




Simulating the Human in HCD with ChatGPT Redesigning Interaction Design with AI

 **Albrecht Schmidt**, LMU Munich,
Passant Elagroudy, German Research Center for Artificial Intelligence,
Fiona Draxler, LMU Munich, **Frauke Kreuter**, LMU Munich,
Robin Welsch, Aalto University

Insights

- Generative AI can enhance the human-centered design process.
- LLMs encode human experiences and can be used to emulate users at a large scale.
- Use of generative AI must be made transparent in human-centered design.
- Generative AI will not replace user studies but rather will enhance the toolkit of HCI researchers and practitioners.

Human-centered design (HCD) puts the human at the center of interactive system design. Can we do that without actively including the human user in the process? Is that still HCD? We believe that large language models (LLMs) and generative AI will fundamentally change the way we design and implement interactive systems. Models are not new to HCI, but the scale at which LLMs can support the design process is changing their value and applicability.

The use of models in design decisions for interactive systems has a long tradition in HCI. Task modeling is commonly used in menu design.

There are attempts to model human physiology as well as cognitive processes, and there is an entire journal (<https://www.springer.com/journal/11257>) and conference (UMAP; <https://www.um.org/umap2023/>) devoted to user modeling. Models are helpful since they do not require the user during the design process and thus potentially speed up that process. Creating models, however, has been difficult and cumbersome, which is why human-centered design, which involves people in the process, has been the most common approach to creating usable systems. Putting the

human at the center meant involving people in the design process.

This will significantly change with LLMs. Artificial intelligence that draws from complex statistical distributions of text to predict the next piece of text [1], LLMs are a particularly powerful tool for human-centered development. The hope is that LLMs provide an advantage similar to that of model-based development, hence reducing the need for direct involvement of humans in the design process. As LLMs are readily available, this advantage comes without the effort and cost of creating specific models. The basic idea is that LLMs encode human experiences, which may be drawn upon in design. If we imagine the model includes a representation of all the forums in which people discuss issues of interacting with computing systems, these models should have plenty of information to offer. Assuming the training data of the LLMs is based on large parts of the Internet, including discussion forums, tutorials, product descriptions, scientific papers, handbooks, product reviews, and support websites, they contain information about issues encountered when setting up WiFi on a phone with a suggested solution, reports about problems encountered when installing an update, or errors and solutions when working with a specific software. If we see LLMs conceptually as a massive database of experiences that people have recorded, it should be possible to use their input to replace human feedback—at least in certain parts of the design process.

In this article, we review how LLMs can be used in the design process and where they can replace and augment human input. As we write this, LLMs are continuing to improve. So does our understanding of their potential and how we can use

them. We offer some of our ideas and suggestions regarding where we should consider using LLMs to rethink HCD. Our take is, if we can get information from a model similar to that which we get from humans, it is preferable to use a model and not bother the humans. But if humans offer better insights, we should not use models as shortcuts and end up with suboptimal solutions. We see this as a starting point for a discussion in the community about how we want to advance our design processes and create more usable and more enjoyable interactive experiences.

Our explorations and experience thus far show that there are many areas where LLMs can significantly ease the design and development process. As noted above, there are also areas where we want to continue to involve humans; the most important issue is to be transparent about where we use AI and how we use it. To offer a concrete example, the results for a design requirements analysis may come from a focus group or from an interaction or series of interactions with ChatGPT. There may be good arguments for producing the requirements list using either approach or a combination of both. The approach taken, however, should be made transparent to the customer or stakeholder in a commercial setting or the reader of a paper in the scientific space. Transparency about where and how the insights were gathered is paramount.

RULE 1: *Be transparent and honest about where AI tools were used and provide information on how they augmented HCD or fell short.*

Concretely, LLMs can support HCD in different ways:

- They can replace humans by generating output from their human-informed models.

- They can add an AI agent in addition to humans in iterative and interactive processes.

- They can extend the existing range of HCD methods, for example, by enabling previously unthinkable or unpractical prototyping methods.

LARGE LANGUAGE MODELS AS WORLD MODELS FOR HCD

LLMs make predictions, which in the case of tools like ChatGPT appear as responses to a prompt, that are based on human experiences captured in writing. Several parts of HCD processes are, on an abstract level, questions and answers. For example:

Question: How might we design adaptive notifications that it's time to take medication?

Answer: If notifications are ignored, use different notification styles (e.g., visual, auditory, haptic).

Hence, these models enable a speedy and scalable simulation of the interaction with humans—and at a low cost in time and resources (in comparison with testing human subjects).

LLMs capture written human experiences. The vast amount of data they are trained from represents a substantial share of the world's collective knowledge and experiences in words [2]. This also includes reports of personal experiences at a large scale, for example, comparing good and bad game reviews. LLMs probabilistically tap this experience information from their world model, which they build based on accessible knowledge (e.g., texts on the Internet) when comprehending and generating texts.

The central question in LLMs for HCD processes is how well they capture what we could otherwise learn from interacting with people. Generally speaking, the amount and diversity of the training data mean that many different perspectives are captured. In particular, the data substantiates tendencies in domains where many personal experience reports exist (e.g., consider travel blogs or a Medium post on a widely used Python tool). Even for edge cases and specific needs (e.g., recommendations for activities in a small town, patients with specific conditions), the available data may be sufficient to derive probable outputs that can inform HCD. In many cases, getting users of a

If we see LLMs conceptually as a massive database of experiences, it should be possible to use their input to replace human feedback in certain parts of the design process.

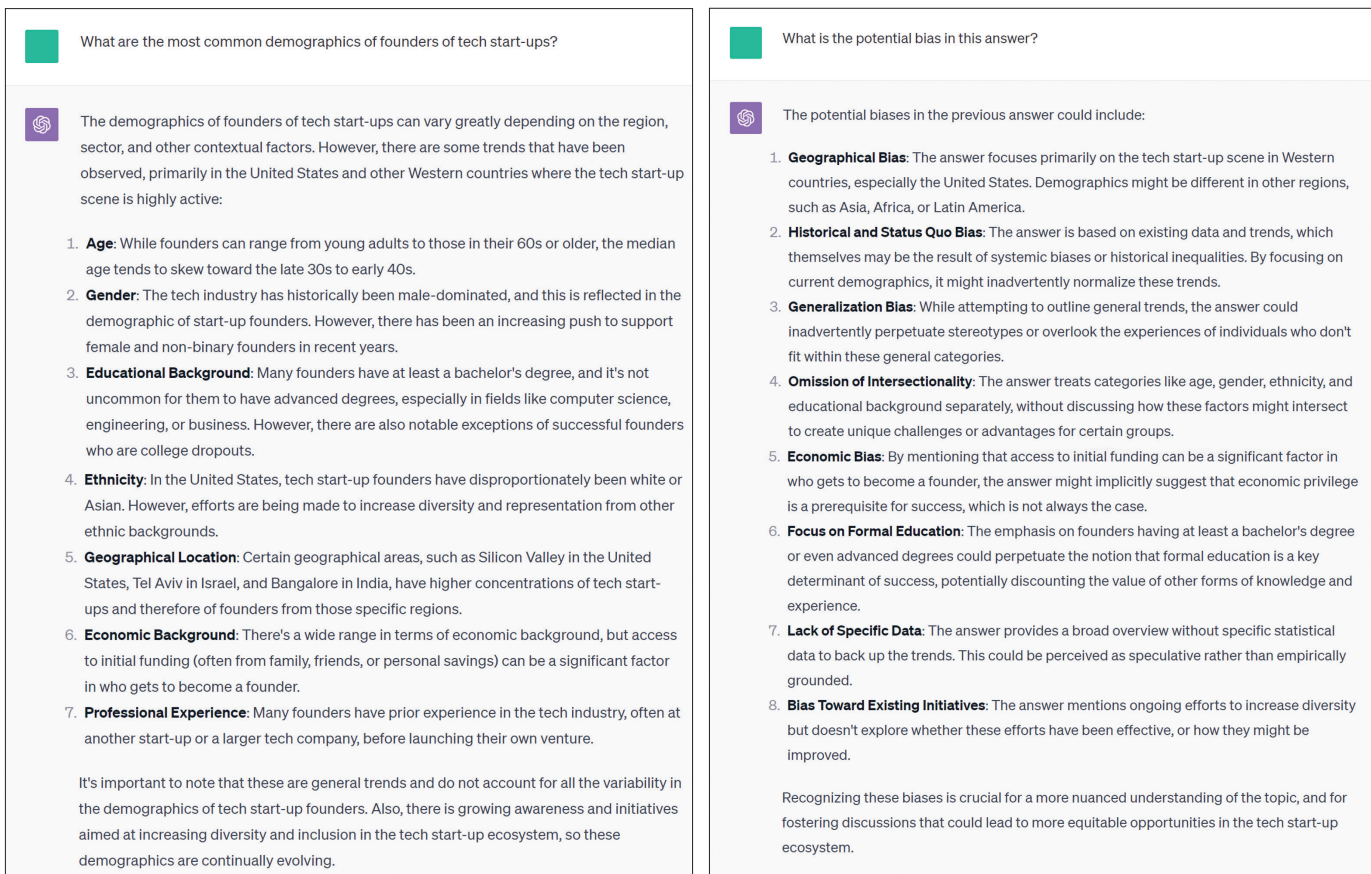


Figure 1. Example of a ChatGPT conversation that asks for information and then inquires about potential bias in the answer.

specific group to take part in user-centered design (UCD) can be hard or even impossible. Most often, however, these same people have shared and discussed their experiences in Internet forums; hence, we may be able to retrieve their views and opinions through the use of LLMs. Such hard-to-get users may include people who consider themselves too busy to take part in a focus group (e.g., for inputs on redesigning the interface of the first-class cabin) or people who do not feel safe in a university environment (e.g., studying how to provide immigrants without a legal status with healthcare).

Nonetheless, applying LLMs/generative AI for HCD comes with certain challenges and caveats, including biases, prompting, and system specifications. Notably, there is a need to understand the consequences of biases inherent to AI. As with many other AI systems, LLMs have been shown to reproduce human biases and stereotypes present in the training data [3]. For example, they may reproduce gender and racial stereotypes. This can be addressed to some extent by specifying a specific person's view (e.g.,

“What requirements for a smartwatch would an older woman with higher education living in a rural area have?” rather than “What are the requirements for a smartwatch?”). However, it is important to be mindful of the output created and check actively for potential biases.

Biases also influence us as users of AI and are giving rise to the field of prompt engineering. Are we capable of identifying biased outputs? Can we prompt LLMs in a sufficiently neutral and open-ended manner to minimize output biases? Or can we insert information in prompts to actively explore and counter biases that we expect (Figure 1)?

RULE 2: *Be aware of limitations and biases. Actively mitigate them in prompts and by checking the examples.*

Furthermore, clear specifications of the target outcome are essential because a vague LLM prompt will yield speculative responses and possibly hallucinations [4]. Few-shot prompting or fine-tuning may be necessary in application scenarios that

are underrepresented in the training data. Similarly, effective role descriptions and background information, such as a document knowledge base, steer outputs into the desired direction. This additional information can be added to the user's prompt programmatically or provided as system messages in ChatGPT. Users also need to evaluate whether a generic or a specific model better fits a concrete goal.

Also, as LLMs start generating more content and potentially interacting with one another, a significant question arises: At what point does the data—and thus the insights derived from it—begin to reflect the LLM's “perspective” more than the human's? The concern is that if the data pool is increasingly populated with LLM-generated content, a trend we can already see in survey research [5], the “human” aspect may become diluted and the AI's output may not depict human experiences. This fundamental issue to language models, however, will not be solved by HCD and HCI designers but rather lies in the hands of AI developers. One option could be that there are future models based on only

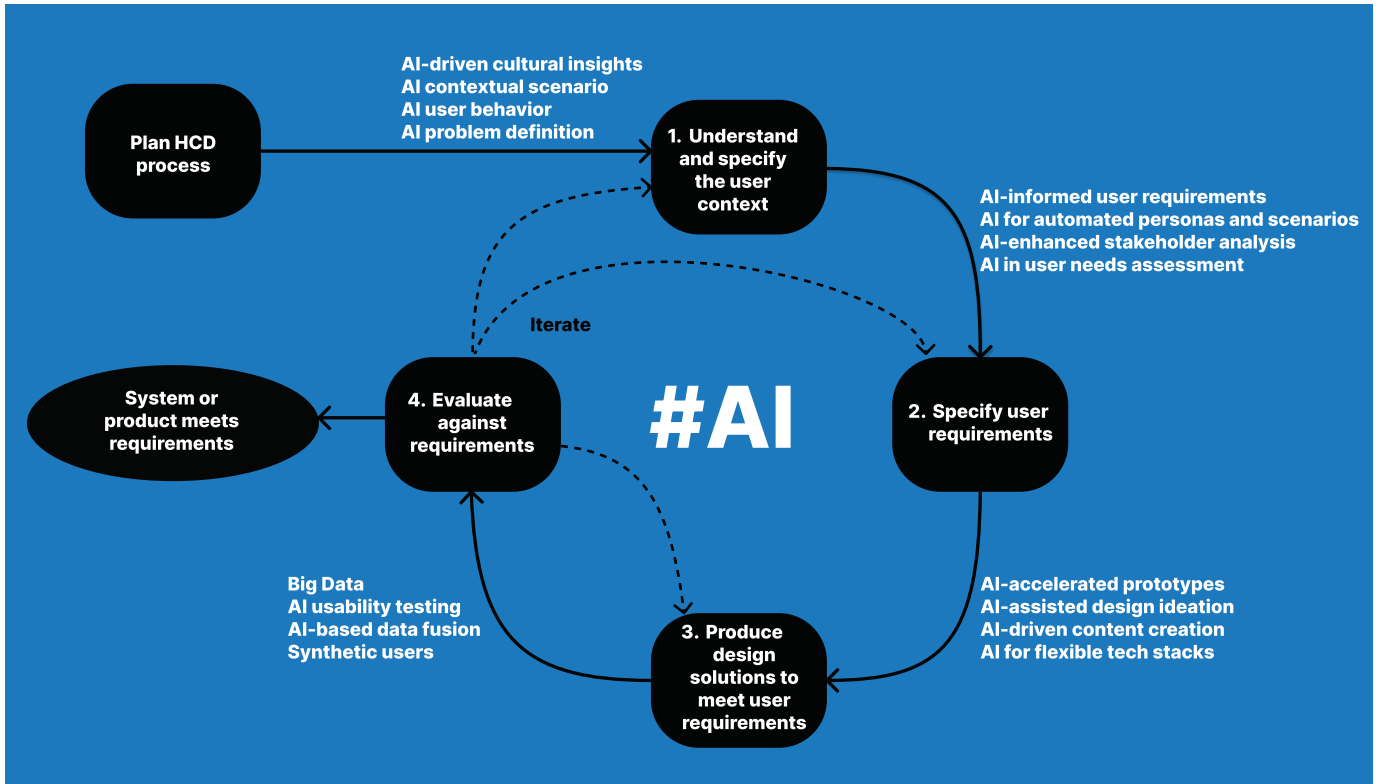


Figure 2. The human-centered design cycle according to ISO 9241-210:2019, with suggestions on how AI may be used in different steps.

human-generated sources and others that include synthetic data generated by LLMs. In such a case, the UX researcher could choose which model to use.

Using LLMs, while understanding the limitations, represents a versatile and effective tool for HCD.

AI TO SUPPORT UNDERSTANDING AND SPECIFYING USER NEEDS

In our work, we have been reflecting on how generative AI (particularly LLMs) can leverage and expand the standardized ISO 9241-210:2019 HCD process (Figure 2). Phase 1 (*Understand and specify the user context*) and Phase 2 (*Specify user requirements*) are covered in this section. Phase 3 (*Produce design solutions to meet user requirements*) is covered in the focus groups section and in later sections about development using LLMs. Last but not least, Phase 4

(*Evaluate against requirements*) is covered in the surveys and interviews section and in a dedicated later section.

Stakeholder identification and needs assessment. Generative AI can revolutionize the way we involve stakeholders in HCD because it increases the scalability and representation of diverse target groups. It can save researchers time and effort in creating personas and scenarios. AI can help researchers auto-generate their personas while embedding deeper backgrounds about marginalized groups, such as users with accessibility needs. It could also change the way we think about personas from a static construct to a live character by allowing developers and product owners to “actively interrogate” a persona during the design process, as the system is now automatically able to understand the character’s history and simulate

reasoning about its motivations, capabilities, likes, and dislikes from the pool of human experiences available. An example of this is asking a persona with accessibility needs whether a proposed design modality is possible from a physical and psychological perspective. What’s even more interesting is that in the future we will need to start designing for AI as a stakeholder and a user in our intelligent systems. Future systems are about human-AI collaboration and cocreation, where the AI is not a basic obedient tool for the user but rather a companion and an assistant that can steer the discussion and the outputs of tasks. The current system designs rarely account for that complex relationship or know how to support it.

Prompt examples:

- Who are the stakeholders when we develop a smart toothbrush for children?
- Create a persona of a mother of two with a university degree who lives in a rural area and works in engineering.
- What would be the main criteria for the created persona when selecting a smart toothbrush for her 9-year-old?

Generative AI can revolutionize the way we involve stakeholders in HCD because it increases the scalability and representation of diverse target groups.

Focus groups and expert interviews. The power of LLMs is

evident in semi-repetitive tasks. Automatically and quickly generating questions for focus groups is at the core of its capabilities. An interesting advancement to explore is to ask LLMs to answer those questions and request automated answers simulating the identity of the generated personas. Alternatively, include one of the personas as an interactive virtual participant in the focus group to ensure the group is kept focused on relevant aspects. Using ideation methods such as storyboards and adversary methods could now be faster. For example, we can generate automated output based on the personas we have. We can also allow participants to cocreate content with AI without prior knowledge of design by describing what they want and using generative AI tools for photos, videos, and sensory experiences to visualize their descriptions. This will also empower researchers to further experiment on the fly with a larger set of higher-fidelity prototypes for participants to choose from during focus groups. Combining advancements in voice recognition, content analysis, and LLMs will also enable researchers to generate new ideas and prompts on demand to steer focus groups with real users. For expert interviews, we can follow a similar process. Here, the level of expertise may vary widely, depending on the domain and topic.

Also, in organizing traditional interviews and focus groups, very mundane tasks can easily be supported, ranging from drafting invitation letters for participants to translation of questions to different languages or educational languages (e.g., to simple English).

Prompt examples:

- *We are running a focus group on casual gaming on smartwatches. What participants should we consider inviting?*
- *Can you create personas for potential participants in the focus group?*
- *What would be good questions to pose and discuss?*
- *Assuming we have the people suggested in the focus group, what would they state as their main concerns?*

Surveys. Similarly, LLMs can help

researchers and practitioners of HCD in generating questions for surveys and interviews. Here, LLMs can be used to 1) speed up the process in supporting different tasks, and 2) explore a broader scope than realistically possible with traditional methods (e.g., produce an infinite number of possible survey items and publish them to AI personas).

It is easy to create a larger number of questions for surveys and interviews. Many of these questions will not be perfect, but iteratively refining them and then making a selection can be helpful to increase the scope and speed. For example, “reminding” the LLM to take a specific method or a school of thought into account can strongly improve the results.

Prompt example [6]:

- *How would we design the survey based on the methodology suggested by Schuman and Presser?*

LLMs can be used as a simulation of the participants of a survey. This is the most controversial usage in our view. For this, we ask the LLM to create the potential answers to the survey questions. These answers are obviously not from humans. They are answers “invented” by the LLM and hence are not the result of a classical survey. Nevertheless, we find the results provide interesting outcomes, as they can hint at potential pitfalls and unanticipated responses. Still, the validity of the generated answers is not clear. That said, research suggests that many surveys using tools and platforms like Prolific (<https://www.prolific.com/>) and Mechanical Turk (<https://www.mturk.com>) also

generate results where validity is hard to prove and often questioned [7]. In Prolific, we can specify the target group (e.g., age, profession, language skills, location where people live, etc.). With ChatGPT, we can do something similar by combining this information in the prompt with the question. It is clearly unethical and fraudulent to “sell” simulated surveys as real user surveys, as there are first reports on single LLM-based bots emulating participants in online surveys [5], but we expect “simulated surveys” will become an additional method in UCD.

RULE 3: *Involving the user is not an end in itself. Hence, if we can get information from LLMs of the same quality, we should not waste human resources.*

Prompt examples:

- *We want to conduct an online survey asking users about their experience with public health websites. Create questions.*
- *We have the following survey question: “You want to get a loan for a car. Whom would you trust to get good advice?” Can you create five answers we could expect when asking a person in Cairo this question?*
 - *[Iterate on 2]: When asking a person in Manchester, U.K., this question?*
 - *[iterate 2 and 3] What would be the most common answer for Cairo and for Manchester? Why?*

AI FOR PROTOTYPING AND IMPLEMENTING INTERFACES AND SYSTEMS

In the stages of prototyping and system implementation, AI serves as a valuable collaborator. Imagine a

EMERGING PRACTICES

Model-based **stakeholder** identification: Using an LLM to identify stakeholders and to describe the concerns and motivations.

Automated creation of **personas and scenarios** (iterative): Using prompts that specify the main aspects and then using an LLM to create elaborate personas and scenarios.

ChatGPT for **ideation**: In different stages, where ideas need to be generated, LLMs can provide a “partner” to get ideas, ranging from asking simple questions about ideas for a certain challenge to asking the LLM to comment on an idea.

Iterative **question design** (interactive, iterative): Interacting with users is at the core of HCD. Here we often ask questions and start discussions. For creating these questions or prompting discussion, LLMs are helpful and can provide valuable input.

Simulated user responses: Instead of users in a interview, focus group, or surveys, we can ask an LLM to simulate the user responses. This is controversial and must be made transparent.

designer leveraging a generative AI to realize a low-fidelity prototype, for instance, using a diffusion model to quickly generate a mock-up of a smartphone app. Within minutes, they have a visually compelling draft to discuss and assess. Subsequently, high-fidelity prototypes can be produced; a prime example would be using an LLM to swiftly construct code for the app. The shift between high- and low-fidelity prototypes thus becomes less about labor-intensive reworking and more about leveraging AI's speed and efficiency, giving new fuel to quick ideation and creativity.

Creating prototypes with some functionality, such as a website or an app, becomes, in many cases, easier and faster with LLMs. Creating HTML pages, CSS design, or JavaScript code for simple applications works very well with ChatGPT or Copilot. Rewriting webpages based on the suggestions of a participant in a design session can often be insidious.

Prompt examples:

- *Create HTML code that shows a timetable with events. Users should be able to select events.*
- *[iterate on 1] Increase the contrast and the font size.*

Looking to the future, multimodal generative models are set to improve our ability to enrich prototypes with varied types of content. This capability not only simplifies the building process but also fuels the creation of more-diverse prototypes.

RULE 4: *Utilize LLMs to create functional prototypes to improve what users experience.*

AI FOR EVALUATION

AI can enhance evaluation processes and methods in HCD. With its ability to process, combine, and generate large

amounts of data, AI can automate and speed up the evaluation process.

By leveraging vast amounts of data, AI can provide recommendations on which system to implement and test. It can augment and automate decision making in implementation processes by weighing a multitude of factors, such as costs, operational efficiency, scalability, and fairness, in real time. AI can thus support and automate decision-making processes where multiple variables of a to-be-implemented design have to be balanced in real time.

Furthermore, AI can play a new significant role in identifying system shortcomings by building user models. It can simulate a wide range of personas, usage scenarios, and interaction with a given system. LLM user models make potential issues and areas of improvement immediately evident and create a large and diverse group of emulated users and experts that are typically not sourced in HCD.

Notably, AI can also facilitate heuristic system evaluation directly, where general rules or principles are used to assess usability. Leveraging AI agents in this context allows for a comprehensive, broad review of the design, identifying strengths, weaknesses, and opportunities for improvement that might be overlooked in a traditional evaluation process. These agents can interact with the system, providing feedback on usability, effectiveness, and overall user experience. This gives designers a unique inside look at how their designs function from a user perspective, thus making the subsequent iteration more attuned to user needs and giving designers a closer look at users' mental models.

DISCUSSION AND ETHICAL IMPLICATIONS

The numerous benefits of using AI in

HCD are evident. AI has the potential to make the HCD process more efficient, inclusive, comprehensive, and adaptable. As these exciting new opportunities emerge, it is important to consider the theoretical, practical, and ethical consequences that come with rethinking HCD with AI.

George Box's aphorism "All models are wrong, but some are useful" captures the state of AI integration in HCD. No model is infallible, but certain AI models, applied with careful consideration of their biases and other shortcomings, could enhance the design process tremendously. It is important, however, to critically evaluate the quality of current AI systems and their applicability to HCD. AI models excel in areas requiring large-scale data analysis, pattern recognition, and scenario simulation [8]. In contrast, core areas of human experience like intuition, empathy, and understanding may not be part of a modern AI training set or could be too context-specific to be emulated by AI. In some fields of HCI like user interface design, empathetic computing, or social VR, we create experiences that support intuitive use, enhance our emotional palette, or create bonds between people. If current AI is devoid of emulating these aspects of human experience, then we might consider refraining from using AI in HCD altogether.

On a larger scale, however, AI automation could make nuances visible in the absence of data resources by emulating user groups that would not be considered otherwise, therefore avoiding bias, for example, by simulating users with red-green color vision deficiency or even the rare yellow-blue color vision deficiency. Thus, as with any tool in HCD, we should apply AI in HCD with reasonable skepticism, that is, by using multimethod approaches combining small-scale user studies with large-scale user simulation, which may help in uncovering and leveling biases introduced by new tools such as AI in HCD.

Finally, while AI can enhance HCD, it is important to not limit innovations and imagination to these models alone. AI-augmented HCD will set a new baseline for the design

The power of LLMs is evident in semi-repetitive tasks. Automatically and quickly generating questions for focus groups is at the core of its capabilities.

process by facilitating the understanding of users, accelerating the design of prototypes, and scaling evaluation. Ultimately, however, AI will be a new tool to adopt; increasing the quality of our work, interaction designers should strive to innovate beyond AI-augmented HCD to continue broadening the boundaries of what is possible for creating, refining, and evaluating human experiences.

Once people are using AI to augment, enhance, and speed up their HCD, it will be hard to opt out and do it the “traditional way,” relying solely on human participants and their input. By using LLMs, we might make UCD cheaper and hence more widely applicable; at the same time, though, we put pressure on the field to move this way in order to stay competitive. Hence, the transparency about how UCD is conducted and to what extent models are used is critical.

Despite the disruptive nature of generative AI use, the shift from traditional, hands-on approaches like focus groups and field studies to AI techniques involving data scraping and linguistic analysis through LLMs will be gradual. It is not just a change in methodology, though; it is embracing synthetic data of the rich information of human experience relevant to technology design, and therefore making HCD more human-centric.

So, when we discuss “human-centric” in the realm of HCD, what do we mean? The human here is far from an archetype. It embodies a vibrant mosaic of all our collective experiences—data that LLMs can approximate. To be human-centric, however, is to hold a commitment to developing technologies that not only fit into human lives but also fundamentally improve them. Thus, in the era of AI-enhanced HCD, AI is not the end-all but a means—an instrument that, while powerful, is employed by designers to augment research.

CONCLUSION AND RECOMMENDATION

The goal of HCD is to create systems that are easy to use, are enjoyable to use, and, most

importantly, enhance people’s lives by helping people meet their needs and achieve their goals, whether they are personal or related to work or to leisure. Involving people—users—in the design and evaluation process has been the key since the beginning of the field.

Involving people/users is not an end in itself, however. It has the purpose of helping us create useful, well-designed interactive systems. We believe that we can transform or at least augment our methods with AI models. If we can create a system with the same or a higher bar of quality using AI models, we need not involve people; we should not waste their time. Of course, if we cannot meet or elevate the quality bar using AI model-based methods, we still need to involve humans.

At this point, the LLMs are changing quickly. Many of our explorations and experiments have been carried out with ChatGPT, but there are more AI-based modeling tools and platforms being developed and made available, and these models are improving. Hence, it is important to experiment and learn where LLMs offer shortcuts and where they are not useful.

Most important in our view is that we keep an open mind, that we understand there are areas where we can benefit, and also tasks for which LLMs are not helpful. The key is to be transparent about how goals were reached, to clearly describe where LLMs were used and what the limitations are.

ENDNOTES

1. Brown, T. et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems* (2020), 1877–1901.
2. Bommasani, R. et al. On the opportunities and risks of foundation models. arXiv:2108.07258, 2021; <https://doi.org/10.48550/ARXIV.2108.07258>
3. Paul, J., Ueno, A., and Dennis, C. ChatGPT and consumers: Benefits, pitfalls and future research agenda. *Int. J. Consum. Stud.* (Mar. 2023). ijcs.12928; <https://doi.org/10.1111/ijcs.12928>
4. Alkaiissi, H. and McFarlane, S.I. Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus* (Feb. 2023); <https://doi.org/10.7759/cureus.35179>
5. Draxler, F. et al. Gender, age, and

technology education influence the adoption and appropriation of LLMs. arXiv:2310.06556, 2023; <https://doi.org/10.48550/ARXIV.2310.06556>

6. This prompt works sufficiently well, even if you know very little about the method. The LLM will know that this prompt refers to Howard Schuman and Stanley Presser. However, we should be aware that the LLM may suggest a very general interpretation of the method.
7. Tang, J., Birrell, E., and Lerner, A. Replication: How well do my results generalize now? The external validity of online privacy and security surveys. *Proc. of the 18th Symposium on Usable Privacy and Security*. ACM, New York, 367–385
8. Jarrahi, M.H. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons* 61, 4 (Jul. 2018), 577–586; <https://doi.org/10.1016/j.bushor.2018.03.007>

Albrecht Schmidt is a professor of computer science at Ludwig Maximilian University Munich (LMU). His research interests are in intelligent interactive systems, interactive AI, ubiquitous computing, digital media technologies, and media informatics. He studied computer science at the University of Ulm and at Manchester Metropolitan University and has a Ph.D. from Lancaster University.

→ albrecht.schmidt@acm.org

Passant Elagroudy is a postdoctoral researcher working at the intersection of HCI and AI with a Ph.D. in computer science from the University of Stuttgart and a robust foundation in media engineering and technology. She focuses on creating ubiquitous and virtual technologies that enhance human cognition and reshape human memories.

→ passant.elagroudy@gmail.com

Fiona Draxler is a recent Ph.D. graduate from LMU Munich. Her research interests include human-AI interaction in application domains such as ubiquitous learning and cowriting with large language models.

→ fiona.draxler@ifi.lmu.de

Frauke Kreuter is the chair of Statistics and Data Science at LMU Munich and is codirector of the Social Data Science Center at the University of Maryland, where she is also a faculty member in the Joint Program in Survey Methodology.

→ frauke.kreuter@lmu.de

Robin Welsch is a professor of engineering psychology at Aalto University. His research focuses on HCI to improve theories and methods in engineering psychology. His current research interests include AI, extended reality, and human augmentation.

→ robin.welsch@aalto.fi

