Nonhumaness and Non-materiality of Synthetic Agents

TAKUYA MAEDA, Western University

1 Context

LLM-based systems are increasingly exhibiting human-like personalities and personas. This is by design, as developers attempt to evoke tailored qualities such as "helpfulness" and "harmlessness" through reinforcement learning [3], human or AI feedback to models, prompt engineering [17], and inference tasks to see if given responses share similar trends as human decisions [9, 11]. The goal, they say, is to align model outputs with human values and behaviors.

These efforts have been largely successful, with LLM-based systems resembling humans so convincingly that some researchers claim synthetic data (or data outputted by LLM-based systems) could be a viable option to increase diversity in datasets, potentially replacing human participants in qualitative research. For instance, previous studies argue that synthetic participants could be beneficial for examining impractical, unethical, or unsafe cases or topics [2, 5].

However, such use cases beg important questions about the legitimacy and authenticity of generated profiles. What, or who, do they actually represent?

2 Considerations

These potential use cases of LLM-based systems and outputs brush up against long-standing debates about the relationship between function vs. form and substance vs. representation. For instance, we know that social and cultural norms can determine how experiences are expressed in language and actions, and how people express their thoughts and beliefs [4, 6]. Moreover, the way that language and actions are subsequently interpreted can depend on how these contents are framed and communicated, potentially diverging from original intentions. Thus, even the practice of "cleaning" *original* user-generated data can distort meaning by shaping content to a particular end.

The proposed use of LLM-based outputs in research, in particular, echoes practices from the early, deeply colonial days of anthropology, where "armchair anthropologists" drew conclusions about peoples from secondary/mediated accounts (generally from colonial officers whose records were intended to support the subjugation of these peoples). Contemporary curricula in material culture anthropology include whole units devoted to the "politics of representation," emphasizing how words, photographs, and artifacts—even those generated by the populations being represented—can take on different meanings when stripped of their original context [7, 10].

Author's Contact Information: Takuya Maeda, tmaeda@uwo.ca, Western University.

These risks are even more pressing when represented peoples are excluded even from the generation of artifacts (data). The outputs generated by LLM-based systems, while seemingly "realistic" or even plausible, are not yielded from the cultural, historical, or socio-economic context of the people in question, but from the foundational dataset from which the LLM calculates patterns to predict likely letter and word sequences. Therefore, at most, synthetic agents are proxies or simulations constrained by existing knowledge that describe someone's experience without context or mimic someone's emotion, thought, or speech without materiality.

Why, then, should we assume that such different processes will yield comparable results[1]—comparable enough that LLM-based outputs could replace individual experience? And what are the potential harms that might arise when the simulation of contexts, experiences, and values misses its mark? If research, at its root, is intended to shed light on the unknown (that which exists beyond existing representations and understandings), then we must carefully interrogate whether and how synthetic agents and synthetic data may be useful in expediting/enabling research, and weigh that against the risk for a kind of intellectual incest—the harvesting of "insights" from data riddled with the researchers' (or third party's) own assumptions about the target population.[2]

3 Proposal

I'm interested in exploring these questions of LLM-generated personas' non-materiality and non-humanness through the lens of Helen Nissenbaum's notion of "contextual integrity" [16], as well as from the lens of existing discourse in critical HCI [8, 19] and material culture anthropology. These fields provide examples for how to discuss/contemplate synthetic agents in ways that situate research methods, goals, and outcomes within human experience and interactions, rather than abstract formulations of human "values."[3] Other relevant discursive lenses include theories of affordance [15] and media equation [14], which distinguish between what systems can generate and what they actually represent, arguing that the ability to produce human-like outputs does not constitute genuine representation of values and experiences.

I believe drawing insights from social science is important to define what it means to participate or be meaningfully represented in the public sphere. Similarly, foregrounding the non-materiality and non-humanness of LLM-based systems is important to understand what types of representations synthetic agents afford. Revisiting the basic ideas of human interaction could reveal why and how synthetic agents fall short of representing human experience or "authentic" human qualities. Altogether, these inquiries could shed light on how, when, and why LLM-based systems and outputs can (or cannot) be meaningfully used in research.

3.1 My Background

I'm a 2nd year (soon-to-be 3rd year) PhD Student in the Faculty of Information and Media Studies at Western University. My research is on the social and ethical impact of anthropomorphic design in AI systems. My work is jointly supervised by Prof. Luke Stark and Prof. Anabel Quan-Haase, incorporating work from HCI, STS, and communication to tackle this topic. Through this workshop, I hope to interact with other researchers to enrich my understanding of synthetic data and to critically examine the purpose and effects of such generated data. I'm currently working on a paper in this topic, too.

---

[1]Indeed, Agnew et al. [1] suggest that contextual representation doesn't seem to create inclusion.
[2]This resembles the kind of abstraction trap described in Selbst et al. [18], where human experience is turned into something processible for computing systems.
[3]Value alignment doesn't provide reciprocity or consideration for diverse social circumstances. This lack of reciprocity and social consideration creates an illusion of care that may encourage users to anthropomorphize and project humanness onto such systems [12, 13].

References

[1] William Agnew, A. Stevie Bergman, Jennifer Chien, Mark Díaz, Seliem El-Sayed, Jaylen Pittman, Shakir Mohamed, and Kevin R. McKee. 2024. The Illusion of Artificial Inclusion. 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 286, 12 pages. doi:10.1145/3613904.3642703

[2] Gati V Aher, Rosa I. Arriaga, and Adam Tauman Kalai. 2023. Using Large Language Models to Simulate Multiple Humans and Replicate Human Subject Studies. In *Proceedings of the 40th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 337–371. https://proceedings.mlr.press/v202/aher23a.html

[3] Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. 2022. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073* (2022).

[4] Emily M. Bender and Alexander Koller. 2020. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 5185–5198. doi:10.18653/v1/2020.acl-main.463

[5] Courtni Byun, Piper Vasicek, and Kevin Seppi. 2023. Dispensing with Humans in Human-Computer Interaction Research. 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 413, 26 pages. doi:10.1145/3544549.3582749

[6] Catherine R Cooper and Jill Denner. 1998. Theories linking culture and psychology: Universal and community-specific processes. *Annual review of psychology* 49, 1 (1998), 559–584.

[7] Helen Davis. 2003. Understanding Stuart Hall. (2003).

[8] Paul Dourish. 2006. Implications for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 541–550. doi:10.1145/1124772.1124855

[9] Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences* 120, 30 (2023), e2305016120.

[10] Stuart Hall. 1989. Cultural identity and cinematic representation. *Framework: The Journal of Cinema and Media* 36 (1989), 68–81.

[11] 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. doi:10.1145/3544548.3580688

[12] Mirabelle Jones, Nastasia Griffioen, Christina Neumayer, and Irina Shklovski. 2025. Artificial Intimacy: Exploring Normativity and Personalization Through Fine-tuning LLM Chatbots. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 793, 16 pages. doi:10.1145/3706598.3713728

[13] Takuya Maeda and Anabel Quan-Haase. 2024. When Human-AI Interactions Become Parasocial: Agency and Anthropomorphism in Affective Design. 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, 1068–1077. doi:10.1145/3630106.3658956

[14] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues* 56, 1 (2000), 81–103.

[15] Gina Neff and Peter Nagy. 2016. Automation, Algorithms, and Politics| Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication* 10 (2016), 17.

[16] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[17] Joon Sung Park, Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2022. Social Simulacra: Creating Populated Prototypes for Social Computing Systems. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 74, 18 pages. doi:10.1145/3526113.3545616

[18] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 59–68. doi:10.1145/3287560.3287598

[19] Lucille Alice Suchman. 2007. *Human-machine reconfigurations: Plans and situated actions.* Cambridge university press.