

科研人员使用 AI 的需求与顾虑

基于 20 份 AI 模拟回答的演示分析

数据性质

20 份 AI 模拟回答，流程演示
不代表真实受访者

样本编号

DEMO-001 — DEMO-020

采集方式

飞书多维表格收集表
匿名访问 · 已停止收集

出品 / 日期

LightRead · 2026年10月

目录

一、这份报告要回答什么	3
二、问卷与模拟样本	3
三、数据质量	5
四、主要统计结果	6
五、开放建议	9
六、可考虑的改进方向	10
七、局限	11
八、下一步真实调研方案	11
附录 A 题目与字段	12
附录 B 复现说明	12

数据性质声明：本报告的 20 份回答全部由 AI 按设定的人物画像生成，经飞书收集表真实提交并回读核验。**20 份 AI 模拟数据，仅为流程演示，不代表真实受访者，也不能外推到科研人群。**

分母说明：总体描述与多选题的比例，分母是 20 份回答；分组比较里的“组内占比”（如某个组内“几乎每天”的比例），分母是该组人数 n，表中逐组给出 n。两类比例不可混用，也不可相加。

一、这份报告要回答什么

这份报告是一次完整调研流程的演示产物：从问卷设计、飞书收集表搭建、模拟填写、停止收集、数据核验，到描述统计与成稿。它要回答的问题很具体——如果对科研人员做一轮关于 AI 使用的问卷，我们会看到什么样的数据结构，哪些指标值得在真实调研里继续追。

因此报告只做描述，不做推断：不给显著性检验，不做因果解释，不估计市场规模。文中所有“更”“高于”“集中在”都是在 20 份回答内部比较，不指向真实群体。

数值显示口径：小数统一按四舍五入保留两位（Decimal ROUND_HALF_UP，不用二进制浮点的默认舍入），百分比统一保留一位。每周节省时间的精确均值为 4.225 小时，报告显示 4.23 小时；stats.json 同时保留精确值与显示值。

二、问卷与模拟样本

2.1 问卷结构

问卷共 9 题，设计目标是 2 分钟内可填完、不收集可识别个人身份的信息。第 1—8 题必填，第 9 题选填。

#	题目	题型	必填	作用
1	演示样本编号	文本	是	逐份追溯，DEMO-001…020
2	数据来源（固定标记）	单选，唯一选项「模拟」	是	每行自带数据性质标记
3	科研身份	单选 7 项	是	分层的基本变量
4	AI 使用频率	单选 5 项	是	区分重度与边缘用户

5	主要使用场景	多选 8 项	是	看需求落在哪些环节
6	主要顾虑	多选 8 项	是	看阻力在哪里
7	每周因使用 AI 节省的时间	数字（小时，1 位小数）	是	定量口径，允许填 0
8	使用满意度	1—5 分评分	是	与顾虑形成对照
9	最希望改进的一件事	文本	否	收一句可直接引用的原话

问卷本身不含姓名、手机号、单位名称字段，也没有人员类字段；本批 20 条记录全部是 AI 模拟数据，表单在收集期间允许匿名访问。需要说清的是：这里能确认的是“问卷不采集身份字段、表单允许匿名填写”，**不对飞书底层的提交日志、访问记录或记录关联关系作任何承诺。**

2.2 样本构成

20 份回答的身份与使用频率分布如下（分母均为 20）。

科研身份	n	占比	AI 使用频率	n	占比
硕士研究生	4	20.0%	几乎每天	7	35.0%
博士研究生	4	20.0%	每周数次	8	40.0%
博士后	3	15.0%	每月数次	3	15.0%
高校教师或科研岗	3	15.0%	极少使用	2	10.0%
科研院所研究人员	3	15.0%	从未使用	0	0.0%
企业研发人员	2	10.0%			
其他	1	5.0%			

学生合计 8 份，教学科研岗（博士后、教师、院所）合计 9 份，企业研发与其他 3 份。频率上以“几乎每天”和“每周数次”为主，合计 15 份；低频端 5 份。**样本里没有「从未使用」**——原因见 3.3。

2.3 模拟样本是怎么造的

20 份回答不是随机勾选，而是先设 20 个人物画像，再让人设内部的各项取值自治：一个“每周用 AI 写综述的博士生”会同时高频率、多场景、较高满意度、较长节省时间；一个“试过两次就没再用的院所研究员”则低频、单一场景、低满意度、几乎

无时间收益。这样做的目的是让演示数据看起来像一个真实的问卷结果，而不是随机数表。

代价也很清楚：内部一致性是**设计出来的**。样本内的所有关联（比如使用频率高的人节省时间也多）都来自这套画像设定，不是从真实行为里观测到的，因此不能用来解释任何真实原因。

三、数据质量

3.1 采集与固定

收集表在提交完成后被关闭（enabled=false，回读确认），随后从飞书整表重新读取，records_count=20、has_more=false，即 20 条记录一次性完整取回。分析只使用这次回读，不使用本地准备提交的答案文件。原始回读与提交回执单独存放，未随公开材料发布。

3.2 完整性核验

检查项	结果
记录数	20
编号完整性	DEMO-001…DEMO-020 齐全、唯一、无重复
数据来源	全部为「模拟」
必填项缺失	8 个必填字段缺失数均为 0
满意度取值	全部为 1—5 整数
节省小时	0.5—10.0，无非负违规
整行空白记录	0

回读记录与提交计划逐条比对（身份、频率、场景、顾虑、小时、满意度、建议），20 条全部一致，没有错位。

3.3 三件需要说明的事

其一，样本里没有「从未使用」。问卷把使用场景设为必填，而从未使用者无法自治地选择场景与满意度；硬填会造出前后矛盾的样本。低频一端由「极少使用」（2 份）承担。这是模拟样本的设计取舍，不是数据缺失；正式版应当为「从未使用」设置跳题。

其二，开放题是编码后才统计的。20 条建议由 AI 按原意归并主题，并由本案例审阅；编码表与数据一同保存，可自行改判。没有把未编码文本直接当分类变量。

其三，多选题的分母是 20。每项比例=选择该项人数 ÷ 20，因此各项比例之和会超过 100%。读图时不要把这些百分比相加；分组比较里的组内占比另有分母，见 4.4 的说明。

四、主要统计结果

4.1 使用场景：入口在文献环节

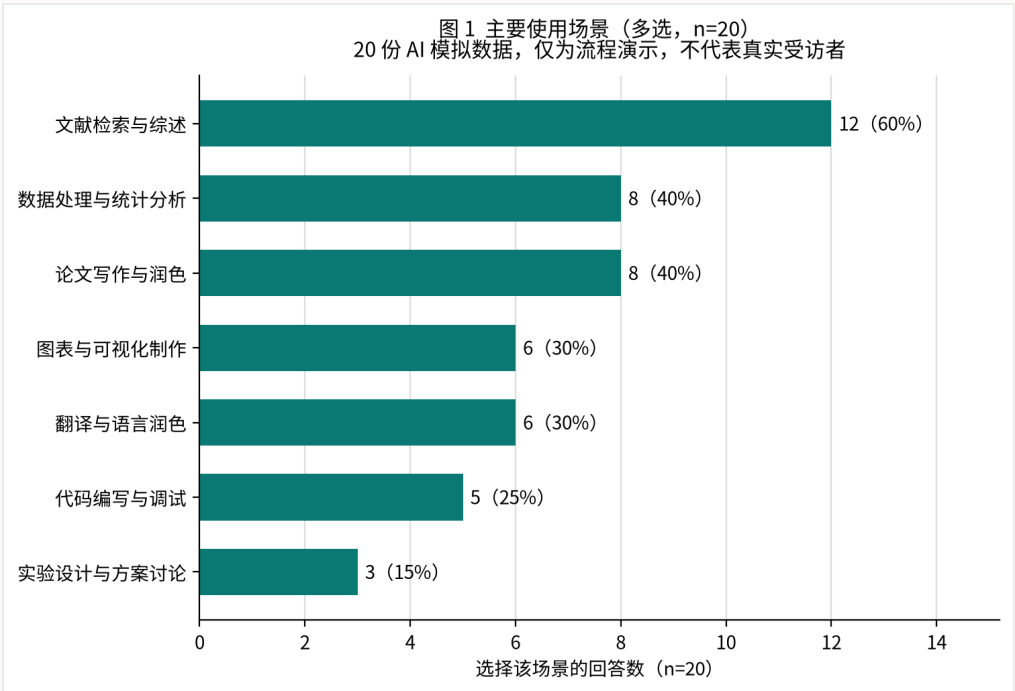


图 1 主要使用场景（多选，n=20）。分母为 20 份回答，各项百分比之和超过 100%。图内已标注数据性质。

12 份（60%）把「文献检索与综述」列为使用场景，是唯一过半的选项；「数据处理与统计分析」与「论文写作与润色」各 8 份（40%）；「图表与可视化制作」「翻译与语言润色」各 6 份（30%）；「代码编写与调试」5 份（25%）；「实验设计与方案讨论」只有 3 份（15%）。

这组数字说的是一件比较朴素的事：这批画像里的 AI 使用集中在读与写，越靠近判断性、设计性的环节越少。真实情况是否也如此，需要真调研才知道。

4.2 顾虑：准确性与引用被选中最多

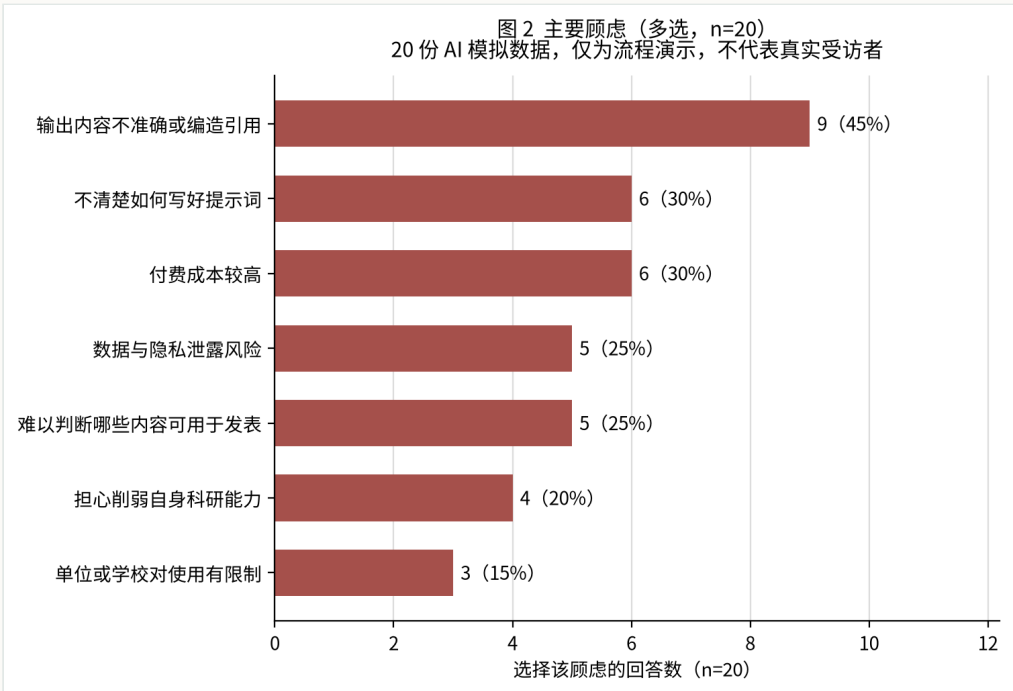


图 2 主要顾虑（多选，n=20）。分母同上，各项比例不可相加。图内已标注数据性质。

「输出内容不准确或编造引用」被 9 份（45%）选中，是唯一超过四成的选项。按被选份数往下数：「不清楚如何写好提示词」与「付费成本较高」各 6 份（30%）；「数据与隐私泄露风险」「难以判断哪些内容可用于发表」各 5 份（25%）；「担心削弱自身科研能力」4 份（20%）；「单位或学校对使用有限制」3 份（15%）。这是多选题，份数只说明有多少人勾选，不构成排序或优先级。

值得注意的是，“不会用”（提示词 30%）与“用不起”（成本 30%）被选份数相同，说明在这批画像里，门槛不只来自技术可信度，也来自操作与预算。

4.3 时间收益与满意度

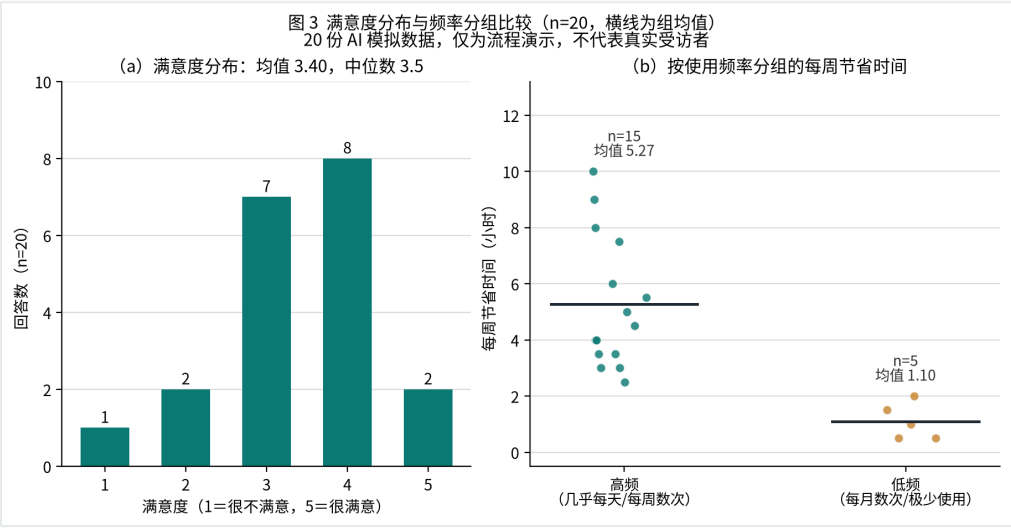


图 3 (a) 满意度分布, 均值 3.40、中位数 3.5; (b) 按使用频率分组的每周节省时间, 横线为组均值, 高频组 n=15、低频组 n=5。

每周因使用 AI 节省的时间: 均值 4.23 小时 (精确值 4.225, 四舍五入保留两位; CSV 原值未改动)、中位数 3.75 小时, P25—P75 为 2.13—5.88 小时, 最小 0.5 小时、最大 10 小时, 20 份合计 84.5 小时/周。分布右偏, 说明少数重度使用者拉高了均值, 所以中位数更值得一看。

满意度是 1—5 序数: 均值 3.40、中位数 3.5。分布上, 1 分 1 份 (5%)、2 分 2 份 (10%)、3 分 7 份 (35%)、4 分 8 份 (40%)、5 分 2 份 (10%); 3—4 分合计 15 份 (75%), 4 分及以上 10 份 (50%)。这些是分布事实; 本案例没有预先定义 “满意” 的阈值, 因此不对整体态度给结论。

序数量表不宜只报均值, 这里的均值仅作概览; 解读以中位数与分布为主。

4.4 分组比较

按使用频率分组:

组	n	节省小时 均值	中位 数	满意度均 值	组内 “几乎每天” 占比
高频 (几乎每天/每周 数次)	15	5.27	4.5	3.80	46.7%
低频 (每月数次/极少 使用)	5	1.10	1.0	2.20	0.0%

按身份分组:

组	n	节省小时 均值	中位 数	满意度均 值	组内“几乎每天” 占比
学生（硕士/博士）	8	4.94	4.5	3.75	62.5%
教学科研岗（博士后/教师/院所）	9	3.33	3.5	3.00	11.1%
企业研发与其他	3	5.00	5.5	3.67	33.3%

表中“组内占比”以该组 n 为分母（例如 $7 \div 15 = 46.7\%$ ），与前面以 20 为分母的总体比例不是同一口径。

高频组的节省时间与满意度都高于低频组，学生组的“几乎每天”占比高于教学科研岗组。这些差值**完全来自模拟样本的人物画像设定**：画像本身规定了频率、场景、时间与满意度的高低搭配，所以它们只是设定的直接结果，既不能识别任何真实原因，也不能作为真实人群的结论。组内 n 最小为 3，表里的数字只用于说明怎么读表。

五、开放建议

20 条建议按原意归并后，主题分布如下（分母 20，编码表与数据一同保存）。

主题	n	占比	示例（模拟编号）
可信度：引用可核验	8	40.0%	DEMO-001、DEMO-005、DEMO-007
成本：价格与单位采购	3	15.0%	DEMO-008、DEMO-012、DEMO-015
合规：使用留痕与声明	2	10.0%	DEMO-016、DEMO-020
数据与隐私：本地化部署	2	10.0%	DEMO-004、DEMO-018
边界：不替代自己的判断	2	10.0%	DEMO-003、DEMO-009
上手门槛：提示词与用法	1	5.0%	DEMO-002
能力：输出可直接执行	1	5.0%	DEMO-011
能力：长文档完整性	1	5.0%	DEMO-006

主题集中在一件事上：能不能核对。八条建议（DEMO-001、005、007、010、013、014、017、019）指向同一类诉求，只是说法不同：

DEMO-001

“希望引用能直接链到原文页码，现在每条都要自己回查一遍。”

DEMO-005

“检索结果要能给出可核查的来源，不然不敢用。”

DEMO-019

“试过两次，给的文献查不到，之后就没再用。”

其次是成本与采购：学生希望有教学版（DEMO-015），院所研究员希望单位统一采购（DEMO-012），企业研发提到团队额度（DEMO-008）。再往后是两类“留痕”诉求——希望标注 AI 参与部分以便在论文里声明（DEMO-016），希望记录用过哪些 AI 片段以便写方法学（DEMO-020）；以及两类“边界”诉求——润色可以，但别替人改论证结构（DEMO-003）；当语言工具可以，选题判断要自己来（DEMO-009）。

六、可考虑的改进方向

下面这些只是针对这个模拟案例的演示建议，用来说明“从数据到动作”可以是怎样一条推导路径。它们不是关于真实科研群体的效果、市场或趋势判断，每一条都需要真实调研验证。

第一，引用与来源的可核验性。开放建议里 8 条（40%，DEMO-001、005、007、010、013、014、017、019）落在这一主题，顾虑题里“输出内容不准确或编造引用”也正好最高（9 份，45%）。两组数字指向同一件事，值得在真实调研里单独设题、看是否稳定复现；本案例不能判断它该不该成为最高优先级。

第二，上手门槛与费用。开放建议里 3 条谈价格与采购（DEMO-008、012、015），1 条谈提示词模板（DEMO-002）；顾虑题里“不清楚如何写好提示词”与“付费成本较高”各 6 份（30%，两者有 2 份重叠：DEMO-002 与 DEMO-015）。费用以个人订阅、单位统一采购还是学生优惠来承担，本案例无法判断，需要在真实样本里分身份问清付款主体。

第三，使用留痕。2 条建议（DEMO-016、020）提到希望标注或记录 AI 参与部分，对应投稿与合规场景。样本里只有 2 条（10%），不足以说明重要性，只能说明这道开放题能收到这类回答；真实调研里应当把“是否需要在论文中声明 AI 使用”做成独立题目。

第四，满意度与使用强度是否同向。本案例里 3—4 分占 15 份（75%），高频组的满意度均值 3.80、低频组 2.20（ $n=15$ 与 $n=5$ ）。这组数值是画像设定的直接产物（画像里低满意度的人物同时也被设成低频低收益），**不能**据此推断真实用户的满意度水平，也不能推断工具进入“日常”或“离不开”的阶段。

七、局限

- **数据是模拟的**：20 份回答由 AI 按画像生成，比例与均值只描述这份演示数据本身，不能作为科研人群的估计。
- **没有抽样框**：不存在“科研人员”总体的抽样设计，也没有加权还原总体的可能。
- **量表与自报量**：满意度是序数，节省时间是自报估计，都没有客观日志对照。
- **组内样本过小**：最小分组 $n=3$ ，组间差值只能作为读表示例，不能做推断。
- **单批次**：只有一次采集，无法看时间变化或重复测量的稳定性。
- **未做检验**：按需求未做显著性检验、未做因果识别。
- **样本关联是设计出来的**：频率、场景、时间、满意度之间的同向关系来自人物画像，不是观测到的真实关联。

八、下一步真实调研方案

如果要把这个案例做成真实调研，建议按下面的顺序推进。

1. **先定目标量**：这份问卷能回答“结构与分布”，回答不了“为什么”。若要解释原因，需要在问卷之外加半结构化访谈。
2. **样本量按精度算，不给统一阈值**：只做比例估计时，按目标置信区间的半宽反推（例如希望某选项比例的 95% 区间半宽控制在 $\pm x$ 个百分点）；要比较两个分层的差异时，按最小可检出差异与检验效能估算；同时受招募渠道、伦理审批与预算约束。本案例的 20 份不具备上述任何条件，只验证了流程能跑通。
3. **补跳题逻辑**：第 4 题选「从未使用」时应跳过使用场景、时间与满意度三题，避免强迫作答产生矛盾记录；本案例的模拟样本因此没有包含「从未使用」。
4. **换掉标记题**：正式版把「演示样本编号」「数据来源」两题去掉，换成匿名编号；数据来源改由采集批次管理，不再让填写者选择。
5. **量化口径收紧**：节省时间改成“上周实际节省的小时数”，并加一道“这个估计有多确定”的自评；满意度后可加一个开放追问，避免只得到分数。
6. **多选题上限按调研目标与预测试决定**：上限设多少取决于要覆盖多少场景、以及预测试里每题的作答负担；上限低会压缩可选范围，上限高会拉长作答时间，具体取多少要看预测试结果。本案例题干写的是“最多选择 4 项”，属于填写说明，未在系统中设置强制上限，也没有验证过上限对作答的影响。
7. **分析计划前置**：在发放前写好分析口径（分母、分组、哪些比较要做、是否校正），避免事后挑结果。
8. **合规与留痕**：正式调研需要知情同意、数据保留期限与使用范围说明；若涉及单位内部信息，还要确认机构对 AI 工具的规定。

附录 A 题目与字段

#	题目	题型	必填	选项数
1	演示样本编号	文本	是	—
2	数据来源（固定标记）	单选	是	1
3	科研身份	单选	是	7
4	AI 使用频率	单选	是	5
5	主要使用场景	多选	是	8
6	主要顾虑	多选	是	8
7	每周因使用 AI 节省的时间	数字	是	—
8	使用满意度	评分	是	5（1—5）
9	最希望改进的一件事	文本	否	—

附录 B 复现说明

分析脚本只读干净 CSV，不访问网络：

```
python3 analyze_survey.py --input clean_responses.csv --
topic-coding topic_coding.csv --outdir . --figdir figures
```

输出 stats.json 与三张图。报告中的所有数字都来自这次运行的输出，可重新计算核对。

显示精度：小数按 Decimal ROUND_HALF_UP 保留两位（stats.json 的 hours.mean=4.23，hours.mean_exact=4.225 同时保留），百分比保留一位；CSV 中保存的是原始填报值，未做任何舍入改动。

数据文件：clean_responses.csv（20 行 × 9 列，UTF-8 BOM，Excel 可直接打开；不含任何飞书记录 ID）。编码表：topic_coding.csv（编号 → 一级主题 → 归并主题）。

本报告为 LightRead 官网博客的流程演示案例。20 份 AI 模拟数据，仅为流程演示，不代表真实受访者。