



# (Very) Recent Research on LLM's Synthetic Respondents

## Sawtooth Software Webinar

Dimitri Liakhovitski, PhD  
Global Data Sciences, NIQ

April 2026

**NIQ**

## Upheaval in Market Research since the first ChatGPT went public (late 2022)

Can LLMs' synthetic respondents replace survey respondents?

- Stage 0: Wow, that's an idea!
- Stage 1: First evidence from LLM community – we probably can (soon)!
- Stage 2: Clients: Give us synthetic respondents now, they are cheap!
- Stage 3: Marketing Scientists: Wait, let me check!
- Stage 4: A lot of new evidence coming in since late 2024...

**Bit question: Can we let LLMs answer our surveys?**

- Let's take a look at the evidence that's coming in...

## Synthetic LLM Responses' Symptoms: Overwhelming base model priors

“Symptoms” of synthetic respondents based on a review of recent Synthetic respondent literature

- Base model weights (pretrained “priors”) “trump” additional information provided → limited *individuation* of responses, lack of variability and diversity
    - Low variance across individuals: multiple variables with little or no variance
    - Normative overcorrection (excessive rationality): Failure to reproduce human errors, biases, logical inconsistencies, etc.
- Can we make an LLM “change its basic opinion” and provide human-like variety of responses?

## Synthetic LLM Responses' Symptoms: WEIRD countries' bias

“Symptoms” of synthetic respondents based on a review of recent Synthetic respondent literature

- LLMs show low representativeness and cultural bias: models were mostly trained on WEIRD\* countries' data and on mostly white people's data.
  - Minorities tend to be cast by the LLMs as "cultural ambassadors" and "resilient survivors."
  - LLMs fail to reproduce anti-social or rude behavior

→ Can we expect to get valid data for rare, “niche” samples?

\*Western, Educated, Industrialized, Rich, Democratic

## **Synthetic LLM Responses' Symptoms: Insensitivity to novelty**

“Symptoms” of synthetic respondents based on a review of recent Synthetic respondent literature

- Inability to track real world changes – in the world + in the minds and hearts of real people: each LLM has a cutoff time point

→ How can we expect to get up-to-date data?

## Synthetic LLM Responses' Symptoms: Instability

“Symptoms” of synthetic respondents based on a review of recent Synthetic respondent literature

- Instability: update- and model-dependence:
  - Change the LLM, change model version, change prompt instructions → get different LLM responses
  - Merely naming the persona can change the LLM responses!

→ Can we get reliable data?

## **Synthetic LLM Responses' Symptoms: Contamination by memorization**

“Symptoms” of synthetic respondents based on a review of recent Synthetic respondent literature

- LLMs are trained and have billions of parameters. They have a good memory!
    - The responses provided might be rather “memorization” of training material than “created” answers
- Will this be truly new data? How about intellectual property questions?
- Qualtrics example



## **“Ceiling effect”: It’s hard to predict human behavior. Period.**

“Symptoms” of synthetic respondents based on a review of recent Synthetic respondent literature

- Even many best psychological measurements combined fail to predict human behaviors:
  - The limited ability of person-related information to predict human behavior has been well established in academic psychology several decades ago (see Person-Situation debate in Furr & Funder).
  - In the field of social psychology and personality, correlations between personality measures and behavioral outcomes are typically in 0.1-0.3 range
  - Correlations of 0.4 are rare and considered truly impressive (see three large recent meta-analyses by Akbari, M. et al. (2019), Soutter, Bates & Möttus (2020), and Nauzeer & Jaunky (2021))

→ How can we expect LLMs to perform better than humans? “Ceiling effect” of ANY person-related information

## Our research to be presented in Berlin in May












Using synthetic respondents for data imputation

From a synthetic respondents' review article:

- "The studies that achieved the highest alignment did it by including the expected results in the prompt ... even giving the model access to qualitative data from real individuals didn't lead to psychometric alignment or realistic variability. So the best way to make synthetic participants match human data is to **give them the human data first.**"

That's essentially what we did: tried to use LLM for data imputation → See our findings in Berlin 😊

While imputation quality is generally not stellar, LLM under performs MICE and XGBOOST on most evaluation metrics. XGBOOST beats MICE on a few metrics

Purpose	Evaluation Metric	   			
		MICE	XGBoost xgboost vote7	LLM with Completes	Random
Do the imputed responses align with the original human responses?	Rescaled Accuracy (individual level) 	0.71	0.70	0.65	0.60
	Glass Delta (sample level) 	0.28	0.19	0.47	0.09
	Corr <sub>true.imp</sub> (individual level) 	0.20	0.21	0.12	0.00
Do we maintain the original relationship among the variables we impute?	MAE of correlations 	0.40	0.10	0.19	0.19
	Correlation of correlations 	0.53	0.49	0.43	-0.02
Do we maintain the original distribution of human responses?	Kolmogorov-Smirnov statistic 	0.18	0.10	0.26	0.05
	% of variables with no variance 	0.0%	0.0%	9.6%	0.0%



## Some references

- Atari, M., Xue, M. J., Park, P. S., Blasi, D. E., & Henrich, J. (2023). „Which humans?” [Manuscript]. Department of Human Evolutionary Biology, Harvard University.
- Binz, M., Akata, E., Bethge, M., Brändle, F., Callaway, F., Coda-Forno, J., Dayan, P., Demircan, C., Eckstein, M. K., Éltető, N., Griffiths, T. L., Haridi, S., Jagadish, A. K., Ji-An, L., Kipnis, A., Kumar, S., Ludwig, T., Mathony, M., Mattar, M., ... Schulz, E. (2025). A foundation model to predict and capture human cognition. *Nature*, 644, 1002–1009.
- Brand, J., Israeli, A., & Ngwe, D. (2025, October 6). Using LLMs for market research (Working Paper No. 23-062). Harvard Business School.
- Dash, T., Karri, D., Vurity, A., Datla, G., Ahmad, T., Rafi, S., & Tangudu, R. (2025, December 16). Polypersona: Persona-grounded LLM for synthetic survey responses. *arXiv*. <https://arxiv.org/abs/2512.14562>
- Furr, R. M. & Funder, D. C. (in press). Persons, Situations, and Person-Situation Interactions. *Handbook of personality: Theory and Research* (4<sup>th</sup> edition), New York: Guildford.
- Li, A., Chen, H., Namkoong, H., & Peng, T. (2025). “LLM-generated persona is a promise with a catch”. Columbia University.
- Maier, B. F., Aslak, U., Fiaschi, L., Rismal, N., Fletcher, K., Luhmann, C. C., Dow, R., Pappas, K., & Wiecki, T. V. (2025, October 27). LLMs reproduce human purchase intent via semantic similarity elicitation of Likert ratings. *arXiv*. <https://arxiv.org/abs/2510.08338>
- Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Willer, R., Liang, P., & Bernstein, M. S. (2024). \*Generative agent simulations of 1,000 people\*. Stanford University.
- Peng, T., Gui, G., Brucks, M., Merlau, D. J., Fan, G. J., Ben Sliman, M., Johnson, E. J., Althenayyan, A., Bellezza, S., Donati, D., Fong, H., Friedman, E., Guevara, A., Hussein, M., Jerath, K., Kogut, B., Kumar, A., Lane, K., Li, H., ... Toubia, O. (2026, January 4). Digital twins as funhouse mirrors: Five key distortions. *arXiv*. <https://arxiv.org/abs/2509.19088>
- Zhang, J., Liang, X., Deng, A., Bonge, N., Tan, L., Zhang, L., & Zarrett, N. (2025). Leveraging interview-informed LLMs to model survey responses: Comparative insights from AI-generated and human data. Preprint manuscript. (Project/code: <https://osf.io/afgg3>)
- UX Research (March 28, 2026). The largest review of synthetic participants ever conducted found exactly what you’d expect. Synthetic users don’t work: [The Largest Review of Synthetic Participants Ever Conducted Found Exactly What You'd Expect. Synthetic Users Don't Work.](#)

**Thank you**

**NIQ**