Reply to Westwood: Questioning the empirical evidence that AI survey contamination is real and substantial

David M. Rothschild*[1], Soubhik Barari[2], Trent D. Buskirk[3], Andrew Gordon[4], and D. Sunshine Hillygus[5]

[1]Microsoft Research
[2]NORC at the University of Chicago
[3]Old Dominion University
[4]Prolific
[5]Duke University

Note: This manuscript was submitted to *Proceedings of the National Academy of Sciences* (PNAS) on February 27, 2026. Following rejection on May 22, 2026, we are publishing it here, as a pre-preprint, as originally submitted.

Competing interests: A.G. works for Prolific and D.M.R. previously advised Prolific.

Westwood [2025], followed closely by Van der Stigchel et al. [2026] and Westwood and Frederick [2026], argues that "AI contamination" poses a "potential existential threat of large language models to online survey research." Although AI (frequently LLMs) poses potential challenges for survey research, the articles overstate their case, conflating distinct risks and advancing claims of field-level vulnerability without (1) clear definitions, (2) appropriate benchmarks, or (3) reproducible demonstrations of real-world impact.

Lack of Clear Definitions: Westwood [2025] (and much of the related literature) **conflate three categorically distinct phenomena.** (i) *LLM-Assisted Responding*: human respondents using LLMs to assist with answers, typically on open-ended items, (ii) *Simple Agentic Survey Completion*: autonomous agents completing surveys, when a human gives them a link to complete, similar to humans, (iii) *Ecosystem-Embedded Autonomous Agentic Survey Completion*: autonomous agents completing surveys within a managed panel ecosystem, without human involvement. Fully autonomous agents are far more difficult to deploy in the field, require different evidence to establish, infrastructure to execute, and countermeasures to mitigate compared to simple agentic survey completions or humans using LLMs for response assistance.

***Simple Agentic Survey Completion* shown in Westwood [2025] is a far narrower finding than demonstrating that it can exist, participate repeatedly, and remain undetected within modern panel infrastructure.** The stronger claim is never actually evaluated in this body of work. First-party platforms such as Prolific and CloudResearch maintain persistent respondent identities, longitudinal behavioral profiles, device- and metadata-level verification, and population-level anomaly detection across survey completions. An agent that passes an isolated quality check must still establish a credible account history and avoid signals specifically designed to detect coordinated or synthetic activity.

*To whom correspondence should be addressed. E-mail: David@ResearchDMR.com

Lack of Appropriate Benchmarks: What Westwood [2025] and Westwood and Frederick [2026] actually explore are legitimate concerns, but **neither are new nor unsolvable.** For example, methods for detecting and mitigating LLM-assistance are well established and empirically validated (e.g., Veselovsky et al. [2025]), but there is no indication that Westwood and Frederick [2026] applied such methods prior to reporting their flagging rates. Thus, their reported rate of presumed LLM-assistance is likely inflated relative to what a typical researcher would observe after routine quality controls.

Lack of Reproducible Demonstrations of Real-World Impact: Westwood [2025] withholds its code, so the central empirical claims cannot be independently validated and stressed, while Westwood and Frederick [2026] does not provide the survey instrument, sampling protocol, or analytic pipeline. Without these standard disclosures it is **impossible to assess whether the reported rate of "AI contamination" reflects the construct under study or a methodological artifact**.

We encourage research that directly tests whether autonomous agents can operate within authentic survey ecosystems, maintain persistent respondent identities under platform scrutiny, and do so at sufficient scale to meaningfully affect inference without producing detectable anomalies. Such work (e.g., Martherus et al. [2025]) would enable the development of targeted, evidence-based solutions while avoiding unnecessary reputational spillover to survey methods largely insulated from these risks. The danger is not that concerns about AI and survey integrity are raised (*existence of risk*), but that extraordinary claims about systemic failures of scientific infrastructure (*existential risk*) are being advanced without necessary evidentiary and disclosure standards, risking a massive misallocation of resources away from the real challenges of AI disruption to survey research.

References

James Martherus, Alexander Podkul, Edgar Cook, and Robert Liebowitz. How to detect ai-assisted interviews in online surveys. *Survey Practice*, 18, 2025.

S Van der Stigchel, C Strauch, B de Zwart, and L Van Maanen. Will online behavioral research follow the fate of online survey research? *Proceedings of the National Academy of Sciences*, 123 (8):e2535585123, 2026.

Veniamin Veselovsky, Manoel Horta Ribeiro, Philip J Cozzolino, Andrew Gordon, David Rothschild, and Robert West. Prevalence and prevention of large language model use in crowd work. *Communications of the ACM*, 68(3):42–47, 2025.

Sean J Westwood. The potential existential threat of large language models to online survey research. *Proceedings of the National Academy of Sciences*, 122(47):e2518075122, 2025.

Sean J Westwood and Samuel Frederick. Reply to van der stigchel et al.: Empirical evidence that ai survey contamination is real and substantial. *Proceedings of the National Academy of Sciences*, 123(8):e2537420123, 2026.