AI-BASED SIMULATION AND PREDICTION

# Population simulation

MATT HAUER, PHD · AARU

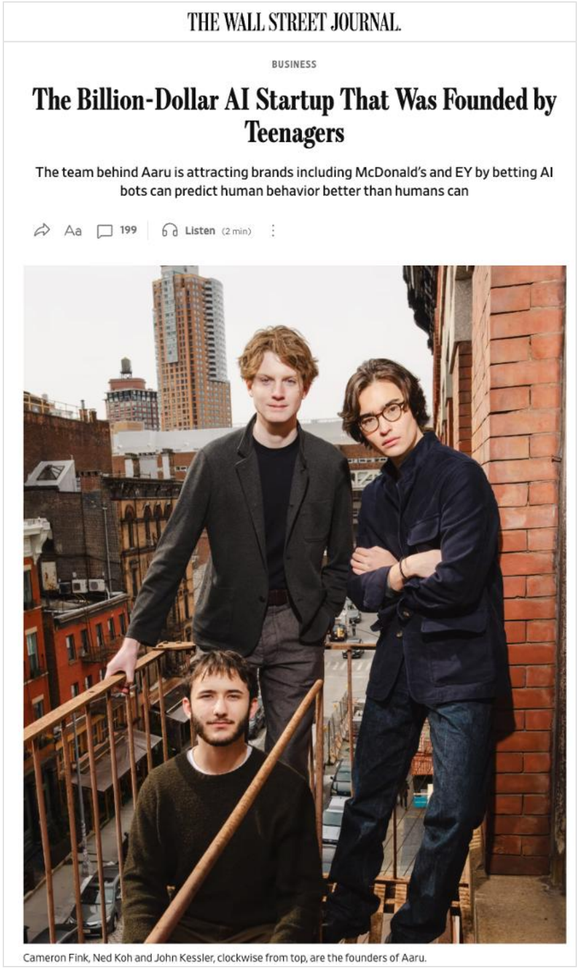
© 2026 AARU

■ WHO IS AARU?

# Aaru simulates populations to predict behavior.

Aaru builds large synthetic populations from ground-truth data, then asks them questions. It replaces surveys and focus groups with simulation.

| | |
|---|---|
| FOUNDED | 2024 · New York |
| WHAT IT DOES | Builds populations of AI agents and asks them, in place of surveys |
| USED FOR | Research, policy, and enterprise decisions |



THE WALL STREET JOURNAL.

BUSINESS

**The Billion-Dollar AI Startup That Was Founded by Teenagers**

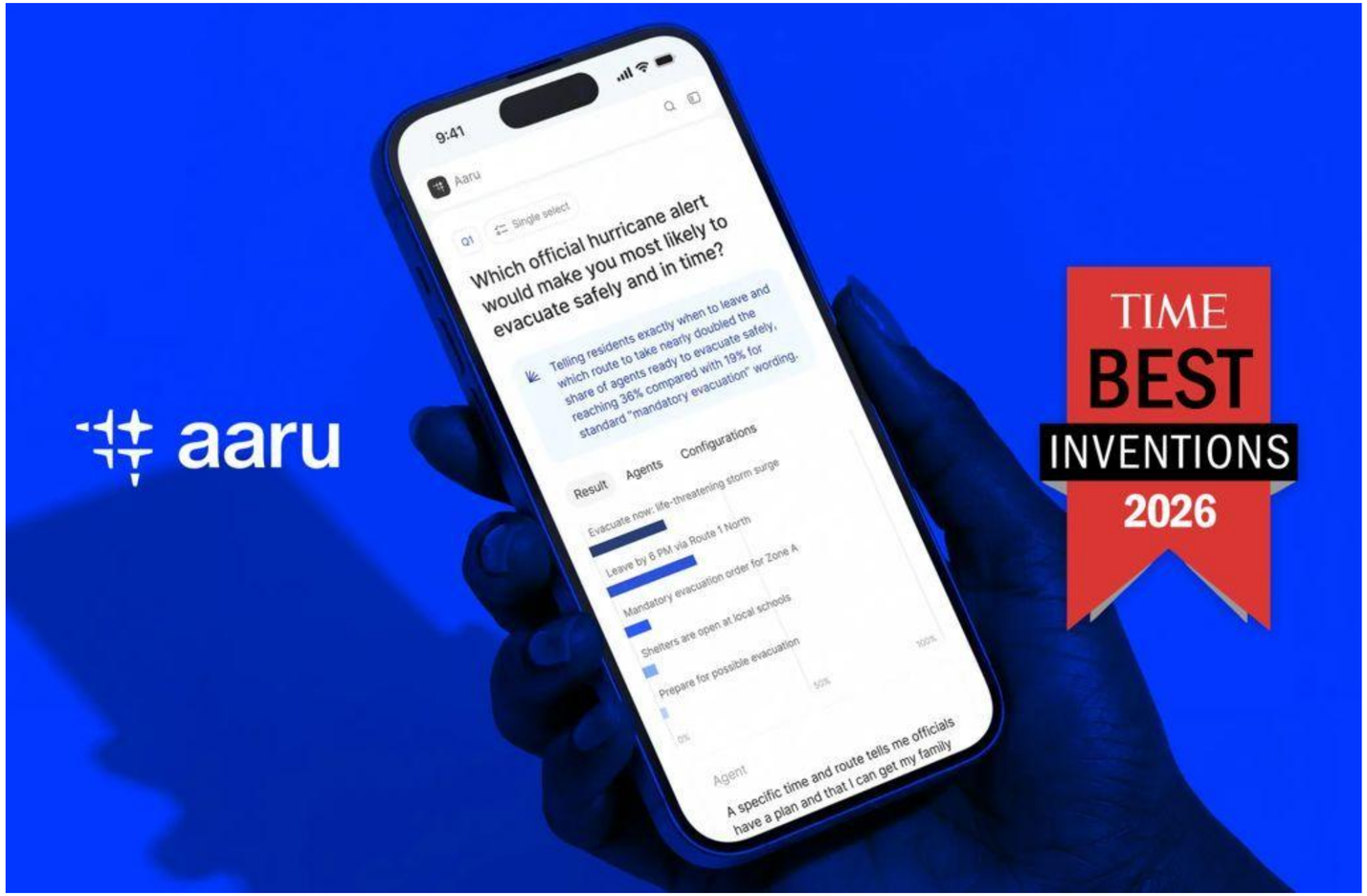
The team behind Aaru is attracting brands including McDonald's and EY by betting AI bots can predict human behavior better than humans can

Aa 199 Listen (2 min)

Cameron Fink, Ned Koh and John Kessler, clockwise from top, are the founders of Aaru.

■ RECOGNITION

# Named one of TIME's Best Inventions of 2026.




## AI human behavior predictor

TIME · AI category ↗

WHY IT WORKS

# People are overlapping identities.

ONE PERSON, ALL AT ONCE

Father | Husband | Demographer | Black-coffee drinker | Bicyclist | BJJ brown belt | AI researcher | University professor | Sociology professor | ...

People are an infinite set of overlapping, often contradictory, identities. These intersections, rather than broad categories, are what actually predict decisions.

We read those identities from many datasets, public and proprietary, and aggregate them up to a population.

It's a small leap from **Matcha-drinking dog walkers** → **Homeowners losing trust in flood insurance**

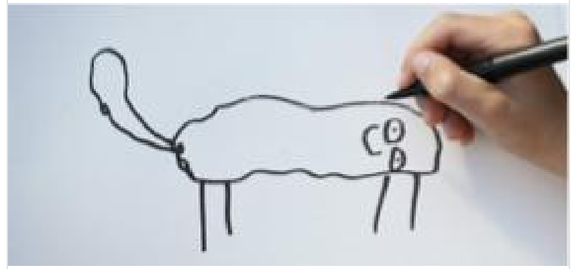
© 2026 AARU

# A model is only as good as its predictions.

"All models are wrong, but some are useful."

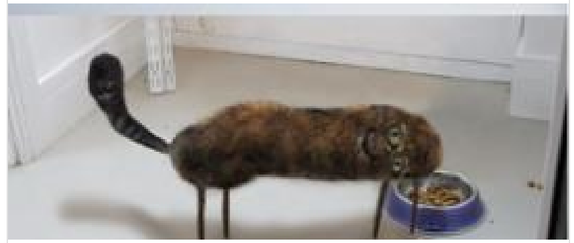
**Better models make better predictions.**

CRUDE MODEL



↓

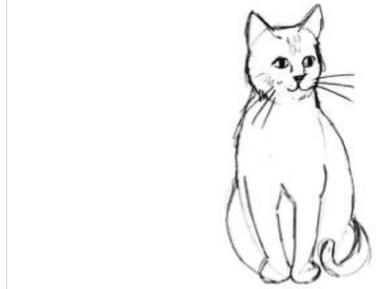
PREDICTS



**Wrong, and far from real.**

instagram.com/thingsihavedrawn

BETTER MODEL



↓

PREDICTS



**Still wrong, but useful.**

© 2026 AARU




"

We let our tech speak for itself... Give us something you know, the survey you ran, don't give us anything else. Just give us the people you talked to and the questions you asked. We will simulate it totally blind and prove more accurate.

**Ned Koh, President · Aaru**

WATCH ON INSTAGRAM ↗

# EY Global Wealth Survey. We replicated it blind.




EY surveyed 3,600 people worth $20M+. Aaru simulated the survey without seeing the results.

**90%**

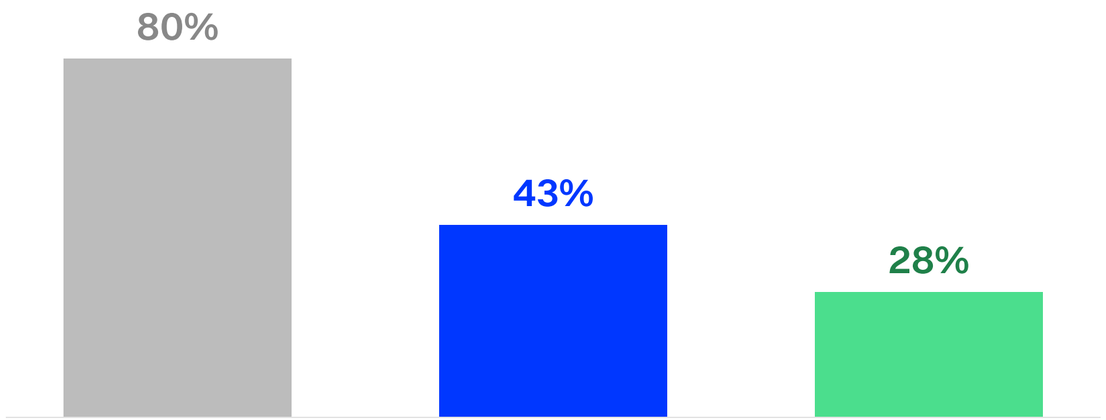
Median correlation across 53 questions

**3 of 3**

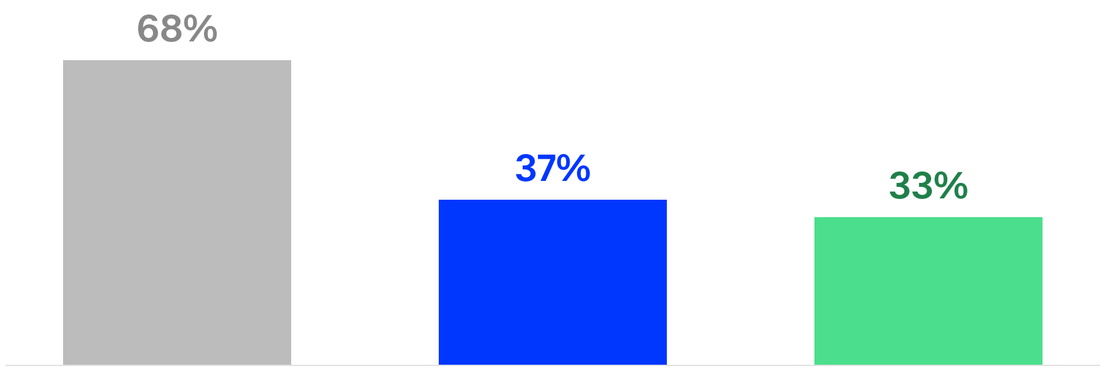
Disagreements where Aaru was closer to real-world data

# Stated versus revealed preferences.

*How likely are heirs to keep their parents' wealth advisor?*




*What is your preference for a single provider?*




EY survey (stated) Aaru simulation Real-world data

© 2026 AARU

# We simulated the endorsement before it happened.

MAY 13

Aaru simulates: if Paxton is endorsed, he wins 65.7% of the two-way vote.

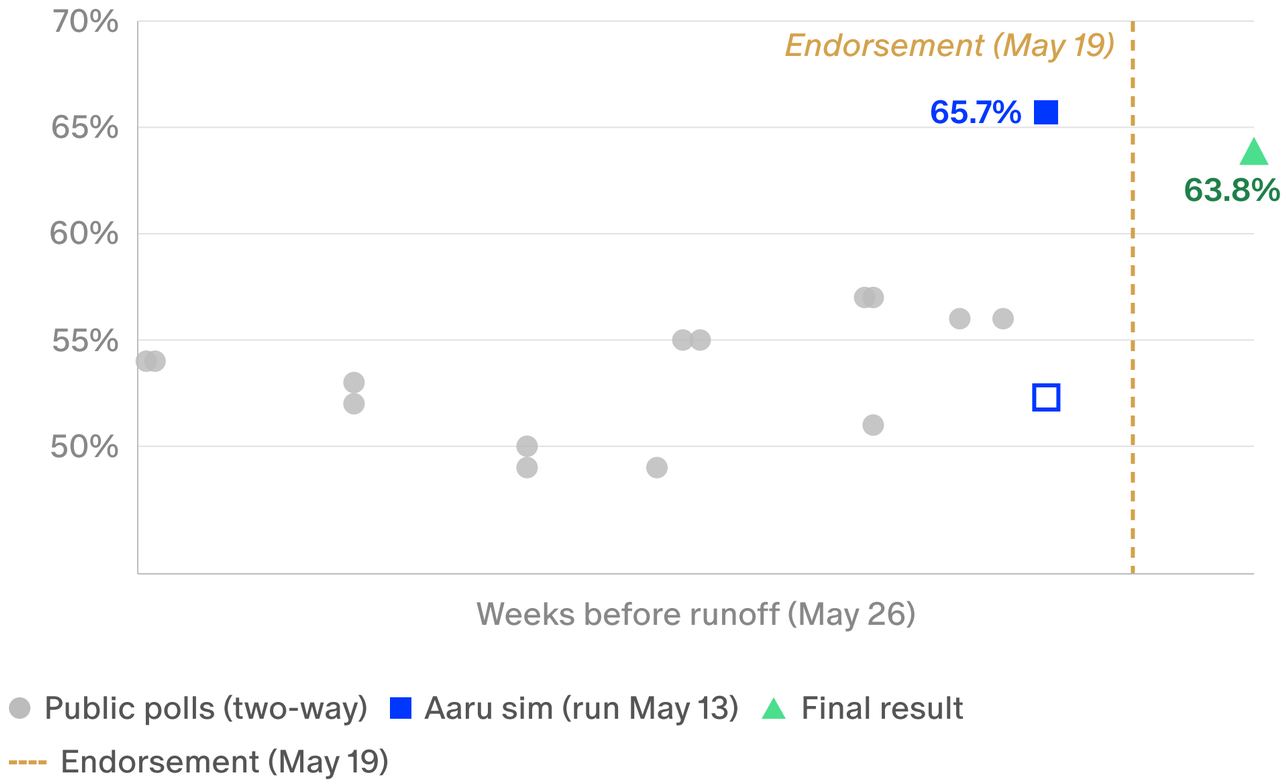
MAY 19

Paxton is endorsed.

MAY 26

Paxton wins with 63.8%.

**Off by 1.9 points, on an event yet to happen.**



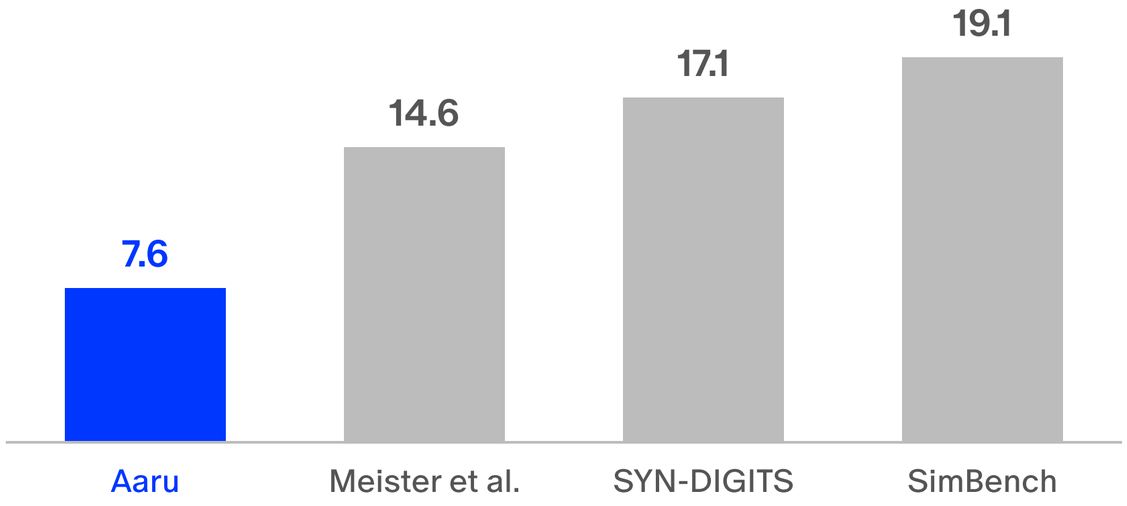

© 2026 AARU

VALIDATION — AT SCALE

# Tested on 2,993 published survey questions.

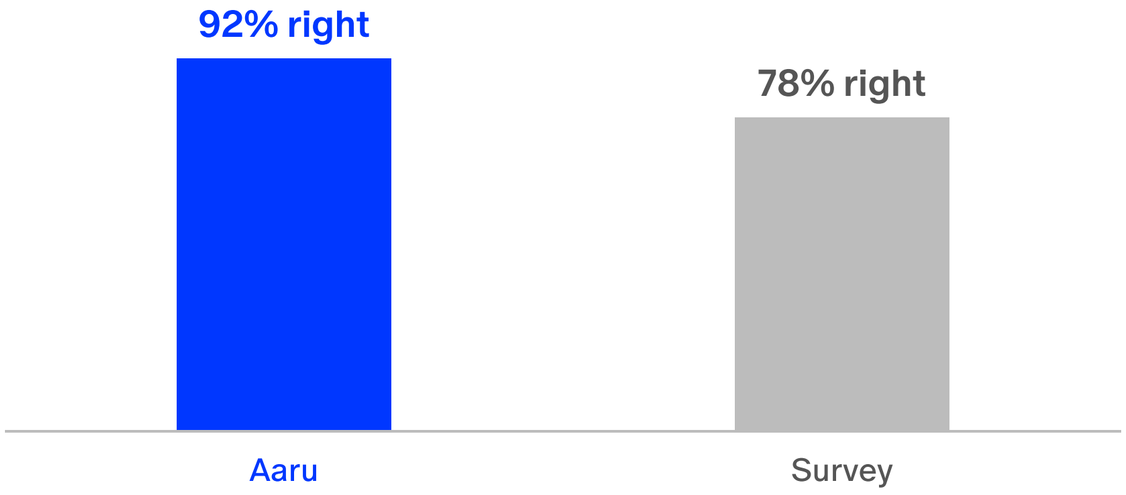
## Half the error of published LLM results

Error on 2,993 published survey questions, in points. Lower is better.




## Which brand do people actually buy more?

Across 340 brand pairs, like Kroger vs. ALDI, checked against credit-card purchases.




Aaru Research, “Population simulation at the replication floor,” Sep 2026 ↗

# Why other approaches fail.

The threat of a

In

Social scienti
datasets inte
requires man
and the amo
examine whe
human data. I
using two soc
rankings, res
substantially
often perform
sample use
configuration
substantially
studies demo
conclusions a
analytic flexib
reduce this th

*[Submitted on 23 Sep 2025 (v1), last revised 19 Apr 2026 (this version, v5)]*

## Digital Twins as Funhouse Mirrors: Five Key Distortions

Tianyi Peng, George Gui, Melanie Brucks, Daniel J. Merlau, Grace Jiarui Fan, Malek Ben Sliman, Eric J. Johnson, Abdullah Althenayyan, Silvia Bellezza, Dante Donati, Hortense Fong, Elizabeth Friedman, Ariana Guevara, Mohamed Hussein, Kinshuk Jerath, Bruce Kogut, Akshit Kumar, Kristen Lane, Hannah Li, Vicki Morwitz, Oded Netzer, Patryk Perkowski, Olivier Toubia

Scientists and practitioners are increasingly moving to deploy digital twins--LLM-based models of real individuals--across social science and policy research. We conduct 19 pre-registered studies spanning 164 diverse outcomes (e.g., attitudes toward hiring algorithms, intentions to share misinformation), comparing human responses to those of their corresponding digital twins, which are trained on each individual's prior responses to over 500 questions. We establish an empirical benchmark for digital twin performance: their predictions are only modestly more accurate than those of a homogeneous base LLM and exhibit weak correlation with human responses (average $r = 0.20$). To inform future development, we identify five systematic distortions in digital twin behavior: (i) insufficient individuation, (ii) stereotyping, (iii) representation bias, (iv) ideological bias, and (v) hyper-rationality. Finally, we release our full dataset and code as a standardized testbed for evaluating and improving digital twin methodologies. Together, our findings caution against premature deployment while laying the groundwork for a transparent, replicable, and iterative science of responsible digital twin development.

es. Human-to-silicon correlations
sible configurations of the

. Sci., 2025 ↗

base LLM, with weak correlation to
systematic stereotyping and

dels and the empirical validity of
t responses that drift from

*rediction. But they often fail individual*

SUMMARY

# Key takeaways.

01 **A model is only as good as its predictions**

Every model is wrong. What matters is whether it predicts what actually happens.

02 **Most current AI approaches fail in practice**

Silicon sampling, digital twins, and generative agents suffer from mode collapse, stereotyping, poor validation, and minimal population-level fidelity.

03 **People are overlapping identities**

Intersections read from many public and proprietary datasets, aggregated to a population, are what predict decisions.

04 **AI-based simulation is possible**

Aaru is validated against the EY Global Wealth Survey (90% correlation), the 2026 TX Senate runoff (off by 1.9 points), and ongoing real-world benchmarks.

# Thank you.