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

Jingyu Shi*
shi537@purdue.edu
Purdue University
West Lafayette, Indiana, USA

Rahul Jain*
jain348@purdue.edu
Purdue University
West Lafayette, Indiana, USA

Hyungjun Doh
hdoh@purdue.edu
Purdue University
West Lafayette, Indiana, USA

Ryo Suzuki
ryo.suzuki@ucalgary.ca
University of Calgary
Calgary, Alberta, Canada

Karthik Ramani
ramani@purdue.edu
Purdue University
West Lafayette, Indiana, USA

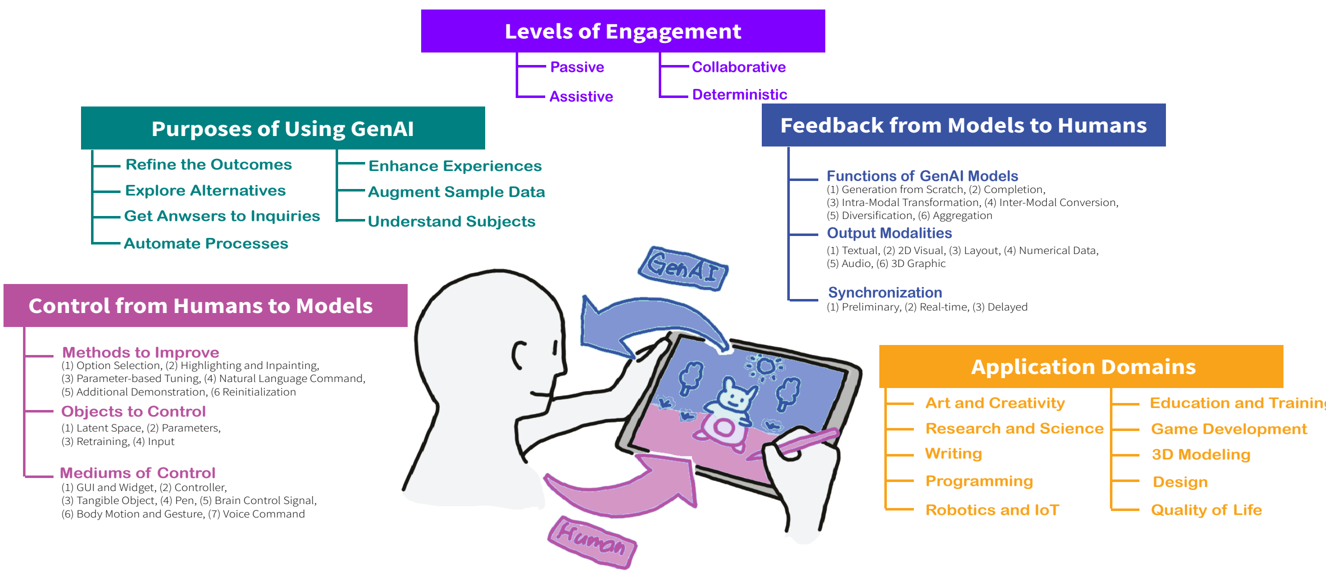

Figure 1: Visual abstract of our survey and taxonomy of Human-Generative-AI Interaction. Our taxonomy summarizes five key dimensions, namely, Purposes of Using GenAI, Feedback from Models to Humans, Control from Humans to Models, Levels of Engagement, and Application Domains.

ABSTRACT

Generative AI (GenAI) has shown remarkable capabilities in generating diverse and realistic content across different formats like images, videos, and text. In Generative AI, human involvement is essential, thus 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. In this paper, we present a survey of 154 papers, providing a novel taxonomy and analysis of Human-GenAI Interactions from both human and Gen-AI perspectives. The dimension of design space includes 1) Purposes of Using Generative AI, 2) Feedback from Models to Users , 3) Control from Users to Models, 4) Levels of Engagement, 5) Application Domains, and 6) Evaluation Strategies. Our work is also timely at the current development stage of GenAI, where the Human-GenAI interaction design is of paramount importance. We also highlight challenges and opportunities to guide the design of Gen-AI systems and interactions towards the future design of human-centered Generative AI applications.

*Both authors contributed equally to this research.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

datasets, neural networks, gaze detection, text tagging

ACM Reference Format:

Jingyu Shi, Rahul Jain, Hyungjun Doh, Ryo Suzuki, and Karthik Ramani. 2018. An HCI-Centric Survey and Taxonomy of Human-Generative-AI Interactions. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 25 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Recently, Generative Artificial Intelligence (GenAI) models have gained immense popularity and are being applied in diverse applications such as art [5, 109], design [84, 162], and entertainment [108, 176]. Current popular GenAI models including Large Language Models (LLM) and Large Visual Models (LVM) are widely deployed on platforms or in software for their capabilities to create imagery content (Dalle-2 [154], Stable Diffusion [157]), writing literature [200], and Question Answering (ChatGpt [140]). The adoption of GenAI models has demonstrated significant advantages, e.g. fostering creativity [148], driving innovation [19], enabling personalized content generation [192], and providing valuable assistance in creation endeavors [122]. 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 Gen-AI models, Human-GenAI interaction techniques such as prompt engineering [114, 148], visualization [183], and interactive interfaces [31, 44], have become popular and effective mediums for humans to interact with GenAI systems. These allow users to collaborate [200], be assisted[3], take suggestions [25] or revise recommendations [180] from GenAI systems.

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 Gen-AI 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 to.

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.

In this paper, we review a corpus of 154 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 Generative AI, 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.

A widely accepted definition [195] of GenAI goes by *the probabilistic models that model the joint distribution* in contrast to discriminate AI models *that model the conditional distribution.* 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) [41] and Long Short-Term Memory (LSTM) networks [58]. Likewise in CV, the concepts of Artificial/Convolutional Neural Networks (ANN [51] and CNN [100]) 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 [124, 172]. Similarly in the CV area, models like Generative Adversarial Networks (GAN) [55], Variational Autoencoders (VAE) [94], and their successors [12, 83–85] enabled diverse applications such as style transferring between images [52, 215], generating images based on texts [216], etc.

Recent research has highlighted and merged the two fields, enabling the multi-modal generative power of GenAI. Works like Transformer [181] and Diffusion Model [67, 134, 158, 160] have built the theoretical foundation for the current stage of GenAI,

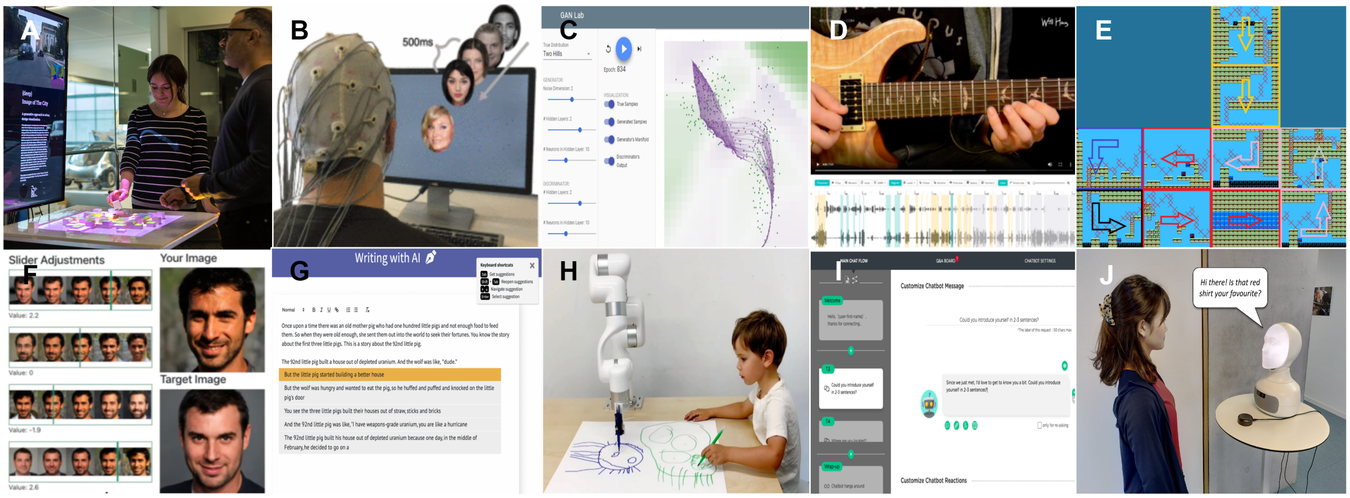

Figure 2: Examples of GenAI applications located in our survey, covering the topics of research: A) Embodied interactions with GenAI [136], B) Direct Control human to AI [167], C) Human Interpretable [81], D) Gen AI enhancing skill [185], E) Automate Process [18], F) Human controllability [31], G) Natural Language Generation [56], H) Human AI collaboration [177], I) Personalization and Adaptation [64], J) Conversational GenAI [74]

where large models such as Generative Pretrained Transformer (GPT) and its successors [14, 87, 150, 151], T5 [152], BERT [36], and CLIP [149] 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 Generative AI. In this paper, we provide a comprehensive summary of diverse GenAI applications and highlight some notable open-source platforms in Table 11 . These platforms are designed and contributed by researchers, developers, and experts, aiming to make GenAI technology accessible to a broader audience. Table 11 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 AI), GenAI systems possess the unique ability to create novel content (i.e. generative AI). 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 [37] 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 [214]. Some work has conducted study [70] and discussion [21] to gain human perspective in using GenAI models. Chen et al. [21] 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 [17]. Prior work has contributed in survey and review papers in the field of GenAI such as GenAI recent developments [208], their technical perspective [65, 205], content generated [17] and application [57, 214]. Some recent work has also proposed design space [70, 126, 191] and design guidelines [114]. Recently, GenAI and human interactions have been a topic in HCI workshops [10, 128, 129].

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 [68]. We systematically created a relevant corpus using PRISMA [143] (a systematic review strategy) guidelines: (1) Search strategy to explore; (2) Identification of the publication outlets; (3) Evidence Screening; (4) Eligibility: Inclusion and Exclusion criteria.

3.1 Search Strategy

We explore existing research on GenAI models and systems to identify keywords that can cover the full spectrum. To achieve this, we have developed two methodologies. Firstly, we developed a methodology to find out advancements in GenAI models in the Machine Learning/Deep Learning conferences. This provides keywords for the models which were further used in developing a methodology to find GenAI systems keywords in the HCI domain. The two methodologies are discussed in detail below:

3.1.1 Methodology to Define GenAI-related Model Keywords. : In our research, we conducted a thorough literature search in the proceedings of three prominent machine learning conferences (ICML, ICLR, and NeurIPS) and three computer vision conferences (CVPR, ICCV, and ECCV). Our focus was on identifying papers related to Generative AI, machine learning, and deep learning. To ensure the relevance of the research, we limited our search to papers published 15 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. Some of the examples of keywords are GAN, Transformer, Diffusion, and BERT.

3.1.2 Methodology to Define GenAI-Related Systems Keywords. : This methodology focuses on finding relevant keywords related to GenAI systems that are interacted with by the users. The process involved conducting a search in five venues: CHI, UIST, CSCW, TVCG, and TOG. Initially, the search used keywords such as "Generative AI," "GenAI," and "Generative Artificial Intelligence" to identify relevant papers. The search was limited to papers published in the last 10 years. During the search process, the title, abstract, and relevant authors of the papers were carefully reviewed. This gives a comprehensive list of keywords that were used to get relevant papers which we filtered based on our own expertise of GenAI knowledge.

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 Generative AI 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," to capture diverse facets of Generative AI 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 from 2014 to the present, ensuring up-to-date coverage. 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 the key strategy. Simultaneously, the authors read the abstracts of the papers to include papers that are relevant to GenAI human interactions. As a result of our comprehensive methodology, we successfully compiled a corpus of 289 papers published in journal articles and conference proceedings.

3.3 Selection process

During our research process, we used a systematic screening process. 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. The overall focus for selection criteria involves papers with only human GenAI system interactions. Then all three authors discussed each paper present in the corpus to finalize a total of 154 papers.

3.4 Eligibility and Selection Criteria

Criteria for including and Excluding specific papers were defined based on the overall theme of our work, i.e. human interaction with generative AI systems. We generated the following criteria for inclusion and exclusion.

- **EC1 GenAI model Technical Improvement:** We excluded papers that solely focus on improving the performance of generative AI models for the applications.
- **EC2 No Human interactions:** We eliminated papers that solely presented applications of GenAI without actual users or humans interacting with the applications.

- **EC3 Opinion:** Literature Review, survey, and opinion papers were not included.
- **EC4 Change of Context:** Words/terms were used in a different context but are still present in the list.
- **EC5 False Search:** No search keywords were present but still were found by the search engine.
- **EC6 Idea Paper:** Short papers, proposals, demos, and position papers were not included.
- **EC7 Non-AI Papers:** We excluded the papers with no generative AI models being used (E.g., generating design using simple if-else conditions).
- **IC1 Interactions:** We included papers where humans are interacting with the AI system, even if they were not focusing on the systems or humans.
- **IC2 Extended Abstracts:** We included papers published as extended abstracts.
- **IC3 Study Based Papers:** We included study study-based paper that involved interaction with GenAI systems.

3.5 Analysis

Our analysis aims to uncover the dimensions and categories related to the design of GenAI systems, specifically focusing on their interaction with humans. To achieve this, we went through reading each paper in the list to gain an understanding of the various components that constitute GenAI systems when humans are involved. The analysis involves multiple stages of review and collaboration among the authors. Initially, one of the authors read a small subset ($N = 25$) of the papers to establish an approximate taxonomy of components and dimensions. This preliminary taxonomy was then discussed among all authors to iteratively refine and enhance it, adding or subtracting components and categories as necessary. Once the components and dimensions were finalized, 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 were not included in the final categorization but were 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. This collaborative approach helped ensure the accuracy and reliability of the categorization process. Ultimately, our analysis resulted in a comprehensive taxonomy of dimensions and categories that shed light on the design aspects of GenAI systems when interacting with human users.

In the following sections, we discuss various components and dimensions of the taxonomies spanning the design space of human GenAI systems. The taxonomy covers perspectives from both sides, The GenAI model side as well as the human side. Later on, we present the future opportunities and challenges in this domain summarized from our literature review. In the appendix, we included tables that contain all the citations and counts of the papers that fall into respective categories and dimensions.

4 PURPOSES OF USING GENAI

GenAI models possess various capabilities, with which the users can combine or iterate to achieve certain purposes in their domains. In this 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.

Purpose-1 Refine the Outcome. 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 [192], melody in a piece of music [120, 170], plots in a story [27], layout in a web application [77, 127], content [43, 148] or subjects [19, 34] 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 [98], efficiency of codes [112], precise layout of cameras in a VR space [199], etc.

Purpose-2 Explore Alternatives. GenAI possesses the abstraction of human knowledge across many disciplines and is capable of converting this knowledge to various modalities of information. Users can actively utilize GenAI to obtain ideas built from the abstraction of knowledge, by viewing multiple generated instances by GenAI. For example, GAN-based applications like GANravel [44] and GANCollage [183] 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 [6] 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 challenge or question, users can leverage GenAI to brainstorm potential solutions or avenues of inquiry. For instance, GenAI helps to generate codes to solve specific problems and then iterate with the users to optimize the codes [86]. Moreover, GenAI can directly generate the answers to the problem input by users [89].

Purpose-4 Understand Subjects. GenAI can significantly enhance users' understanding of various subjects by providing insights, generating examples, and offering new perspectives. Such understanding can be of the process of the model itself (e.g., GANslider [31] 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. [114] 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 [204] generates graphs and charts based on input data from the users in order to help understand the data.).

Purpose-5 Automate Processes. GenAI has a broad spectrum of applications when it comes to automation. GenAI can be applied to generating control sequences of robots [13, 71, 110, 155] 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. [119] propose fully autonomous programming with

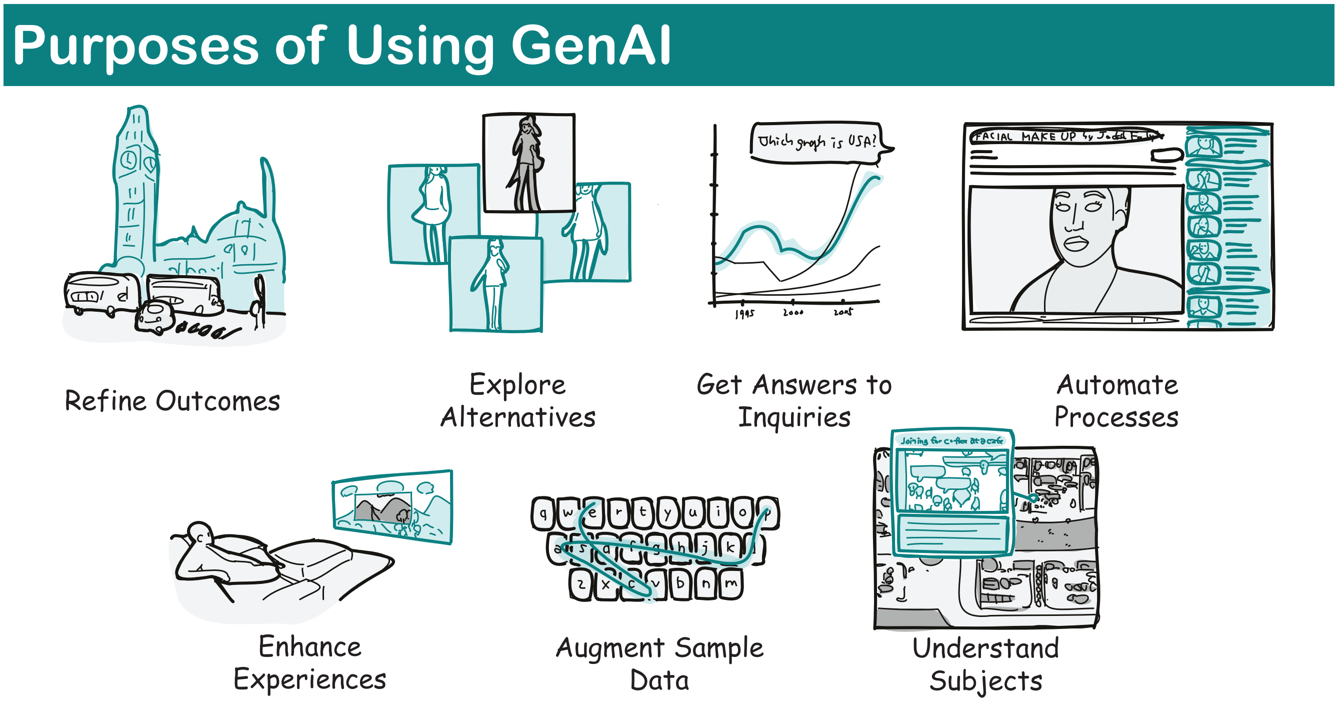

Figure 3: Purposes of Using GenAI depict the users' intention of the interactions and the high-level capabilities of the applications, consisting of Refine Outcomes [9], Explore Alternatives [183], Get Answers to Inquiries [89], Automate Processes [175], Enhance Experiences [165], Augment Sample Data [25], and Understand [144]

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 [93], make language in articles user-friendly [168], or modify the user input for more efficient communication [180, 194] 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 [7] adapts to the contexts of the users to provide personalized experiences of learning foreign vocabulary. AdaptiFont [80] 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 [25] utilizes GAN to generate synthetic gesture data for training keyboard gesture recognition models. Deepwriting [3] 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. [145] and Hamalainein et al. [63] experimented with AI-generated for research in social computing and HCI respectively.

5 FEEDBACK FROM MODELS TO USERS

In this section, we discuss the feedback from GenAI models to the users. We identified three dimensions to depict the current landscape of feedback techniques, namely, 1) output modalities, 2) functions of the models, and 3) Output synchronization.

Dimension-1 Output Modalities. The output modality of a generative AI 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 [76, 112, 119], the handwriting of texts [3], and fonts [80]. Specifically, natural language in texts can be chats [64, 72, 202], descriptions of a problem [86], or pieces of literature [27, 138, 168].

2D Visual Generative AI 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

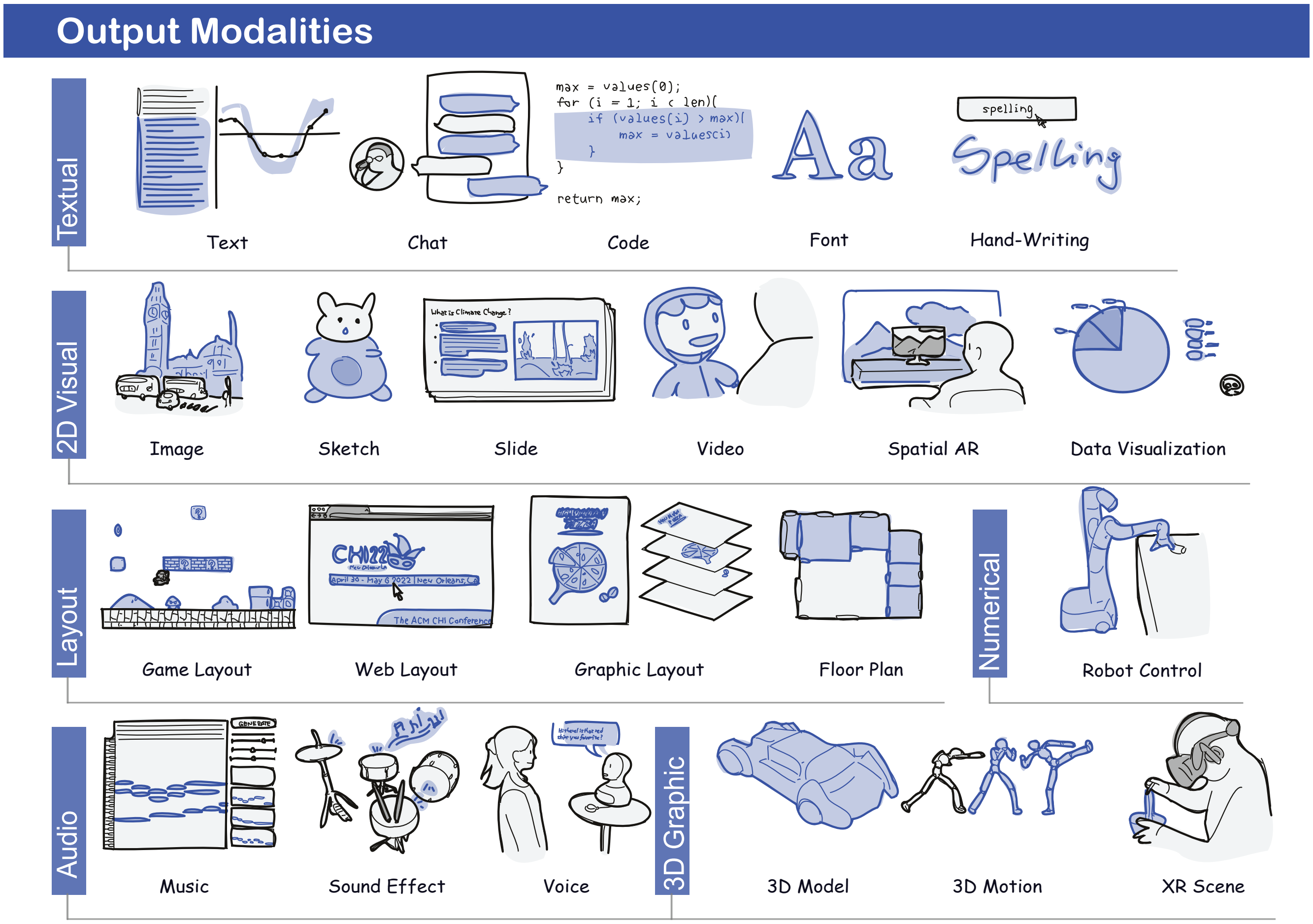

Figure 4: Output Modalities consists of four categories, namely textual (text [27], chat [78], code [86], font [80], and hand-writing [3]), 2D visual (image [9], sketch [45], slide [6], video [199], spatial AR [93], and visualizations of data [64]), layout (game layout [182]), web layout [91], graphic layout [60], and floor plan [66]), numerical data (robot control sequence [71]), audio (music [170], sound effect [20], and voice [74]), and 3D graphics (3D model [117], 3D motion [196], and XR scene [131])

training. 2D visual outputs consist of images [19, 114, 118, 159, 161], sketches [22, 207], videos [115, 175], 2D visualization of data [103, 204], and spatial AR [92, 93].

3D Graphic 3D graphic outputs consist of 3D models, 3D motion of various objects, and XR scenes. For example, Koyama et al. [98] utilize generated 3D models to provide suggestions during 3D design. Yoo et al. [199] generate VR camera layouts by referring to a clip of a film.

Audio Audio output consists of music [48, 120, 170], sound effects [138], or natural language voice synthesis [74].

Layout GenAI is capable of generating layout information, widely deployed in designing game layouts [127], web layouts [186], graphic layouts [60, 77], and more domain-specific layout designs (e.g. game layout [18, 163, 182]).

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 [25], control sequence to a robot [13, 155], and hierarchical representations of concepts [103].

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.

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,

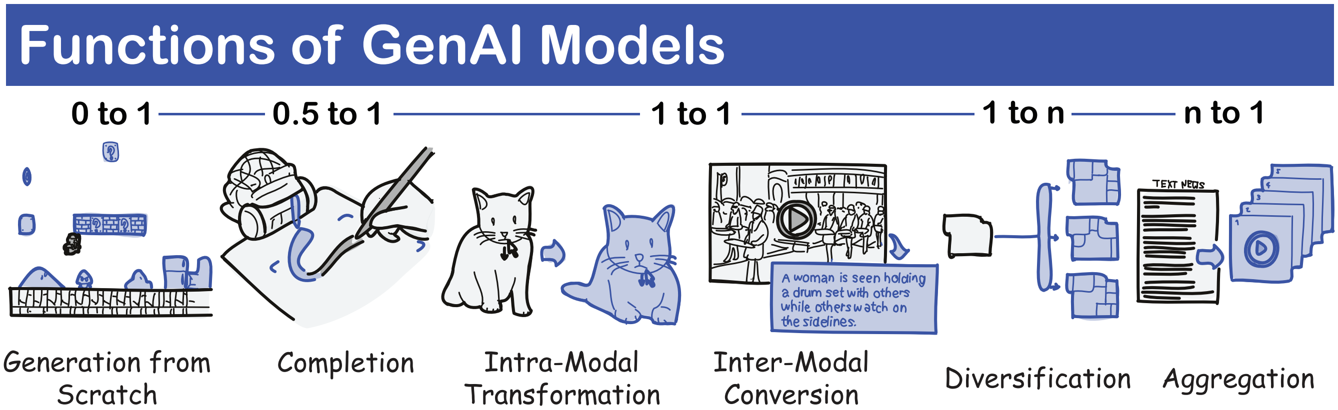

Figure 5: Functions of GenAI Models describe the capabilities of GenAI models, consisting of Generation from Scratch [182], Completion [110], Intra-Modal Transformation [42], Inter-Modal Conversion [189], Diversification [69], and Aggregation [138].

and the model operates more autonomously. With this autonomy, this method of input initialization can benefit (1) GenAI systems with non-expert [80, 120, 170] users who do not possess knowledge of proper input, (2) systems that passively assist the users [7, 80], or (3) systems that help the users with ideation within a specific domain through vast collections of examples [44, 96, 142, 182]

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 [34, 73] 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 [45] or provide a different perspective on the subject [6].

Intra-Modal Transformation Intra-modal transformation refers to the function of GenAI to change within the same input modality to produce 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. [168] propose an LLM-based method to modify the article for friendliness to people for minority and De et al. [34] enable human portraits editing with brain signals.). Such transformation usually results in changes in styles [60, 148], content [148, 200], or quality [93, 98, 194] 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. [22] enable image editing via texts to convert textual descriptions of a design into a visual representation of the design. Similarly, Yoo et al. [199] 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 [76, 86, 112, 119]. 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 [19].

Diversification GenAI possesses the function of diversification, by generating multiple diverse outputs from a single input. The outputs can be of the same modality. 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 [44, 183, 206], generating longer music given a short clip of melody [120, 170], generating textual content such as NPC quests in games [8] or (fake) news [212], or designs for different game layouts [18, 163, 182]. Moreover, the outputs can be of different modalities. In this case, GenAI is converting the input inter-modally to multiple outputs. For example, Jing et al. [77] 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 [187] blends the concepts from texts and images into a new image, to generate the best representations of an idea. Huber et al. [72] 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 [192] 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 [146].

Dimension-3 Output Synchronization. GenAI systems also utilize different output synchronization strategies. 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 [7] passively or as is [78, 185]. Another scenario with preliminary output is the human-GenAI collaboration tasks where GenAI takes the first move to inspire [103] or motivate [6] 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 [11, 73, 180] and auto completion [105]. 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 [31, 44, 196], 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 [96], co-writing [11], and programming assistance [46, 147].

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 [74, 130] or clicking on a button to input a set of parameters [59]. 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 [22], artistic image generation considering multiple attributes [96], style merging requiring multiple inputs [192], 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

This section discusses how users can control the GenAI and its output. To this end, this section delves into the common ways by which humans can provide feedback to the GenAI system. 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. This subsection includes methods to improve the output generated from GenAI.

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 [120, 137, 183]. 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 [43]. 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 [44], text [29, 49], or document [122]. Users can highlight or color paint emphasizing particular details, areas, or objects to either add [9],erase [9, 49], modify [43, 96] or keep [101] 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 influence or refine the output [31, 98, 163, 170, 178]. It provides users with a granular level of control over the output using slider [170] or numerical input [178] to semantically change a generation of output. Parameter tuning is helpful in image input-output target matching [31, 159], controlling the randomness of the generated output such as LLMs [107], changing the style of output [153] e.g. graphic design [178], and editing the content but preserving style [3, 153].

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 [22, 71, 155]. Users can provide commands sequentially to steer the output generation [118]. These commands are commonly used in Large Language models [62], Chatbots [64], and visual design assistant [22].

Additional Demonstration Users can further provide additional information to the GenAI system by drawing sketches [59, 211], outline [164], copy & paste [24], handwriting [2], images [148] and keywords [116] to narrow the scope of the generated output. Users can provide specific preferences [211], additional context [192, 209], and set constraints [27] to generate a more relevant and accurate output.

Re-intiliazation Users can re-initialize the generation process over, with new or some adjustments [45, 110, 177] in the input. This iterative and adaptable approach allows users to fine-tune content generation effectively [139]. Reinitializing also allows for experimentation to see how different approaches or inputs affect the output [111].

Dimension-2 Objects to Control. This section covers what part of GenAI with which humans are interacting to control the output. We broadly categorize objects of control into four based on the focus of humans either in controlling the GenAI model or with input.

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 [44] 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

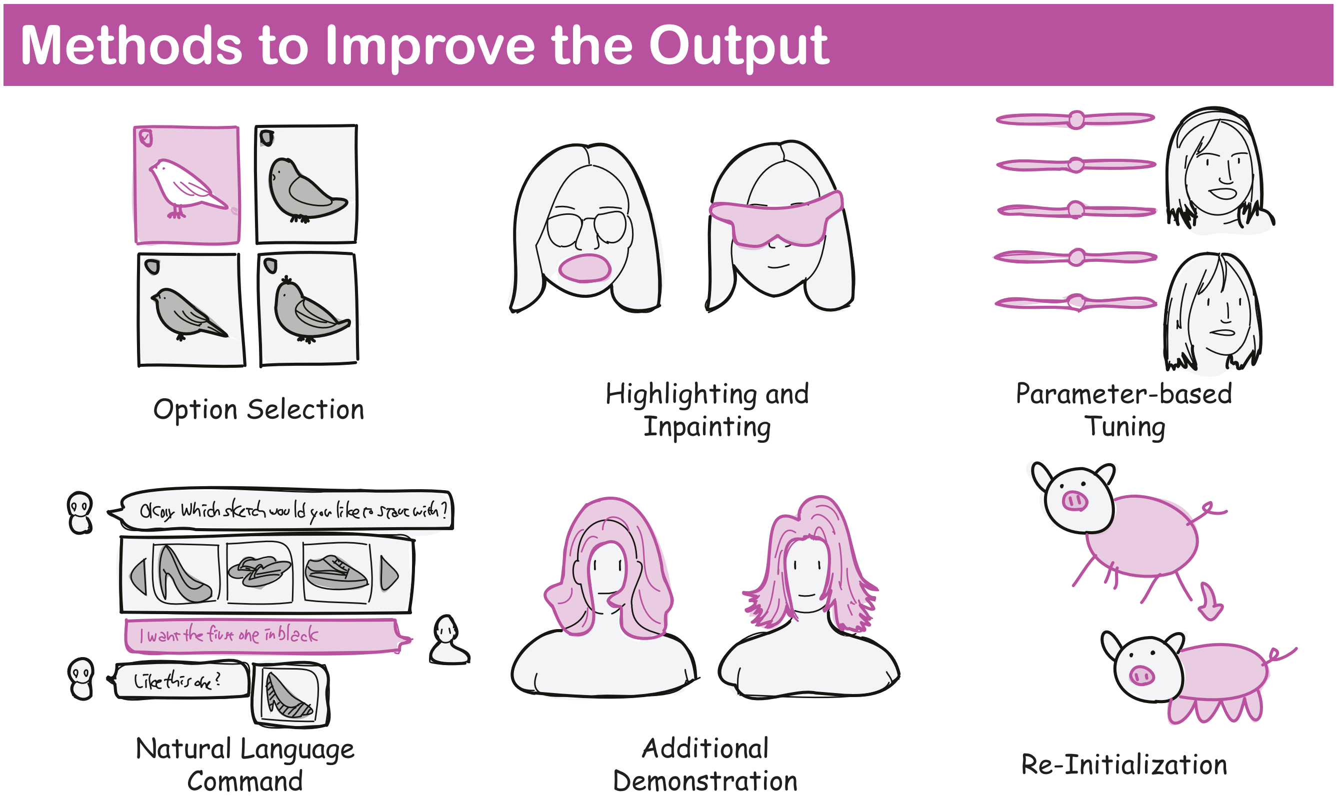

Figure 6: Methods to Improve the Output from the user's perspective include Option Selection [206], Highlighting and In-painting [44], Parameter-based Tuning [31], Natural Language Command [22], Additional Demonstration [164], and Re-Initialization [45].

freedom to visualize the direction of output [81], explore for diverse generated outputs [33, 206], and transform [127] 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 [44] or moving sliders [31, 96]. In addition, latent space allows users to potentially influence the types [77, 200], quantities [206], and levels of variability [44] present in the system's outputs.

Parameters These GenAI model parameters allow users to control the creativity(randomness) in the outputs generated from the GenAI systems [107]. Some GenAI system includes *Temperature* parameter [120, 180], frequency penalty [101] and random seed during the development of GenAI system [107] to control the variability in the generated output. For example, Louie et al. [170] use a temperature parameter to generate conventional or surprising music. Hyperparameters can be changed by the end-user using tools that allow changing values using sliders [120] or text editor [107].

Retraining Retraining the GenAI system involves fine-tuning the GenAI model either with few or zero-shot learning [14, 174]. It allows the system to become adaptive and personalized for each individual user [167]. It also allows users to build this GenAI system unique for their own process [174]. Retraining of GenAI system makes them personalized to the user by improving task-specific capabilities and domain-specific knowledge of the GenAI system.

Input Input control provides the unique ability for the user to interact with GenAI without retraining or changing parameters [202]. The quality [114] and relevance [103] 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 [192]. 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 [75, 198]. For example, Jiang et al. [76] used LLMs to support software development. Users can also develop an automatic method of designing such input prompts instead of manually specifying [90] to get more relevant output. Some of the examples are by suggesting prompts [103], constructing prompt templates [111], combining multiple prompts primitive [75], reformulating prompts suiting GenAI system [122], and transforming prompt to different input modality [27].

Dimension-3 Mediums of Control. We discuss what are the mediums, humans provide input to and control the GenAI system.

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. Buttons [30],

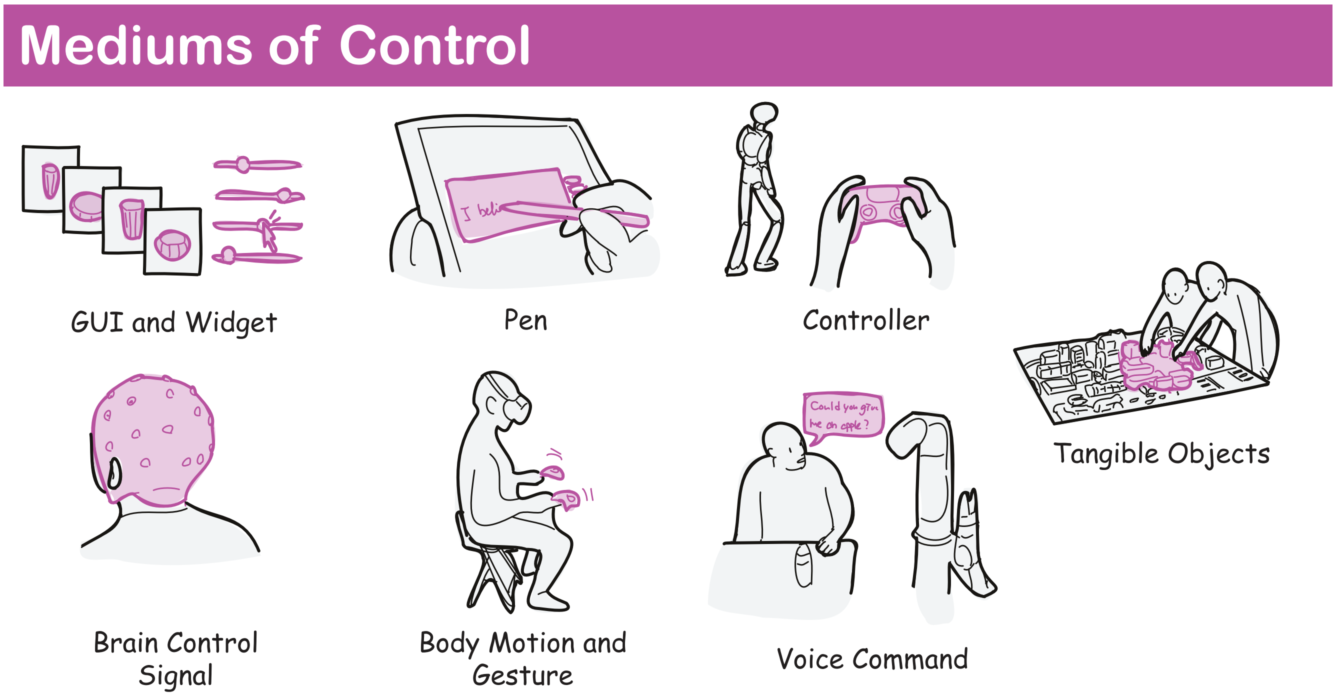

Figure 7: Mediums of Controlling GenAI models include GUI and Widget [98], Pen [3], Controller [196], Brain Control Signal [82], Tangible Objects [136], Body Motion and Gesture [20], and Audio Command [1].

sliders [31], mouse clicks and drags [42], and tool pallets [206] are often used to provide input or any modifications to GenAI system. Image editors [43] and text editors [193] are also used to display outputs and provide or edit the inputs. Interfaces are also used in displaying the 3D content [117] Information panels and menus are used to display information that guides users to use the GenAI system [170]. Canvas is also used to provide multiple [43, 44, 183] or different modalities [148] 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. For instance, Xu et al. [196] used a game controller to control human motion generation.

Tangible Object The user can control GenAI systems by moving objects in the real world [136]. The tangible object includes interactions with physical objects such as static objects[165], dynamic objects[131], and remote objects [40].

Pen Users can use a physical pen or pencil to draw [110, 177], write text [2, 3], or make sketches [96, 209] on screens [66] and paper [177].

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) [32, 82, 167]. Brain responses are directly connected to the internal parameters of GenAI models such as latent space [34] for providing implicit feedback. For example, Spape et al. [167] 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. Gesture movement can be captured by interactive surfaces such as mobiles and tablets [25, 26]. Also, body movements are useful in interacting with GenAI in immersive environments[20]. Face tracking, facial emotion, and expressions are also used in designing natural interactions with GenAI[79]. Body motion and Gesture are the most engaging mechanisms while interactive with GenAI [79]

Audio Audion includes both human voice[1, 74] and music [48]. 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 [112], 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 [78] and intuitive manner [112]. Additionally, audio can also be used to provide additional information to GenAI for music generation [48]

7 LEVELS OF ENGAGEMENT

In this section, we report our categorization based on the level of engagement of the human-GenAI interaction. We identified four levels of engagement, namely, *Passive Engagement*, *Assistive Engagement*, *Collaborative Engagement*, and *Deterministic Engagement*.

Level-1 Passive Engagement: Passive engagement depicts the systems with which users receive information or content generated

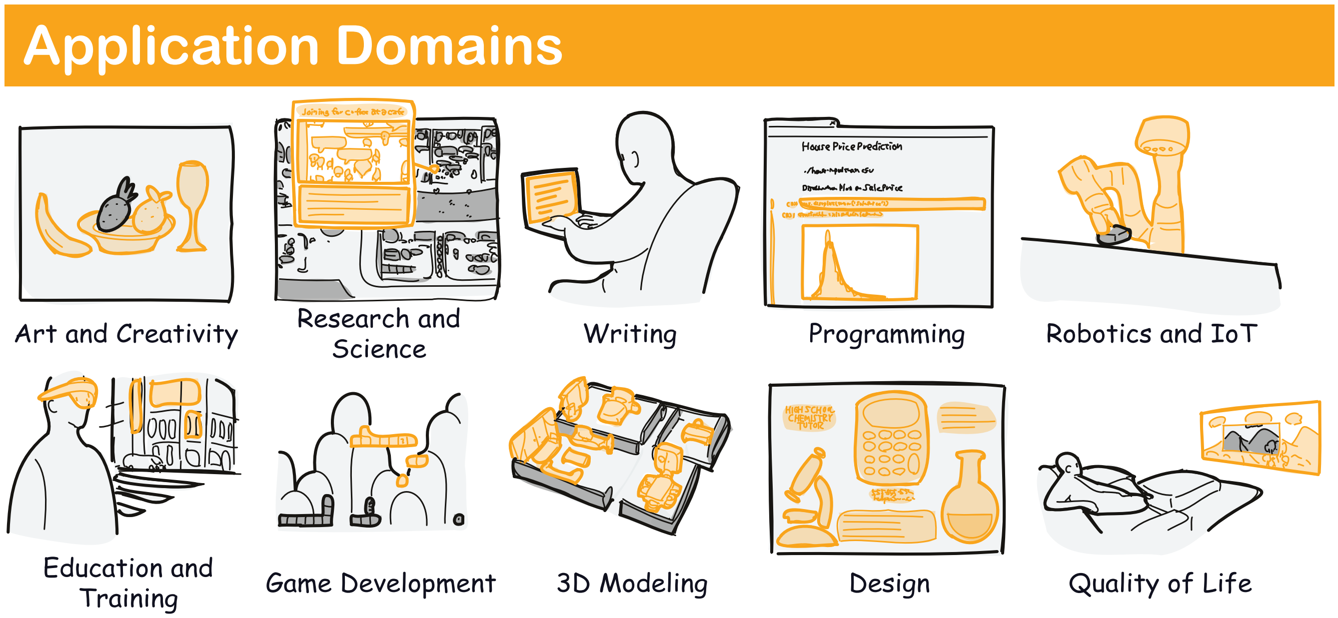

Figure 8: Through our discovery of the literature, the application domains of GenAI include Art and Creativity [139], Research and Science [144], Writing [7], Programming [184], Robotics and IoT [1], Education and Training [7], 3D Modeling [50], Design [137], and Quality of Life [165].

by the AI without direct interaction. Example system designs with passive engaging interactions fulfill the tasks of immersive news writing [138] and immersive vision system [92, 93], where the users are passively engaged with GenAI and its product without explicit interactions to guide the output.

Level-2 Deterministic Engagement: As its name conveys, in a GenAI system where the engagement is deterministic, the outcome is largely determined by the AI's inherent logic, instructions, or a predetermined set of rules, rather than being shaped by the users' interactions. 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. 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 [7], AI-generated hierarchical tutorials for the users to follow [175], or an adaptive font generator that evaluates the users' performance and autonomously generates the best font for reading [80].

Level-3 Assistive Engagement: Assistive engagement allows the GenAI system to generate content to assist the users in the creation process, not necessarily of the same content. In other words, the output of assistive GenAI systems does not substantially contribute to the final product of the interactions but rather conceptually or abstractly contributes to the creation process. An example of an assistive system can be an auto-completion assistant in writing [73], a contextual provider of suggestions [180], or an online debugger for an ongoing programming [147].

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. 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 [15, 75, 125, 197, 200], 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 [31, 42, 44, 118, 127, 188]. 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. Examples of this method can be jointly creating slides for a presentation [6], finishing a sketch by GenAI adding details [45], and composing a piece of melody by both users and generated music [120, 170].

8 APPLICATION DOMAINS

Through our exploration, we identified a range of diverse application domains of human-GenAI systems. Figure 8 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 9.

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 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 are following the classification by Suzuki et al [173]: (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 [50, 106, 212], 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 [32, 66, 74, 190]. The evaluations can be conducted on annotated datasets by the researchers themselves and the technical statistics of the datasets are also reported [3, 185].

Evaluation-2 Demonstration. Evaluations through *demonstrations* assess the system performance under specific conditions. Common methods consist of generalizability demonstration [98], proof of concept demonstration [104, 168, 202], demonstration through an example use case [19, 93, 201].

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 [31, 114], qualitative lab studies [74] for rich insights into design and contextual understanding, quantitative lab studies [43] for objective measurement and generalizability, and interviews with both experts [142, 171] and novices [193, 203] in the subject matter.

10 FINDINGS

In this section, we discuss the standard strategies and gaps that we identify through our extensive analysis of the literature.

Finding-1 Mediums of Control: Direct but not Intuitive. Through our literature review, we notice that direct control modalities are preferred (e.g. widgets, controllers, drawings and highlighting, and text, $N = 122$) over the intuitive ones (e.g. gestures, brain signals, and voice, $N = 32$) 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 [43, 44, 183] is direct yet not intuitive, because 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, 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 ($N = 89$). 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 I/O design space of the GenAI systems, which is a significant missing piece in the current research. It is important that users understand the process of using a system, i.e. what consequences result from each of their interactions. 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, also referred to as Foundation Models [95] along with Large Vision Models, 63 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 [91], and vice versa [74]. Similarly, images can be summarized by models and translated into text for conversation as well [189]. Enlightened by this finding, we discuss the future opportunities of research in interaction design in section 11

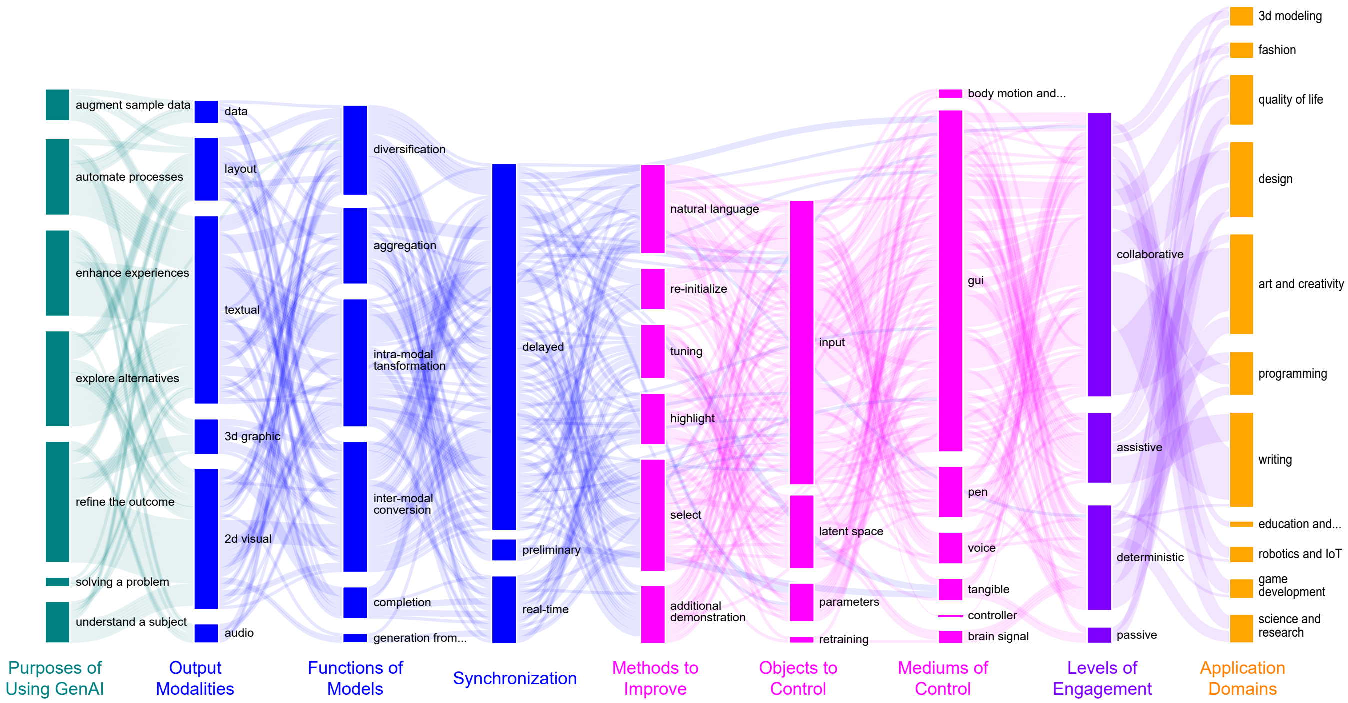

Figure 9: An alluvial diagram of the characteristics from our survey across all dimensions.

Finding-4 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 154 research papers, only 11 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 [38, 47, 135], opinionated bias in GenAI system [73], and gender bias in Natural Language generation [168]. We will be detailing the future opportunity of investigating how to tack GenAI ethical problems in section 11.

11 FUTURE OPPORTUNITIES

Opportunity-1 Bridging User Interactions with the AI Output. As aforementioned in section 10, intuitive mappings from the users' interaction to the models' output are necessary for designing a GenAI system. We envision two major directions to bridge these two aspects, namely *Exploring Control of Internal Parameters* from the AI side and *Exploring Novel Interactions* from the Human side.

—*Exploring Control of Internal Parameters* Interacting with internal parameters of the GenAI system allows users to explore GenAI model capabilities[31]. Only 15 papers out of 156 allow users to control the model parameters. Generally speaking, all systems should allow users to guide the behavior of output to align with their preferences. This involves adjusting certain GenAI model parameters and exploring the full capabilities of GenAI models. The question remains unanswered: **What are the correspondences between the model parameters and the model capabilities?** Secondly, users should be provided with options to control internal parameters. The core question to be addressed here is: **Which capabilities (and their corresponding parameters) are to be made optional for users to control, considering the particular application domains the systems serve?** Third of all, all GenAI models are not easy to control and sometimes it is cumbersome for developers to achieve total controllability. Exploring the workload distribution for target user groups is essential: **Does this approach make it easier for non-technical users to interact with the system?** Or **Does it hinder the ability of more technical users to fine-tune or troubleshoot the system?** Lastly, blindly adding controllability to the GenAI model complexities the usage. 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 intuitive interactions 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?** 2) How do we accurately contextualize and adapt the generated content to users' natural interactions as input?

Opportunity-2 Designing and Exploring Interactions with Foundation Models. *—Various I/O Modalities through Foundation Models* We identified promising usage of Foundation Models in our survey. 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 [117, 161, 190]. 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. To advance in the intuitive and user-friendly design of interactions with Foundation Models, further research may need to address questions such as (1) **What are the I/O modalities that are the most intuitive for a specific application?** (2) **What are the interactions that suit the applications with specific I/O modalities?** (3) **What are the metrics to evaluate the interactions?** (4) **What are the general patterns we can conclude from the designs addressing the aforementioned questions?**

—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-3 Explainable AI from the Users' perspective. Traditionally, explainable AI has often been discussed in terms of making machine learning models understandable to developers, researchers, or regulators. We envision further discussion, particularly over GenAI applications, from the end-users' perspectives emphasizing the importance of making GenAI systems understandable and controllable.

—Real-time feedback to learn the AI's behaviors From the literature, we discover that most systems utilize a delayed synchronization strategy ($N = 117$), where users finish their interactions before an output is generated and fed back to the users. This results in a discontinuity in the user experience, because they only see a delayed product once they fix the prompts or attribute setting, which sets a cognitive offset between the users' interaction and the corresponding outcome. We envision the use of real-time feedback to tackle this cognitive offset. Real-time feedback reflects continuously on how user input affects the system's responses, allowing users to **iteratively shape the AI's behavior**. For example, an auto-completion system for writing should not wait until the users finish their type, but should rather simultaneously suggest possible completion choices, with which the users can smoothly comprehend how their input changes the generation of the choices.

—Foster a user-friendly learning curve of the system As was pointed out in the section 10, it is technically cumbersome to make the end-users understand the process of the GenAI models, for it is mostly a black box. However, we foresee the value of researching the learning curve of the I/O mappings. It is important for the black box users to understand what output their input leads to (i.e. prompting) [19]. This can be done by highlighting the change brought by the users' interaction or comparing the differences between output from two iterations. Through specific visualization or tweaking of the output, the users obtain a smooth learning curve of the I/O pattern of the systems, which can foster a quick mastering of the usage of the systems. To accomplish this goal, the research question has yet to be addressed: **How do we guide the users to give the best input leading toward their desired output?**

Opportunity-4 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 [54], intrusion into copyrights and intellectual property [135], and generation of illegal content [4, 28]. 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 generative AI 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-generative-AI 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 generative AI and the common ground of application design. Further, we discuss future opportunities in this topic, namely, (1) bridging between user interactions and AI output, (2) designing interactions for Foundation Models, (3) explainable AI from the Users' perspective, and (4) ethical discussion on GenAI. We conclude with a discussion on the negative externalities of GenAI with a possible reduction of the importance of Humans with GenAI evolution. We hope our research will guide and inspire future work on human-generative-AI interaction.

ACKNOWLEDGMENTS

We wish to thank all readers for reading our paper. This work is partially supported by the NSF under the Future of Work at the Human-Technology Frontier (FW-HTF) 1839971. We also acknowledge the Feddersen Distinguished Professorship Funds and a gift from Thomas J. Malott. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

REFERENCES

[1] 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).

[2] 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.

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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).

[10] 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).

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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*.

[17] 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).

[18] 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.

[19] 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.

[20] 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.

[21] 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).

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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)*.

[27] 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.

[28] 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).

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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).

[37] 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

[38] 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

[39] 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.

[40] 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.

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

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

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

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

[52] 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.

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

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

[55] 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).

[56] 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.

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

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

[59] 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.

[60] 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.
[61] 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.
[62] 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.
[63] 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.
[64] 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.
[65] 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.
[66] 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.
[67] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. arXiv:2006.11239 [cs.LG]
[68] 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.
[69] 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.
[70] 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).
[71] 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).
[72] 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.
[73] 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.
[74] 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.
[75] 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.
[76] 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.
[77] 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.
[78] 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.
[79] 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.
[80] 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.
[81] 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.
[82] Lauri Kangassalo, Michiel Spapé, and Tuukka Ruotsalo. 2020. Neuroadaptive modelling for generating images matching perceptual categories. *Scientific reports* 10, 1 (2020), 14719.
[83] 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).
[84] 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.
[85] 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.
[86] 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.
[87] 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).
[88] 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.
[89] 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.
[90] 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.
[91] 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.
[92] 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.
[93] 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.
[94] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
[95] 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).
[96] 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.
[97] 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.
[98] 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.
[99] 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).
[100] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.
[101] 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.
[102] 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.
[103] 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.
[104] 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.

[105] 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.

[106] 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).

[107] 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.

[108] 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.

[109] 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.

[110] 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.

[111] 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.

[112] 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.

[113] 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.

[114] 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.

[115] 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).

[116] 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.

[117] 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.

[118] 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.

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

[120] 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.

[121] 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.

[122] 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.

[123] 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.

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

[125] 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.

[126] 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).

[127] 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.

[128] 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.

[129] 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.

[130] 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.

[131] 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.

[132] 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.

[133] 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.

[134] 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]

[135] 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.

[136] 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.

[137] 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.

[138] 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.

[139] 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.

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

[141] 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.

[142] 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.

[143] 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 arXiv:https://www.bmj.com/content/372/bmj.n71.full.pdf

[144] 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).

[145] 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.

[146] 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.

[147] 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).

[148] 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.

[149] 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]

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

[151] 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.

[152] 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]

[153] 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.

[154] 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).

[155] 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.

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

[157] 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.

[158] 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.

[159] 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.

[160] 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

[161] 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.

[162] 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.

[163] 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.

[164] 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.

[165] 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.

[166] 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).

[167] 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).

[168] 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.

[169] 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.

[170] 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.

[171] 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.

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

[173] 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.

[174] 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.

[175] 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.

[176] 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.

[177] 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.

[178] 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.

[179] 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.

[180] 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.

[181] 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).

[182] 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.

[183] 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.

[184] 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.

[185] 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.

[186] 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.

[187] 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.

[188] 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).

[189] 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.
[190] 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.
[191] 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).
[192] 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.
[193] 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.
[194] 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.
[195] Jungang Xu, Hui Li, and Shilong Zhou. 2015. An overview of deep generative models. *IETE Technical Review* 32, 2 (2015), 131–139.
[196] 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.
[197] 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.
[198] 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.
[199] 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.
[200] 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.
[201] Paulina Yurman and Anuradha Venugopal Reddy. 2022. Drawing Conversations Mediated by AI. In *Proceedings of the 14th Conference on Creativity and Cognition*. 56–70.
[202] 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).
[203] 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.
[204] 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.
[205] 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).
[206] 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.
[207] 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.
[208] 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).
[209] 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.
[210] 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.
[211] 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.
[212] 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.
[213] 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.
[214] 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.
[215] 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.
[216] 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.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

Table 1: Appendix Table: Purposes of Using GenAI

Category	Count	Citation
Augment Sample Data	10	Figure [25]: [3, 8, 18, 25, 62, 63, 145, 163, 182, 212]
Automate Processes	26	Figure [175]: [13, 15, 40, 45, 71, 74, 78, 88, 91, 102, 102, 106, 107, 110, 117, 119, 123, 139, 147, 155, 166, 175, 179, 184, 189, 196]
Enhance Experiences	27	Figure [165]:[2, 7, 16, 20, 29, 30, 73, 75, 76, 80, 92, 93, 115, 116, 131–133, 138, 165, 168, 169, 174, 180, 193, 194, 202, 203]
Explore Alternatives	31	Figure [183]:[6, 11, 23, 33, 39, 44, 48–50, 53, 66, 69, 90, 97, 99, 101, 104, 105, 113, 125, 137, 142, 146, 156, 178, 183, 187, 197, 201, 206, 211]
Refine Outcomes	38	Figure [9]:[9, 19, 22, 24, 27, 32, 34, 42, 43, 56, 59, 60, 77, 79, 82, 96, 98, 112, 118, 120, 121, 127, 141, 148, 153, 161, 164, 167, 170, 171, 188, 190, 192, 199, 200, 207, 209, 213]
Get Answers for Inquiries	4	Figure [89]:[35, 46, 86, 89]
Understand Subjects	14	Figure [144]:[31, 64, 72, 81, 103, 114, 136, 144, 159, 175, 185, 186, 204, 210]

Table 2: Appendix Table: Output Modalities

Category	Count	Citation
2D Visuals	48	Figure [19][45][6][199][93][64]: [6, 9, 19, 22–24, 31–34, 42–45, 64, 79, 81, 82, 92, 93, 96, 97, 103, 114–116, 118, 122, 139, 142, 148, 153, 159, 161, 165, 167, 175, 177, 183, 187, 190, 192, 199, 201, 204, 206, 207, 209]
3d Graphic	12	Figure [117][196][131]: [40, 59, 98, 117, 121, 131–133, 136, 164, 196, 199]
Audio	7	Figure [170][20][74]: [20, 48, 74, 120, 138, 170, 213]
Numerical Data	8	Figure [71]: [13, 25, 71, 78, 102, 103, 155, 196]
Layout	20	Figure [182][91][60][66]: [6, 18, 50, 60, 66, 69, 77, 89, 91, 102, 127, 137, 163, 178, 182, 185, 186, 188, 210, 211]
Textual	62	Figure [27][78][86] [80][3]: [2, 3, 7, 8, 11, 15, 16, 27, 29, 30, 35, 39, 46, 49, 53, 56, 62–64, 72–76, 78, 80, 86, 88, 90, 99, 101, 104–107, 111, 112, 119, 123, 125, 138, 141, 146, 147, 156, 166, 168, 169, 171, 174, 179, 180, 184, 189, 193, 194, 197, 198, 200, 202, 203, 212]

Table 3: Appendix Table: Functions of Models

Category	Count	Citation
Aggregation	26	Figure [138]: [6, 15, 29, 35, 46, 64, 72, 78, 88, 99, 101, 103, 106, 107, 116, 123, 125, 138, 145, 146, 175, 184, 187, 192, 210, 211]
Completion	12	Figure [110]: [6, 30, 34, 45, 73, 104, 105, 110, 139, 147, 171, 177]
Diversification	30	Figure [69]: [3, 8, 18, 25, 31, 33, 43, 44, 48, 53, 62, 63, 69, 77, 81, 90, 96, 97, 113, 120, 137, 163, 170, 178, 180, 182, 183, 188, 206, 212]
Generation from Scratch	10	Figure [182]: [7, 44, 80, 82, 96, 120, 142, 167, 170, 182]
Inter-modal Conversion	42	Figure [189]: [19, 20, 22, 24, 27, 40, 50, 66, 69, 71, 74, 76, 86, 89, 102, 102, 112, 114, 115, 117–119, 121, 122, 127, 131–133, 136, 138, 155, 161, 164, 179, 185, 186, 189, 190, 196, 199, 204, 207]
Intra-modal Transformation	40	Figure [42]: [2, 7, 9, 11, 16, 32, 34, 39, 42, 49, 56, 59, 60, 75, 79, 80, 91–93, 98, 141, 142, 148, 153, 156, 159, 165, 166, 168, 169, 174, 193, 194, 197, 200–203, 209, 213]

Table 4: Appendix Table: Synchronization

Category	Count	Citation
Preliminary	8	[6, 7, 78, 92, 102, 103, 165, 185]
Real-time	26	[11, 15, 16, 30–32, 34, 44, 46, 49, 73, 82, 96, 101, 104, 105, 131, 132, 147, 159, 166, 167, 171, 180, 184, 196]
Delayed	118	[2, 3, 8, 9, 18–20, 22–25, 27, 29, 33, 35, 40, 42–46, 48, 50, 53, 56, 59, 60, 62–64, 66, 69, 71, 72, 74–77, 79–81, 86, 88–91, 93, 96–99, 102, 106, 107, 110, 112–123, 125, 127, 130, 133, 136–139, 141, 142, 145, 146, 148, 153, 155, 156, 161, 163, 164, 168–170, 174, 175, 177–180, 182, 183, 185–190, 192–194, 197, 199, 201, 203, 204, 206, 207, 209–213]

Table 5: Appendix Table: Methods to Improve the Output

Category	Count	Citation
Additional Demonstration	21	Figure [164]: [2, 19, 24, 26, 27, 40, 59, 61, 69, 113, 115–117, 121, 132, 148, 164, 188, 192, 209, 211]
Highlighting and In-painting	16	Figure [44]: [3, 9, 11, 16, 29, 30, 43, 44, 49, 56, 88, 96, 101, 103, 122, 200]
Natural Language Commands	28	Figure [22]: [6, 7, 22, 25, 39, 62–64, 71, 72, 74, 76, 78, 80, 99, 102, 112, 114, 118, 119, 155, 168, 171, 193, 194, 198, 202, 203]
Re-Initialization	15	Figure [45]: [27, 45, 53, 66, 110, 111, 136, 139, 145, 177, 185, 190, 196, 197, 201]
Option Selection	37	Figure [206]: [15, 23, 32, 33, 43, 48, 50, 60, 73, 82, 86, 90, 91, 97, 102, 104, 105, 120, 125, 127, 131, 137, 141, 142, 146, 147, 156, 166, 167, 169, 175, 183, 187, 204, 206, 207, 210]
Parameter-based Tuning	19	Figure [31]: [2, 3, 24, 31, 34, 42, 75, 81, 89, 98, 107, 153, 159, 163, 170, 174, 178, 183, 213]

Table 6: Appendix Table: Objects to Control

Category	Count	Citation
Latent Space	25	[3, 9, 24, 31–34, 42–44, 77, 81, 82, 96, 98, 127, 153, 159, 163, 178, 183, 200, 206, 211, 213]
Parameters	12	[15, 75, 101, 106, 107, 120, 125, 147, 156, 170, 180, 203]
Retrainnig	3	[14, 167, 174]
Input	93	[2, 6, 7, 11, 16, 19, 20, 23, 25–27, 29, 30, 35, 39, 40, 45, 48, 49, 53, 56, 60–63, 66, 69, 71, 72, 74–76, 78, 79, 86, 88–91, 97, 99, 102, 102–105, 110–117, 121–123, 130–132, 137, 139, 141, 145, 146, 148, 155, 161, 164, 166, 168, 169, 171, 175, 177, 179, 185, 186, 188–190, 192–194, 196–202, 207, 209]

Table 7: Mediums to Control

Category	Count	Citation
Brain Signal	4	Figure [82]: [32, 34, 82, 167]
Controller	1	Figure [196]: [196]
Gesture	4	Figure [20]: [20, 25, 26, 79]
GUI and Widgets	110	Figure [98]: [6, 7, 9, 11, 15, 16, 18, 19, 22–24, 29–31, 33, 35, 39, 42–44, 46, 49, 50, 53, 56, 60–62, 64, 66, 69, 72, 73, 75–77, 80, 81, 86, 88–90, 97–99, 101, 102, 102–107, 111, 113, 114, 116–121, 123, 125, 127, 130, 137, 138, 141, 142, 145–148, 153, 155, 156, 159, 161, 163, 166, 168–171, 174, 175, 178–180, 182–184, 186, 187, 189, 190, 192–194, 197–200, 202–204, 206, 207, 210]
Pen	17	Figure [3]: [2, 3, 26, 27, 45, 59, 66, 96, 110, 132, 139, 164, 177, 188, 201, 209, 211]
Tangible	7	Figure [136]: [40, 92, 93, 131, 133, 136, 165]
Audio Command	10	Figure [1]: [1, 48, 71, 74, 78, 91, 112, 115, 185, 213]

Table 8: Appendix Table: Levels Of Engagement

Category	Count	Citation
Colloborative	91	[6, 9, 15, 16, 18–20, 22–24, 26, 27, 29, 31, 33, 39, 40, 42–45, 48, 53, 59, 60, 63, 64, 71, 75, 76, 79, 81, 86, 90, 96, 98, 99, 101–103, 110–122, 125, 127, 130–133, 139, 141, 146, 148, 153, 155, 156, 159, 161, 163, 164, 168–171, 177, 178, 182, 183, 188, 192–194, 196, 197, 200, 202, 203, 206, 207, 210, 212]
Deterministic	33	[7, 8, 25, 32, 34, 50, 61, 62, 66, 72, 74, 77, 78, 80, 82, 88, 89, 106, 123, 145, 167, 175, 184, 186, 187, 189, 190, 198, 199, 201, 204, 209, 211]
Assistive	22	[2, 3, 11, 30, 46, 49, 56, 69, 73, 91, 104, 105, 107, 137, 142, 147, 166, 174, 179, 180, 185, 213]
Passive	5	[92, 93, 102, 138, 165]

Table 9: Appendix Table: Application Domains

Category	Count	Citation
3D Modeling	7	Figure [50]: [50, 59, 69, 98, 121, 164, 199]
Art and Creativity	33	Figure [139]: [3, 9, 19, 20, 24, 31, 32, 34, 42, 43, 45, 48, 82, 96, 97, 110, 115, 116, 120, 122, 139, 148, 153, 161, 167, 170, 177, 183, 185, 190, 201, 206, 213]
Design	24	Figure [137]: [6, 23, 60, 66, 77, 89, 91, 93, 113, 114, 117, 118, 127, 137, 138, 175, 178, 186–188, 204, 207, 210, 211]
Education and Learning	3	Figure [7]: [7, 81, 103]
Fashion	5	[22, 33, 142, 192, 209]
Game Development	7	Figure [163]: [8, 18, 79, 99, 163, 182, 196]
Programming	14	Figure [184]: [46, 75, 76, 86, 106, 107, 112, 119, 123, 147, 171, 179, 184, 193]
Quality of Life	16	Figure [165]: [2, 7, 25, 40, 74, 78, 80, 92, 102, 130–133, 165, 189, 203]
Robotics and IOT	6	Figure [1]: [1, 64, 71, 72, 102, 155]
Science And Research	10	Figure [144]: [39, 44, 61–63, 144, 145, 159, 198, 212]
Writing	31	Figure [7]: [7, 11, 15, 16, 26, 27, 29, 30, 49, 53, 56, 73, 88, 90, 101, 104, 105, 111, 125, 141, 146, 156, 166, 168, 169, 174, 180, 194, 197, 200, 202]

Table 10: Appendix Table: Evaluation Methodologies

Category	Count	Citation
Technical Evaluation	50	**Quantitative**: [1, 3, 9, 18, 22, 24–27, 32, 34, 39, 40, 42, 53, 59, 61, 63, 66, 69, 71, 72, 74, 77, 82, 90, 92, 101, 102, 105, 111, 118, 119, 121, 122, 127, 155, 161, 164, 165, 167, 175, 178, 182, 185, 187, 188, 190, 198, 213]
	40	**Qualitative**: [1, 7–9, 22, 24, 25, 27, 32, 39, 42, 45, 50, 53, 59, 66, 71, 72, 92, 105, 106, 110, 118, 121, 122, 145, 146, 155, 161, 164, 165, 167, 175, 178, 187, 188, 190, 192, 212, 213]
Demonstration	12	[19, 34, 92, 93, 98, 104, 106, 165, 168, 190, 201, 202]
User Evaluation	108	[2, 3, 6–8, 11, 15, 16, 18–20, 23, 24, 26, 27, 29–31, 33, 42–46, 48, 49, 53, 56, 60, 62, 64, 69, 73–80, 86, 88–93, 96, 99, 101, 103, 105, 107, 110, 112–117, 119, 120, 123, 125, 127, 130, 131, 137–139, 141, 142, 145–148, 153, 156, 159, 163, 165–167, 169–171, 174, 179, 180, 183–187, 189, 192–194, 197, 199, 200, 203, 204, 206, 207, 209, 210, 213]

Table 11: 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	Flemigo	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	OpenFlemingo	https://github.com/mlfoundations/open_flamingo