CATALYSTMR
transforming market research data into marketing solutions

METHODOLOGY PAPER

DATA QUALITY · FRAUD & AI DETECTION

Survey *Fraud*

AI-Generated Responses and New Detection Methods

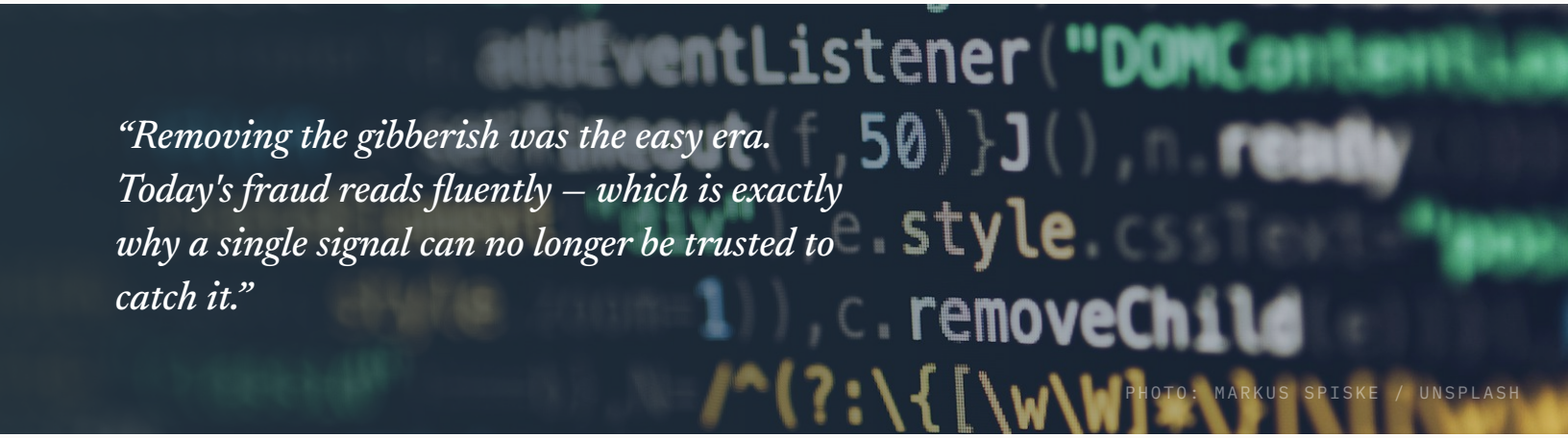
Survey fraud has evolved beyond duplicates and obvious speeders. Synthetic identities, automated scripts, and generative AI can now pass screeners and produce plausible-looking answers. This is a vendor-neutral guide to the modern threat — and the layered stack of behavioural and technical signals, backed by human review, that detects it.

PUBLISHED BY
CatalystMR Research Team

SERIES
Methodology Papers

READING TIME
~16 minutes

EDITION
2026



“Removing the gibberish was the easy era. Today's fraud reads fluently — which is exactly why a single signal can no longer be trusted to catch it.”

PHOTO: MARKUS SPISKE / UNSPLASH

CATALYSTMR
transforming market research data into marketing solutions

ABSTRACT

For years, fraudulent survey respondents gave themselves away — duplicate IPs, impossible completion times, gibberish open-ends. That era is ending. Bad actors now use VPNs and synthetic identities, automated scripts and survey farms, and generative AI that produces fluent, on-topic open-ends able to pass naïve checks. The tells researchers relied on are exactly the ones modern fraud has learned to avoid.

This paper is a vendor-neutral guide to the current threat and how to meet it. It maps how survey fraud has escalated; explains why **no single signal is sufficient** any longer; sets out a **layered detection stack** of behavioural and technical signals; argues why **human review** remains decisive against fluent fraud; and frames detection as an ongoing arms race rather than a fixed checklist. It is the threat-and-detection companion to the eleven-point QC framework of Paper No. 137.

CONTENTS

WHO THIS IS FOR

Client-side insights and research buyers who need confidence in the data behind a decision; **agency and brand researchers** fielding online studies; and **procurement and data-quality leaders** evaluating how a provider detects fraud. No technical background is assumed; practitioners can skip to Section 04.

01 THE THREAT

Fraud no longer looks like fraud.

Historically, many fraudulent respondents were caught by the obvious: duplicate IPs, impossible completion times, or open-ends full of gibberish. Those signals still matter — but on their own they now catch only the laziest fraud. The economics of paid research attract organised, technically capable bad actors, and the tools available to them have changed faster than many QC routines have.

Three shifts that broke the old tells

SHIFT 1

Synthetic identities

VPNs, proxies, disposable emails, and fabricated profiles let one actor appear as many distinct, plausible respondents.

SHIFT 2

Automation at scale

Scripts and bots — some able to defeat naïve CAPTCHAs — complete surveys en masse, and survey farms coordinate human click-work.

SHIFT 3

Generative AI

AI can produce fluent, on-topic, non-duplicated open-ends that read like a real respondent and sail past exact-match checks.

Peer-reviewed research into online recruitment has documented just how quickly bots and bad actors can overwhelm a study, and how sophisticated their evasion has become — some bots bypassing CAPTCHAs and generating false contact details.[4] The shift is not that fraud appeared; it is that fraud learned to look legitimate.

KEY INSIGHT

The fluency of modern fraud is the whole problem. When a fraudulent open-end is grammatical, on-topic, and unique, the single checks that once defined QC — de-duplication, a gibberish scan, a speed threshold — quietly stop working. Detection has to move from spotting obvious defects to reading patterns across many signals at once.

02 ESCALATION

Every old tell has a modern evasion.

The clearest way to see why detection had to change is to set the fraud of a few years ago beside its 2025 form. In each row, the defence that used to suffice on its own is now just one signal among many — because the threat has adapted around it.

DIMENSION	THEN — THE EASY ERA	NOW — FLUENT FRAUD
Identity	One IP, one obvious duplicate	**VPNs, proxies & synthetic identities** mask one actor as many
Volume	A person submitting twice	**Bots & scripts** at scale; coordinated survey farms
Open-ends	Gibberish or copy-paste	**AI-generated text** — fluent, on-topic, non-duplicated
Speed	Impossibly fast completes	**Human-paced bots** that mimic realistic timing
Screening	Obvious wrong answers	**Coached / pattern-matched** screener answers that qualify

WHAT THE ESCALATION IMPLIES

If each row's old defence can now be individually defeated, then any QC approach resting on one or two signals is, by construction, beatable. The response is not a better single test — it is **layering**, so that defeating one signal still leaves a respondent exposed by others. That is the subject of the next two sections.

03 THE PRINCIPLE

Detection is a question of patterns, not a single test.

Every individual fraud signal has a blind spot a determined actor can exploit. A VPN check is defeated by residential proxies; a speed threshold by a bot that waits; an exact-match de-duplication by AI that paraphrases. The power of modern detection comes not from any one check being perfect, but from **combining independent signals** so the evasions required to beat all of them at once rarely co-occur.

Independent signals, compounding odds

A respondent who clears the device-fingerprint check may still straight-line the grid; one who writes a fluent open-end may still arrive on a flagged proxy from an out-of-quota geography. Each signal a fraudster must satisfy multiplies the effort and cost of passing — until, for most, the study is no longer worth gaming. Layering does not need any single signal to be infallible; it needs the signals to fail independently.

Why AI raised the stakes

Generative AI specifically neutralised the open-end — historically one of the strongest single tells, because gibberish and copy-paste were easy to spot. Now that a fraudulent verbatim can be fluent and unique, the open-end must be read for pattern (uniformity, generic phrasing, mismatch with stated experience) and corroborated against behavioural and technical signals — not trusted on its surface.

> THE SHIFT IN ONE LINE
>
> Old QC asked, of each complete, **"is this answer obviously bad?"** Modern detection asks **"do these signals, together, describe a real person?"** — and treats the open-end as evidence to be corroborated, not a verdict on its own.

04 THE STACK

Eight signals, layered — no respondent judged on one.

Modern fraud detection combines technical and behavioural signals across the field. Each layer below catches something the others can miss; a complete is judged on the weight of the stack, not any single flag. The first layers act before and during fielding; the last reaches into the finished data.

01 **Device fingerprinting** — recognise the same device or browser behind ostensibly separate respondents. PRE-FIELD / IN-FIELD

02 **VPN / proxy detection** — flag known anonymising infrastructure and out-of-market geographies. PRE-FIELD / IN-FIELD

03 **Duplicate detection** — across sessions and across blended sources, not just within one. IN-FIELD

04 **Response-time analysis** — speeders and, increasingly, suspiciously uniform human-paced timing. IN-FIELD

05 **Straight-lining & pattern detection** — flat or templated answering across grids. IN-FIELD

06 **Screener consistency** — logically linked questions whose answers must agree. IN-FIELD

07 **Open-end quality & AI-text review** — read for gibberish, copy-paste, and generated-text patterns. POST-FIELD

08 **Source-level anomaly monitoring** — spikes in flags concentrated in one source or batch. IN-FIELD / POST-FIELD

NO SINGLE LAYER QUALIFIES OR REJECTS A RESPONDENT — THE STACK DOES, TOGETHER

05 HUMAN REVIEW

The last layer is a person — because fluent fraud is built to pass machines.

Automated signals are essential and do most of the volume, but they are tuned to patterns; the most advanced fraud is engineered specifically to satisfy those patterns. Experienced reviewers catch what automated validation cannot — context. The ICC/ESOMAR Code's emphasis on **human oversight of automated processes** is not a formality here; against generated text it is the decisive layer.[1]

What a human reviewer catches

- **Terminology misuse** — answers that use the words but not the meaning of a field or condition.
- **Conflicting context** — verbatims that contradict screener answers or each other.
- **Unnatural phrasing** — text that is fluent yet generic, evasive, or oddly uniform across respondents.
- **Plausible-but-wrong** — completes that pass every automated check and still do not read like a real person.

Machine and human, not one or the other

The point is not that humans outperform automation — at scale they cannot. It is that the two fail differently: automation handles volume and catches the technical signals; human review catches meaning and context. Generated text is designed to beat the first, which is precisely why the second has become indispensable rather than optional.

> BUYER'S QUESTION
>
> Ask: **"What share of open-ends does a person actually read — and who?"** A provider relying on automated scoring alone is, against AI-generated text, fighting the last war.

06 THE ARMS RACE

Every counter invites the next evasion.

Detection is not a checklist you complete once; it is a moving contest in which each defence prompts a new evasion, and each evasion a new defence. The pairs below show the current state of play — and why the only durable posture is a layered stack that is monitored and updated, not a fixed test.

THREAT		COUNTER
AI-generated open-ends that are fluent and unique	→	**Pattern & human review** of verbatims, not exact-match de-duplication
Residential proxies that defeat simple VPN lists	→	**Device fingerprinting** + geo/quota anomaly monitoring across sources
Human-paced bots that mimic realistic timing	→	**Behavioural & consistency signals** the timing fix alone doesn't satisfy

Two questions to put to a provider

- ✕ **How many independent signals stand between a fraudster and a delivered complete** — and do they include AI-text and source-level anomaly review, not just duplicates and speed?
- ✕ **How is the detection stack kept current** as evasion tactics change — is it monitored and updated, or a fixed routine set up once and left?

WHERE THIS SERIES GOES DEEPER

No. 137	**The QC framework.** The eleven-point pre-field / in-field / post-field process these signals live inside.
No. 142	**Respondent validation.** The end-to-end QC playbook for confirming a respondent is who they claim.
No. 143 · 144	**Speeder detection & straight-lining.** Two stack signals examined in depth.

CATALYSTMR
transforming market research data into marketing solutions

▲ CONCLUSION

Fluent fraud needs layered detection.

Survey fraud has crossed a threshold: it now reads like a real respondent. The defences that once worked alone — de-duplication, a gibberish scan, a speed limit — still belong in the kit, but no longer suffice on their own. The durable response is a layered stack of independent behavioural and technical signals, an open-end discipline that reads for generated-text patterns rather than trusting fluency, a human reviewer as the last layer, and the humility to treat detection as an arms race that must be monitored and updated. The ICC/ESOMAR Code, ESOMAR 37, and ISO 20252 frameworks exist precisely so buyers can ask for this rigour in consistent terms — and recognise it when they see it.[1,2,3]

§ REFERENCES

[1] ICC/ESOMAR. **International Code on Market, Opinion and Social Research and Data Analytics.** ICC & ESOMAR. (Transparency and human oversight of automated processing.)

[2] ESOMAR. **37 Questions to Help Buyers of Online Samples.** Amsterdam: ESOMAR, 2023. esomar.org

[3] International Organization for Standardization. **ISO 20252:2019 — Market, opinion and social research, including insights and data analytics — Vocabulary and service requirements.** Geneva: ISO, 2019.

[4] Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsky JA, Hong F, et al. **Threats of Bots and Other Bad Actors to Data Quality Following Research Participant Recruitment Through Social Media: Cross-Sectional Questionnaire.** J Med Internet Res 2020;22(10):e23021. doi:10.2196/23021. (PMID 33026360; PMCID PMC7578815.)

Standards and codes are cited for the conduct and quality-management frameworks referenced; the peer-reviewed study [4] is cited for the documented nature and speed of bot/bad-actor threats to online research — not for any rate figure transferred to this paper. The detection signals named are established industry techniques. This paper publishes no fraud, contamination, or detection-rate figures; any such metric should be measured from the specific study's own data, and any vendor "fraud-free" or detection-rate guarantee requested and avoided.

§ ABOUT

CatalystMR

CatalystMR is a global market-research panel and fieldwork partner specialising in **hard-to-reach B2B, healthcare, and niche audiences**. We layer technical and behavioural fraud signals across pre-field, in-field, and post-field stages — backed by experienced human open-end review — and apply them to every project as the standard, not an optional tier.

We publish our own responses to ESOMAR's 37 Questions and treat fraud detection as a monitored, evolving discipline rather than a fixed end-of-field scrub.

Compliance posture: our methodology is aligned to the ESOMAR Code and Guidelines and the ISO 20252 framework, and we are certified under the EU-U.S., UK, and Swiss Data Privacy Frameworks, with personal data siloed from response data.

SURVEY FRAUD · AI DETECTION · BOTS & SYNTHETIC IDS · OPEN-END REVIEW · ESOMAR 37 · ISO 20252

Ask us how we'd detect fraud on your study — the signals we layer, what a person reviews, and how we keep the stack current — and we'll return a modelled feasibility range, typically within 24 hours.

sales@catalystmr.com
1-800-819-3130
catalystmr.com