The Quiet Take-Over of Surveys?*

How AI Agents are Infiltrating Surveys and Pathways Forward

Public Opinion Analytics Lab (POAL) Methods Brief Series

Lauren Leek

July 2025

Executive Summary

This brief outlines the emerging methodological challenge posed by Large Language Model (LLM) agents completing online surveys at scale - a trend that risks undermining the validity, representativeness, and credibility of survey findings. Drawing on interviews with leading UK public opinion firms, it finds that while some see AI-generated responses as an urgent threat, others emphasise that the deeper risk lies in broader panel integrity and recruitment: sustained respondent engagement, careful recruitment, and long-term trust often do as much to protect data quality as any bot-detection tool. Across the industry, quality assurance is shifting from static, one-off safeguards to adaptive, layered systems combining pre-survey gatekeeping, in-survey behavioural diagnostics, and post-survey coherence checks. These layers increasingly target a spectrum of fraud - from geo-spoofing and coordinated sign-ups to fully synthetic responses. Challenges remain - including false positives, resource demands, and the rapid evolution of both LLMs and fraud tactics - yet with continual updating and a focus on respondent relationships, surveys can remain a dependable instrument for political polling, market research, and public policy analysis.

1 Introduction

Surveys have served as the "workhorse" of social insight for nearly a century. While the modes and platforms have evolved - from door-to-door interviews to mobile push notifications - the foundational premise remains: ask a (probabilistic) sample and generalise to a population. Historically, the principal threat to this premise was non-response bias or simply 'bad' responses; today an additional adversary has arrived in the form of synthetic respondents powered by AI.

The barriers to creating an autonomous *survey bot* have collapsed. A competent programmer can now assemble a functioning pipeline in a single afternoon, using only off-the-shelf components. In practice, the interface automation - rather than the language generation - is the dominant time sink; the underlying natural-language tasks are trivial for current models. Estimating the prevalence of this is difficult since it is highly domain and survey dependent.

*This brief is based on posts from *Lauren's Data Substack*: part one and part two.

One study found the number of AI bots to be more than 90% in online panels.[1], however, the overall developments confirm that the core engineering hurdles are no longer technical but *institutional*: detection protocols, incentive redesign, and transparent provenance auditing must keep pace with the accelerating supply of synthetic respondents.

Rising synthetic completes matters downstream. Nowhere is this sharper than in political polling: turnout and vote-share models hinge on post-stratification assumptions that fail when bots inflate "typical" or median response patterns, dampen variance, and wash out hard-to-reach electorates. Weighting then "corrects" to the wrong signal, yielding stable yet systematically biased estimates that can misinform campaigns, media narratives, and even seat projections. Similar dynamics extend to market research (products built for a statistical average that no segment actually matches) and public policy (resource formulas blind to vulnerable groups when synthetic data fills gaps unevenly). Understanding these downstream distortions is essential: quality breaches at the response stage compound through modelling pipelines and can ultimately misallocate political attention, capital, and services.

2 The method: layered bot detection

Sophisticated fraud screening is no longer optional; it is now an *always-on, multi-layer workflow.* Survey companies run an end-to-end security pipeline to make sure real people - not automated "bots" - answer their questionnaires. In simple terms, there are three "layers" a respondent must walk through:

1. **Layer 1: Pre-survey sign-up checks.** When someone first joins a panel they must show they are real - for example by confirming an e-mail address, entering an SMS code, or uploading photo-ID. Some firms block known VPNs or suspicious internet providers at this stage. If this step is outsourced, accountability obligations to the sample providers are of importance.

2. **Layer 2: In-survey layered detection.** Traditional measures such as trap questions, CAPTCHA, and two-factor authentication are increasingly inadequate against sophisticated LLM-based bots. Modern detection systems now rely on behavioural telemetry - tracking indicators like typing speed, scroll depth, and response timing - to flag anomalous patterns. In parallel, consistency checks compare answers across the survey to identify contradictions (e.g., a respondent identifying as vegan but later reporting frequent steak consumption). These signals often feed into a multi-factor scoring system (e.g., on a 0-100 scale), with respondents falling below a defined threshold subject to real-time verification prompts or flagged for manual review.

3. **Layer 3: Post-survey cleaning.** After data collection, researchers conduct a final validation step using machine learning models to reassess response patterns and estimate the likelihood of fraud. This phase allows for recalibrating fraud probabilities, re-weighting

[1]Brittney Goodrich, Marieke Fenton, Jerrod Penn, John Bovay, and Travis Mountain. Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys. Applied Economic Perspectives and Policy, 2023. doi: 10.1002/aepp.13353

cases as necessary, and flagging residual inconsistencies or incoherent responses that may have bypassed earlier detection layers.

The table below summarises the key approaches both in the front-end and back-end taken by major survey companies to detect AI bots.

Platform	**Front-end gate**	**Back-end signals (examples)**
Opinium	Panel-provider audit (device fingerprint, recruitment source), mandatory metadata before launch	Real-time 0-100 risk score mixing IP overlap, typing rhythm, scroll depth, answer consistency; sub-threshold gets challenge question or manual review; second-pass forensic cleaning after fieldwork
YouGov	Focus on long-term panel relationships and preventative measures in addition to AI checks	Proprietary ML-driven fraud detection system, with ID verification for non-geo cases; robust geo-fraud monitoring (e.g., detecting organised sign-ups from influencer/Telegram groups); dedicated ethical-hacking team to red-team surveys and uncover vulnerabilities
Verian	Careful random recruitment via postal addresses (using the Postcode Address File database)	Sampling addresses ensures physical location is in the UK and by limiting the number of joiners from a single address, thus low impact if someone is sampled with technical know-how how to use AI bots. Regular data quality checks.
Qualtrics	Google Invisible reCAPTCHA v3 flags as *probable bot*	Duplicate-ID checks, device fingerprint, completion time, BallotBox stuffing flag (Qualtrics Survey Checker)
Prolific	Identity gate with multiple steps (e-mail, SMS, photo-ID) + ISP / VPN check on every login	Hand-graded writing sample, IP velocity, open-text coherence; $< 2\%$ of onboarded accounts later purged for quality (Prolific: Bots and Data Quality on Crowdsourcing Platforms)
SurveyMonkey	"Build-with-AI" coherence engine + duplicate-ID filter	Honeypots, speeding flags, profanity scan; internal red-team bot recorded *zero* final completes (SurveyMonkey: Market Research Data Quality)

Table 1: UK industry defences against AI bots. The insights are derived from interviews with Opinium, YouGov and Verian methodologists and online available information for Qualtrics, Prolific and SurveyMonkey.

3 Implementation guidelines

Safeguarding data quality is no longer a single checkpoint but an *end-to-end pipeline* - one that reallocates substantial resources from sampling to continuous fraud monitoring and respondent engagement.

Recommended Practice

1. **Careful recruitment:** if possible recruit randomly via postal addresses (e.g., from the Postcode Address File database).
2. **Stack complementary signals:** CAPTCHA → IP/device blocklists → behavioural telemetry → latent-semantic checks.
3. **Store raw para-data:** millisecond timestamps, focus-change events, typing cadence - essential for audit trails and justifying exclusions.
4. **Budget for manual review:** even a 1-2% flag rate on a 50k respondent study yields hundreds of borderline cases for online panels.
5. **Refresh rules quarterly:** bot tactics evolve in lock-step with LLM releases; a static rule set leaks within months.
6. **Invest in long-term panel engagement:** strong relationships with respondents - through careful recruitment, retention incentives, and regular communication - reduce incentives for fraud and improve overall data quality.