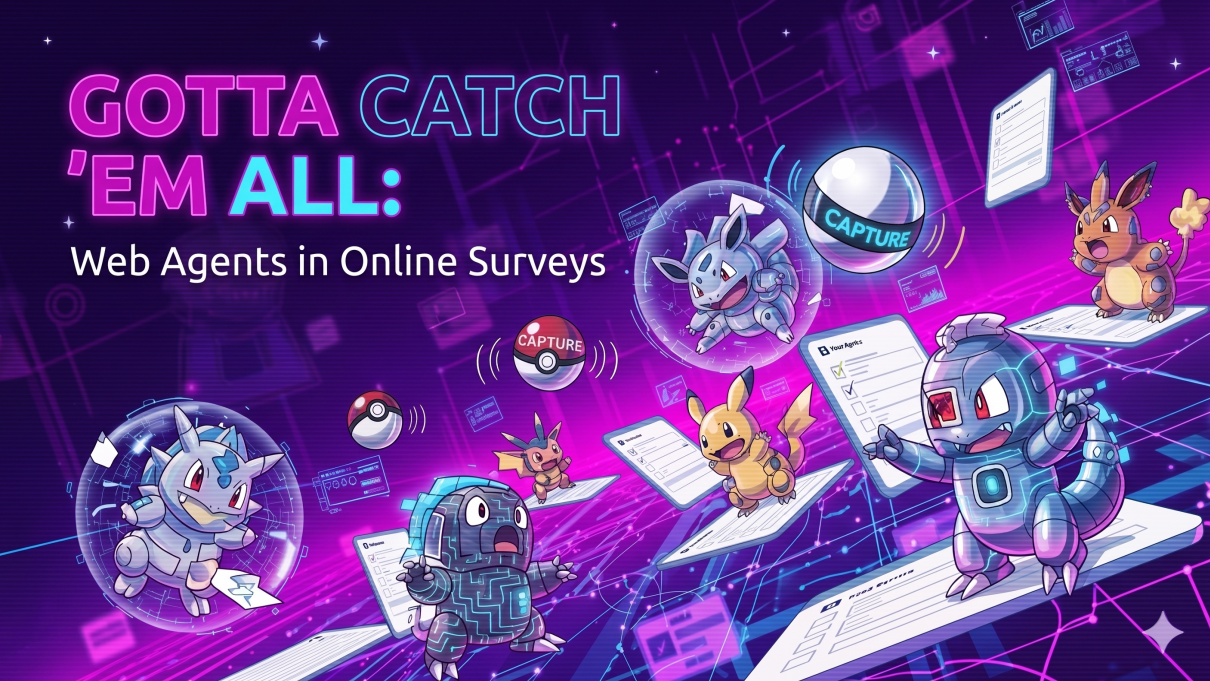


GOTTA CATCH 'EM ALL:

Web Agents in Online Surveys



MISSION MI POSSIBLE:

The Collection of High-Quality Data.



Data Quality Concerns

Inattention

Account Fraud

LLM Assistance

AI Web Agents

Data Quality Checks

- Inattention* → Standard attention-check questions
- Account Fraud* → IP address, geolocation clustering
- LLM Assistance* → Keystrokes, copy/paste, mouse movements, Pangram
- AI Web Agents* → Video check

Data Quality Checks

Inattention → Standard attention-check questions

Account Fraud → IP address, geolocation clustering

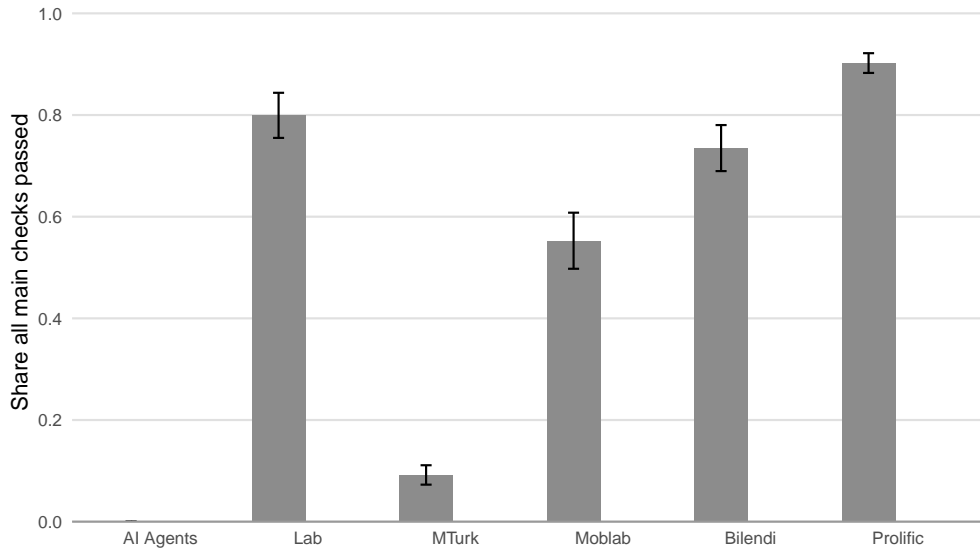
LLM Assistance → Keystrokes, copy/paste, mouse movements, Pangram

AI Web Agents → Video check

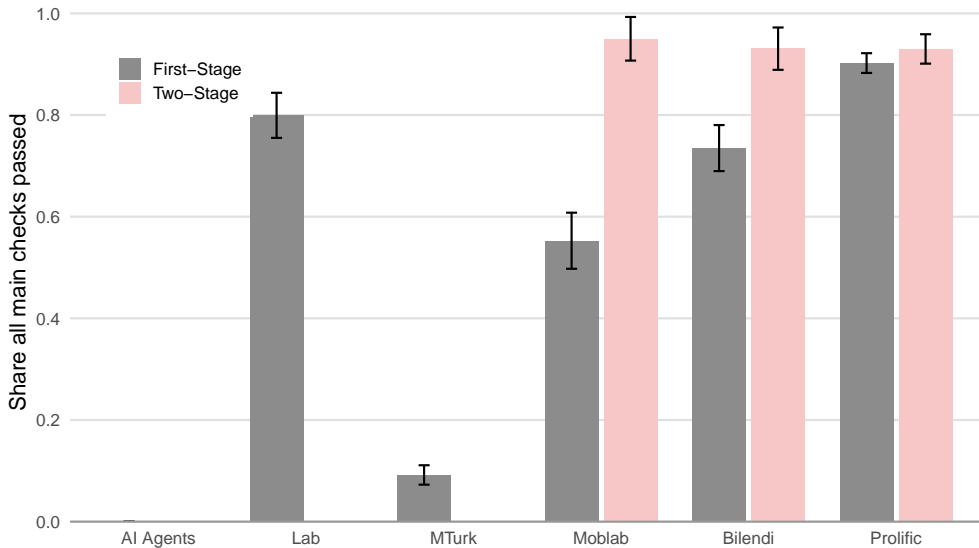


~5 min

Mission Possible: Main Results



Mission Possible: 2-Stage Procedure





Track and compare data quality across online survey platforms.

Researchers submit their study results — the dashboard updates as new studies are added.

[Submit Results](#)[GitHub](#)[Join Discussions](#)

1 Platforms · 3 Studies · 232,344 Participants · ● Live

Data Quality Dashboard

Means by Platform

Trends

Overall

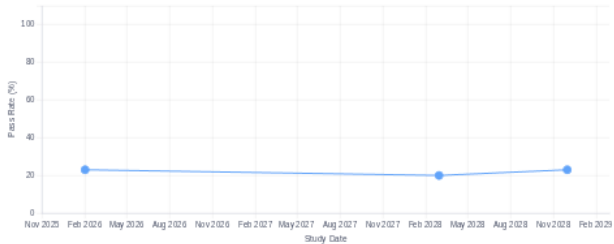
Attention

AI and Bots

Account Fraud

ⓘ

Filters



PLATFORM ● Prolific

RECRUITMENT ● Standard

Web Agents

Core

- Image processing
- Planning
- Text generation
- Decision

Actions

- Screenshots
- Cursor movement
- Click
- Write, paste, etc.

Research Question

What task can a human do that a Web agent cannot?

Research Question

What task can a human do that a Web agent cannot?

- **Condition 1:** easy for humans

Research Question

What task can a human do that a Web agent cannot?

- **Condition 1:** easy for humans
- **Condition 2:** the agent does not suspect it is failing

Mission Possible: Video Check

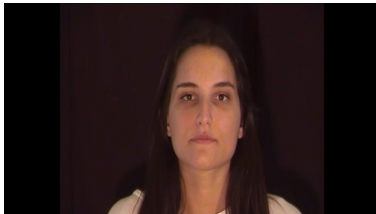
3

Video Check +

3 5 4 5

McGurk Effect

McGurk & MacDonald (1976)



See: “ka”

Hear: “pa”

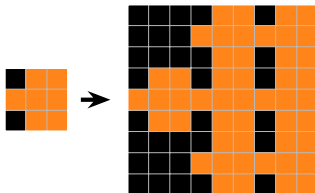
Perceive: “ta”

Not reliable — effect rate varies 30–98% across studies.

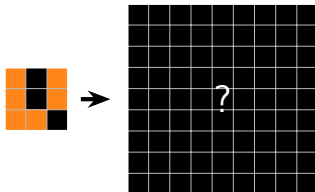
ARC-AGI

Chollet (2019), *On the Measure of Intelligence*

Example



Test



Easy to see, tedious to answer — 81 cells to paint in, one by one.



`can-celebi.github.io/games`