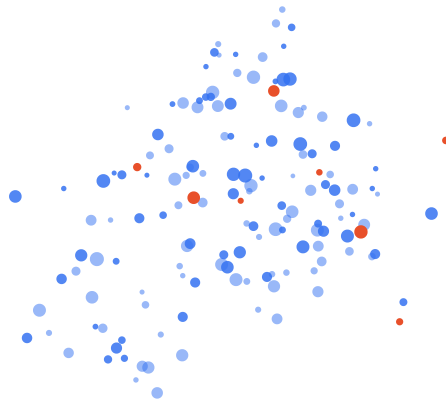




EXPERT INTELLIGENCE LAYER OF AI

Deep Choice

deepchoice

Deep-research data built for
high-stakes, real-life decisions.

When there is no correct answer, does AI answer well?

On a math problem or a code bug, today's AI is already smart enough. But everyday questions are rarely standardized: how do I make the most of a 2-hour commute? With sensitive skin, what's an affordable toner-and-lotion routine under 300 RMB?

These questions have no single right answer. Users don't want **"technically-correct fluff"** — **they want advice grounded in real experience, with judgment, ready to act on.** This dataset tests whether a model can leave its "safe zone" and solve problems like an experienced person would.

To make subjective "usefulness" measurable, experts in each vertical wrote full-mark answers for real scenarios, then split their experience into dozens of scoring rules — from "did it surface the implicit need" to "is the advice actionable" — defining, point by point, what a good answer looks like.

**Users don't want a generic template —
they want expert-grade decision support.**

COVERAGE

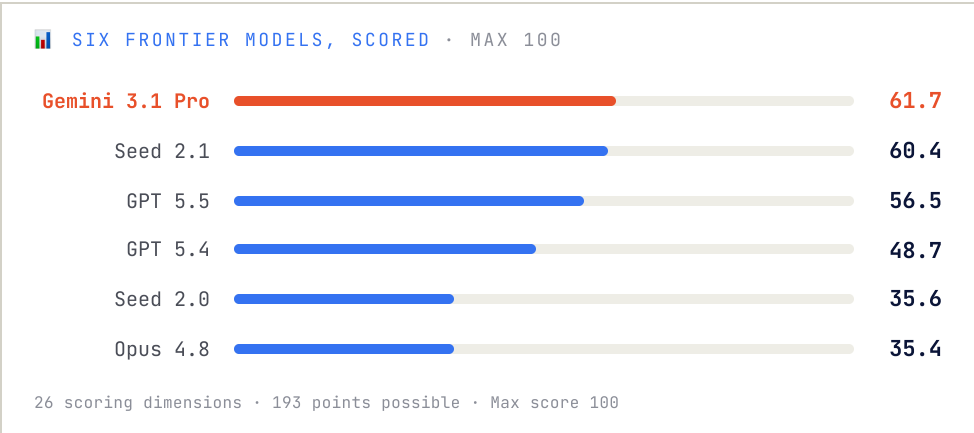
Nine verticals, drilled down to real sub-topics.

The bank spans verticals from women's health to careers, pets, and skincare, drilling into specifics — **from treatment choices for young breast-cancer patients to odor control in multi-cat homes**. Full two-level map below:

●	Career & Money	CAREER & MONEY		424
Job search & career direction 157 • Side gigs • pivots • startups 127 • Personal brand / creator growth 59 • Financial planning & choices 56 • Legal career development 25				
●	Health & Medical	HEALTH & MEDICAL		382
Chronic disease & checkup metrics 137 • Sedentary strain & posture care 65 • Nodule & serious-illness decisions 63 • Oral health & relief 57 • Healthy aging 36 • Seasonal health protection 24				
●	Interests & Picks	INTERESTS & PICKS		341
Off-the-beaten-path travel 103 • Budget phone picks 58 • Car budget & choice 49 • Practical, thoughtful gifts 42 • Arts & culture picks 39 • Consumer-rights guides 35 • Game picks & dev 15				
●	Home & Living	HOME & LIVING		225
Indoor air quality 66 • Renting • renovation • storage 60 • Food recommendations 46 • Cat care & health 32 • Drinks 101 & choices 21				
●	Wellness & Fitness	WELLNESS & FITNESS		145
Fat-loss snacking guides 54 • Better sleep, scientifically 48 • Workout weight-loss plans 43				
●	Learning	LEARNING		105
Effective study methods 76 • History learning paths 29				
●	Parenting & Kids	PARENTING & KIDS		79
Parenting & child development 79				
●	Emotion & Mind	EMOTION & MIND		76
Relationship communication & choices 42 • Mood & focus regulation 34				
●	Beauty & Skincare	BEAUTY & SKINCARE		37
Sensitive-skin skincare guide 37				

Making a daily commute productive

I commute 2 hours a day and want to use that time to improve myself — what learning methods or content would you recommend?



✓ SCORING RULES • 5 OF 26 SHOWN

- Frames it as "cognitive activation + time-into-gains": match the method to the real commute environment.
- Modes by scenario: "ears only" when crowded or driving; "immersive reading" when seated and steady.
- Hard skills (coding / data analysis): learn theory on the commute, save hands-on work for uninterrupted blocks.
- Holds the safety line: never watch a screen while driving.
- Must not recommend nonexistent or fabricated apps / tools / platforms (factual penalty).

Dimensions: 26 • ● Pass ● Critical miss

★ EXPERT REFERENCE • GOLDEN • ANSWER EXCERPT

GOLDEN RESPONSE

BY THE EXPERT • EXCERPT

A 2-hour commute is overlooked wealth — ~500 hours a year. What matters isn't *what* you learn, but **matching method to your commute.**

"Ears only" mode Crowded / driving / cycling Universal, eye-friendly — podcasts, audiobooks, safe voice memos.

"Immersive reading" mode Seated • steady Read-later tools plus spaced-repetition flashcards turn fragments into structured input.

Also covers: three tracks • five strategies • a 1-hour template • startup tips.

EVALUATION SAMPLE · WHEN NO MODEL REACHES THE CEILING

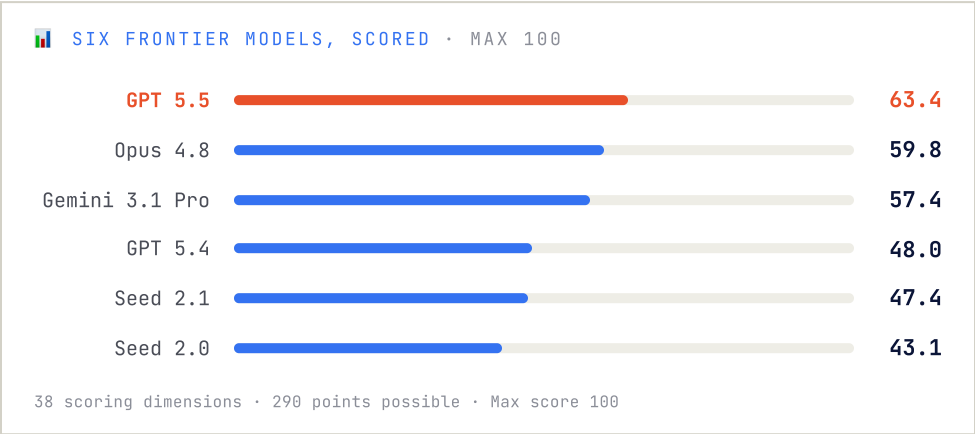
Sample 08

Beauty & Skincare

Sensitive skin

Sensitive skin: a gentle set under 300 RMB

Woman aged 20–30 with sensitive skin: how do I pick a gentle, hydrating, good-value skincare set on a budget under 300 RMB?



✓ SCORING RULES · 5 OF 38 SHOWN

- Centers on barrier repair, minimalist routine (cleanse / hydrate / repair only).
- Must avoid alcohol and fragrance (prone to redness and stinging).
- Prioritizes repairing, soothing ingredients like ceramides and centella.
- Must not recommend soap-based cleansers as a daily go-to for sensitive skin (penalty).
- Must not claim "within budget" when items sum over 300 RMB (hard-constraint violation, penalty).

Dimensions: 38 · ● Pass ● Critical miss

★ EXPERT REFERENCE · GOLDEN · ANSWER EXCERPT

GOLDEN RESPONSE

BY THE EXPERT · EXCERPT

The approach must center on **barrier repair**. Fragile skin needs a strict **minimalist routine** — cleanse, hydrate, repair — budget firmly **under 300 RMB**.

Avoid Ingredient blacklist Alcohol / fragrance / harsh preservatives, high-strength acids, soap-based cleansers, scrubs.

Seek out Repair whitelist Prioritize ceramides, squalane, panthenol, centella, and hyaluronic acid.

Also covers: three under-300-RMB combos (Curel / Simplex / Avène + Winona) · daily principles · pitfalls.

What a single record looks like.

Three core delivery fields (in blue), plus metadata for evaluation, indexing, and provenance — ready to plug straight into a training / evaluation pipeline.

<code>prompt</code>	The deep-decision question itself, with a clear line of inquiry and real business-scenario friction. <code>core</code>
<code>golden_response</code>	A benchmark answer written by a domain expert, serving as the highest-quality anchor for RLHF / SFT. <code>core</code>
<code>rubrics</code>	Dozens of weighted scoring rules with a strict criteria system, add/deduct points, and factual sources — a perfect fit for automated evaluation. <code>core</code>
<code>domain</code>	Nine top-level domains + 30+ sub-topic tags, for precise indexing by business line.
<code>model_scores</code>	A quantitative score matrix across leading frontier models, an out-of-the-box competitive-alignment reference.

THE EXPERT WORKFORCE

AI's generic talk can't fool a real expert.

200K+

Verified Experts

50+

Disciplines

9

Domain Tracks

On **Zhihu**'s 15-year expert community, SoTALab strictly credential-verifies every contributor. The reference answers come from those who truly own the know-how —

EXPERTS WHO SET THE STANDARD FOR DEEP-DECISION PROBLEMS • MAPPED TO DOMAINS

● Health & Medical	Attending / chief physicians at top hospitals (oncology • surgery • internal medicine)
● Career & Money	CFA / FRM charterholders • IB / PE analysts • CPA advisors
● Interests	Cross-border e-commerce operators • management consultants
● Emotion & Mind	Psychology researchers (SSCI-published)
● Legal & Rights	Litigators at top law firms • patent attorneys
● Lifestyle & Beauty	Home, parenting, wellness & beauty life specialists

Not low-cost crowdsourced writers —
a network of experts who truly own the know-how.

DELIVERY • PUT THE WORKFORCE BEHIND YOUR MODELS

Turn this expert network into your model's foundational moat.

1,800+

Real deep-decision problems, expert-calibrated

50

Disciplines • cross-domain depth

9

Domain tracks, customizable mix

Experts

Calibrated by senior practitioners in each vertical

Model Training

Deep experience training / evaluating frontier models

Infrastructure

Industry-leading data production and evaluation infrastructure

Stacking parameters won't buy industry know-how — train the next generation of top general intelligence with the best human intuition. Tell us your training / evaluation direction, and we'll quickly assemble a matching expert-data and evaluation pipeline.

Request full samples • book a walkthrough →

The Expert Intelligence Layer of AI

Smart $\neq$ Useful.
It takes SoTA to train SOTA.



BY ZHIHU