Mensch und Computer 2026 · Workshop · Duisburg · 30.08.2026

The Median Has No Edge Cases

An Evidence Map for Synthetic Users in UX Research

Prof. Dr. Andreas Hinderks · Hochschule Hannover

jan-andreas.hinderks@hs-hannover.de

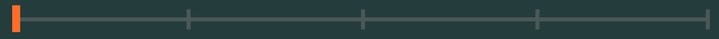
Synthetic users promise real insight without talking to real users.

(Rosala & Moran, 2024)

1 · The question

A question with no single answer

Do synthetic users work?



User research does three different jobs

- **Averages.** A mean score. What most people prefer. Which way an effect goes.
- **Variation.** How groups differ. Which needs pull apart.
- **Lived experience.** What something means to a person, in their world.

2 · The evidence

What comparative studies show

Sorted by job, is the evidence still in conflict?

Averages come out right — under conditions

- **What comes out right.** How groups voted. Classic lab experiments. Which way an effect goes, and roughly how big.
- **Only under three conditions.** The result is a group average. The task is common in the training data. The prompt holds rich data from real people.

(Argyle et al., 2023; Aher et al., 2023; Hewitt et al., 2024)

Hewitt et al.: preprint, not peer-reviewed

Evidence base: 182 studies, reviewed by Kuric et al. (2026). A preprint. The authors work for a UX research vendor.

Same mean

(Bisbee et al., 2024)

different distribution

Bisbee and colleagues repeated the silicon sampling study. The simulated means matched the human means. But the spread was far too small, the links between answers were wrong, and the results changed with the model version. Two people can read the same simulation and both be right. It depends on which number they report.

The questionnaire fails too

- **Gao et al. (2025)** — the model spread its answers almost evenly. Humans spread widely. And the errors change with language, role and safety settings. You cannot predict them.
- **Domínguez-Olmedo et al. (2024)** — the order and the labels of the answer options mattered more than the question itself. A direct warning for anyone who gives a UEQ or a SUS to a model.

Almost none of this work comes from UX research. The mechanism still applies. But nobody has tested the UX tasks themselves.

And the people are not there

- **Misportrayal and flattening** — the model repeats what outsiders say about a group. Inside the group, everyone sounds the same.
- **Culture** — a synthetic sample built for a German election study leaned toward Green and Left. It found the typical voter and missed the rest.
- **Interviews** — nineteen researchers rebuilt their own interview study. The stories sounded real at first. Then: no consent, no person behind them, no context.

(Wang et al., 2025; von der Heyde et al., 2026; Kapania et al., 2025)

3 · Why

Why the split is built in

A young technology, or the design itself?

What the training predicts

- **Step one.** The model learns to predict the next word from a huge amount of text. What it learns is an average. Its most likely answers sit in the middle.
- **Step two.** Training with human feedback makes the range even narrower. The model learns what human raters like.

(Ouyang et al., 2022)

The median has a face: middle-aged, able-bodied, native-born, Caucasian, atheistic, male.

(Tan & Lee, 2025)

Persona prompting moves the middle. It does not bring back the range.

(Cheng et al., 2023; Wang et al., 2025)

4 · The map

An evidence map for synthetic users

Where exactly is the line?

Nine activities, placed by two questions

How much depends on the result →

Kind of knowledge the activity needs →

	Averages aggregate scores, majority directions	**Variation & subgroups** differences, response structure	**Lived experience** meaning, context, representation
Substantive claims about users, products; design decisions	Summative evaluation from synthetic panels (UEQ, SUS); market estimates	Personas and segmentations presented as empirical findings about real users	Insight generation; interviews as data; research on marginalized populations
Directional what to study next; prioritization, sampling	Hypothesis triage; effect-direction screening; proto-personas as hypotheses	Segment prioritization or sampling plans derived from simulated heterogeneity	Scoping research directions for underrepresented groups from simulated voices
Procedural does the study run? instruments, training	Pilot instruments and survey logic; dry-run guides and analysis pipelines	Heterogeneous test inputs for branching and edge-case checks	Moderator training; role-played difficult interviews

Fig. 1: The placement, before the verdict. Hinderks (2026), redrawn.

One cell works. Three work with limits. Five do not.

How much depends on the result →

Kind of knowledge the activity needs →

	Averages aggregate scores, majority directions	Variation & subgroups differences, response structure	Lived experience meaning, context, representation
Substantive claims about users, products; design decisions	**Not defensible** Summative evaluation from synthetic panels (UEQ, SUS); market estimates	**Not defensible** Personas and segmentations presented as empirical findings about real users	**Not defensible** Insight generation; interviews as data; research on marginalized populations
Directional what to study next; prioritization, sampling	**Conditional** Hypothesis triage; effect-direction screening; proto-personas as hypotheses Constraint: human validation required	**Not defensible** Segment prioritization or sampling plans derived from simulated heterogeneity	**Not defensible** Scoping research directions for underrepresented groups from simulated voices
Procedural does the study run? instruments, training	**Defensible** Pilot instruments and survey logic; dry-run guides and analysis pipelines	**Conditional** Heterogeneous test inputs for branching and edge-case checks Constraint: tests mechanics, not real variance	**Conditional** Moderator training; role-played difficult interviews Constraint: builds skill, produces no data

Fig. 1: Verdicts as supported by the evidence. Hinderks (2026), redrawn.

The line, and two rules

- **Rehearsal is fine.** Test your questionnaire. Try out your interview guide. Practice your analysis. You only ask: does the study run?
- **Findings are not.** Nothing that leaves this room as a claim about real users. No scores from synthetic panels. No personas sold as facts.
- **Say where the data came from** — every time you report a result.
- **Write down your setup** — model, version, prompt, settings. A new model version gives you a different answer.

(Kuric et al., 2026; Rosala & Moran, 2024; Batzner et al., 2025)

Synthetic users can rehearse research. They cannot perform it.

The median has no edge cases. And edge cases are where user research really matters.

References

Aher, G., Arriaga, R. I., & Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. In *Proceedings of the 40th International Conference on Machine Learning* (PMLR 202, pp. 337–371).

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis, 31*(3), 337–351. https://doi.org/10.1017/pan.2023.2

Batzner, J., Stocker, V., Tang, B., Natarajan, A., Chen, Q., Schmid, S., & Kasneci, G. (2025). Whose personae? Synthetic persona experiments in LLM research and pathways to transparency. In *Proceedings of the 2025 AAAI/ACM Conference on AI, Ethics, and Society*. https://doi.org/10.1609/aies.v8i1.36553

Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., & Larson, J. M. (2024). Synthetic replacements for human survey data? The perils of large language models. *Political Analysis, 32*(4), 401–416.

Cheng, M., Durmus, E., & Jurafsky, D. (2023). Marked personas: Using natural language prompts to measure stereotypes in language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*.

Domínguez-Olmedo, R., Hardt, M., & Mendler-Dünner, C. (2024). Questioning the survey responses of large language models. In *Advances in Neural Information Processing Systems*.

Gao, Y., Lee, D., Burtch, G., & Fazelpour, S. (2025). Take caution in using LLMs as human surrogates. *Proceedings of the National Academy of Sciences, 122*(24), e2501660122. https://doi.org/10.1073/pnas.2501660122

Hewitt, L., Ashokkumar, A., Ghezae, I., & Willer, R. (2024). *Predicting results of social science experiments using large language models*. Preprint, not peer-reviewed.

References

Kapania, S., Agnew, W., Eslami, M., Heidari, H., & Fox, S. E. (2025). “Simulacrum of stories”: Examining large language models as qualitative research participants. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3706598.3713220

Kuric, E., Demcak, P., & Krajcovic, M. (2026). *Synthetic participants generated by large language models: A systematic literature review*. Research Square. Preprint, not peer-reviewed. https://doi.org/10.21203/rs.3.rs-9057643/v1

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., … Lowe, R. (2022). Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*.

Rosala, M., & Moran, K. (2024). *Synthetic users: If, when, and how to use AI-generated “research”*. Nielsen Norman Group. https://www.nngroup.com/articles/synthetic-users/

Tan, B. C. Z., & Lee, R. K.-W. (2025). Unmasking implicit bias: Evaluating persona-prompted LLM responses in power-disparate social scenarios. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the ACL: Human Language Technologies* (Vol. 1, pp. 1075–1108). ACL.

von der Heyde, L., Haensch, A.-C., & Wenz, A. (2026). Vox populi, vox AI? Using large language models to estimate German vote choice. *Social Science Computer Review, 44*(3). https://doi.org/10.1177/08944393251337014

Wang, A., Morgenstern, J., & Dickerson, J. P. (2025). Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence*. https://doi.org/10.1038/s42256-025-00986-z

Thank you.

Prof. Dr. Andreas Hinderks

Hochschule Hannover, Fakultät IV

jan-andreas.hinderks@hs-hannover.de · hinderks.org