

EXPERT INTELLIGENCE LAYER OF AI

深度决策

deepchoice

High-Stakes Life Decision
Deep Research Data

面向重大生活决策场景的
深度研究数据解决方案

数据样本 · 生活重大决策

当问题没有标准答案， AI 答得好吗？

当面对数学题或代码 bug 时，现在的 AI 已经表现得足够聪明。但在真实的日常生活中，我们抛给 AI 的往往是非标问题：每天通勤2小时怎么提升自己？敏感肌如何用300元搭配平价水乳？

这些问题没有绝对的标准答案。用户需要的，不是四平八稳、面面俱到的「正确的废话」，而是带有真实生活经验、有判断力、能直接落地的建议。我们构建这套数据的初衷，就是想看看模型能不能跳出"安全区"，像一个有经验的真人那样去解决实际问题。

为了让主观的"有用性"变得可衡量，我们邀请了各垂类领域经验丰富的专家，结合真实场景撰写了满分回答。他们将自己的实践经验，拆解为数十项细粒度的评分规则——从"是否识别了隐性需求"到"建议是否具备实操性"，逐一定义出好回答的标准。

用户要的不是泛化的模板，
而是专家级的决策支持。

EXPERIENCE · JUDGMENT · ACTIONABILITY

九大垂直赛道， 下沉到真实子话题。

题库横跨女性健康、职场搞钱、萌宠生活、护肤干货等多个垂类，并下沉到极具体的生活语境——如年轻乳腺癌患者的治疗抉择、多猫家庭的异味治理。下方两级图谱即其全貌：

职场搞钱

CAREER & MONEY

424

求职与职业方向选择 157 · 副业 · 转型 · 创业 127 · 个人 IP / 自媒体成长 59 · 理财规划与选择 56 · 法律职业发展与学习 25

健康医疗

HEALTH & MEDICAL

382

慢病与体检指标管理 137 · 久坐劳损与体态调理 65 · 结节与重疾治疗决策 63 · 口腔健康与舒缓 57 · 老年健康生活 36 · 季节性健康防护 24

兴趣种草

INTERESTS & PICKS

341

小众旅行推荐 103 · 预算购机推荐 58 · 购车预算与选择 49 · 实用心意礼物推荐 42 · 文艺作品推荐 39 · 维权指南合集 35 · 游戏推荐与开发 15

居家生活

HOME & LIVING

225

室内空气质量监测 66 · 租房 · 装修 · 收纳 60 · 美食推荐与分享 46 · 猫咪养护与健康 32 · 饮品入门与选择 21

养生健身

WELLNESS & FITNESS

145

减脂加餐指南 54 · 科学调整睡眠 48 · 运动减肥攻略 43

教育学习

LEARNING

105

高效学习方法 76 · 历史学习路径 29

母婴育儿

PARENTING & KIDS

79

育儿与孩子成长 79

情感心理

EMOTION & MIND

76

情感沟通与选择 42 · 情绪与专注调节 34

美妆护肤

BEAUTY & SKINCARE

37

敏感肌护肤指南 37

合计 1,800+ 条 · 完整可交互气泡图见官网 [sotalab.ai](#)

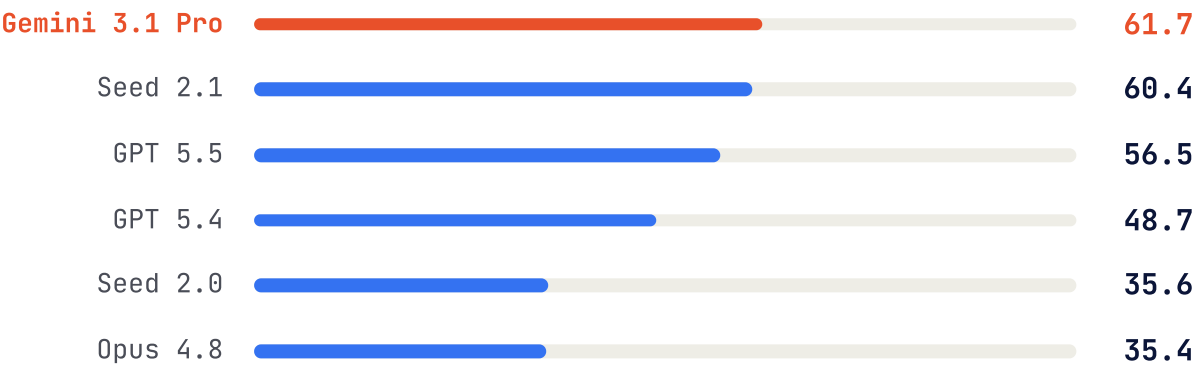
评测样本 · 真实问题 + 多模型量化得分

样本 20 职场搞钱 高效学习方法

通勤时间的高效利用方案

每天通勤 2 小时，想利用这段时间提升自己，有什么适合的学习方式或内容推荐？

六款前沿模型量化得分 · 评分上限 100



评测维度 26 项 · 正向满分 193 分 · 评分上限 100

评分细则 · 节选 5 / 26 项

- 以「认知唤醒 + 时间转化」破题：核心是根据真实通勤环境精准匹配学习方式。
- 按场景分模式：拥挤 / 驾使用「只带耳朵」，有座平稳用「沉浸阅读」。
- 硬核技能（编程 / 数据分析）通勤只听理论，动手实操留给整块时间。
- 守住安全底线：驾驶通勤状态下绝对不可观看屏幕。
- 不得推荐根本不存在、纯属捏造虚构的 APP / 工具 / 平台名称（事实性扣分）。

评测维度：26 项 · 通过 关键失分

专家标准回答 · GOLDEN · 节选

GOLDEN RESPONSE 专家亲笔 · 节选

每天 2 小时通勤，是一笔被忽视的隐形财富。一年 250 个工作日即 500 小时，足以让你在一个领域从小白长成半个专家。关键不在「学什么」，而在如何根据真实通勤环境匹配学习方式。

「只带耳朵」模式 拥挤 / 驾驶 / 骑行
最普适护眼，不受姿势屏幕限制——播客听深度内容、App 听书、安全场景语音备忘录轻量复盘。

「沉浸阅读」模式 有座 · 平稳 稍后读工具集中处理长文 + 闪卡间隔重复，把碎片时间变成结构化输入。

正文另含：三大高回报方向（语言 / 职场 / 通讯）
· 五大落地策略 · 单程 1 小时拆解模板 · 启动建议。正式版含全文。

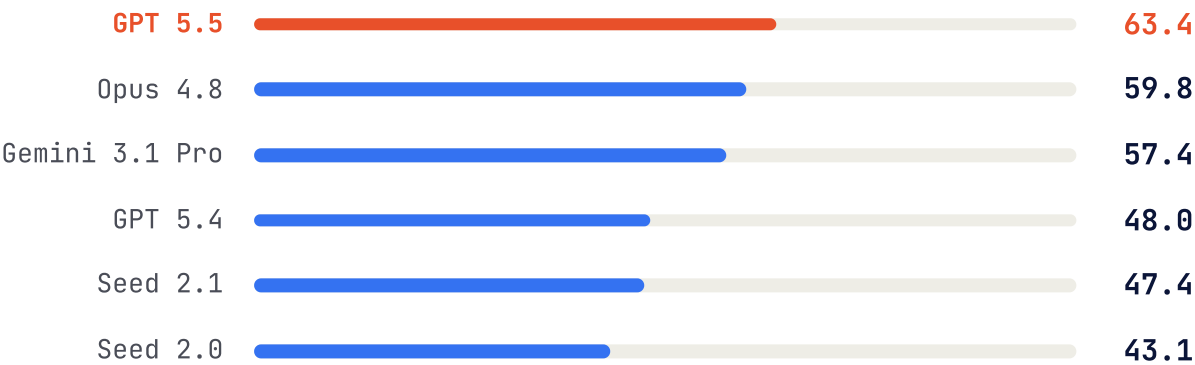
评测样本 · 当所有模型都够不到天花板

样本 08 美妆护肤 敏感肌护肤指南

敏感肌，300 元内选购一套温和补水护肤产品

20~30 岁女性，敏感肌，怎样选购一套温和补水、性价比高的护肤产品，预算 300 元以内？

六款前沿模型量化得分 · 评分上限 100



评测维度 38 项 · 正向满分 290 分 · 评分上限 100

评分细则 · 节选 5 / 38 项

- 以屏障修护为核心、精简护肤步骤为总原则（只留清洁 / 保湿 / 修护）。
- 必须避开含酒精、香精成分（易引发泛红与刺痛）。
- 优先选择含神经酰胺、积雪草等修护舒缓成分。
- 不得把皂基洁面推荐为敏感肌日常首选（扣分）。
- 不得在单品相加超 300 元时仍声称「符合预算」（违反硬约束，扣分）。

评测维度：38 项 · ● 通过 ● 关键失分

专家标准回答 · GOLDEN · 节选

GOLDEN RESPONSE 专家亲笔 · 节选

整体思路必须以修护肌肤屏障为主。这个阶段皮肤脆弱，严格遵循**精简护肤**——只留清洁、保湿、修护三步，总预算严格控制在**300 元以内**。

避雷 成分黑名单 酒精 / 香精 / 强防腐、高浓度果酸与水杨酸、皂基洁面与磨砂——都在剥脱本就脆弱的屏障。

寻宝 修护白名单 优先神经酰胺、角鲨烷、泛醇、积雪草、玻尿酸；泛红不适优先含积雪草的舒缓成分。

正文另含：三套 300 元内黄金搭配（珂润 / 溪木源 / 雅漾 + 薇诺娜）· 日常护肤准则 · 避坑提示。正式版含全文。

一条数据长什么样。

三个核心交付字段（标蓝），加一组用于评测、索引与溯源的元数据——开箱即用，可直接接入训练 / 评测管线。

prompt	深度决策难题题目本身，具备明确探究方向与商业场景对抗性。 核心
golden_response	领域专家亲笔撰写的标杆回答，作为 RLHF / SFT 的最高质量锚点。 核心
rubrics	数十项加权评分细则，含严格判据体系、加扣分项与事实出处，完美适配自动化评测流。 核心
domain	九大一级领域 + 30+ 二级子话题标签，支持按业务线精准索引。
model_scores	多款主流前沿模型的量化跑分矩阵，提供开箱即用的竞品对齐参照。

专家网络 · THE EXPERT WORKFORCE

AI 的泛泛谈资， 欺骗不了真正的行家。

200K+认证专家
Verified Experts**50+**覆盖学科
Disciplines**9**领域赛道
Domain Tracks

SoTALab 依托知乎 15 年沉淀的专业社区，对每一位数据贡献者进行严苛的身份与履历核验（credential-verified）。深度决策难题的标准回答，出自其中真正掌握行业 Know-how 的专家——

为深度决策难题定标的专家 · MAPPED TO DOMAINS

- **健康医疗** 三甲医院主治 / 主任医师（肿瘤 · 外科 · 内科）
- **职场搞钱** CFA / FRM 持证人 · 投行 / PE 分析师 · CPA 税务顾问
- **兴趣种草** 跨境电商操盘手 · 管理咨询顾问
- **情感心理** 心理学研究者（SSCI 发表）
- **法律维权** 顶级律所诉讼律师 · 专利代理人
- **居家 · 育儿 · 养生 · 美妆** 各领域具备丰富实践与生活经验的资深达人

不是低价众包写手，
是真正掌握行业 Know-how 的专家网络。

交付 · PUT THE WORKFORCE BEHIND YOUR MODELS

把这支专家网络， 沉淀为大模型的底层护城河。

1,800+

专家定标的真实深度决策难题

50

覆盖学科 · 跨领域专业纵深

9

领域赛道，配比可定制

行业专家

各垂类资深从业者亲笔定标

模型训练

深度参与前沿大模型训练 / 评测经验

领先基建

行业领先的数据生产与评测基础设施

堆砌参数无法带来行业 Know-how——用最顶尖的人类直觉，训练下一代最顶尖的通用智能。告知我们您的训练 / 评测方向，我们将快速组建对口的专家数据与评测管线。

索取完整样本 · 预约 walkthrough →

contact@sotalab.ai · 微信 / 飞书联系对接人

The Expert Intelligence Layer of AI

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



BY ZHIHU