Curompt: A Spatially Situated Interface for Generative AI in 3D Design Software

Mitchell Foo
Autodesk Research
Toronto, Canada
mitchellcfoo@gmail.com

Kendra Wannamaker
Autodesk Research
Toronto, Canada
kendra.wannamaker@autodesk.com

Jo Vermeulen
Autodesk Research
Toronto, Canada
jo.vermeulen@autodesk.com

George Fitzmaurice
Autodesk Research
Toronto, Canada
george.fitzmaurice@autodesk.com

Justin Matejka
Autodesk Research
Toronto, Canada
justin.matejka@autodesk.com

***Abstract*—Generative AI tools are increasingly being integrated into various workflows. Many powerful AI systems, such as ChatGPT, Gemini, and Grok, are accessed through simple chat interfaces. While these language-based interactions may seem intuitive and futuristic, they also bear a resemblance to early computer command lines. With our prototype system, Curompt (combined cursor+prompt), we explore how to integrate conversational interaction powered by generative AI into familiar graphical interfaces with direct manipulation. We focused on the 3D environment, where relying on language-based interactions alone proves to be challenging.**

***Index Terms*—3D design, AI co-creation, Cursor interaction, Large language models, Generative AI**

I. Introduction

While generative AI technologies are solidifying their place in future design workflows, the primary way we interact with these systems is through a floating chat window, where the user types prompts that the AI model responds to. This way of interacting is detached from the context of the design space, and as such, ill-suited to tasks in graphical and spatial domains, such as when generating or manipulating a 3D scene of objects. In this video showcase, we present our explorations to make interfaces for Generative AI systems more spatially situated. We developed the Curompt (combined cursor+prompt) concept and prototype, which deeply integrates the common prompting interaction within the sophisticated graphical interfaces and design canvas that users are already familiar with.

II. Related Work

Some of the first work in computer interaction in natural language, as seen in Put That There [1], conceptualized the potential of semantics and gesturing to work in tandem with natural language commands. Now, with Large Language Models (LLMs) capable of translating ambiguous commands into actions, they can make spatially informed responses when provided structured scene descriptions [4].

With plugins for 3D design software that integrate generative AI capabilities into the workspace [2], [6], these initial interfaces allow users to query LLMs to make edits to and assist with their 3D design. Further research on such interfaces [3], [5] expand on this by connecting prompts to cursor interactions in the design space. This demonstrates the potential of pairing 2D canvas manipulation with a traditional prompt to better express user intent to an LLM performing edits.

III. Curompt

In our video showcase, we show a variety of features and 3D interaction enabled by our Curompt interaction within a 3D environment. We demonstrate instances of using AI prompting within a 3D design canvas, pairing intuitive cursor interactions with text-based prompting. As with a regular 3D design canvas, users can directly move around the space, manipulate objects, and undo actions. To interact with the generative AI, users can select any number of objects within the scene, which will make a text-box appear connected to all the selections. Once the user enters a prompt, the system converts the scene context into a JSON scene description that the LLMs can understand and decide which model and functions are most appropriate for the request (ChatGPT or DALL-E).

When users hover over 3D objects, a floating indicator appears. We use a "$>$" symbol to signal creating a new Curompt and a speech bubble with the number of messages to indicate revisiting a prior Curompt for that object. Users can also create a Curompt based on selecting multiple objects, selecting a target based on a 3D point in space, or a general Curompt without a selected object. Lastly, users can manually manipulate 3D objects between AI-assisted edits. We demonstrate in the video how the design of these interactions are paired with implicit text inputs that enable novel collaborations between the user and AI assistant.

The models, without any fine-tuning or specialized knowledge, can modify the scene context in several compelling ways, including adding, swapping, and arranging objects within the scene, changing the properties of objects (size, material, and color), and generating new textures, as shown in

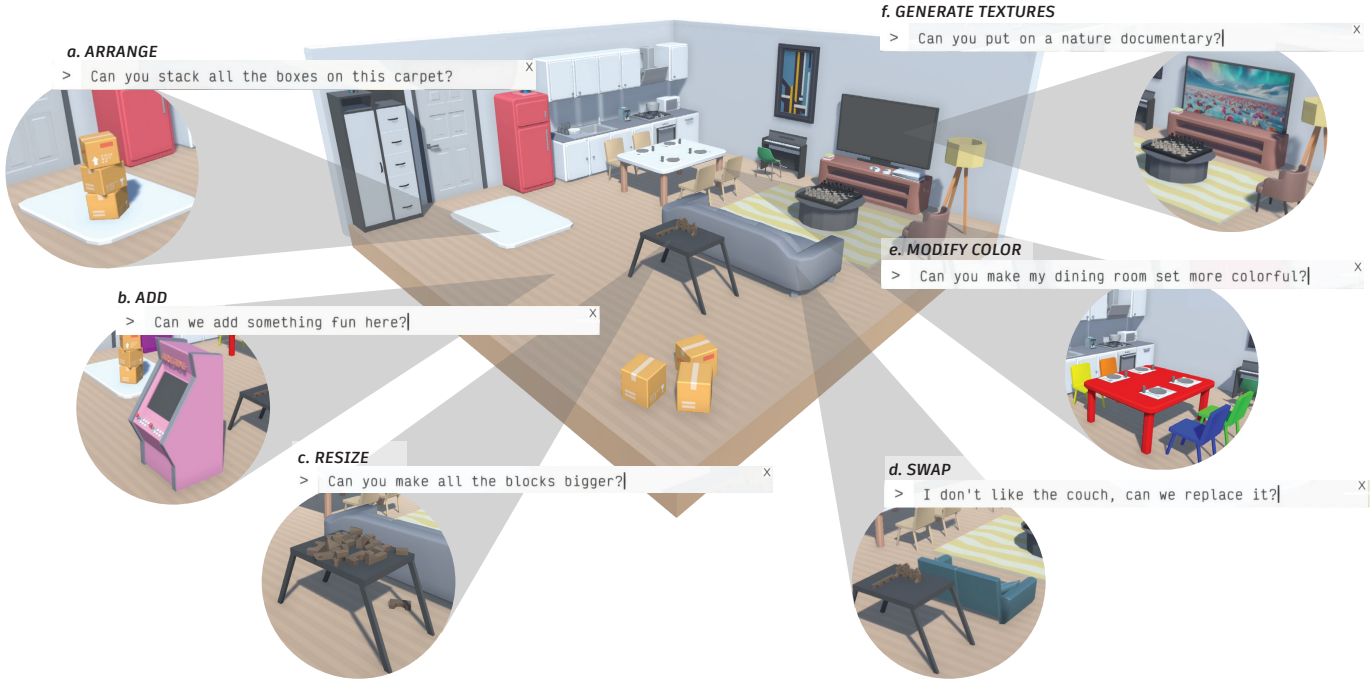

Fig. 1. The Curompt prototype explores spatially situated generative AI-assisted interactions, as shown here in a 3D scene of an apartment. Users can enact spatial scene edits with high-level descriptions with a *Curompt*, a combined cursor+prompt — such as (a) arranging; (b) adding; (c) resizing; (d) swapping; (e) modifying the color of; and (f) generating textures for objects in the scene.

Figure 1. In our video showcase, we demonstrate the system's ability to convert both hyper-specific directions and open-ended requests into a series of changes within the scene. To maintain understandability and editable history, the assistant writes out what it is changing, each object has its own chat log, and the system maintains a global timeline that can be used to revert any AI action.

IV. Conclusion

Our video showcase for Curompt introduces a novel interface for integrating AI assistants spatially contextualized inside a designer's familiar 3D design canvas, promoting open-ended interactions and explorations for human-AI co-creation. We aim to present novel visions for potential new workflows and interfaces that empower 3D designers rather than replace their creative decision-making. We believe strongly that future interactions can balance traditional software interfaces, like physically navigating with a cursor, with implicit natural-language interfaces powered by AI.

In future work, we would like to explore integrating cursor prompts within more complex interaction patterns common in 3D design. We envision that these interaction techniques will also help facilitate synchronous and asynchronous collaborations between multiple human designers and AI assistants.

References

[1] Richard A Bolt. Put-that-there: Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, pages 262–270. ACM, 1980.

[2] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 3dall-e: Integrating text-to-image ai in 3d design workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 1955–1977, New York, NY, USA, 2023. Association for Computing Machinery.

[3] Zhaoyang Liu, Yinan He, Wenhai Wang, Weiyun Wang, Yi Wang, Shoufa Chen, Qinglong Zhang, Zeqiang Lai, Yang Yang, Qingyun Li, Jiashuo Yu, et al. Interngpt: Solving vision-centric tasks by interacting with chatgpt beyond language. *arXiv preprint arXiv:2305.05662*, 2023.

[4] Rui Ma, Akshay Gadi Patil, Matthew Fisher, Manyi Li, Sören Pirk, Binh-Son Hua, Sai-Kit Yeung, Xin Tong, Leonidas Guibas, and Hao Zhang. Language-driven synthesis of 3d scenes from scene databases. In *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*. ACM, 2018.

[5] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. Directgpt: A direct manipulation interface to interact with large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.

[6] NVIDIA Omniverse. Kit extension sample: Ai room generator. https://github.com/NVIDIA-Omniverse/kit-extension-sample-airoomgenerator, 2023. Accessed: 2025-01-20.