COMPOSABLE AI × RESEARCH METHODOLOGY

# How Sunzu simulates human behavior.

How composable AI, customer evidence and model checks support simulations for product, design and marketing research.

By **Alex Street, Yoann Lemain & Ludovic Lafole** / Updated September 29, 2026 / sunzu.com

Sunzu simulates human behavior with composable AI: 20+ layers, including algorithms and small models, working together to support customer research.

IN BRIEF

01 Customer evidence grounds the audience. AI simulates how members of that audience might respond.

02 Source links, confidence ratings and explicit gaps help teams review the basis for a finding.

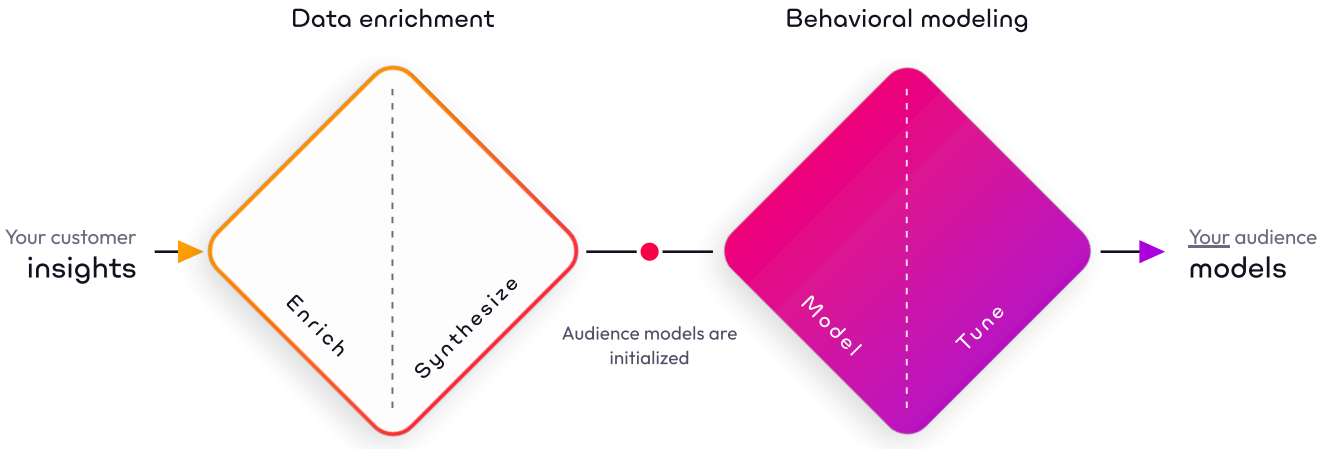
03 Coherence and alignment checks assess the personas. Human reference findings provide a separate basis for benchmarking.

04 The reported results come from Sunzu's own research on proprietary data and apply to specific comparisons.

Our first paper explains where simulations can help a team make better-informed decisions. This paper describes the evidence that grounds those simulations, the audience and persona models, and how to interpret our internal benchmarks.

A useful simulation must do more than produce a plausible conversation. Research identifies risks including stereotypes, caricatures and weak validation.[4,9] Asking an AI to speak for a demographic group does not establish that it reflects the group's opinions.[8]

Sunzu combines customer evidence, audience modeling and response checks within its composable AI system. The layer count describes the system's composition. Its usefulness for a research task is assessed through the findings it produces and their relationship to human evidence.




Two stages: organize the evidence, then build and validate personas.

---

STAGE ONE · ENRICH → SYNTHESIZE

# 01 Modeling the audience

## Customer evidence defines the audience's context. The simulation explores how people in that context might respond.

We identify customer groups supported by the supplied research and documents. Interviews, surveys, support themes and other customer evidence help establish their circumstances, motivations and constraints.

Research by Park and colleagues found that agents grounded in interviews or surveys predicted responses more accurately than agents based on demographics alone.[10] This supports the approach of grounding simulations in customer evidence. It does not independently validate Sunzu's implementation.

Findings extracted for the audience model retain source links so reviewers can check them and update the model when evidence changes. Relevant public reviews, forums, news and published studies provide additional context. Findings are checked against customer evidence and assigned confidence ratings.

Sunzu's behavioral corpus contains **1M+ profiles, 34M+ attributes and 57M+ data points**, according to our internal research. It supplements the context available from customer evidence. These figures describe the underlying evidence base, not recruited participants or the sample size of a benchmark. The underlying proprietary data is not publicly disclosed.

Broader evidence can suggest what to investigate when customer data is sparse. Its relevance depends on the audience and question. Source links and confidence ratings support review; they do not, by themselves, establish that a simulated answer is correct.

The audience model has four components. These describe its contents, rather than the 20+ layers of the AI system:

| | | |
|---|---|---|
| 01 | Demographic profile<br>Who they are | The group's documented background and circumstances, using characteristics supported by the source material. |
| 02 | Psychographic insights<br>How they think | Motivations, attitudes and preferences, with evidence and confidence ratings for each finding. |
| 03 | Scenario specifics<br>What is at stake | The needs, pain points and constraints relevant to the question being tested. |
| 04 | Abstentions<br>What we do not know | Gaps in the evidence, marked explicitly so missing information is not presented as a finding. |

The four parts of the audience model.

STAGE TWO · MODEL → TUNE

# 02 Generating the persona

## A persona represents an individual within the audience model. Its responses are AI-generated simulations.

Each persona starts with a defined role, situation, attitudes and likely decision factors. Contextual detail supports responses across the situations being tested.

The persona retains links to the audience model and its evidence. This lets reviewers trace its basis while distinguishing a source finding from a modeled response.

Persona responses should be read as simulated reactions. A realistic portrait or fluent answer does not add human observations to the study.

Illustrative portrait of a modeled persona.

## Each persona must pass coherence and alignment checks before it is used in a study.

Gate 01 / Coherence

### Does it hold together?

We check that the persona's attitudes, background and behavior are internally consistent.

Gate 02 / Alignment

### Does it fit the audience?

We test responses across several turns of conversation against the audience model and its evidence. Alignment includes variation between responses: matching a group average can still hide differences in the distribution.[12]

Coherence and alignment assess consistency with the audience evidence. Predictive validity asks how closely findings match what people report or do.

Back-testing informs model refinement. For your own evaluation, keep the human findings used to judge results separate from the material used to build or adjust the audience.

BENCHMARKS AGAINST HUMAN FINDINGS

The following results come from Sunzu's internal research on proprietary data. They measure reference findings recovered in specific back-tests. Concept, UX and ad testing use research with 500+ humans as the reference; depth interviews use human insights retested within two weeks. Each human reference is normalized to 100%. Results vary by audience and task.

01 DISCOVER

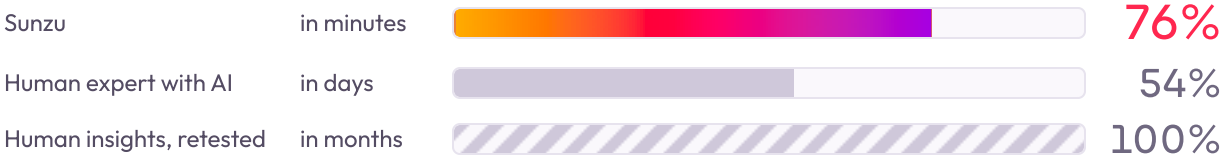
## Depth interviews

Understand problems, priorities and the language your audience uses.

REPORTED SAMPLE EQUIVALENCE

10+ humans

THEMES COVERED




02 FRAME

## Concept testing

Compare ideas and identify the strongest direction for your audience.

REPORTED SAMPLE EQUIVALENCE

30+ humans

INSIGHTS FOUND

03 SHAPE

## UX testing

Find confusing steps and usability problems before release.

REPORTED SAMPLE EQUIVALENCE

20+ humans

DEFECTS DETECTED

| | | |
|---|---|---|
| Sunzu | in minutes | 75% |
| Human expert with AI | in days | 34% |
| Research, 500+ humans | in months | 100% |

04 LAUNCH

## Ad testing

Compare messages and creative before spending on a campaign.

REPORTED SAMPLE EQUIVALENCE

20+ humans

INSIGHTS FOUND

| | | |
|---|---|---|
| Sunzu | in minutes | 69% |
| Human expert with AI | in days | 55% |
| Research, 500+ humans | in months | 100% |

Read the scope below, or see the benchmark comparison table.

EVIDENCE AND INTERPRETATION

# 03 What the benchmarks establish

## Each result measures a specific comparison with human findings. Use the benchmark that matches your research task.

**Coverage of reference findings.** The percentages describe themes covered, insights found or defects detected. The human reference is set to 100% to express relative coverage. That baseline does not imply that the reference research captured every possible finding.

**Coverage and correctness.** Recovering a reference finding and avoiding an unsupported finding are different measures. Coverage alone does not state how many additional findings were incorrect, how serious a missed defect was or whether a proposed decision improved a business outcome.

**Sample equivalence.** The figures shown are Sunzu's reported estimates of result equivalence within these comparisons. They are not counts of recruited participants, a statistical margin of error or a guarantee that another study will produce the same result.

**Research time.** Minutes, days and months describe the compared workflows. Assess total elapsed time for your own process, including preparation, setup and review. The two-week interval belongs to the depth-interview retest reference, rather than the duration of a Sunzu study.

**Proprietary research.** The underlying datasets and detailed evaluation protocols are proprietary and are not published here. The results are reported by Sunzu from its own research. The external publications cited in this paper provide context for the approach; they are not independent audits of these results.

**Application to your audience.** A benchmark is a starting point for evaluation. Compare findings with your own human evidence, review misses and unsupported additions, and assess whether the simulation helps your team reach a useful decision.

LIMITS

# 04 When to work with people

## Use modeled audiences to sharpen early decisions and focus the research you take to customers.

Work directly with people when a decision depends on lived experience, accessibility, cultural context or behavior that the available evidence does not capture. Unknowns need to remain visible when the evidence cannot support a conclusion.

Some studies show promising results on specific tasks. Maier and colleagues reported purchase-intent predictions close to human test-retest reliability for personal-care product surveys.[11] That is evidence for their method and dataset, not a guarantee for every audience or research question.

Research on designers' use of persona-based chatbots also examines how these tools fit into design practice.[6] We use audience models alongside human research, with their sources and limits visible to the team.

Start with a defined audience, a decision and human evidence suitable for evaluating the result. Review the simulation's sources and limits, then measure the time and effort required to reach a useful conclusion. Our practical guide explains how to connect that evaluation to business value.

# Get user insights in minutes.
# Shape products people love.

GET STARTED

## Test a decision with your audience.

hello@sunzu.com

SOURCES & FURTHER READING


01 Bain & Company. How Synthetic Customers Bring Companies Closer to the Real Ones.

02 Making Science. Enhancing UX/UI Research with Synthetic Users.

03 Interaction Design Foundation. Are AI-Generated Synthetic Users Replacing Personas?

04 arXiv (2025). Creating and Evaluating Personas Using Generative AI.

05 Snowflake. What Is Synthetic Data? Examples and Use Cases.

06 AI EDAM, Cambridge University Press. Synthetic users: insights from designers' interactions with persona-based chatbots.

07 MJV Innovation. How are AI models used to create synthetic users for research?

08 arXiv (2023), Santurkar et al. Whose Opinions Do Language Models Reflect?

09 arXiv (2023), Cheng, Piccardi & Yang. CoMPosT: Characterizing and Evaluating Caricature in LLM Simulations.

10 arXiv (2024), Park et al. LLM Agents Grounded in Self-Reports Enable General-Purpose Simulation of Individuals.

11 arXiv (2025), Maier et al. LLMs Reproduce Human Purchase Intent via Semantic Similarity Elicitation of Likert Ratings.

12 arXiv (2026), Moon et al. Beyond Averages: Evaluating LLMs on Human Survey Replication at the Distributional Level.


sunzu


© 2026 Sunzu, Inc. All rights reserved. Services outside the UK, Switzerland and EEA are provided by Sunzu, Inc., a Delaware corporation (file #10756833). UK, Swiss and EEA services are provided by its wholly owned subsidiary, Alpha Base OS Ltd, registered in England and Wales (company #16212608; registered office: 71–75 Shelton Street, Covent Garden, London WC2H 9JQ, United Kingdom; VAT GB488054656; ICO C1103931).

Cookie preferences