An HCI-Centric Survey and Taxonomy of Human-Generative-AI Interactions

JINGYU SHI*, Department of Electrical and Computer Engineering, Purdue University, USA

RAHUL JAIN*, Department of Electrical and Computer Engineering, Purdue University, USA

HYUNGJUN DOH, Department of Electrical and Computer Engineering, Purdue University, USA

RYO SUZUKI, Department of Computer Science, University of Calgary, Canada

KARTHIK RAMANI, Department of Mechanical Engineering, Purdue University, USA

Generative AI (GenAI) has shown remarkable capabilities in generating diverse and realistic content across different formats. HCI literature has investigated how to effectively create collaborations between humans and GenAI systems. However, the current literature lacks a comprehensive framework to better understand human-GenAI interactions, as the holistic aspects of human-centered GenAI systems are rarely analyzed systematically. We present a survey of 291 papers, providing a novel taxonomy and analysis of Human-GenAI Interactions from both human and GenAI perspectives. We highlight challenges and opportunities to guide the design of GenAI systems and interactions toward the future design of human-centered GenAI applications.

CCS Concepts: • **Human-centered computing** → **Interaction design theory, concepts and paradigms**.

Additional Key Words and Phrases: generative AI, human-AI interaction, interactive machine learning

1 INTRODUCTION

Recently, Generative Artificial Intelligence (GenAI) models have gained immense popularity and are being applied in diverse applications such as art [10, 184], design [144, 274], and entertainment [183, 297]. Current popular GenAI models including Large Language Models (LLM) and Large Vision Models (LVM) are widely deployed on platforms or in software for their capabilities to create imagery content (Dalle-2 [259], Stable Diffusion [265]), writing literature [344], and Question Answering (ChatGPT [238]). The adoption of GenAI models has demonstrated significant advantages, e.g. fostering creativity [252], driving innovation [39], enabling personalized content generation [323], and providing valuable assistance in creation endeavors [208]. As a result, these models have become ubiquitous, emphasizing the need for well-crafted and compelling interactions between humans and GenAI.

To take advantage of the generative power of GenAI models, Human-GenAI interaction techniques such as prompt engineering [195, 252], visualization [308], and interactive interfaces [55, 73], have become popular and effective mediums for humans to interact with GenAI systems. These allow users to collaborate [344], be assisted[8], take suggestions [48] or revise recommendations [304] from GenAI systems.

---

*Both authors contributed equally to this research.

However, existing research on Human-GenAI interactions focuses on each individual aspect and domain. The key design considerations, common practices, and future research opportunities are still hidden and scattered across many broad topics embedded in GenAI and its applications. To keep pace with the development of GenAI models and their new out-of-the-box capabilities, we identify a need to systematically analyze the research in this field, particularly from an interaction design perspective, to assist the HCI community to innovate and explore new interaction design techniques for the best utilization of GenAI capabilities. Furthermore, this view of GenAI will also foster new emerging applications to consider key vantage points where our framework will point.

Inspired by the above, we aimed to lay the groundwork for further developments in the field of human-GenAI interactions by systematically synthesizing all the current research and consolidating the existing knowledge and approaches in this domain.
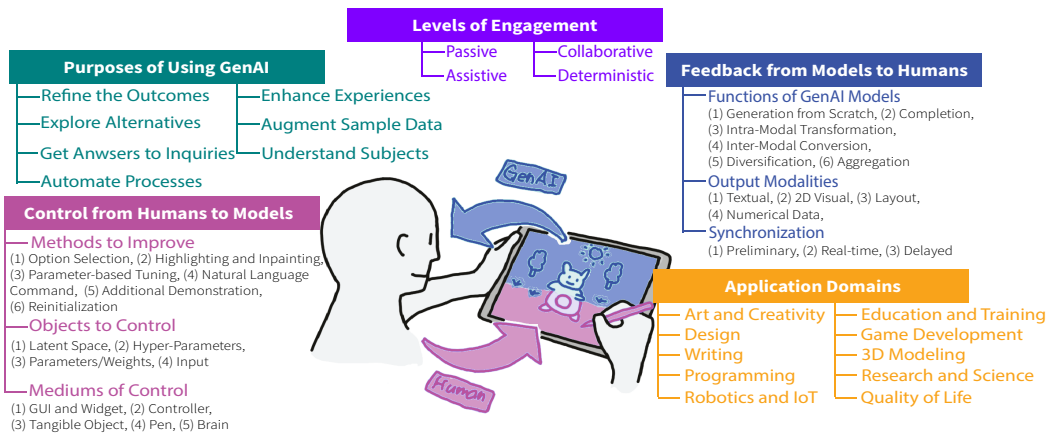

Fig. 1. Visual abstract of our survey and taxonomy of Human-GenAI Interaction. Our taxonomy summarizes five key dimensions, namely, Purposes of Using GenAI, Feedback from Models to Humans, Control from Humans to Models, Levels of Engagement, and Application Domains.

In this paper, we review a corpus of 291 papers for synthesizing the taxonomy of human-GenAI interactions. Specifically, we synthesize the research fields from both the user and GenAI perspectives (briefly shown in Figure 1) into the following dimensions of the design space: 1) Purposes of Using GenAI, 2) Feedback from Models to Users, 3) Control from Users to Models, 4) Levels of Engagement, 5) Application Domains, and 6) Evaluation Strategies.

Our main goal is to present a comprehensive overview of recent developments in and research on AI-model analysis, interaction designs, visualization techniques, and application domains of GenAI-based systems. By compiling the state-of-the-art advancements in these areas, we aim to provide a valuable resource for researchers to understand the current landscape and situate their own work within a broader design space. To achieve the goals above, we summarize a taxonomy from the literature, offering a holistic view that encompasses perspectives from both the GenAI model side (e.g. I/O design, capabilities, and volumes) and the human side (e.g. evaluation strategies and application domains), as well as the interactions between them (e.g. interfaces to control, visualization technique, and feedback design). This taxonomy will enable readers to gain a deeper understanding of the intricacies involved in creating effective and meaningful human-GenAI interaction systems, fostering future evolution and innovation in the design of GenAI technologies. Additionally, we identify open research questions, challenges, and opportunities in the future design

of GenAI systems and interactions. By highlighting these areas of exploration, we aim to guide researchers in their pursuit of addressing crucial issues and uncovering new possibilities, so that the HCI community paces and also identifies vantage points for creating new GenAI, with the rapidly evolving technology of GenAI.

2 BACKGROUND, SCOPES, AND CONTRIBUTIONS

In this section, we cover the developments in GenAI models in prior research as well as open-source platforms and software.
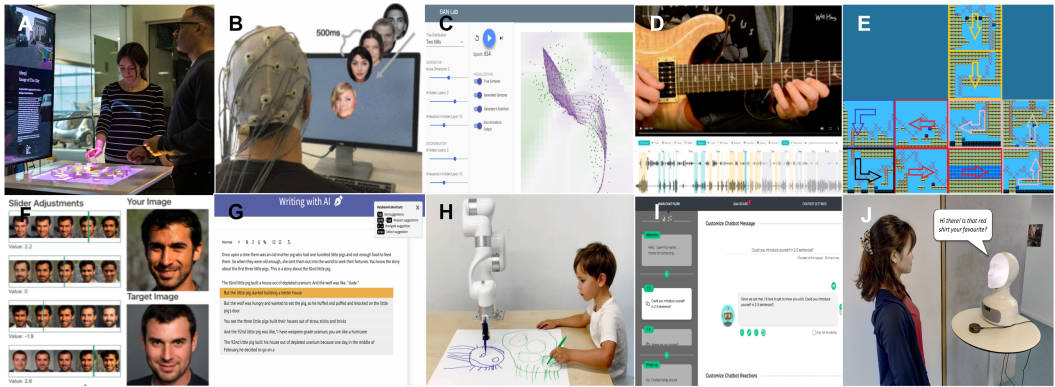

Fig. 2. Examples of GenAI applications located in our survey, covering the topics of research: A) Embodied interactions with GenAI [233], B) Direct Control human to AI [284], C) Human Interpretable [139], D) Gen AI enhancing skill [310], E) Automate Process [38], F) Human controllability [55], G) Natural Language Generation [94], H) Human AI collaboration [299], I) Personalization and Adaptation [107], J) Conversational GenAI [126]

A widely accepted definition [330] of GenAI goes by *the probabilistic models that model the joint distribution* in contrast to discriminate AI models *that model the conditional distribution*. From a high-level understanding, GenAI is defined as "an AI that uses existing media to create new, plausible media" [222]. As a sub-topic in the domain of AI, GenAI has a rich history of development.

In the early stage of its development, the metaphor of GenAI advanced independently in two major domains, namely Natural Language Processing (NLP) and Computer Vision (CV). In NLP, the generation of nature languages by AI was handled by early implementations of Recurrent Neural Networks(RNN) [70] and Long Short-Term Memory (LSTM) networks [97]. Likewise in CV, the concepts of Artificial/Convolutional Neural Networks (ANN [86] and CNN [170]) were applied to models that generate images. Nevertheless, in both fields, the NN-based methods were greatly limited by the hardware conditions back then.

It was not until the early 2010s that the breakthrough in hardware technology enabled the explosion in GenAI research with increased computational power in both NLP and CV fields. Long sentence generation and sequence-to-sequence generation were then achieved by RNNs with much larger sizes and computational power [217, 290]. Similarly in the CV area, models like Generative Adversarial Networks (GAN) [93], Variational Autoencoders (VAE) [159], and their successors [28, 143–145] enabled diverse applications such as style transferring between images [89, 366], generating images based on texts [367], etc.

Recent research has highlighted and merged the two fields, enabling the multi-modal generative power of GenAI. Works like Transformer [305] and Diffusion Model [113, 229, 266, 269] have built

the theoretical foundation for the current stage of GenAI, where large models such as Generative Pretrained Transformer (GPT) and its successors [31, 149, 254, 255], T5 [256], BERT [62], and CLIP [253] enable diverse applications with content of higher quality and multiple modalities.

Empowered by the models, an increasing number of software and interactive platforms are being developed to cater to various fields, such as art, design, education, and algorithm development, democratizing access to the creative potential of GenAI. In this paper, we provide a comprehensive summary of diverse GenAI applications and highlight some notable open-source platforms in Table 15 . These platforms are designed and contributed by researchers, developers, and experts, aiming to make GenAI technology accessible to a broader audience. Table 15 serves as a valuable resource for readers to gain insights into the versatility of GenAI applications and discover open-source platforms that can facilitate their creative pursuits.

2.1 Scope

2.1.1 GenAI vs AI. GenAI, as its name suggests, represents a category of AI that goes beyond traditional models by focusing on generating new data rather than solely analyzing or making predictions. In our research, we place a particular emphasis on GenAI models that excel at generating fresh content. While traditional AI models are designed to perform specific tasks or offer predefined responses based on data patterns and algorithms (i.e. discriminative), GenAI systems possess the unique ability to create novel content (i.e. generative). By enabling users to influence the generated content through inputs like prompts, these interactions become more dynamic and creative.

2.1.2 GenAI Systems. Among extensive existing research and work on the applications of GenAI systems, we have chosen to narrow the scope of our research to focus exclusively on GenAI systems that are developed using deep generative models and specifically designed for user interactions, because of their overwhelming generative power [63] and rapid improvement in recent years.

Notably, our paper does not encompass the usage of GenAI models where no user interaction exists, i.e. research that focuses on only the model performance and architectures rather than interactions or applications. We have also deliberately chosen not to delve into the detailed formulations of GenAI models and their creation process. Similarly, we do not discuss the specific methodologies for creating GenAI models, improving their performance, or training and collecting datasets. While the technical details of creating and deploying GenAI models are undoubtedly essential and relevant in other contexts, our research emphasizes the human perspective such as the utilization and interaction of these systems by users and the impact of GenAI systems on user experiences, creativity, and decision-making.

2.2 Contributions

The GenAI has been explored in various other papers from both sides human and GenAI model [365]. Some work has conducted study [118] and discussion [44] to gain human perspective in using GenAI models. Chen et al. [44] conducted a discussion with researchers and presented a roadmap for future directions from the technical (GenAI models) side aligning with human values and accommodating human intent [37]. Prior work has contributed in survey and review papers in the field of GenAI such as GenAI recent developments [357], their technical perspective [109, 351], content generated [37] and application [96, 365]. Some recent work has also proposed design space [118, 219, 320] and design guidelines [195]. Recently, GenAI and human interactions have been a topic in HCI workshops [21, 221, 222].

Building upon prior work, this paper offers the following significant contributions.

Firstly, it presents a comprehensive **taxonomy of the design space**, considering both the human perspective and the GenAI perspective. This taxonomy provides a detailed and in-depth

view of various dimensions and categories, with a specific focus on human interactions with GenAI systems. By examining the design aspects from these two perspectives, the paper sheds light on the dynamic and creative interactions between users and GenAI systems, providing valuable insights into the user-centric nature of these systems.

Secondly, the paper represents a pioneering effort as the **first comprehensive literature survey** on human-GenAI interaction systems. The literature survey serves as a valuable resource for HCI researchers, offering a well-organized and insightful compilation of the current state of human-GenAI interaction systems. Researchers can draw from this survey to gain a deeper understanding of the design space and the nuances of interactions between users and GenAI systems. Moreover, the survey provides a solid foundation for further research and exploration of novel design possibilities in this rapidly evolving area.

Thirdly, we end our paper with discussions over **directions for future investigations**, helping researchers identify **unexplored opportunities and challenges** in human-GenAI interactions. The discussions and insights are derived from the collections of the papers and the high-level summarization we identified.

3 METHODOLOGY

We aim to identify and collect a large representative set of state-of-the-art GenAI models and GenAI systems using systematic search techniques [114]. We systematically created a relevant corpus using PRISMA [241] guideline: (1) Search strategy to explore; (2) Identification of the publication outlets; (3) Evidence Screening; (4) Eligibility; (5) Inclusion.

3.1 Search Strategy

We performed two search waves to select the relevant keywords for our survey. First search waves identify keywords for advancements in GenAI **models** in the Machine Learning/Deep Learning conferences. This also includes open-source software built on GenAI models such as ChatGPT. This first spectrum of keywords provides models that were further used in developing a methodology to find GenAI systems corpus in the HCI domain. The two search strategies are discussed in detail below:

3.1.1 Methodology to Define GenAI-related Model Keywords. : The GenAI field is broad and continuous works are being published in general machine learning areas and domain-specific areas such as CV and NLP. Therefore, we performed a broad search to identify keywords related to GenAI in order to maximize the inclusion of relevant papers. Focusing on GenAI ***models***, we used different expressions and abbreviations of "Generative Artificial intelligence". The list of keywords used is shown in Table 3. We searched the title, abstract, and keywords to get relevant papers related to GenAI. The search was conducted in the proceedings of prominent machine learning conferences (ICML, ICLR, and NeurIPS), computer vision conferences (CVPR, ICCV, and ECCV), and NLP conferences (ACL, EMNLP, and HLT-NAACL). To ensure the relevance of the research specifically to Deep GenAI models, we limited our search to papers published 10 years prior to our work. During the search process, we carefully examined author keywords, abstracts, and titles to extract additional relevant keywords related to GenAI models. We also added more keywords based on the three authors' knowledge about the recent advancements in GenAI algorithms and software.

3.1.2 Methodology to Define GenAI-Related Systems Keywords. : Then, we focus on finding relevant keywords related to GenAI ***systems that are interacted with*** by the users. As the focus of the papers should involve Human-GenAI interactions, we used combinations of keywords: "Generative

AI models" AND "Humans" OR "Human" OR "Users" OR "User" for our search. The complete list of keywords is shown in Table 4.

3.2 Identification

We meticulously executed a systematic search strategy using relevant keywords from renowned publication platforms, including ACM Digital Library, IEEE Xplore, MDPI, Springer, and Elsevier. Employing the OR operator between keywords ensured a comprehensive exploration of the literature on GenAI models and systems. Additionally, we proactively searched for variations and synonyms of the keywords, encompassing terms such as "GAN," "StyleGAN," "CycleGAN," "Transfer learning," and "ChatBots," and many more to capture diverse facets of GenAI research. To focus on the most pertinent content, we applied filters to restrict the search to the title, abstract, and authors' keywords of the articles. Considering the rapid advancements in GenAI models, we narrowed our search to papers published in the last 10 years because of the advancements in Deep GenAI models. Moreover, we prioritized open-access papers and those accessible via institutional subscriptions, broadening the availability of our research findings. We also included relevant citations in our corpus from the papers we found from our key strategy.

3.3 Screening

Three authors screened the initial corpus of 12076 papers individually based on the title and abstract. The entire corpus was divided into 3 equal parts and each part was screened by one author. The authors read the title and abstract of the papers assigned to them and screened the paper based on the following exclusion and inclusion criteria: 1) We focus on including papers that were only written in English. 2) We included peer-reviewed conference papers such as journals, conferences, and workshop papers. We excluded opinion papers, idea papers, workshop proposals, patents, non-peer-reviewed papers (e.g. arXiv papers), posters, surveys, and literature review papers. 3) We removed false and change-of-context papers that were not focused on Human-GenAI interaction but were present because of keywords. 4) we excluded the papers with no Deep GenAI model papers used. 5) papers were included that focused on human GenAI interactions rather than on GenAI model performance improvement. If the authors were unsure about the paper, a discussion among all the authors was considered for the final assessment. After the screening process, we kept 797 papers.

3.4 Eligibility

To determine the eligibility of papers each author individually read the entire papers. As the primary interest is in human GenAI interactions, we included additional inclusion and exclusion criteria (as shown in Table 1) that were also used when reading the entire paper: 1) we removed papers that were focusing on only technical improvements of the GenAI model. 2) we excluded works without human interactions with the GenAI model or system. 3) we included papers where humans were interacting with the AI system, even if the papers were not focusing on the GenAI system or humans. 4) We included study-based papers that involved interaction with GenAI systems. These criteria along with screening criteria, three authors reviewed the entire corpus individually by going through the entire paper independently and following the selection criteria to exclude out-of-scope papers. We determined inter-rater agreement among the three authors using Fleiss' Kappa Measure during the eligibility of the final corpus. The agreement score of $\kappa = 0.79$ indicates substantial agreement. A final discussion was performed to solve any discrepancies. The overall focus for selection criteria involves papers with only human GenAI system interactions. Then all three authors discussed each paper in the corpus to finalize a total of 291 papers.

Table 1. Exclusion Criteria and Inclusion Criteria authors followed during the Confirmation process of the eligibility of the papers.

	Exclusion Criteria (EC)
EC1 GenAI Model Technical Improvement	Papers that solely focus on improving the performance of GenAI models for the applications
EC2 No Human Interactions	Papers that solely present applications of GenAI without actual users or humans interacting with the applications.
EC3 Opinion	Literature Review, survey, and opinion papers (reserved)
EC4 Change of Context:	Papers with words/terms from different contexts from GenAI
EC5 False Search:	Papers with no desired keywords but found by the search engine
EC6 Idea Paper:	Short papers, proposals, demos, and position papers (reserved)
EC7 Non-AI Papers:	Papers with no GenAI models being used (E.g., generating design using simple if-else conditions).
	Inclusion Criteria (IC)
IC1 Interactions:	Papers where humans are interacting with the AI system, even if they are not focusing on the systems or humans
IC2 Extended Abstracts:	Papers published as extended abstracts
IC3 Study-based Papers	Study-based papers that involve interaction with GenAI systems.

3.5 Inclusion

Our final paper has a total of 291 papers. The collected corpus varies from 2013 to 2024. A larger majority of the papers have been published in ACM (55%)and IEEE (29%). Table 2 shows the filtering of manuscripts at various stages for the final corpus.

Despite our systematic methodology for collecting representative papers of the existing work, we acknowledge that the resulting corpus might not contain every piece of the development of GenAI, due to the mass popularity of GenAI in every other domain such as medicine, manufacturing, and others. Further, there might be some papers that are on the borderline of Exclusion/Inclusion criteria that were not included in the corpus.

Nevertheless, we argue that this corpus is large and comprehensive enough to provide representative subsets of each dimension for the most relevant papers for our taxonomy. To collectively address our limitations, we will make our coding, dataset, and tagging system open-source and available for researchers to iterate, suggest new categories, expand our current taxonomy, and advance.

Table 2. Manuscripts filtered at various stages. Id. for Identification, S. for Screening, E. for Eligibility.

Process	Criteria	ACM	Springer	IEEE	S.D.	MDPI	T&F	Others	Total
Identification	Search term	1279	1048	6832	1334	1164	443	14	12114
Identification	Filtering	1265	1048	6812	1332	1164	441	14	12076
Screening	Keyword and abstract	138	35	394	47	13	46	14	797
Eligibility	Full paper	160	6	84	7	11	9	14	291

3.6 Analysis

As discussed in the former sections, we aim to present a comprehensive taxonomy of human-GenAI interaction in order to help future research and development in designing interactive GenAI systems. Over the collected corpus of literature, we conduct a collaborative analysis to identify the key dimensions of the taxonomy.
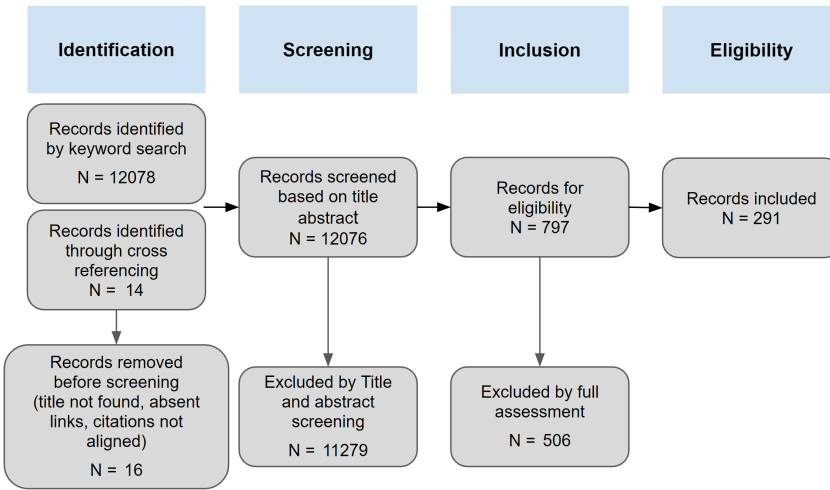
Fig. 3. Flowchart for our paper selection process

The analysis involves multiple stages of review and collaboration among the authors. Initially, each of the authors reads and summarizes a small subset ($N = 25$) of the papers to establish an approximate taxonomy of components and dimensions. These preliminary taxonomies are then discussed among all authors to iteratively refine and enhance, by highlighting the overlapping dimensions, adding or subtracting components and categories as necessary, and eventually agreeing upon one revised taxonomy. Upon revision, three authors individually read the entire list of papers thoroughly to assign them to their respective categories and dimensions. During this phase, workshop proposals, surveys, and literature reviews are not included in the final categorization but are used as supplementary references to guide our analysis and concretize the design components and categories. To ensure consistency and resolve any conflicts, the three authors subsequently engaged in discussions to finalize the tagging of the papers, by iteratively opening up new discussions upon any author identifying a paper that does not properly fit into the existing taxonomy. At the end of each iteration, authors decide if a new dimension is needed for the taxonomy or agree on the position of the paper in the existing taxonomy. This collaborative approach helps ensure the accuracy and reliability of the categorization process.

Eventually, our analysis results in a comprehensive taxonomy of dimensions from an HCI design perspective. The HCI design logic behind this taxonomy suggests that developers and researchers in this field practice User-Centered Design [3] by

(1) orienting the system through the needs of the end-users (section 4)
(2) considering the capabilities and functions of the GenAI models (section 5)
(3) exploring the possible actions of the users, the corresponding results of the actions, and the mapping between the actions and the results(section 6)
(4) enabling different levels of engagement in the interaction for different tasks (section 7)
(5) deciding the application domain the system can be generalized into (section 8),
(6) and evaluating the system via designated metrics (section 9).

In the following sections, we discuss various components in the dimensions and the motivations for considering them when designing a system. Later on, we present the future opportunities and challenges in this domain summarized from our literature review. In the appendix, we included tables containing all the papers' citations and counts that fall into respective categories and dimensions.

4 PURPOSES OF USING GENAI

The initial step of designing an interactive GenAI system is to determine the end goals of (i.e. purposes of using) the system from the user's perspective. Characterizing such purposes sets the fundamental notes on the design of the systems and answers the essential questions on what types of tasks the system is to fulfill and what the users of the system expect from it. To this end, in the very first section, we categorize the purposes of users of GenAI applications. On a high level, we identify the purposes falling into the following categories: 1) Refine the Outcome, 2) Explore Alternatives, 3) Get Answers for Inquiries, 4) Understand a Subject, 5) Automate Processes, 6) Enhance Experiences, and 7) Augment Sample Data.
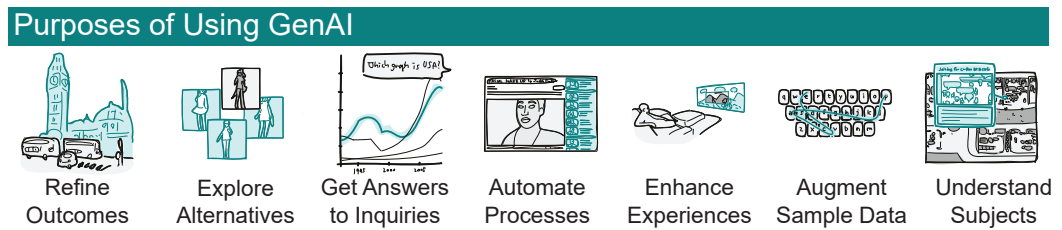

Fig. 4. Purposes of Using GenAI depict the users' intention of the interactions and the high-level capabilities of the applications, consisting of Refine Outcomes [19], Explore Alternatives [308], Get Answers to Inquiries [151], Automate Processes [296], Enhance Experiences [278], Augment Sample Data [48], and Understand [244]

Purpose-1 Refine the Outcome. Users possess certain expectations of an instance or outcome to be created. With a specific objective, users utilize GenAI applications to generate instances to meet their qualitative or quantitative expectations. Qualitative expectations of the users encompass subjective properties of the instances, such as style of a fashion design [323], melody in a piece of music [203, 287], plots in a story [50], layout in a web application [133, 220], content [72, 252] or subjects [39, 58] in an image, etc. Quantitative expectations depict the objective metrics that the generated instances are to satisfy, such as parametric designs of a 3D model [163], efficiency of codes [193], precise layout of cameras in a VR space [342], etc.

Purpose-2 Explore Alternatives. Users can actively utilize GenAI to generate multiple instances of an idea, building from the abstraction of knowledge learned by the GenAI models. Users then explore the generated alternatives for ideation, choosing among, or learning the patterns. For example, GAN-based applications like GANravel [73] and GANCollage [308] enable the users to generate multiple images similar to the input and explore the gallery of images to decide the best design of the images. GenAI can also passively assist users in their ideation process. For example, CatAlyst [14] motivates the users to continue their unfinished presentations by completing part of their work to provide new ideas.

Purpose-3 Get Answers for Inquiries. When faced with a specific challenge or question, users can leverage GenAI to brainstorm potential solutions or avenues of inquiry. The inquiries put forward can be of single or multiple modalities and from diverse domains. To design GenAI applications fulfilling this purpose requires considerations over both the delivery of the answers and the capability of providing correct or rational answers. For instance, users can ask GenAI systems a specific question and have it generate programming codes to solve this question and iterate with the system to optimize the codes [147]. Moreover, GenAI can directly generate the answers in natural language to the problem input by users [151].

Purpose-4 Understand Subjects*.* Different from getting answers to specific inquiries, users can utilize the generative powers of models to obtain their understanding of various subjects in a comprehensive manner, without users explicitly deciding the form or content generated to contribute to the understanding. The understanding is enhanced by GenAI through diverse methodologies such as providing insights, generating examples, or offering new perspectives. Such understanding can be of the process of the model itself (e.g., GANslider [55] utilizes filmstrips of screenshots to illustrate the process of GAN-based transferring of images.), the knowledge of a concept in specific domains (e.g., Liu et al. [195] demonstrates how users of text-to-image GenAI can effectively prompt by observing vast amount of text-image-pairs.), the nature or phenomena conveyed by data (e.g., Vis Ex Machina [349] generates graphs and charts based on input data from the users in order to help understand the data.).

Purpose-5 Automate Processes*.* Users can utilize GenAI as an automation tool in a broad spectrum of applications. GenAI is specifically strong in automating tasks that involve the generation of content. GenAI can be applied to generating control sequences of robots [30, 119, 188, 260] in various scenarios based on users' textual or voice commands. While traditional automation focuses on executing repetitive tasks based on specific instructions, GenAI can also introduce creativity, adaptability, and decision-making into automation processes. For example, Liventsev et al. [201] propose fully autonomous programming with LLMs, where the models are able to rationalize and determine the best practice of coding.

Purpose-6 Enhance Experiences*.* GenAI, given its ability to generate content and adapt to user input, can significantly enhance user experiences across various platforms and applications. In general, GenAI is capable of enhancing the experience by improving the quality of the generated content based on diverse metrics. For instance, GenAI can extend the visual experience of the users [158], make language in articles user-friendly [285], or modify the user input for more efficient communication [304, 329] A particular key aspect in GenAI's enhancement of experience is personalization, where GenAI adapts its output based on users' profiles, preferences, or states. For example, VocabEncounter [15] adapts to the contexts of the users to provide personalized experiences of learning foreign vocabulary. AdaptiFont [137] generates adaptive fonts according to users' reading speed to maximize the reading experience.

Purpose-7 Augment Sample Data*.* GenAI has become a powerful tool for data augmentation, a process used to increase the amount and diversity of data. AI-generated data can be used as training data to build a new AI model. For example, Word-Gesture-GAN [48] utilizes GAN to generate synthetic gesture data for training keyboard gesture recognition models. Deepwriting [8] uses GAN to generate handwriting data to train style-transfer models. Enabled by the large corpus of knowledge learned by LLMs and LVMs, research has also investigated the possibility of deploying AI-generated data of various modalities into other research domains. For instance, Park et al. [245] and Hamalainein et al. [106] experimented with AI-generated for research in social computing and HCI respectively.

5 FEEDBACK FROM MODELS TO USERS

To better understand the essence of human-GenAI interaction, we investigate the field from two major aspects. The first aspect is the feedback from Models to Users. In this aspect, we majorly consider the capabilities of the models, the range of modalities of the models, and the methodologies of delivering the output of the models. These considerations over the models are essential to designing a GenAI system in order to fulfill the purposes aforementioned in section 4. Prior work [338] in general human-AI interaction has also addressed that the key design challenge

stems from understanding AI's capabilities, especially from a user-centered perspective. To this end, we identified three dimensions to depict the current landscape of feedback techniques of GenAI systems, namely, 1) Output Modalities, 2) Functions of the Models, and 3) Output Synchronization.
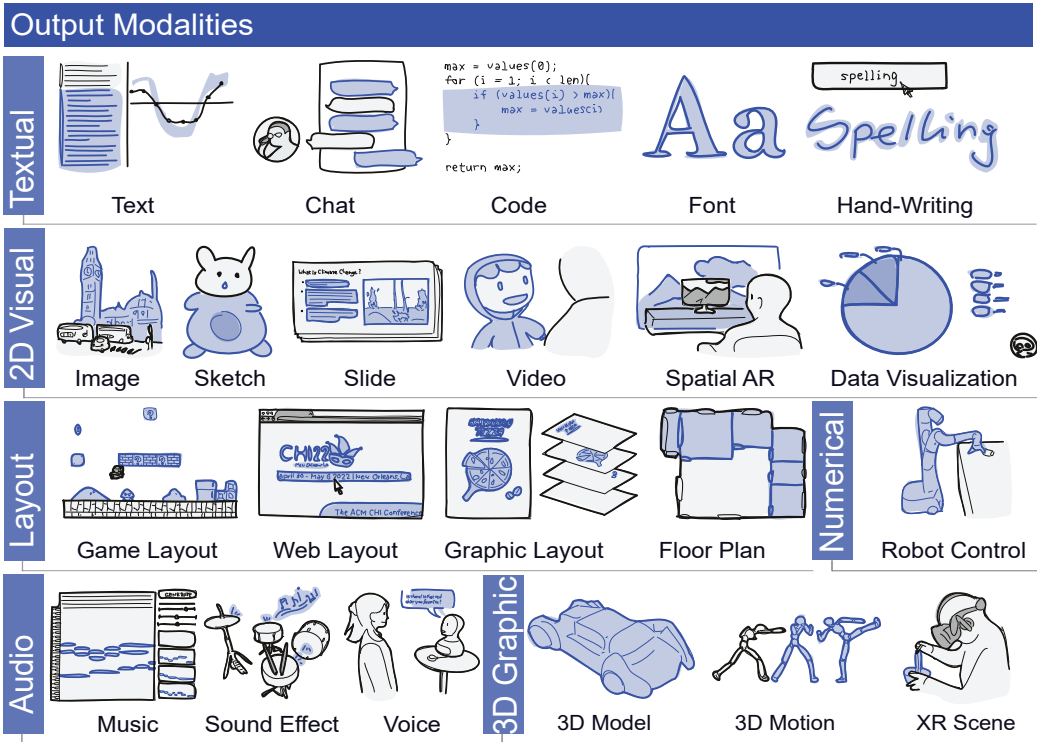

Fig. 5. Output Modalities consists of four categories, namely textual (text [50], chat [134], code [147], font [137], and hand-writing [8]), 2D visual (image [19], sketch [75], slide [14], video [342], spatial AR [158], and visualizations of data [107]), layout (game layout [306]), web layout [156], graphic layout [102], and floor plan [110]), numerical data (robot control sequence [119]), audio (music [287], sound effect [40], and voice [126]), and 3D graphics (3D model [198], 3D motion [331], and XR scene [226])

Dimension-1 Output Modalities. The output modality of a GenAI model refers to the type or form of data that the model produces. Generative models can produce a variety of outputs, and the modality is determined by the type of data they are trained on and designed to generate. In our research, output modality determines the modality of the feedback presented to the users (i.e. they are identical), because we have not identified a case where the output of the system is not presented to the users.

—Textual Textual output encompasses natural language in texts, programming code [128, 193, 201], the handwriting of texts [8], and fonts [137]. Specifically, natural language in texts can be chats [107, 120, 346], descriptions of a problem [147], or pieces of literature [50, 236, 285].

2D Visual GenAI models that produce 2D visual outputs are typically trained on large datasets of images to learn the underlying patterns and can create novel images based on their training. 2D visual outputs consist of images [39, 195, 200, 267, 272], sketches [45, 356], videos [196, 296], 2D visualization of data [173, 349], and spatial AR [157, 158].

3D Graphic 3D graphic outputs consist of 3D models, 3D motion of various objects, and XR scenes. For example, Koyama et al. [163] utilize generated 3D models to provide suggestions during 3D design. Yoo et al. [342] generate VR camera layouts by referring to a clip of a film.

Audio Audio output consists of music [81, 203, 287], sound effects [236], or natural language voice synthesis [126].

Layout GenAI is capable of generating layout information, widely deployed in designing game layouts [220], web layouts [311], graphic layouts [102, 133], and more domain-specific layout designs (e.g. game layout [38, 275, 306]).

Numerical Data All modalities of input can be fundamentally regarded as numerical data in the computer science and engineering realm. In addition to the aforementioned modalities that contain high-level information that can be directly perceived by humans, we identify the numerical data otherwise conveying information and being used as inputs to GenAI e.g., gestural data [48], control sequence to a robot [30, 260], and hierarchical representations of concepts [173].
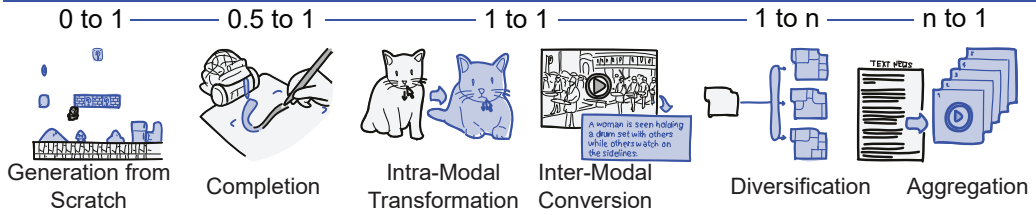

Fig. 6. Functions of GenAI Models describe the capabilities of GenAI models, consisting of Generation from Scratch [306], Completion [188], Intra-Modal Transformation [71], Inter-Modal Conversion [316], Diversification [117], and Aggregation [236].

Dimension-2 Functions of the Models. As was previously elucidated, the core capability of GenAI is to generate new data samples that are similar in distribution and characteristics to the training data. Based on this capability, a range of functions of GenAI models are developed. We categorized the most common six functions in the literature, namely: 1) Generation from Scratch, 2) Completion, 3) Intra-Modal Transformation, 4) Inter-Modal Conversion, 5) Diversification, and 6) Aggregation. The categorization of the functions of GenAI models aims to bring a clear vision of designing a GenAI system to fulfill the users' needs, utilizing a specific or a combination of functions.

Generation from Scratch GenAI applications are capable of producing entirely new content without specific input. This could be through utilizing pre-trained patterns, internal algorithms, or some combination of initial states or conditions within the model itself. The initial generated content is not directly guided by a user's input, and the model operates more autonomously. With this autonomy, this method of input initialization can benefit (1) GenAI systems with non-expert [137, 203, 287] users who do not possess knowledge of proper input, (2) systems that passively assist the users [15, 137], or (3) systems that help the users with ideation within a specific domain through vast collections of examples [73, 161, 240, 306]

Completion In some scenarios, GenAI is required to finish an incomplete product from the user. For example, GenAI can generate auto-completion or suggestions for an ongoing writing task to compose a piece of literature or a story [58, 125] with designated plots or opinions to inspire or lead the writers. Moreover, based on what users have input, generated content can be as good

as the user input in terms of quality in some cases, [75] or provide a different perspective on the subject [14].

Intra-Modal Transformation Intra-modal transformation refers to the function of GenAI to change within the same input modality to generate a different output in the same modality. Systems that leverage intra-modal transformation typically include modifying the details in the content to meet the users' preferences or experiences (e.g. Strengers et al. [285] propose an LLM-based method to modify the article for friendliness to people for minority and De et al. [58] enable human portraits editing with brain signals.). Such transformation usually results in changes in styles [102, 252], content [252, 344], or quality [158, 163, 329] of the output.

Inter-Modal Conversion Inter-modal conversion refers to the function of GenAI models to convert between different input and output modalities. This function to convert abstract knowledge or representations among diverse modalities has fostered a promising quantity of possibilities for GenAI applications. This is because human knowledge and information can now be instantiated to the best modality to either 1) be conveyed efficiently or 2) fit the platforms of the applications. For example, Cheng et al. [45] enable image editing via texts to convert textual descriptions of a design into a visual representation of the design. Similarly, Yoo et al. [342] utilize GenAI to generate VR camera layouts given a reference video, obtaining a unique output for VR applications. Moreover, this function of inter-modal conversion has lessened the barriers of expertise requirements in many domains for novices, particularly thanks to its capability to convert ideas and information from intuitive modality, e.g. natural language and sketches, to exclusive modalities, such as programming language, artistic work, or domain-specific designs. For example, text-to-code applications allow conversion from simple descriptions of tasks in natural language to codes to handle the tasks [128, 147, 193, 201]. This benefit is also manifested in text-to-image and sketch-to-image applications, where users with no artistic skills can instantiate their intuition or idea, and eventually compose an artistic painting [39].

Diversification GenAI possesses the function of diversification, by generating multiple diverse outputs from a single input. The outputs can be of the same modality (1 to n intra-modal transformation). In this case, GenAI is capable of generating instances with variations in details. For example, generating images of the same content but with different viewpoints or features [73, 308, 352], generating longer music given a short clip of melody [203, 287], generating textual content such as NPC quests in games [16] or (fake) news [362], or designs for different game layouts [38, 275, 306]. Moreover, the outputs can be of different modalities (1 to n inter-modal conversion). In this case, GenAI is converting the input inter-modally to multiple outputs. For example, Jing et al. [133] enable the generation of diverse layout designs from a scenario constraint for mobile shopping applications.

Aggregation Finally, GenAI is capable of taking multiple inputs and synthesizing them into a single concise output, which we refer to as aggregation. GenAI can aggregate inputs of different modalities of inputs to an output of specific modalities. For instance, PopBlends [314] blends the concepts from texts and images into a new image, to generate the best representations of an idea. Huber et al. [120] make possible the aggregation from texts and images to emotional dialogues. GenAI can also aggregate inputs to outputs of uniform modalities, focusing on refining the information within. For example, StyleMe [323] enables users to merge the outlines and styles from two fashion designs into a new one. AngleKindling allows journalists to take different angles in writing a journal by summarizing the ideas from the text [246].

Dimension-3 Output Synchronization. GenAI systems also utilize different output synchronization strategies. The synchronization strategies of the GenAI systems investigated follow certain objectives. Understanding the logic of selecting the synchronization strategy is therefore critical to

the system design. We identify three strategies based on the output timing with respect to the user interaction timing.

Preliminary This category describes a GenAI system with output prior to the user interaction. The preliminary output strategy is usually utilized in the GenAI systems where users absorb the AI-generated content [15] passively or as is [134, 310]. Another scenario with preliminary output is the human-GenAI collaboration tasks where GenAI takes the first move to inspire [173] or motivate [14] the human.

Real-time Real-time output is generated concurrently with human interaction with the GenAI systems. This strategy benefits the systems with the requirement of immediate responses such as in writing suggestions [23, 125, 304] and auto completion [175]. Moreover, real-time output is preferred when the systems consist of interaction modalities that tweak the direction, attributes, or details of the generated content. In these systems, users expect real-time feedback generated when they are interacting. For example, in GAN-based image generation applications [55, 73, 331], when a user is dragging a slider controlling the direction of the GAN models, the visualization of the generated images is expected to be dynamic and aligned with the slider movement. This advantage of real-time feedback can be identified through other collaborative applications such as webtoon sketch creation [161], co-writing [23], and programming assistance [79, 249].

Delayed A delayed output is generated after an explicit mark of the end of the users' interaction, e.g., hitting enter when chatting with a chatbot [126, 224] or clicking on a button to input a set of parameters [99]. Delay can be intentionally designed for users to either modify their input or confirm the parameters. This strategy is common in most human-GenAI interactive applications that require descriptions of human expectations of the output, such as fashion design containing multiple layers [45], artistic image generation considering multiple attributes [161], style merging requiring multiple inputs [323], etc. When there are multiple elements to be considered by the humans in the loop, the delayed output prioritizes users' decision-making on the final output. Some interaction techniques require delayed output by nature. For example, interaction with a chatbot requires input and output on a conversational basis which goes one by one. Nevertheless, the computation cost is a major reason for some applications resorting to delayed output. Although from a design perspective, real-time output is preferred for the reasons aforementioned, subject to the model size and constraints on the computational power, most image-based GenAI systems resort to delayed output for consistent user interactions.

6 CONTROL FROM USERS TO MODELS

The second aspect of the interaction comes from the human side. We seek the answers to the questions of what and how users can control so that the models arrive at an expected output. These answers serve as critical human factors in designing the interactions in a GenAI system. To this end, this section delves into the common ways by which humans can provide feedback to the GenAI system, as well as the categorization of feedback to be provided. Broadly, we categorize into three categories: *How* users take actions to navigate or adjust GenAI, *what* the objects in GenAI systems are controlled, and the *mediums* to provide feedback.

Dimension-1 Methods to Improve the Output. In this subsection, we aim to reveal the answers to the essential question – how do users get the expected generated content from GenAI? Looking at the question from an HCI perspective, we summarize the high-level interaction design of the GenAI systems in the literature, to identify the methodologies of user interaction to prompt or improve the output towards their expectations. The design of the interactions follows specific design goals, which are crucial considerations for the HCI design of a GenAI system.
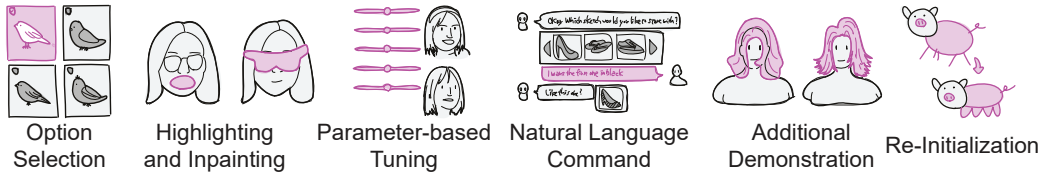
Fig. 7. Methods to Improve the Output from the user's perspective include Option Selection [352], Highlighting and In-painting [73], Parameter-based Tuning [55], Natural Language Command [45], Additional Demonstration [276], and Re-Initialization [75].

Options Selection Users can select their preferred output from a range of options generated by the AI, allowing them to choose the result that best aligns with their needs [203, 234, 308]. Output selection can also serve as a feedback mechanism to train or fine-tune GenAI to get better results in the future. Additionally, users can also select one of the outputs from GenAI model intermediate layers guiding the direction to the final desired output [72]. This allows the user to iterate or build on the intermediate output refining them further to achieve the final result.

Highlighting and Inpainting It allows users to point out specific areas that need modification or replacement in the input such as image [73], text [53, 82], or document [208]. Users can highlight or color paint emphasizing particular details, areas, or objects to either add [19],erase [19, 82], modify [72, 161] or keep [171] specific regions of the input in generating the final output.

Parameter based Tuning GenAI system provides users with a unique ability to access and manipulate the intermediate layers to indirectly modify (compared to composing the content artificially) the output [55, 163, 275, 287, 301]. It provides users with a granular level of control over the output using slider [287] or numerical input [301] to semantically change a generation of output. Parameter tuning is helpful in image input-output target matching [55, 267], controlling the randomness of the generated output such as LLMs [177], changing the style of output [257] e.g. graphic design [301], and editing the content but preserving style [8, 257].

Natural language Commands Natural language guidance from humans either text or voice allows them to guide the output from the GenAI system using commands or instructions [45, 119, 260]. Users can provide commands sequentially to steer the output generation [200]. These commands are commonly used in Large Language models [105], Chatbots [107], and visual design assistant [45].

Additional Demonstration Users can further provide additional information to the GenAI system by drawing sketches [99, 361], outline [276], copy & paste [47], handwriting [7], images [252] and keywords [197] to narrow the scope of the generated output. Users can provide specific preferences [361], additional context [323, 358], and set constraints [50] to generate a more relevant and accurate output.

Re-intiliazation Users can re-initialize the generation process, with new or some adjustments [75, 188, 299] in the input. This iterative and adaptable approach allows users to fine-tune content generation effectively [237]. Reinitializing also allows for experimentation to see how different approaches or inputs affect the output [190].

Dimension-2 Objects to Control. In this subsection, we elaborate on the objects inside the GenAI models that are actually controlled by the users during interaction. We broadly identify four categories, namely, the latent space, the hyper-parameters, the weights/parameters, and the input, which are commonly but not systematically identified objects controlled by the users of an AI system by prior works on general human-AI interaction [11, 68, 112, 275]. As a sub-set of general AI

systems, GenAI inherits these identical properties in terms of interactive components. To provide concrete and novel insights based on prior work by Morris et al. [219] that touches upon how GenAI models reach *evolution* of themselves (i.e. how GenAI models improve the output through interactions), we further investigate the distinctions between controlling GenAI and controlling general AI in the following discussion.

Latent Space Latent spaces are high-dimension representations of the input given by the users. Modifying these high-dimension states allows users a semantically meaningful way to control the style of the output. For example, GANravel [73] allows users to edit face features such as adding glasses to their eyes or making the person smile keeping the rest of the face the same. Such controls of latent space representation in GenAI systems give users the freedom to visualize the direction of output [139], explore for diverse generated outputs [57, 352], and transform [220] the latent space to get the desired output. In most cases, users may not directly manipulate the latent space so they are mapped to UI elements such as clicking a button [73] or moving sliders [55, 161]. In addition, latent space allows users to potentially influence the types [133, 344], quantities [352], and levels of variability [73] present in the system's outputs.

Parameters/Weights Parameters/Weights refer to the weights and coefficients that the algorithm extracts (learns) from the data during the training process. The approach to changing the parameters of a model is addressed as fine-tuning. Fine-tuning the GenAI system involves re-training the GenAI model either with few or zero-shot learning [31, 292]. It allows the system to become adaptive and personalized for each individual user [284]. Fine-tuning GenAI systems makes them personalized to the user by improving task-specific capabilities and domain-specific knowledge of the GenAI system, consequently enabling guided generation of content, in terms of a range of elements in the content, e.g. the subjects or the style of the generated content.

Hyper-Parameters Hyper-parameters refer to the type of parameters that are second-level tuning parameters that decide the training and inferencing process other than the first-level parameters or model weights [78, 250]. Specific designs to interact with these GenAI model hyper-parameters allow users to control the creativity(randomness) in the outputs generated from the GenAI systems [177]. Some GenAI system includes *Temperature* parameter [203, 304], frequency penalty [171] and random seed during the development of GenAI system [177] to control the variability in the generated output. For example, Louie et al. [287] use a temperature parameter to generate conventional or surprising music. Hyper-parameters can be changed by the end-user using tools that allow changing values using sliders [203] or text editor [177].

Input Input control provides the unique ability for the user to interact with GenAI without changing the parameters or the hyper-parameters [346]. The quality [195] and relevance [173] of the input prompt influence the output generated from GenAI systems. Users can provide input prompts to the model to get the desired generated output [323]. Directly giving input prompts from the user to the model often generates out-of-context output. To eliminate this users can additionally provide a few examples of the input-output to guide the model in generating output in a way the user wants [127, 339]. For example, Jiang et al. [128] used LLMs to support software development. Users can also develop an automatic method of designing such input prompts instead of manually specifying [153] to get more relevant output. Some of the examples are by suggesting prompts [173], constructing prompt templates [190], combining multiple prompts primitive [127], reformulating prompts suiting GenAI system [208], and transforming prompt to different input modality [50].

Dimension-3 Mediums of Control. The interfaces through which users interact with the GenAI models are essential in the HCI design of GenAI systems. Under a framework of a generic taxonomy of human-computer input mediums, we characterize the interface design considerations for GenAI systems.
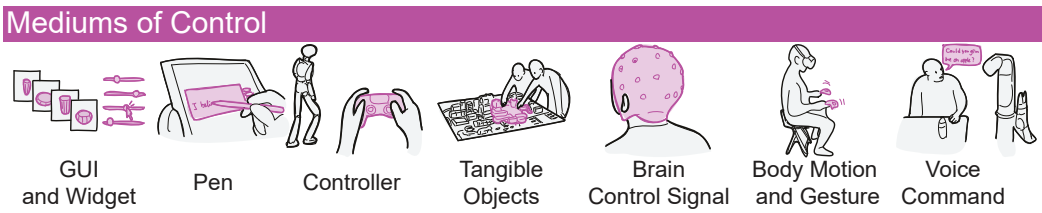
Fig. 8. Mediums of Controlling GenAI models include GUI and Widget [163], Pen [8], Controller [331], Brain Control Signal [142], Tangible Objects [233], Body Motion and Gesture [40], and Audio Command [6].

GUI and Widget UI and widgets are the most common design elements in the GenAI system which allow users to understand the displayed information and interact with them, because of the coverage over the most common modalities of data of GenAI systems. GUIs provide straightforward visualization and mediums of input for textual and imagery data. Similarly, widgets provide also straightforward adjustments to objects being controlled. Buttons [54], sliders [55], mouse clicks and drags [71], and tool pallets [352] are often used to provide input or any modifications to GenAI system. Image editors [72] and text editors [328] are also used to display outputs and provide or edit the inputs. Interfaces are also used in displaying the 3D content [198] Information panels and menus are used to display information that guides users to use the GenAI system [287]. Canvas is also used to provide multiple [72, 73, 308] or different modalities [252] to the user for viewing and selection of the input-output.

Controller The controller is a user interface example that allows the user to provide input and can be used to control the GenAI system. This medium is specialized for the generation of control sequences for robots, game control, or human motion. For instance, Xu et al. [331] used a game controller to control human motion generation.

Tangible Object The user can control GenAI systems by moving objects in the real world [233]. The tangible object includes interactions with physical objects such as static objects[278], dynamic objects[226], and remote objects [67]. Users' interactions with the tangible objects in the 3D space, virtual or real, are usually mapped as the input to the generation of 3D content, with consistency in I/O spatiality.

Pen For imagery or visual-text-based GenAI systems, having the users draw their ideation or instance is a popular way of controlling the models. Users can use a physical pen or pencil to draw [188, 299], write text [7, 8], or make sketches [161, 358] on screens [110] and paper [299].

Brian Control Signal Brian signals are another medium that helps the user control the GenAI system using brain signals. Some works record these signals using electroencephalography(EEG) [56, 142, 284]. Brain responses are directly connected to the internal parameters of GenAI models such as latent space [58] for providing implicit feedback. For example, Spape et al. [284] used a brain interface for generating personalized attractive images.

Body Motion and Gesture Gesture interaction involves the utilization of physical gestures and movements as a means of engaging with the GenAI system, commonly carrying spatial-temporal information that guides the generation of content. Gesture movement can be captured by interactive surfaces such as mobiles and tablets [48, 49]. Also, body movements are useful in interacting with GenAI in immersive environments [40]. Face tracking, facial emotion, and expressions are also used in designing natural interactions with GenAI [135], which usually carries sentimental or semantical information that guides the generation of content.

Audio Audio includes both human voice[6, 126] and music [81]. Human voice commands offer a dynamic approach to manipulating and directing the outcomes of GenAI models. Through vocal

prompts, users can effectively steer the generated output [193], leveraging their spoken instructions to guide the GenAI output. This innovative interaction method harnesses the potential of natural language and empowers users to shape the GenAI output in a more personalized [134] and intuitive manner [193]. On the other hand, music or sound effects as the medium of control usually serve as the reference of styles for generated content.

7 LEVELS OF ENGAGEMENT

With the design human-GenAI interaction investigated from both the human side and the GenAI model side, we then examine the high-level characterization of the design of the interaction as an entity. In this section, we follow the most commonly used definition of engagement as "*the process by which interactors start, maintain, and end their perceived connections to each other during an interaction.*" [235, 281]. We identify four levels of engagement in human-GenAI interaction, namely, *Passive Engagement*, *Deterministic Engagement*, *Collaborative Engagement*, and *Assistive Engagement*, with an increasing connection between the generated output and the user interaction within the engagement. Put simply, with a design of passive engagement, users do not substantially contribute to the generation of the content but the models generate the actual output. On the other hand, in assistive engagement, the models do not substantially contribute but only help in ideation for users to create the content. A comprehensive metaphor could be writing a story, where the ideation is to come up with the flow and plot of the story and substantial contribution means writing down the story in full-text in any form.

The design of engaging patterns of a GenAI system critically determines the user groups for the particular part of the output users can contribute and the expense of mental load they can afford [68]. It also determines the application scenarios because the level of engagement limits the choices of interaction mediums. Finally, the originality or creativity of the content created is also determined by the level of engagement, as pointed out by many prior works [216, 273, 295] that the process of the work assigns the credit for generated work by a human-AI team.

Level-1 Passive Engagement: Passive engagement depicts the systems with which users receive information or content generated by the AI without direct interaction. In this case, users do not create or contribute to the generated content directly but their profiles or preferences are passively taken by the models to direct the output. Example system designs with passive engaging interactions fulfill the tasks of immersive news writing [236] and immersive vision system [157, 158], where the users are passively engaged with GenAI and its product without explicit interactions to guide the output.

With only passive input from the users, passively engaging GenAI systems are expected to generate satisfactory products through an end-to-end pipeline. The design for possible passive input from the users is also tricky and thus challenging due to its implicity and constraints.

Level-2 Deterministic Engagement: As its name conveys, in a GenAI system where the engagement is deterministic, the outcome is determined by the AI's inherent logic, instructions, or a predetermined set of rules, rather than being shaped by the users' interactions. Similar to passively engaging systems, deterministic GenAI systems usually consider users' profiles and preferences as part of the input and directly generate content to meet the users' requirements, resulting in limited contribution from the users to the final result. Yet Deterministic Engagement differs from Passive Engagement with active interactions from the users. User interactions in this type of engagement are usually just instructions to stop and start the generation. Examples can be a foreign language dictionary for the users to learn but in AI-decided contexts [15], AI-generated hierarchical tutorials for the users to follow [296], or an adaptive font generator that evaluates the users' performance and autonomously generates the best font for reading [137].

An end-to-end pipeline for generation is also expected for deterministic GenAI systems. In this setup, users do not need to possess knowledge of the domain to contribute to the generation but are supposed to simply notify the systems when the content is needed by them.

Level-3 Assistive Engagement: Assistive engagement allows the GenAI system to generate content to assist the users in the creation process in the form of suggestions, summarization, or ideation that do not substantially contribute to the final product of the interactions but rather conceptually or abstractly contribute to the creation process. An example of an assistive system can be an auto-completion assistant in writing [125], a contextual provider of suggestions [304], or an online debugger for an ongoing programming [249].

Assistive engagement requires the systems to possess the capabilities to abstractly understand users' needs and produce high-level assistance with the tasks. In the possible application scenarios, users of such systems are expected to complete the generation of the final output with the generated high-level assistance from GenAI models.

Level-4 Collaborative Engagement: Collaborative engagement is the most common design in the current deployment of GenAI systems. In these systems, GenAI and users work collaboratively on a task, with both sides contributing to the final product. One major method of these systems is to collaborate through interactive two-way conversation, exchanging information, and user iterating based on the responses. This method is widely deployed among the systems based on large language models [33, 127, 218, 337, 344], where conversation in natural language dialogue is possible, and some GAN-based applications, where GenAI provides hints or visualizations on the generation direction in response to user queries [55, 71, 73, 200, 220, 315]. Another method is cooperation, in which the users and the GenAI share the same goal and substantially contribute to the final results in the same format or modality. Examples of this method can be jointly creating slides for a presentation [14], finishing a sketch by GenAI adding details [75], and composing a piece of melody by both users and generated music [203, 287].

Collaboratively engaging GenAI systems require the users to possess the skills and abilities to create the entity or part of content themselves, with GenAI models sharing the same tasks with them. It also requires the GenAI models to possess a similar set of capabilities as users in the designated application scenarios.

8 APPLICATION DOMAINS

Through our exploration, we identified a range of diverse application domains of human-GenAI systems. Figure 9 summarizes the categories of domains and lists the related papers correspondingly. We classified existing works into the following high-level application domains: (1) *Art and Creativity*, (2) *Science and Research*, (3) *Writing*, (4) *Programming*, (5) *Robotics/IoT*, (6) *Education and Training*, (7) *Game Development* (8) *3D Modeling*, (9) *Design*, and (10) *Quality of Life*. A detailed list of references in each of the domains above can be found in Table 13.

Art and Creativity is the domain where most applications emerge. The generative power of GenAI has changed the game in the art industry, covering lots of aspects of artistic creation across the disciplines of visual arts, music, literature, and filming. In general, GenAI can contribute to the processes of ideation, variation, and polishing the artwork. Such contribution will be further refined by the improvement of the interaction design of human-GenAI. GenAI also manifests promising potential in the realm of design. In *Art* and *Design*, where the visual components are generated by GenAI and then evaluated by human designers, we anticipate further research on the interaction designs in this context from both micro (e.g. efficiency of certain methods for visualizing designs) and macro perspective (e.g. conceptual processes in designs that can be enhanced by GenAI and how). In our discovery, there are more domains that are less investigated or not investigated best
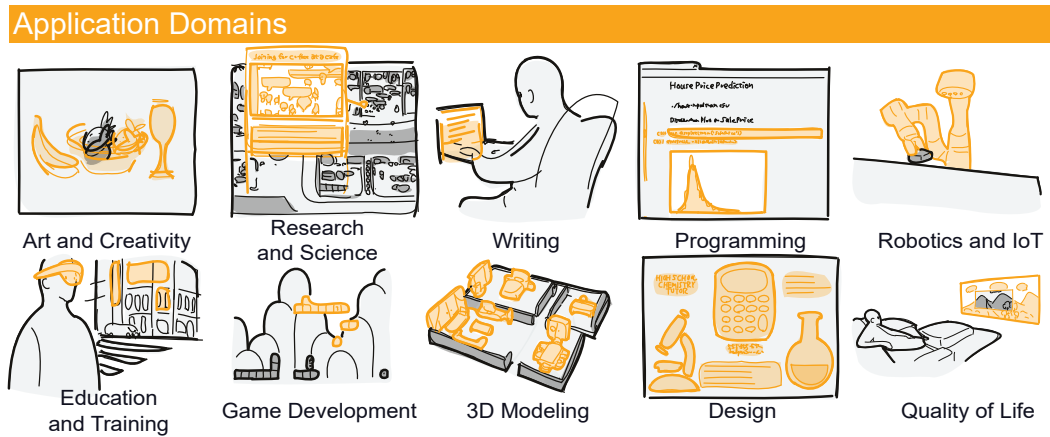
Fig. 9. Through our discovery of the literature, the application domains of GenAI include Art and Creativity [237], Research and Science [244], Writing [15], Programming [309], Robotics and IoT [6], Education and Training [15], 3D Modeling [83], Design [234], and Quality of Life [278].

to our knowledge, e.g. *Education and Learning*. We foresee further exploration and exploitation in these domains based on the insights into specific patterns and methodologies depicted by our taxonomy.

9   EVALUATION METHODOLOGIES

In this section, we report our categorization of evaluation strategies for GenAI systems. The main categories we identified following the classification by Suzuki et al [291]: (1) technical evaluations, (2) evaluation through demonstration, and (3) user evaluations. Through this section, we aim to provide references for future research on GenAI systems, specifically for deciding the evaluation techniques for future systems.

Evaluation-1 Technical Evaluation. *Technical Evaluation* focuses on the performance of the backend, the algorithm, and the model of a system. Typical technical evaluation methods on system performance are qualitative assessment of the output [83, 176, 362], and quantitative measurement of the output via computing distance (e.g. BLEU for text and FID for images) between the generated and the expected in public datasets [56, 110, 126, 317]. The evaluations can be conducted on annotated datasets by the researchers themselves and the technical statistics of the datasets are also reported [8, 310].

Evaluation-2 Demonstration. Evaluations through *demonstrations* assess the system performance under specific conditions. Common methods consist of generalizability demonstration [163], proof of concept demonstration [174, 285, 346], demonstration through an example use case [39, 158, 345].

Evaluation-3 User Evaluation. *User evaluation* refers to measuring the performance of a system through user studies, focusing mostly on the effectiveness of the interaction designs in the system, which is hard to technically evaluate through uniform metrics. Common methods for user evaluation are questionnaires carefully designed to assess how well the design goals of the systems are satisfied [55, 195], qualitative lab studies [126] for rich insights into design and contextual understanding, quantitative lab studies [72] for objective measurement and generalizability, and interviews with both experts [240, 289] and novices [328, 347] in the subject matter.

10 FINDINGS

Based on the analysis of our taxonomy, Figure 10 shows a summary of the number of papers for each dimension In this section, we discuss the standard strategies and gaps that we identify through our analysis of the literature.
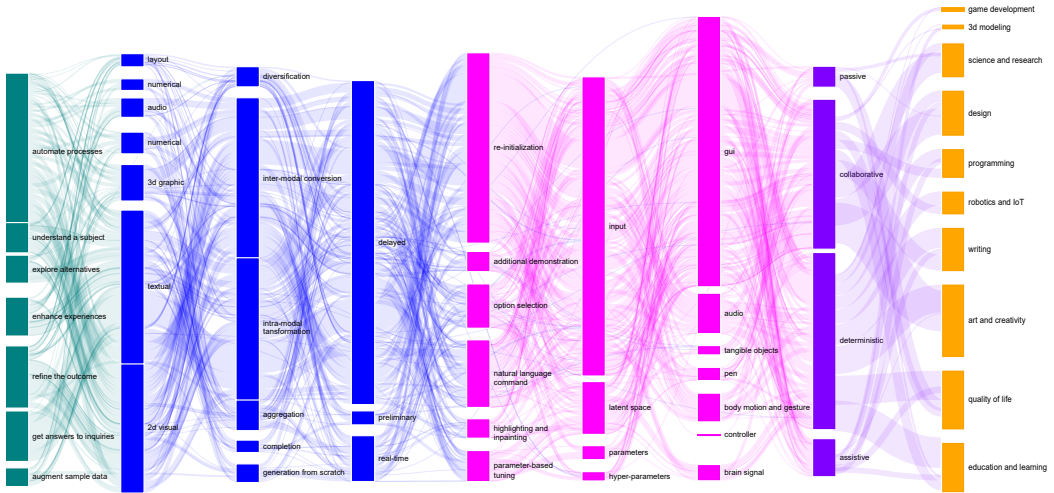

Fig. 10. An alluvial diagram of the characteristics from our survey across all dimensions.

Finding-1 Mediums of Control: Direct but not Intuitive. Through our literature review, we notice that direct control modalities are more applied (e.g. widgets, controllers, drawings and highlighting, and text, $N = 234$) than the intuitive ones (e.g. gestures, brain signals, and voice, $N = 49$) to control the output of the GenAI system. Direct control modalities allow users to modify the attributes of the models or data straightforwardly, while intuitive control modalities require mappings from the users' intuition to the functions of the models.

This inclination highlights that GenAI systems align more with the need to tweak the GenAI models for specific functions directly while overlooking the users' need for intuitive interactions. For example, using sliders to adjust the weights of the attributes of a GAN model [72, 73, 308] is direct yet not intuitive, because novice users (who do not know AI) do not possess the technical knowledge to understand the correspondences between the attributes and the outputs. When users interact with systems with straightforward interactions, they need to build the mapping between their interaction input and the output [3], while with intuitive interactions, researchers have preset this mapping for the users. Put simply, considering the Gen-AI systems as black boxes with unknown I/O correspondence, intuitive interactions foster smoother learning curves of this I/O correspondence than straightforward interactions. This observation suggests that intuitive human interactions are not yet the mainstream mediums for controlling the GenAI models.

Finding-2 Visualizing the Results rather than the Process. We notice that most GenAI systems do not reveal the intermediate layers or output to the users ($N = 231$). This is hard to accomplish from the AI-developing side, given the fact that it is hard for end-users to comprehend the mathematical functions that lie within the intermediate layers of the users. However, from an HCI perspective, we highlight the necessity of investigating the design space of interacting with the intermediate layers of the GenAI systems, which is a significant missing piece in the current

research. It is important that users understand the intention-action mapping [3, 232] of using a system, i.e. what consequences result from each of their interactions. From the designer's perspective, it is required to understand the complexity of the models so that designers can "conceptualize the system's behaviors in order to choreograph its interactions" [338]. From the user's perspective, users need to interpret outputs and express feedback to the models [68], which can benefit from a multi-layer design of the system compared to an end-to-end one. Starting from this consideration, we further discuss the future directions to address this concern in section 11

Finding-3 the Use of Foundation Models. We observed that a major GenAI utilized in the research is Large Language Models (LLMs, a significant subset of Foundation Models, which are defined as models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks [27, 160, 219] in 82 papers). Large language models have gained significant popularity due to their unparalleled ability to understand and produce human-like text. We also observed that most LLM-based applications utilize textual conversation as their interaction modality, implying that the users of these LLM-based systems interact by text input. This interaction follows the most instinctive patterns for LLMs, which are, after all, models of language. However, considering LLMs' overwhelming generative power and multi-modal potential, we suggest taking one step back and reconsidering the possible interaction modalities applicable to LLMs. For example, a voice command can be converted to text to converse [156], and vice versa [126]. Similarly, images can be summarized by models and translated into text for conversation as well [316]. Enlightened by this finding, we discuss the future opportunities of research in interaction design in section 11

Finding-4 Missing Discussion over Ethics. Through our discovery in the papers, we identify the missing piece of discussion over the ethical problems induced by the widespread application of GenAI. Out of the corpus of 291 research papers, only 21 papers discuss the potential ethical problems induced or tackled by their systems or studies. From our previous analysis of the papers, we identified similar patterns in the topics, methodologies, or application domains. We conclude that the applications of GenAI share similar ethical concerns that are yet to be addressed through further research. Examples of GenAI ethical problems we have located include GenAI plagiarism [64, 80, 231], opinionated bias in GenAI system [125], and gender bias in Natural Language generation [285]. We will be detailing the future opportunity of investigating how to tack GenAI ethical problems in section 11.

11   FUTURE OPPORTUNITIES

Opportunity-1 Designing and Exploring Interactions with Foundation Models. —*Various I/O Modalities through Foundation Models* In our survey, we identified promising usage of Foundation Models, as mentioned in section 10. While the Foundation Models have enabled diverse applications in domains associated with texts and images, we argue that further research can aim toward ***more intuitive modalities***, considering the cross-modality potential shown from both applications we've investigated [198, 272, 317]. Human conveys information through diverse means in addition to text and image. For example, the audio of natural language speaking can be converted into text as an approach to converse, a gesture or sign language may contain the information needed for instructing a robot, or a human gaze can guide the foundation models to generate descriptions of an object or an event in sight for educational purpose. AI and ML community have been widely discussed over the capabilities of Foundation Models learning the abstraction of human knowledge and generating representations of the knowledge in diverse modalities [27, 219]. With our proposed taxonomy identifying the design considerations of interaction mediums and modalities of the GenAI models, questions are to be addressed such as (1)**What are the possible modalities**

of Foundation Models that can be applied in interactive systems? For example, there is a limited amount of research that tackles the 3D content generation for AR/VR applications through Foundation Models. (2)**What are the effects of the conversion of modalities by Foundation Models on the users?** The abstraction of knowledge can be instantiated by Foundation Models across modalities, but what are the effects of the conversion on the users? How much of the information or knowledge is preserved through the generated modalities? How much of it is actually perceived by the users? Answering those questions will significantly guide the direction of future GenAI application development. (3) The former questions lead to the investigation of another question**What are the metrics to evaluate the interactions?** Since the popularity and continuing growth of Foundation Models are still novel, there is not yet a set of dedicated metrics to evaluate the quality of the interaction (not just of the model performance or the generated content as proposed in [221]) design for them. Finally, (4) **What are the general patterns we can conclude from the designs addressing the aforementioned questions?** A general framework for designing interactions with Foundation Models is called for the future development of this promising direction.

—*Diverse Applications through Foundation Models* Further from above, we argue that more diverse applications of Foundation Models can be introduced by future endeavors. First of all, through the capability of Foundation Models to handle diverse modalities I/O, **we anticipate consideration of the formats of data that were unable to be generated by the predecessors of current GenAI models**, which can be specifically used in a certain task. For example, there can be an application to generate a blueprint of a novel refrigerator (sketch, numerical data, and text as output) given users' routines of menus (text and image as input). Secondly, **we suggest that the interaction with Foundation Models (or GenAI in general) should not be constrained to merely collaborative tasks, but can also be applied to tasks with passive or deterministic engagement.** To be specific, with the strong generative power and capability to consume data in diverse forms, Foundation Models are able to actively understand the environment or context of the users and generate content that is to be passively consumed by the users. For example, a GenAI-based instructional AR system can scan the vision or environment of the users and detect the elements in the context (e.g. tools, furniture, and appliances), based on which it will predict the intended tasks of the users and generate corresponding AR instructions. To embrace the promising possibility of diverse applications through Foundation Models, questions remain unanswered **What are the types of information that can be passively perceived by the users and meanwhile be generated by the Foundation Models (or GenAI in general)? What are the types of contextual, semantic, or environmental information that can be used as the input to the models?**

Opportunity-2 Designing Interactions for Understanding GenAI. Pointed out by many prior works on human-AI interaction, a key consideration in designing an interface for AI-based applications is to enable the users to understand AI, in terms of the capabilities or potential mistakes to make [11]. Some [68] also argue that the users require the "*ability to retrace steps in the event that recent actions have resulted in an undesired outcome*" and that it is necessary to consider how to reduce the effort of the users to interpret the outputs and express the feedback. Based on these general discussions over human-AI interaction design, we investigate what open research questions are yet to be addressed specifically about GenAI.

—*Users Understanding the model capabilities* Different from predictive or discriminative AI models that perform a discrimination behavior, solving classification or regression problems [95], GenAI models generate new media. This difference creates a divergence between the interactions with predictive AI and those with GenAI. The capabilities of GenAI to be understood are vaguer and

more subjective than those of predictive AI since the latter mostly focuses on or works towards a specific or explicit goal while the former focuses on creativity or implicit tries to meet the users' expectations [222]. The questions remain unanswered: (1) **What are the metrics to evaluate how close the generated content is to users?** With the designated sets of metrics, we can investigate (2) **How can users themselves learn to interact with the systems to bring the generated content closer to their expectations?** To address this question, the interfaces or designers need to make clear to the users why the system did what it did [11]. For example, there are already a massive amount of empirical tutorials, blogs, posts, and even academic papers [211, 261, 322] on how to prompt with ChatGPT to direct the output of it. What are the essential HCI problems behind the empirical studies that can be applied to other GenAI systems? Answering this question leads to a clearer vision of how interactions should be designed so that the users understand the capabilities of the GenAI models they are interacting with. (3) **What are the correspondences between the model parameters and the model capabilities? And which capabilities (and their corresponding parameters) are to be made optional for users to control, considering the particular application domains the systems serve?** Interacting with internal parameters of the GenAI system allows users to explore GenAI model capabilities[55]. Only 18 papers out of 291 allow users to control the model parameters. One of the principles of User-centered Design is that the systems should allow users to guide the behavior of output to align with their preferences [3]. This involves adjusting certain GenAI model parameters and exploring the full capabilities of GenAI models. The correspondence between the GenAI model parameters and the actual generated content is different from that between predictive AI parameters and the if-else-rule-based prediction. A promising direction we see is in the domain of Explainable GenAI, where the correspondence can be studied and applied to the interaction design of actual systems. It is also crucial for system designers to understand the trade-off between the controllability and the complexity of the GenAI system [338]. Reducing the choices of parameters for the end-users, on the one hand, enhances user experience and increases efficiency, but on the other hand, lessens customization or adaptability of the system. A balance between the degree of freedom and efficiency of the system has yet to be revealed by future research.

—*Exploring Novel Interactions* We found limited numbers of dedicated mediums of interaction between the users and the GenAI systems. With natural interaction such as gesture-based and brain-controlled interfaces, users interact with devices and systems through modalities as intuitive as moving the hands or thinking about a picture, to obtain a desired output. We foresee the potential novel interactions with such modalities that **reduce the cognitive offset between user-expected output (resulting from the interactions) and the actual output.** For example, enabling the usage of brain signals to control the modifications to a generated image waives the cognitive cost of learning the correspondence between traditional GUI and output.

Furthermore, this allows the deployment of GenAI systems in more natural and immersive platforms, particularly in virtual reality (VR) and augmented reality (AR) applications. Also, BCIs have the potential to enable interactions without any physical movement, opening up possibilities for users with disabilities and new modes of interaction. Finally, natural interactions often come with technological difficulties such as user adaptation and personalization, subject to variability in user input and interaction. For example, people tend to express their feelings in different ways. Some prefer informal language, while others make ambiguous gestures. These differences pose challenges for designing an adaptive system that relies on users' expressions as input, say, to generate an image that describes their mood. The research questions to be addressed in these scenarios are: 1) **How do we integrate natural interactions with current GenAI models? What can be the connections between the natural interactions and the generation of the**

content? 2) **How do we accurately contextualize and adapt the generated content to users' natural interactions as input?**

Opportunity-3 Investigating the Uncertainty of GenAI. The uncertainty of an AI system lies in two major aspects. First, as mentioned in [338] the capabilities of the AI system are uncertain to both the HCI designers and the end-users. Second, the systems are prone to be more uncertain when the users come into the picture as the human factors [11]. To tackle both aspects of uncertainty, the AI systems are called to be adaptive (as discussed in the last subsection) and explainable respectively.

In the scope of GenAI, the domain of explainability in GenAI is still underexplored, although Explainable AI (XAI) has been discussed for general AI over the years. The explainability of AI is domain-specific and user-oriented. Few works like [289] investigated the need for explainability for a specific application scenario as well as its impact on the HCI considerations of the systems.

Opportunity-4 Multi-party Interaction Design. A less general yet interesting direction in future discussions of human-GenAI interaction is the multi-party interaction design. We anticipate a less constrained workflow with multiple users or multiple GenAI models engaging towards the same generative task. In this workflow, the users and the models can function in the same roles or not. For example, there can be multiple users brainstorming over a topic while a LLM-based system functions as a summarizer of the discussions from the users. In addition to the parties, there can be another diffusion-model-based system generating visualization based on the summarization from the LLM-based system. We propose to investigate the following research question in the potential scenario (1) **What is the difference in user experience when interacting with multiple GenAI models, when they are of the same and different roles?** (2) **What are the characterizations of the interactions between the multiple models in the workflow?** More interestingly, (3) **Can the roles setup be agnostic of whether they are filled by models or humans?** And finally, (4) **How much can models and humans each contribute to the final content generation?**

Opportunity-5 Human-GenAI Ethics Discussion. As stated in the Discussion and Findings, we identified the missing pieces of discussion over ethical problems induced by GenAI. The impact of the problems varies across different applications, such as shrinking the job market [91], intrusion into copyrights and intellectual property [231], and generation of illegal content [9, 52]. We will describe two common problems as examples to open the floor for further discussion for researchers to take into consideration when addressing more foreseen ethical concerns.

—*Credit Assignment between GenAI and Human* In the GenAI applications where the final output is of market values or artistic attributes, it is still vague and undefined how the credit between GenAI and Human creators should be rigorously distributed, despite the heated discussion over this topic. The credit assignment dynamic between GenAI systems and human users exemplifies a modern collaboration where innovation is nurtured through a symbiotic relationship. For example, if an artist creates a painting using a GAN-based system, does he or the GenAI system deserve credit for this artwork? It would be unrealistic to claim that GenAI should take full credit, for the fact that there would not be art without humans as long as human interactions are the external force fostering this artistic creation. Yet, one can easily see the flaws and unfairness in giving the human artist the full credit, because, in the creation process, GenAI contributes to the final result, whether in ideation, styling, or any fundamental stroke. It is also a weak argument that GenAI (or AI in general) is not human and deserves no credit in human work, considering the human efforts in implementing the model and creating the artwork sample training this model. With all these being said, we propose to take the middle ground that both sides share the credit. The credit lies in the harmonious exchange: AI offers a canvas, while humans contribute a vivid

palette of experiences, cultural nuances, and depth of understanding. However, a rigorous pattern for credit assignment will not emerge until the following questions are addressed: (1) **What is the definition of creativity in the context of human-GenAI collaboration?** (2) **Should the data being used to train GenAI be considered contributing to the generated content?** (3) **What is the taxonomy of human-GenAI interactions that can help define the contribution of a work?**

  —*Inappropriate Use of GenAI* Generated content can be harmful in many possibilities, such as generating biased or opinionated data for educational content, overlooking the needs of minority groups, generating illegal content that poses threats to society (e.g. rumors), or breaching basic human rights (e.g. identity theft in fake content). We call for rigorous and clarified rules, regulations, and laws in the domain, which are also considered significant parts of human-GenAI interactions. Only with clear-defined appropriate applications and usages of GenAI, shall we foster a positive impact of GenAI on the existing human industries and communities.

12 CONCLUSION

In this paper, we present a survey on existing GenAI applications and research, deriving a taxonomy of human-GenAI interactions. We synthesize the existing research in this scope and discuss their (1) Purposes of Interacting with GenAI, (2) Feedback from Models to Users, (3) Control from Users to Models, (4) Levels of Engagement, (5) Application Domains, and (6) Evaluation Strategies. Our research aims to provide an overview of the landscape of the topic of human-GenAI interaction and the common ground of application design. Further, we discuss future opportunities in this topic, namely, (1) designing interactions for Foundation Models, (2) designing interactions for understanding GenAI, (3) investigating the Uncertainty of GenAI, (4) multi-party interaction design, and (5) ethical discussion on GenAI. We hope our research will guide and inspire future work on human-GenAI interaction.

REFERENCES

[1] Rajagopal A, Nirmala V, Immanuel Johnraja Jebadurai, Arun Muthuraj Vedamanickam, and Prajakta Uthaya Kumar. 2023. Design of Generative Multimodal AI Agents to Enable Persons with Learning Disability. In *Companion Publication of the 25th International Conference on Multimodal Interaction* (Paris, France) *(ICMI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 259–271. https://doi.org/10.1145/3610661.3617514

[2] Rameen Abdal, Peihao Zhu, John Femiani, Niloy Mitra, and Peter Wonka. 2022. Clip2stylegan: Unsupervised extraction of stylegan edit directions. In *ACM SIGGRAPH 2022 conference proceedings*. 1–9.

[3] Chadia Abras, Diane Maloney-Krichmar, Jenny Preece, et al. 2004. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications* 37, 4 (2004), 445–456.

[4] Telmo Adão, João Oliveira, Somayeh Shahrabadi, Hugo Jesus, Marco Fernandes, Ângelo Costa, Vânia Ferreira, Martinho Fradeira Gonçalves, Miguel A Guevara Lopéz, Emanuel Peres, et al. 2023. Empowering Deaf-Hearing Communication: Exploring Synergies between Predictive and Generative AI-Based Strategies towards (Portuguese) Sign Language Interpretation. *Journal of Imaging* 9, 11 (2023), 235.

[5] Obinna Agbodike, Chiao-Hua Huang, and Jenhui Chen. 2020. Cognitive Attention Network (CAN) for Text and Image Multimodal Visual Dialog Systems. In *2020 6th International Conference on Applied System Innovation (ICASI)*. IEEE, 37–41.

[6] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, et al. 2022. Do as i can, not as i say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691* (2022).

[7] Emre Aksan and Otmar Hilliges. 2021. Generative Ink: Data-Driven Computational Models for Digital Ink. *Artificial Intelligence for Human Computer Interaction: A Modern Approach* (2021), 417–461.

[8] Emre Aksan, Fabrizio Pece, and Otmar Hilliges. 2018. Deepwriting: Making digital ink editable via deep generative modeling. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.

[9] Mohammad Al-Rubaie and J Morris Chang. 2019. Privacy-preserving machine learning: Threats and solutions. *IEEE Security & Privacy* 17, 2 (2019), 49–58.

[10] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, et al. 2022. Flamingo: a visual language model for few-shot learning. *Advances in Neural Information Processing Systems* 35 (2022), 23716–23736.

[11] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

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

[13] Kota Arai, Yutaro Hirao, Takuji Narumi, Tomohiko Nakamura, Shinnosuke Takamichi, and Shigeo Yoshida. 2023. TimToShape: Supporting Practice of Musical Instruments by Visualizing Timbre with 2D Shapes based on Crossmodal Correspondences. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 850–865.

[14] Riku Arakawa, Hiromu Yakura, and Masataka Goto. 2023. 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*. 1–19.

[15] Riku Arakawa, Hiromu Yakura, and Sosuke Kobayashi. 2022. VocabEncounter: NMT-powered vocabulary learning by presenting computer-generated usages of foreign words into users' daily lives. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–21.

[16] Trevor Ashby, Braden K Webb, Gregory Knapp, Jackson Searle, and Nancy Fulda. 2023. Personalized Quest and Dialogue Generation in Role-Playing Games: A Knowledge Graph-and Language Model-based Approach. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[17] Eran Bamani, Nadav D Kahanowich, Inbar Ben-David, and Avishai Sintov. 2021. Robust multi-user in-hand object recognition in human-robot collaboration using a wearable force-myography device. *IEEE Robotics and Automation Letters* 7, 1 (2021), 104–111.

[18] Junseong Bang, Byung-Tak Lee, and Pangun Park. 2023. Examination of Ethical Principles for LLM-Based Recommendations in Conversational AI. In *2023 International Conference on Platform Technology and Service (PlatCon)*. IEEE, 109–113.

[19] David Bau, Hendrik Strobelt, William Peebles, Jonas Wulff, Bolei Zhou, Jun-Yan Zhu, and Antonio Torralba. 2020. Semantic photo manipulation with a generative image prior. *arXiv preprint arXiv:2005.07727* (2020).

[20] Jesse Josua Benjamin, Heidi Biggs, Arne Berger, Julija Rukanskaitė, Michael B Heidt, Nick Merrill, James Pierce, and Joseph Lindley. 2023. The Entoptic Field Camera as Metaphor-Driven Research-through-Design with AI Technologies. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[21] Michael S Bernstein, Joon Sung Park, Meredith Ringel Morris, Saleema Amershi, Lydia Chilton, and Mitchell L Gordon. 2018. Architecting Novel Interactions With Generative AI Models. (2018).

[22] Rishabh Bhargava, Russel Lobo, Rushabh Shah, Nishank Shah, and Sindhu Nair. 2022. Easier Web Navigation Using Intent Classification, Web Scraping and NLP Approaches. In *2022 5th International Conference on Advances in Science and Technology (ICAST)*. 286–290. https://doi.org/10.1109/ICAST55766.2022.10039559

[23] Advait Bhat, Saaket Agashe, Parth Oberoi, Niharika Mohile, Ravi Jangir, and Anirudha Joshi. 2023. Interacting with Next-Phrase Suggestions: How Suggestion Systems Aid and Influence the Cognitive Processes of Writing. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 436–452.

[24] Bhavya Bhavya, Jinjun Xiong, and Chengxiang Zhai. 2023. Cam: A large language model-based creative analogy mining framework. In *Proceedings of the ACM Web Conference 2023*. 3903–3914.

[25] Yali Bian, Chris North, Eric Krokos, and Sarah Joseph. 2021. Semantic explanation of interactive dimensionality reduction. In *2021 IEEE Visualization Conference (VIS)*. IEEE, 26–30.

[26] Vijaylaxmi Bittal, Anuradha Sanjay Bachhav, Makarand Shahade, Pradnya Rajendra Chavan, Bhavesh Anil Nikam, and Ajinkya Anil Pawar. 2023. Speech To Image Translation Framework for Teacher-Student Learning. In *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*. IEEE, 1515–1520.

[27] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).

[28] Andrew Brock, Jeff Donahue, and Karen Simonyan. 2018. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096* (2018).

[29] Lennart Brocki, George C Dyer, Anna Gładka, and Neo Christopher Chung. 2023. Deep learning mental health dialogue system. In *2023 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 395–398.

[30] Anthony Brohan, Yevgen Chebotar, Chelsea Finn, Karol Hausman, Alexander Herzog, Daniel Ho, Julian Ibarz, Alex Irpan, Eric Jang, Ryan Julian, et al. 2023. Do as i can, not as i say: Grounding language in robotic affordances. In *Conference on Robot Learning*. PMLR, 287–318.

[31] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[32] Arthur Bucker, Luis Figueredo, Sami Haddadinl, Ashish Kapoor, Shuang Ma, and Rogerio Bonatti. 2022. Reshaping robot trajectories using natural language commands: A study of multi-modal data alignment using transformers. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 978–984.

[33] Daniel Buschek, Martin Zürn, and Malin Eiband. 2021. The impact of multiple parallel phrase suggestions on email input and composition behaviour of native and non-native english writers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[34] Tomas Cabezon Pedroso, Javier Del Ser, and Natalia Díaz-Rodríguez. 2022. Capabilities, limitations and challenges of style transfer with CycleGANs: a study on automatic ring design generation. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 168–187.

[35] Zefeng Cai and Zerui Cai. 2022. PEVAE: A Hierarchical VAE for Personalized Explainable Recommendation.. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 692–702.

[36] Alex Calderwood, Vivian Qiu, Katy Ilonka Gero, and Lydia B Chilton. 2020. How Novelists Use Generative Language Models: An Exploratory User Study.. In *HAI-GEN+ user2agent@ IUI*.

[37] Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S Yu, and Lichao Sun. 2023. A comprehensive survey of ai-generated content (aigc): a history of generative ai from gan to chatgpt. *arXiv preprint arXiv:2303.04226* (2023).

[38] Benjamin Capps and Jacob Schrum. 2021. Using multiple generative adversarial networks to build better-connected levels for mega man. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 66–74.

[39] Minsuk Chang, Stefania Druga, Alexander J Fiannaca, Pedro Vergani, Chinmay Kulkarni, Carrie J Cai, and Michael Terry. 2023. The Prompt Artists. In *Proceedings of the 15th Conference on Creativity and Cognition*. 75–87.

[40] Minwook Chang, Youngwon Ryan Kim, and Gerard Jounghyun Kim. 2018. A Perceptual Evaluation of Generative Adversarial Network Real-Time Synthesized Drum Sounds in a Virtual Environment. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 144–148.

[41] Chien-Wen Chen, Wen-Cheng Chen, and Min-Chun Hu. 2018. Doodle master: a doodle beautification system based on auto-encoding generative adversarial networks. In *Proceedings of the 2018 International Joint Workshop on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia*. 2–7.

[42] Jiali Chen, Changjie Fan, Zhimeng Zhang, Gongzheng Li, Zeng Zhao, Zhigang Deng, and Yu Ding. 2021. A music-driven deep generative adversarial model for guzheng playing animation. *IEEE Transactions on Visualization and Computer Graphics* (2021).

[43] Jin Chen, Guangyi Xiao, Xu Han, and Hao Chen. 2021. Controllable and Editable Neural Story Plot Generation via Control-and-Edit Transformer. *IEEE Access* 9 (2021), 96692–96699.

[44] Xiang'Anthony' Chen, Jeff Burke, Ruofei Du, Matthew K Hong, Jennifer Jacobs, Philippe Laban, Dingzeyu Li, Nanyun Peng, Karl DD Willis, Chien-Sheng Wu, et al. 2023. Next Steps for Human-Centered Generative AI: A Technical Perspective. *arXiv preprint arXiv:2306.15774* (2023).

[45] Yu Cheng, Zhe Gan, Yitong Li, Jingjing Liu, and Jianfeng Gao. 2020. Sequential attention GAN for interactive image editing. In *Proceedings of the 28th ACM international conference on multimedia*. 4383–4391.

[46] Li-Yuan Chiou, Peng-Kai Hung, Rung-Huei Liang, and Chun-Teng Wang. 2023. Designing with AI: An Exploration of Co-Ideation with Image Generators. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 1941–1954.

[47] Toby Chong, I-Chao Shen, Issei Sato, and Takeo Igarashi. 2021. Interactive Optimization of Generative Image Modelling using Sequential Subspace Search and Content-based Guidance. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 279–292.

[48] Jeremy Chu, Dongsheng An, Yan Ma, Wenzhe Cui, Shumin Zhai, Xianfeng David Gu, and Xiaojun Bi. 2023. WordGesture-GAN: Modeling Word-Gesture Movement with Generative Adversarial Network. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[49] John Joon Young Chung, Minsuk Chang, and Eytan Adar. 2021. Gestural Inputs as Control Interaction for Generative Human-AI Co-Creation. In *Workshops at the International Conference on Intelligent User Interfaces (IUI)*.

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

[51] Azelle Courtial, Guillaume Touya, and Xiang Zhang. 2023. Deriving map images of generalised mountain roads with generative adversarial networks. *International Journal of Geographical Information Science* 37, 3 (2023), 499–528.

[52] George Danezis, Josep Domingo-Ferrer, Marit Hansen, Jaap-Henk Hoepman, Daniel Le Metayer, Rodica Tirtea, and Stefan Schiffner. 2015. Privacy and data protection by design-from policy to engineering. *arXiv preprint*

arXiv:1501.03726 (2015).

[53] Hai Dang, Karim Benharrak, Florian Lehmann, and Daniel Buschek. 2022. Beyond text generation: Supporting writers with continuous automatic text summaries. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[54] Hai Dang, Sven Goller, Florian Lehmann, and Daniel Buschek. 2023. Choice over control: How users write with large language models using diegetic and non-diegetic prompting. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[55] Hai Dang, Lukas Mecke, and Daniel Buschek. 2022. GANSlider: How Users Control Generative Models for Images using Multiple Sliders with and without Feedforward Information. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[56] Keith M Davis, Carlos de la Torre-Ortiz, and Tuukka Ruotsalo. 2022. Brain-supervised image editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18480–18489.

[57] Richard Lee Davis, Thiemo Wambsganss, Wei Jiang, Kevin Gonyop Kim, Tanja Käser, and Pierre Dillenbourg. 2023. Fashioning the Future: Unlocking the Creative Potential of Deep Generative Models for Design Space Exploration. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–9.

[58] Carlos de la Torre-Ortiz, Michiel M Spapé, Lauri Kangassalo, and Tuukka Ruotsalo. 2020. Brain relevance feedback for interactive image generation. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 1060–1070.

[59] Niklas Deckers, Maik Fröbe, Johannes Kiesel, Gianluca Pandolfo, Christopher Schröder, Benno Stein, and Martin Potthast. 2023. The Infinite Index: Information Retrieval on Generative Text-To-Image Models. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 172–186.

[60] Paul Denny, Viraj Kumar, and Nasser Giacaman. 2023. Conversing with copilot: Exploring prompt engineering for solving cs1 problems using natural language. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*. 1136–1142.

[61] Joseph F DeRose, Jiayao Wang, and Matthew Berger. 2020. Attention flows: Analyzing and comparing attention mechanisms in language models. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 1160–1170.

[62] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[63] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion Models Beat GANs on Image Synthesis. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.), Vol. 34. Curran Associates, Inc., 8780–8794. https://proceedings.neurips.cc/paper_files/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf

[64] Joseph Dien. 2023. Editorial: Generative artificial intelligence as a plagiarism problem. *Biological Psychology* 181 (2023), 108621. https://doi.org/10.1016/j.biopsycho.2023.108621

[65] Zijian Ding, Arvind Srinivasan, Stephen MacNeil, and Joel Chan. 2023. Fluid Transformers and Creative Analogies: Exploring Large Language Models' Capacity for Augmenting Cross-Domain Analogical Creativity. In *Proceedings of the 15th Conference on Creativity and Cognition*. 489–505.

[66] Hanxiang Du, Wanli Xing, and Bo Pei. 2023. Automatic text generation using deep learning: providing large-scale support for online learning communities. *Interactive Learning Environments* 31, 8 (2023), 5021–5036.

[67] Tinglin Duan, Parinya Punpongsanon, Sheng Jia, Daisuke Iwai, Kosuke Sato, and Konstantinos N Plataniotis. 2019. Remote environment exploration with drone agent and haptic force feedback. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 167–1673.

[68] John J Dudley and Per Ola Kristensson. 2018. A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 2 (2018), 1–37.

[69] Rinky Dwivedi, Anushka Jadhav, Prachi Jha, and Anubhav. 2022. Predicting 3D Model of Images for Augmented Reality Applications. In *International Conference on Computing in Engineering & Technology*. Springer, 395–404.

[70] Jeffrey L Elman. 1990. Finding structure in time. *Cognitive science* 14, 2 (1990), 179–211.

[71] Yuki Endo. 2022. User-Controllable Latent Transformer for StyleGAN Image Layout Editing. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 395–406.

[72] Noyan Evirgen and Xiang'Anthony' Chen. 2022. GANzilla: User-Driven Direction Discovery in Generative Adversarial Networks. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–10.

[73] Noyan Evirgen and Xiang'Anthony Chen. 2023. GANravel: User-Driven Direction Disentanglement in Generative Adversarial Networks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[74] Fatemeh Fahimi, Strahinja Dosen, Kai Keng Ang, Natalie Mrachacz-Kersting, and Cuntai Guan. 2020. Generative adversarial networks-based data augmentation for brain–computer interface. *IEEE transactions on neural networks and learning systems* 32, 9 (2020), 4039–4051.

[75] Judith E Fan, Monica Dinculescu, and David Ha. 2019. Collabdraw: an environment for collaborative sketching with an artificial agent. In *Proceedings of the 2019 on Creativity and Cognition*. 556–561.

[76] Ahmed Fares, Sheng-hua Zhong, and Jianmin Jiang. 2020. Brain-media: A dual conditioned and lateralization supported GAN (DCLS-GAN) towards visualization of image-evoked brain activities. In *Proceedings of the 28th ACM International Conference on Multimedia*. 1764–1772.

[77] Faraz Faruqi, Ahmed Katary, Tarik Hasic, Amira Abdel-Rahman, Nayeemur Rahman, Leandra Tejedor, Mackenzie Leake, Megan Hofmann, and Stefanie Mueller. 2023. Style2Fab: Functionality-Aware Segmentation for Fabricating Personalized 3D Models with Generative AI. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[78] Matthias Feurer and Frank Hutter. 2019. Hyperparameter optimization. *Automated machine learning: Methods, systems, challenges* (2019), 3–33.

[79] James Finnie-Ansley, Paul Denny, Andrew Luxton-Reilly, Eddie Antonio Santos, James Prather, and Brett A Becker. 2023. My AI Wants to Know if This Will Be on the Exam: Testing OpenAI's Codex on CS2 Programming Exercises. In *Proceedings of the 25th Australasian Computing Education Conference*. 97–104.

[80] Errol Francke and Alexander Bennett. 2019. The potential influence of artificial intelligence on plagiarism: A higher education perspective. In *European Conference on the Impact of Artificial Intelligence and Robotics (ECIAIR 2019)*. 131–140.

[81] Emma Frid, Celso Gomes, and Zeyu Jin. 2020. Music creation by example. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[82] Liye Fu, Benjamin Newman, Maurice Jakesch, and Sarah Kreps. 2023. Comparing Sentence-Level Suggestions to Message-Level Suggestions in AI-Mediated Communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.

[83] Qiang Fu, Xiaowu Chen, Xiaotian Wang, Sijia Wen, Bin Zhou, and Hongbo Fu. 2017. Adaptive synthesis of indoor scenes via activity-associated object relation graphs. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 1–13.

[84] Yuki Fujita, Haoran Xie, and Kazunori Miyata. 2019. Perceptual font manifold from generative model. In *2019 Nicograph International (NicoInt)*. IEEE, 41–48.

[85] Yuuya Fukumoto, Daiki Shimizu, and Chihiro Shibata. 2018. Generation of character illustrations from stick figures using a modification of generative adversarial network. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, Vol. 2. IEEE, 183–186.

[86] Kunihiko Fukushima. 1988. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks* 1, 2 (1988), 119–130.

[87] Karolina Gabor-Siatkowska, Marcin Sowański, Rafał Rzatkiewicz, Izabela Stefaniak, Marek Kozłowski, and Artur Janicki. 2023. AI to Train AI: Using ChatGPT to Improve the Accuracy of a Therapeutic Dialogue System. *Electronics* 12, 22 (2023), 4694.

[88] Xingen Gao, Changle Zhou, Fei Chao, Longzhi Yang, Chih-Min Lin, and Changjing Shang. 2019. A robotic writing framework–learning human aesthetic preferences via human–machine interactions. *IEEE Access* 7 (2019), 144043–144053.

[89] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2414–2423.

[90] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for science writing using language models. In *Designing interactive systems conference*. 1002–1019.

[91] Avijit Ghosh and Genoveva Fossas. 2022. Can there be art without an artist? *arXiv preprint arXiv:2209.07667* (2022).

[92] Yan Gong and Georgina Cosma. 2023. Boon: A Neural Search Engine for Cross-Modal Information Retrieval. *arXiv preprint arXiv:2307.14240* (2023).

[93] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[94] 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. 2022. 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*. 1–18.

[95] Roberto Gozalo-Brizuela and Eduardo C Garrido-Merchan. 2023. ChatGPT is not all you need. A State of the Art Review of large Generative AI models. *arXiv preprint arXiv:2301.04655* (2023).

[96] Roberto Gozalo-Brizuela and Eduardo C Garrido-Merchán. 2023. A survey of Generative AI Applications. *arXiv preprint arXiv:2306.02781* (2023).

[97] Alex Graves and Alex Graves. 2012. Long short-term memory. *Supervised sequence labelling with recurrent neural networks* (2012), 37–45.

[98] Antonio Greco, Antonio Roberto, Alessia Saggese, and Mario Vento. 2022. Efficient Transformers for on-robot Natural Language Understanding. In *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*. IEEE, 823–828.

[99] Éric Guérin, Julie Digne, Eric Galin, Adrien Peytavie, Christian Wolf, Bedrich Benes, and Benoît Martinez. 2017. Interactive example-based terrain authoring with conditional generative adversarial networks. *ACM Trans. Graph.* 36, 6 (2017), 228–1.

[100] Jingtao Guo and Yi Liu. 2020. Attributes guided facial image completion. *Neurocomputing* 392 (2020), 60–69.

[101] Rui Guo, Ivor Simpson, Chris Kiefer, Thor Magnusson, and Dorien Herremans. 2022. MusIAC: An extensible generative framework for Music Infilling Applications with multi-level Control. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)*. Springer, 341–356.

[102] Shunan Guo, Zhuochen Jin, Fuling Sun, Jingwen Li, Zhaorui Li, Yang Shi, and Nan Cao. 2021. Vinci: an intelligent graphic design system for generating advertising posters. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–17.

[103] Shuyu Guo, Shuo Zhang, Weiwei Sun, Pengjie Ren, Zhumin Chen, and Zhaochun Ren. 2023. Towards Explainable Conversational Recommender Systems. *arXiv preprint arXiv:2305.18363* (2023).

[104] Tao Guo, Song Guo, and Junxiao Wang. 2023. pFedPrompt: Learning Personalized Prompt for Vision-Language Models in Federated Learning. In *Proceedings of the ACM Web Conference 2023*. 1364–1374.

[105] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2022. Neural Language Models as What If?-Engines for HCI Research. In *27th International Conference on Intelligent User Interfaces*. 77–80.

[106] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating large language models in generating synthetic hci research data: a case study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[107] Xu Han, Michelle Zhou, Matthew J Turner, and Tom Yeh. 2021. Designing effective interview chatbots: Automatic chatbot profiling and design suggestion generation for chatbot debugging. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[108] Rola Harmouche, Aidan Lochbihler, Francis Thibault, Gino De Luca, Catherine Proulx, and Jordan L Hovdebo. 2020. A virtual assistant for cybersickness care. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 384–387.

[109] GM Harshvardhan, Mahendra Kumar Gourisaria, Manjusha Pandey, and Siddharth Swarup Rautaray. 2020. A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review* 38 (2020), 100285.

[110] Feixiang He, Yanlong Huang, and He Wang. 2022. iPLAN: interactive and procedural layout planning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7793–7802.

[111] Yingxu He, Lizi Liao, Zheng Zhang, and Tat-Seng Chua. 2021. Towards enriching responses with crowd-sourced knowledge for task-oriented dialogue. In *Proceedings of the 2nd ACM multimedia workshop on multimodal conversational AI*. 3–11.

[112] Keita Higuchi, Shotaro Sano, and Takeo Igarashi. 2021. Interactive Hyperparameter Optimization with Paintable Timelines. In *Designing Interactive Systems Conference 2021*. 1518–1528.

[113] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. arXiv:2006.11239 [cs.LG]

[114] Masoumehsadat Hosseini, Tjado Ihmels, Ziqian Chen, Marion Koelle, Heiko Müller, and Susanne Boll. 2023. Towards a Consensus Gesture Set: A Survey of Mid-Air Gestures in HCI for Maximized Agreement Across Domains. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–24.

[115] Wenjun Hou and Yiming Cao. 2021. Intelligent Question Answering System Based on Knowledge Graph of Beijing Opera. In *2021 3rd International Conference on Natural Language Processing (ICNLP)*. IEEE, 87–92.

[116] Chao-Chun Hsu, Yu-Hua Chen, Zi-Yuan Chen, Hsin-Yu Lin, Ting-Hao Huang, and Lun-Wei Ku. 2019. Dixit: Interactive visual storytelling via term manipulation. In *The World Wide Web Conference*. 3531–3535.

[117] Ruizhen Hu, Zeyu Huang, Yuhan Tang, Oliver Van Kaick, Hao Zhang, and Hui Huang. 2020. Graph2plan: Learning floorplan generation from layout graphs. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 118–1.

[118] Yongquan Hu, Mingyue Yuan, Kaiqi Xian, Don Samitha Elvitigala, and Aaron Quigley. 2023. Exploring the Design Space of Employing AI-Generated Content for Augmented Reality Display. *arXiv preprint arXiv:2303.16593* (2023).

[119] Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, et al. 2022. Inner monologue: Embodied reasoning through planning with language models. *arXiv preprint arXiv:2207.05608* (2022).

[120] Bernd Huber, Daniel McDuff, Chris Brockett, Michel Galley, and Bill Dolan. 2018. Emotional dialogue generation using image-grounded language models. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[121] Shunki Itadera, Taisuke Kobayashi, Jun Nakanishi, Tadayoshi Aoyama, and Yasuhisa Hasegawa. 2021. Towards physical interaction-based sequential mobility assistance using latent generative model of movement state. *Advanced Robotics* 35, 1 (2021), 64–79.

[122] Koichiro Ito, Masaki Murata, Tomohiro Ohno, and Shigeki Matsubara. 2021. Estimating the Generation Timing of Responsive Utterances by Active Listeners of Spoken Narratives. In *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 495–502.

[123] Anish Jain, Diti Modi, Rudra Jikadra, and Shweta Chachra. 2019. Text to image generation of fashion clothing. In *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 355–358.

[124] Naman Jain, Skanda Vaidyanath, Arun Iyer, Nagarajan Natarajan, Suresh Parthasarathy, Sriram Rajamani, and Rahul Sharma. 2022. Jigsaw: Large language models meet program synthesis. In *Proceedings of the 44th International Conference on Software Engineering*. 1219–1231.

[125] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-writing with opinionated language models affects users' views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[126] Ruben Janssens, Pieter Wolfert, Thomas Demeester, and Tony Belpaeme. 2022. 'Cool glasses, where did you get them?" Generating Visually Grounded Conversation Starters for Human-Robot Dialogue. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 821–825.

[127] Ellen Jiang, Kristen Olson, Edwin Toh, Alejandra Molina, Aaron Donsbach, Michael Terry, and Carrie J Cai. 2022. Promptmaker: Prompt-based prototyping with large language models. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–8.

[128] Ellen Jiang, Edwin Toh, Alejandra Molina, Kristen Olson, Claire Kayacik, Aaron Donsbach, Carrie J Cai, and Michael Terry. 2022. Discovering the syntax and strategies of natural language programming with generative language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

[129] Ming Jiang and Li Zhang. 2021. An Interactive Evolution Strategy based Deep Convolutional Generative Adversarial Network for 2D Video Game Level Procedural Content Generation. In *2021 International Joint Conference on Neural Networks (IJCNN)*. 1–6. https://doi.org/10.1109/IJCNN52387.2021.9533847

[130] Peiling Jiang, Jude Rayan, Steven P Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. *arXiv preprint arXiv:2305.11473* (2023).

[131] Xin Jin, Chenyu Fan, Biao Wang, Chaoen Xiao, and Chao Xia. 2023. Human Pose Recommendation and Profession-alization. In *Proceedings of the 1st International Workshop on Multimedia Content Generation and Evaluation: New Methods and Practice*. 91–98.

[132] Yuxi Jin, Ping Li, Wenxiao Wang, Suiyun Zhang, Di Lin, and Chengjiu Yin. 2023. GAN-based pencil drawing learning system for art education on large-scale image datasets with learning analytics. *Interactive Learning Environments* 31, 5 (2023), 2544–2561.

[133] Qianzhi Jing, Tingting Zhou, Yixin Tsang, Liuqing Chen, Lingyun Sun, Yankun Zhen, and Yichun Du. 2023. Layout Generation for Various Scenarios in Mobile Shopping Applications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[134] Eunkyung Jo, Daniel A Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[135] Brian Jordan, Nisha Devasia, Jenna Hong, Randi Williams, and Cynthia Breazeal. 2021. PoseBlocks: A toolkit for creating (and dancing) with AI. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 15551–15559.

[136] Julian Just, Thomas Ströhle, Johann Füller, and Katja Hutter. 2023. AI-based novelty detection in crowdsourced idea spaces. *Innovation* (2023), 1–28.

[137] Florian Kadner, Yannik Keller, and Constantin Rothkopf. 2021. 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*. 1–11.

[138] Minsuk Kahng and Duen Horng Polo Chau. 2020. How does visualization help people learn deep learning? Evaluating GAN Lab with observational study and log analysis. In *2020 IEEE Visualization Conference (VIS)*. IEEE, 266–270.

[139] Minsuk Kahng, Nikhil Thorat, Duen Horng Chau, Fernanda B Viégas, and Martin Wattenberg. 2018. Gan lab: Understanding complex deep generative models using interactive visual experimentation. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 310–320.

[140] Harshil T Kanakia and Suraj Prakkash Nair. 2023. Designing a User-Friendly and Responsive AI based Image Generation Website and Performing Diversity Assessment of the Generated Images. In *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE, 1090–1097.

[141] Hayata Kaneko, Yuuya Yoshizu, Ryuto Ishibashi, and Lin Meng. 2023. An Attempt at Zero-shot Ancient Documents Restoration Based on Diffusion Models. In *2023 International Conference on Advanced Mechatronic Systems (ICAMechS)*.

IEEE, 1–6.

[142] Lauri Kangassalo, Michiel Spapé, and Tuukka Ruotsalo. 2020. Neuroadaptive modelling for generating images matching perceptual categories. *Scientific reports* 10, 1 (2020), 14719.

[143] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).

[144] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4401–4410.

[145] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8110–8119.

[146] Isaak Kavasidis, Simone Palazzo, Concetto Spampinato, Daniela Giordano, and Mubarak Shah. 2017. Brain2image: Converting brain signals into images. In *Proceedings of the 25th ACM international conference on Multimedia*. 1809–1817.

[147] Majeed Kazemitabaar, Justin Chow, Carl Ka To Ma, Barbara J Ericson, David Weintrop, and Tovi Grossman. 2023. Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–23.

[148] Abe Kazemzadeh. 2022. BERT-Assisted Semantic Annotation Correction for Emotion-Related Questions. In *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 242–247.

[149] Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. 2019. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858* (2019).

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

[151] Dae Hyun Kim, Enamul Hoque, and Maneesh Agrawala. 2020. Answering questions about charts and generating visual explanations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[152] Hanseob Kim, Ghazanfar Ali, Seungwon Kim, Gerard J Kim, and Jae-In Hwang. 2021. Auto-generating virtual human behavior by understanding user contexts. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 591–592.

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

[154] Jiin Kim, Heekyung Yang, and Kyungha Min. 2022. 24-GAN: Portrait Generation with Composite Attributes. *Mathematics* 10, 20 (2022), 3887.

[155] Jong Wook Kim and Beakcheol Jang. 2022. Deep learning-based privacy-preserving framework for synthetic trajectory generation. *Journal of Network and Computer Applications* 206 (2022), 103459.

[156] Tae Soo Kim, DaEun Choi, Yoonseo Choi, and Juho Kim. 2022. Stylette: Styling the web with natural language. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[157] Naoki Kimura, Michinari Kono, and Jun Rekimoto. 2019. Deep dive: Deep-neural-network-based video extension for immersive head-mounted display experiences. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays*. 1–7.

[158] Naoki Kimura and Jun Rekimoto. 2018. ExtVision: augmentation of visual experiences with generation of context images for a peripheral vision using deep neural network. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–10.

[159] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[160] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. *arXiv preprint arXiv:2304.02643* (2023).

[161] Hyung-Kwon Ko, Subin An, Gwanmo Park, Seung Kwon Kim, Daesik Kim, Bohyoung Kim, Jaemin Jo, and Jinwook Seo. 2022. We-toon: A Communication Support System between Writers and Artists in Collaborative Webtoon Sketch Revision. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[162] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. 2019. May AI? Design ideation with cooperative contextual bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[163] Yuki Koyama and Masataka Goto. 2022. BO as Assistant: Using Bayesian Optimization for Asynchronously Generating Design Suggestions. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[164] BMUS Kularatne, BMJN Basnayake, PDLAM Sathmini, GVU Sewwandi, Samantha Rajapaksha, and Dilshan De Silva. 2022. Elderly Care Home Robot using Emotion Recognition, Voice Recognition and Medicine Scheduling. In *2022 7th International Conference on Information Technology Research (ICITR)*. IEEE, 1–6.

[165] Oh-Hyun Kwon and Kwan-Liu Ma. 2019. A deep generative model for graph layout. *IEEE Transactions on visualization and Computer Graphics* 26, 1 (2019), 665–675.

[166] Pier Luca Lanzi and Daniele Loiacono. 2023. Chatgpt and other large language models as evolutionary engines for online interactive collaborative game design. *arXiv preprint arXiv:2303.02155* (2023).

[167] Shahid Latif, Shivam Agarwal, Simon Gottschalk, Carina Chrosch, Felix Feit, Johannes Jahn, Tobias Braun, Yanick Christian Tchenko, Elena Demidova, and Fabian Beck. 2021. Visually connecting historical figures through event knowledge graphs. In *2021 IEEE Visualization Conference (VIS)*. IEEE, 156–160.

[168] Tomas Lawton, Kazjon Grace, and Francisco J Ibarrola. 2023. 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*. 1978–1996.

[169] Tomas Lawton, Francisco J Ibarrola, Dan Ventura, and Kazjon Grace. 2023. Drawing with Reframer: Emergence and Control in Co-Creative AI. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 264–277.

[170] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

[171] Mina Lee, Percy Liang, and Qian Yang. 2022. 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*. 1–19.

[172] Younghoon Lee, Sungzoon Cho, and Jinhae Choi. 2019. Smartphone help contents re-organization considering user specification via conditional GAN. *International Journal of Human-Computer Studies* 129 (2019), 108–115.

[173] Yoonjoo Lee, John Joon Young Chung, Tae Soo Kim, Jean Y Song, and Juho Kim. 2022. Promptiverse: Scalable generation of scaffolding prompts through human-AI hybrid knowledge graph annotation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[174] Florian Lehmann. 2023. Mixed-Initiative Interaction with Computational Generative Systems. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–6.

[175] Florian Lehmann, Niklas Markert, Hai Dang, and Daniel Buschek. 2022. Suggestion lists vs. continuous generation: Interaction design for writing with generative models on mobile devices affect text length, wording and perceived authorship. In *Proceedings of Mensch und Computer 2022*. 192–208.

[176] Juho Leinonen, Paul Denny, Stephen MacNeil, Sami Sarsa, Seth Bernstein, Joanne Kim, Andrew Tran, and Arto Hellas. 2023. Comparing code explanations created by students and large language models. *arXiv preprint arXiv:2304.03938* (2023).

[177] Juho Leinonen, Arto Hellas, Sami Sarsa, Brent Reeves, Paul Denny, James Prather, and Brett A Becker. 2023. Using large language models to enhance programming error messages. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*. 563–569.

[178] Adam D Lelkes, Vinh Q Tran, and Cong Yu. 2021. Quiz-style question generation for news stories. In *Proceedings of the Web Conference 2021*. 2501–2511.

[179] Chen Li, Dimitrios Chrysostomou, and Hongji Yang. 2022. A Natural Language-enabled Virtual Assistant for Human-Robot Interaction in Industrial Environments. In *2022 IEEE 22nd International Conference on Software Quality, Reliability, and Security Companion (QRS-C)*. IEEE, 673–678.

[180] Jianan Li, Jimei Yang, Jianming Zhang, Chang Liu, Christina Wang, and Tingfa Xu. 2020. Attribute-conditioned layout gan for automatic graphic design. *IEEE Transactions on Visualization and Computer Graphics* 27, 10 (2020), 4039–4048.

[181] Kunpeng Li, Chen Fang, Zhaowen Wang, Seokhwan Kim, Hailin Jin, and Yun Fu. 2020. Screencast tutorial video understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12526–12535.

[182] Nanxing Li, Bei Liu, Zhizhong Han, Yu-Shen Liu, and Jianlong Fu. 2019. Emotion reinforced visual storytelling. In *Proceedings of the 2019 on International Conference on Multimedia Retrieval*. 297–305.

[183] Shuyu Li and Yunsick Sung. 2021. INCO-GAN: variable-length music generation method based on inception model-based conditional GAN. *Mathematics* 9, 4 (2021), 387.

[184] Yanghao Li, Haoqi Fan, Ronghang Hu, Christoph Feichtenhofer, and Kaiming He. 2023. Scaling language-image pre-training via masking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 23390–23400.

[185] Zhiyu Li, Yanfang Chen, Xuan Zhang, and Xun Liang. 2023. BookGPT: A General Framework for Book Recommendation Empowered by Large Language Model. *Electronics* 12, 22 (2023), 4654.

[186] Jinfeng Lin, Yalin Liu, Jin Guo, Jane Cleland-Huang, William Goss, Wenchuang Liu, Sugandha Lohar, Natawut Monaikul, and Alexander Rasin. 2017. Tiqi: A natural language interface for querying software project data. In *2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 973–977.

[187] Ji Lin, Richard Zhang, Frieder Ganz, Song Han, and Jun-Yan Zhu. 2021. Anycost gans for interactive image synthesis and editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14986–14996.

[188] Yuyu Lin, Jiahao Guo, Yang Chen, Cheng Yao, and Fangtian Ying. 2020. It is your turn: Collaborative ideation with a co-creative robot through sketch. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.

[189] Yu-Ching Lin, Tang-Jie Cai, Meng-Hua Yen, Chi-Chun Chen, Guo-Shing Huang, and Cheng-Yu Peng. 2023. Automatic Feeding system with High accuracy intelligent product defection function. In *2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*. IEEE, 1–5.

[190] Stephan Linzbach, Tim Tressel, Laura Kallmeyer, Stefan Dietze, and Hajira Jabeen. 2023. Decoding Prompt Syntax: Analysing its Impact on Knowledge Retrieval in Large Language Models. In *Companion Proceedings of the ACM Web Conference 2023*. 1145–1149.

[191] Bo Liu, Lejian He, Yafei Liu, Tianyao Yu, Yuejia Xiang, Li Zhu, and Weijian Ruan. 2022. Transformer-Based Multimodal Infusion Dialogue Systems. *Electronics* 11, 20 (2022), 3409.

[192] Jerry Liu, Fisher Yu, and Thomas Funkhouser. 2017. Interactive 3D modeling with a generative adversarial network. In *2017 International Conference on 3D Vision (3DV)*. IEEE, 126–134.

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

[194] Vivian Liu. 2023. Beyond Text-to-Image: Multimodal Prompts to Explore Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–6.

[195] Vivian Liu and Lydia B Chilton. 2022. Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–23.

[196] Vivian Liu, Tao Long, Nathan Raw, and Lydia Chilton. 2023. Generative Disco: Text-to-Video Generation for Music Visualization. *arXiv preprint arXiv:2304.08551* (2023).

[197] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal image generation for news illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–17.

[198] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating text-to-image AI in 3D design workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 1955–1977.

[199] Yiting Liu, Liang Li, Beichen Zhang, Shan Huang, Zheng-Jun Zha, and Qingming Huang. 2023. MaTCR: Modality-Aligned Thought Chain Reasoning for Multimodal Task-Oriented Dialogue Generation. In *Proceedings of the 31st ACM International Conference on Multimedia*. 5776–5785.

[200] Zhenhuan Liu, Jincan Deng, Liang Li, Shaofei Cai, Qianqian Xu, Shuhui Wang, and Qingming Huang. 2020. Ir-gan: Image manipulation with linguistic instruction by increment reasoning. In *Proceedings of the 28th ACM International Conference on Multimedia*. 322–330.

[201] Vadim Liventsev, Anastasiia Grishina, Aki Härmä, and Leon Moonen. 2023. Fully Autonomous Programming with Large Language Models. *arXiv preprint arXiv:2304.10423* (2023).

[202] Asier Lopez Zorrilla, M Ines Torres, Heriberto Cuayahuitl, et al. 2022. Audio Embedding-Aware Dialogue Policy Learning. *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING* 31 (2022), 525–538.

[203] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J Cai. 2020. Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[204] Shihan Lu, Mianlun Zheng, Matthew C Fontaine, Stefanos Nikolaidis, and Heather Culbertson. 2022. Preference-driven texture modeling through interactive generation and search. *IEEE Transactions on Haptics* 15, 3 (2022), 508–520.

[205] Shuyang Luo, Haoran Xie, and Kazunori Miyata. 2021. Sketch-based anime hairstyle editing with generative inpainting. In *2021 Nicograph International (NicoInt)*. IEEE, 7–14.

[206] Yuyu Luo, Nan Tang, Guoliang Li, Jiawei Tang, Chengliang Chai, and Xuedi Qin. 2021. Natural language to visualization by neural machine translation. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 217–226.

[207] Qing Lyu, Menglei Chai, Xiang Chen, and Kun Zhou. 2020. Real-time hair simulation with neural interpolation. *IEEE Transactions on Visualization and Computer Graphics* 28, 4 (2020), 1894–1905.

[208] Yiyang Ma, Huan Yang, Bei Liu, Jianlong Fu, and Jiaying Liu. 2022. Ai illustrator: Translating raw descriptions into images by prompt-based cross-modal generation. In *Proceedings of the 30th ACM International Conference on Multimedia*. 4282–4290.

[209] Stephen MacNeil, Andrew Tran, Arto Hellas, Joanne Kim, Sami Sarsa, Paul Denny, Seth Bernstein, and Juho Leinonen. 2023. Experiences from using code explanations generated by large language models in a web software development e-book. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*. 931–937.

[210] Paula Maddigan and Teo Susnjak. 2023. Chat2vis: Generating data visualisations via natural language using chatgpt, codex and gpt-3 large language models. *IEEE Access* (2023).

[211] Raquel Magalhães. 2023. How to Write Killer Prompts for Your GPT Chatbot. Landbot. https://landbot.io/blog/how-to-write-prompts-for-chat-gpt Accessed: 2023-11-26.

[212] Mehran Maghoumi, Eugene Matthew Taranta, and Joseph LaViola. 2021. DeepNAG: Deep non-adversarial gesture generation. In *26th International Conference on Intelligent User Interfaces*. 213–223.

[213] Javier Maldonado-Romo, Alberto Maldonado-Romo, and Mario Aldape-Pérez. 2022. Path Generator with Unpaired Samples Employing Generative Adversarial Networks. *Sensors* 22, 23 (2022), 9411.

[214] Gianfranco Mauro, Mateusz Chmurski, Lorenzo Servadei, Manuel Pegalajar-Cuellar, and Diego P Morales-Santos. 2022. Few-shot user-definable radar-based hand gesture recognition at the edge. *IEEE Access* 10 (2022), 29741–29759.

[215] Silvan Mertes, Marcel Strobl, Ruben Schlagowski, and Elisabeth André. 2023. ASMRcade: Interactive Audio Triggers for an Autonomous Sensory Meridian Response. In *Proceedings of the 25th International Conference on Multimodal Interaction*. 70–78.

[216] Elzė Sigutė Mikalonytė and Markus Kneer. 2022. Can artificial intelligence make art?: Folk intuitions as to whether AI-driven robots can be viewed as artists and produce art. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 4 (2022), 1–19.

[217] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[218] Piotr Mirowski, Kory W Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–34.

[219] Meredith Ringel Morris, Carrie J Cai, Jess Holbrook, Chinmay Kulkarni, and Michael Terry. 2023. The design space of generative models. *arXiv preprint arXiv:2304.10547* (2023).

[220] Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin LC Guo. 2022. GANSpiration: Balancing Targeted and Serendipitous Inspiration in User Interface Design with Style-Based Generative Adversarial Network. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[221] Michael Muller, Lydia B Chilton, Anna Kantosalo, Q Vera Liao, Mary Lou Maher, Charles Patrick Martin, and Greg Walsh. 2023. GenAICHI 2023: Generative AI and HCI at CHI 2023. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.

[222] Michael Muller, Lydia B Chilton, Anna Kantosalo, Charles Patrick Martin, and Greg Walsh. 2022. GenAICHI: generative AI and HCI. In *CHI conference on human factors in computing systems extended abstracts*. 1–7.

[223] Vitobha Munigala, Abhijit Mishra, Srikanth G Tamilselvam, Shreya Khare, Riddhiman Dasgupta, and Anush Sankaran. 2018. Persuaide! an adaptive persuasive text generation system for fashion domain. In *Companion Proceedings of the The Web Conference 2018*. 335–342.

[224] Shabnam Najafian, Geoff Musick, Bart Knijnenburg, and Nava Tintarev. 2023. How do people make decisions in disclosing personal information in tourism group recommendations in competitive versus cooperative conditions? *User Modeling and User-Adapted Interaction* (2023), 1–33.

[225] Kizashi Nakano, Daichi Horita, Nobuchika Sakata, Kiyoshi Kiyokawa, Keiji Yanai, and Takuji Narumi. 2019. Deep-Taste: Augmented reality gustatory manipulation with GAN-based real-time food-to-food translation. In *2019 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE, 212–223.

[226] Kizashi Nakano, Daichi Horita, Nobuchika Sakata, Kiyoshi Kiyokawa, Keiji Yanai, and Takuji Narumi. 2019. Enchanting your noodles: GAN-based real-time food-to-food translation and its impact on vision-induced gustatory manipulation. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1096–1097.

[227] Shu Naritomi, Gibran Benitez-Garcia, and Keiji Yanai. 2021. Ketchup As You Like: Drawing Editor for Foods. In *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 207–209.

[228] Shu Naritomi, Ryosuke Tanno, Takumi Ege, and Keiji Yanai. 2018. FoodChangeLens: CNN-based food transformation on HoloLens. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 197–199.

[229] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. 2022. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. arXiv:2112.10741 [cs.CV]

[230] ARM Nizzad and Samantha Thelijjagoda. 2022. Designing of a Voice-Based Programming IDE for Source Code Generation: A Machine Learning Approach. In *2022 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, Vol. 5. IEEE, 14–21.

[231] Javier Díaz Noci. 2023. Merging or plagiarizing? The role of originality and derivative works in AI-aimed news production. *Hipertext. net* 26 (2023), 69–76.

[232] Donald A Norman. 1988. *The psychology of everyday things.* Basic books. 188 pages.

[233] Ariel Noyman and Kent Larson. 2020. Deepscope: HCI platform for generative cityscape visualization. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–9.

[234] Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. 2015. Designscape: Design with interactive layout suggestions. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 1221–1224.

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

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

[237] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[238] OpenAI. 2021. ChatGPT based on the GPT-4 architecture. https://openai.com. Accessed: 2023-09-09.

[239] Hiroyuki Osone, Jun-Li Lu, and Yoichi Ochiai. 2021. 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*. 1–10.

[240] Aadarsh Padiyath and Brian Magerko. 2021. desAIner: Exploring the Use of" Bad" Generative Adversarial Networks in the Ideation Process of Fashion Design. In *Creativity and Cognition*. 1–3.

[241] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372 (2021). https://doi.org/10.1136/bmj.n71

[242] Xingang Pan, Ayush Tewari, Thomas Leimkühler, Lingjie Liu, Abhimitra Meka, and Christian Theobalt. 2023. Drag your gan: Interactive point-based manipulation on the generative image manifold. In *ACM SIGGRAPH 2023 Conference Proceedings*. 1–11.

[243] Cheonbok Park, Inyoup Na, Yongjang Jo, Sungbok Shin, Jaehyo Yoo, Bum Chul Kwon, Jian Zhao, Hyungjong Noh, Yeonsoo Lee, and Jaegul Choo. 2019. Sanvis: Visual analytics for understanding self-attention networks. In *2019 IEEE Visualization Conference (VIS)*. IEEE, 146–150.

[244] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442* (2023).

[245] Joon Sung Park, Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2022. Social simulacra: Creating populated prototypes for social computing systems. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–18.

[246] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. Anglekindling: Supporting journalistic angle ideation with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[247] A Ponmalar, K Kausalya, GKR Bhavana, S Aarthi, D Gokulakrishnan, and A Anand Jose. 2022. Implementing Virtual Personal Assistant Through Artificial Intelligence Requirements. In *2022 International Conference on Computer, Power and Communications (ICCPC)*. IEEE, 552–556.

[248] Matei-Dan Popovici. 2023. ChatGPT in the Classroom. Exploring Its Potential and Limitations in a Functional Programming Course. *International Journal of Human–Computer Interaction* (2023), 1–12.

[249] James Prather, Brent N Reeves, Paul Denny, Brett A Becker, Juho Leinonen, Andrew Luxton-Reilly, Garrett Powell, James Finnie-Ansley, and Eddie Antonio Santos. 2023. " It's Weird That it Knows What I Want": Usability and Interactions with Copilot for Novice Programmers. *arXiv preprint arXiv:2304.02491* (2023).

[250] Philipp Probst, Anne-Laure Boulesteix, and Bernd Bischl. 2019. Tunability: Importance of hyperparameters of machine learning algorithms. *The Journal of Machine Learning Research* 20, 1 (2019), 1934–1965.

[251] Ilham Qasse, Shailesh Mishra, Björn Þór Jónsson, Foutse Khomh, and Mohammad Hamdaqa. 2023. Chat2Code: A Chatbot for Model Specification and Code Generation, The Case of Smart Contracts. In *2023 IEEE International Conference on Software Services Engineering (SSE)*. IEEE, 50–60.

[252] Han Qiao, Vivian Liu, and Lydia Chilton. 2022. Initial images: using image prompts to improve subject representation in multimodal ai generated art. In *Proceedings of the 14th Conference on Creativity and Cognition*. 15–28.

[253] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020 [cs.CV]

[254] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training. (2018).

[255] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[256] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. arXiv:1910.10683 [cs.LG]

[257] Janet Rafner, Steven Langsford, Arthur Hjorth, Miroslav Gajdacz, Lotte Philipsen, Sebastian Risi, Joel Simon, and Jacob Sherson. 2021. Utopian or Dystopian?: using a ML-assisted image generation game to empower the general public to envision the future. In *Creativity and Cognition*. 1–5.

[258] Nina Rajcic and Jon McCormack. 2020. Mirror ritual: An affective interface for emotional self-reflection. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[259] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* (2022).

[260] Allen Z Ren, Bharat Govil, Tsung-Yen Yang, Karthik R Narasimhan, and Anirudha Majumdar. 2023. Leveraging language for accelerated learning of tool manipulation. In *Conference on Robot Learning*. PMLR, 1531–1541.

[261] Reid Robinson. 2023. How to write an effective GPT-3 or GPT-4 prompt. Zapier. https://zapier.com/blog/gpt-prompt/ Accessed: 2023-11-26.

[262] Melissa Roemmele. 2021. Inspiration through observation: Demonstrating the influence of automatically generated text on creative writing. *arXiv preprint arXiv:2107.04007* (2021).

[263] R Rohit, SA Gandheesh, KS Suriya, and Peeta Basa Pati. 2023. System for Enhancing Accuracy of Noisy Text using Deep Network Language Models. In *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*. IEEE, 1–7.

[264] Mahdin Rohmatillah and Jen-Tzung Chien. 2023. Hierarchical Reinforcement Learning With Guidance for Multi-Domain Dialogue Policy. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 31 (2023), 748–761.

[265] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[266] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695.

[267] Andrew Ross, Nina Chen, Elisa Zhao Hang, Elena L Glassman, and Finale Doshi-Velez. 2021. Evaluating the interpretability of generative models by interactive reconstruction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[268] Steven I Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D Weisz. 2023. 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*. 491–514.

[269] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 36479–36494. https://proceedings.neurips.cc/paper_files/paper/2022/file/ec795aeadae0b7d230fa35cbaf04c041-Paper-Conference.pdf

[270] Aneeshan Sain, Ayan Kumar Bhunia, Yongxin Yang, Tao Xiang, and Yi-Zhe Song. 2021. Stylemeup: Towards style-agnostic sketch-based image retrieval. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8504–8513.

[271] Abraar Raza Samar, Bostan Khan, and Adeel Mumtaz. 2022. Context-Based Narrative Generation Transformer (NGen-Transformer). In *2022 19th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*. IEEE, 256–261.

[272] Téo Sanchez. 2023. Examining the Text-to-Image Community of Practice: Why and How do People Prompt Generative AIs?. In *Proceedings of the 15th Conference on Creativity and Cognition*. 43–61.

[273] Advait Sarkar. 2023. Exploring Perspectives on the Impact of Artificial Intelligence on the Creativity of Knowledge Work: Beyond Mechanised Plagiarism and Stochastic Parrots. (2023).

[274] Othman Sbai, Mohamed Elhoseiny, Antoine Bordes, Yann LeCun, and Camille Couprie. 2018. Design: Design inspiration from generative networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 0–0.

[275] Jacob Schrum, Jake Gutierrez, Vanessa Volz, Jialin Liu, Simon Lucas, and Sebastian Risi. 2020. Interactive evolution and exploration within latent level-design space of generative adversarial networks. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*. 148–156.

[276] Yuefan Shen, Changgeng Zhang, Hongbo Fu, Kun Zhou, and Youyi Zheng. 2020. Deepsketchhair: Deep sketch-based 3d hair modeling. *IEEE transactions on visualization and computer graphics* 27, 7 (2020), 3250–3263.

[277] Seitaro Shinagawa, Koichiro Yoshino, Seyed Hossein Alavi, Kallirroi Georgila, David Traum, Sakriani Sakti, and Satoshi Nakamura. 2020. An Interactive Image Editing System Using an Uncertainty-Based Confirmation Strategy. *IEEE Access* 8 (2020), 98471–98480.

[278] Muhammad Ayaz Shirazi, Riaz Uddin, and Min-Young Kim. 2021. Supervised Learning Based Peripheral Vision System for Immersive Visual Experiences for Extended Display. *Applied Sciences* 11, 11 (2021), 4726.

[279] Abdulhadi Shoufan. 2023. Exploring Students' Perceptions of CHATGPT: Thematic Analysis and Follow-Up Survey. *IEEE Access* (2023).

[280] M Siddharth and R Aarthi. 2023. Blended multi-class text to image synthesis GANs with RoBerTa and Mask R-CNN. *Procedia Computer Science* 218 (2023), 845–857.

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

[282] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L Glassman. 2022. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Transactions on Computer-Human Interaction* (2022).

[283] Li Siyao, Weijiang Yu, Tianpei Gu, Chunze Lin, Quan Wang, Chen Qian, Chen Change Loy, and Ziwei Liu. 2022. Bailando: 3d dance generation by actor-critic gpt with choreographic memory. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 11050–11059.

[284] Michiel Spape, Keith Davis, Lauri Kangassalo, Niklas Ravaja, Zania Sovijarvi-Spape, and Tuukka Ruotsalo. 2021. Brain-computer interface for generating personally attractive images. *IEEE Transactions on Affective Computing* (2021).

[285] Yolande Strengers, Lizhen Qu, Qiongkai Xu, and Jarrod Knibbe. 2020. Adhering, steering, and queering: Treatment of gender in natural language generation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–14.

[286] Hendrik Strobelt, Albert Webson, Victor Sanh, Benjamin Hoover, Johanna Beyer, Hanspeter Pfister, and Alexander M Rush. 2022. Interactive and visual prompt engineering for ad-hoc task adaptation with large language models. *IEEE transactions on visualization and computer graphics* 29, 1 (2022), 1146–1156.

[287] Minhyang Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 2021. AI as social glue: uncovering the roles of deep generative AI during social music composition. In *Proceedings of the 2021 CHI conference on human factors in computing systems.* 1–11.

[288] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. *arXiv preprint arXiv:2305.11483* (2023).

[289] Jiao Sun, Q Vera Liao, Michael Muller, Mayank Agarwal, Stephanie Houde, Kartik Talamadupula, and Justin D Weisz. 2022. Investigating explainability of generative AI for code through scenario-based design. In *27th International Conference on Intelligent User Interfaces.* 212–228.

[290] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems* 27 (2014).

[291] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. 2022. Augmented reality and robotics: A survey and taxonomy for ar-enhanced human-robot interaction and robotic interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* 1–33.

[292] Ben Swanson, Kory Mathewson, Ben Pietrzak, Sherol Chen, and Monica Dinalescu. 2021. Story centaur: Large language model few shot learning as a creative writing tool. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations.* 244–256.

[293] Idan Szpektor, Deborah Cohen, Gal Elidan, Michael Fink, Avinatan Hassidim, Orgad Keller, Sayali Kulkarni, Eran Ofek, Sagie Pudinsky, Asaf Revach, et al. 2020. Dynamic composition for conversational domain exploration. In *Proceedings of The Web Conference 2020.* 872–883.

[294] Yuki Tanaka and Seiichiro Katsura. 2023. A Voice-Controlled Motion Reproduction Using Large Language Models for Polishing Robots. In *2023 IEEE International Conference on Mechatronics (ICM).* IEEE, 1–6.

[295] Paul Todorov. 2019. A game of dice: Machine learning and the question concerning art. *arXiv preprint arXiv:1904.01957* (2019).

[296] Anh Truong, Peggy Chi, David Salesin, Irfan Essa, and Maneesh Agrawala. 2021. Automatic generation of two-level hierarchical tutorials from instructional makeup videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–16.

[297] Jonathan Tseng, Rodrigo Castellon, and Karen Liu. 2023. Edge: Editable dance generation from music. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 448–458.

[298] Yamei Tu, Olga Li, Junpeng Wang, Han-Wei Shen, Przemek Powałko, Irina Tomescu-Dubrow, Kazimierz M Slomczynski, Spyros Blanas, and J Craig Jenkins. 2023. SDRQuerier: A Visual Querying Framework for Cross-National Survey Data Recycling. *IEEE Transactions on Visualization and Computer Graphics* (2023).

[299] Robert Twomey. 2022. Three stage drawing transfer: collaborative drawing between a generative adversarial network, co-robotic arm, and five-year-old child. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 5, 4 (2022), 1–7.

[300] Takahisa Uchida, Hiroshi Ishiguro, and Peter Ford Dominey. 2020. Improving quality of life with a narrative robot companion: II–Creating group cohesion via shared narrative experience. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 906–913.

[301] Michihiko Ueno and Shin'ichi Satoh. 2021. Continuous and Gradual Style Changes of Graphic Designs with Generative Model. In *26th International Conference on Intelligent User Interfaces*. 280–289.

[302] Afamefuna P Umejiaku, Prastab Dhakal, and Victor S Sheng. 2023. Balancing Password Security and User Convenience: Exploring the Potential of Prompt Models for Password Generation. *Electronics* 12, 10 (2023), 2159.

[303] Priyan Vaithilingam, Tianyi Zhang, and Elena L Glassman. 2022. Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models. In *Chi conference on human factors in computing systems extended abstracts*. 1–7.

[304] Stephanie Valencia, Richard Cave, Krystal Kallarackal, Katie Seaver, Michael Terry, and Shaun K Kane. 2023. "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*. 1–14.

[305] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[306] Vanessa Volz, Jacob Schrum, Jialin Liu, Simon M Lucas, Adam Smith, and Sebastian Risi. 2018. Evolving mario levels in the latent space of a deep convolutional generative adversarial network. In *Proceedings of the genetic and evolutionary computation conference*. 221–228.

[307] Hendric Voß and Stefan Kopp. 2023. AQ-GT: a Temporally Aligned and Quantized GRU-Transformer for Co-Speech Gesture Synthesis. *arXiv preprint arXiv:2305.01241* (2023).

[308] Qian Wan and Zhicong Lu. 2023. GANCollage: A GAN-Driven Digital Mood Board to Facilitate Ideation in Creativity Support. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 136–146.

[309] April Yi Wang, Dakuo Wang, Jaimie Drozdal, Michael Muller, Soya Park, Justin D Weisz, Xuye Liu, Lingfei Wu, and Casey Dugan. 2022. Documentation matters: Human-centered AI system to assist data science code documentation in computational notebooks. *ACM Transactions on Computer-Human Interaction* 29, 2 (2022), 1–33.

[310] Bryan Wang, Meng Yu Yang, and Tovi Grossman. 2021. Soloist: Generating mixed-initiative tutorials from existing guitar instructional videos through audio processing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[311] Fengjie Wang, Xuye Liu, Oujing Liu, Ali Neshati, Tengfei Ma, Min Zhu, and Jian Zhao. 2023. Slide4N: Creating Presentation Slides from Computational Notebooks with Human-AI Collaboration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[312] Ker-Jiun Wang, Caroline Yan Zheng, Mohammad Shidujaman, Maitreyee Wairagkar, and Mariana von Mohr. 2020. Jean Joseph v2. 0 (REmotion): make remote emotion touchable, seeable and thinkable by direct brain-to-brain telepathy neurohaptic interface empowered by generative adversarial network. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 3488–3493.

[313] Ker-Jiun Wang, Caroline Yan Zheng, Maitreyee Wairagkar, and Mariana von Mohr. 2020. Jean Joseph ZERO: Let Memories Come to Life with Music. In *2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*. IEEE, 1–2.

[314] Sitong Wang, Savvas Petridis, Taeahn Kwon, Xiaojuan Ma, and Lydia B Chilton. 2023. PopBlends: Strategies for conceptual blending with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[315] Shidong Wang, Wei Zeng, Xi Chen, Yu Ye, Yu Qiao, and Chi-Wing Fu. 2021. ActFloor-GAN: Activity-Guided Adversarial Networks for Human-Centric Floorplan Design. *IEEE Transactions on Visualization and Computer Graphics* (2021).

[316] Yujia Wang, Wei Liang, Haikun Huang, Yongqi Zhang, Dingzeyu Li, and Lap-Fai Yu. 2021. Toward automatic audio description generation for accessible videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[317] Yunlong Wang, Shuyuan Shen, and Brian Y Lim. 2023. RePrompt: Automatic Prompt Editing to Refine AI-Generative Art Towards Precise Expressions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–29.

[318] Jialiang Wei, Anne-Lise Courbis, Thomas Lambolais, Binbin Xu, Pierre Louis Bernard, and Gérard Dray. 2023. Boosting GUI Prototyping with Diffusion Models. *arXiv preprint arXiv:2306.06233* (2023).

[319] Lingyu Wei, Liwen Hu, Vladimir Kim, Ersin Yumer, and Hao Li. 2018. Real-time hair rendering using sequential adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 99–116.

[320] Justin D Weisz, Michael Muller, Jessica He, and Stephanie Houde. 2023. Toward general design principles for generative AI applications. *arXiv preprint arXiv:2301.05578* (2023).

[321] Justin D Weisz, Michael Muller, Stephanie Houde, John Richards, Steven I Ross, Fernando Martinez, Mayank Agarwal, and Kartik Talamadupula. 2021. Perfection not required? Human-AI partnerships in code translation. In *26th International Conference on Intelligent User Interfaces*. 402–412.

[322] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C Schmidt. 2023. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382* (2023).

[323] Di Wu, Zhiwang Yu, Nan Ma, Jianan Jiang, Yuetian Wang, Guixiang Zhou, Hanhui Deng, and Yi Li. 2023. StyleMe: Towards Intelligent Fashion Generation with Designer Style. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[324] Hui Wu, Yupeng Gao, Xiaoxiao Guo, Ziad Al-Halah, Steven Rennie, Kristen Grauman, and Rogerio Feris. 2021. Fashion iq: A new dataset towards retrieving images by natural language feedback. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*. 11307–11317.

[325] Jian-Lin Wu, Pei-Chen Chang, Chao Wang, and Ko-Chih Wang. 2023. ATICVis: A Visual Analytics System for Asymmetric Transformer Models Interpretation and Comparison. *Applied Sciences* 13, 3 (2023), 1595.

[326] Qi Wu, Peng Wang, Chunhua Shen, Ian Reid, and Anton Van Den Hengel. 2018. Are you talking to me? reasoned visual dialog generation through adversarial learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6106–6115.

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

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

[329] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–22.

[330] Jungang Xu, Hui Li, and Shilong Zhou. 2015. An overview of deep generative models. *IETE Technical Review* 32, 2 (2015), 131–139.

[331] Pei Xu and Ioannis Karamouzas. 2021. A GAN-Like Approach for Physics-Based Imitation Learning and Interactive Character Control. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 4, 3 (2021), 1–22.

[332] Tao Xu, Xu Wang, Jiabao Wang, and Yun Zhou. 2021. From textbook to teacher: an adaptive intelligent tutoring system based on BCI. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 7621–7624.

[333] Kanta Yamaoka, Ko Watanabe, Koichi Kise, Andreas Dengel, and Shoya Ishimaru. 2022. Experience is the Best Teacher: Personalized Vocabulary Building Within the Context of Instagram Posts and Sentences from GPT-3. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*. 313–316.

[334] Takato Yamazaki, Katsumasa Yoshikawa, Toshiki Kawamoto, Tomoya Mizumoto, Masaya Ohagi, and Toshinori Sato. 2023. Building a hospitable and reliable dialogue system for android robots: a scenario-based approach with large language models. *Advanced Robotics* 37, 21 (2023), 1364–1381.

[335] Zihan Yan, Chunxu Yang, Qihao Liang, and Xiang'Anthony' Chen. 2023. XCreation: A Graph-based Crossmodal Generative Creativity Support Tool. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[336] Chaojie Yang, Hanhui Li, Shengjie Wu, Shengkai Zhang, Haonan Yan, Nianhong Jiao, Jie Tang, Runnan Zhou, Xiaodan Liang, and Tianxiang Zheng. 2022. BodyGAN: general-purpose controllable neural human body generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7733–7742.

[337] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and Ray LC. 2022. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the ACM IUI Workshops*, Vol. 10. CEUR-WS Team.

[338] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.

[339] Yuting Yang, Wenqiang Lei, Pei Huang, Juan Cao, Jintao Li, and Tat-Seng Chua. 2023. A Dual Prompt Learning Framework for Few-Shot Dialogue State Tracking. In *Proceedings of the ACM Web Conference 2023*. 1468–1477.

[340] Ning Ye, Ariel Fuxman, Vivek Ramavajjala, Sergey Nazarov, J Patrick McGregor, and Sujith Ravi. 2018. Photoreply: Automatically suggesting conversational responses to photos. In *Proceedings of the 2018 World Wide Web Conference*. 1893–1899.

[341] Yang Ye, Hengxu You, and Jing Du. 2023. Improved trust in human-robot collaboration with ChatGPT. *IEEE Access* (2023).

[342] Jung Eun Yoo, Kwanggyoon Seo, Sanghun Park, Jaedong Kim, Dawon Lee, and Junyong Noh. 2021. Virtual camera layout generation using a reference video. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.

[343] Han Yu, Aiping Li, Rong Jiang, Yan Jia, Xiaojuan Zhao, and Weihong Han. 2019. HDGS: A Hybrid Dialogue Generation System using Adversarial Learning. In *2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC)*. IEEE, 135–141.

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

[345] Paulina Yurman and Anuradha Venugopal Reddy. 2022. Drawing Conversations Mediated by AI. In *Proceedings of the 14th Conference on Creativity and Cognition*. 56–70.

[346] JD Zamfirescu-Pereira, Heather Wei, Amy Xiao, Kitty Gu, Grace Jung, Matthew G Lee, Bjoern Hartmann, and Qian Yang. 2023. Herding AI Cats: Lessons from Designing a Chatbot by Prompting GPT-3. (2023).

[347] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.

[348] Matteo Zaramella, Irene Amerini, and Paolo Russo. 2023. Why Don't You Speak?: A Smartphone Application to Engage Museum Visitors Through Deepfakes Creation. In *Proceedings of the 5th Workshop on analySis, Understanding and proMotion of heritAge Contents*. 29–37.

[349] Rachael Zehrung, Astha Singhal, Michael Correll, and Leilani Battle. 2021. Vis ex machina: An analysis of trust in human versus algorithmically generated visualization recommendations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[350] Yanhong Zeng, Jianlong Fu, Hongyang Chao, and Baining Guo. 2022. Aggregated contextual transformations for high-resolution image inpainting. *IEEE Transactions on Visualization and Computer Graphics* (2022).

[351] Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, and In So Kweon. 2023. Text-to-image diffusion model in generative ai: A survey. *arXiv preprint arXiv:2303.07909* (2023).

[352] Enhao Zhang and Nikola Banovic. 2021. Method for exploring generative adversarial networks (gans) via automatically generated image galleries. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[353] Wenbo Zhang. 2020. Sketch-To-Color Image with GANs. In *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*. IEEE, 322–325.

[354] Zhe Zhang, Wei Chai, and Jiankun Wang. 2023. Mani-GPT: A Generative Model for Interactive Robotic Manipulation. *Procedia Computer Science* 226 (2023), 149–156.

[355] Zhenyu Zhang, Bowen Yu, Haiyang Yu, Tingwen Liu, Cheng Fu, Jingyang Li, Chengguang Tang, Jian Sun, and Yongbin Li. 2022. Layout-aware information extraction for document-grounded dialogue: Dataset, method and demonstration. In *Proceedings of the 30th ACM International Conference on Multimedia*. 7252–7260.

[356] Nanxuan Zhao, Nam Wook Kim, Laura Mariah Herman, Hanspeter Pfister, Rynson WH Lau, Jose Echevarria, and Zoya Bylinskii. 2020. Iconate: Automatic compound icon generation and ideation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[357] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* (2023).

[358] Zhenjie Zhao and Xiaojuan Ma. 2018. A compensation method of two-stage image generation for human-ai collaborated in-situ fashion design in augmented reality environment. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 76–83.

[359] Chengbo Zheng, Dakuo Wang, April Yi Wang, and Xiaojuan Ma. 2022. Telling stories from computational notebooks: Ai-assisted presentation slides creation for presenting data science work. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–20.

[360] Qinkai Zheng, Xiao Xia, Xu Zou, Yuxiao Dong, Shan Wang, Yufei Xue, Lei Shen, Zihan Wang, Andi Wang, Yang Li, et al. 2023. Codegeex: A pre-trained model for code generation with multilingual benchmarking on humaneval-x. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5673–5684.

[361] Xinru Zheng, Xiaotian Qiao, Ying Cao, and Rynson WH Lau. 2019. Content-aware generative modeling of graphic design layouts. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–15.

[362] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic lies: Understanding ai-generated misinformation and evaluating algorithmic and human solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[363] Le Zhou, Qiu-Feng Wang, Kaizhu Huang, and Cheng-Hung Lo. 2019. An interactive and generative approach for chinese shanshui painting document. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 819–824.

[364] Yijun Zhou, Yuki Koyama, Masataka Goto, and Takeo Igarashi. 2021. Interactive exploration-exploitation balancing for generative melody composition. In *26th International Conference on Intelligent User Interfaces*. 43–47.

[365] Yutong Zhou and Nobutaka Shimada. 2023. Vision+ Language Applications: A Survey. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 826–842.

[366] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*. 2223–2232.

[367] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. 2019. Dm-gan: Dynamic memory generative adversarial networks for text-to-image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5802–5810.

[368] Daniel Zimmermann and Anne Koziolek. 2023. Automating GUI-based Software Testing with GPT-3. In *2023 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*. IEEE, 62–65.

[369] Daniel Zimmermann and Anne Koziolek. 2023. GUI-Based Software Testing: An Automated Approach Using GPT-4 and Selenium WebDriver. In *2023 38th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*. IEEE, 171–174.

[370] Xu Zou, Da Yin, Qingyang Zhong, Hongxia Yang, Zhilin Yang, and Jie Tang. 2021. Controllable generation from pre-trained language models via inverse prompting. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2450–2460.

Table 3. Appendix Table: Keywords for locating papers about GenAI models

"Generative AI" OR "Generative Artificial Intelligence" OR "Deep Generative Modelling" OR "Large language models" OR "LLMs" OR "LLM" OR "Generative Deep Neural Network" OR "Deep Generative Models" OR "Generative Deep Neural Networks" OR "Generative Deep Neural Network" OR "Large Vision Models"

Table 4. Appendix Table: Keywords used to search for relevant papers of GenAI systems that are interacted with by the users.

ACM
("Generative AI" OR "Generative Artificial Intelligence" OR "GAN" OR "Generative Adversarial Network" OR "Diffusion Model" OR "language model" OR "Large Language Model "OR "Deep Generative Model" OR "Prompt" OR "LLM" OR "Generative Deep Neural Network" OR "Transformer" OR "GPT" OR "BART" OR "Dalle" OR "CLIP" OR "Codex" OR "Copilot" OR "VAE" OR"Variational Autoencoder" OR "LLama" OR "BLIP" OR "Flamingo" OR "StarCoder" OR "Code T5" OR "Imagen" OR "GLIDE" OR "BERT" OR "Sparrow") AND("User" OR "Human")

IEEE
("Abstract": "Generative AI" OR "Abstract":"Generative Artificial Intelligence" OR "Abstract":"GAN" OR "Abstract":"Generative Adversarial Network" OR "Abstract":"Diffusion Model" OR "Abstract":"language model" OR "Abstract":"Deep Generative Model" OR "Abstract":"Prompt" OR "Abstract":"LLM" OR "Abstract":"Generative Deep Neural Network" OR "Abstract": "Deep Generative Models" OR "Abstract":"Transformer" OR "Abstract":"GPT" OR "Abstract":"BART" OR "Abstract":"Dalle" OR "Abstract":"CLIP" OR "Abstract":"Codex" OR "Abstract":"Copilot" OR "Abstract":"VAE" OR "Abstract":"Variational Autoencoder" OR "Abstract":"LLama" OR "Abstract":"Flamingo" OR "Abstract":"BLIP" OR "Abstract":"StarCoder" OR "Abstract":"Code T5" OR "Abstract":"Imagen" OR "Abstract":"GLIDE" OR "Abstract":"BERT" OR "Abstract":"Sparrow") AND ("Abstract":"User" OR "Abstract":"Human")

Springer
(Generative AI OR Generative Artificial Intelligence OR GAN OR Generative Adversarial Network OR Diffusion Model OR language model OR OR Deep Generative Model OR Prompt OR LLM OR Generative Deep Neural Network OR Transformer OR GPT VAE OR Variational Autoencoder) AND (User OR Human)

ScienceDirect
("Generative AI" OR "Generative Artificial Intelligence" OR "GAN" OR "Generative Adversarial Network" OR "Diffusion Model" OR "language model" OR "Deep Generative Model" OR "Prompt" OR "LLM" OR "VAE" OR "Generative Deep Neural Network" OR "Transformer" OR "GPT" OR "BART" OR "Dalle" OR "CLIP" OR "Codex" OR "Copilot" OR "Variational Autoencoder" OR "LLama" OR "Flamingo" OR "BLIP" OR "StarCoder" OR "Code T5" OR "Imagen" OR "GLIDE" OR "BERT" OR "Sparrow") AND ("User" OR "Human")

MDPI
("Generative AI" OR "Generative Artificial Intelligence" OR "GAN" OR "Generative Adversarial Network" OR "Diffusion Model" OR "language model" OR "Deep Generative Model" OR "Prompt" OR "LLM" OR "VAE" OR "Generative Deep Neural Network" OR "Transformer" OR "GPT" OR "BART" OR "Dalle" OR "CLIP" OR "Codex" OR "Copilot" OR "Variational Autoencoder" OR "LLama" OR "Flamingo" OR "BLIP" OR "StarCoder" OR "Code T5" OR "Imagen" OR "GLIDE" OR "BERT" OR "Sparrow") AND ("User" OR "Human")

Talyor & Francis
[[Abstract: "generative ai"] OR [Abstract: "generative artificial intelligence"] OR [Abstract: "gan"] OR [Abstract "generative adversarial network"] OR [Abstract: "diffusion model"] OR [Abstract: "language model"] OR [Abstract: "deep generative model"] OR [Abstract: "prompt"] OR [Abstract: "llm"] OR [Abstract: "vae"] OR [Abstract: "generative deep neural network"] OR [Abstract: "transformer"] OR [Abstract: "gpt"] OR [Abstract: "bart"] OR [Abstract: "dalle OR [Abstract: "clip"] OR [Abstract: "codex"] OR [Abstract: "copilot"] OR [Abstract: "variational autoencoder OR [Abstract: "llama"] OR [Abstract: "flamingo"] OR [Abstract: "blip"] OR [Abstract: "starcoder"] OR [Abstract: "code t5 OR [Abstract: "imagen"] OR [Abstract: "glide"] OR [Abstract: "bert"] OR [Abstract: "sparrow"]] AND [[Abstract: "user"] OR [Abstract: "human"]] AND [Article Type: Article] AND [Language: English] AND [Publication Date: (01/01/2013 TO 12/31/2024)]

Table 5. Appendix Table: Purposes of Using GenAI

Category	Count	Citation
Augment Sample Data	17	Figure [48]: [8, 16, 38, 48, 51, 66, 74, 105, 106, 148, 155, 189, 212, 245, 275, 306, 327, 362]
Automate Processes	86	Figure [296]: [4, 13, 17, 20, 22, 26, 30, 32–34, 42, 43, 67, 69, 75, 77, 85, 88, 98, 108, 116, 119, 122, 123, 126, 131, 132, 134, 150, 150, 152, 156, 168, 169, 172, 172, 176, 176–179, 182, 186, 188, 198, 198, 201, 209, 213, 214, 230, 237, 247–249, 251, 258, 260, 260, 263, 270, 271, 282, 283, 294, 296, 300, 303, 309, 312, 313, 316, 318, 319, 331, 333–336, 340, 341, 348, 353, 368–370]
Enhance Experiences	38	Figure [278]:[7, 15, 29, 35, 36, 40, 53, 54, 87, 121, 125, 127, 128, 130, 137, 157, 158, 164, 196, 197, 202, 215, 223, 225–228, 236, 278, 285, 286, 292, 304, 321, 328, 329, 346, 347]
Explore Alternatives	36	Figure [308]:[2, 14, 23, 46, 57, 65, 73, 81–84, 90, 110, 117, 140, 153, 154, 162, 166, 171, 174, 175, 194, 218, 234, 240, 246, 262, 301, 307, 308, 314, 337, 345, 352, 361]
Refine Outcomes	61	Figure [19]:[12, 19, 39, 41, 45, 47, 50, 56, 58, 59, 71, 72, 76, 94, 99–102, 124, 129, 133, 135, 141, 142, 146, 161, 163, 180, 187, 192, 193, 200, 203–207, 220, 239, 242, 252, 257, 272, 276, 277, 284, 287, 289, 315, 317, 323, 324, 332, 342, 344, 350, 356, 358, 360, 363, 364]
Get Answers for Inquiries	22	Figure [151]:[5, 18, 60, 79, 92, 103, 111, 115, 147, 151, 167, 191, 199, 210, 264, 268, 279, 293, 326, 343, 354, 355]
Understand Subjects	26	Figure [244]:[24, 25, 55, 61, 107, 120, 138, 139, 139, 165, 173, 181, 185, 195, 233, 243, 244, 267, 288, 296, 298, 310, 311, 325, 349, 359]

Table 6. Appendix Table: Output Modalities

Category	Count	Citation
2D Visuals	93	Figure [39][75][14][342][158][107]: [2, 12–14, 19, 20, 25, 26, 39, 41, 45–47, 51, 55–59, 61, 71–73, 75, 76, 85, 100, 107, 116, 123, 131, 132, 135, 138–142, 146, 154, 157, 158, 161, 162, 168, 173, 181, 187, 189, 195–197, 200, 205, 206, 208, 210, 237, 240, 242, 243, 252, 257, 267, 270, 272, 277, 278, 280, 284, 296, 298, 299, 308, 312–314, 317, 323–325, 332, 335, 342, 345, 348–350, 352, 353, 356, 358, 363]
3d Graphic	26	Figure [198][331][226]: [34, 42, 67, 69, 77, 99, 108, 152, 163, 192, 198, 204, 207, 213, 225–228, 233, 276, 283, 307, 319, 331, 336, 342]
Audio	15	Figure [287][40][126]: [1, 40, 81, 87, 101, 122, 126, 203, 215, 236, 247, 287, 300, 334, 364]
Numerical Data	21	Figure [119]: [17, 30, 32, 48, 74, 88, 98, 119, 121, 134, 155, 164, 172, 173, 179, 212, 214, 260, 294, 331, 341]
Layout	24	Figure [306][156][102][110]: [14, 38, 83, 102, 110, 117, 129, 133, 151, 156, 165, 172, 180, 220, 234, 275, 301, 306, 310, 311, 315, 318, 359, 361]
Textual	112	Figure [50][134][147] [137][8]: [4, 5, 7, 8, 15, 16, 18, 22–24, 29, 33, 35, 36, 43, 50, 53, 54, 60, 65, 66, 79, 82, 84, 90, 92, 94, 103, 105–107, 111, 115, 120, 124–128, 130, 134, 136, 137, 147, 148, 150, 153, 166, 167, 169, 171, 174–177, 182, 185, 186, 190, 191, 193, 199, 201, 202, 209, 218, 223, 230, 236, 239, 246, 248, 249, 251, 258, 262–264, 268, 271, 279, 282, 285, 286, 288, 289, 292, 302–304, 309, 316, 321, 326–329, 333, 337, 339, 340, 343, 344, 346, 347, 354, 355, 360, 362, 368–370]

Table 7. Appendix Table: Functions of Models

Category	Count	Citation
Aggregation	33	Figure [236]: [14, 33, 35, 53, 60, 61, 79, 107, 120, 134, 150, 152, 166, 171, 173, 176, 177, 197, 209, 213, 218, 223, 236, 245, 246, 296, 309, 314, 323, 325, 340, 359, 361]
Completion	14	Figure [188]: [14, 54, 58, 75, 101, 125, 174, 175, 178, 188, 237, 249, 289, 299]
Diversification	36	Figure [117]: [8, 16, 38, 48, 55, 57, 66, 72, 73, 81, 90, 105, 106, 117, 133, 138, 139, 153, 161, 162, 165, 189, 194, 203, 234, 275, 287, 301, 302, 304, 306, 308, 315, 318, 352, 362]
Generation from Scratch	17	Figure [306]: [15, 17, 51, 73, 74, 129, 137, 142, 155, 161, 203, 212, 240, 284, 287, 306, 332]
Inter-modal Conversion	109	Figure [316]: [1, 4, 5, 13, 22, 25, 25, 26, 32, 34, 39, 40, 42, 45, 47, 47, 50, 59, 67, 69, 76, 76, 83, 85, 98, 98, 108, 110, 116, 117, 119, 121, 123, 126, 128, 132, 140, 146, 146, 147, 151, 164, 168, 169, 172, 172, 179, 181, 182, 186, 193, 195, 196, 198, 200, 201, 204, 206–208, 210, 214, 215, 215, 220, 220, 225–228, 230, 233, 236, 243, 243, 247, 248, 251, 258, 260, 263, 272, 276, 280, 283, 294, 298, 298, 300, 303, 307, 310, 311, 316, 317, 319, 324, 326, 331, 333–336, 341, 342, 348, 349, 356, 360]
Intra-modal Transformation	82	Figure [71]: [7, 12, 15, 18–20, 23, 36, 41, 56, 58, 65, 71, 77, 82, 84, 87, 94, 99, 100, 102, 103, 111, 115, 122, 124, 127, 130, 135, 137, 148, 154, 156–158, 163, 167, 180, 185, 191, 192, 192, 199, 199, 202, 239, 240, 242, 252, 257, 262, 264, 267, 268, 278, 279, 282, 285, 286, 288, 292, 293, 312, 313, 321, 328, 329, 337, 343–347, 353, 354, 354, 355, 355, 358, 364, 368, 369]

Table 8. Appendix Table: Synchronization

Category	Count	Citation
Preliminary	14	[13−15, 101, 134, 157, 157, 164, 172, 173, 278, 278, 310, 333]
Real-time	44	[1, 23, 33, 35, 36, 54, 55, 55, 56, 58, 73, 74, 79, 82, 88, 108, 121, 122, 125, 142, 161, 171, 174, 175, 215, 223, 225−227, 249, 267, 267, 282, 284, 289, 289, 294, 304, 309, 331, 332, 334, 340, 355]
Delayed	239	[2, 4, 5, 7, 8, 12, 16−20, 22, 24−26, 29, 32, 34, 38−43, 45−48, 50, 51, 53, 57, 59−61, 66, 67, 69, 71−73, 75−77, 79, 81, 83−85, 87, 90, 92, 94, 98−100, 102, 103, 105−107, 110, 111, 115−117, 119, 120, 123, 124, 126−133, 135−141, 146−148, 150−156, 158, 158, 161−163, 165−169, 172, 176−182, 185−189, 191−207, 207−210, 212−214, 218, 220, 224, 228, 230, 233, 233, 234, 236, 236, 237, 239, 240, 242, 243, 245−248, 251, 252, 257, 258, 260, 262−264, 268, 270−272, 275−277, 279, 280, 283, 285−288, 292, 293, 296, 298−304, 306−308, 310−319, 321, 323−329, 335−337, 341−343, 345, 347−350, 352−354, 356, 358−364, 368−370]

Table 9. Appendix Table: Methods to Improve the Output

Category	Count	Citation
Additional Demonstration	27	Figure [276]: [7, 39, 47, 49, 50, 67, 99, 104, 117, 152, 168, 194, 196−198, 207, 207, 215, 227, 252, 276, 315, 323, 335, 358, 361, 363]
Highlighting and In-painting	21	Figure [73]: [8, 19, 23, 36, 53, 54, 59, 72, 73, 82, 94, 100, 141, 150, 161, 171, 173, 208, 321, 344, 350]
Natural Language Commands	57	Figure [45]: [1, 5, 12, 14, 15, 18, 24, 29, 32, 43, 45, 48, 48, 65, 98, 105−107, 116, 119, 120, 126, 128, 134, 137, 140, 166, 172, 193, 195, 199−201, 210, 230, 248, 251, 260, 271, 279, 285, 288, 289, 293, 324, 327, 328, 328, 329, 336, 339, 346, 347, 354, 355, 360, 370]
Re-Initialization	89	Figure [75]: [4, 13, 17, 20, 26, 42, 50, 51, 61, 66, 69, 74−77, 85, 87, 90, 92, 103, 108, 110, 111, 115, 121−124, 131, 132, 136, 146, 148, 155, 164, 167, 179−182, 185, 186, 188−192, 202, 212−214, 223, 233, 233, 237, 243, 245, 247, 258, 263, 264, 268, 270, 280, 283, 294, 299, 300, 302, 307, 310, 312, 313, 317, 319, 325, 326, 331−334, 337, 341, 343, 345, 348, 353, 368, 369]
Option Selection	50	Figure [352]: [33, 35, 46, 56, 57, 72, 81, 83, 84, 88, 102, 125, 129, 130, 142, 147, 153, 154, 156, 162, 165, 172, 174, 175, 178, 187, 203, 204, 218, 220, 225, 226, 234, 239, 240, 246, 249, 262, 277, 282, 284, 286, 296, 308, 314, 340, 349, 352, 356, 359]
Parameter-based Tuning	28	Figure [55]: [2, 7, 8, 25, 34, 47, 55, 58, 71, 101, 127, 138, 139, 151, 163, 177, 205, 242, 257, 267, 275, 287, 292, 298, 301, 308, 318, 364]

Table 10. Appendix Table: Objects to Control

Category	Count	Citation
Latent Space	46	[2, 8, 12, 19, 34, 47, 55–59, 71–73, 84, 100, 129, 133, 138, 139, 142, 154, 161, 163, 165, 187, 192, 204, 205, 215, 220, 242, 257, 267, 275, 277, 301, 308, 312, 313, 318, 336, 344, 352, 361, 364]
Hyper-Parameters	13	[33, 101, 127, 171, 176, 177, 203, 218, 249, 262, 287, 304, 347]
Parameters/Weights	7	[25, 31, 88, 243, 284, 292, 298]
Input	200	[1, 4, 5, 7, 13–15, 17, 18, 20, 22–24, 26, 29, 32, 35, 36, 39–43, 46, 48–51, 53, 54, 60, 61, 65–67, 69, 74–77, 81, 82, 85, 87, 90, 92, 94, 98, 102–106, 108, 110, 111, 115–117, 119–124, 126–128, 130–132, 134–136, 140, 141, 146–148, 150–153, 155, 156, 162, 164, 166, 167, 169, 172, 172–175, 178–182, 185, 186, 188–191, 193–199, 202, 206–210, 212–214, 223–227, 230, 234, 237, 239, 245–248, 251, 252, 258, 260, 263, 264, 268, 270–272, 276, 279, 280, 282, 283, 285, 286, 288, 289, 293, 294, 296, 299, 300, 302, 303, 307, 310, 311, 315–317, 319, 321, 323, 326–329, 331–335, 337, 339–346, 348, 350, 353–356, 358, 360, 363, 368–370]

Table 11. Mediums to Control

Category	Count	Citation
Brain Signal	10	Figure [142]: [56, 58, 74, 76, 142, 146, 284, 312, 313, 332]
Controller	1	Figure [331]: [331]
Gesture	14	Figure [40]: [4, 17, 40, 48, 49, 69, 121, 131, 135, 155, 212–214, 258]
GUI and Widgets	207	Figure [163]: [2, 5, 12, 14, 15, 18–20, 23–25, 29, 33, 34, 36, 38, 39, 42, 43, 45–47, 51, 53–55, 57, 59–61, 65, 66, 71–73, 77, 79, 82–85, 88, 90, 92, 94, 100–105, 107, 110, 111, 115–117, 120, 123–125, 127–130, 132, 133, 136–141, 147, 148, 150, 151, 153, 154, 162, 163, 165–169, 171, 172, 172–178, 180–182, 185–187, 189–192, 194, 195, 197–203, 205–207, 209, 210, 215, 218, 220, 223, 224, 234, 236, 239, 240, 242, 243, 245–249, 251, 252, 257, 260, 262–264, 267, 268, 271, 272, 275, 277, 279, 280, 282, 283, 285–289, 292, 293, 296, 298, 301–304, 306, 308, 309, 311, 314, 316–319, 321, 323–329, 333, 335–337, 339, 340, 342–344, 346, 347, 349, 350, 352–356, 359, 360, 363, 368? –370]
Pen	19	Figure [8]: [7, 8, 41, 49, 50, 75, 99, 110, 161, 188, 227, 237, 270, 276, 299, 315, 345, 358, 361]
Tangible	9	Figure [233]: [67, 157, 158, 204, 225, 226, 228, 233, 278]
Audio Command	28	Figure [6]: [1, 6, 13, 22, 26, 32, 81, 87, 98, 108, 119, 122, 126, 134, 152, 156, 179, 193, 196, 230, 294, 300, 307, 310, 334, 341, 348, 364]

Table 12. Appendix Table: Levels Of Engagement

Category	Count	Citation
Collaborative	143	[2, 12, 14, 17, 19, 25, 29, 32–34, 36, 38–40, 43, 45–47, 49, 50, 53, 55, 57, 59, 65, 67, 71–73, 75–77, 81, 90, 92, 99–102, 106, 107, 119, 124, 127–129, 135, 136, 138, 139, 141, 146–148, 153, 161, 163, 166, 169, 171–173, 178, 180, 185, 187, 188, 190, 193–198, 200, 201, 203–208, 210, 215, 218, 220, 224–228, 237, 239, 242, 243, 246, 248, 251, 252, 257, 260, 262, 267, 270–272, 275–277, 285–289, 298, 299, 301, 302, 306, 308, 315, 318, 319, 323, 324, 327–329, 331, 335–337, 344, 346, 347, 350, 352, 356, 359, 360, 362, 363, 370]
Deterministic	100	[4, 5, 13, 15, 16, 18, 20, 22, 24, 26, 41, 42, 48, 56, 58, 61, 66, 69, 74, 83, 85, 87, 88, 103–105, 108, 110, 111, 115, 116, 120–123, 126, 130–134, 137, 140, 142, 150–152, 154, 165, 167, 168, 176, 179, 181, 182, 186, 189, 191, 202, 209, 213, 214, 230, 245, 247, 258, 263, 264, 268, 279, 280, 283, 284, 294, 296, 300, 307, 309, 311–314, 316, 317, 325, 326, 332, 334, 339, 341–343, 345, 348, 349, 353, 358, 361, 368, 369]
Assistive	31	[7, 8, 23, 54, 79, 82, 84, 94, 98, 117, 125, 156, 174, 175, 177, 192, 199, 234, 240, 249, 282, 292, 293, 303, 304, 310, 321, 340, 354, 355, 364]
Passive	13	[1, 35, 51, 155, 157, 158, 164, 172, 212, 223, 236, 278, 333]

Table 13. Appendix Table: Application Domains

Category	Count	Citation
3D Modeling	9	Figure [83]: [83, 99, 117, 152, 163, 192, 207, 276, 342]
Art and Creativity	60	Figure [237]: [2, 8, 12, 19, 20, 26, 39–42, 47, 55, 56, 58, 59, 71, 72, 75, 76, 81, 100, 101, 136, 140, 142, 146, 154, 161, 162, 168, 169, 187, 188, 196, 197, 203, 205, 208, 237, 242, 252, 257, 272, 277, 280, 283, 284, 287, 299, 308, 310, 317, 335, 345, 348, 350, 352, 353, 363, 364]
Design	34	Figure [234]: [14, 34, 46, 77, 84, 85, 102, 110, 133, 151, 156, 158, 180, 194, 195, 198, 200, 204, 220, 234, 236, 270, 296, 301, 311, 314, 315, 318, 319, 336, 349, 356, 359, 361]
Education and Learning	22	Figure [15]: [13, 15, 22, 24, 69, 74, 92, 130, 132, 138, 139, 173, 181, 185, 199, 264, 288, 332, 333, 343, 354, 369]
Fashion	8	[45, 57, 123, 131, 240, 323, 324, 358]
Game Development	8	Figure [275]: [16, 38, 129, 135, 166, 275, 306, 331]
Programming	24	Figure [309]: [79, 124, 127, 128, 147, 176, 177, 186, 193, 201, 206, 209, 230, 248, 249, 251, 268, 289, 303, 309, 321, 328, 360, 368]
Quality of Life	39	Figure [278]: [1, 4, 7, 15, 29, 35, 48, 51, 66, 67, 87, 108, 122, 126, 134, 137, 157, 172, 213–215, 223–228, 247, 258, 278, 300, 302, 307, 312, 313, 316, 340, 347, 355]
Robotics and IOT	16	Figure [6]: [6, 17, 32, 88, 107, 119–121, 172, 179, 189, 260, 294, 326, 334, 341]
Science And Research	25	Figure [244]: [5, 25, 61, 65, 73, 104–106, 141, 148, 155, 165, 202, 206, 210, 212, 243–245, 267, 298, 325, 327, 339, 362]
Writing	40	Figure [15]: [15, 23, 33, 36, 43, 49, 50, 53, 54, 82, 90, 94, 103, 111, 125, 150, 153, 167, 171, 174, 175, 178, 190, 191, 218, 239, 246, 262, 263, 271, 282, 285, 286, 292, 304, 329, 337, 344, 346, 370]

Table 14. Appendix Table: Evaluation Methodologies

Category	Count	Citation
Technical Evaluation	50	**Quantitative**: [6, 8, 19, 38, 45, 47–50, 56, 58, 65, 67, 71, 90, 99, 104, 106, 110, 117, 119, 120, 126, 133, 142, 153, 157, 171, 172, 175, 190, 200, 201, 207, 208, 220, 260, 272, 276, 278, 284, 296, 301, 306, 310, 314, 315, 317, 339, 364]
	99	**Qualitative**: [5, 6, 12, 15–17, 19, 22, 26, 32, 35, 43, 45, 47, 48, 50, 56, 65, 66, 71, 75, 76, 83, 88, 90, 92, 99, 100, 108, 110, 116, 119–121, 129, 136, 141, 146, 148, 155, 157, 175, 176, 178–181, 187, 188, 192, 199, 200, 202, 204, 206, 207, 207, 208, 212–214, 223, 230, 242, 245, 246, 260, 263, 270–272, 276, 278, 280, 283, 284, 294, 296, 298, 301, 302, 307, 312–315, 317, 319, 323, 324, 326, 336, 340, 350, 354, 355, 360, 362, 364]
Demonstration	43	[1, 25, 29, 34, 35, 39, 58, 59, 61, 69, 84, 92, 101, 103, 111, 123, 129, 140, 157, 158, 163, 167, 174, 176, 178, 189, 191, 210, 223, 230, 243, 278, 285, 317, 318, 325, 345, 346, 348, 353, 363, 368, 370]
User Evaluation	176	[4, 7, 8, 13–16, 18, 20, 23, 24, 32, 33, 36, 38–40, 42, 43, 46, 47, 49–51, 53–55, 57, 61, 66, 71–75, 77, 79, 81, 82, 85, 87, 88, 90, 94, 98, 102, 105, 107, 115, 117, 122, 124–138, 147, 150–154, 156–158, 161, 165, 166, 168, 169, 171, 173, 175, 177, 178, 180, 182, 185, 186, 188, 193–198, 201, 203–206, 209, 212, 218, 220, 224–226, 234, 236, 237, 239, 240, 245, 246, 248, 249, 251, 252, 257, 258, 262, 264, 267, 268, 275, 277–280, 282–284, 286–289, 292, 293, 298, 300, 302–304, 307–311, 314, 316, 321, 323, 326, 328, 329, 332–337, 341–344, 347, 349, 350, 352, 356, 358–360, 364, 369]

Table 15. State-of-the-Art Commonly Used GenAI Applications

Input	Output	Application	Model	Link
Text	Text	Chatbot	LLAMA	https://github.com/facebookresearch/llama
Text	Text	Chatbot	ChatGPT	https://github.com/lencx/ChatGPT
Text	Text	Chatbot	Sparrow	https://arxiv.org/pdf/2209.14375.pdf
Text	Text	Chatbot	BART	https://github.com/huggingface/transformers/tree/main/src/transformers/models/bart
Text	Image	Art	Dalle -1	https://github.com/lucidrains/DALLE-pytorch
Text	Image	Art	Dalle- 2	https://github.com/lucidrains/DALLE2-pytorch
Text	Image	Art	CLIP	https://github.com/openai/CLIP
Text	Image	Art	VisualBERT	https://github.com/uclanlp/visualbert
Text	Image	Art	GLIDE	https://github.com/bumptech/glide
Text	Image	Art	Imagen	https://imagen.research.google/
Text	Image	Art	CM3Leon	https://github.com/kyegomez/CM3Leon
Text	Image	Art	Stable Diffusion	https://github.com/CompVis/stable-diffusion
Text	Code	Programming	ChatGPT	https://github.com/lencx/ChatGPT
Text	Code	Programming	Codex	https://platform.openai.com/docs/guides/code
Text	Code	Programming	Copilot	https://github.com/github/copilot-docs
Text	Code	Programming	Code Interpreter	https://github.com/ricklamers/gpt-code-ui
Text	Code	Programming	Code T5	https://github.com/salesforce/codet5
Text	Code	Programming	Code LLama	https://github.com/facebookresearch/codellama
Text	Code	Programming	StarCoder	https://github.com/bigcode-project/starcoder
Text	Motion	Animation	MDM	https://github.com/GuyTevet/motion-diffusion-model
Text	Motion	Animation	Natural Motion	https://github.com/EricGuo5513/text-to-motion
Text	Audio	Music	AudioLDM	https://audioldm.github.io/
Text	Audio	Music	Make-An-Audio	https://github.com/Text-to-Audio/Make-An-Audio
Text	Audio	Music	AudioCraft	https://github.com/facebookresearch/audiocraft
Image	Image	Art	GAN	https://github.com/eriklindernoren/PyTorch-GAN
Image	Image	Art	StyleGAN	https://github.com/NVlabs/stylegan
Image	Image	Art	BigGAN	https://github.com/ajbrock/BigGAN-PyTorch
Image	Image	Art	CycleGAN	https://github.com/junyanz/CycleGAN
Image	Image	Art	DenoisingGAN	https://github.com/NVlabs/denoising-diffusion-gan
Image	Image	Art	VAE	https://github.com/AntixK/PyTorch-VAE
Image	Image	Art	InstaFormer	https://github.com/KU-CVLAB/InstaFormer
Image	Image	Art	Stable Diffusion	https://stablediffusionweb.com/
Image	Image	Art	Cold Diffusion	https://github.com/arpitbansal297/Cold-Diffusion-Models
Image	Image	Art	DiffusionVAE	https://github.com/kpandey008/DiffuseVAE
Image	Image	Art	CM3Leon	https://github.com/kyegomez/CM3Leon
Image	Text	Image Description	Flamigo	https://github.com/lucidrains/flamingo-pytorch
Image	Text	Image Description	BLIP	https://github.com/huggingface/transformers
Image	Text	Image Description	GROUNDING	https://github.com/kohjingyu/fromage
Image	Text	Image Description	AltCLIP	https://github.com/automatic1111/stable-diffusion-webui
Image	Text	Image Description	GIT	https://github.com/microsoft/GenerativeImage2Text
Image	Text	Image Description	M-GPT	https://github.com/open-mmlab/Multimodal-GPT
Audio	Audio	Voice Conversion	WaveNet	https://github.com/vincentherrmann/pytorch-wavenet
Audio	Video	Animation	VAE	https://github.com/NVlabs/Dancing2Music
Video	Text	Video Description	VideoBERT	https://github.com/ammesatyajit/VideoBERT
Video	Text	Video Description	CoMVT	https://google.github.io/look-before-you-speak/
Video	Text	Video Description	MVGPT	https://github.com/open-mmlab/Multimodal-GPT
Video	Text	Video Description	UniVL	https://github.com/microsoft/UniVL
Video	Text	Video Description	OpenFlamingo	https://github.com/mlfoundations/open_flamingo