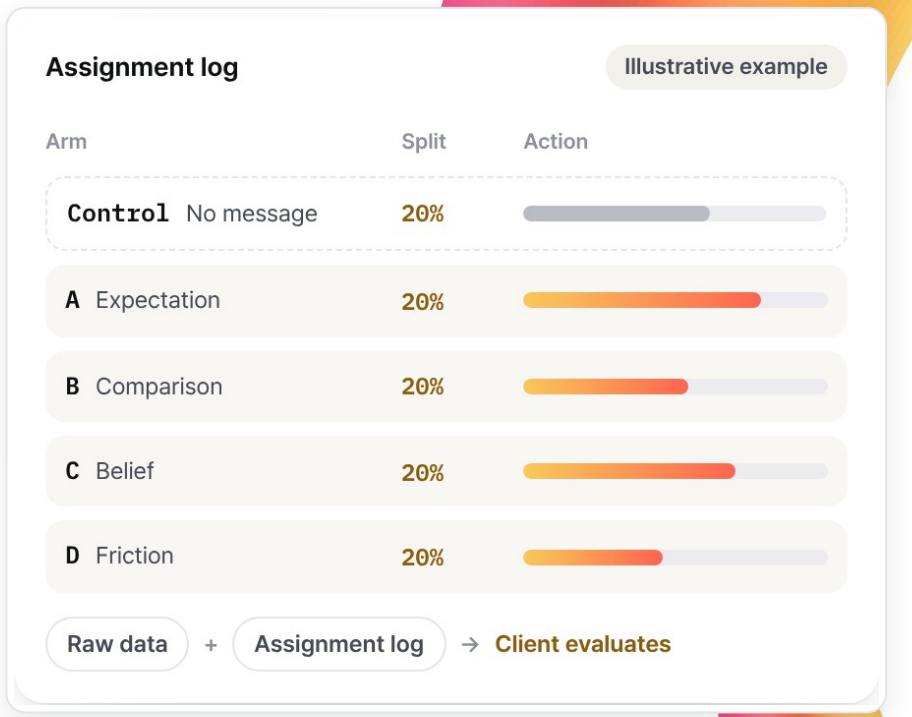
DGC

GroundProof Overview

GroundProof Data Solution

Consumer experiment data for calibration and validation

We run client-designed experiments in the real market and return the assignment log with raw data.



GroundProof The experiment channel at a glance

Who uses it, what it does, what you receive, and the rules it keeps.

For

AI companies predicting consumer response, and research labs

What it does

Runs and measures client-designed experiments in the real market

You receive

Raw data · assignment log · per-view metrics

Rules

The client evaluates; results stay private

Three offerings

Custom Experiment

Client-set factor combinations, produced as independent creatives and posted at random

Holdout Data

Raw data on new messages the model has never seen, with a control group

Purchase-Linked Experiment

Pairs behavior records with order outcomes

Digicon **runs the experiment**; the client uses the data to **calibrate and validate its model**.

Rise of behavior-simulating AI

Capital flows to models that mimic human responses: Simile alone raised USD 300M+ in 2026.

Simile

USD 2B

Valuation 2026.7

Series A and B, USD 300M+ in 2026

Ground Truth Gallup interviews

Aaru

USD 1B

Headline value · 2025.12

Series A · About USD 88M total

Ground Truth EY survey

Electric Twin

USD 10M

Latest · 2026.2

Total USD 14M · value private

Ground Truth News UK panel

Their Ground Truth	Stated responses
Behavior from real ads	No public deals

Capital flows to behavior AI, but **real-market Ground Truth** stays rare.

Sources: Simile (2026.7.30), TechCrunch (2025.12.5), PitchBook (via Fortune 2026.6.17), Electric Twin (2026.2.11), Gallup (2026.5.11), EY (2026.9.30), Research Live (2026.5.1), Neudata (2025.6.10). Ground Truth: measured answer key.

Studies demanding behavior checks

LLMs track survey experiments well but miss more in field experiments on behavior.

Hewitt et al., Stanford, 2024 preprint

LLM accuracy on message effects per study · correlation



Like predicting per-question exam scores. Real exams are much harder than surveys. Closer to 1 is more accurate.

Note · 0.85 across 70 survey experiments

Park · 2024.11

Agents built on 2-hour interviews of 1,052 people reached 85% of self-consistency after 2 weeks.

Kutzner · 2026.9

Set observed behavior over statements as the validation standard.

NeurIPS 2025 · ICML 2026

“Social world model” appears in workshop and main-track papers

AI finds **behavior** harder than **answers**; **records** fill the gap.

Sources: Hewitt et al. (2024 preprint, Nature), Park et al. (2024.11), Kutzner et al. (2026.9), Zhou et al. (NeurIPS 2025 workshop), Yu et al. (ICML 2026). Megastudy: meta-analytic mean correlation, uncorrected.

GroundProof Prediction models without real behavior

AI that predicts surveys well loses accuracy on real behavior.

How well AI predictions match results

Closer to 1 is more accurate

Survey experiments like a mock exam



Field experiments · real behavior like the real exam



Like predicting a real exam from a mock exam.

Why random assignment matters

Share of traits that differ sharply between the two groups

Meta A/B tests 181,890 tests · audience differs by ad



Meta Lift tests 3,204 tests · random control



A fair split would sit near zero. Give each judge different diners, and you cannot tell the dish from the diner.

Experiment data **randomly assigned in a real market** closes that gap.

Sources: Hewitt et al. (2024 preprint, megastudy correlations); Burtch et al. (arXiv 2508.21251, 2025.8, not peer reviewed). Right-hand bars end at 25%.

GroundProof The client designs, Digicon runs

The client runs its own evaluation, and Digicon returns raw data with the assignment log.

Four steps

01 Design Client

Sets stimuli and factors; predictions locked inside first

02 Produce and post Digicon

Independent creatives, posted by random assignment

03 Measure Digicon

Assignment log and actions on a test landing page

04 Hand over Client

Raw data with the assignment log; the client evaluates

Three offerings

Custom Experiment

Client-set factor combinations, produced as independent creatives and posted at random. Includes cross-creator runs of one script.

Holdout Data

Raw data on new messages the model has never seen, measured against a control group.

Purchase-Linked Experiment

Pairs behavior records with order outcomes, on projects with consent to link orders.

Results stay **with the client**, and Digicon **does not publish** them.

GroundProof Three offerings of the experiment channel

The client designs, Digicon runs it in a real market and returns the raw data.

Custom Experiment

Client-set factor combinations produced as independent creatives; accounts and posting order assigned at random.

Client does
Set factors · lock predictions

Digicon does
Produce · post at random · log

Client gets
Raw data · assignment log · per-view metrics

Includes cross-creator runs of one script

Holdout Data

New messages the model has never seen, measured with a control group on a test landing page.

Client does
Lock predictions · evaluate

Digicon does
Make messages · assign · measure

Client gets
Action data by message · probabilities

Every arm enters through the same ad

Purchase-Linked Experiment

Pairs behavior records with order outcomes.

Client does
Order-link consent · sales totals

Digicon does
Purchase codes · pair action and order

Client gets
Action and order pairs by arm

Runs where order-link consent is in place

The client **runs its own evaluation**, and results stay **with the client**.

Resale limits and use scope for data built from a client design are set in the contract data clause.

GroundTruth Record One entry of experiment data

Message, entry and assignment, attributes, response and purchase signal in one entry.

Example, not real

Field	Example
Market	US · Cleansing balm
Entry	One static ad click · source marked (paid or organic)
Assign	Message C · chance 25% · control 20%
Copy	"Takes off sunscreen in one step." **Expectation**
Attributes	Age 25 to 34 · foam cleanser · de-ID
Response	Dwell 48s · to store
Verbatim	"Worried it will leave my skin greasy." **Friction**
Purchase	Own-site sales by message (brand consent, no link)
Verdict/basis	Adopt · control design

Included

- Opt-in test-page responses
- Random assignment, probability
- Traffic source (paid or organic)
- Sales total, brand consent
- Five driver tags
- Scope : research, calibration, validation

Excluded

- PII
- Buyer-level order data
- Resale of raw platform data

Verdict records come only from **test landing pages**; organic response is used **in aggregate only**.

Example values are fictional. Format and data card are fixed before the first cell runs. Cell: one category in one market.

Sources vs. in-market records

Existing sources miss or limit real response, random assignment or licensing.

Question

Which source has real response, message-level comparison, licensing?

Source	In-market	Random assignment	Licensing
Surveys	× Stated	○	○
Platform A/B	○	△ Ad audiences vary	× Barred by terms
Purchase	○ Actual	△ Campaign-level	△ Owner-held
AI-made	× Modeled	N/A	○
Digicon (design)	○ Ad-click visitors	○	○ Consent/terms

Surveys ask **"Would you buy?"**. Test landing pages record **what real ad visitors clicked**.

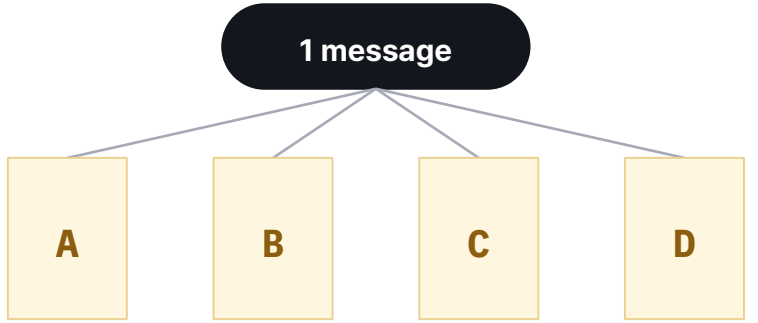
○ Met · △ Partly · × Not met. Digicon row is design basis; terms pending legal review. Sources: Meta Platform Terms (rev. 2026.2.3), Burtch et al. (2025.8), Circana NCSolutions, Kantar (2024.2.15)

Video sharpens response

Each video impression adds time metrics stills lack.

Signal	Stills	Video
Like/share/save	Record	Record
Watch time	None	Log · IG Reels
2s · 6s video views	None	Log · TikTok Ads
25/50/75/100% video views	None	Log · TikTok Ads
First 3s skip %	None	Log · IG Reels
Playtime with replays	None	Log · TikTok Ads
Engaged views	None	Log · YouTube

One message, many creatives



Views of one video are not independent, so read message strength across creatives to rule out one video's luck. One person asked 100 times is not 100 people.

Test **one message as independent creatives**, with a frequency cap.

Sources: Instagram Platform API, Reels Insights (2025.8); TikTok Ads Manager Help (2025.11); YouTube Help; Judd, Westfall, Kenny (2017); Campbell, Keller (2003). Repeated exposure weakens ad effect. YouTube counts first frame from 2026.8.24.

Five Purchase Drivers Stop Points

Stop points differ by market, so we start diagnosis where the most valuable persona stops.

01	02	03	04	05
Trigger	**Belief**	**Comparison**	**Expectation**	**Friction**
Category enters life	**Beliefs on cause and fix**	**Alternatives and criteria**	**Result and how judged**	**What blocks purchase**
Theory	Theory	Theory	Theory	Theory
JTBD: Push · Pull Anxiety · Habit	**Means-end chain Feature → Value**	**POP · POD Ref. price**	**Expectancy gap Result/Emotion**	**Behavioral econ friction Hassle · Pay pain · Doubt**
“I started daily sunscreen”	“Remove oil with oil”	“Why not foam cleanser twice?”	“Sunscreen off in one step”	“Shipping takes 2 weeks”

Rule Findings are logged with insight and verbatim VOC, tracked to a second repurchase as reviews trigger others.

A **stop point** per market, a **message** per stop point.

Illustrative US cleansing balm examples. Theory: Christensen et al. 2016, Gutman 1982, Keller et al. 2002, Oliver 1980, Prelec and Loewenstein 1998.

Data rules What we keep, and the limits

The safeguards behind data quality, and what the data does not represent.

What we keep

Random assignment

Every arm arrives through the same ad; assignment probability is logged on a test landing page.

Criteria fixed in advance

The action to measure and the minimum sample are set in writing; messages short of it are put on hold.

Double-checked tags

Two people tag 10% of the sample separately, and agreement is reported by type.

Private results

The client holds the raw data, and Digicon does not publish results.

Limits

Representativeness

Records reflect visitors who came through an ad, not a probability sample. Inference scope, source and exclusion rate go into the data card.

Platform data

Raw platform data is never resold; organic response is used in aggregate only.

Consent and contracts

Consent wording and data clauses are fixed after legal review per market.

Trust in the data comes from **procedures fixed before launch**.

Tag check: 10% double-coded, agreement reported by type. Basis: model-only tagging without human review led about 31% of hypothesis tests to wrong conclusions (Baumann et al., arXiv 2509.08825, 2025, top models).

Operations High-volume production, three hubs

We make the independent creatives experiments need, at volume, and run them in local languages.

2,000 a month

AI short-form production

Three hubs

Korea sales and R&D · Singapore HQ · Vietnam production

AI executes, people judge

Agents do repeat work; people set criteria and review

Validation plan

2026 Q4 — **First records in the first cell**

Record format and data card fixed

2027 Q1 — **Tag reliability, second cell**

Driver tagging agreement of 0.80 or higher

2027 H1 — **Rules published**

New markets and categories

Running one script with **many creators and accounts** takes a **high-volume production system**.

Production volume is our own count (as of Oct 2026). Dates are the validation plan as of Oct 2026. Cell: one category in one market.

Getting started A first custom experiment

Start by picking the stimuli to test and the test market.

First project

A first custom experiment of your design

Agreed before kickoff

- Stimuli and factors to test
- Test market and period
- Action to measure and minimum sample
- Scope of raw data use

What follows

Holdout Data · Purchase-Linked Experiment

Agreed together after the first results.

Contact

Jay Jin, CEO

jay@digicon.pro · ams.digicon.pro

Digicon Holdings Pte. Ltd. · Singapore

Results stay **with the client**, and Digicon **does not publish** them.



Digital Growth Intelligence Continuum

ams.digicon.pro