The Design Space of Generative Models

Meredith Ringel Morris
Google Research
Seattle, WA, USA
merrie@google.com

Carrie J. Cai
Google Research
Mountain View, CA, USA
cjcai@google.com

Jess Holbrook[*]
Meta
Seattle, WA, USA
jessholbrook@fb.com

Chinmay Kulkarni
Google Research
Atlanta, GA, USA
ckulkarni@google.com

Michael Terry
Google Research
Cambridge, MA, USA
michaelterry@google.com

Abstract

Card *et al.'s* classic paper "The Design Space of Input Devices" [4] established the value of design spaces as a tool for HCI analysis and invention. We posit that developing design spaces for emerging pre-trained, generative AI models is necessary for supporting their integration into human-centered systems and practices. We explore what it means to develop an AI model design space by proposing two design spaces relating to generative AI models: the first considers how HCI can impact generative models (i.e., interfaces for models) and the second considers how generative models can impact HCI (i.e., models as an HCI prototyping material).

1 Introduction

Among the most significant advances in machine learning in recent years are the introduction of *pre-trained, generative AI models* (sometimes called "foundation models" [1]); these general-purpose models are trained on large data sets and can generate novel text (e.g., [2], [11]), imagery (e.g., [9], [10]), or other media that is in many cases indistinguishable in quality from human-created content [5]. Because of the relative ease with which they can be customized and controlled (e.g., through text-based "prompts"), these models offer a number of potential benefits to human-centered computing, particularly for the field of Human-Computer Interaction (HCI). More specifically, generative models have the potential to revolutionize design and evaluation methodologies for a wide variety of interactive systems, as well as to support novel interaction paradigms (e.g., generative language models' potential to support fast, accurate communication for people with mobility [3] and/or cognitive [6] disabilities).

In this position paper, we propose two preliminary "design spaces" [4]: one for conceptualizing how HCI may impact the burgeoning field of generative model research, and a second for theorizing how generative models may impact HCI practices. We do not argue that these proposed design spaces are definitive or complete; rather, we hope that these serve as useful artifacts for reflection, discussion, and debate during the NeurIps 2022 Workshop on Human-Centered AI.

2 Design Space for Interfaces to Generative Models

Our first design space considers the application of HCI to generative AI models. The interfaces designed for interacting with such models have the potential to influence who is able to use the models

---

[*]work done while employed at Google Research

(e.g., ML engineers vs. general users), model safety (e.g., supporting identification or interrogation of undesired outputs), and model applications (e.g., by altering the time and effort involved in prompt engineering). Our proposed taxonomy (Table 1) considers both the design possibilities for interfaces for providing input to models and those for presenting model outputs.

Input Dimensions				
Media	language	image	video	audio	multimodal
Format	free-form	restricted choice	templated	edits	
Explainability	input tokenization	input frequency	blocklists		
Evolution	prompting	prompt-tuning	fine-tuning	retraining	

Output Dimensions				
Media	language	image	video	audio	multimodal
Product	artifact (*text*, *image*, etc.)	customized model	prompt(s)		
Timing	asynchronous	real-time			
Operation	classification	completion	creation	information extraction	transformation
Interactions	quality control (*bug reporting*, *safety reporting*)	input iteration	query		
Explainability	link output to input	link output to training data	metadata embedding	transparency (*constrained decoding*, *safety filtering*)	
Format	detail (*safety warnings*, *confidence values*, etc.)	quantity (*top-1*, *top-n*)	stylistic choices (e.g., *anthropomorphization*)	standardization	

Table 1: Our proposed design space for interfaces to generative models; the first section enumerates the range of design parameters for input interfaces to models, while the second section demonstrates the range of design possibilities for presenting model outputs.

2.1 Design Space for Model Inputs

When considering model inputs, different *media* are possible; note that the input media and the output media may not necessarily match (e.g., the use of language prompts to generate images in status quo systems like DALL-E [9] and Imagen [10], though as our design space suggests, one can imagine alternative media input to output pairings such as submitting a sketch or musical score as input to generate an image).

Varying input *formats* may also be suitable for different purposes – while many of today's language models (e.g., [2], [11]) allow free-form input in the form of an open-ended textbox, alternatives might include restricting choices via a menu (e.g., similar to many customer service chatbots or as done in the original Dall-E announcement blog[2]) or offering a template to help make the many potential parameters of complex prompts more apparent to novice users. Editing an existing artifact or collection of artifacts (including those previously output by the system) may be another way to

---

[2]https://openai.com/blog/dall-e/

specify model inputs (e.g., circling the portions of a generated image that did not meet the user's initial expectations).

While status quo interfaces to models offer little or no feedback to end-users, future tools might provide *explainability* features for model inputs, such as providing the end-user with information about input tokenization, frequency, block-listed items, etc., which might grow the user's mental model of the system and support improved utilization.

Finally, our taxonomy accounts for the ways in which the input might support the *evolution* of the model itself, whether via prompting, prompt-tuning, fine-tuning, or retraining.

2.2   Design Space for Model Outputs

When considering model outputs, in addition to generating artifacts comprising one or more *media*, our design space also considers other kinds of *products* that may result from the user's interactions; for instance, while GPT-3 might produce a paragraph of text as an artifact, end-users may be equally or more interested in the customized model created via a given prompting strategy, or in the prompt(s) used to produce a given artifact.

Timing of outputs may also be varied; while real-time results are necessary for supporting many interactions and experiences, asynchrony may be necessary for computationally expensive models (e.g., video generation), yet may offer new design opportunities (e.g., previewing or giving feedback on partial outputs).

Interfaces for model output may also account for the *operation* performed by the model; common operations include classification (e.g., a language model that can tell you if an input word is or is not a type of food), completion (e.g., a model that can provide a punchline given the setup for a joke), creation (e.g., producing a short story), information extraction (e.g., a question-answering system that allows the user to probe the underlying training data), and transformation (e.g., a model that turns English text into French). Additional fundamental operations are likely to be discovered as new use cases and model capabilities evolve.

Today's interfaces to model outputs tend to be static, but we envision a future where interfaces afford *interactions* with outputs, including allowing the end-user to give quality control feedback (e.g., report a bug or safety issue), to iterate by refining or editing the output and resubmitting it to the model for further processing, or to query the output in order to understand more about why a particular output or portion of an output was generated.

Interactive querying is one possible instantiation of *explainability* - the presence or absence of explainability features (and their manner of presentation to the end-user) should be a key dimension of model output interfaces, including mechanisms for explaining how outputs link to inputs and how they link to training data. Further explainaiblity features might include whether metadata (such as the model version, prompt, and any model parameters) are embedded in resultant artifacts, and whether the system offers transparency into any filtering stages that may impact output (e.g., constrained decoding, safety filtering, etc.).

Finally, there are many options for outputs' *format* to the end-user. This includes stylistic choices, such as whether to employ anthropomorphization (as is common in many chatbots). It also includes concerns around the level of detail to present to the end-user, such as whether and how to present safety warnings, confidence metrics, or other details related to the inner workings of the model. Additionally, the choice of the quantity of alternative outputs produced (e.g., is the user shown a single, canonical output or a set of n to choose among) can greatly impact the user's experience and mental model regarding the determinism of the system. Finally, as the variety of available models continues to proliferate, it is important to consider *standardization* of output formats, such that the output of one model might be consumed by other models (e.g., to support pipelining between models, and/or to other related tools).

2.3   Applying the HCI-for-Models Design Space

As an example of how this taxonomy can be used to describe particular interfaces, consider Prompt-Maker [7], an application intended to support prompt programming. Along the *input* dimensions, PromptMaker accepts language (text) as input (media) in a free-form format, and provides explicit

3

support for both zero- and few-shot prompting. It provides no explainability capabilities or evolution of the prompts (an oversight our taxonomy might make visible to the PromptMaker creators, perhaps suggesting an area for future development!). Along the *output* dimensions, PromptMaker produces language (text) as output (media), where this text is the final artifact (product). It is real-time (timing), and can support a number of types of operations (classification, completion, creation, information extraction). Built into the interface are safety reporting mechanisms (interactions) and safety warnings about the potential for undesirable content (presentation).

In addition to being descriptive, our taxonomy can also be generative, identifying new opportunities for design and research. For example, status quo model interfaces typically produce a static output, but *interactive outputs*, such as those that support input iteration, could dramatically change the pace at which people are able to experiment with and create content using generative models. Altering the *media* used as input could also afford novel experiences - sketch-based, photograph-based, or even audio inputs rather than natural-language strings could all afford new means of expression and creativity with generative image models.

Many considerations, including the use context and target end-users, will factor into how to select among these design parameters when developing interfaces for generative models. While some dimensions may seem upon first glance to be inherently "good" or "bad," such choices are not always clear-cut. For instance, while there are many risks in inappropriate or undisclosed *anthropomorphization* of AI, this may nevertheless be a desirable interface for some customer service or entertainment applications. On the other hand, while *explainability* is generally viewed positively, consider the scenario of whether to include the *output explainability* feature of metadata embedding in a generative image model. If the use context is in an application for professional artists creating original works, embedding the prompt as metadata may be undesirable, since the artist may view the prompt as their intellectual property. However, if the use context is in an application for generating custom clip-art for use in documents and slide decks, then embedding the prompt as metadata may be highly beneficial, since it could function as alternative text for screen readers, thereby ensuring that documents produced with this clip art are accessible to blind end-users.

3 Design Space for Generative Models as an HCI Prototyping Material

We believe that the increasing variety and power of generative models will fundamentally change HCI research and practice by enabling a new generation of assets, tools, and methods. These artifacts will impact the full spectrum of HCI, including new tools to support design through creative ideation [13], new tools to support building rapid prototypes of novel, interactive experiences [6], and new tools to support evaluation of interfaces and ecosystems, such as through simulation [8]. Table 2 offers a starting point for such a design space.

Dimension	Design choices		
Goal	capture rich context of use ("Field" [14])	co-create with users ("Studio" [14])	study specific phenomena ("Lab" [14])
Role of AI	as-is capabilities	stand-in for future model	synthetic data (*simulating input*, *simulating content*)
Context	context-free	minimal context	deep context
Lifecycle	inspiration	building	evaluation
Fidelity	proving	scaling	serving
Media	text	visual media (*images*, *video*)	sound

Table 2: Our proposed design space of generative models as a tool for human-computer interaction practices. The shaded cells depict one particular choice, as in [8], of a system that simulates other users to speed up prototyping of a social media platform interface.

3.1 Dimensions of the Models-for-HCI Design Space

In this taxonmy, the *goal* dimension acknowledges that generative models can be used for prototyping different kinds of design explorations. For instance, borrowing terminology from [14], generative models could allow designers to capture how a new AI-based system might behave in the rich

context of the "field," without building it out fully. Examples of prototypes with such a "field" goal include [8], which uses an AI model to simulate the rich social context of the real world in which newsgroups and other feed interfaces are deployed. Instead of simulation, other systems focus on co-creation ("the studio" in [14]'s terminology); for instance, Wordcraft [13] uses a generative model to augment a story-writing interface. In this case, the goal of the prototype is to co-create fiction stories with the user (note how Wordcraft, unlike "social simulacra" [8], cannot fulfill its prototyping goal without real users' input). Finally, generative models may also help prototyping by narrowing the focus of investigation, or studying specific phenomena. One example of this "lab" goal is "AI Chains" [12], which studied whether breaking up the operation of a language model into smaller steps and aggregating them allows for greater human control and transparency.

In using generative models as an HCI prototyping material, the *role of AI* may also vary greatly. Models could be used for their currently known capabilities, as with many AI systems today using GPT-3. They could also represent a less developed version of a yet-to-be-created model, to investigate whether it is worth the effort of creating a specialized model. Finally, they could allow researchers access to synthetic data, for instance by simulating user actions or populating a system with content produced by models.

The *context* dimension captures the idea that generative models can be used to prototype interactions with differing access to context. For instance, models may primarily perform tasks that are essentially context-free (e.g., rewriting a text snippet to be more polite, generating images from text prompts, etc.). They may also work with minimal context (for instance, recommending a movie based on a list of movies the user likes or generating a proposed chat reply based on several turns of conversational history). Finally, generative models can use deep context accumulated over a long period of time, for instance, encoding rhetorical preferences in writing learned over months or years of interaction with a particular author.

The *lifecycle* dimension reflects the range of HCI practices that may be impacted by emerging models. For instance, models could support the early parts of the design process by producing content to inspire creativity (e.g., suggesting design scenarios); they could support the building stage of HCI systems by standing in for a more mature system component such as a fine-tuned or personalized model; or they could support the evaluation stage of HCI work by simulating user-generated content or interaction sequences.

A related aspect is the *fidelity* dimension, which suggests how emerging AI models can be suited to different aspects of HCI practice that require different levels of robustness, speed, safety, or other dimensions of performance. Appropriate fidelity levels may differ for scenarios such as to prove or disprove a design hypothesis, to scale an interaction to more people than possible through other prototyping methods (such as Wizard of Oz), or serving "in production," for users to experience as a completed product.

Finally, the *media* (or combination of multiple media) supported by a model will, of course, shape its suitability and role as an HCI design material.

3.2 Applying the Models-for-HCI Design Space

As with our "interfaces for model inputs & outputs" taxonomy, this second design space can also stimulate research, design, and reflection by supporting both descriptive and generative functions. The shading in (Table 2) reflects a descriptive example, illustrating a classification for Park et al.'s "Social Simulacra" work [8], which used GPT-3 to populate parallel hypothetical futures of proposed social media systems in order to allow social media designers to tweak system features to facilitate desired interaction styles.

This ontology can also suggest directions for future investigation of the potential for generative models to further the field of HCI. For instance, models that generate text could be used in the design stage of the HCI lifecycle to create personas that would inspire a practitioner to consider a wide range of design concepts. Similarly, language models could be used to co-create productivity tools with users, for instance, embedding the context of their work into writing aids in word processors.

4 Conclusion

Card et al. [4] established the value of design spaces and similar ontologies for analysis and invention. In this position paper, we seek to advance discussion and reflection around the marriage of HCI and emerging generative AI models by proposing two design spaces: one for HCI applied to generative models (Table 1) and one for generative models applied to HCI (Table 2). We hope these frameworks spark additional research, both in further refining these taxonomies as well as in spurring the development of novel interfaces for generative models and novel AI-powered design and evaluation tools for HCI broadly.

References

[1]   Rishi Bommasani et al. "On the Opportunities and Risks of Foundation Models". In: *CoRR* abs/2108.07258 (2021). arXiv: 2108.07258. URL: https://arxiv.org/abs/2108.07258.

[2]   Tom B. Brown et al. "Language Models are Few-Shot Learners". In: *CoRR* abs/2005.14165 (2020). arXiv: 2005.14165. URL: https://arxiv.org/abs/2005.14165.

[3]   Shanqing Cai et al. "Context-Aware Abbreviation Expansion Using Large Language Models". In: *Proceedings of NAACL 2022*. 2022. DOI: 10.48550/ARXIV.2205.03767. URL: https://arxiv.org/abs/2205.03767.

[4]   Stuart K. Card, Jock D. Mackinlay, and George G. Robertson. "The Design Space of Input Devices". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '90. Seattle, Washington, USA: Association for Computing Machinery, 1990, pp. 117–124. ISBN: 0201509326. DOI: 10.1145/97243.97263. URL: https://doi.org/10.1145/97243.97263.

[5]   Elizabeth Clark et al. "All That's 'Human' Is Not Gold: Evaluating Human Evaluation of Generated Text". In: *CoRR* abs/2107.00061 (2021). arXiv: 2107.00061. URL: https://arxiv.org/abs/2107.00061.

[6]   Steven Goodman and et al. "LaMPost: Design and Evaluation of an AI-assisted Email Writing Prototype for Adults with Dyslexia." In: *Proceedings of ASSETS 2022*. ACM, 2022. URL: https://arxiv.org/pdf/2207.02308.pdf.

[7]   Ellen Jiang et al. "PromptMaker: Prompt-Based Prototyping with Large Language Models". In: *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI EA '22. New Orleans, LA, USA: Association for Computing Machinery, 2022. ISBN: 9781450391566. DOI: 10.1145/3491101.3503564. URL: https://doi.org/10.1145/3491101.3503564.

[8]   Joon Sung Park et al. *Social Simulacra: Creating Populated Prototypes for Social Computing Systems*. 2022. DOI: 10.48550/ARXIV.2208.04024. URL: https://arxiv.org/abs/2208.04024.

[9]   Aditya Ramesh et al. *Hierarchical Text-Conditional Image Generation with CLIP Latents*. 2022. DOI: 10.48550/ARXIV.2204.06125. URL: https://arxiv.org/abs/2204.06125.

[10] Chitwan Saharia et al. *Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding*. 2022. DOI: 10.48550/ARXIV.2205.11487. URL: https://arxiv.org/abs/2205.11487.

[11] Romal Thoppilan et al. *LaMDA: Language Models for Dialog Applications*. 2022. DOI: 10.48550/ARXIV.2201.08239. URL: https://arxiv.org/abs/2201.08239.

[12] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. "Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts". In: *CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–22.

[13] Ann Yuan et al. "Wordcraft: Story Writing With Large Language Models". In: *27th International Conference on Intelligent User Interfaces*. IUI '22. Helsinki, Finland: Association for Computing Machinery, 2022, pp. 841–852. ISBN: 9781450391443. DOI: 10.1145/3490099.3511105. URL: https://doi.org/10.1145/3490099.3511105.

[14] John Zimmerman and Jodi Forlizzi. "Research through design in HCI". In: *Ways of Knowing in HCI*. Springer, 2014, pp. 167–189.