

虚拟人仿真调研 · 方法与边界白皮书

—— 我们不卖「更接近真人」；我们卖的是「你能亲手验证」

白皮书 v1.0 | 修订日期 2026-10-16 | 对应调研平台版本 V1.6.1 | 文中所有数字均注明取样条件

两个版本号分别代表什么（先说清，避免误以为本文档过期）：

「白皮书 v1.0」是本文档自己的修订号；「V1.6.1」是调研平台的版本号。

两者节奏不同：本文档多数平台版本发布时并不需要改。判断本文档是否过期，请以「对应调研平台版本」这一行为准。

零、这份文件为什么这样写

我们是小公司，没有真人面板可以背书。我们能提供的只有一件事 ——

让每一个说法都能被你亲手验证。

因此这份白皮书里：该说清的限制一条不省；每个实测数字都带取样条件；并且我们主动引用了一篇打到我们自己的研究（见第三节第 4 条）。

一、我们是什么（一句话）

**我们不卖「更接近真人」。我们提供的是：一个可复现、可回查、边界写在报告里的合成受访者体系 ——

用于在花真钱做真人调研之前，先把问卷与判断练一遍。 **

二、它是怎么工作的（只写能被验证的）

部件	事实	你怎么验证
虚拟城市人口	一座持续运行的虚拟城市中的市民人口，按普查与统计公报口径建模	见第五节「口径核对」——我们连取样条件一起给你
唯一身份	每位受访者有唯一编号，可在平台上回查他的档案与生活轨迹	报告导出里带编号与回查入口
抽样可复现	同一套参数取样，拿到的是同一批人	同参数取样两次，名单逐位相同（我们提供复现脚本）
作答可复现	同一套参数重跑，结果逐格一致	跑两次同样的问卷，分布与逐人答案完全相同
两种模式，代价讲清	仿真版：秒级、零边际成本、完全可复现，但不做语义理解、开放题不作答；模型版：由大模型逐人作答、更会说人话，但有波动、不可逐字复现	报告里写明本次是哪种作答路径
作答路径公开	如果你的账号选的是	报告「口径与边界」页 + PDF 首页警示框

	模型版、但本次没有任何题目由模型作答，报告会明确告诉你	
--	-----------------------------	--

三、我们实测到的（每条都带局限）

1. 仿真版对措辞改写不敏感

同一批市民、同一议题、三种措辞（原／同义改写／原措辞重复）分别施测：

三份分布完全相同，逐人答案完全相同。

⇒ 仿真版的结论稳、可复现；换一种说法不会制造出假差异。

局限：单一议题、样本较小；这是仿真版结论，不能外推到模型版。

2. 模型版对措辞敏感

同义改写题干（语义与选项顺序不变）后重新施测，分布漂移中位约 5 个百分点、最大约 13 个百分点，明显大于重复跑的波动。

⇒ 同一轮研究内请不要中途改动题干措辞；跨轮比较请确认措辞一致。

局限：仅适用于模型版；仿真版不受此影响。

3. 模型版不可逐字复现，但抖动有量级

同参数重跑，各选项占比波动约 2 个百分点；同一个人同一道题重复作答，约有四分之一的题次会给出不同答案（与温度设置无关）。

⇒ 请把模型作答当作方向性参考，不要做小数点级比较。

4. 我们主动引用一篇打到我们自己的研究

Maiorano (2026) 用 Upworthy 的真实 A/B 测试（实测点击率）作对照，发现：

按真实受众人口统计构造的「十个人格方案」，其预测准确性低于「不设人格」的简单基线

（排序相关 $\tau = 0.084$ 对 0.361 ；首选命中率 34.6% 对 49.2% ，置信区间不重叠，换模型可复现）。

该研究还指出：多数 A/B 测试本身没有统计上可区分的赢家。

我们怎么用这条：

- 我们不声称我们的结果比"不设人格"更准，也不声称接近真人
- 我们提供的是分布推演 + 可核验性，不是对真实参与度的预测
- 该研究测的是文案/标题排序，不是分布类问卷调研 —— 我们不夸大它的适用范围

四、我们不声称的（与上面成对出现）

不声称	原因
任何真人效度数字	我们没有真人面板，也没有真人锚点数据
替代真人调研	同上；且业界结论互相矛盾
「比不设人格更准」	见第三节第 4 条，人格式模拟可能降低准确性
「模型记得上次答过什么」	不成立 —— 编号连续，但经验不连续（我们已决定不做跨问卷记忆）

「人口越多越准」	实测边际信息差别不大；规模不是我们的方向
----------	----------------------

另外三条我们必须一贯讲明：

- ① 我们**没有任何**对真人的效度数字；
- ② 让市民"更像活的"**没有**让答案更分散（实测 8% 对 10%）；
- ③ **情感与社会互动、意外发现，我们不做。**

五、口径核对（我们把取样条件一起给你）

项	值
指标	城市成年人口中 大专及以上 占比
取样条件	n=2000 2026-10-16 对应平台 V1.6.1
本次实测	34.55% （高中及以下 65.45% 大专 18.80% 本科 12.95% 硕士及以上 2.80%）
普查口径参考	33.12% （第七次全国人口普查； 分母与口径需以官方原文为准 ）
偏差	1.43 个百分点

⚠ 我们特意保留一个"曾经的错误写法"作对照：

我们最初把「大专及以上 33.05%、与普查差 0.07 个百分点」当成一个**恒定数字**写进材料。

写复现脚本时才证明这是错的 —— 同一天换取样参数，实测就是 34.55%（差 1.43pp）。

⇒ **因此本白皮书里所有此类数字都带取样条件；该数字随取样变化。**

六、你能亲手验证的事（我们提供复现脚本）

脚本	验证什么	我们的实测结果
取样可复现	同参数取样两次，名单是否逐位相同	n=30/100/300 均逐位相同（并含 阴性对照 ）
同参重跑一致	跑两次同样的问卷，分布与逐人答案是否相同	分布逐格相同、逐人答案 0 差异
口径核对	学历结构与普查口径的偏差	34.55% vs 33.12%（差 1.43pp）

每个脚本里都写明了「判据纪律」 —— 包括我们自己踩过的坑（例如：**不能拿"同义改写的题干"当阴性对照**，

因为仿真版本来就该给出相同分布）。⇒ **你跑一遍，就能验证或推翻我们上面每一条。**

七、我们适合用在什么场景 / 不适合什么

适合

- 在正式投放真人调研**之前**，先跑一版看方向、检查问卷与选项设计是否有坑
- 需要**完全可复现**的分布推演（仿真版）
- 需要**逐人可回查**的样本（每一位都有编号、有档案、有轨迹）
- 预算或时间不允许做真人调研，但**需要一个可复现的参考基线**

不适合

- 需要真人效度的任何结论（我们不具备）
- 需要预测真实点击率/销量类结果（见第三节第 4 条）
- 情感与社会互动、意外发现类研究（我们不做）

八、一句话总结

**我们的价值不在"更接近真人"，而在"每一个数字你都能自己复跑、每一位受访者你都能自己回查、

以及我们主动告诉你哪里不能用"。 **

在一个谁都可以声称"我的合成更真"的市场里，这是我们唯一能保证的事。

*本文件所引第三方研究为公开资料，我们未对其做独立复核。所有实测数字均注明取样条件与对应版本。 *