Interacting with AI by Manipulating Intents

Tae Soo Kim
taesoo.kim@kaist.ac.kr
School of Computing, KAIST
Daejeon, Republic of Korea

Abstract

Advanced AI models allow users to perform diverse tasks by simply expressing their high-level intents, without performing low-level operations. However, users can struggle to fully form and effectively express their intents, and inspecting and evaluating model outputs to verify whether their intents have been satisfied incurs significant cognitive load. My PhD research introduces the concept of *intent manipulation*, where user intents are externalized as interactive objects, allowing for direct exploration and iteration on both intents and model outputs. I explore three forms of intent manipulation: intent curation, disentangle intents into palettes users can curate their intent with; intent assembly, creating intent blocks that users can combine and experiment with; and intent framing, helping users inspect outputs through the lens of their intents. This work contributes to human-AI interaction by suggesting how interfaces can be designed to support iterative exploration and sensemaking of one's own intents and the AI models in parallel.

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; *Empirical studies in HCI*; • **Computing methodologies → Natural language processing**.

Keywords

Human-AI Interaction, Intent, Natural Language Interfaces, General Purpose AI, Large Language Models

ACM Reference Format:
Tae Soo Kim. 2025. Interacting with AI by Manipulating Intents. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3706599.3707611

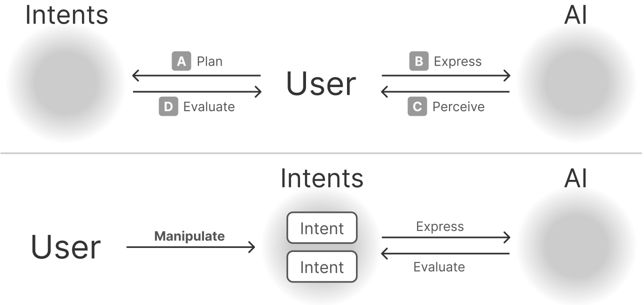


Figure 1: Top. Traditional interaction path through which users interact with AI through their intents: (a) the user plans their intents, (b) expresses these to the AI, (c) perceives or parses the AI's output, and (d) then evaluates whether the output aligns with their intents. Bottom. My thesis explores the concept of *intent manipulation* where, instead of relying on a fuzzy internal representation of their own intents, interfaces externalize user intents into objects and support novel interactions on or through these objects. My thesis investigates how this can support users' exploration, iteration, and sensemaking of AI behaviors and outcomes.

1 Introduction

State-of-the-art AI models have shifted the paradigm for interactions between users and computers. Instead of translating their high-level intents into low-level operations that a computer should perform, users can now simply state their intents [33]. Through advancements in natural language (NL) understanding and instruction following capabilities, recent models can perform tasks from user's high-level, NL inputs. For example, models can generate music [2, 9], images [32, 36], and even videos [6, 51] from user's descriptions. More advanced models, like Large Language Models (LLMs), possess general purpose capabilities that allow users to perform ever more complex and unique tasks, from writing full research papers [29] to developing interactive applications [34]. This paradigm shift expands the space of possibilities of what one can achieve through computers and how one should interact with the computer to reach these possibilities [44].

The path to these possibilities, however, becomes *fuzzier* in this new paradigm. Consider the interaction path (Fig. 1) where a user plans their intents, executes action based on this intent, perceives the result from the model, and then evaluates whether these results align with their intents:

- **Plan** (Fig. 1a): As users employ models for more *ill-defined* tasks without clear goals but lack awareness on the models' capabilities [41, 43], the user's intents may not be fully formed at the start.
- **Express** (Fig. 1b): Even with fully formed intents, users may fail to successfully express their intents as these models are sensitive to inputs formats [30, 48], but provide no affordances on how to write these inputs [41].
- **Perceive** (Fig. 1c): As these models are stochastic (i.e., generate diverse outputs for the same input [25, 41, 44]), users must inspect outputs to understand the result of their inputs. As outputs can be challenging to parse and process (e.g., long-form text), this is cognitively demanding [21, 43].

CHI EA '25, April 26-May 1, 2025, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3707611

- **Evaluate** (Fig. 1d): Without fully formed intents, users may be uncertain about how to evaluate outputs as the standards or criteria that these outputs should follow are also uncertain [41].

In my thesis, I explore **interaction techniques to navigate these fuzzy paths to their intended possibilities with these models**. Namely, my work explores the concept of ***intent manipulation*** (Fig. 2): representing user intents into *first-class objects* in interfaces and supporting novel interactions with these intents. This concept takes inspiration from *direct manipulation*, which suggests continuous representation of the objects of interest and supporting rapid incremental reversible actions on these objects [18, 40]. As interaction with recent AI models requires extensive operation with one's intents (e.g., planning, expressing, evaluating), my work proposes that interfaces should consider intents to be main objects of interest and provide interaction mechanisms that facilitate acting on and through these objects. Further, I take inspiration from the design process [10, 13, 37], where externalizing and iterating on designs allows one to explore and *illuminate* the problem space (i.e., what should be done, intents) and the solution space (i.e., how it should be done, outcomes). My work proposes that, besides the final model outcomes, intents themselves can be considered to be design artifacts that the user has to explore and iterate on—requiring interfaces to support these interactions.

Through my research, I explore various forms of *intent manipulation* by proposing novel representations of intents and interactions on these intents—aiming to support users in interacting with recent AI models in various tasks. First, I explore *intent curation* [19] (Fig. 2a): a user's high-level and abstract intent is disentangled into a *palette* of diverse but plausible low-level operations, which the user can explore and select to curate their desired outcome. Second, I propose *intent assembly* [20] (Fig. 2b): intents are represented as *blocks* that users can assemble and re-assemble into more complex intents. Third, I propose *intent framing* [21] (Fig. 2c): users can create multiple *lenses* that represent distinct intents, which the user can put on and switch around to inspect how model outputs align with these intents. In future work, I aim to develop this concept further by investigating more *complex intents* that entail comprehensive workflows rather than singular tasks, and exploring manipulation of *implicit intents in long-term human-AI interactions.*

2 Research Methods and Contributions

I am primarily a systems researcher. I design and develop novel interactive systems that support users to interact with state-of-the-art AI models to fulfill their intents in diverse tasks (e.g., design, writing). To drive these systems, I also design and implement AI-based techniques and computational pipelines. To evaluate these systems, I conduct mixed-method user studies and technical evaluations for the underlying techniques and pipelines. Beyond these methods shared across my work, I also construct datasets to train and evaluate pipelines [19], conduct interviews and workshops with practitioners to gain in-depth qualitative insights [20, 21], and propose design frameworks to guide system design [20].

2.1 Stylette: Intent Revision for Website Styling (Figure 2a)

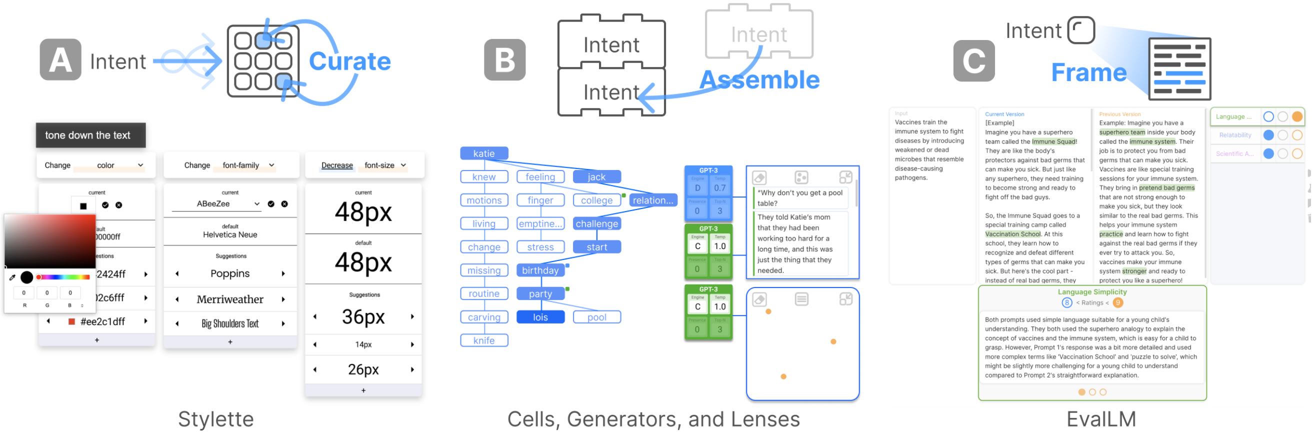
Despite the inherent *malleability* of the web, end-users may lack the ability to modify the look and form of websites to their needs. End-users are frequently novices in terms of both design and programming knowledge. As a result, they may struggle to decompose their high-level intents into specific low-level modifications [1, 23] and to then translate these modifications into the underlying coding language [26]. To understand how novices want to modify websites and how they express these intents, I conducted novice-expert sessions where participants verbalized desired changes to a website's design and an expert performed these changes in real-time. Beyond the challenges from prior work, this study revealed that novices can be purposefully vague and abstract as their intents were not fully formed and they wanted to explore diverse options before refining their intents.

To support this, I designed Stylette [19], a natural language interface that allows users to change website styles by clicking components and verbally expressing their intents (e.g., "more modern"). Instead of directly editing the component, Stylette instead encodes the user's intent into a *styling palette* that provides an assortment of style properties (e.g., font family) and values (e.g., Helvetica) that *could* align with the user's intent. This palette allows users to directly manipulate and experiment with style changes to explore the design space and revise their intent. To create these palettes, I designed a computational pipeline that uses an LLM with prompt tuning [28] to process the user's intent to infer relevant style properties, and a Variational Autoencoder (VAE) model [22] to process the component and suggest the style values of similar components in a large-scale dataset. A comparative user study demonstrated that Stylette helped participants to fulfill their intents more quickly and successfully, while also enabling them to experiment more with styling properties.

2.2 Cells, Generators, and Lenses: Intent Assembly for Writing (Figure 2b)

LLMs enable users to write without actually writing. Users can delegate the effort of actually producing the text to these models, and employ them to facilitate other writing adjacent tasks (e.g., ideation [7], editing [14], reflection [11]). Despite these models' benefits, users must dedicate significant effort in iteratively configuring the models' inputs and parameters until they return outputs that align with their intents [12, 24]. However, as existing interfaces only provide a single input area and a single set of model parameters, every change overwrites previous configurations, which adds friction to experimentation and prohibits parallel testing and reuse of previous ideas—inhibiting iteration, overall.

To address these limitations, I proposed that a paradigm shift was needed in the design of LLM-powered writing interfaces. Specifically, I proposed a framework for designing interfaces that support object-oriented interaction with LLMs through *cells, generators, and lenses* [20]. My framework suggests that interfaces should enable users to express their intents and to reify [5] these into persistent objects: *cells* that represent model inputs, *generators* that contain model parameters, and *lenses* that encompass spaces that represent and visualize model outputs. By creating multiple of these objects



"Why don't you get a pool table?

They told Katie's mom that they had been working too hard for a long time, and this was just the thing that they needed.

Vaccines train the immune system to fight diseases by introducing weakened or dead microbes that resemble disease-causing pathogens.

[Example]
Imagine you have a superhero team called the Immune Squad! They are like the body's protectors against bad germs that can make you sick. But just like any superhero, they need training to become strong and ready to fight off the bad guys.

So, the Immune Squad goes to a special training camp called Vaccination School. At this school, they learn how to recognize and defeat different types of germs that can make you sick. But here's the cool part - instead of real bad germs, they

Example: Imagine you have a superhero team inside your body called the immune system. Their job is to protect you from bad germs that can make you sick. Vaccines are like special training sessions for your immune system. They bring in pretend bad germs that are not strong enough to make you sick, but they look similar to the real bad germs. This helps your immune system practice and learn how to fight against the real bad germs if they ever try to attack you. So, vaccines make your immune system stronger and ready to protect you like a superhero!

Both prompts used simple language suitable for a young child's understanding. They both used the superhero analogy to explain the concept of vaccines and the immune system, which is easy for a child to grasp. However, Prompt 1's response was a bit more detailed and used more complex terms like 'Vaccination School' and 'puzzle to solve', which might be slightly more challenging for a young child to understand compared to Prompt 2's straightforward explanation.

Figure 2: Each work in my thesis explores a different actualization of how *intent manipulation* can support users when interacting with AI models: (a) Stylette [19] represents intents as palettes that the user can curate their intent with, (b) Cells, Generators, and Lenses [20] proposes a framework for interfaces that represent intents as blocks that the user can assemble into diverse configurations, and (c) EvalLM [21] represents intents as lenses that the user can use to distinctly frame and inspect model outputs.

and assembling and reassembling them into various configurations, users can more rapidly iterate and experiment with both the AI model and also with their own intended ideas. I demonstrate the framework's (1) *generalizability* by developing three interfaces that support distinct writing tasks, (2) *effectiveness* through a user study that showed that participants could experiment with more configurations and explore more outputs with our framework, and (3) *usability* through a workshop with interface designers that showed that our framework bootstrapped and inspired designers to better support user iteration.

2.3 EvalLM: Intent Framing for Prompt Engineering (Figure 2c)

General purpose AI models have catalyzed the creation of a wide array of novel applications. For example, by composing a prompt for an LLM, designers and developers can guide the model to perform new tasks to power these applications. Designing these successful prompts (i.e., prompt engineering [45]) multiple iterations of revising the prompt, testing it with diverse samples, and evaluating the model's outputs to verify whether they align with the designer's intents [27, 48, 49]. The evaluation step, in particular, incurs significant cognitive load. As the considered tasks are novel and frequently subjective, no automated metrics may exist that can adequately assess performance and designers frequently need to manually inspect the outputs themselves [8, 21].

To facilitate evaluation and iteration during prompt engineering, I proposed EvalLM [21], an interactive system that supports evaluation of LLM outputs through user-defined criteria. In EvalLM, users can simply write criteria that describe the requirements or standards that they intend the LLM. Through an adaptation of the *LLM-as-a-judge* technique [47, 50], an LLM-based evaluation assistant then evaluate each output on its performance on each criterion. The system intends to allow users to inspect and make sense of outputs through the lens of their intents, which are reflected in their criteria. Specifically, the system provides summaries of criterion-wise performance and highlights fragments in each output relevant to each criterion. Furthermore, to help users refine their intents and criteria, the system provides a criteria review tool that suggests how to decompose, merge, or revise criteria. A comparative user study revealed that EvalLM helped users to gain a high-level understanding of their prompt's performance, assisted them in focusing their own inspections of outputs, and supported iterative cycles through which they co-evolved their prompts and criteria.

3 Future Directions

To further my thesis, I plan to extend the concept of *intent manipulation* to encapsulate more complex and implicit intents.

3.1 Intents as Functions to Agents

The scope of the intents covered by my prior work resemble *software functions*: perform a singular task or action. These models, however, can also be employed to drive more complex "programs": *agents* [15] that can autonomously perform dynamic workflows composed of multiple tasks in diverse situations [39, 46, 52]. While designing personal agents could augment users' own workflows, designing these agents entails a complex workflow in itself: planning how the agent should act in diverse situations, translating this plan into model inputs, and testing behaviors in these situations [16, 31, 35]. To facilitate user-driven agent creation, my ongoing work aims to build a system that supports users to design agents through *intent augmentation and navigation*. As it is challenging for users to design every agent behavior for each situation, I will explore techniques to augment users' feedback—expressed for limited behaviors and situations—into generalizable principles and criteria that guide and verify agent's behavior, respectively. Then, to support scalable oversight of the agent [3], I plan to design interactive visualization techniques that leverages the user's intents (i.e., principles and criteria) as *anchors* to navigate, examine, and refine the agent's demonstrated behaviors in diverse situations.

3.2 Implicit Intents from Long-Term Interactions

To date, my research has proposed approaches that allow users to manipulate intents that they have explicitly expressed. When interacting with AI models, users also possess intents that are implicit, which the user forgets to verbalize or expects the model to infer [42]. In human-human interaction, we learn to infer others' intents by developing representations of their mental states (e.g., beliefs, desires) from prior observations—a process referred to as *"theory-of-mind"* or *"mind reading"* [4, 17, 38]. Similarly, if AI models could infer user's mental states from prior interactions, this could enable them to infer their intents in the future. In future work, I plan to investigate this capability in LLMs. Through this, I aim to design a system where a user interacts with an LLM while the model continuously builds and updates a representation of the user's inferred mental states. Beyond using this representation to guide the model's behaviors in future interactions, the system can enable users to interactively inspect and manipulate this representation, serving as a mechanism for explainability and controllability.

4 Dissertation Status and Long-Term Goals

I am currently a fourth-year PhD student in the School of Computing at KAIST in the Republic of Korea, advised by Professor Juho Kim. I have completed all required coursework and have passed our program's qualification exam. The expected completion date of my PhD studies is Spring 2026. Upon graduation, I plan to seek research positions in industry where I can continue investigating how to empower users to interact with and fulfill their goals through state-of-the-art (and future) AI models.

Acknowledgments

I am deeply grateful to my advisor, Juho Kim, for his invaluable support and guidance throughout this journey. I would also like to thank all of my mentors, collaborators, and labmates for their time, encouragement, and insights. This work was supported by NAVER-KAIST Hypercreative AI Center. This work was also supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-01347,Video Interaction Technologies Using Object-Oriented Video Modeling).

References

[1] Eytan Adar, Mira Dontcheva, and Gierad Laput. 2014. CommandSpace: modeling the relationships between tasks, descriptions and features. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 167–176.

[2] Andrea Agostinelli, Timo I Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. 2023. Musiclm: Generating music from text. *arXiv preprint arXiv:2301.11325* (2023).

[3] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565* (2016).

[4] Chris L Baker, Rebecca Saxe, and Joshua B Tenenbaum. 2009. Action understanding as inverse planning. *Cognition* 113, 3 (2009), 329–349.

[5] Michel Beaudouin-Lafon and Wendy E Mackay. 2000. Reification, polymorphism and reuse: three principles for designing visual interfaces. In *Proceedings of the working conference on Advanced visual interfaces*. 102–109.

[6] Tim Brooks, Bill Peebles, Connor Holmes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, Clarence Ng, Ricky Wang, and Aditya Ramesh. 2024. Video generation models as world simulators. (2024). https://openai.com/research/video-generation-models-as-world-simulators

[7] 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*.

[8] Elizabeth Clark, Tal August, Sofia Serrano, Nikita Haduong, Suchin Gururangan, and Noah A Smith. 2021. All that's' human'is not gold: Evaluating human evaluation of generated text. *arXiv preprint arXiv:2107.00061* (2021).

[9] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2024. Simple and Controllable Music Generation. arXiv:2306.05284 [cs.SD] https://arxiv.org/abs/2306.05284

[10] Nigel Cross. 1982. Designerly ways of knowing. *Design studies* 3, 4 (1982), 221–227.

[11] 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.

[12] 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.

[13] Steven P Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L Schwartz, and Scott R Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 4 (2010), 1–24.

[14] Wanyu Du, Zae Myung Kim, Vipul Raheja, Dhruv Kumar, and Dongyeop Kang. 2022. Read, revise, repeat: A system demonstration for human-in-the-loop iterative text revision. *arXiv preprint arXiv:2204.03685* (2022).

[15] Stan Franklin and Art Graesser. 1996. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In *International workshop on agent theories, architectures, and languages*. Springer, 21–35.

[16] Nitesh Goyal, Minsuk Chang, and Michael Terry. 2024. Designing for Human-Agent Alignment: Understanding what humans want from their agents. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–6.

[17] Mark K Ho, Rebecca Saxe, and Fiery Cushman. 2022. Planning with theory of mind. *Trends in Cognitive Sciences* 26, 11 (2022), 959–971.

[18] Edwin L Hutchins, James D Hollan, and Donald A Norman. 1985. Direct manipulation interfaces. *Human–computer interaction* 1, 4 (1985), 311–338.

[19] 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* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 5, 17 pages. https://doi.org/10.1145/3491102.3501931

[20] Tae Soo Kim, Yoonjoo Lee, Minsuk Chang, and Juho Kim. 2023. Cells, Generators, and Lenses: Design Framework for Object-Oriented Interaction with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 4, 18 pages. https://doi.org/10.1145/3586183.3606833

[21] Tae Soo Kim, Yoonjoo Lee, Jamin Shin, Young-Ho Kim, and Juho Kim. 2024. EvalLM: Interactive Evaluation of Large Language Model Prompts on User-Defined Criteria. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 306, 21 pages. https://doi.org/10.1145/3613904.3642216

[22] Diederik P Kingma. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[23] Gierad P Laput, Mira Dontcheva, Gregg Wilensky, Walter Chang, Aseem Agarwala, Jason Linder, and Eytan Adar. 2013. Pixeltone: A multimodal interface for image editing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2185–2194.

[24] 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.

[25] Yoonjoo Lee, Kihoon Son, Tae Soo Kim, Jisu Kim, John Joon Young Chung, Eytan Adar, and Juho Kim. 2024. One vs. Many: Comprehending Accurate Information from Multiple Erroneous and Inconsistent AI Generations. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 2518–2531. https://doi.org/10.1145/3630106.3662681

[26] 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.

[27] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35.

[28] Xiao Liu, Yanan Zheng, Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang. 2023. GPT understands, too. *AI Open* (2023).

[29] Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. 2024. The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. arXiv:2408.06292 [cs.AI]

[30] Yao Lu, Max Bartolo, Alastair Moore, Sebastian Riedel, and Pontus Stenetorp. 2021. Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. *arXiv preprint arXiv:2104.08786* (2021).

[31] Qianou Ma, Weirui Peng, Hua Shen, Kenneth Koedinger, and Tongshuang Wu. 2024. What you say= what you want? Teaching humans to articulate requirements for LLMs. *arXiv preprint arXiv:2409.08775* (2024).

[32] Midjourney. 2024. Midjourney. https://www.midjourney.com/.

[33] Jakob Nielsen. 2023. AI: First New UI Paradigm in 60 Years. https://www.nngroup.com/articles/ai-paradigm/.

[34] OpenAI. 2024. Introducing OpenAI o1-preview. https://openai.com/index/introducing-openai-o1-preview/.

[35] Savvas Petridis, Benjamin D Wedin, James Wexler, Mahima Pushkarna, Aaron Donsbach, Nitesh Goyal, Carrie J Cai, and Michael Terry. 2024. ConstitutionMaker: Interactively Critiquing Large Language Models by Converting Feedback into Principles *(IUI '24)*. Association for Computing Machinery, New York, NY, USA, 853–868. https://doi.org/10.1145/3640543.3645144

[36] 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.

[37] Donald A Schön. 2017. *The reflective practitioner: How professionals think in action*. Routledge.

[38] Karen Shanton and Alvin Goldman. 2010. Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science* 1, 4 (2010), 527–538.

[39] Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2024. Hugginggpt: Solving ai tasks with chatgpt and its friends in hugging face. *Advances in Neural Information Processing Systems* 36 (2024).

[40] Ben Shneiderman. 1983. Direct manipulation: A step beyond programming languages. *Computer* 16, 08 (1983), 57–69.

[41] Hari Subramonyam, Roy Pea, Christopher Pondoc, Maneesh Agrawala, and Colleen Seifert. 2024. Bridging the Gulf of Envisioning: Cognitive Challenges in Prompt Based Interactions with LLMs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.

[42] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The metacognitive demands and opportunities of generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–24.

[43] Michael Terry, Chinmay Kulkarni, Martin Wattenberg, Lucas Dixon, and Meredith Ringel Morris. 2024. Interactive AI Alignment: Specification, Process, and Evaluation Alignment. arXiv:2311.00710 [cs.HC] https://arxiv.org/abs/2311.00710

[44] Justin D Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. 2024. Design Principles for Generative AI Applications. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.

[45] 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.

[46] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629* (2022).

[47] Seonghyeon Ye, Doyoung Kim, Sungdong Kim, Hyeonbin Hwang, Seungone Kim, Yongrae Jo, James Thorne, Juho Kim, and Minjoon Seo. 2023. Flask: Fine-grained language model evaluation based on alignment skill sets. *arXiv preprint arXiv:2307.10928* (2023).

[48] 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. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 2206–2220.

[49] 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.

[50] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems* 36 (2023), 46595–46623.

[51] Zangwei Zheng, Xiangyu Peng, Tianji Yang, Chenhui Shen, Shenggui Li, Hongxin Liu, Yukun Zhou, Tianyi Li, and Yang You. 2024. *Open-Sora: Democratizing Efficient Video Production for All*. https://github.com/hpcaitech/Open-Sora

[52] Shuyan Zhou, Frank F Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Tianyue Ou, Yonatan Bisk, Daniel Fried, et al. 2023. Webarena: A realistic web environment for building autonomous agents. *arXiv preprint arXiv:2307.13854* (2023).