.1), connecting (Sec. 3.4.2) and resizing (Sec. 3.4.3).

Motivation: There is likely no generative AI system that utilizes all of the techniques that will be outlined, nor should there be. The goal of these user-guided techniques is to ensure a seamless and user-friendly navigation of the respective systems to improve the generative process. Moreover, surveying such a wide array of user-guided techniques in generative AI will help the reader understand which generative AI techniques are most appropriate in a number of different situations. The goal is also to expose several novel user interaction techniques that are not widely known and to empower designers and developers to leverage them to create powerful and accessible generative AI systems.

Variability: As such, we survey works that have used any of the UI interactions (*e.g.*, text, visual prompts, drag-and-drop, sliders, selection, in-painting, and so on), and categorize the systems that use each, and how they use them, and the different ways each are used to guide the generation process. This is useful to understand how each UI interaction (*e.g.*, sliders) was used by existing systems (*e.g.*, in the case of sliders, some work used sliders to adjust the generative model hyperparameters, while others used it to adjust the attention weights of specific user-guided selections such as text that was generated).

3.1 Prompting

Definition & Scope: Prompting is a user-guided interaction in which a user asks or "prompts" the system to complete a certain task or job. Prompting, specifically text-based prompting, is often thought of as the primary interaction method utilized to interact with generative AI systems. Prompting is different than just inputting content like images or videos, as the user is explicitly asking the system to perform a task. An

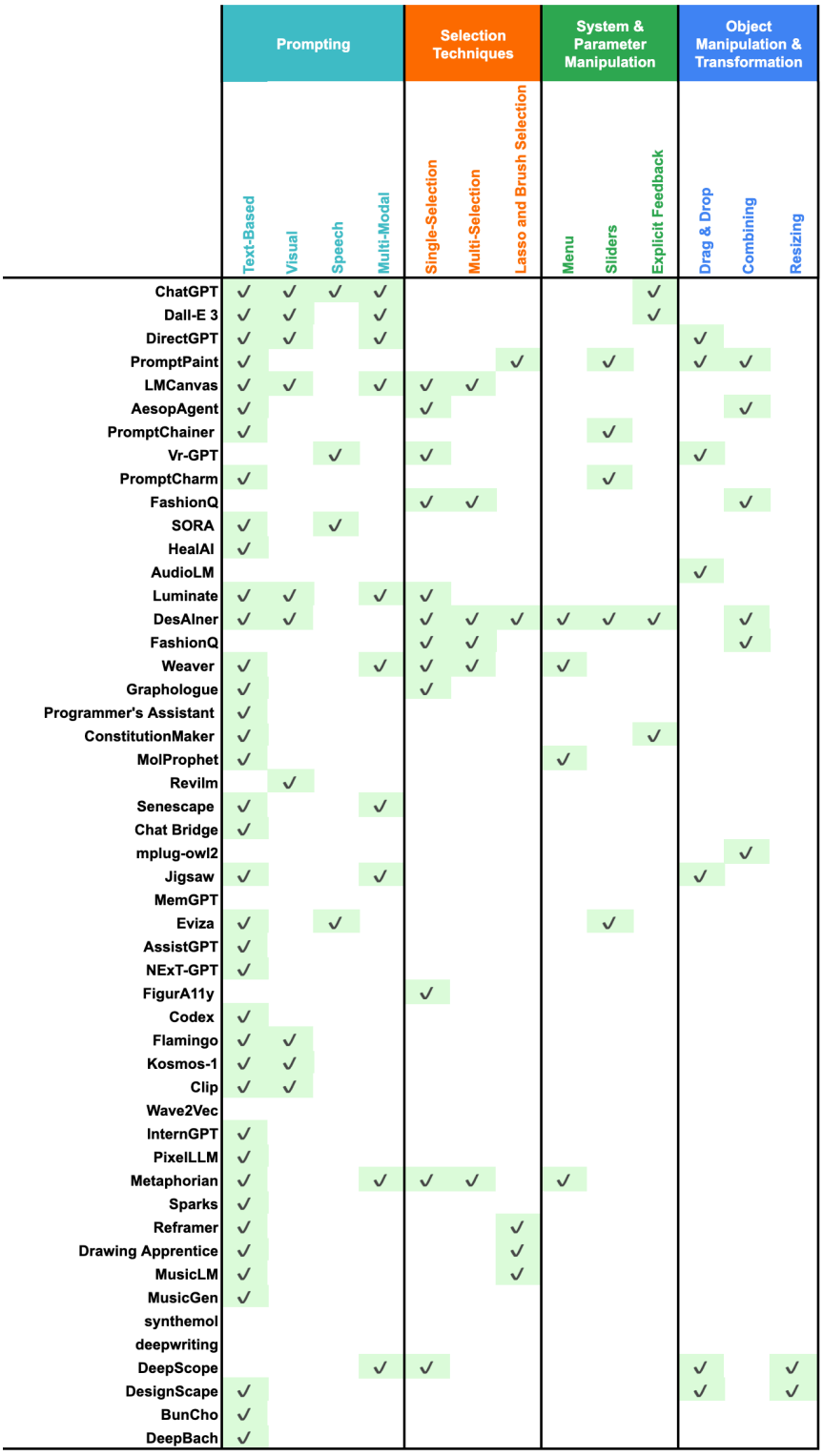
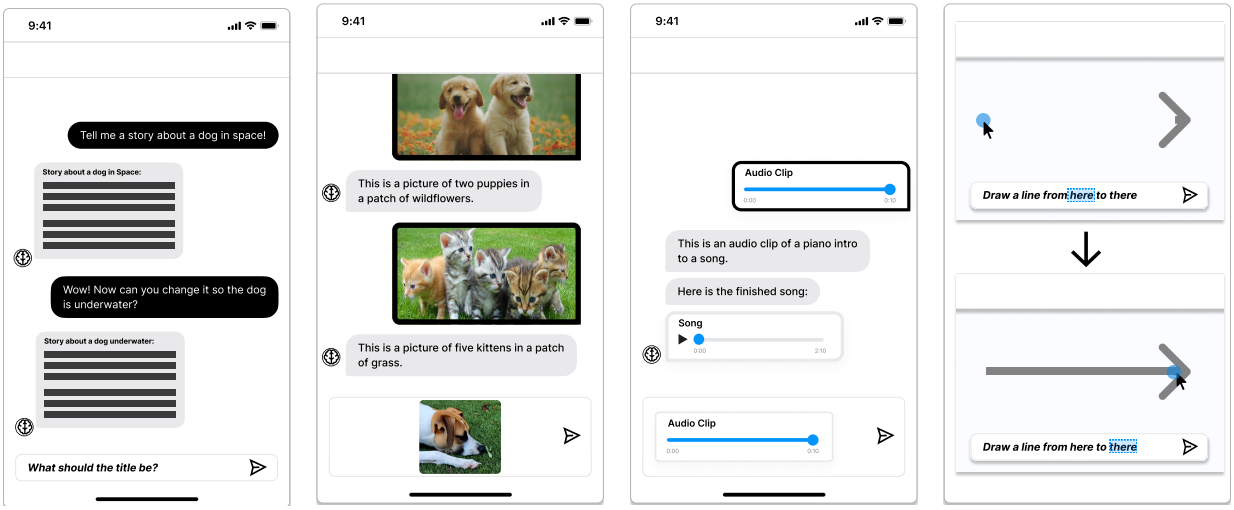


Figure 4: **User-Guided Interaction Taxonomy.** Generative AI systems and tools are summarized using the proposed user-guided interaction taxonomy (Sec. 3).

important distinction is that in terms of generative AI user interactions, a prompt is not the same as an input. For example, if there is an AI system like video editing, the input could be a video, but the user-guided prompt can be text, such as "edit the video so it is no longer than 10 seconds." Thus, the inputs and prompts are distinct, where the prompt is a user-guided interaction that acts on the input video.



(a) Text-based Prompt (§3.1.1) (b) Visual Prompts (§.3.1.2) (c) Audio Prompts (§.3.1.3) (d) Multi-Modal Prompts (§3.1.4)

Figure 5: **Prompting Visual Summary (Sec. 3.1): An overview of the four main prompting subcategories. Prompting is a user-guided interaction where a user asks or "prompts" the generative AI system to complete a certain task.**

- **Text-based Prompts (Sec. 3.1.1; Fig. 5(a))**: Text-based prompts consist of using written text, often in the form of natural language, to prompt the system to complete a certain task.
- **Visual Prompts (Sec. 3.1.2; Fig. 5(b))**: Visual prompting consists of using visual communication, like gestures, to prompt the system to complete a certain task.
- **Audio Prompts (Sec. 3.1.3; Fig. 5(c))**: Audio prompting consists of using speech or any other type of audio to prompt the system to complete a certain task.
- **Multimodal Prompts (Sec. 3.1.4; Fig. 5(d))**: Multimodal prompting consists of using a mix of the previous methods to prompt the system to complete a certain task.

3.1.1 Text-based Prompts

Text-to-text prompting: Text-based prompting is the most common interaction medium a user uses when interacting with Generative AI systems and has a wide range of applications (Goyal et al., 2024; Achiam et al., 2023; Betker et al., 2023; Gero et al., 2022). While there are several prompt mediums, systems like ChatGPT (Achiam et al., 2023) and Dall-E 3 (Betker et al., 2023) heavily rely on text-based inputs. In ChatGPT (Achiam et al., 2023), for example, users primarily input prompts like "explain the Krebbs cycle to me as if I were a child," and the system will output an explanation of the Krebbs Cycle in plain English (Fig. 5(a)). Meanwhile, in systems like Chen et al. (2021), users can use text prompts to ask generative models to create or continue a section of code for them.

Text-to-multimodal prompting: Meanwhile, multi-modal systems can also utilize text-based prompts and produce outputs of a different medium like visuals (Alayrac et al., 2022; Liu et al., 2023b) or audio (Copet et al., 2023; Agostinelli et al., 2023; Hadjeres et al., 2017). Dall-E 3, for example, incorporates a text-to-image generation system that takes text-based prompts and outputs an image. Another example is text-to-video, where generative AI systems, like AESOP's agent (Wang et al., 2024b), take text-based prompts and create AI videos. Text prompts are common in many multimodal generative AI system interactions, with some examples consisting of, but not limited to: text-to-image, text-to-video, text-to-audio, and so on. Similarly, some generative AI applications (Zhao et al., 2023; Ren et al., 2023) allow text-based prompts to ask questions about or manipulate multi-modal inputs like an image or video. All in all, text prompts are often the backbone of most user interactions because of their versatility and the large amounts of different actions that can be performed.

3.1.2 Visual Prompts

Visual manipulation: Visual prompting consists of using visual communication, like object manipulation, to prompt the system to complete a certain task (Fig. 5(b)). For example, Jigsaw (Lin & Martelaro, 2024) created a set of puzzle pieces that each have a corresponding instruction that the system can complete. So visually, the user can manipulate, drag and drop, and connect these puzzle pieces (each puzzle piece has a partial prompt) to create a new, larger prompt. In doing so, they are visually manipulating UI elements to prompt the system uniquely.

Image-Prompts: As mentioned, there is a distinct difference between prompts and inputs. A prompt is used to ask the system to complete a certain task, and some inputs can be used to do the same. In turn, this makes it possible for some inputs, like visuals, to explicitly prompt the system to perform a certain task. This is most common in few-shot learning examples where a model can generate content based on a limited number of training inputs (Alayrac et al., 2022; Wei et al., 2023a; Radford et al., 2021). In Alayrac et al. (2022), an image of a solved math equation can be inputted (i.e. 1+1=2, 1+2=3, etc.). Then, using another image of an unsolved math equation (i.e., 1+3=?) would prompt the system to generate a fill to that last equation. Whereas the original picture of solved math equations would purely be seen as inputs, the image of the unsolved math equations would be considered a prompt, as it is *prompting* the system to provide an answer to the unsolved equation.

3.1.3 Audio Prompts

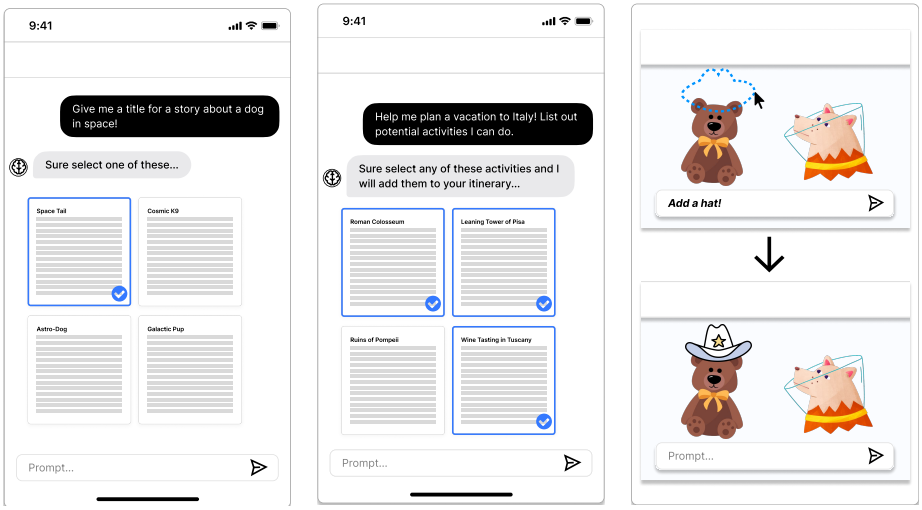
Speech: When a user is embedded in a virtual environment and do not have access to traditional text-based inputs, audio and speech become one of the strongest ways to interact with AI models. In Konenkov et al. (2024), the user can interact with the visual language model (VLM) by speaking into the VR headset's microphone. As it is a VR training application, users can prompt the system by verbally asking it questions like "what is my next step." In situations like this, a user relies on speech as the primary way to prompt the VLM for a response. Just as LLMs such as ChatGPT rely on text-based prompts, VLMs embedded in VR environments often rely on spoken prompts in the same way.

Audio Files: Some models (Borsos et al., 2023; Schneider et al., 2019), can be prompted with just an audio clip. AudioLM is a novel system that uses two audio clips, one being the input clip and the other being the prompt clip, to extend the prompted audio clip. So for example, if a user has a ten-second speech (inputted clip) they can then prompt the system with the first three seconds of the clip and then the system will generate a new and unique continuation from those original three seconds. This type of prompting is interesting because there is no natural language involved. Because the system's only function is to generate a continuation of the inputted audio, the only prompting needed is the truncation of the original clip. This is significant because few systems do not require either a spoken or written prompt. This system is only able to use this novel system of prompting because it only has a single function with essentially no extra parameters, so it does not require extra written or spoken directions.

3.1.4 Multi-Modal Prompts

Dual-modality: DirectGPT (Masson et al., 2023) illustrates that it is common for selection techniques to exemplify different modalities. For example, (Masson et al., 2023) explains how the DirectGPT system incorporates a text-based and selection-based input simultaneously. In this instance, users can highlight multiple words simultaneously and then type a prompt that encourages the system to replace those words with a synonym. Using a hybrid of several interaction techniques is especially useful when a single-dimensional interaction does not address the customer's need on its own. Furthermore, hybrid interactions can significantly reduce the amount of time it takes for a user to complete a task, as seen in Masson et al. (2023) where users completed tasks 50 percent faster than those who utilized single dimensional interactions like in ChatGPT.

Multi-modal LLMs: Meanwhile, multi-modal LLMs (MM-LMMs) are becoming more common, as they are almost universal generative tools that work in a number of different situations (Alayrac et al., 2022; Achiam et al., 2023; Wu et al., 2023) NExT GPT (Wu et al., 2023) is an empowered MM-LLM, meaning that it takes



(a) Single Selection (b) Multi-Selection (c) Lasso and Brush Selection

Figure 6: **Selection Techniques (Sec. 3.2): Selecting, in terms of generative AI systems, consists of choosing or highlighting a specific UI element in order to further interact with it.**

inputs from any modality and can output in any modality. It relies on a mix of searching for keywords and detecting the input modalities to decide what the output modality should be.

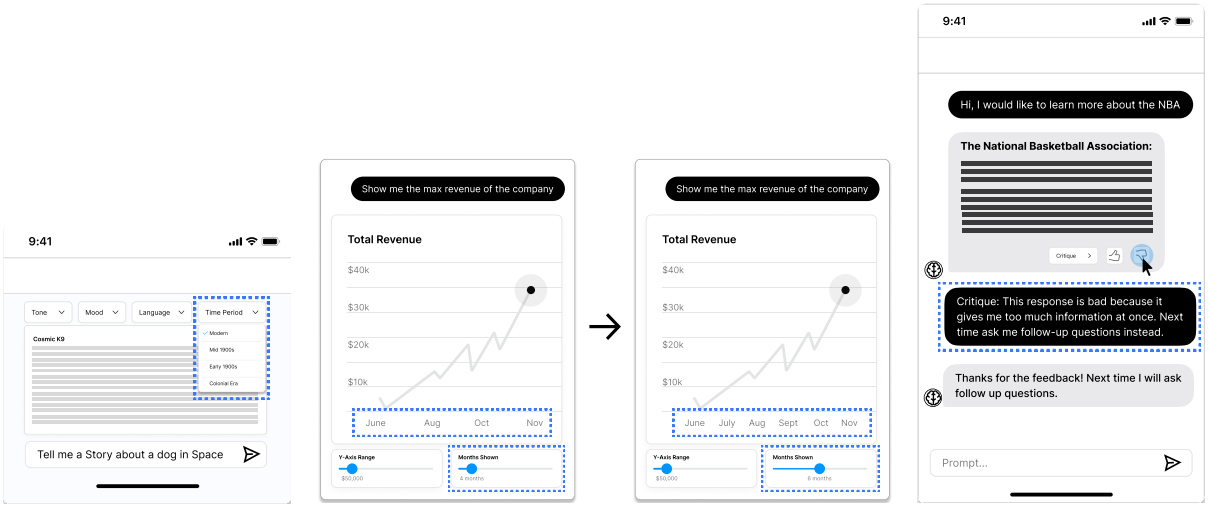
3.2 Selection Techniques

Definition & Scope: Selecting, in terms of generative AI systems, consists of choosing or highlighting a specific UI element to further interact with it. Selecting UI elements and utilizing selection tools has become a major user interaction method utilized in many generative AI systems. Selectable UI elements and buttons, selectable inputs and outputs, and dropdown menus all offer a way to directly interact with the system. Meanwhile, selection tools such as selection boxes, lassos, or marquee tools enable users to choose precise areas or objects generated by the AI. By incorporating these user-interaction techniques, users are able to select content with more control and accuracy.

- **Single Selection (Sec. 3.2.1; Fig. 6(a))**: A single-selection interaction consists of clicking or choosing a single GUI element that will be interacted with further. An example would be a user choosing one of 3 outputs that they wish to iterate on further.
- **Multi-Selection (Sec. 3.2.2; Fig. 6(b))**: A multi-selection interaction consists of clicking or choosing multiple UI elements that will be interacted with further.
- **Lasso and Brush Selection (Sec. 3.2.3; Fig. 6(c))**: Lasso and brush selections are selection techniques where a lasso or brush is used to create a bounding box that controls the region where a specific prompt is applied (Figure 6(c)).

3.2.1 Single-Selection

A single-selection interaction consists of clicking or choosing a single UI element (*e.g.*, selecting one of several outputs, choosing part of an image, etc) that will be interacted with further (Lee et al., 2023; Suh et al., 2023a). Luminate (Suh et al., 2023a) is an LLM that intakes a prompt from an author writing a story and outputs several different story options based on the original prompt. Utilizing single-selection, the user clicks on and chooses which of the story options they want to further iterate on. So if a user prompts the system to write a short story about an astronaut, the system will output several stories. From there, the user can click and select the story that best fits their needs. Selection interactions, as opposed to prompting, are often utilized because of their ease of use and efficiency.



(a) Menus (b) Sliders (c) Explicit Feedback

Figure 7: **System and Parameter Manipulation (Sec. 3.3): User interaction techniques that allow the user to adjust the parameters, settings, or functions of an overall generative AI system.**

3.2.2 Multi-Selection

Multi-Select to Alter: Multi-selection consists of interactions where users can select multiple elements simultaneously with the goal of interacting with multiple elements at once. For example, DirectGPT (Masson et al., 2023) allows the user to make one or more selections within an inputted text and create a tool that affects all of the selected words. For example, say a user is writing a story, but they feel that their story is a bit bland, and they want to make the vocabulary more descriptive. They can use multi-select the words that they want to make more interesting, such as “run”, "eat," and "said," and then they can type "replace with synonyms" in the prompt module. The system will take all of the words that they selected and replace them with more interesting words like "dashed," "devoured," and "exclaimed." In using the multi-select interaction, users can find and select multiple elements or words that they want to interact with further.

Multi-Select to connect: Furthermore, multi-selection is often used to select multiple outputs so that the user can connect them together (Padiyath & Magerko, 2021; Jeon et al., 2021). For example, Padiyath & Magerko (2021) is a generative fashion AI system that creates several dress outputs for a designer. The user can then use multi-select to select multiple aspects from different dresses and add them together to create a new dress. So, the fashion designer could select the sleeves they like from one dress, the corset they like from another, and the straps from another output and use all these parts of a dress to prompt the system to create new designs. In cases like this one, multi-select is used to select and add together multiple elements to prompt the generative AI system with the sum of those elements.

3.2.3 Lasso and Brush Selection

Inpainting: Lasso and brush selection, also known as inpainting, allows users to use a tool to select very specific parts of a larger element. Whereas other selection methods make you choose an entire element, inpainting allows users to fine-tune and edit their selection. PromptPaint (Chung & Adar, 2023) focuses on enhancing the generation of images by brushing over the image to better control the generation process where multiple prompts can be mixed. Different prompts can be applied to different regions of the images. The paint-like interactions can be seen as a type of visual selection via lasso, where the lasso bounding box controls the region that a specific prompt is applied. Another work called PromptCharm (Wang et al., 2024d) allows the user to create a mask M over a generated image I by brushing over the area of interest, and then a new image I' is generated by modifying only the pixels of I that lie within the mask M. The user can guide the inpainting process by typing a text prompt $\mathbf{x}_M$ that is applied only to the masked region of I.

3.3 System and Parameter Manipulation

Definition & Scope: The System and parameter manipulation category consists of user interaction techniques that allow the user to adjust the parameters, settings, or functions of an overall generative AI system. These interactions are often used to personalize generated outputs to meet the needs of a user. Some examples of this type of user-guided interaction consist of Menus (Section 3.3.1), Sliders (Section 3.3.2), and Explicit Feedback (Section 3.3.3).

- **Menus (Sec. 3.3.1; Fig. 7(a))**: when a user either inputs their own parameters or chooses from preset options to change the parameters of the generative process. An example would be a user using dropdown menu options to alter the output parameters.
- **Sliders (Sec. 3.3.2; Fig. 7(b))**: A UI element that can be "slid" to adjust the parameters of the generative AI system.
- **Explicit Feedback (Sec. 3.3.3; Fig. 7(c))**: Explicit feedback (i.e., thumbs up/down, written critiques, etc.) is a user interaction that is used to expressly personalize the system to the user's preferences.

3.3.1 Menus

Preset Option Menus: Menus are user-guided interaction features that allow users to either input their own parameters or choose from preset options (Wang et al., 2024c; Suh et al., 2023a; Jiang et al., 2023; Kim et al., 2023a). A common example of a parameter-selection menu is a drop-down menu that allows users to adjust the parameters of a system with preset parameters. An example of this can be seen in Suh et al. (2023a), a system that helps users write stories or poems. In this system's UI, there are drop-down menus that allow users to impact the output's tone, mood, and overall structure. In menus like these, the user can choose between pre-set parameters to adjust the overall output. Being able to utilize UI features to quickly adjust output parameters drastically speeds up the user experience and lowers the user's cognitive load.

Input Menus: While drop-down menus are a common interaction used to adjust parameters, some systems have menus that solely rely on manual user input. These menus allow the user to input text parameters that work in tandem with the prompt to create a concise output. For example, in Setlur et al. (2016), users can create a data visualization given a data file input and can adjust the generated data visualization by changing typing in different parameters. So if a user prompts the system to "map out all earthquakes in California," there is a parameter menu next to visualization where they can manually type parameters like time range, location range, etc. that will affect the generated visualization. Similarly, PromptCharm (Wang et al., 2024d) is an image generation platform that allows users to generate images with text prompts and menus. In terms of menus, for example, PromptCharm has an "image style" menu where a user can input text separate from the original prompt that allows them to specifically impact the image style. Overall, menus allow users to add another dimension to their already existing prompts and guide them to manipulate specific parameters within the generative process.

3.3.2 Sliders

Range Adjustments: Sliders are visual UI elements that are able to manipulate the parameters of the generative AI system (Chung & Adar, 2023; Setlur et al., 2016; Wu et al., 2022) For example, in Eviza (Setlur et al., 2016), sliders are used to control the parameters of the type of data that is being shown. Eviza is a data visualization software that overlays data onto a map and the user can interact with the sliders to affect the types of data that is being shown. So if the system generates a historical map of earthquakes in California, the user can manipulate the output by using the sliders to ensure that only earthquakes with magnitudes greater than 5.0 are shown. Essentially, in cases like this one, sliders are used to manipulate data being and work in tandem with the original prompt.

Attention Weights: During the generation process, PromptCharm (Wang et al., 2024d) allows users to select specific tokens in the prompt to then adjust their attention weights via a slider that is dynamically shown for the token selected. This is another user-guided interaction to better control the generated image and how the model attends to the specific tokens in the prompt, which can be viewed as importance/influence

scores. Hence, the user can ensure that the tokens most important to the user are also consistent with the importance in the model, giving another axis of fine-grained controllability during the generation process. This can also be viewed as an explainability feature that can help the user refine the prompt directly as well (*e.g.*, tokens in the current prompt can be highlighted via a colormap that corresponds with the attention weights of each, and then the user can modify the prompt by adding, removing, or rephrasing the prompt accordingly).

3.3.3 Explicit Feedback

Binary Feedback: A user's explicit feedback is a user interaction type that can change a system's parameters and, therefore, the generative system as a whole (Petridis et al., 2024; Achiam et al., 2023). Giving feedback to the system can be anything from giving an answer a thumbs up or thumbs down (Achiam et al., 2023), to explicitly instructing the system to respond a certain way, i.e. "Critique: when I talk about my mom, ask how she is doing" (Petridis et al., 2024). For example, systems like Achiam et al. (2023) offer binary feedback on whether the answer that the generative model gave was good or bad. More specifically, the system provides a thumbs up or down option to allow the user to give explicit feedback as to whether or not they were satisfied with the outputted answer.

Comprehensive Feedback: Meanwhile, systems like Petridis et al. (2024) allow users to give more specific feedback than just a binary "good" or "bad" response. Constitutionmaker allows users to rate an output by giving it kudos, critiquing it, or rewriting it. As seen in Figure 7(c), a user can write out a specific critique, such as "Critique: This response is bad because it gives me too much information at once. Next time ask me follow-up questions instead." The system will then take this feedback and add it to its own constitution, and all of the subsequent outputs will follow the rules set forth in the constitution. Essentially, feedback techniques allow a user to adjust the parameters of a system to their liking.

3.4 Object Manipulation and Transformation

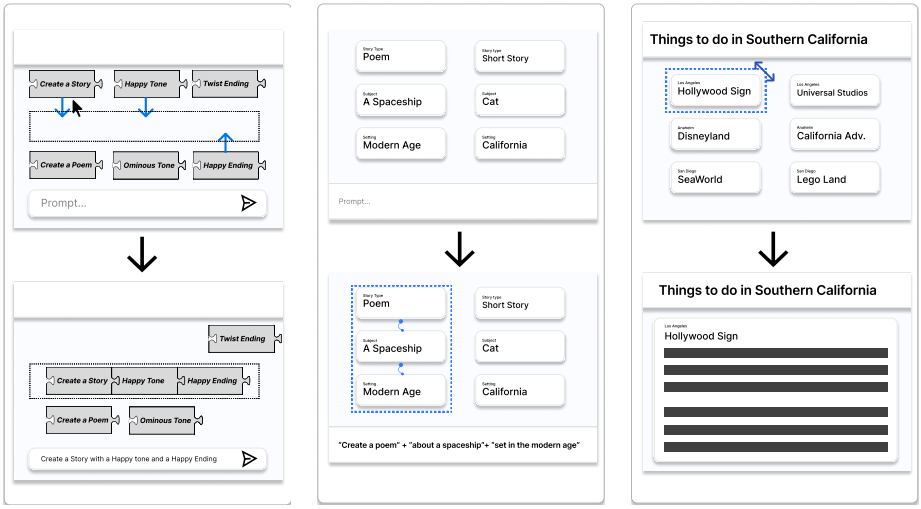
Definition & Scope: Object manipulation and transformation interactions involve the user directly modifying, adjusting, or transforming a specific UI element. Doing so gives the user a deeper control over the system and allows them to interact with the UI elements in a unique and novel way. Some examples of this type of user-guided interaction consist of Dragging and Dropping UI elements (Section 3.4.1) and Resizing a UI element within the system, and (Section 3.4.3).

- **Drag and Drop (Sec. 3.4.1; Fig. 8(a))**: Moving an element to a specific location or in a way that manipulates the generative system.
- **Connecting (Sec. 3.4.2; Fig. 8(b))** In connecting interactions, a user stacks and connects two UI elements in a way that affects the overall generative process.
- **Resizing (Sec. 3.4.3; Fig. 8(c))**: Altering the size of a UI element in a way that changes its function in the generative process.

3.4.1 Drag and Drop

For creation: Drag and drop interactions consist of moving an element to a specific location or in a way that manipulates the generative system. One work by Masson et al. (2023) leverages drag-and-drop interactions to help in modifying a vector graphic, *e.g.*, they type "add a line from", then use drag-and-drop on the vector graphic to refer to the specific locations, which are then included in the prompt being created via direct manipulations.

To Connect Dragging and dropping, as an interaction, often goes hand in hand with the subsequent user interaction technique, connecting (Sec. 3.4.2) (Kim et al., 2023b; Lin & Martelaro, 2024). Oftentimes, users will drag and drop UI elements toward each other with the end goal of connecting them. Take Jigsaw (Lin & Martelaro, 2024) for example, a system that created a set of puzzle pieces that each have a corresponding system instruction that the system can complete. The user can then drag and drop the puzzle pieces to combine them, which simulates connecting system instructions within a prompt. For example, a user might Connect a puzzle piece that has the prompt "Upload an image" with two other puzzle pieces that say "remove



(a) Drag and Drop (b) Connecting (c) Resizing

Figure 8: **Object Manipulation and Transformation (Sec. 3.4): User interaction techniques that modify, adjust, and/or transform a specific UI element, like a building block, puzzle piece, or similar entity.**

people" and "increase resolution." By dragging and dropping these puzzle pieces together, the user is essentially creating a prompt without having to type it out. Dragging and dropping puzzle pieces with different system instructions on them helps spur novel prompt creations and helps inform the user what the application is capable of.

To Disconnect While the drag and drop interaction is often used to connect two UI components, often to prompt a system in a unique way (Lin & Martelaro, 2024; Kim et al., 2023b), dragging and dropping can be used to disconnect components. LMCanvas (Kim et al., 2023b) is a system that uses sentence blocks as starting points that users can interact with and generate content. In LMCanvas (Kim et al., 2023b), users can write a sentence in a sentence block, use multi-select to select half of the sentence, and then drag that half of the sentence away from the other half of the sentence. In doing so, the user has split the original sentence block, creating two sentence blocks that they can further interact with. Essentially, drag and drop to split is another way that users can interact with prompts in the generative system. All in all, object manipulation techniques, such as this one, may be more appealing to users who are more hands-on or are not as familiar with the capabilities of generative AI.

3.4.2 Connecting

Prompt-to-input: As mentioned, connecting goes hand in hand with drag-and-drop interactions. In connecting interactions, a user stacks and connects two UI elements in a way that affects the overall generative process. For example, Kim et al. (2023b) aims to facilitate an easier writing process by allowing users to create prompts and input "blocks" that can be connected to simulate connecting a prompt to an input. For instance, there may be a prompt/model block that says, "translate this story from English to French." Then, a user can drag an input block with an English story onto that system instruction block, and the system will recognize this as the user prompting the system to complete the task or translating the inputted story to French. Interactions like these are beneficial for visual or kinesthetic learners who may not be familiar with the basic functionalities of traditional generative AI models like ChatGPT.

Prompt-to-prompt: A similar application that has already been talked about, Promptpaint (Chung & Adar, 2023), allows users to combine existing prompts together in a novel way. As mentioned, Promptpaint is a generative AI system that allows users to utilize multiple prompts to create an image. A user can pick and choose which prompts on a prompt list they want to apply to a generated image. An interesting feature that Promptpaint has is *prompt mixing*, where a user can add prompts to a painter's palette where each prompt is

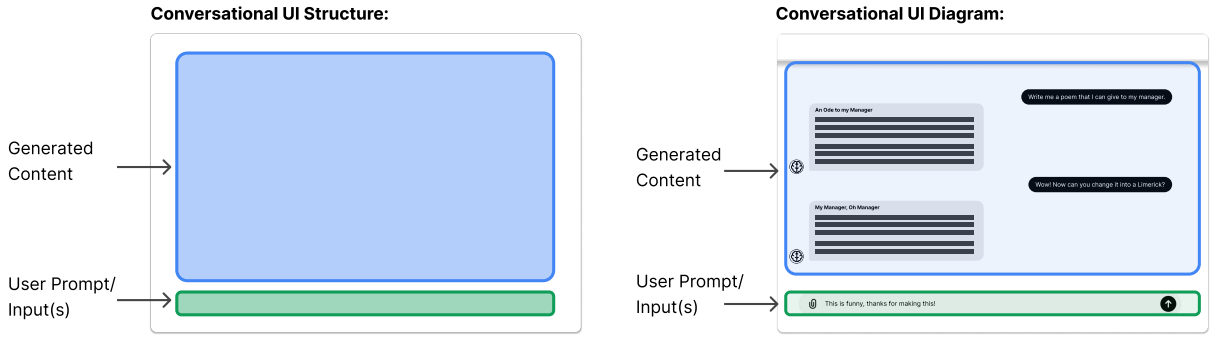


Figure 9: **Conversational UI: A conversational UI is structured so that a user interacts with the user prompt/input box. From there, their output(s) and output history exist in a larger space within the UI (Sec. 4.1).**

assigned a sort of color blob, just as a normal painter's palette would have. From there, the user can "connect" the prompts together, which essentially connects the contents of the prompt to create an image. This visual representation of connecting prompts helps users visualize the connecting of two often dissimilar prompts.

3.4.3 Resizing

General resizing In terms of generative UI interactions, resizing an object can have drastically different outcomes depending on the system. Suh et al. (2023b)'s proposed system, Senescape, is a generative AI system that utilizes a resizeable canvas to make digesting information much easier for the user. In Senescape, the user can resize the view port of the canvas to adjust the amount of information that is generated and subsequently shown to them. By essentially zooming in on an element on the canvas, the system will generate more detailed information about the respective element. Whereas if a user zooms out, they will see only high-level information. In essence, using a resizing user-interaction technique allows users to control the amount of information that is shown to them. Doing so significantly reduces cognitive load and effectively gives the user control over the system.

4 User Interface Layouts for Generative AI

Definitions and Scope: In this section, we propose a taxonomy based on different high-level user interface (UI) categories, looking at the overall structures of generative AI user interfaces. Generally, a user interface (UI) is defined as a means by which a human and a computer communicate with one another (Norman, 1988; Chignell, 1990). In the context of generative AI, we organize UIs into the following categories: conversational user interface (Section 4.1), canvas user interface (Section 4.2), modular user interface (Section 4.4), and simulated user interfaces (Section 4.5).

1. **Conversational User Interface (Sec. 4.1, Fig. 9)**: A conversational interface is an input-output based interface where users interact with the AI on a turn-based cadence. This structure mimics human conversation, allowing users to ask questions, input media, and receive responses in a sequential format. Examples include AI assistants like chatbots, zero-shot agents, and other turn-based AI agents.
2. **Canvas User Interface (Sec. 4.2), Fig. 10)**: A canvas interface often revolves around generating and interacting with a central piece of content. In this interface, the content in question is often in the center of the UI, and generative tools are often on the peripheral. This interface category includes generative systems that edit and generate content like images, videos, word documents, code, data visualizations, and audio.

3. **Contextual User Interface (Sec. 4.3, Fig. 11)**: Contextual UIs consist of a user interface where the generative element is in a smaller UI element that exists, structurally, in line with a particular part of the larger UI. Oftentimes, this contextual generative UI element exists near where the user is interacting with the larger application.
4. **Modular User Interface (Sec. 4.4, Fig. 12)**: Modular UIs consist of user interfaces that are broken up into multiple main interaction areas, with each of these interaction areas having a different function in the generative process. Modular user interfaces are often used in systems with multiple levels of generation.
5. **Simulated Environment (Sec. 4.5)**: A simulated environment UI replicates real-world or hypothetical scenarios in a virtual space, allowing users to interact with and navigate through these scenarios as if they were real. This type of UI is commonly used in virtual reality (VR), augmented reality (AR), and training simulations, where users can immerse themselves in the environment and interact with objects and elements within it. The goal is to provide a realistic and interactive experience that can be used for training, entertainment, or exploration.

Motivation: We include this taxonomy as a compendium that designers can use when designing a new generative system. The UIs that follow all have specific use cases where they serve the user best, and using the correct UI in the appropriate scenarios will create a better user experience overall. The goal of this section is to explain how and when each UI structure can be used and to give real-world examples of how these UI structures are being used today.

4.1 Conversational User Interface

A conversational UI is characterized as a user interface that is visually structured in a way that mimics a conversation (Lister et al., 2020; Achiam et al., 2023; Fu et al., 2023; Miller et al., 2017), with there being an exchange between a user and the AI system following a turn-based cadence. This interaction space often consists of an input or prompt box along with a section of the UI for interaction history (Fig. 11). Furthermore, this category will focus on conversational user interfaces within GUIs, and not VUIs like Amazon Alexa and Apple's Siri.

In terms of generative AIs, visual conversational user interfaces primarily have the same or similar user interface structures. Generally, much of the focus is on an input or prompting box where a user can prompt the system to complete a certain task or ask it a question (Achiam et al., 2023; Fu et al., 2023). In conversational UIs a majority of user-guided interactions will occur in this prompt or input box, as it is the primary interaction space. Given this, the secondary interaction space in conversational user interfaces is the chat history or output section of the UI (Fig. 9). Given a prompt, the output or chat history section houses the output and/or a chat history of past interactions. This section can store anything from past input and act as a chat history (Fu et al., 2023; Achiam et al., 2023) or even hold a single or a gallery of user outputs that a user can select from (Betker et al., 2023; Lee et al., 2023; Wang et al., 2024b).

Although conversational UIs are mostly constrained by a turn-based cadence and an input-and-output model, they are capable of a variety of use cases. For example, systems like Fu et al. (2023)'s Gemini can be used to generate text and have general conversations, while systems like Lee et al. (2023)'s BOgen can generate a gallery of 3D models. Both of these scenarios also encapsulate one of the major strengths of conversational UI. Conversational UIs excel at asking the system to make continuous incremental changes as the generative process goes on (Achiam et al., 2023; Betker et al., 2023; Fu et al., 2023; Liu et al., 2023b). Similarly, conversational UIs excel at recalling information from earlier in the conversation to inform more recent outputs. For example, systems like Alayrac et al. (2022)'s Flamingo learn from earlier conversations and inputs to perform one-shot tasks where they learn how to identify images and patterns using past chat history. This is a strength of visual conversational UIs, as both the system and the user can reference earlier conversations. All in all, conversational UIs are versatile in that they can be used to perform a variety of different tasks.

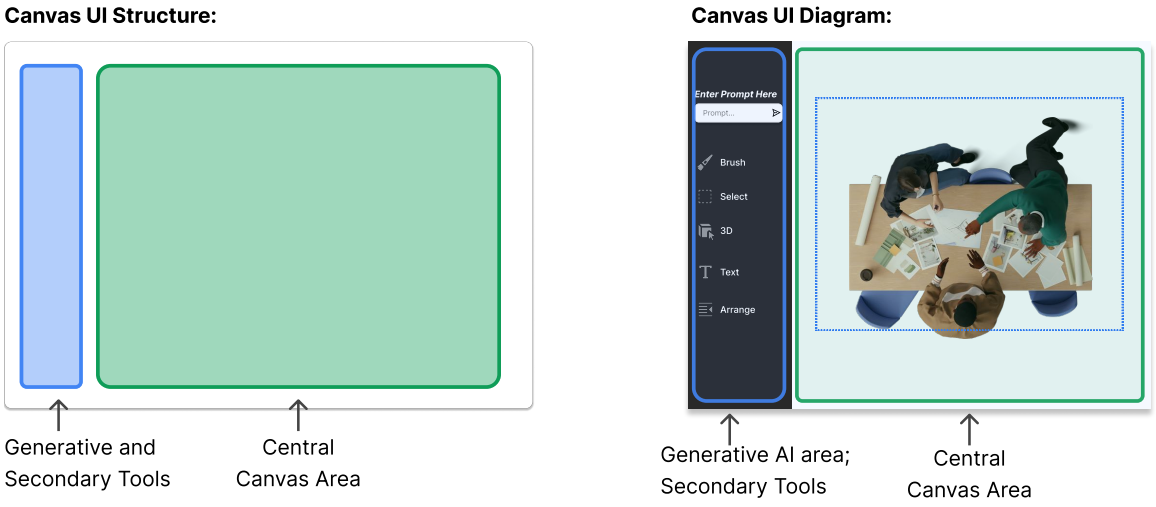


Figure 10: **Canvas User Interface: A UI structure with a central canvas area that houses the primary content. The generative and other tools are often in the periphery or off to the side. (Sec. 4.2).**

4.2 Canvas User Interface

As it pertains to the user interfaces of generative AI, we define canvas-focused user interfaces as those that are structured with a focus on a central canvas area within the UI. Structurally, a majority of the user interactions occur in this central canvas, with the generative tools existing in the periphery (Fig. 10). Canvas-focused user interfaces are broken into two subcategories: content canvases and information visualization canvases.

Content Canvases: Content Canvas user interfaces are a subcategory of canvas UIs Where the primary canvas area is occupied by a piece of content such as an image or drawing (Chung & Adar, 2023; Liu et al., 2023a; Lawton et al., 2023; Du et al., 2024; Padiyath & Magerko, 2021) a set of text (Shi et al., 2022), code (Ross et al., 2023b; Prather et al., 2023; Barke et al., 2023b), or data visualizations (Setlur et al., 2016). An example of a content canvas is Du et al. (2024)'s DeepThInk. DeepthInk (Du et al., 2024) is an AI art therapy tool that enables the average person to draw in art therapy sessions. Its UI is structured, so there is a central canvas, and the generative and other art tools are on the periphery. So the user essentially primarily interacts with the art in the middle of the screen on the central canvas and can elicit help from the generative tools on the periphery whenever it is appropriate. By using this structure, most of the interaction is funneled to the central canvas, whereas secondary interactions, like AI generation, occur on the periphery (Fig. 10).

Information Visualization Canvases: Information visualization canvases focus more on visualizing input and output interactions on a central canvas. While content canvases focus on a central piece of content, information visualization canvases are essentially sandboxes where users can directly manipulate components that alter the system. These systems are used to help users visualize the interplay between either the inputs (Lin & Martelaro, 2024; Masson et al., 2023; Kim et al., 2023b) or the outputs (Suh et al., 2023a; Kim et al., 2023a; Jiang et al., 2023; Suh et al., 2023b). Take Jiang et al. (2023)'s Graphologue, for example. Graphologue (Jiang et al., 2023) is a system that takes common prompts like "plan a vacation for me in San Diego" and outputs a treemap that represents an itinerary for said trip. Whereas conversational UIs (Section 4.1) primarily output a block of text, visualization canvases UIs output the same information in an easy-to-digest chart, graph, or other visualization. Doing so lowers the cognitive load needed to digest the outputted information and allows the user to interact with specific parts of the output as needed. In the case of Graphologue (Jiang et al., 2023), the system generates a canvas that visualizes the output as a hierarchical treemap in which every "branch" represents a part of the itinerary. Each branch can be expanded to reveal a list of restaurants, a list of hikes, etc.. Again, this type of UI can be utilized to break large pieces of information into easy-to-digest segments, lowering the cognitive load of the user

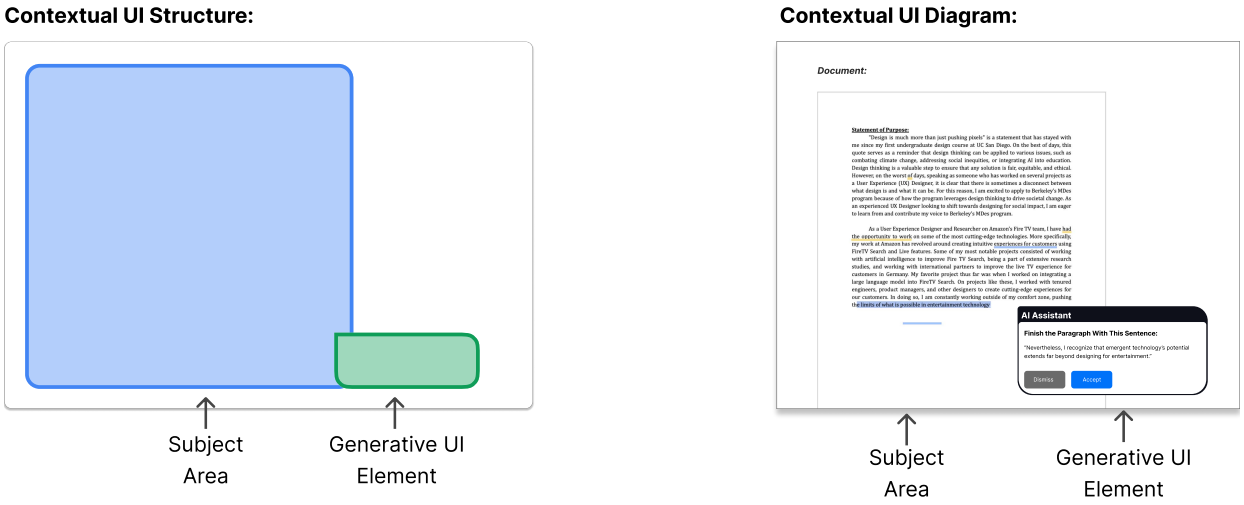


Figure 11: **Contextual User Interface: UI Structures where the generative UI element appears inline in context to the primary user interaction space (Sec. 4.3).**

4.3 Contextual User Interface

As it pertains to generative AI, a contextual user interface is a type of UI where the generative interaction occurs in line with a specific aspect of the larger subject area (Fig. 11). Unlike canvas UIs, contextual UIs's generative action does not occur in the periphery and instead occurs in line where the user is most likely looking. Oftentimes, the generative UI elements within contextual UIs appear unprompted, but instead, as a result of something the user does within the subject area (Jakesch et al., 2023; Chang et al., 2023; Fitria, 2021; Packer et al., 2024). Other times, this UI structure is used to put prompting or other toolbar interactions contextually within the subject area, as seen in Masson et al. (2023). All in all, this type of UI structure is an effective strategy for lowering the cognitive load of the user as it displays relevant interactions in context to the user.

Take, for example, Packer et al. (2024)'s MemGPT, a generative AI system that contextually gives reminders to the user about the person that they are talking to. This system works contextually within a messaging application and "remembers" relevant facts about the person the user is speaking with. Then, at appropriate times, a pop-up window appears that reminds the user of relevant information about the person they are chatting with. So, for example, if a user was talking to their friend on their birthday, the system may contextually remind them to say happy birthday and can go so far as to tell them what birthday plans this person might enjoy doing. All of these interactions happen contextually within the system and occur in what is essentially another chat bubble. Contextual user interfaces are especially useful in situations such as this one, where the generated output is relevant in context to a specific unprompted input.

4.4 Modular User Interface

As it pertains to generative AI, modular UIs consist of user interfaces that are broken up into multiple main interaction areas, with each of these interaction areas having a different function in the generative process (Fig. 12). Modular user interfaces are often used in systems with multiple levels of generation. These systems necessitate a modular design as each "module" is handling a different level of interaction (Yan et al., 2023; Wang et al., 2024d; Singh et al.; Petridis et al., 2024). Take, for example, Wang et al. (2024d)'s PromptCharm. PromptCharm (Wang et al., 2024d) is a text-to-image generative system with a modular user interface. It is unique in that it has an interstitial step where the system helps refine the user prompt to allow the text-to-image part of the system to better capture the user's original intent. This system necessitates a modular design because it has two generative functions: refining the user's original prompt and

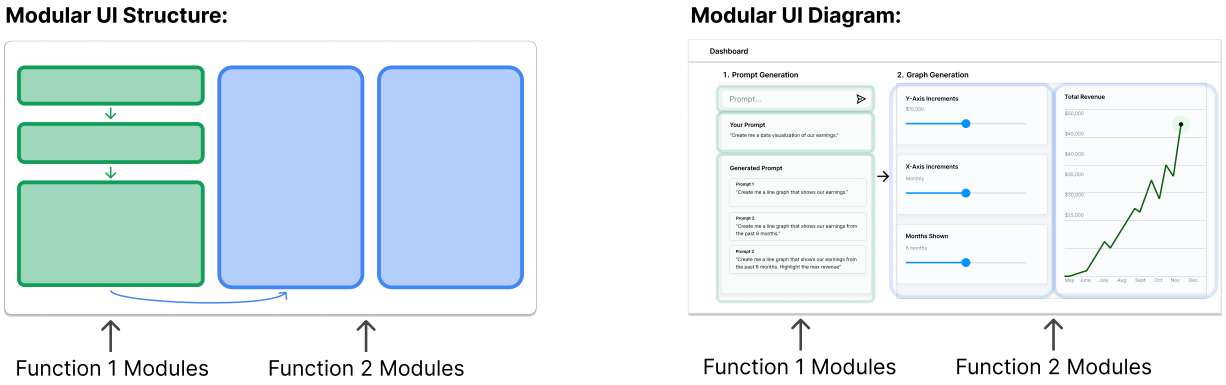


Figure 12: **Modular User Interface: A user interface layout that is broken up into several interaction spaces, each with a different function (Sec. 4.4).**

generating an image based on said prompt. For this reason, PromptCharm has three "modules" dedicated to the generation and refinement of the original prompt and two modules dedicated to the generation and refinement of the image. Given this, modular UIs are useful in that they are versatile and can handle several levels of interaction. Designers should consider utilizing them to design interfaces for multi-level LLMs and other multi-functional generative systems.

4.5 Simulated User Interface

Simulated User Interfaces allow users to interact with generative systems in either virtual reality or some form of augmented reality. Simulated UIs are often necessary when traditional GUIs are incapable of helping a user complete a specific task. These interfaces allow users to interact with the generative system in a way that can teach them how to perform a certain task in a simulated environment (Konenkov et al., 2024; Giunchi et al., 2024) or interact with data in a tangible way (Doe et al., 2019). Take, for example, Doe et al. (2019)'s DeepScope, a tangible user interface created in augmented reality that simulates the urban planning of a particular city. Deepscope essentially helps users create a 3 dimensionsional blueprint of a city that can be used to discuss and visualize urban planning concepts. While they are currently more difficult to create than GUIs, simulated interfaces, such as DeepScope, are especially useful as they offer a tangible interaction that GUIs lack. In turn, they are able to better train users how to complete certain tasks or help them interact with "tangible" data in real-time

5 Human-AI *Engagement* Taxonomy: From Passive to Collaborative

In this section, we propose a taxonomy based on the different levels of engagement in human-GenAI interaction, going from least user interaction to increasingly more. More formally, engagement is defined as the process by which interactors start, maintain, and end their perceived connections with each other during an interaction (Oertel et al., 2020; Sidner et al., 2003; Salam et al., 2023).

Specifically, we propose a spectrum of engagement levels going from passive to fully collaborative. In particular, the main engagement types identified include passive engagement (Sec. 5.1), deterministic engagement (Sec. 5.2), assistive engagement (Sec. 5.3), and sequential collaborative engagement (Sec. 5.4), and simultaneous collaborative engagement (Sec. 5.5). We provide an intuitive summary of the engagement level taxonomy in Table 1 and provide intuitive examples of each in Figure 13. The engagement level dictates the application scenarios supported by the generative AI system and the interaction techniques.

5.1 Passive Engagement

Passive engagement is defined as a system that generates content based solely on implicit information gained by the user. The implicit inputs can be anything from usage patterns, user preferences, or user search history. Passive engagement consists of using these implicit inputs to generate content, and in these scenarios, there is no active interaction by the user. The systems generate content independent of user-guided interactions and are often agent-initiated systems. Examples of passive engagements in generative AI consist of social media engagement systems that measure and generate user engagement metrics (Gatti et al., 2014; Dorri et al., 2018), predictive AI medical models (Dogheim & Hussain, 2023; Farrokhi et al., 2024; Jeddi & Bohr, 2020), systems that curate personalized news based on user preferences (Kim & Lee, 2019; Oh et al., 2020), and systems that recommend personalized design assets (Cai et al., 2022; Kadner et al., 2021).

An example of a system that leverages passive engagement is PINGS (Personalized and Interactive News Generation System) (Kim & Lee, 2019). This system automatically generates personalized sports news stories based on a user's preferences. Systems that utilize passive engagement are often used to create outputs that rely heavily on user preferences or other implicit interactions. The best passive systems successfully integrate themselves into user's lives and require little interaction from the user. Furthermore, the success of these systems is often measured by how well they can interpret a user's passive inputs.

5.2 Deterministic Engagement

Deterministic engagement consists of a user-AI interaction cadence where the AI system works almost entirely, void of user input or interaction. In these scenarios, the system is provided preset parameters that it uses to complete predetermined task(s). Often times the extent of user interaction at this level of engagement consists of a user inputting preset parameters and/or telling the system to START or STOP. Deterministic generative AIs often only have a single use that can range from anything from news gathering (Nishal & Diakopoulos, 2024), chemical synthesis (Yang et al., 2024), and educational content generation (Truong et al., 2021; Arakawa et al., 2022),

Table 1: **Taxonomy of Human-GenAI Engagement.** We summarize the main categories of human-GenAI engagement and provide intuitive definitions and examples of each.

Engagement	Definition	Examples
Passive Engagement (§5.1)	No direct user interaction during the generation process leverages only user profile and preferences	- immersive news writing (Oh et al., 2020) - personalized curated sports articles (Kim & Lee, 2019) - AI-generated user engagement metrics (Gatti et al., 2014)
Deterministic Engagement (§5.2)	Similar to passive, though user provides basic instructions to the genAI model to start or stop the generative process.	- AI generated hierarchical tutorials (Truong et al., 2021) - automated newsgathering (Nishal & Diakopoulos, 2024) - chemical synthesis (Truong et al., 2021)
Assistive Engagement (§5.3)	Offers indirect assistance to users such as making suggestions. Systems using assistive engagement must understand the user intentions and high-level goals.	- follow-up question generation (Valencia et al., 2023b) - autocompletion (Jakesch et al., 2023) - writing suggestions (Fitria, 2021)
Turn-based Collaborative Engagement (§5.4)	The generative process between the user and generative model occurs in a sequential fashion (*i.e.*, turn-based)	Turn-based conversational interfaces where the user makes a request, then AI generates content, and the process repeats in a turn-based fashion.
Simultaneous Collaborative Engagement (§5.5)	User and GenAI work together in parallel to generate the final content	A drawing system where user and generative AI draw concurrently in real-time (Lawton et al., 2023)

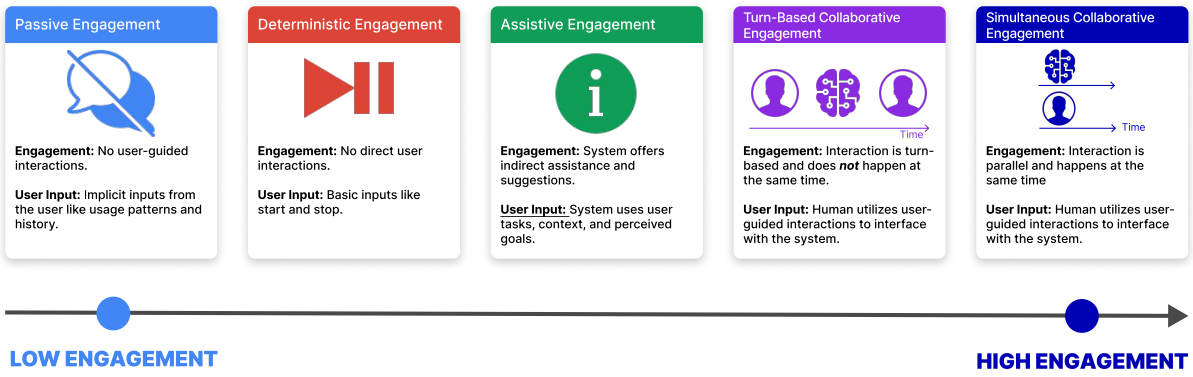


Figure 13: **Human-AI Engagement Taxonomy: A high-level visual summary of the different Human-AI Engagement Levels (Sec. 5).**

A novel example of this is a generative system that is responsible for a "newsgathering" task at a newspaper outlet (Nishal & Diakopoulos, 2024). Essentially, this generative system is responsible for gathering leads, data, and potentially newsworthy leads within a given news cycle. This process is automated, and the generative system determines what is newsworthy based on preset parameters. When it is finished, it presents the topics that it feels to be newsworthy. This is a wholly automated process, with the only human engagement happening when the system starts and when it stops.

5.3 Assistive Engagement

The assistive engagement level is characterized by interactions that offer indirect assistance to the users such as helping them complete a question (Fitria, 2021; Jakesch et al., 2023), auto-completing incomplete code (Barke et al., 2023a; Ross et al., 2023a; Prather et al., 2023), and helping users keep their focus (Arakawa et al., 2023). This level of engagement enhances a user's output without taking creative control. Generative AI that engages at an assistive level is incapable of functioning as an independent agent and is not prompted to act by the user. Assistive engagement is often ongoing and can occur several times while a user is working on a project

Assistive engagement is especially valuable in creative or academic use cases where users need to generate unique or personal content. In these scenarios, users can benefit from generative AI without being flagged for plagiarism or feeling disconnected from the creative process. For instance, a person designing art software might receive suggestions about color palettes to use, or a computer science student might be offered fixes for compiling errors. Essentially, assistive-level engagement is utilized in designing for use cases where users desire autonomy over the generative process while still benefiting from generative AI.

5.4 Turn-based Collaborative Engagement

Turn-based collaborative engagement consists of a user working in tandem with generative systems to complete a task or create a final product. Commonly, the UI structure of this collaboration method follows an almost sequential, turn-based cadence with the user and the system taking turns inputting and outputting information (Achiam et al., 2023; Masson et al., 2023; Petridis et al., 2024; Ross et al., 2023b; Betker et al., 2023). Scenarios such as these involve information exchanges and are typically conversational in style. Moreover, the human and AI are working towards a single goal such as creating and editing a data visualization (Setlur et al., 2016), editing a piece of visual content (Lawton et al., 2023; Wang et al., 2024b; Davis et al., 2015; Yan et al., 2023; Bangerl et al., 2024), or writing a story or other piece of creative content (Suh et al., 2023a; Wang et al., 2024c; Jiang et al., 2023; Kim et al., 2023a). Furthermore, this interaction dynamic requires the user to have some level of subject matter experience to guide the system. In order for the user and the system to work collaboratively with one another, there may be instances where the user may have to act as a guide.

While common turn-based collaborative models include the likes of ChatGPT (Achiam et al., 2023) and Dall-E (Betker et al., 2023), one novel turn-based collaborative system can be found in Davis et al. (2015)'s Drawing Assistant. This system is a drawing model that generates art asynchronously with its user. Essentially, the human and the AI take turns drawing, with the AI basing its drawing on the user input. The system operates on a turn-based cadence, with the human and AI not working simultaneously. While collaborative interactions do not need to occur in real-time, this section is underscored by a back-and-forth interaction between the user and the system and is the most involved of any interaction level.

5.5 Simultaneous Collaborative Engagement

In this section, we discuss simultaneous collaborative engagement that occurs at the same time. More formally, the human and AI are collaborating concurrently on a given task (Deshpande, 2020; Lawton et al., 2023). As an example, a user and AI may be working simultaneously on an image editing task where the user and AI are making changes to different parts of the image in parallel. This is in contrast to turn-based collaborative engagement discussed previously in Section 5.4 where the user and AI take turns.

One novel engagement related to simultaneous engagement can be found in Reframer (Lawton et al., 2023), a human-AI drawing system that enables the user and generative model to draw concurrently in real-time, where the user interacts with the system through drawing, and these interactions, in turn impact what the AI draws, and vice-versa. The system also includes a prompt, drawing models, such as draw, focus, and explore, and sliders to control more advanced drawing features, such as a slider to allow the AI more freedom, a recency slider that indicates to the AI to either use more history when drawing concurrently or the immediate past and so on.

More recently, there has been work on multi-agent approaches (Wei et al., 2023b). In terms of simultaneous collaborative engagement, there is also the situation where multiple AIs or agents (multi-agent)can work together simultaneously towards the same goal. Take Repilot (Wei et al., 2023b) for example, a multi-agent approach that works side by side with Barke et al. (2023b)'s Co-pilot, refining its outputs. Essentially, Repilot was created as a generative system that takes Co-pilot's code suggestions and outputs as its own inputs. Then it refines Co-pilot's suggestions and iterates on them to reduce hallucinations and improve the accuracy of the code. Overall, multi-agent approaches, such as Repilot, are effective simultaneous engagements that add another level of AI generation to an already existing generative process.

6 Applications

Generative AI is transformative not only due to its flexibility in engagement but also due to its wide range of applications. As seen in Figure 6.1, the ability to be an effective tool in everything from content creation (Sec. 6.1), to data analysis and forecasting (Sec. 6.2), to research and development (Sec. 6.3), to task automation (Sec. 6.4), and to personal assistance (Sec. 6.5) makes it important to categorize them in this way. In doing so, we can highlight key applications of generative AI and explicitly explore the specific benefits that every application type provides users. Furthermore, we will focus on the interplay between these applications and the impact it has on the different UI techniques used by the generative system.

6.1 Content Creation

Overview and Examples Content creation in the context of generative AI consists of prompting the generative system to create a specific piece of generated material with specific parameters. Content creation consists of anything from creating or editing visual media (Liu et al., 2023a; Tang et al., 2024; Jeon et al., 2021; Davis et al., 2015; Wang et al., 2024d; Chung & Adar, 2023) to creating written content (Achiam et al., 2023; Yuan et al., 2022; Chung et al., 2022; Wang et al., 2024b; Suh et al., 2023a; Wang et al., 2024c) or audio content (Agostinelli et al., 2023; Copet et al., 2023; Borsos et al., 2023). The incredible part about generative AI in content creation is that it lowers the barrier to entry for many creatives. Take, for example, Tang et al. (2024)'s RealFill, a generative fill system that responsively and accurately "fills" in gaps in existing images or expands them in a way that is in line with the user's requests. This case also exemplifies that content creation generative systems can be used both to edit content and create it from scratch. Furthermore, content

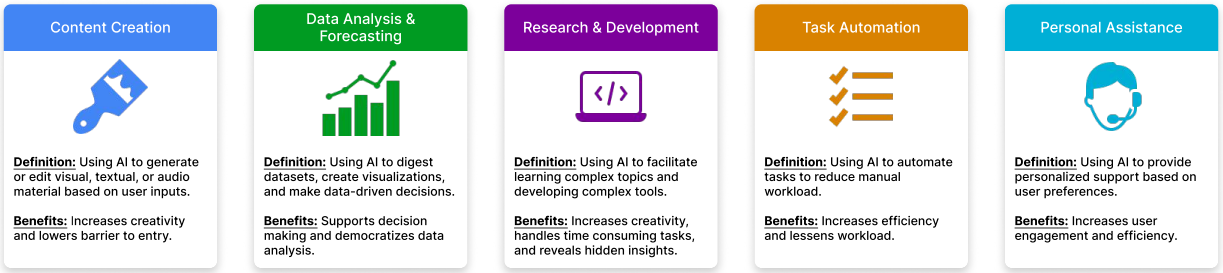


Figure 14: **Applications Taxonomy: A high-level visual summary of the different generative AI Applications (Sec. 6).**

creation systems are especially effective at lowering the barrier to entry for many creative domains as they make it easier for the average user to be creative without having to be an expert in the domain. Furthermore, it also speeds up the creative process as the user can spend less time writing, editing, and pushing pixels, and more time creating and brainstorming.

Common User Interactions Content creation platforms can essentially come in all shapes and sizes, but there are some UI layouts (Sec. 4) that, through our literature reviews, we have found to be more common. For example, content creation often goes hand-in-hand with conversational user interfaces (Sec. 4.1) and canvas user interfaces (Sec. 4.2). The conversational user interface is especially useful in generating written content but can also be used to generate images. Meanwhile, canvas user interfaces are often best used in generating visual content. Furthermore, content creation systems often use a mix of different user-guided techniques, but there is often a large emphasis on prompting (Sec. 3.1). All this to say, that while content creation systems use a wide range of user interaction techniques, these were the ones found to be the most common.

6.2 Data Analysis and Forecasting

Overview and Examples Data analysis and forecasting is one of the applications where generative AI has the potential to make the largest impact. With data becoming more and more valuable and data sets becoming larger and larger, gleaning insights from them has simultaneously become more important and more time-consuming. Generative AI can help data experts both digest and glean insights from data (Goyal et al., 2024; Achiam et al., 2023; Fu et al., 2023) and also help them visualize and present data in an easy-to-understand format (Singh et al.; Setlur et al., 2016). In doing so, generative AI makes it easier to make data-backed decisions and increases the efficiency and speed at which those decisions can be made.

Common User Interactions Data Analysis and Forecasting generative platforms take many different shapes, but through our survey, we found that there were many different throughlines. In terms of UI layouts, we found that most data-focused generative systems have mostly are either information visualization canvases, modular user interfaces, or even conversational user interfaces (Sec. 4). Modular user interfaces were effective at acting almost as a dashboard and allowing users to adjust many parameters at once, while information visualization canvases were more focused on a single visualization. Conversational UIs were most effective when a user was attempting to ask for specific insights about the data. In terms of common user-guided interactions, data-focused systems often use many system and parameter manipulation techniques to adjust the data's parameters (Sec. 3.3). All in all, there is no one correct way to create a data analysis and forecasting generative system, these were just the common through-lines that existed for these types of applications.

6.3 Research and Development

Overview and Examples Generative AI has begun to revolutionize research and development in much of the same ways that it revolutionizes other fields. In terms of research, generative AI has made it easier to learn complex topics and skills as it allows users to learn in a personalized environment (Cao et al., 2024;

Malandri et al., 2023). Meanwhile, it also can help develop products and do research-related tasks that normally would be extremely time-consuming (Petridis et al., 2023; Chang et al., 2023; Liang et al., 2024b; Yen et al., 2023). Like many of the other applications, generative AI in the research and development field makes processes more efficient and enables developers and researchers to spend less time on menial tasks.

Common User Interactions Like all applications, research and development systems use a wide range of user interaction techniques. However, we found that there were common design patterns that were used in many of the research and development applications. For one, research and development systems often rely on conversational user interfaces as these interfaces synergize well with the question and answer cadence common in research and learning (Sec. 4.1). Meanwhile, the contextual UI structure can be used in development and research writing settings to give relevant edits and suggestions to the users' development process within the context of what they are working on (Sec. 4.3).

6.4 Task Automation

Overview and Examples Task automation has become one of the strongest applications of generative AI. Essentially, this application consists of generative AI automating often repetitive tasks that a human would normally do. However, just because the job is repetitive, does not mean it is low-skill, as generative AI is able to automate what are usually considered high-skill tasks (Dogheim & Hussain, 2023; Farrokhi et al., 2024; Jeddi & Bohr, 2020; Kim & Lee, 2019; Oh et al., 2020). Like many of the other applications, automating tasks increases efficiency and eliminates menial tasks. In doing so, this gives users more time to focus on the larger task they are trying to complete and leaves more time for decision-making.

Common User Interactions Since task automation, by nature, is designed not to have a large level of human interaction, one can imagine that there would not be a large amount of user-guided interactions. For the most part, task automation is mostly done in conversational UIs (Sec. 4.1) as this is where the user initiates the automation. Naturally, in the same way, the most common user-guided interaction is prompting (Sec. 3.1), as the user will commonly use prompts to initiate the automation. All in all, while many different user interactions can be used to create a task automation system, these two are the most common.

6.5 Personal Assistance

Overview and Examples One of the strongest aspects of generative AI is its ability to give personalized assistance to each of its users based on the user's preferences and individual needs. Most commonly, it can act as a chatbot that can converse with the user and help them with personalized information or advice (Achiam et al., 2023; Fu et al., 2023; Ang et al., 2023) but can also act in place of customer support to increase engagement and streamline interactions (Brynjolfsson et al., 2023; Verma & Kumari, 2023). Furthermore, it is often commonly used in line to spell check, provide edits when creating something, or offer personalized, contextual help (Fitria, 2021; Jakesch et al., 2023) The benefit of this is that users can access personalized and often professional-level assistance at little to no cost. This application is especially impressive as it sometimes serves as a stand-in for low-level medical, legal, or professional advice (Li et al., 2023b; Yue et al., 2023). All in all, the application of generative AI as a personal assistant changes the way that users can access personalized advice and makes "expert" level guidance more accessible to those who may not normally be able to obtain it.

Common User Interactions While personal assistance applications come in many different forms, these applications skew heavily toward having a conversational user interface (Sec. 4.1). This should come as no surprise, as much of the interaction with personal assistance synergizes well with conversational UIs, which are essentially just conversations. Naturally, this also coincides with the fact that the most common user-guided technique is prompting, most specifically text or speech prompts (Sec. 3.1). However, there are some personal assistant systems that leverage contextual user interfaces (Sec. 4.3), effectively providing in-context personal assistance to the user in real-time. Using contextual user interfaces is effective when creating personal assistants, as they understand the context in which the user is working and can give them specific instructions inline with their task.

7 Open Problems & Challenges

Generative AI is a fast-growing space with many implications for many different fields. For this reason, it is important that research continues in this area, specifically looking at the role that user interfaces and interactions play in generative AI development. In this section, we discuss open problems and highlight important challenges for future work.

7.1 Accessibility

7.1.1 Designing for Disabilities

Generative artificial intelligence has the opportunity to revolutionize how people with disabilities interact with technology, offering tools that can help disabled users with common difficult tasks like writing or "seeing" who is at a meeting (Goodman et al., 2022; Iyer, 2023). Furthermore, Glazko et al. (2023) talks about how generative AI can be used to benefit neurodivergent users who have trouble using the correct tone in different messages at work. They can use products like ChatGPT (Achiam et al., 2023) or Gemini (Fu et al., 2023) to help adjust the tones of their messages. However, the user experience for disabled users is still lackluster in that many disabled users cannot use it entirely independently. Many times, disabled users need assistance from coworkers to verify that the system understood their original query (Glazko et al., 2023).

Future generative artificial intelligence applications need to take more care of users with disabilities when designing interfaces. Many of the works surveyed did not address how their system could be used by users with disabilities, nor did they address how accessibility would play a role in the future. A simple solution is to involve designers or users more in the design process and to center their needs while designing the system's user interfaces. However, extensive research has been done on the role accessibility should play when designing *traditional* user interfaces and experiences (Petrie & Bevan, 2009; Sauer et al., 2020; Aizpurua et al., 2016), but the field still lacks extensive literature on how designers should specifically design generative AI user interfaces for those with disabilities.

7.1.2 Designing for Limited Technical Literacy

Generative artificial intelligence has the opportunity to level the playing field for those traditionally marginalized by technology access. It does this by lowering the barrier to entry for many fields of practice like data science, illustration generation, fictional and nonfictional writing, and so much more. However, many users lack the technical literacy to take full advantage of the aforementioned benefits which leads users to be frustrated with or abandon generative applications altogether.

The user interfaces of common generative AI applications are easy to use, as many times, the UI elements are limited to chat boxes and conversational history (Sec. 4.1) (Achiam et al., 2023; Fu et al., 2023; Betker et al., 2023). However, many of the applications we surveyed have complicated user interfaces that are most likely targeted at users who already have some understanding of the process. This can be done if the designs place a greater emphasis on educational design elements or even design elements that make discoverability much easier. User interfaces of generative AI have the opportunity to democratize their systems capabilities by designing for users who exhibit a large range of technical literacy, not just ones who are already capable of completing a task without generative AI. Therefore, future research should focus on how user interfaces should be designed to better assist users and give personalized experiences to users with both high and low amounts of technical literacy.

7.2 The Future of Generative AI

7.2.1 Designing for Future User Interfaces

Just as quickly as it rose to prominence, generative AI will continue to grow and embed itself within new technology. We have already begun to see how generative AI plays roles in virtual (Konenkov et al., 2024; Giunchi et al., 2024) and tangible user interfaces (Doe et al., 2019). This evolution of generative AI technology will necessitate much more research into how user interfaces should be developed for these new

mediums. Because these interactions often occur in three dimensions, many of the best practices for designing two-dimensional applications are not applicable. Therefore, additional work needs to be done to outline and survey the user interactions of new technologies as they emerge.

On a similar note, as mentioned, multi-agent approaches, like Wei et al. (2023b)'s Repilot, are becoming more common. Essentially, these multi-agent systems consist of when generative AI agents interact with one another rather than directly with the user. Therefore, as these multi-agent systems emerge, there should be extensive work surveying the common user interactions that exist within them. The user interfaces of multi-agent systems should be transparent and clearly communicate to users what is happening and which agent is acting at what time. In continuing research on emerging generative AI technology, such as this one, we can get a better understanding of the user interactions that are needed to best serve the user.

7.2.2 Designing for Growth and Scalability

As generative AI continues to grow and is forecasted to grow significantly in the coming decade (Halal et al., 2016), it is important the user interfaces grow and evolve with it. In almost every field, generative AI is forecasted to grow, and with that comes a larger need for the application to handle a more diverse set of users and use cases. Furthermore, it is important to retain consistent user interaction patterns as the application grows, preventing users from having to continuously relearn how to use the app with every iteration. One of the major difficulties that comes with scaling up a generative AI application is the increased abilities and complexities that the application must now account for. At the same time, they must account for these complexities while also ensuring that the user interface is straightforward and easy to use. This will address the concern that user interfaces could become overwhelming and cognitively complex with more features being added to the generative AI application. All in all, as generative AI applications grow in size, designers must adapt to the capabilities of the system while also continuing to meet the needs of the user.

7.3 Ethics

7.3.1 Designing for Harmful Bias Mitigation

Current literature has addressed and surveyed the current state of both bias and harmful bias mitigation in generative AI. Literature like Gallegos et al. (2024) detail how generative applications inherit biases present within the training data that they were provided. Therefore, generative applications are known to propagate harmful stereotypes, incorrect data, and skewed answers. Given this, there is still a need to explore the role that good design and effective user interaction design techniques can play in mitigating bias in generative artificial intelligence. Most of the user interfaces that were surveyed in this paper do not address that their generative systems may be biased in some harmful way, despite the fact that many of them are biased (Gallegos et al., 2024).

Future user interface design approaches can potentially address this by designing in features that are transparent about the bias that the system might exhibit. Furthermore, more research can be done to discover if there are ways to leverage user interaction techniques to mitigate harmful biases, such as being able to give explicit feedback on whether or not an answer is explicitly biased. More work can be done to leverage a collaborative approach between users and generative systems to mitigate harmful bias.

7.3.2 Designing to Prevent Misuse

Generative AI is very much a two-sided coin: on one hand, it is an amazing invention capable of lowering the barrier to entry for many complex fields. On the other hand, it can be extensively misused. If misused, generative AI can facilitate everything from opinion manipulation and misinformation to plagiarism and much more (Marchal et al., 2024; Ferrara, 2024; Wach et al., 2023). This misuse can have long-lasting and profound negative impacts on society; for this reason, developers can and should design with these considerations in mind when creating generative AI systems.

Given this, it is crucial for future works to consider designing user interfaces that include elements aimed at preventing misuse. Among the generative AI systems we surveyed, there was little to no warning from these systems that a specific query could lead to misuse. For instance, if a user asks a generative system to create

a chatbot that responds to social media posts with misinformation, the system should inform the user that this behavior can be considered misuse and explain why. By incorporating such warnings, generative systems would have built-in protections that could prevent misuse before it occurs. This, in turn, could prevent many of the negative societal impacts that generative AI could have moving forward.

7.4 Data Privacy

As generative AI grows larger and larger, many systems are integrating into various aspects of daily life. Through this process, users are more likely to surrender sensitive personal data to generative systems, either knowingly or unknowingly. Often, this is for good reason, as generative applications thrive when they can provide personalized experiences to each user. However, there should be guardrails in place to prevent generative AI applications from collecting overly sensitive user data without the user's permission.

That being said, there are few design patterns specifically created for data transparency. For this reason, more research is needed on best practices for communicating to users how their data is being used. There should be user interaction features aimed explicitly at ensuring that users understand what data is being collected and how it is utilized. Furthermore, designers should explore more solutions that allow users to choose which personal data is shared. They can also investigate educational design flows that help users understand the best practices for sharing data with generative systems. By empowering users, designers can ensure that they are well-informed and have control over their data privacy preferences.

7.5 Open Problems in Designing for Generative AI Interpretation

With the rise of generative AI, numerous methods have emerged to interpret different types of generative models, such as language models, text-to-image models, and multimodal language models. These methods mainly aim to identify internal components or representations within the generative models that are causally responsible for specific outputs. The ultimate goal is to enhance transparency for end-users while empowering them to control various aspects of the generated content.

In language models, studies like (Meng et al., 2023a;b; Hartvigsen et al., 2022) have pinpointed causal components that can be adjusted to alter final outputs. In the context of text-to-image models, research such as (Basu et al., 2024b;a; Arad et al., 2024; Gandikota et al., 2023) has identified interpretable subspaces that can influence the generation of specific concepts. Despite significant technical advancements in understanding model components, there remains a gap in the availability of comprehensive, end-to-end systems that allow users to interactively engage with these methods. Recent research has extended these methods to focus on understanding the computational sub-graphs, or "circuits" within generative models (Elhage et al., 2021; Prakash et al., 2024; Hanna et al., 2023; Wang et al., 2022). This increased complexity in circuit analysis has highlighted the importance of designing interactive systems to facilitate deeper insights. Recently the community has shifted towards finding controllable steering vectors which when added to the language model can elicit certain behaviours (e.g., improving refusal rate to harmful queries) to an end-user (Arditi et al., 2024; Bricken et al., 2023; Li et al., 2024; Stolfo et al., 2024; Cunningham et al., 2023). There are existing interactive systems for steering vectors in language models, but many are closed-source[1]. In contrast, Transluce[2], an open and independent technology lab, has initiated efforts to develop interactive tools for understanding language models, aiming to provide users with greater transparency and control.

With the rise of advanced mechanistic interpretability tools for generative AI, the next logical step is to develop unified interactive systems that allow users to engage with these tools seamlessly. This approach not only enhances user interaction but also paves the way for innovative and robust interactive system development.

[1] https://goodfire.ai/
[2] https://transluce.org/

8 Conclusion

We have presented a comprehensive survey detailing various dimensions of generative AI user interactions and the different design techniques used to facilitate them. We began by expanding on common user interaction concepts, defining a *user-guided interaction* as a form of interaction that is explicitly and deliberately initiated by the user. Furthermore, we detailed the different input mediums that users can utilize when interacting with generative artificial intelligence. Next, we presented four instructional taxonomies that outline how current generative systems are designed for generative AI. These taxonomies include a taxonomy of user-guided interaction techniques for generative AI systems, a taxonomy of common user interface layouts, a taxonomy of human-AI engagement levels, and a taxonomy of applications and use cases for generative AI systems. Our first taxonomy categorizes different types of generative AI user-guided interactions, focusing on interaction patterns explicitly initiated by the user. The second taxonomy addresses key user interface layouts, specifically examining how and when they are employed. Thirdly, the taxonomy on human-AI engagement levels for generative AI systems explores the extent of user involvement in the generative process and the levels of deliberate interaction taking place. Finally, the taxonomy of generative AI applications highlights the various ways generative AIs are utilized and the user interfaces that best align with these use cases. We conclude our survey with several actionable open problems identified during our exploration of generative applications. Through this work, we aim to create a compendium of common generative AI user interactions to lower the barrier to entry for designers and non-designers alike.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Forest Agostinelli et al. Musiclm: Generating music from text. *arXiv preprint arXiv:2301.11325*, 2023. URL https://google-research.github.io/seanet/musiclm/examples/.

Amaia Aizpurua, Simon Harper, and Markel Vigo. Exploring the relationship between web accessibility and user experience. *International Journal of Human-Computer Studies*, 91:13–23, 2016.

Emre Aksan, Fabrizio Pece, and Otmar Hilliges. Deepwriting: Making digital ink editable via deep generative modeling. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2018. doi: 10.1145/3173574.3173708.

Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, et al. Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35:23716–23736, 2022.

Paulo Alonso Gaona García, David Martín-Moncunill, Salvador Sánchez-Alonso, and Ana Fermoso Garcia. A usability study of taxonomy visualisation user interfaces in digital repositories. *Online Information Review*, 38(2):284–304, 2014.

Shivangi Aneja, Justus Thies, Angela Dai, and Matthias Nießner. Clipface: Text-guided editing of textured 3d morphable models. In *ACM SIGGRAPH 2023 Conference Proceedings*, pp. 1–11, 2023.

Beng Heng Ang, Sujatha Das Gollapalli, and See Kiong Ng. Socratic question generation: A novel dataset, models, and evaluation. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 147–165, 2023.

Victor Nikhil Antony and Chien-Ming Huang. Id.8: Co-creating visual stories with generative ai. 2024. ISSN 2160-6455. doi: 10.1145/3672277. URL https://doi.org/10.1145/3672277.

Dana Arad, Hadas Orgad, and Yonatan Belinkov. Refact: Updating text-to-image models by editing the text encoder, 2024. URL https://arxiv.org/abs/2306.00738.

Riku Arakawa, Hiromu Yakura, and Sosuke Kobayashi. Vocabencounter: Nmt-powered vocabulary learning by presenting computer-generated usages of foreign words into users' daily lives. New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. doi: 10.1145/3491102.3501839. URL https://doi.org/10.1145/3491102.3501839.

Riku Arakawa, Hiromu Yakura, and Masataka Goto. Catalyst: domain-extensible intervention for preventing task procrastination using large generative models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–19, 2023.

Andy Arditi, Oscar Obeso, Aaquib Syed, Daniel Paleka, Nina Panickssery, Wes Gurnee, and Neel Nanda. Refusal in language models is mediated by a single direction, 2024. URL https://arxiv.org/abs/2406.11717.

Mia Magdalena Bangerl, Katharina Stefan, and Viktoria Pammer-Schindler. Explorations in human vs. generative ai creative performances: A study on human-ai creative potential. 2024.

Soumya Barikeri, Anne Lauscher, Ivan Vulić, and Goran Glavaš. Redditbias: A real-world resource for bias evaluation and debiasing of conversational language models. *arXiv preprint arXiv:2106.03521*, 2021.

Shraddha Barke, Michael B. James, and Nadia Polikarpova. Grounded copilot: How programmers interact with code-generating models. 7(OOPSLA1), 2023a. doi: 10.1145/3586030. URL https://doi.org/10.1145/3586030.

Shraddha Barke, Michael B James, and Nadia Polikarpova. Grounded copilot: How programmers interact with code-generating models. *Proceedings of the ACM on Programming Languages*, 7(OOPSLA1):85–111, 2023b.

Samyadeep Basu, Keivan Rezaei, Priyatham Kattakinda, Vlad I Morariu, Nanxuan Zhao, Ryan A. Rossi, Varun Manjunatha, and Soheil Feizi. On mechanistic knowledge localization in text-to-image generative models. In *Forty-first International Conference on Machine Learning*, 2024a. URL https://openreview.net/forum?id=fsVBsxjRER.

Samyadeep Basu, Nanxuan Zhao, Vlad I Morariu, Soheil Feizi, and Varun Manjunatha. Localizing and editing knowledge in text-to-image generative models. In *The Twelfth International Conference on Learning Representations*, 2024b. URL https://openreview.net/forum?id=Qmw9ne6SOQ.

James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. Improving image generation with better captions. *Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf*, 2(3):8, 2023.

Elisabeth André (Augsburg University Germany) Birgit Endrass (Augsburg University, Germany) and Denmark) Matthias Rehm (Aalborg University. Towards culturally-aware virtual agent systems. *Handbook of Research on Culturally-Aware Information Technology: Perspectives and Models*, 2011.

Zalán Borsos, Raphaël Marinier, Damien Vincent, Eugene Kharitonov, Olivier Pietquin, Matt Sharifi, Dominik Roblek, Olivier Teboul, David Grangier, Marco Tagliasacchi, et al. Audiolm: a language modeling approach to audio generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.

Trenton Bricken, Adly Templeton, Joshua Batson, Brian Chen, Adam Jermyn, Tom Conerly, Nick Turner, Cem Anil, Carson Denison, Amanda Askell, Robert Lasenby, Yifan Wu, Shauna Kravec, Nicholas Schiefer, Tim Maxwell, Nicholas Joseph, Zac Hatfield-Dodds, Alex Tamkin, Karina Nguyen, Brayden McLean, Josiah E Burke, Tristan Hume, Shan Carter, Tom Henighan, and Christopher Olah. Towards monosemanticity: Decomposing language models with dictionary learning. *Transformer Circuits Thread*, 2023. https://transformer-circuits.pub/2023/monosemantic-features/index.html.

Erik Brynjolfsson, Danielle Li, and Lindsey R Raymond. Generative ai at work. Working Paper 31161, National Bureau of Economic Research, April 2023. URL http://www.nber.org/papers/w31161.

Tianyuan Cai, Shaun Wallace, Tina Rezvanian, Jonathan Dobres, Bernard Kerr, Samuel Berlow, Jeff Huang, Ben D Sawyer, and Zoya Bylinskii. Personalized font recommendations: Combining ml and typographic guidelines to optimize readability. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, pp. 1–25, 2022.

Chuxue Cao, Ziqing Yuan, and Hailiang Chen. Scholargpt: Fine-tuning large language models for discipline-specific academic paper writing. 2024.

Joseph Chee Chang, Amy X Zhang, Jonathan Bragg, Andrew Head, Kyle Lo, Doug Downey, and Daniel S Weld. Citesee: Augmenting citations in scientific papers with persistent and personalized historical context. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2023.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021.

Zhiyuan Chen and Bing Liu. *Lifelong machine learning*. Springer Nature, 2022.

Mark H Chignell. A taxonomy of user interface terminology. *ACM SIGCHI Bulletin*, 21(4):27, 1990.

Joseph Cho, Fachrina Dewi Puspitasari, Sheng Zheng, Jingyao Zheng, Lik-Hang Lee, Tae-Ho Kim, Choong Seon Hong, and Chaoning Zhang. Sora as an agi world model? a complete survey on text-to-video generation. *arXiv preprint arXiv:2403.05131*, 2024.

John Joon Young Chung and Eytan Adar. 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*, pp. 1–17, 2023.

John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. Talebrush: Sketching stories with generative pretrained language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–19, 2022.

Jade Copet, Alexandre Defossez, Yossi Adi, and Gabriel Synnaeve. Musicgen: Simple and controllable music generation. *arXiv preprint arXiv:2301.11325*, 2023. URL https://huggingface.co/facebook/musicgen-small.

Hoagy Cunningham, Aidan Ewart, Logan Riggs, Robert Huben, and Lee Sharkey. Sparse autoencoders find highly interpretable features in language models, 2023. URL https://arxiv.org/abs/2309.08600.

Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, Sanat Moningi, and Brian Magerko. Drawing apprentice: An enactive co-creative agent for artistic collaboration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, pp. 185–186, 2015.

Nicholas Davis, Safat Siddiqui, Panote Karimi, Mary Lou Maher, and Kazjon Grace. Creative sketching partner: A co-creative sketching tool to inspire design. In *Proceedings of the 10th International Conference on Computational Creativity*, pp. 358–359, 2019. doi: 10.1007/978-3-031-12807-3_11.

H Onan Demirel, Molly H Goldstein, Xingang Li, and Zhenghui Sha. Human-centered generative design framework: an early design framework to support concept creation and evaluation. *International Journal of Human–Computer Interaction*, 40(4):933–944, 2024.

Gelei Deng, Yi Liu, Yuekang Li, Kailong Wang, Ying Zhang, Zefeng Li, Haoyu Wang, Tianwei Zhang, and Yang Liu. Jailbreaker: Automated jailbreak across multiple large language model chatbots. *arXiv preprint arXiv:2307.08715*, 2023.

Manoj Deshpande. Towards co-build: An architecture machine for co-creative form-making. Master's thesis, The University of North Carolina at Charlotte, 2020.

Zijian Ding. Advancing gui for generative ai: Charting the design space of human-ai interactions through task creativity and complexity. In *Companion Proceedings of the 29th International Conference on Intelligent User Interfaces*, pp. 140–143, 2024.

John Doe, Jane Smith, and Kevin Lee. Deepscope: Hci platform for generative cityscape visualization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 123–132, 2019. doi: 10.1145/3313831.3376722.

Gaidaa Maher Dogheim and Abrar Hussain. Patient care through ai-driven remote monitoring: Analyzing the role of predictive models and intelligent alerts in preventive medicine. *Journal of Contemporary Healthcare Analytics*, 7(1):94–110, 2023.

Ali Dorri, Salil S. Kanhere, and Raja Jurdak. Multi-agent systems: A survey. *IEEE Access*, 6:28573–28593, 2018. doi: 10.1109/ACCESS.2018.2831228.

Xuejun Du, Pengcheng An, Justin Leung, April Li, Linda E Chapman, and Jian Zhao. Deepthink: Designing and probing human-ai co-creation in digital art therapy. *International Journal of Human-Computer Studies*, 181:103139, 2024.

Nelson Elhage, Neel Nanda, Catherine Olsson, Tom Henighan, Nicholas Joseph, Ben Mann, Amanda Askell, Yuntao Bai, Anna Chen, Tom Conerly, Nova DasSarma, Dawn Drain, Deep Ganguli, Zac Hatfield-Dodds, Danny Hernandez, Andy Jones, Jackson Kernion, Liane Lovitt, Kamal Ndousse, Dario Amodei, Tom Brown, Jack Clark, Jared Kaplan, Sam McCandlish, and Chris Olah. A mathematical framework for transformer circuits. *Transformer Circuits Thread*, 2021. https://transformer-circuits.pub/2021/framework/index.html.

Babajide Tolulope Familoni and Nneamaka Chisom Onyebuchi. Advancements and challenges in ai integration for technical literacy: a systematic review. *Engineering Science & Technology Journal*, 5(4):1415–1430, 2024.

Mehrdad Farrokhi, Fatemeh Taheri, Amir Moeini, Masoud Farrokhi, Mousavi Zadeh Sayed Alireza, Maryam Farahmandsadr, Ehsan Bahrami Hezaveh, Ali Davoodi, Sepideh Niknejad, Mahmonir Bayanati, et al. Artificial intelligence for remote patient monitoring: Advancements, applications, and challenges. *Kindle*, 4 (1):1–261, 2024.

Emilio Ferrara. Genai against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, pp. 1–21, 2024.

James Finnie-Ansley, Paul Denny, Brett A Becker, Andrew Luxton-Reilly, and James Prather. The robots are coming: Exploring the implications of openai codex on introductory programming. In *Proceedings of the 24th Australasian Computing Education Conference*, pp. 10–19, 2022.

Tira Nur Fitria. Grammarly as ai-powered english writing assistant: Students' alternative for writing english. *Metathesis: Journal of English Language, Literature, and Teaching*, 5(1):65–78, 2021.

Chaoyou Fu, Renrui Zhang, Zihan Wang, Yubo Huang, Zhengye Zhang, Longtian Qiu, Gaoxiang Ye, Yunhang Shen, Mengdan Zhang, Peixian Chen, Sirui Zhao, Shaohui Lin, Deqiang Jiang, Di Yin, Peng Gao, Ke Li, Hongsheng Li, and Xing Sun. A challenger to gpt-4v? early explorations of gemini in visual expertise, 2023. URL https://arxiv.org/abs/2312.12436.

Isabel O Gallegos, Ryan A Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, and Nesreen K Ahmed. Bias and fairness in large language models: A survey. *Computational Linguistics*, pp. 1–79, 2024.

Rohit Gandikota, Hadas Orgad, Yonatan Belinkov, Joanna Materzyńska, and David Bau. Unified concept editing in diffusion models, 2023. URL https://arxiv.org/abs/2308.14761.

Difei Gao, Lei Ji, Luowei Zhou, Kevin Qinghong Lin, Joya Chen, Zihan Fan, and Mike Zheng Shou. Assistgpt: A general multi-modal assistant that can plan, execute, inspect, and learn, 2023. URL https://arxiv.org/abs/2306.08640.

Maíra Gatti, Paulo Cavalin, Samuel Barbosa Neto, Claudio Pinhanez, Cícero dos Santos, Daniel Gribel, and Ana Paula Appel. Large-scale multi-agent-based modeling and simulation of microblogging-based online social network. In *Multi-Agent-Based Simulation XIV: International Workshop, MABS 2013, Saint Paul, MN, USA, May 6-7, 2013, Revised Selected Papers*, pp. 17–33. Springer, 2014.

Katy Ilonka Gero, Vivian Liu, and Lydia B. Chilton. Sparks: Inspiration for science writing using language models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, pp. 1002–1019, 2022. doi: 10.1145/3532106.3533455.

Daniele Giunchi, Nels Numan, Elia Gatti, and Anthony Steed. Dreamcodevr: Towards democratizing behavior design in virtual reality with speech-driven programming. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 579–589. IEEE, 2024.

Kate S. Glazko, Momona Yamagami, Aashaka Desai, Kelly Avery Mack, Venkatesh Potluri, Xuhai Xu, and Jennifer Mankoff. An autoethnographic case study of generative artificial intelligence in accessibility. *ACM Digital Library*, October 2023. URL https://dl.acm.org/doi/fullHtml/10.1145/3597638.3614548.

Frederic Gmeiner, Humphrey Yang, Lining Yao, Kenneth Holstein, and Nikolas Martelaro. Exploring challenges and opportunities to support designers in learning to co-create with ai-based manufacturing design tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–20, 2023.

Steven M Goodman, Erin Buehler, Patrick Clary, Andy Coenen, Aaron Donsbach, Tiffanie N Horne, Michal Lahav, Robert MacDonald, Rain Breaw Michaels, Ajit Narayanan, et al. Lampost: Design and evaluation of an ai-assisted email writing prototype for adults with dyslexia. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1–18, 2022.

Raghav Goyal, Effrosyni Mavroudi, Xitong Yang, Sainbayar Sukhbaatar, Leonid Sigal, Matt Feiszli, Lorenzo Torresani, and Du Tran. Minotaur: Multi-task video grounding from multimodal queries. *arXiv preprint arXiv:2302.08063*, 2023.

Sagar Goyal, Eti Rastogi, Sree Prasanna Rajagopal, Dong Yuan, Fen Zhao, Jai Chintagunta, Gautam Naik, and Jeff Ward. Healai: A healthcare llm for effective medical documentation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pp. 1167–1168, 2024.

Gaëtan Hadjeres, François Pachet, and Frank Nielsen. Deepbach: a steerable model for bach chorales generation. In *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*, 2017. URL https://www.researchgate.net/publication/332141615_DEEPBACH_A_STEERABLE_MODEL_FOR_BACH_CHORALES_GENERATION.

William Halal, Jonathan Kolber, and Owen Davies. Forecasts of ai and future jobs in 2030: Muddling through likely, with two alternative scenarios. *Journal of futures studies*, 21(2), 2016.

Michael Hanna, Ollie Liu, and Alexandre Variengien. How does GPT-2 compute greater-than?: Interpreting mathematical abilities in a pre-trained language model. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=p4PckNQR8k.

Yaru Hao, Zewen Chi, Li Dong, and Furu Wei. Optimizing prompts for text-to-image generation. *Advances in Neural Information Processing Systems*, 36, 2024.

Thomas Hartvigsen, Swami Sankaranarayanan, Hamid Palangi, Yoon Kim, and Marzyeh Ghassemi. Aging with grace: Lifelong model editing with discrete key-value adaptors. *ArXiv*, abs/2211.11031, 2022. URL https://api.semanticscholar.org/CorpusID:253735429.

Tomayess Issa and Pedro Isaias. Usability and human–computer interaction (hci). In *Sustainable design: HCI, usability and environmental concerns*, pp. 23–40. Springer, 2022.

IXDF. Human-ai interaction (hax), Apr 2024. URL https://www.interaction-design.org/literature/topics/human-ai-interaction.

Chitra Iyer. How ai can help with digital workplace accessibility. *Reworked*, September 2023. URL https://www.reworked.co/digital-workplace/how-ai-can-help-with-digital-workplace-accessibility/.

Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. Co-writing with opinionated language models affects users' views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. ACM, April 2023. doi: 10.1145/3544548.3581196. URL http://dx.doi.org/10.1145/3544548.3581196.

Anna Jaruga-Rozdolska. Artificial intelligence as part of future practices in the architect's work: Midjourney generative tool as part of a process of creating an architectural form. *Architectus*, (3 (71):95–104, 2022a.

Anna Jaruga-Rozdolska. Artificial intelligence as part of future practices in the architect's work: Midjourney generative tool as part of a process of creating an architectural form. *Architectus*, (3 (71):95–104, 2022b.

Zineb Jeddi and Adam Bohr. Remote patient monitoring using artificial intelligence. In *Artificial intelligence in healthcare*, pp. 203–234. Elsevier, 2020.

Seonmin Jeon, Seungbae Jin, Taehyun Lee, Seongmin Kim, Youngkeun Park, and Byungjoo Lee. Fashionq: An ai-driven creativity support tool for facilitating ideation in fashion design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021. URL https://www.researchgate.net/publication/345324055_FashionQ_An_Interactive_Tool_for_Analyzing_Fashion_Style_Trend_with_Quantitative_Criteria.

Hyeonhak Jeong, Minki Chun, Hyunmin Lee, Seung Young Oh, and Hyunggu Jung. Wataa: Web alternative text authoring assistant for improving web content accessibility. In *Companion proceedings of the 28th international conference on intelligent user interfaces*, pp. 41–45, 2023.

Peiling Jiang, Jude Rayan, Steven P. Dow, and Haijun Xia. Graphologue: Exploring large language model responses with interactive diagrams. New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701320.

Florian Kadner, Yannik Keller, and Constantin Rothkopf. Adaptifont: Increasing individuals' reading speed with a generative font model and bayesian optimization. In *Proceedings of the 2021 chi conference on human factors in computing systems*, pp. 1–11, 2021.

Dongwhan Kim and Joonhwan Lee. Designing an algorithm-driven text generation system for personalized and interactive news reading. *International Journal of Human–Computer Interaction*, 35(2):109–122, 2019.

Eunseo Kim, Jeongmin Hong, Hyuna Lee, and Minsam Ko. Colorbo: Envisioned mandala coloringthrough human-ai collaboration. New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391443. URL https://doi.org/10.1145/3490099.3511135.

Jeongyeon Kim, Sangho Suh, Lydia B. Chilton, and Haijun Xia. Metaphorian: Leveraging large language models to support extended metaphor creation for science writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, pp. 115–135, 2023a. doi: 10.1145/3563657.3595996.

Tae Soo Kim, Arghya Sarkar, Yoonjoo Lee, Minsuk Chang, and Juho Kim. Lmcanvas: Object-oriented interaction to personalize large language model-powered writing environments. *arXiv preprint arXiv:2303.15125*, 2023b.

Taewan Kim, Donghoon Shin, Young-Ho Kim, and Hwajung Hong. Diarymate: Understanding user perceptions and experience in human-ai collaboration for personal journaling. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2024.

Mikhail Konenkov, Artem Lykov, Daria Trinitatova, and Dzmitry Tsetserukou. Vr-gpt: Visual language model for intelligent virtual reality applications. *arXiv preprint arXiv:2405.11537*, 2024.

Tomas Lawton, Kazjon Grace, and Francisco J Ibarrola. When is a tool a tool? user perceptions of system agency in human–ai co-creative drawing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, pp. 1978–1996, 2023.

Hung Le, Hailin Chen, Amrita Saha, Akash Gokul, Doyen Sahoo, and Shafiq Joty. Codechain: Towards modular code generation through chain of self-revisions with representative sub-modules. *arXiv preprint arXiv:2310.08992*, 2023.

Mina Lee, Percy Liang, and Qian Yang. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pp. 1–19, 2022.

Seung Won Lee, Jiin Choi, and Kyung Hoon Hyun. Bogen: Generating part-level 3d designs based on user intention inference through bayesian optimization and variational autoencoder. *arXiv preprint arXiv:2312.02557*, 2023.

Jingyao Li, Pengguang Chen, and Jiaya Jia. Motcoder: Elevating large language models with modular of thought for challenging programming tasks. *arXiv preprint arXiv:2312.15960*, 2023a.

Kenneth Li, Oam Patel, Fernanda Viégas, Hanspeter Pfister, and Martin Wattenberg. Inference-time intervention: Eliciting truthful answers from a language model, 2024. URL https://arxiv.org/abs/2306.03341.

Yunxiang Li, Zihan Li, Kai Zhang, Ruilong Dan, Steve Jiang, and You Zhang. Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge. 2023b. URL https://arxiv.org/abs/2303.14070.

Jenny T Liang, Chenyang Yang, and Brad A Myers. A large-scale survey on the usability of ai programming assistants: Successes and challenges. In *Proceedings of the 46th IEEE/ACM International Conference on Software Engineering*, pp. 1–13, 2024a.

Zhaohui Liang, Xiaoyu Zhang, Kevin Ma, Zhao Liu, Xipei Ren, Kosa Goucher-Lambert, and Can Liu. Storydiffusion: How to support ux storyboarding with generative-ai. *arXiv preprint arXiv:2407.07672*, 2024b.

David Chuan-En Lin and Nikolas Martelaro. Jigsaw: Supporting designers to prototype multimodal applications by chaining ai foundation models. New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. URL https://doi.org/10.1145/3613904.3641920.

Zhiyu Lin, Upol Ehsan, Rohan Agarwal, Samihan Dani, Vidushi Vashishth, and Mark Riedl. Beyond prompts: Exploring the design space of mixed-initiative co-creativity systems. *arXiv preprint arXiv:2305.07465*, 2023.

Kate Lister, Tim Coughlan, Francisco Iniesto, Nick Freear, and Peter Devine. Accessible conversational user interfaces: considerations for design. In *Proceedings of the 17th international web for all conference*, pp. 1–11, 2020.

Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 3dall-e: Integrating text-to-image ai in 3d design workflows, 2023a. URL https://arxiv.org/abs/2210.11603.

Xin Liu and Kati London. Tai: a tangible ai interface to enhance human-artificial intelligence (ai) communication beyond the screen. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pp. 281–285, 2016.

Zhaoyang Liu, Yinan He, Wenhai Wang, Weiyun Wang, Yi Wang, Shoufa Chen, Qinglong Zhang, Zeqiang Lai, Yang Yang, Qingyun Li, et al. Interngpt: Solving vision-centric tasks by interacting with chatgpt beyond language. *arXiv preprint arXiv:2305.05662*, 2023b. URL https://arxiv.org/abs/2305.05662.

Yuwen Lu, Yuewen Yang, Qinyi Zhao, Chengzhi Zhang, and Toby Jia-Jun Li. Ai assistance for ux: A literature review through human-centered ai. *arXiv preprint arXiv:2402.06089*, 2024.

Stephen MacNeil, Andrew Tran, Joanne Kim, Ziheng Huang, Seth Bernstein, and Dan Mogil. Prompt middleware: Mapping prompts for large language models to ui affordances. *arXiv preprint arXiv:2307.01142*, 2023.

Lorenzo Malandri, Fabio Mercorio, Mario Mezzanzanica, and Navid Nobani. Convxai: a system for multimodal interaction with any black-box explainer. *Cognitive Computation*, 15(2):613–644, 2023.

Nahema Marchal, Rachel Xu, Rasmi Elasmar, Iason Gabriel, Beth Goldberg, and William Isaac. Generative ai misuse: A taxonomy of tactics and insights from real-world data. *arXiv preprint arXiv:2406.13843*, 2024.

Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. Directgpt: A direct manipulation interface to interact with large language models. *arXiv preprint arXiv:2310.03691*, 2023.

Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. Locating and editing factual associations in gpt, 2023a. URL https://arxiv.org/abs/2202.05262.

Kevin Meng, Arnab Sen Sharma, Alex Andonian, Yonatan Belinkov, and David Bau. Mass-editing memory in a transformer, 2023b. URL https://arxiv.org/abs/2210.07229.

Alexander H Miller, Will Feng, Adam Fisch, Jiasen Lu, Dhruv Batra, Antoine Bordes, Devi Parikh, and Jason Weston. Parlai: A dialog research software platform. *arXiv preprint arXiv:1705.06476*, 2017.

Alex Nichol, Heewoo Jun, Prafulla Dhariwal, Pamela Mishkin, and Mark Chen. Point-e: A system for generating 3d point clouds from complex prompts. *arXiv preprint arXiv:2212.08751*, 2022.

Sachita Nishal and Nicholas Diakopoulos. Envisioning the applications and implications of generative ai for news media. *arXiv preprint arXiv:2402.18835*, 2024.

Donald A Norman. *The psychology of everyday things*. Basic books, 1988.

Catharine Oertel, Ginevra Castellano, Mohamed Chetouani, Jauwairia Nasir, Mohammad Obaid, Catherine Pelachaud, and Christopher Peters. Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI*, 7:92, 2020.

Changhoon Oh, Jinhan Choi, Sungwoo Lee, SoHyun Park, Daeryong Kim, Jungwoo Song, Dongwhan Kim, Joonhwan Lee, and Bongwon Suh. Understanding user perception of automated news generation system. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.

Katsumi Okuda and Saman Amarasinghe. Askit: Unified programming interface for programming with large language models. In *2024 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)*, pp. 41–54. IEEE, 2024.

Hiroyuki Osone, Jun-Li Lu, and Yoichi Ochiai. Buncho: Ai supported story co-creation via unsupervised multitask learning to increase writers' creativity in japanese. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021. URL https://digitalnature.slis.tsukuba.ac.jp/2021/05/buncho_chi2021/.

Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G. Patil, Ion Stoica, and Joseph E. Gonzalez. Memgpt: Towards llms as operating systems, 2024. URL https://arxiv.org/abs/2310.08560.

Aadarsh Padiyath and Brian Magerko. desainer: Exploring the use of "bad" generative adversarial networks in the ideation process of fashion design. In *Proceedings of the 2021 Creativity and Cognition Conference*, pp. 42:1–42:3, 2021. URL https://www.researchgate.net/publication/352662786_desAIner_Exploring_the_Use_of_Bad_Generative_Adversarial_Networks_in_the_Ideation_Process_of_Fashion_Design.

Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. Anglekindling: Supporting journalistic angle ideation with large language models. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, pp. 1–16, 2023.

Savvas Petridis, Benjamin D Wedin, James Wexler, Mahima Pushkarna, Aaron Donsbach, Nitesh Goyal, Carrie J Cai, and Michael Terry. Constitutionmaker: Interactively critiquing large language models by converting feedback into principles. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*, pp. 853–868, 2024.

Helen Petrie and Nigel Bevan. The evaluation of accessibility, usability, and user experience. *The universal access handbook*, 1:1–16, 2009.

Nikhil Prakash, Tamar Rott Shaham, Tal Haklay, Yonatan Belinkov, and David Bau. Fine-tuning enhances existing mechanisms: A case study on entity tracking, 2024. URL https://arxiv.org/abs/2402.14811.

James Prather, Brent N Reeves, Paul Denny, Brett A Becker, Juho Leinonen, Andrew Luxton-Reilly, Garrett Powell, James Finnie-Ansley, and Eddie Antonio Santos. "it's weird that it knows what i want": Usability and interactions with copilot for novice programmers. *ACM Transactions on Computer-Human Interaction*, 31(1):1–31, 2023.

Cheng Qian, Chi Han, Yi R Fung, Yujia Qin, Zhiyuan Liu, and Heng Ji. Creator: Tool creation for disentangling abstract and concrete reasoning of large language models. *arXiv preprint arXiv:2305.14318*, 2023.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pam Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. *arXiv preprint arXiv:2103.00020*, 2021. URL https://arxiv.org/abs/2103.00020.

Ori Ram, Yoav Levine, Itay Dalmedigos, Dor Muhlgay, Amnon Shashua, Kevin Leyton-Brown, and Yoav Shoham. In-context retrieval-augmented language models. *Transactions of the Association for Computational Linguistics*, 11:1316–1331, 2023.

Zhongwei Ren, Zhicheng Huang, Yunchao Wei, Yao Zhao, and Dongmei Fu. Pixellm: Pixel reasoning with large multimodal model. *arXiv preprint arXiv:2312.02228*, 2023. URL https://arxiv.org/abs/2312.02228.

Jeba Rezwana and Mary Lou Maher. 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): 1–28, 2023.

Steven I. Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D. Weisz. The programmer's assistant: Conversational interaction with a large language model for software development. 2023a. ISBN 9798400701061.

Steven I Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D Weisz. The programmer's assistant: Conversational interaction with a large language model for software development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pp. 491–514, 2023b.

Hanan Salam, Oya Celiktutan, Hatice Gunes, and Mohamed Chetouani. Automatic context-aware inference of engagement in hmi: A survey. *IEEE Transactions on Affective Computing*, 2023.

Juergen Sauer, Andreas Sonderegger, and Sven Schmutz. Usability, user experience and accessibility: towards an integrative model. *Ergonomics*, 63(10):1207–1220, 2020.

Steffen Schneider, Alexei Baevski, Ronan Collobert, and Michael Auli. wav2vec: Unsupervised pre-training for speech recognition. *arXiv preprint arXiv:1904.05862*, 2019. URL https://arxiv.org/abs/1904.05862.

Abigail Sellen and Eric Horvitz. The rise of the ai co-pilot: Lessons for design from aviation and beyond. *Communications of the ACM*, 67(6), Jun 2024. URL https://cacm.acm.org/opinion/the-rise-of-the-ai-co-pilot-lessons-for-design-from-aviation-and-beyond/.

Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. Eviza: A natural language interface for visual analysis. New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341899. URL https://doi.org/10.1145/2984511.2984588.

Hua Shen and Tongshuang Wu. Parachute: Evaluating interactive human-lm co-writing systems. *arXiv preprint arXiv:2303.06333*, 2023.

Jingyu Shi, Rahul Jain, Hyungjun Doh, Ryo Suzuki, and Karthik Ramani. An hci-centric survey and taxonomy of human-generative-ai interactions. *arXiv preprint arXiv:2310.07127*, 2023.

Shuming Shi, Enbo Zhao, Duyu Tang, Yan Wang, Piji Li, Wei Bi, Haiyun Jiang, Guoping Huang, Leyang Cui, Xinting Huang, et al. Effidit: Your ai writing assistant. *arXiv preprint arXiv:2208.01815*, 2022.

Kurt Shuster, Jing Xu, Mojtaba Komeili, Da Ju, Eric Michael Smith, Stephen Roller, Megan Ung, Moya Chen, Kushal Arora, Joshua Lane, et al. Blenderbot 3: a deployed conversational agent that continually learns to responsibly engage. *arXiv preprint arXiv:2208.03188*, 2022.

Candace L Sidner, Christopher Lee, and Neal Lesh. Engagement when looking: behaviors for robots when collaborating with people. In *Diabruck: Proceedings of the 7th workshop on the Semantic and Pragmatics of Dialogue*, pp. 123–130. Citeseer, 2003.

Auste Simkute, Lev Tankelevitch, Viktor Kewenig, Ava Elizabeth Scott, Abigail Sellen, and Sean Rintel. Ironies of generative ai: Understanding and mitigating productivity loss in human-ai interactions. *arXiv preprint arXiv:2402.11364*, 2024.

Nikhil Singh, Lucy Lu Wang, and Jonathan Bragg. Figura11y: Ai assistance for writing scientific alt text.

Alex W Stedmon and Robert J Stone. Re-viewing reality: human factors of synthetic training environments. *International Journal of Human-Computer Studies*, 55(4):675–698, 2001.

Alessandro Stolfo, Vidhisha Balachandran, Safoora Yousefi, Eric Horvitz, and Besmira Nushi. Improving instruction-following in language models through activation steering, 2024. URL https://arxiv.org/abs/2410.12877.

Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. Structured generation and exploration of design space with large language models for human-ai co-creation. *arXiv preprint arXiv:2310.12953*, 2023a.

Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. Sensecape: Enabling multilevel exploration and sensemaking with large language models. New York, NY, USA, 2023b. Association for Computing Machinery. ISBN 9798400701320. URL https://doi.org/10.1145/3586183.3606756.

Simeng Sun, Wenlong Zhao, Varun Manjunatha, Rajiv Jain, Vlad Morariu, Franck Dernoncourt, Balaji Vasan Srinivasan, and Mohit Iyyer. Iga: An intent-guided authoring assistant. *arXiv preprint arXiv:2104.07000*, 2021.

Kyle Swanson, George Liu, David B. Catacutan, Alex Arnold, James Zou, et al. Generative ai for designing and validating easily synthesizable and structurally novel antibiotics. *Nature Machine Intelligence*, 6:338–353, 2024. URL https://www.genengnews.com/topics/infectious-diseases/drug-resistant-bacteria-stymied-by-ai-designed-antibiotics/.

Luming Tang, Nataniel Ruiz, Qinghao Chu, Yuanzhen Li, Aleksander Holynski, David E Jacobs, Bharath Hariharan, Yael Pritch, Neal Wadhwa, Kfir Aberman, et al. Realfill: Reference-driven generation for authentic image completion. *ACM Transactions on Graphics (TOG)*, 43(4):1–12, 2024.

Michael Terry, Chinmay Kulkarni, Martin Wattenberg, Lucas Dixon, and Meredith Ringel Morris. Ai alignment in the design of interactive ai: Specification alignment, process alignment, and evaluation support. *arXiv preprint arXiv:2311.00710*, 2023.

Anh Truong, Peggy Chi, David Salesin, Irfan Essa, and Maneesh Agrawala. Automatic generation of two-level hierarchical tutorials from instructional makeup videos. New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445721. URL https://doi.org/10.1145/3411764.3445721.

Theophanis Tsandilas, Olivier Chapuis, Emmanuel Pietriga, and Michel Beaudouin-Lafon. Designscape: Design with interactive layout suggestions. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*, pp. 535–544, 2015. doi: 10.1145/2807442.2807451.

Stephanie Valencia, Richard Cave, Krystal Kallarackal, Katie Seaver, Michael Terry, and Shaun K. Kane. "the less i type, the better": How ai language models can enhance or impede communication for aac users. New York, NY, USA, 2023a. Association for Computing Machinery. ISBN 9781450394215. doi: 10.1145/3544548.3581560. URL https://doi.org/10.1145/3544548.3581560.

Stephanie Valencia, Richard Cave, Krystal Kallarackal, Katie Seaver, Michael Terry, and Shaun K Kane. "the less i type, the better": How ai language models can enhance or impede communication for aac users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2023b.

Ramesh Kumar Verma and Nalini Kumari. Generative ai as a tool for enhancing customer relationship management automation and personalization techniques. *International Journal of Responsible Artificial Intelligence*, 13(9):1–8, 2023.

Krzysztof Wach, Cong Doanh Duong, Joanna Ejdys, Rūta Kazlauskaitė, Pawel Korzynski, Grzegorz Mazurek, Joanna Paliszkiewicz, and Ewa Ziemba. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of chatgpt. *Entrepreneurial Business and Economics Review*, 11(2):7–30, 2023.

Bryan Wang, Yuliang Li, Zhaoyang Lv, mit Xia, Yan Xu, and Raj Sodhi. Lave: Llm-powered agent assistance and language augmentation for video editing. New York, NY, USA, 2024a. Association for Computing Machinery. ISBN 9798400705083. URL https://doi.org/10.1145/3640543.3645143.

Jiuniu Wang, Zehua Du, Yuyuan Zhao, Bo Yuan, Kexiang Wang, Jian Liang, Yaxi Zhao, Yihen Lu, Gengliang Li, Junlong Gao, et al. Aesopagent: Agent-driven evolutionary system on story-to-video production. *arXiv preprint arXiv:2403.07952*, 2024b.

Kevin Wang, Alexandre Variengien, Arthur Conmy, Buck Shlegeris, and Jacob Steinhardt. Interpretability in the wild: a circuit for indirect object identification in gpt-2 small, 2022. URL https://arxiv.org/abs/2211.00593.

Tiannan Wang, Jiamin Chen, Qingrui Jia, Shuai Wang, Ruoyu Fang, Huilin Wang, Zhaowei Gao, Chunzhao Xie, Chuou Xu, Jihong Dai, et al. Weaver: Foundation models for creative writing. *arXiv preprint arXiv:2401.17268*, 2024c.

Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. Promptcharm: Text-to-image generation through multi-modal prompting and refinement. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–21, 2024d.

Furu Wei et al. Language is not all you need: Aligning perception with language models. *arXiv preprint arXiv:2302.14045*, 2023a. URL https://arxiv.org/abs/2302.14045.

Yuxiang Wei, Chunqiu Steven Xia, and Lingming Zhang. Copiloting the copilots: Fusing large language models with completion engines for automated program repair. New York, NY, USA, 2023b. Association for Computing Machinery. ISBN 9798400703270. doi: 10.1145/3611643.3616271. URL https://doi.org/10.1145/3611643.3616271.

Justin D Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. Design principles for generative ai applications. *arXiv preprint arXiv:2401.14484*, 2024.

Sharon Whitfield and Melissa A Hofmann. Elicit: Ai literature review research assistant. *Public Services Quarterly*, 19(3):201–207, 2023.

Allison Woodruff, Renee Shelby, Patrick Gage Kelley, Steven Rousso-Schindler, Jamila Smith-Loud, and Lauren Wilcox. How knowledge workers think generative ai will (not) transform their industries. *arXiv preprint arXiv:2310.06778*, 2023.

Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. Next-gpt: Any-to-any multimodal llm, 2023. URL https://arxiv.org/abs/2309.05519.

Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J Cai. Promptchainer: Chaining large language model prompts through visual programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–10, 2022.

Zihan Yan, Chunxu Yang, Qihao Liang, and Xiang'Anthony' Chen. Xcreation: A graph-based crossmodal generative creativity support tool. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–15, 2023.

Keda Yang, Zewen Xie, Zhen Li, Xiaoliang Qian, Nannan Sun, Tao He, Zuodong Xu, Jing Jiang, Qi Mei, Jie Wang, et al. Molprophet: A one-stop, general purpose, and ai-based platform for the early stages of drug discovery. *Journal of Chemical Information and Modeling*, 64(8):2941–2947, 2024.

Zhuolin Yang, Wei Ping, Zihan Liu, Vijay Korthikanti, Weili Nie, De-An Huang, Linxi Fan, Zhiding Yu, Shiyi Lan, Bo Li, et al. Re-vilm: Retrieval-augmented visual language model for zero and few-shot image captioning. *arXiv preprint arXiv:2302.04858*, 2023.

Qinghao Ye, Haiyang Xu, Jiabo Ye, Ming Yan, Haowei Liu, Qi Qian, Ji Zhang, Fei Huang, and Jingren Zhou. mplug-owl2: Revolutionizing multi-modal large language model with modality collaboration. *arXiv preprint arXiv:2311.04257*, 2023. URL https://arxiv.org/abs/2311.04257.

Catherine Yeh, Gonzalo Ramos, Rachel Ng, Andy Huntington, and Richard Banks. Ghostwriter: Augmenting collaborative human-ai writing experiences through personalization and agency. *arXiv preprint arXiv:2402.08855*, 2024.

Ryan Yen, Jiawen Zhu, Sangho Suh, Haijun Xia, and Jian Zhao. Coladder: Supporting programmers with hierarchical code generation in multi-level abstraction. *arXiv preprint arXiv:2310.08699*, 2023.

Nur Yildirim, Hannah Richardson, Maria Teodora Wetscherek, Junaid Bajwa, Joseph Jacob, Mark Ames Pinnock, Stephen Harris, Daniel Coelho De Castro, Shruthi Bannur, Stephanie Hyland, et al. Multimodal healthcare ai: identifying and designing clinically relevant vision-language applications for radiology. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–22, 2024.

Eric York. Evaluating chatgpt: Generative ai in ux design and web development pedagogy. In *Proceedings of the 41st ACM International Conference on Design of Communication*, pp. 197–201, 2023.

Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. Wordcraft: story writing with large language models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*, pp. 841–852, 2022.

Shengbin Yue, Wei Chen, Siyuan Wang, Bingxuan Li, Chenchen Shen, Shujun Liu, Yuxuan Zhou, Yao Xiao, Song Yun, Wei Lin, et al. Disc-lawllm: Fine-tuning large language models for intelligent legal services. *arXiv preprint arXiv:2309.11325*, 2023.

Yan Zeng, Xinsong Zhang, Hang Li, Jiawei Wang, Jipeng Zhang, and Wangchunshu Zhou. X 2-vlm: All-in-one pre-trained model for vision-language tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*, 2019.

Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 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*, pp. 1–30, 2023.

Zhenyu Zhang, Ying Sheng, Tianyi Zhou, Tianlong Chen, Lianmin Zheng, Ruisi Cai, Zhao Song, Yuandong Tian, Christopher Ré, Clark Barrett, et al. H2o: Heavy-hitter oracle for efficient generative inference of large language models. *Advances in Neural Information Processing Systems*, 36, 2024.

Zijia Zhao, Longteng Guo, Tongtian Yue, Sihan Chen, Shuai Shao, Xinxin Zhu, Zehuan Yuan, and Jing Liu. Chatbridge: Bridging modalities with large language model as a language catalyst. 2023. URL https://arxiv.org/abs/2305.16103.

Pengfei Zhu, Chao Pang, Yekun Chai, Lei Li, Shuohuan Wang, Yu Sun, Hao Tian, and Hua Wu. Ernie-music: Text-to-waveform music generation with diffusion models. *arXiv preprint arXiv:2302.04456*, 2023.