Transforming HCI Research Cycles using Generative AI and "Large Whatever Models" (LWMs)

Passant Elagroudy
passant.elagroudy@gmail.com
German Research Center for Artificial Intelligence (DFKI),RPTU Kaiserslautern-Landau
Kaiserslautern, Germany

Jie Li
jie_li@epam.com
EPAM
Hoofddorp, Netherlands

Kaisa Väänänen
kaisa.vaananen@tuni.fi
Tampere University
Tampere, Finland

Paul Lukowicz
Paul.Lukowicz@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany

Hiroshi Ishii
Ishii@media.mit.edu
MIT Media Lab
Cambridge, USA

Wendy E. Mackay
mackay@lri.fr
ExSitu, Inria,LISN, Université Paris-Saclay
Paris, France

Elizabeth F Churchill
churchill@acm.org
Google LLC
Mountain View, USA

Anicia Peters
apeters@nust.na
National Commission for Research, Science and Technology
Windhoek, Namibia

Antti Oulasvirta
antti.oulasvirta@aalto.fi
Aalto University
Otaniemi, Finland

Rui Prada
rui.prada@tecnico.ulisboa.pt
INESC-ID,Instituto Superior Técnico, Universidade de Lisboa
Lisbon, Portugal

Alexandra Diening
alexandra_diening@epam.com
EPAM
Zurich, Switzerland

Giulia Barbareschi
barbareschi@kmd.keio.ac.jp
Keio Graduate School of Media Design, Keio University
Greater Tokyo, Japan

Agnes Gruenerbl
agnes.gruenerbl@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany

Midori Kawaguchi
midori@kmd.keio.ac.jp
Keio University Graduate School of Media Design
Greater Tokyo, Japan

Abdallah El Ali
Abdallah.El.Ali@cwi.nl
Centrum Wiskunde & Informatica (CWI)
Amesterdam, Netherlands

Fiona Draxler
fiona.draxler@uni-mannheim.de
University of Mannheim
Mannheim, Germany

Robin Welsch
robin.welsch@aalto.fi
Aalto University
Otaniemi, Finland

Albrecht Schmidt
albrecht.schmidt@ifi.lmu.de
LMU Munich
Munich, Germany

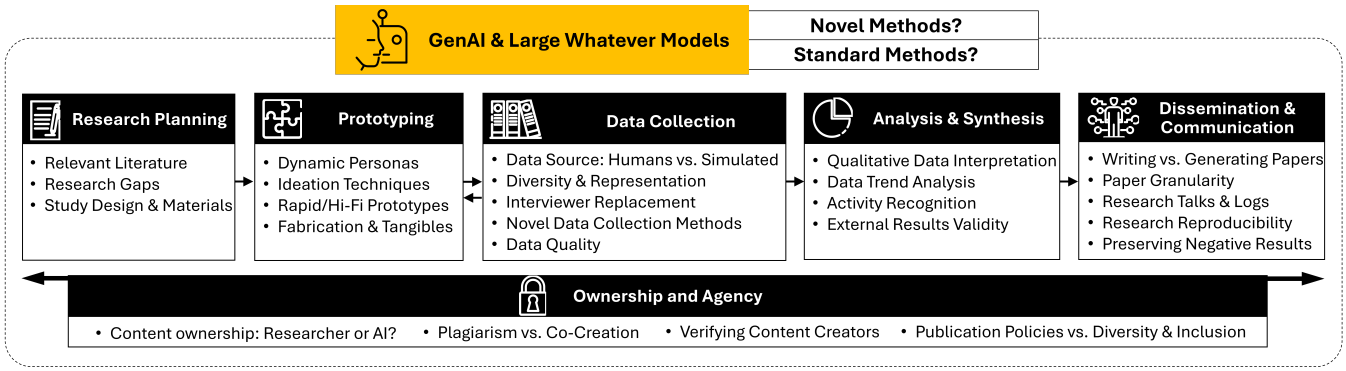

Figure 1: Sample of the topics that interest the SIG showing the five-staged research cycle for HCI research projects.

ABSTRACT

This Special Interest Group (SIG) explores the transformative impact of Generative Artificial Intelligence (GenAI) on Human-Computer Interaction (HCI) research processes. The theme here is to answer "question zero": when to use and when to refrain from using AI tools during the research cycle? The discussion is guided by five research phases commonly used in HCI: research planning, prototyping, data collection, analysis and synthesis, and dissemination and communication. We investigate how GenAI accelerates project cycles, enhances reproducibility, and influences inclusivity in research. We also address the challenging ethical considerations about the ownership of generated content. Our goal is to build a community of HCI enthusiasts to harness the early advantages of the recent groundbreaking technology and foresee challenges arising from its prevalence in the scientific community.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; *Interaction paradigms*; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

Generative AI, Large Language Models, Large Multimodal Models, ChatGPT, research processes, science, HCI research

ACM Reference Format:

Passant Elagroudy, Jie Li, Kaisa Väänänen, Paul Lukowicz, Hiroshi Ishii, Wendy E. Mackay, Elizabeth F Churchill, Anicia Peters, Antti Oulasvirta, Rui Prada, Alexandra Diening, Giulia Barbareschi, Agnes Gruenerbl, Midori Kawaguchi, Abdallah El Ali, Fiona Draxler, Robin Welsch, and Albrecht Schmidt. 2024. Transforming HCI Research Cycles using Generative AI and "Large Whatever Models" (LWMs). In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3613905.3643977

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3643977

1 INTRODUCTION AND MOTIVATION

Generative artificial intelligence (GenAI) provides novel ways for users to create text, images, and various multimedia. There is great excitement about the potential of this technology to support a wide range of creative and analytical tasks. For researchers, especially in human-computer interaction, intriguing questions arise: How is AI changing the overall human-centred design process [14] and the way we do research in general? What if we use large language models (LLMs), such as ChatGPT, to simulate users' responses to interview questions [4]? What if we employ synthetically generated data to train algorithms, addressing issues of data scarcity and bias in user data [9]? What if LLMs support manuscript writing and conducting reviews of research publications, potentially alleviating the heavy workload of academic researchers [8]? It is foreseeable that those models will allow different approaches to conduct research broadly defined, including social sciences and human-computer interaction.

HCI research contributions [1, 17] independent of the domain usually involve some or all of these phases (see Figure 1 and 2): 1) research planning, where a topic is selected and a research method is chosen, 2) prototyping, where intervention or an apparatus is built, 3) data collection, where the researchers gather insights, 4) analysis and synthesis, where researchers use quantitative, qualitative, or mixed methods approaches to make sense of the data and expand the scientific knowledge, and finally 5) dissemination and communication, where the results are communicated internally within research teams and externally to the scientific community for archival and discussion [7, 10].

In principle, GenAI can speed up project cycles by creating and reusing content at different phases using iterative constraints rather than step-by-step creation. For example, Park et al. [12] leveraged GPT-3 to simulate synthetic users and conversations over social computing platform prototypes, where they found that generated responses are hard to differentiate from actual community behaviour. GenAI interactions can also inspire researchers about research gaps through iteratively summarizing and analyzing larger bodies of knowledge. For example, Oppenlaender and Hämäläinen [11] evaluated the combination of ChatGPT and GPT-4 for mining insights

from a text corpus of the CHI 2023 conference proceedings to identify research challenges in HCI, and conclude that such an approach enables a cost-efficient means for mining insights in academia and practice at scale. They can also support researchers in including hard-to-reach communities in research projects. Moreover, they can impact research reproducibility by offering simulated users' data and fundamentally changing the way we need to communicate and preserve our scientific results. An example is Kim and Lee's work [6] demonstrating a methodological framework that integrates LLMs and social surveys to accurately predict individual responses to survey questions that were partially missing. While these advantages are notable, they come at a cost. AI systems are not infallible; they can produce hallucinations, such as content that diverges from the user input, contradicts previously generated context, or misaligns with established world knowledge (as observed in LLMs [19]), and factual errors, often reflecting biases present in their training data [4]. The tension between genuine co-creation and plagiarism is also an ongoing discussion within the HCI community. Many researchers struggle to determine ethically under what conditions must they disclose their use of AI tools in their research work, particularly in light of ACM policies and ethical guidelines (e.g., ACM Policies on Authorship[1]). Feeding research data to GenAI applications to aid researchers also poses serious concerns about data security and privacy.

In this SIG, we want to bring researchers together to envision **how will GenAI support our research processes and identify the limitations of co-working with such systems?** While previous SIGs (e.g. [5, 15]) discussed the use of AI as a tool for building innovative HCI systems, our goal here is to complement this work and focus on the fundamental upcoming changes to the scientific processes in the HCI community with the prevalence of GenAI.

We specifically focus on *two* questions: 1) How will GenAI change the way we perform standard HCI methods? and 2) Is there space for new approaches to do HCI research that were not possible before GenAI? We want to reflect on both questions through the research cycle phases (see Figure 1) focusing on five perspectives:

(1) The power of GenAI to create, simulate and reuse existing content.
(2) The challenges and limitations of relying on the created content.
(3) The impact of using such models on inclusion and diversity of participants and researchers.
(4) Governance models and ethics of the Human-AI interaction in the context of research processes.
(5) The short and long-term impact of different styles of interaction with GenAI on researchers.

2 CONTEXTUAL OVERVIEW OF GENERATIVE AI

AI research has already spanned about six decades [3]. It's only been recently that the launch of commercial GenAI applications such as ChatGPT [2] has transformed the landscape of AI. Versatile foundation models, such as large language models (LLMs, e.g., GPT-3, BERT), visual models (e.g., DALL-E, Florence), or large multimodal models (LMMs, e.g., UniLM, Gemini), built on deep neural networks and self-supervised learning, have gained widespread adoption through several applications [2], including ChatGPT [3], DALL-E 3 [4], Stable Diffusion [5], and Midjourney [6]. We coin the term *Large Whatever Models (LWM)* as an umbrella to refer to all those models employing any type of big data for predictions. We use "GenAI" for the remainder of the proposal to reference the output of LWMs because it is more widely understood. Metaphorically, LLMs are essentially sophisticated autocomplete tools [16] that generate creative texts such as poems, code, or answer questions about a wide range of topics. LWMs are an extension of LLMs that are designed to process and generate content in multiple modalities, including text, images, and audio [18]. For example, the recently introduced LMM *Google Gemini* [7] is described as "a significant leap forward" in AI. It seems that Gemini not only can explain reasoning in math and physics and create games on the go, but also "understands" user intent to generate tailored experiences. Such technologies will change the way we conduct research, and we aim that this SIG puts our community at the forefront of that change.

3 SIG TOPICS

There is the claim that GenAI can speed up the process of creating interactive systems [13]. In the SIG, we use the research cycle phases to guide the discussion (see Figure 1). We specifically want to look into this in detail and discuss the opportunities and limitations for the following tasks:

Topic 1: Research planning Possible discussion questions include: can GenAI help researchers find better research gaps? How to design studies and experiments using GenAI? Which study materials can be generated by GenAI? What is the role of GenAI in uncovering related work?

Topic 2: Prototyping Possible discussion questions include: Can we better understand our stakeholders by developing dynamic personas that designers can converse with? Can we change the way we use ideation methods like storyboards by automatically generating them based on persona descriptions? How will GenAI change the speed of experiencing high-fidelity prototypes and democratize their production? Will GenAI help non-expert researchers use fabrication and tangible prototyping as common research tools? How can we effectively incorporate participation by both human users and intelligent systems into a participatory design process?

Topic 3: Data collection Possible discussion questions include: when to collect data from participants and generate simulation data from GenAI? Can GenAI help in replacing user evaluation for developed prototypes? What is the impact of simulated data on diversity and inclusion of participants' attitudes? Can we better represent communities that are traditionally marginalized through simulated data in contexts such as focus groups? What are the implications of replacing

[1] ACM Policies on Authorship and FAQs: https://www.acm.org/publications/policies/frequently-asked-questions
[2] Launched in November 2022
[3] ChatGPT: https://chat.openai.com
[4] DALL-E 3: https://openai.com/dall-e-3
[5] Stable Diffusion: https://stablediffusionweb.com
[6] Midjourney: https://www.midjourney.com
[7] Google Gemini is a multimodal language model launched on December 6, 2023: https://deepmind.google/technologies/gemini/#introduction

HCI Contribution	Research Planning	Prototyping	Data Collection	Analysis & Synthesis	Dissemination / Communication
Empirical	X	optional	X	X	X
Artifact	X	X	optional	optional	X
Methodological	X	optional	X	X	X
Theoretical	X	optional	optional	X	X
Datasets	X	optional	X	optional	X
Survey	X		optional	X	X
Opinion	X				X

Figure 2: Summary of the research phases in HCI contributions. (X) denotes a used phase from the project's research cycle.

interviewing researchers with conversational agents? Does GenAI enable participants to better express themselves in ways that were not accessible before such as drawing concepts in real-time by describing them? How to ensure data quality and counteract malicious usage such as mechanical turk farms of imaginary users using GenAI?

Topic 4: Analysis and synthesis Possible discussion questions include: how can we utilize GenAI to code and understand qualitative data? How can GenAI help researchers in distilling data trends without biasing their interpretation of the data? How will GenAI impact developments in activity recognition and understanding of human behaviour? How do AI hallucinations impact the analysis validity and how to deal with them? How can GenAI support the external validity of research results through simulated data?

Topic 5: Dissemination and Communication Possible discussion questions include: when should we write papers from scratch and when can we use GenAI to generate text? Should we still publish papers in the same format or should we use alternative formats such as key bullet points that are expanded based on the reader's knowledge? How can GenAI help us give better research talks? What should we preserve about our research processes: the papers or more comprehensive logs? How can GenAI support the reproducibility of research? Should we find indexable alternatives for publications that support the preservation of negative research results?

Topic 6: Ownership and Agency Who owns content created by ChatGPT: the researcher who wrote the prompt or the AI? To what extent does the participation of GenAI agents in research projects qualify as plagiarism, and when does it transition into a collaborative process of human-AI co-creation? How to check the content source in a research project, whether it is human-generated or AI-generated? In what ways might current publication policies regarding GenAI usage discriminate against or benefit authors depending on their demographics? How will GenAI benefit non-native speakers writing papers or researchers with accessibility needs?

4 SIG GOAL AND OUTCOME

Our goal here is to start the discussion and build a persistent support community to harness the power of AI in evolving academic research, foresee the challenges and try to circumvent them. While our discussion primarily revolves around the research-creation cycle, it's essential to acknowledge that GenAI also influences other facets of the scientific ecosystem, including review generation and publication production. The SIG is relevant to authors in SIGCHI conferences as it targets the pillars of scientific projects. Examples of relevant communities include but are not limited to system designers, empirical researchers, social scientists, fabrication and tangibles researchers, and theory researchers. The content is also relevant to the academic and industrial audience doing research.

The SIG will be divided into two parts. The first part is a panel-like format where selected experts provide a pitch vision about the potential of GenAI to support each of the research phases. Afterwards, we will moderate group activities with the attendees such as re-imagining best papers methodology using GenAI and collection of AI tools for HCI with pros and cons. We will also have free time to discuss the next steps.

For content preservation beyond the SIG, we will create a website[8] to publicize the SIG content and host all the post-SIG resources. We will also start a chat space like Discord or Slack and publicize it before, during, and after the SIG. Domain-specific channels such as qualitative analysis and quantitative analysis will be created. After the SIG, we want to consolidate the discussion in different formats for consumption. First, we will compile a list of AI tools for research categorized by the five aforementioned phases and publish them as a resource for the community. We may also submit a report about the results of the SIG as a publication. Additionally, we will summarize the output of the SIG in a short video that can be widely disseminated to the larger scientific audience as a resource. We hope to create a starter resources kit that the community can utilize to experiment with GenAI.

ACKNOWLEDGMENTS

We would like to acknowledge that ChatGPT-3.5 has been used to restructure some sentences in this manuscript following current ACM policies. This work is partially supported and funded by the following entities: the HumanE AI Network under the European Union's Horizon 2020 ICT programme (grant agreement no. 952026), national funds through FCT, Fundação para a Ciência e a Tecnologia, under project UIDB/50021/2020 (DOI:10.54499/UIDB/50021/2020),

[8] https://www.humane-ai.eu/event/chi24sig-lwms-hci/

and the ERC European Research Council Advanced grant No 321135 (CREATIV: Creating Co-Adaptive Human-Computer Partnerships)

REFERENCES

[1] 2024. *Contributions to CHI.* https://chi2024.acm.org/submission-guides/contributions-to-chi/ Accessed: December 15, 2023.

[2] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, et al. 2021. On the Opportunities and Risks of Foundation Models. *arXiv preprint arXiv:2108.07258* (2021). https://crfm.stanford.edu/assets/report.pdf

[3] Michael Haenlein and Andreas Kaplan. 2019. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review* 61, 4 (2019), 5–14.

[4] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating Large Language Models in Generating Synthetic HCI Research Data: A Case Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 433, 19 pages. https://doi.org/10.1145/3544548.3580688

[5] Jun Kato, Jonas Frich, Zhicong Lu, Jennifer Jacobs, Kumiyo Nakakoji, and Celine Latulipe. 2023. Special Interest Group on Creativity and Cultures in Computing. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 520, 4 pages. https://doi.org/10.1145/3544549.3583175

[6] Junsol Kim and Byungkyu Lee. 2023. AI-Augmented Surveys: Leveraging Large Language Models and Surveys for Opinion Prediction. arXiv:2305.09620 [cs.CL]

[7] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction.* Morgan Kaufmann.

[8] Weixin Liang, Yuhui Zhang, Hancheng Cao, Binglu Wang, Daisy Ding, Xinyu Yang, Kailas Vodrahalli, Siyu He, Daniel Smith, Yian Yin, et al. 2023. Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *arXiv preprint arXiv:2310.01783* (2023).

[9] Selina Meyer, David Elsweiler, Bernd Ludwig, Marcos Fernandez-Pichel, and David E. Losada. 2022. Do We Still Need Human Assessors? Prompt-Based GPT-3 User Simulation in Conversational AI. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) *(CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 8, 6 pages. https://doi.org/10.1145/3543829.3544529

[10] Donald A Norman. 1984. Stages and levels in human-machine interaction. *International journal of man-machine studies* 21, 4 (1984), 365–375.

[11] Jonas Oppenlaender and Joonas Hämäläinen. 2023. Mapping the Challenges of HCI: An Application and Evaluation of ChatGPT and GPT-4 for Mining Insights at Scale. arXiv:2306.05036 [cs.HC]

[12] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 22 pages. https://doi.org/10.1145/3586183.3606763

[13] Albrecht Schmidt. 2023. Speeding Up the Engineering of Interactive Systems with Generative AI. In *Companion Proceedings of the 2023 ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (Swansea, United Kingdom) *(EICS '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 7–8. https://doi.org/10.1145/3596454.3597176

[14] Albrecht Schmidt, Passant Elagroudy, Fiona Draxler, Frauke Kreuter, and Robin Welsch. 2024. Simulating the Human in HCD with ChatGPT - Redesigning Interaction Design with AI. *Interactions* (Jan 2024).

[15] Mohammad Tahaei, Marios Constantinides, Daniele Quercia, Sean Kennedy, Michael Muller, Simone Stumpf, Q. Vera Liao, Ricardo Baeza-Yates, Lora Aroyo, Jess Holbrook, Ewa Luger, Michael Madaio, Ilana Golbin Blumenfeld, Maria De-Arteaga, Jessica Vitak, and Alexandra Olteanu. 2023. Human-Centered Responsible Artificial Intelligence: Current & Future Trends. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 515, 4 pages. https://doi.org/10.1145/3544549.3583178

[16] James Vincent. 2023. Introducing the AI Mirror Test, which very smart people keep failing. *The Verge* (2023). https://www.theverge.com/23604075/ai-chatbots-bing-chatgpt-intelligent-sentient-mirror-test

[17] Jacob Wobbrock and Julie Kientz. 2016. Research contribution in human-computer interaction. *interactions* 23 (04 2016), 38–44. https://doi.org/10.1145/2907069

[18] Weihao Yu, Zhengyuan Yang, Linjie Li, Jianfeng Wang, Kevin Lin, Zicheng Liu, Xinchao Wang, and Lijuan Wang. 2023. Mm-vet: Evaluating large multimodal models for integrated capabilities. *arXiv preprint arXiv:2308.02490* (2023).

[19] Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, et al. 2023. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *arXiv preprint arXiv:2309.01219* (2023).