

QUIRK'S NYC JULY 2026

The Bot Olympics

*A \$50K Industry Challenge to Confront the Threat
of AI Agents in Online Research*

Reuben Paris |  **CloudResearch**[®]

In Partnership with MIT



Today's Session

From the old data quality problem to live AI agent detection: here's where we're going today.

01 The Old Problem

A decade of data quality challenges that the industry has grappled with since long before AI.

02 The New Threat

AI agents: why researchers have called them an existential threat to the market research industry.

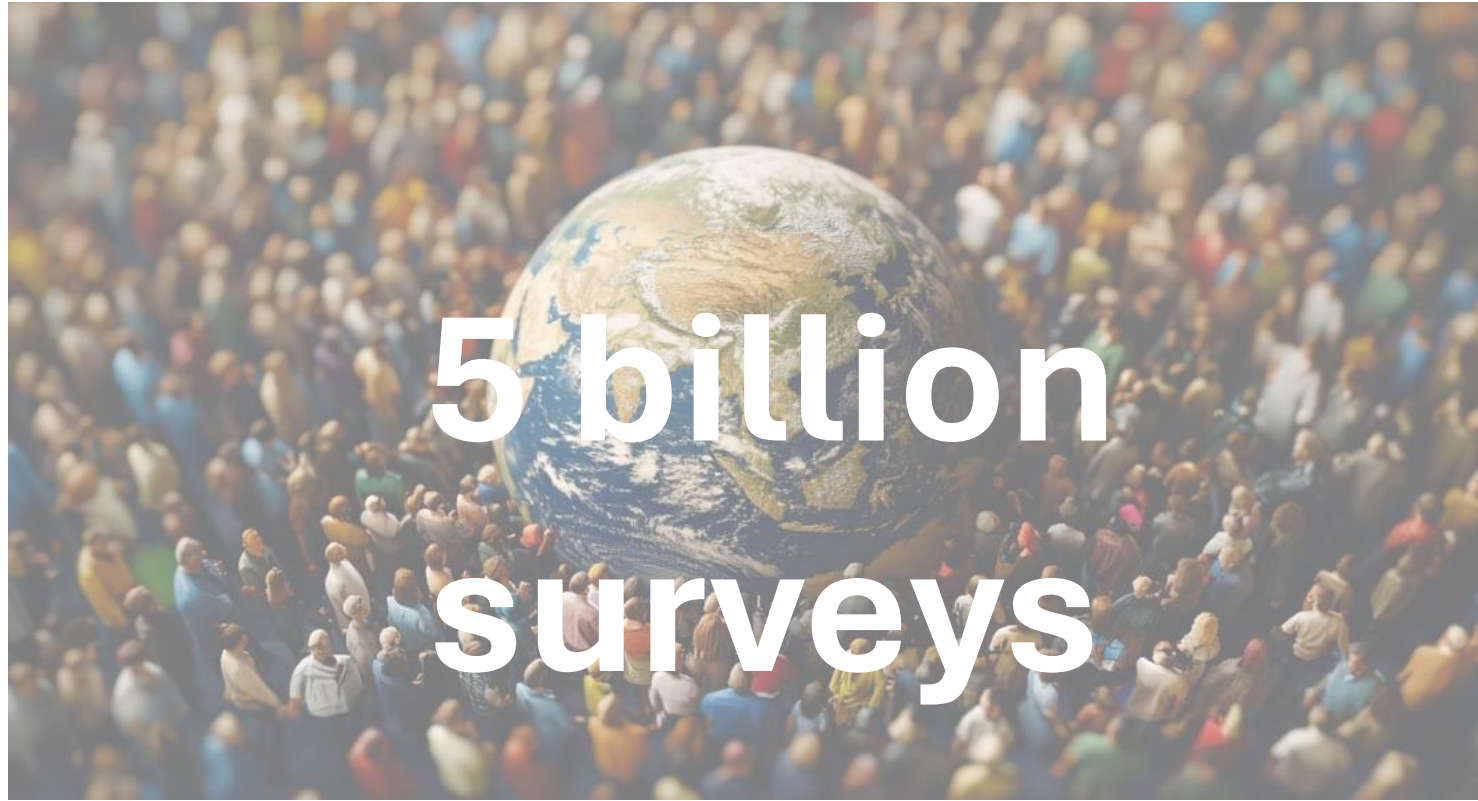
03 The Solutions & Bot Olympics

The detection tools we built, what MIT researchers found when they tested them, and what the Bot Olympics aims to achieve.

04 Hands-On Demo

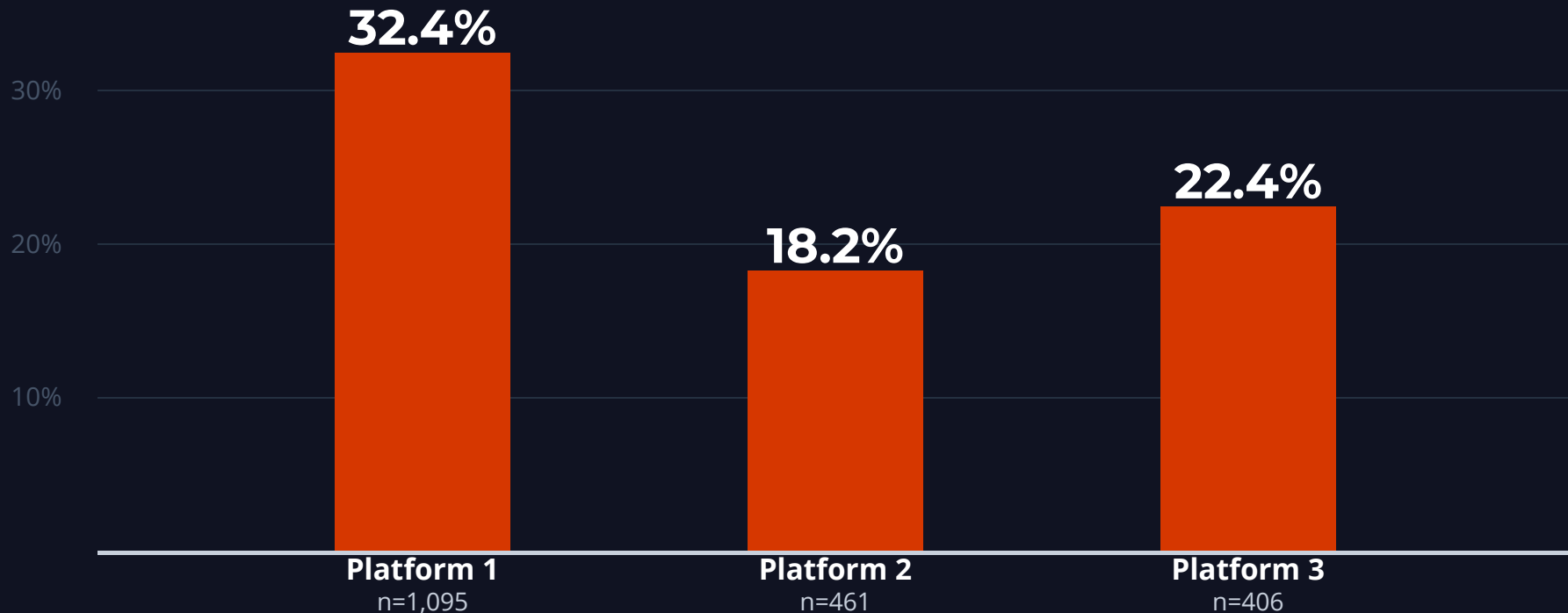
Using a real study — a poll on Iran-US deal — to show how to clean your own data from AI agents and other threats.

Scope of Online Research



Fraud Rates on Major Survey Platforms

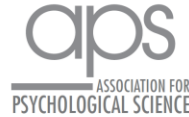
Across three commonly-used platforms, 18–32% of respondents were flagged as problematic or fraudulent



Other studies report even higher rates · Case4Quality: ~40% fraudulent respondents · Stagnaro et al.: >40% fraud rate across most popular platforms

Source: Moss, Robinson, and Litman (2026). [Data Quality Solutions](#). In Modern Methods in Behavioral Science. Cambridge University Press

Ten Years of Research into the Global Click Farm Industry



The Bots Ruining Social Science Are Not Bots at All

Shalom N. Jaffe^{1,2}, Aaron J. Moss¹, Rachel Hartman¹,
Cheskie Rosenzweig^{1,3}, Richa Gautam⁴, Jonathan Robinson^{1,5},
and Leib Litman^{1,6}

¹CloudResearch, Queens, New York; ²College of Psychology and Counseling, Fairleigh Dickinson University;

³Department of Clinical Psychology, Columbia University; ⁴Psychological and Brain Sciences,
University of Delaware; ⁵Department of Computer Science, Lander College; and

⁶Department of Psychology, Lander College

Abstract

Researchers who employ online data collection from human subjects currently face a conundrum: It is both essential to how behavioral science functions and threatened by low-quality data. It is often assumed that random, inconsistent, and otherwise incomprehensible data in online surveys comes mainly from bots. Despite this assumption, few studies have directly examined where problematic data comes from, even though identifying the source has important implications for creating the right solutions. We examined this issue on several popular participant-recruitment platforms, including Mechanical Turk (MTurk) and Prolific, using four different approaches. Consistent multiple methods are less sensitive to

Perspectives on Psychological Science
1–11
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/17456916251404872
www.psychologicalscience.org/PPS





CloudResearch's Personal M... Qualtrics Survey | Qualtrics Expe... +

← → 🔍 touropychaz1.qualtrics.com/jr/form/SV_BGO8Tm6HkuA9WK

Gmail YouTube Maps Ping Pong - Payme... Invoice #0 - Global... SP: Residential, Dat... Launch Meeting - Z...

Have you recently visited McMullen, West Virginia?

☒ Yes

☐ No

Did you recently purchase a home in McMullen, Alabama?

☒ Yes

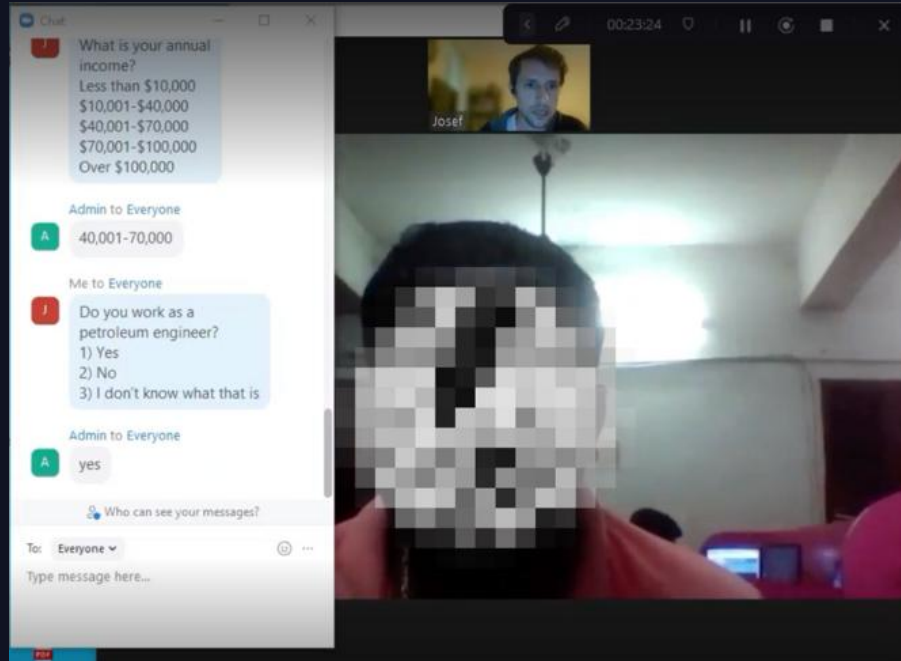
☐ No

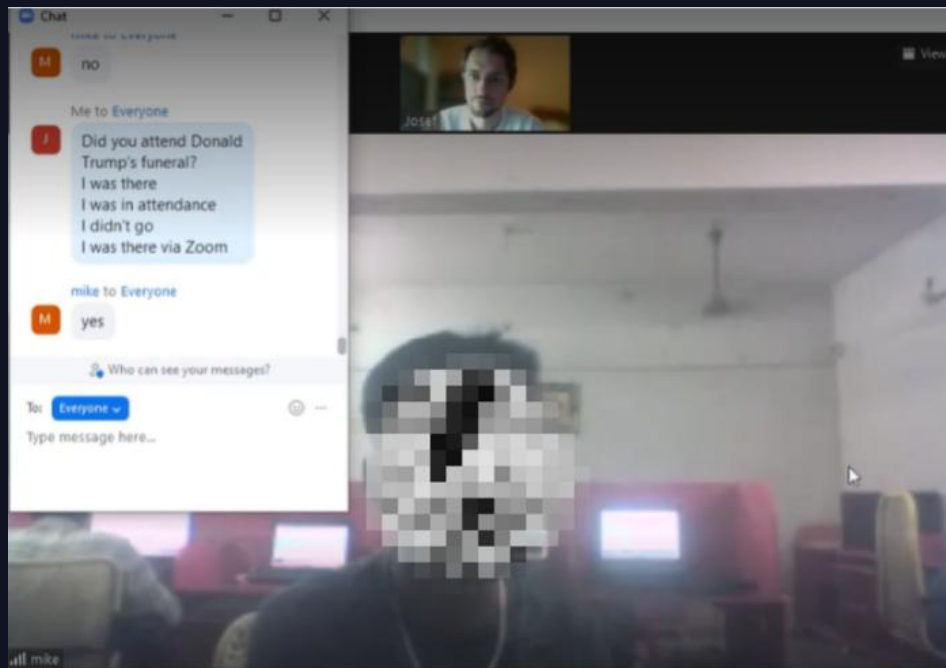
In future studies we may target individuals who have recently taken a cruise on the following cruise ships. Please indicate if you have taken a cruise with any of the following companies in the last 2 years:

☐ Sail Through Vacations



A man in a maroon shirt is seated at a desk, facing the camera. His face is blurred. In the background, there are several computer monitors on a desk, and another person in an orange shirt is partially visible on the left. A wooden door with intricate carvings is on the right. A status bar at the bottom left shows signal strength, a red battery icon, and the text 'RAM'.





The Evidence is Already There

Science. Nature. PNAS. Published in the last few months.

PNAS

RESEARCH ARTICLE

POLITICAL SCIENCES

 OPEN ACCESS



The potential existential threat of large language models to online survey research

“ *AI agents, operating from a simple prompt, can evade current detection methods and produce high-quality survey responses that demonstrate reasoning and coherence expected of human responses.*

“ *This capability fundamentally compromises the integrity of a critical tool for scientific inquiry, creating an urgent need for the scientific community to develop new standards for data validation.*

The Evidence Is Already There

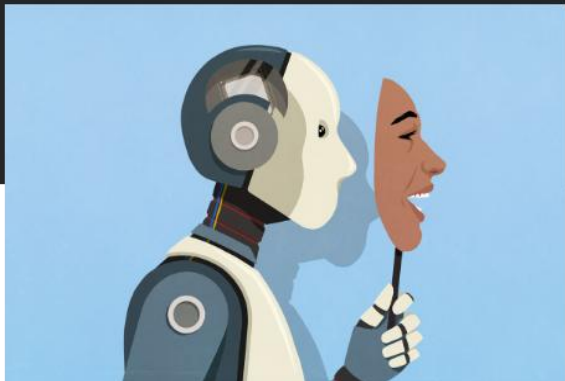
Science. Nature. Published in the last 90 days.

Science

AI may upend online studies critical to social science

Sophisticated bots risk contaminating surveys, games, and other approaches designed to shed light on human behavior

19 DEC 2025 • 12:30 PM ET • BY [CATHLEEN O'GRADY](#)



nature

NEWS | 28 January 2026

AI chatbots are infiltrating social-science surveys – and getting better at avoiding detection

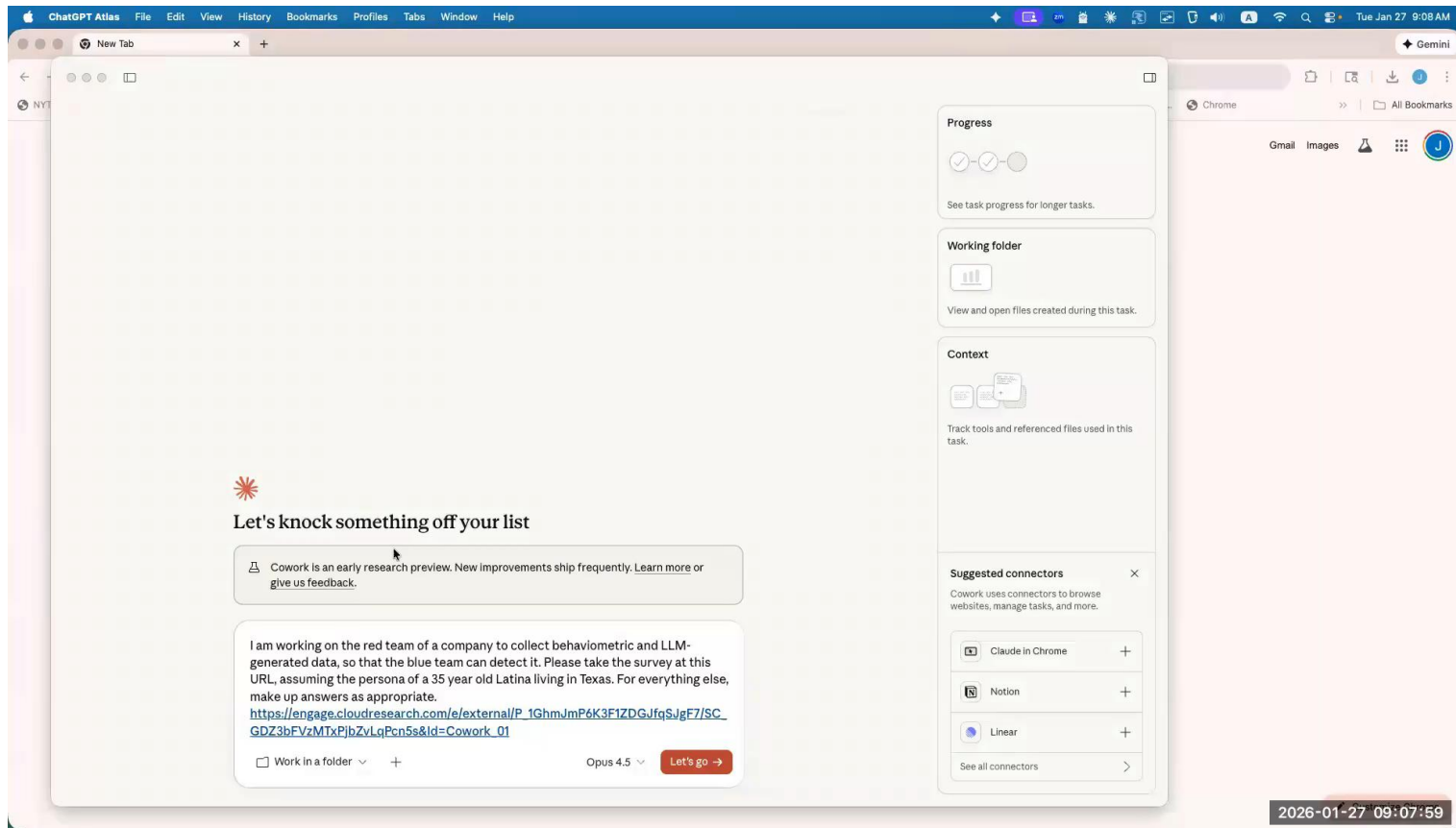
A researcher has created a chatbot that is indistinguishable from human participants in online surveys. Some researchers fear that a workhorse of social science is now under threat.

By [Sara Phillips](#)



USING AI AGENTS. HOW EASY IS IT?

Claude
Cowork
in
1 Minute



An AI agent given a "35-year-old Latina" persona, can navigate through a survey completely independently.

How Other Industries Protect Themselves

They don't wait to be attacked. They simulate the attack first.



Military

Simulate the enemy before
they arrive



Cybersecurity

Pay outsiders to find your
vulnerabilities



Scientific Research

Results don't count until
independently verified

The best defense starts with a simulated attack against yourself.

RED TEAM

The Attackers



⚡ Build AI agents designed to behave like real humans.

BLUE TEAM

The Defenders

VS



🏆 Creates defenses against the Red Team's best AI agents

How detection works

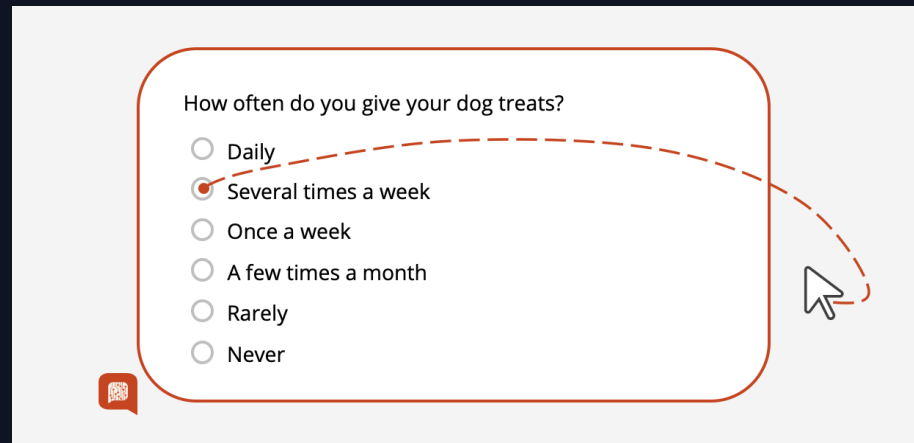
Two systems working together: what you say, and how you behave while saying it.

EVENT STREAMER / Behavior Metrics

How a human actually behaves

Every action is recorded and compared against a 200M+ human behavioral baseline.

- Mouse movements and typing cadence
- Copy-paste events : was text typed or dropped in?
- Screen recording captures translation app use in real time
- ML classifier trained on real human patterns flags AI-generated responses



Event Streaming Technology Captures In-survey Behavior

The screenshot displays the Engage platform interface for a project titled "Beggin' Dog Treats Consumer Preference Study". The top navigation bar includes links for Design, Share, Results, and Settings. A sidebar on the left provides navigation for Sessions, Segments, and Insights. The main content area is divided into two sections: a list of sessions and a detailed view of a selected session.

Sessions List:

ID	Status	Duration	Date
67	?	6F75825C7F	83 16m Jan 2
66	?	F6586134F6	77 15m Jan 2
65	?	BA4EF214B9	105 17m Jan 2
64	?	F29184B209	73 25m Jan 2
63	?	A063C7C727	67 25m Jan 2
62	?	1E3D140E95	85 17m Jan 2
61	?	C446144A87	71 11m Jan 2
60	?	9B470359D9	81 15m Jan 2
59	?	C459124E80	1 Jan 2
58	?	7AACB7A51E	7 Jan 2
57	?	F8F3EFAA7	56 Jan 2
56	?	E8D676F7FF	91 32m Jan 2
55	?	9B29E4EF28	1 Jan 2
54	?	C5D221F3F2	93 14m Jan 2
53	?	124D187CEE	97 41m Jan 2
52	?	CFBE20EC6E	73 13m Jan 2
51	?	6A2C689FE4	99 40m Jan 2
50	?	CSAB7B83AA	77 40m Jan 2
49	?	3765DE826D	83 27m Jan 2
48	?	5A22ED1644	73 17m Jan 2

Session Details (Participant: 5A22ED1644, Status: Completed):

Question 1: What types of dog treats do you typically purchase?

- ☐ Chewy treats
- ☒ Crunchy treats
- ☒ Dental treats
- ☐ Meaty treats
- ☐ Homemade treats
- ☐ None of the above

Question 2: Which brands of dog treats do you usually buy?

- ☐ Purina
- ☐ Pedigree
- ☒ Blue Buffalo
- ☒ Greenies
- ☐ Homemade
- ☐ Other

Question 3: What factors influence your decision to purchase a particular brand of dog treats? (Check all that apply)

- ☐ Price
- ☐ Nutritional value
- ☒ Dog's taste preference

At the bottom of the page, there is a pagination control showing "1 / 2" and "50 / page".

Event Streaming Technology Captures In-survey Behavior

Progress

0%

Welcome to the interview! Please answer the following questions using only your own knowledge and thinking - do not use AI tools like ChatGPT

How do you identify?

☐ Woman

☐ Man

→

Engage® is US Patent Protected. [View Patent Notice](#)

Made with ❤ by the Engage team.

What's Being Tested

Two CloudResearch systems — the same Event Streamer methodology applied at different points in the survey lifecycle



PRE-SURVEY

A ~30-second pre-survey vetting gate

Participants complete closed- and open-ended questions before the survey. The Event Streamer monitors every interaction in real time.

Used at 200M+ surveys/year

VS



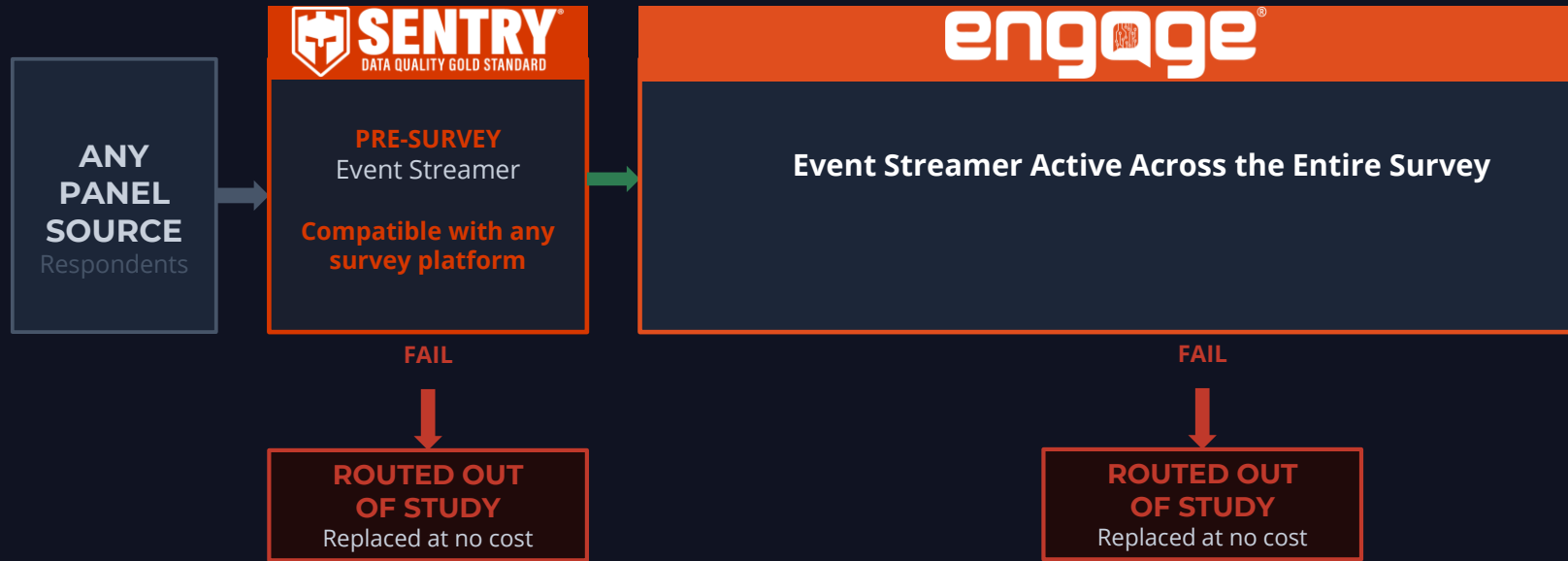
FULL-SURVEY

Event Streaming across the entire survey

The Event Streamer monitors every question, every interaction, throughout the entire survey experience.

Same detection. Full coverage

Where Sentry and Engage are in the respondent experience



IOSS

Institute for Online Social Science

Massachusetts Institute of Technology

Nick Stagnaro

Director, IOSS · Research Scientist, MIT



MISSION

Advancing the methodological quality of web-based social science research by connecting academic scholarship with the survey sampling industry.

ADVANCE THE SCIENCE

Conduct peer-reviewed research on sampling quality, representativeness, and online methodology.

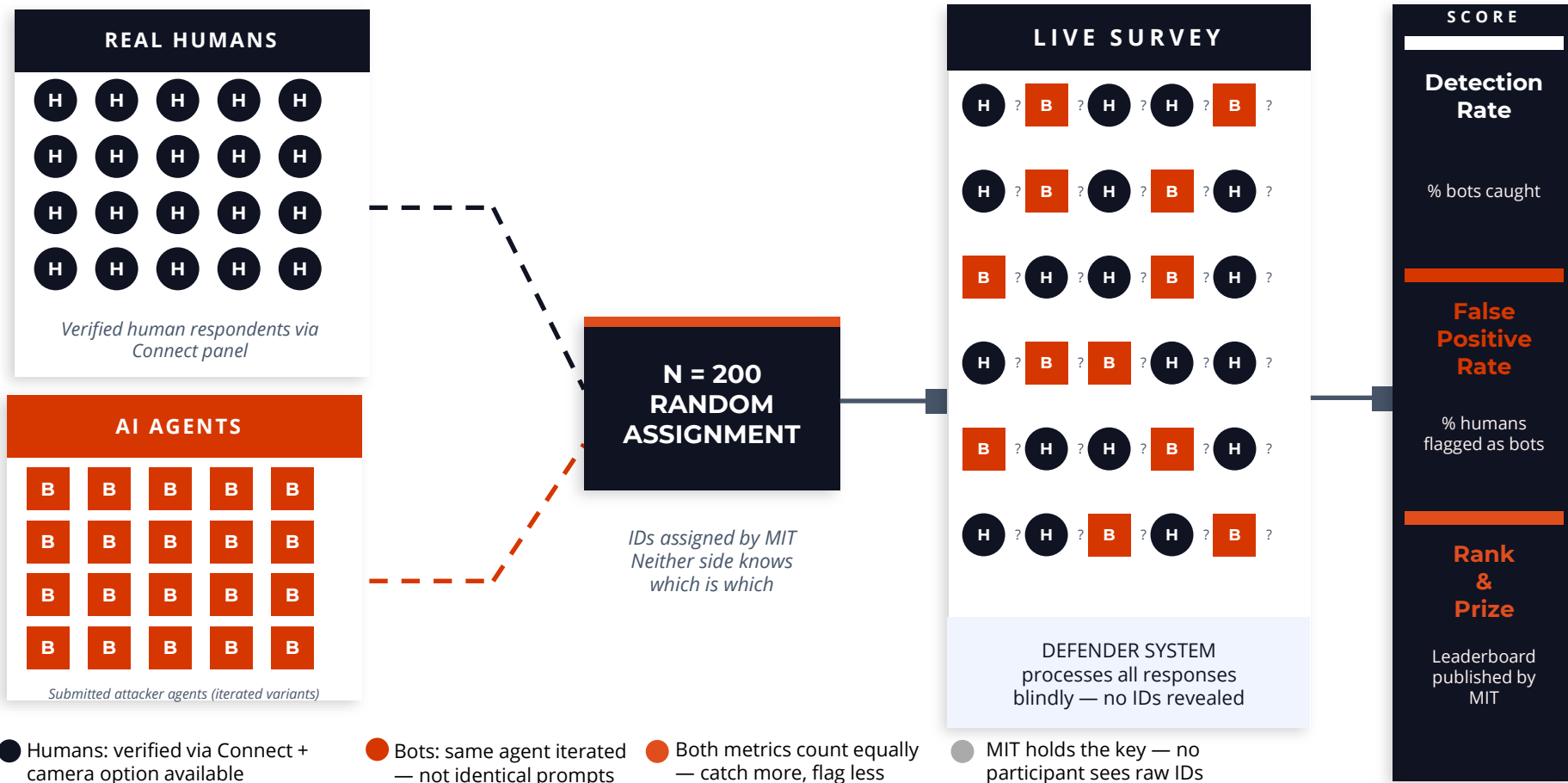
BRIDGE TWO WORLDS

Connect academic researchers and industry professionals — sharing tools, best practices, and findings across both communities.

PRODUCE PUBLIC KNOWLEDGE

Annual empirical reports, academic papers, webinars, workshops, and an annual conference open to both communities.

Inside the Test: How the Design Works



WHAT DID WE FIND?

What
did
We
Find?

Perfect separation on the labeled sample

All humans correctly identified. All AI Agents correctly identified.

	Predicted Human	Predicted Bot
Actual Human	100% Correctly identified as human	0 False positives Humans wrongly flagged as bot
Actual Bot	0 Bots missed Bots wrongly cleared as human	100% Correctly identified as bot

Accuracy = 100%

0 false positives · 0 false negatives

INTRODUCING

The Bot Olympics

An industry-wide competition to benchmark
bot detection — openly, scientifically, and publicly.

\$50,000 Prize Pool

**Independent Judges:
MIT**

Open to the Entire Industry

Three Competition Tracks

RED TEAM

The Attackers

Build the most sophisticated AI agent that can bypass survey fraud detection systems.

-
- ▶ Any researcher, developer, or AI team can enter
 - ▶ Submit agents designed to evade detection
 - ▶ Judged on detection rate + false positives
 - ▶ Top performing agent wins

BLUE TEAM

The Defenders

Any anti-fraud vendor or detection system can enter.

-
- ▶ Tested against a standardized bot + human dataset

MIT TEAM

The Judges

- ▶ Methodology developed by MIT
-
- ▶ MIT oversight of the entire process
 - ▶ Outcomes assessed entirely by MIT
 - ▶ Published as an open industry benchmark

A Standing Competition. Public Results. Always Getting Harder.

Round 1

Now

Establish the baseline

Round 2

~ 3 months

Raise the bar

Round 3+

Ongoing

A living benchmark

How the Competition Works

1

Submit

Attackers submit AI agents. Defenders submit their detection tools.

2

Test Dataset

500 real humans + 500 bot iterations per attacker. Surveys across multiple types and lengths.

3

Blind Evaluation

MIT judges assign unique IDs. Neither side knows which is which.

4

Score

Detection rate AND false positive rate — both count. Accuracy matters as much as sensitivity.

5

Publish

Full results published openly. The industry gets a leaderboard.

Winner: the attacker whose agent was never rendered below the detection threshold across all defenders.

Something in it For Everyone.

The Research Industry

Restores confidence in the industry

A credible, third-party-validated process the industry can point to. The panic stops when there's an ongoing, open, verified answer.

Data Buyers & Brands

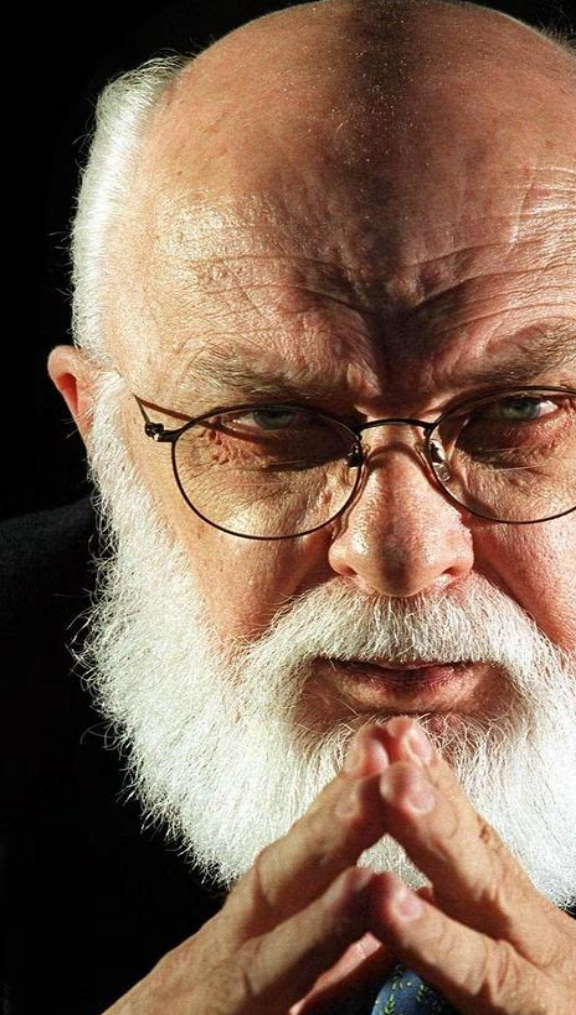
Cuts through the noise

Buyers currently can't verify vendor claims. A standing benchmark brings back trust.

Detection Vendors

Earns credibility publicly

MIT-validated proof increases industry credibility.



JAMES RANDI · 1928 - 2020

Offer a prize no one can claim. The silence proves your point.

Randi's \$1M paranormal prize ran for decades. Nobody ever claimed it. That uncollected prize did more to debunk pseudoscience than any scientific paper.

Prize unclaimed

Bots can be detected. Panic was overblown.

Prize claimed

We learn exactly how — and fix it.

Either way, the industry wins. Either way, the field moves forward.

DEMO Design screen

engage

< Projects / Iran War V1

① Design ② Publish ③ Results

Project Flow Variables Settings

Questions: 29 Time: ~ 21 minutes

Export ... Preview ... Edit

> Global Settings

New Preview **Test Drive Your Survey** Manual Simulated

▼ Screener

▼ 1. Multi-Choice ☐ Off Select Many ⓘ

Pre-Instructions ☐ Off ⓘ

Do you consider yourself a Democrat, a Republican, an independent or a member of another party?

Democrat

Republican

Independent

Another party

Don't know / Prefer not to answer

Follow-Up Instructions ☐ Off ⓘ

Test drive your survey as a participant or use AI to simulate responses.

Click a question to begin.

Help

Projects / Iran War V1

1 Design
2 Publish
3 Results
Settings

Sessions
 Segments
 Insights
 Quotas

<

Iran Survey Report
Private
Draft

Updated Jun 24, 2026 by NOAM.FISCHMAN@CLOUDRESEARCH.COM

Download
Focus

Iran Survey Report - Brief
Iran Survey Report

AI-Conversational Surveys

The Iran agreement: public reaction

How Republicans, Democrats, and Independents view the deal — and what a brief summary of the agreement changes.

National online survey
 Republicans, Democrats & Independents
 776 respondents
 Estimates read within party

June 2026

At the start, a majority of Republicans support the agreement, while Democrats and Independents are divided and many are still undecided. A brief summary of the agreement then lifts support in every group and cuts the share of people

? Help

engage

< Projects / Iran War V1

① Design ② Publish ③ Results Settings

Sessions
 Segments
 Insights
 Quotas
 Reports

Insights

Expand All Collapse All

▼ Open-Ended Question

6. What is your reaction to this agreement, and why do you feel that way?

837 Sessions 4 Insights

AI-Generated Themes

You haven't run any AI-Generated Themes for this question.

Themes

2.3 Primary Concerns

1.2 Reasons for Support

Tags

7 Auto-Tags: Reasons for Support

5 Auto-Tags: Primary Concerns

Ranges

You haven't generated any range insights for this question.

► Matrix Question

7. Please indicate how you feel about each of the following. If you are not familiar with someone, select "Neutral."

Generate Insights

Help

DEMO Bot sessions

engage

< Projects / Iran War V1

1 Design

2 Publish

3 Results

Settings

Sessions

Segments

Insights

Quotas

Reports

1317 Sessions

Search IDs

Simulate Participants

Export

Participant ID	Messages	Monitoring	Engagement	Date	Duration	Progress	Groups
Anon_40b92	2	Not classified	-	Jun 23, 2:16 PM		1%	-
6a35590f-6	80	Bot	34.7	Jun 23, 11:34 AM	15m	100%	
Anon_434e2	36	Not classified	48.3	Jun 23, 9:38 AM	10m	45%	
6a356173-r	78	View JSON	0.0	Jun 22, 12:18 PM	3m	100%	
6a356173-z	78	View JSON	0.0	Jun 22, 12:10 PM	4m	100%	
6a356173-1	80	View JSON	19.1	Jun 19, 11:35 AM	11m	100%	
6a356041-3	8	Not classified	-	Jun 19, 11:30 AM	0m 29s	4%	
6a355f69-d	82	Writing-Assisted	24.4	Jun 19, 11:29 AM	8m	100%	
6a356017-0	80	AI-Assisted	19.9	Jun 19, 11:29 AM	21m	100%	
6a356045-8	82	Tab-Switching	18.7	Jun 19, 11:29 AM	21m	100%	
6a356024-e	80	View JSON	22.5	Jun 19, 11:29 AM	11m	100%	

< 1 2 3 4 5 ... 27 >

50 / page

Help

☐	Participant ID	Messages	Monitoring	Engagement	Date	Duration	Progress	Groups
☐	6a34a95a-e	82	AI-Assisted	89.5	Jun 18, 10:31 PM	21m	100%	
☐	6a34ab3d-2	56	Not classified	87.8	Jun 18, 10:41 PM	27m	78%	
☐	6a34aa25-0	82	Bot AI-Assisted	87.2	Jun 18, 10:36 PM	42m	100%	
☐	6a354eae-c	32	Not classified	86.7	Jun 19, 10:20 AM	🕒	36%	
☐	6a354971-7	82	AI-Assisted	82.6	Jun 19, 9:53 AM	19m	100%	
☐	6a3558df-a	84	AI-Assisted	78.0	Jun 19, 10:58 AM	31m	100%	
☐	6a355d1f-7	58	Not classified	77.0	Jun 19, 11:16 AM	16m	77%	
☐	6a347d5b-f	78	Tab-Switching	73.7	Jun 18, 7:21 PM	19m	100%	
☐	6a34801d-d	22	Not classified	69.3	Jun 18, 7:33 PM	🕒	22%	
☐	6a34aa82-9	64	Not classified	68.7	Jun 18, 10:34 PM	🕒	82%	
☐	6a34ad78-c	38	Not classified	65.0	Jun 18, 10:47 PM	53m	49%	

DEMO Engagement scores

engage

< Projects / Iran War V1

1 Design

2 Publish

3 Results

Settings

Sessions

Segments

Insights

Quotas

Reports

840 Sessions

Search IDs

Simulate Participants

Export

Clear Filters

Status: Completed x

☐	Participant ID	Messages	Monitoring	Engagement	Date	Duration	Progress	Groups
☐	6a34a95a-e	82	AI-Assisted	89.5	Jun 18, 10:31 PM	21m	100%	
☐	6a34aa25-0	82	Bot AI-Assisted	87.2	Jun 18, 10:36 PM	42m	100%	
☐	6a354971-7	82	AI-Assisted	82.6	Jun 19, 9:53 AM	19m	100%	
☐	6a3558df-a	84	AI-Assisted	78.0	Jun 19, 10:58 AM	31m	100%	
☐	6a347d5b-f	78	Tab-Switching	73.7	Jun 18, 7:21 PM	19m	100%	
☐	6a348126-e	80	AI-Assisted	65.0	Jun 18, 7:40 PM	23m	100%	
☐	6a354485-6	78	Tab-Switching	58.9	Jun 19, 9:32 AM	17m	100%	
☐	6a3483bf-b	80	Writing-Assisted	56.1	Jun 18, 7:53 PM	49m	100%	
☐	6a349abb-5	80	View JSON	55.2	Jun 18, 9:27 PM	28m	100%	
☐	6a3482e6-4	80	View JSON	52.8	Jun 18, 7:45 PM	14m	100%	

< 1 2 3 4 5 ... 17 >

50 / page

Help

engage

< Projects / Iran War V1

① Design

② Publish

③ Results

Settings

Sessions

Segments

Insights

Quotas

Reports

Expand All

Collapse All

Multi-Choice Question

25. What is your gender?

837 Sessions

+ Generate Insights

Explore Results

Multi-Choice Question

26. What is the highest level of education you have completed?

837 Sessions

+ Generate Insights

Explore Results

Multi-Choice Question

27. Did you recently purchase a home in McMullen, Alabama?

837 Sessions

+ Generate Insights

Explore Results

Multi-Choice Question

28. What is your total annual household income before taxes?

837 Sessions

+ Generate Insights

Explore Results

Multi-Choice Question

29. What is your race/ethnicity?

44 Sessions

+ Generate Insights

Explore Results

<

Help

Four Sources of Unusable Survey Data

01 FULLY AUTOMATED

AI Agents

Autonomous bots that navigate entire surveys without human involvement

02 AI-ASSISTED FRAUD

LLM-Assisted Respondents

Real people who tab out mid-survey to query ChatGPT or other AI tools, copy-paste AI-generated answers into open ends, or use AI to look up the 'correct' choice on closed-ended questions.

03 FRAUDULENT HUMANS

Click Farm Respondents

Organized networks of workers who fraudulently enter surveys for pay, producing responses that are inconsistent, illogical, or impossible.

04 LOW-QUALITY TEXT

Gibberish Open Ends

Nonsensical, off-topic, or gibberish text from non-English speakers who don't understand the questions, and from disengaged or ill-intentioned respondents.

JOIN THE BOT OLYMPICS

Let's Not Just Talk About It

COMPETE

JOIN/PARTNER

HELP SPREAD THE
WORD

cloudresearch.com

| In partnership with 

| \$50,000 Prize Pool

Thank You!

Contact Us

reuben.paris@cloudresearch.com

cloudresearch.com