

Methodological Challenges and Ethical Tensions

Session 5, Indira Sen and Kristina Gligoric

Some of the many issues...

- **Issues with simulations**
 - Are the rules of interaction realistic?
 - Issues in modeling
 - How is the final outcome variable measured? [[Lundberg et al., 2021](#)]
 - Issues at different time steps
 - Consistency? [[Windrum et al., 2007](#), [Troitzsch, 2014](#)]
 - Performative Prediction [[Hardt et al., 2025](#)] – when prediction influences the outcome

Some of the many issues...

- **Issues with simulations**

- Are the rules of interaction realistic?
- Issues in modeling
- How is the final outcome variable measured? [[Lundberg et al., 2021](#)]
- Consistency? [[Windrum et al., 2007](#), [Troitzsch, 2014](#)]
- Performative Prediction [[Hardt et al., 2025](#)] – when prediction influences the outcome

- **Issues with LLM Simulations**

- Simulation bias [[Anthis et al. 2025](#), [Larooji and Törnberg 2025](#)]
- Prompt brittleness [[Shu et al., 2024](#)]
- Sycophancy [[Anthis et al. 2025](#)]
- Hyper Accuracy bias [[Aher et al., 2023](#)] / alien intelligence [[Kozłowski and Evans 2025](#)]
- Carbon footprint [[Ren et al., 2024](#)]
- Data Leakage [[Barrie and Törnberg, 2025](#)]
- Temperature fluctuations
- Lack of traceability [[Rothschild et al., 2026](#)]
- Persona construction [[Lutz et al., 2025](#)]
- Answer extraction [[Ahnert et al., 2026](#)]
- ...



Real surveys are also imperfect, but...

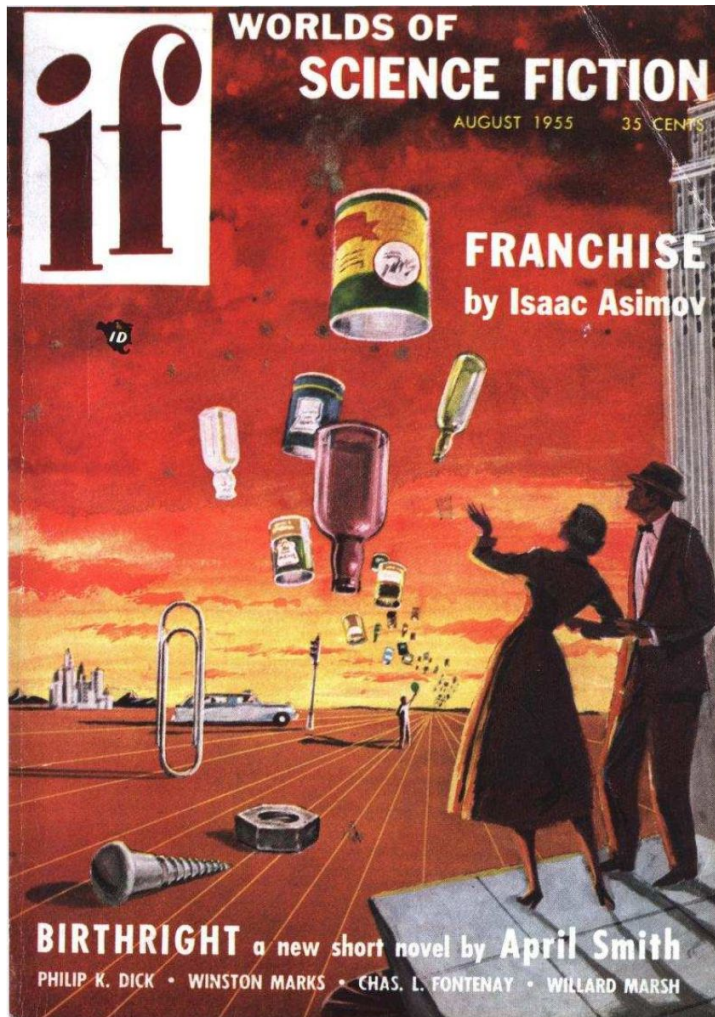
Real vs. Simulated Surveys: Validation

- With real survey data, quality indicators do not entail checking individual responses
- There is also no other ground truth (some exceptions, e.g., elections)
- Validation of survey responses
 - Process validation
 - Transparency
 - Traceability
- Clear polling methodology
- Transparent weights
- LLMs currently fall short of process transparency due to inherent opacity

Real vs. Simulated Surveys: Validation

- With real survey data, quality indicators do not entail checking individual responses
- There is also no other ground truth (some exceptions, e.g., elections)
- Validation of survey responses
 - Process validation
 - Transparency
 - Traceability
- Clear polling methodology
- Transparent weights
- LLMs currently fall short of process transparency due to inherent opacity

Let's say, we manage to fix all of this.



"The first computers were much smaller than Multivac. But the machines grew bigger and they could tell how the election would go from fewer and fewer votes. Then, at last, they built Multivac and it can tell from just one voter."

Linda smiled at having reached a familiar part of the story and said, "That's nice."

Matthew frowned and said, "No, it's not nice. I don't want a machine telling me how I would have voted just because some joker in Milwaukee says he's against higher tariffs. Maybe I want to vote cockeyed just for the pleasure of it. Maybe I don't want to vote. Maybe—"

[https://en.wikipedia.org/wiki/Franchise_\(short_story\)](https://en.wikipedia.org/wiki/Franchise_(short_story))

Ethical Tensions

- Do **respondents** want their survey responses to be simulated?
 - Reduces response burden
 - Response rates have been declining—people might not want to answer surveys
 - Transfers time to answering harder questions more in-depth
 - Risk of ‘AI ventriloquism’
 - Even in hybrid setups, risk of overreliance (‘rubber stamping’)
 - Economic concerns (Prolific and the gig economy)
- Do **users** of the data want to use simulated data?

Andrew Orlowski

Starmer’s ‘synthetic voters’ show Downing Street’s lost the plot

Using AI chatbots to reflect the electorate’s thinking is a fundamentally flawed idea



Andrew Orlowski

Andrew Orlowski is a technology journalist who writes a weekly Telegraph column every Monday. See more ✓

Published 07 July 2025 11:00am BST

🔖 424 📤

📁 Gift this article free

🔍 Add us as preferred source



Sir Keir Starmer’s advisers think AI chatbots could provide immediate feedback on policy ideas Credit: Jordan Pettitt/PA Wire

[The Telegraph](https://www.telegraph.co.uk/technology/2025/07/07/starmer-synthetic-voters-downing-street-lost-the-plot/)

Ethical Tensions

- Collective harms and LLM pollution
 - Misrepresenting a marginalized group in data that feeds policy
- Privacy and consent:
 - Is it okay to fine-tune or steer LLMs, especially commercial ones, on past survey data?
 - Should we simulate individuals?
 - How do we manage informed consent and revocation of consent?

Should We Worry About Problematic Response Behaviour in Social Media Surveys? Understanding the Impact of Social Group Cues in Recruitment

Survey Methods: Insights from the Field, Volume 4, issue 1 (2026)

Zaza Zindel 

How to cite this article : Zindel Z. (2026). Should We Worry About Problematic Response Behaviour in Social Media Surveys? Understanding the Impact of Social Group Cues in Recruitment. Survey Methods: Insights from the Field. Retrieved from <https://surveyinsights.org/?p=21395>.

DOI : 10.13094/SMIF-2026-00015

Copyright : © the authors 2026. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Abstract : Social media advertising has become a common tool for participant recruitment, especially when targeting hard-to-reach or underrepresented populations. A common strategy to boost engagement is the use of social group cues – textual or visual references to religion, gender identity, or ethnicity. While these cues can enhance recruitment efficiency, their impact on response data quality remains poorly understood.

This study investigates how social group cues affect problematic response behaviour in a Facebook-recruited survey on labour market experiences in Germany. Respondents were recruited via four distinct ad conditions: two referencing Muslim identity, and two with neutral framing. Seven indicators were used to assess response quality, capturing satisficing (e.g., speeding, non-differentiation) and potential misrepresentation (e.g., implausible entries, inconsistent metadata).

Results show that social group cues, particularly when targeting Muslim men, are associated with elevated rates of problematic response behaviour. Latent class analysis reveals three behavioural

Risks of misrepresentation can be heightened in LLM-generated surveys:

<https://surveyinsights.org/?p=21395&print=pdf>

Ethical Tensions

- Dual-use and manipulation:
 - Synthetic personas for message-testing, astroturfing, persuasion
- Accessibility, Transparency, and Reproducibility
 - How should end users access LLMs trained on personal data?
 - How do we prevent access for malicious actors?

scientific reports

OPEN

The potential of generative AI for personalized persuasion at scale

S. C. Matz^{1,2,3,4}, J. D. Teeny³, S. S. Vaid⁴, H. Peters¹, G. M. Harari⁵ & M. Cerf¹

Matching the language or content of a message to the psychological profile of its recipient (known as “personalized persuasion”) is widely considered to be one of the most effective messaging strategies. We demonstrate that the rapid advances in large language models (LLMs), like ChatGPT, could accelerate this influence by making personalized persuasion scalable. Across four studies (consisting of seven sub-studies; total $N = 1788$), we show that personalized messages crafted by ChatGPT exhibit significantly more influence than non-personalized messages. This was true across different domains of persuasion (e.g., marketing of consumer products, political appeals for climate action), psychological profiles (e.g., personality traits, political ideology, moral foundations), and when only providing the LLM with a single, short prompt naming or describing the targeted psychological dimension. Thus, our findings are among the first to demonstrate the potential for LLMs to automate, and thereby scale, the use of personalized persuasion in ways that enhance its effectiveness and efficiency. We discuss the implications for researchers, practitioners, and the general public.

Financial analysts have described people’s digital behavioral data as “more valuable than oil”^{1,2}. This is, in part, because such records afford one of the most effective forms of influence: *personalized persuasion*^{3,4}. Compared to non-personalized communication, matching the content of a persuasive message (e.g., its language or visuals) to the psychological profile of its recipient enhances its effectiveness (e.g.^{4,5}). On the one hand, such personalization offers tremendous opportunities to promote desired behaviors, including a healthy lifestyle^{6–8}, financial saving⁹, or support for environmentalism¹⁰. On the other hand, it can have a pernicious effect on societies¹¹, for example, increasing the spread of disinformation¹², manipulating political preferences^{13,14}, or promoting maladaptive consumer decision-making^{15,16}. We provide the first empirical evidence demonstrating how content generated by artificial intelligence (AI) can scale personalized persuasion by automating the creation of such messages with only limited information about the message recipient. As legislators increasingly consider whether (and how) to regulate generative AI¹⁷, our work suggests that AI-automated, personalized persuasion is poised to create an inflection point for the implementation and effectiveness of this influence tactic.

Risks of accurate representation can lead to highly personalized vectors of attack:

<https://www.nature.com/articles/s41598-024-53755-0>

Potential Ways Forward for LLM-generated Surveys

- Thinking about the goal of the survey
 - Pretesting and low-stakes vs. high-stake decisions
- Process traceability
 - Understand where can errors creep in the LLM-generated survey lifecycle
 - End-to-end benchmarks accounting for different design choices
- Disclosure and documentation of LLM-generated surveys
 - All design choices: LLMs, hyperparameters, prompts, personas

Table 5: Enhanced Disclosures

Disclosure Item	What to Report	Why it Matters
Access and Infrastructure		
Method of access	Disclose how the AI model was accessed: • Direct access (e.g., API) • First-party platform or tool (e.g. a provider-hosted chatbot or interface) OR • Embedded within a third-party platform or tool (<i>see next item</i>)	Access methods shape researcher control, transparency and reproducibility and may introduce distinct sources of error. For example, an interviewer bot embedded within

American Association of Public Opinion Research [Report on 'Responsible AI Integration in Surveys'](#)

Potential Ways Forward for LLM-generated Surveys

- Revisiting ethical norms for ‘human-subject’ studies
 - Trade-offs between privacy and LLM performance
 - Issues of informed consent
- Participatory design; include different stakeholders
 - Respondents, survey designers, Users of the survey data, ...
- Closing the validation-simulation loop
 - We need setups where we don't need to do the full human-generated survey to validate the LLM-generated survey

Potential Ways Forward for LLM-generated Surveys

- Revisiting ethical norms for ‘human-subject’ studies
 - Trade-offs between privacy and LLM performance
 - Issues of informed consent
- Participatory design
 - Respondents, survey designers, Users of the survey data, ...
- Closing the validation-simulation loop
 - We need setups where we don't need to do the full human-generated survey to validate the LLM-generated survey

Think-Pair-Share activity

Think-Pair-Share Discussion

- Turn to the person beside you and discuss one or more the following for the next 5 minutes:
 1. What are your concerns or reservations about LLM-generated surveys?
 2. *If* we could have valid and reliable simulation of any behavioral response, what would you be excited about?
 3. What do you think the field should focus working on?