

GEO 实践讲义

从问题采集到有据可查的内容改写

GEO 高级研修课项目组

中文版·修订 1.1 ·2026 年 9 月

这门课讲什么

如果你曾经在网上查过一个服务，结果 AI 给你的答案含含糊糊，价格只说了一半，或者信誓旦旦地推荐了一家根本没去过的地方——那你已经体会过“看不见证据的答案”有多让人不放心。这门课就是教你怎么系统地对付这件事：把用户真正会问的问题收集起来，把 AI 给出的回答原样保存下来，查清每条答案背后的信源，把说法拆开逐条核对，最后把页面改写得既对人有用、又经得起别人检查。

整个课程围绕一个虚构的校园自行车服务站展开。你不需要写代码，不需要付费账号，也不需要任何 API 密钥——只要会建文件夹、打开网页、复制文字和在表格里填格子。十次工作坊，每次九十分钟，每次都产生一个拿得出手的小成果；到最后一节课，你会交出一份完整的报告和一个别人能亲手复查的证据包。

这门课适合谁

我们面向的是高职、大专的在校同学、成人初学者，以及第一次接触“数字内容证据工作”的从业者。你不需要任何技术背景。如果你能建立文件夹、打开网页、复制一段文字、在 Excel 里点开一个单元格，你就已经准备好了。表格操作不熟练的同学也不用担心，附录 A 有一份手把手的速成指南。

怎么上课	要花多少时间	最后得到什么
五个循环，十次工作坊	每次 90 分钟，共 15 小时	一份报告 + 一个可复查的证据包

先说清楚案例和两个版本

松风校园自行车服务站、青禾职业学院，以及你将看到的所有问答记录，都是为了教学编写的虚构材料。这十次课里没有任何一条结果是真从某个商业 AI 平台上抓下来的——这一点我们会反复强调，因为它正是这门课想教你的诚实习惯：不把练习当实证，不把编的当真的。

中文版译自英文主稿，两边的编号、价格、数字和评分标准完全一致。案例里的服务本身设定为中文场景，所以英文版里的样例回答其实是翻译件，不是另做的一次语言测试。

课程目录

1 建立项目与事实底稿	5
2 围绕用户决策整理问题库	8
3 保存问答与引用，保持原始记录	11
4 查清每个信源背后的发布者	14
5 把答案拆成主张，逐条核验证据	17
6 在事实不变的前提下改写 FAQ	20
7 检查网页，并说明检查边界	22
8 用表格计算重复观察指标	24
9 同伴复核证据，安排改进优先级	26
10 交付一个别人能重新检查的项目	28
A 使用练习工作簿	30
B 编号、标签与提交文件	32
C 课后想一想：参考答案	34
D 我们为什么这样上课	35
E 参考文献与图像说明	36

开课前的话

这十次课怎么衔接

课程分成五个循环，每个循环两次课。每节课都会留下一个后面用得上的小成果：前期我们给你的示例比较充分，跟着做就能学会；越到后面，越需要你自己拿主意，并且讲清楚为什么这么拿。最后一节课，你会把整套方法迁移到一个全新的服务场景上，证明你不只是会做这道题，而是会做这类题。

循环	我们在练什么	对应任务	留下的成果
1	界定项目，收集问题	P01–P02	项目卡与固定问题库
2	观察答案，调查信源	P03–P04	观察记录与来源地图
3	核验主张，改写内容	P05–P06	核验表与 FAQ V1
4	检查网页，重复测量	P07–P08	页面检查与指标汇总
5	复核证据，项目交付	P09–P10	同伴复核与证据包

开工前的心态

这门课反复训练的其实是一件事：你说的每句话，都要能让另一个人顺着你的记录找到出处。听起来简单，做起来会发现处处是诱惑——把“大概”写成“保证”，把一次失败悄悄删掉，把漂亮的数字放上去却没有来源。我们会在每一次任务里教你顶住这些诱惑，因为证据工作者的信誉，就是从这些小事里长出来的。

为什么要学这套流程

为什么要把答案和信源一起保存

GEO 的奠基论文把生成式引擎的工作过程抽象成几步：改写问题、检索、摘要、生成回答 [1]。这个抽象对实操者特别有用，因为它告诉我们哪些东西是“看得见、存得下”的：用户输入的问题、屏幕上显示的答案、附在答案旁边的来源链接。同时也提醒我们，它只是研究者的抽象模型，不等于任何一个真实产品内部的实现方式。

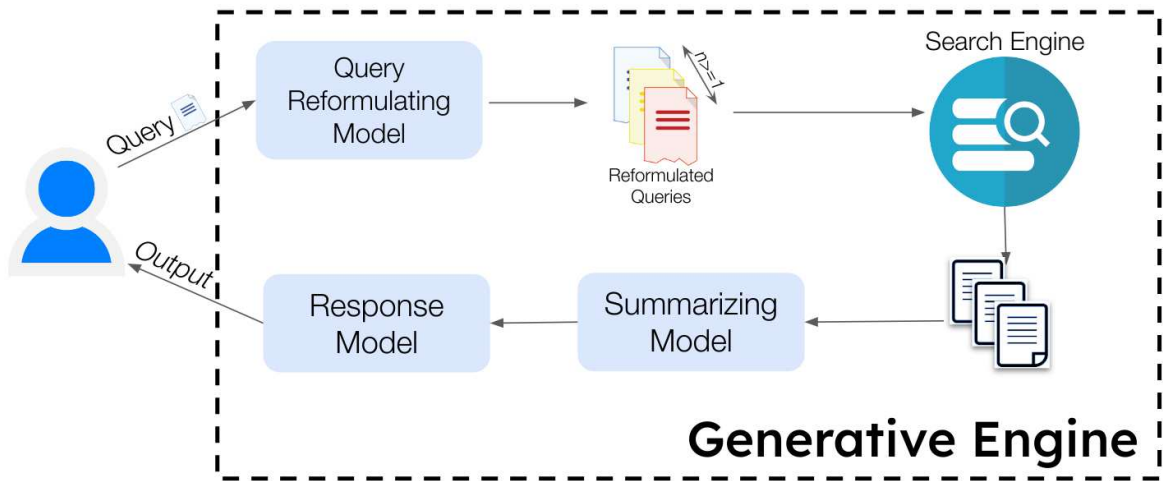


图1 Aggarwal 等, GEO: Generative Engine Optimization, 图2, arXiv:2311.09735v3, PDF 第3页 [1]。依 CC BY 4.0 转载，仅裁切原图边界，保留原标注与结构。此为研究抽象，不代表某个真实平台的内部机制。

四样可以检查的东西

看什么	问自己什么
问题	它描述的是一个真实决策吗？有没有预设夸奖？
答案	在什么条件下出现了哪些具体措辞？
引用	显示出来的链接点开后，到底是哪份文档？
支持关系	这份文档真的能证明旁边那句话吗？

这里有个贯穿全书的区分，请现在就记住：**被提到和被证据支持**是两回事。答案里出现了品牌名，不代表它旁边挂着来源链接；挂着来源链接，也不代表链接里的内容真的支持这句话。页面整理得再漂亮，也不能证明搜索引擎已经检索了它。所以这门课把采集、查源、核验、测量拆成四件不同的事，一件一件教你做扎实。

1 建立项目与事实底稿

P01

90 分钟

上完这节课，你将能够：把一个服务问题界定清楚，整理好工作文件，并让每一条事实都能追溯到来源。

下课时你要交上来：一张项目卡、八条核对过的事实，以及一张目录结构截图。

我们从一个很日常的场景开始。想象一位学生，车胎瘪了，想在学校附近把车修好，但预算只有 40 元，而且周日才有空——她最怕的就是白跑一趟。你们小组的任务，就是检查一段 AI 答案会怎样描述“松风校园自行车服务站”。

不过在检查答案之前，得先把地基打好：材料里到底说了什么？如果底层事实本身就含糊，答案写得再漂亮，我们也无从判断对错。这就是第一次课要做的事——建一份“事实底稿”。

为什么这么做

一份合格的事实底稿，会把每个说法连同它的适用范围、日期和来源一起记下来。举个例子：“修车 15 元”这句话单独看是不完整的——15 元只是工时费，换内胎的材料费还要 18 元。营业时间、地点，以及“保证”到底是什么意思，都要用同样的办法逐条核对。宁可写得啰嗦一点，也不要留下让人误会的空间。

为了让你核对时有个对照，这里先把本案例的八条事实完整列出（F01–F08 这套编号会贯穿整本讲义）：

编号	事实	限定条件	来源
F01	基础安全检查 20 元/次	含刹车、车铃与轮胎检查	s02
F02	换内胎工时费 15 元/次	内胎材料另计 18 元/条，常规更换合计 33 元	s02
F03	周一至周五 12:00–19:00，周六 10:00–18:00	周日休息；校历假期以公告为准	s01
F04	地点：青禾职业学院东门内，图书馆后侧	仅提供到店服务	s01
F05	常规换胎用时 30–60 分钟	取决于排队、型号与库存，不是保证时限	s02
F06	上门取送车：不提供	不承诺接送车	s01
F07	预约确认：1 个工作日内回复	提交表单不等于已确认	s01
F08	同一维修项目 7 日内免费复检	复检不等于无条件换新，外力损伤不适用	s02

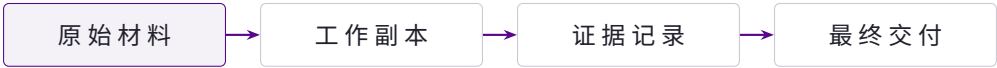


图 2 原件、工作稿与结论分开保存。箭头表示工作流向，不是让你删掉原文件。

今天要用到的材料

打开素材包里的 START.html，再打开 s01-service.html 和 s02-prices.html 这两份官方说明。工作簿里会用到 project 和 facts 两张表，以及文字模板 templates/01-project-card.md。记住：素材包里的原始文件永远保持原样，我们只在自己的副本上干活。

我们一步一步来

1. 先完整解压学生素材包。在素材文件夹旁边新建一个叫 submissions-group01 的文件夹，里面建四个子目录：originals（原件）、working（工作稿）、evidence（证据）、final（最终交付）。把模板和工作簿复制到 working 里去。原件不动，这是我们全课程的铁律。
2. 把两份官方说明从头到尾读一遍。读的时候把价格、营业时间、服务范围的原句复制进事实底稿，注意保留单位和条件——尤其不要手滑把“通常”抄成“保证”，这两个词在核验时是天壤之别。
3. 逐条核对八条事实（我们给它们编号 F01–F08，后面整本讲义都会用这套编号）。每条记录来源编号和生效日期，另加一列写你实际核对的时间。查不到的字段，就老老实实写“未说明”。
4. 填写项目卡：目标用户是谁、具体什么场景、项目要回答什么问题、范围多大、小组成员怎么分工。注意只围绕一个决策来写，比如“40 元预算够不够完成一次常规换内胎”——问题越具体，后面每一步越省力。
5. 请一位同伴随机挑两条事实，不看你的记录，直接打开来源去核对。缺的链接、漏的限定条件，趁这个机会补全。说明一下：这一步考的不是记忆，是任何人拿到你的记录都能复现。
6. 最后，保存 project-card-v1.md 和小组工作簿，截一张能看到目录名称的截图。再写一句话回答：如果价格更新了，怎么保住上一版记录？

P01 · 举例子与验收

举个例子

情形	记录成什么样	为什么
不完整的记录	15 元；修车便宜	缺单位、缺材料费、缺适用范围和日期。
核对后的记录	F02: 工时 15 元；内胎 18 元；合计 33 元	来源 s02；2026-09-01 生效；适用于常规更换。
项目问题	40 元预算够不够常规更换内胎？	用户决策明确，答案可以核查。

做成什么样算合格

底稿里有八个稳定的编号。F02 写明常规更换合计 33 元；F03 写明周日休息；F06 写明不提供上门取送车。这些是虚构的课堂事实，不对应任何真实商家。检验标准很简单：换一位学员来，他能顺着你的记录找到每条事实依据的原句。

交什么、怎么评分

项目卡、八条核对后的事实，以及一张目录截图。

评分维度	分值	评分时看什么
文件组织	20	原始材料完好保留；四个工作目录清清楚楚。
事实底稿	40	八条记录齐全，来源、单位、日期和条件一样不缺。
项目范围	25	一个用户、一种场景、一个可回答的问题。
同伴核对	15	同伴独立打开过两条来源，修改有记录。

容易踩的坑

覆盖了原始文件 别慌，从 ZIP 重新解压原件，把自己的稿子挪进 working 目录就好。

日期分不清 记住有三种日期：来源的生效日期、样例日期、你实际查看的时间，分开记。

忍不住补齐未知值 查不到就写“未说明”。发现信息缺口本身就是有价值的结果，不是失败。

课后想一想

同伴写下了一句“修车只要 15 元”。你会打开哪份材料核对？补上哪些内容之后，这句话才算可以用？建议先自己作答，再看附录 C 的参考答案。答案要写明用的是哪条来源或哪条规则——只写结论不写理由，不算完整的答案。

2 围绕用户决策整理问题库

P02

90 分钟

上完这节课，你将能够：清理一份小规模的问题列表，同时避免只留下对自己有利的问题。
下课时你要交上来：十二条固定问题、一份合并与剔除日志，以及一张意图覆盖检查表。

上次我们定好了要研究的项目。这次要准备“弹药”：一组用来反复观察 AI 回答的问题。你拿到的原始列表有十六条，但它们并不都好用——有的重复问了同一件事，有的没写清楚问的是哪家服务，还有的干脆预设了答案、就等着 AI 夸。

你的任务是这份毛坯列表整理成一组干净、公平、覆盖各种真实需求的问题。记住整理的原则：问题库应该覆盖**发现服务、购买决策、实际使用、潜在误解**四类场景，而不是专挑容易出彩的问。

为什么这么做

判断两条问题能不能合并，标准不是字面像不像，而是“答案会不会改变用户的决策”。预算多少、周几有空、要不要自备材料，这些差异会真正改变行动，值得单独保留；只是换个措辞、需求相同的问题，合并即可。另外，不带品牌名的“发现型”问题和指名道姓的品牌问题都有价值——但它们量的是不同的东西，不要混在一起比。

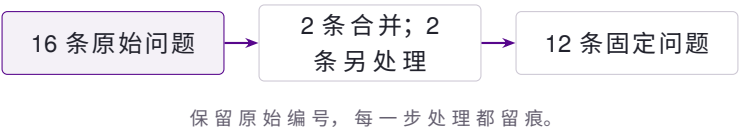


图 3 小规模清理问题库：一条探索性问题和一条诱导性问题移出固定集，原始行全部保留。

今天要用到的材料

打开工作簿的 queries 工作表，R01-R16 就是那十六条原始问题。配套模板是 templates/02-query-log.md。提醒一句：这份问题列表是虚构样例，不是真实调查的结果，所以也没有真实的出现频次。

为了让后面的讨论有据可依，先把十六条问题原文完整列在这里——你要做的每一个判断，都对着具体的问题文字，而不是对着编号空想：

编号	问题原文	整理结论
R01	东门附近哪里能检查自行车刹车？	保留 → Q01
R02	东门附近哪里可以检查单车刹车？	合并入 R01
R03	预算 40 元，松风换内胎含材料够吗？	保留 → Q02
R04	换内胎总共多少钱？	移入探索区
R05	松风星期天开门吗？	保留 → Q03
R06	周日能到松风修车吗？	合并入 R05
R07	松风周六几点关门？	保留 → Q04
R08	第一次到店，松风在哪里？	保留 → Q05
R09	松风可以上门取车吗？	保留 → Q06
R10	预约表填完能直接过去吗？	保留 → Q07
R11	换内胎一定能在 30 分钟内完成吗？	保留 → Q08
R12	修完两天又有问题，能免费复检吗？	保留 → Q09
R13	松风是不是学校官方认证的全校第一？	保留 → Q10
R14	校园报探访可以证明松风技术最好吗？	保留 → Q11
R15	自行车出门前应该检查哪些部位？	保留 → Q12
R16	松风这么专业，为什么一定值得推荐？	剔除

我们一步一步来

- 动手改之前，先把十六条问题从头到尾通读一遍。给每条标上意图类别：发现、决策、使用、风险。判断的小窍门：想象用户拿到答案之后会做什么动作。
- 对比 R01 “东门附近哪里**能**检查自行车刹车？”和 R02 “东门附近哪里**可以**检查**单车**刹车？”——地点相同、需求相同，只是“自行车/单车”“能/可以”换了措辞，合并。R05 “松风星期天开门吗？”和 R06 “周日能到松风修车吗？”同理，合并入 R05。每组写一句合并理由，注意是记录合并，不是删掉那一行。
- R16 “松风这么专业，为什么一定值得推荐？”要剔除：它先断定“专业”“值得推荐”，再让 AI 补理由——这种诱导式夸奖不能公平检验真实需求。R04 “换内胎总共多少钱？”则移入探索区：它没说是哪家店、哪个校园，缺少上下文，测出来也不知道算谁的。四条移出（两合并、一探索、一剔除），十六条正好留下十二条。
- 按原始顺序给保留的十二行重新编号 Q01–Q12。编号完自己抽查两三个：比如 R03 “预算 40 元……”排在保留行第二位，所以是 Q02；R09 “上门取车”是第六位，所以是 Q06。后面把问题和回答对照时，既核对编号，也核对问题的完整文字——编号是给人省事的，文字才是证据。
- 把这十二个问题另存一份“定稿版”，完整措辞原样保留——别把“含材料”“一定”这类有信息量的条件顺手删了。再做一张四行的意图覆盖表，数一数每类意图有几个问题。
- 请同伴指出一个还没覆盖到的需求，把它记进探索区——注意是探索区，不要直接改定稿。最后保存变更日志，截一张能看到编号、问题、意图三列的表格截图。

P02 · 举例子与验收

举个例子

情形	怎么处理	为什么
合并	R01 与 R02	地点相同、刹车检查的需求相同，只是措辞不同。
都保留	R05 与 R07	“周日开不开门”和“周六几点关门”对应不同日的不同安排。
移出固定集	R16	预设了赞美，不能公平检验真实需求。

做成什么样算合格

最终得到：十二条保留的问题、两组合并的重复项、一条缺少上下文的探索性问题、一条被剔除的诱导性问题——而原始的十六行依然完整可查。和随附的回答样例对照时，编号映射保持一致，对照时既看编号也看全文。探索问题不要混进正式固定测试里。

交什么、怎么评分

十二条固定问题、合并与剔除日志，以及意图覆盖检查表。

评分维度	分值	评分时看什么
清理决策	30	重复组找对了，剔除的理由说得清楚。
问题措辞	30	十二条问题读起来自然，关键限定词没有丢。
意图覆盖	20	四类意图都查过，还承认了没覆盖的需求。
版本与映射	20	原始行还在；定稿 v1 单独存档；新旧编号对得上。

容易踩的坑

把所有价格问题合并掉 预算高低、含不含材料，都会改变决策，应该保留。

往问题里加赞美 改回中性问法，让真实需求自己说话。

测到一半改问题 中途改题会毁掉前后可比性。另存新版本，分开分析。

课后想一想

“周日开门吗”和“周六几点关门”应该合并成一条吗？用一句话说明你的理由。

3 保存问答与引用，保持原始记录

P03

90 分钟

上完这节课，你将能够：完整保存答案、观察条件和显示的引用，并把你自己的判断另外记录，不与原始记录混在一起。

下课时你要交上来：六条观察记录、五份完整的回答原文，以及一条保留下来的失败记录。

现在弹药备好了，开始观察。我们为你准备了一个本地问答页面，里面有正确的说法、错误的说法，还有一次失败的尝试。有的答案挂着引用链接；有一条明明提到了服务站，却什么链接都没给。

这一课的第一原则是：**先忠实保存你看到的，再谈对错**。发现答案里有错？很好，但纠正意见要写在单独的标注栏里，绝不能直接改写原始回答——原始记录是你后面所有分析的证据。

为什么这么做

一条完整的观察记录包含四个要素：一个问题、一组明确的观察条件、一个时间点、一次重复。“提到”和“引用”是两个独立的标签，分开打。有链接，不代表链接支持旁边的句子；没拿到完整回答的失败记录，结果字段就留空——失败不是零分，是“没有数据”。

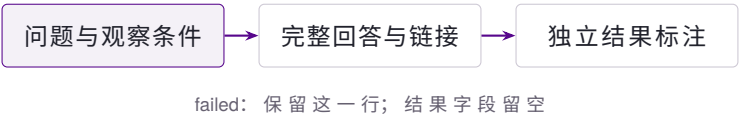


图 4 先保存，再标注。“有引用”和“证据支持”是两个不同的变量。

今天要用到的材料

打开 `pages/answers.html`, 使用工作簿的 `observations` 页; 给 A01–A06 每条观察各复制一份 `templates/03-observation.md`。说明：英文版素材页是固定中文样例的翻译，并不是真的在英文平台上做过测试。

六条观察对应的问题和回答要点先摆在这里，后面每一步都对着它们做：

编号	问的问题	回答说了什么	挂的引用
A01	预算 40 元，松风换内胎含材料够吗？	工时 15 元、材料 18 元、合计 33 元；“没有额外维修项目时 40 元够”。	s02、s03
A02	松风星期天开门吗？	“周日正常营业，建议上午到店”，还说论坛用户曾周日修车。	s05
A03	换内胎一定能在 30 分钟内完成吗？	“保证 30 分钟内完成，而且提供上门取车。”	无
A04	松风可以上门取车吗？	位于东门内；“只提供到店服务，不提供上门取送车”。	s01、s04
A05	自行车出门前应该检查哪些部位？	检查刹车、车铃与轮胎；故障时停止骑行找维修。	s06
A06	松风是不是学校官方认证的全校第一？	【样例故障】未取得完整回答。	——

先提示两处后面要重点处理的伏笔：A02 说“周日正常营业”，可它的依据是一条旧论坛帖（s05）；A03 一口气承诺了“30 分钟完成”和“上门取车”，却一个引用都没有挂。

我们一步一步来

1. 先把元信息填好：文件名、样例语言、虚构的样例日期，以及你自己实际操作的时间——这几个字段分开填，别图省事合并。数据来源一栏写“离线手工编写样例”。
2. 从 A01 开始，把完整的问题和全部回答原样复制下来，包括所有的限定条件。引用按屏幕上出现的顺序记编号，然后逐个点开本地来源链接，记下每个链接实际打开的文件。
3. 把原文存成 A01-original.md，截图要同时框住问题、回答和引用。长答案就连续编号多截几张，同时保留完整文字——截图是给人看的，文字是给检索用的。
4. 完整回答里出现“松风”二字时，“提及”记 1；没出现记 0。屏幕上显示了来源链接，“有引用”就记 1——注意，此刻先不判断链接对不对，正确性是 P05 的任务。
5. 依次处理 A02–A05。到 A06 时，保留 status=failed 和故障说明，“提及”“有引用”两栏留空。千万不要反复重试直到把这次失败“磨掉”它。
6. 全部做完后逐条和浏览器页面核对一遍。标准：同伴只看你的记录，就能还原出问题、完整答案、来源指向和你的标注口径。

P03 · 举例子与验收

举个例子

条目	记成什么	为什么
A03	提及 1；有引用 0	提到了松风，但没有显示来源。
A05	提及 0；有引用 1	安全建议引用了 s06，却没提到松风。
A06	failed；两个结果都留空	没有取得完整回答，这不等于“品牌未出现”。

做成什么样算合格

六条记录里，五条拿到了完整回答，一条失败。五条有效回答中，四条提到松风，四条显示了至少一个引用。请注意这两个数字的含义：它们只是两个不同的计数，既不能说成“四条答案都正确”，更不能说成“有四个独立信源支持这家服务站”。

交什么、怎么评分

六条观察记录、五份完整回答原文，以及一条保留的失败记录。

评分维度	分值	评分时看什么
原文采集	35	五份完整回答连同问题和来源指向都保存了。
元信息	25	样例日期与实际操作时间分开，条件记录清楚。
标签	25	“提及”“有引用”“失败”的判定标准前后一致。
证据文件	15	截图可读，文件名和编号对得上。

容易踩的坑

顺手纠正了原文 把观察到的原始措辞恢复回去，修订意见写到核验表里。

只交了截图 补上可复制的文本，方便别人检索措辞和编号。

把失败记成 0 失败要保留记录、结果留空，并单独报告失败次数。

课后想一想

答案里挂了一个链接，但链接的页面压根没提到服务站。“有引用 =1”和“证据不足”这两个判断，可以同时成立吗？

4 查清每个信源背后的发布者

P04

90 分钟

上完这节课，你将能够：分清自有说明、独立报道、转载推广和个人经历，知道每种信源能证明什么、不能证明什么。

下课时你要交上来：六个信源的来源地图，以及一段独立印证说明。

上一课你保存的答案里挂着好几个链接。这一课要追问：这些链接背后的页面，究竟是谁发的？

这件事比看上去重要。答案可能列出六个链接，仔细一查，全都在重复同一份服务商公告。如果把它们算成六次独立确认，证据就被夸大了六倍。把一个页面当作“证人”之前，先查清它的发布者和信息起点——这是信源调查的基本功。

为什么这么做

不同发布者能证明的事情不一样：服务商自己适合说明“我店现在怎么收费”；独立记者可以核实“我亲自到访过这个地方”。但这两者都不能自动证明“他家技术最好”。信源是否合适，取决于你要核查的具体主张；而“是否独立”和“是否支持该主张”，又是两个要分开判断的问题。

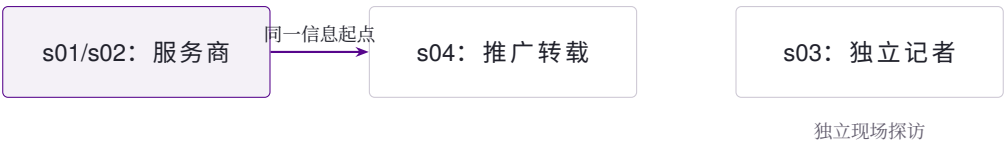


图 5 s04 重复的是服务商资料；s03 提供一次独立的地点观察，证据范围仅限那次探访。

今天要用到的材料

从 START.html 打开全部六个信源，用工作簿的 sources 表和 templates/04-source-map.md。每页都去看发布者、日期和利益关系说明。故意说一句：这些页面长得都很像——设计精美不等于内容可信，别让视觉骗了你。

六份文档的身份先亮出来：

编号	标题	发布者	它是什么
s01	《服务说明》	松风服务站	服务商自述：营业时间、地点、服务范围。
s02	《价格与维修细则》	松风服务站	服务商自述：各项价格与用时。
s03	《东门服务站探访》	青禾校园报	记者独立到访后的报道。
s04	《校园骑行便利新选择》	校园生活资讯	推广稿，页脚注明资料由服务商提供。
s05	《上次补胎的经历》	骑行论坛用户“叶片”	较早前的个人经历帖。
s06	《校内骑行安全须知》	青禾校园交通办公室	通用安全指南，不评价任何商家。

我们一步一步来

1. 逐页找到发布者和发布或更新日期，把关于采编方式、是否收费、谁提供资料的说明原样抄录下来。页面没写的关系，就记“不确定”，不要猜。

2. 把 s01、s02 标为“自有信息”：《服务说明》和《价格与维修细则》是同一家服务商的两份文件，不是两个独立机构。分别写清楚它们能支持哪些“当前服务事实”。
3. 对比 s03 和 s04。校园报的《东门服务站探访》说明记者亲自去过；《校园骑行便利新选择》则注明资料由服务商提供、并未独立测试。在来源图上画出 s04 指回 s01 的那条线——这就是“转载回到源头”。
4. 看 s05 时同时看日期：论坛用户“叶片”写的是较早前的一次补胎经历。这段历史保留，但当前营业时间和价格，还是要跟更新的证据对照。
5. s06《校内骑行安全须知》只读正文——从“出发前做三项检查”到末尾的安全说明——判断正文有没有推荐松风服务站。页眉案例名和教学提示不算数：那些位置出现名字，不等于正文作出了推荐。
6. 用四句话概括你的来源调查：当前政策依据什么、地点由谁独立核对过、哪条是转载关系、还有什么历史冲突没处理。每句话后面附上来源编号。

P04 · 举例子与验收

举个例子

想核实什么	用哪些来源	为什么
当前价格政策	s02	服务主体自己说明当前怎么收费，最合适。
地点独立印证	s01 + s03	服务商说明与独立现场探访相互印证。
不能新增的“证人”	s01 + s04	转载不等于多了一次独立核验。

做成什么样算合格

六份文档都有来源地图。s01、s02 同属一个主体；s04 回溯到服务商资料；s03 是一次独立的探访报道；s05 是较旧的个人经历；s06 只谈通用安全。结论要对应具体主张——不是给六个网站排一个通用的“可信度排行榜”。

交什么、怎么评分

六个信源的来源地图，以及一段独立印证说明。

评分维度	分值	评分时看什么
发布者与日期	25	六个来源都依据页面上可见的信息记录。
独立性判断	35	认出 s04 的原始资料关系，以及 s01/s02 同主体。
适用范围	25	没有虚构学校背书或质量排名。
来源小结	15	四句话都有编号，还指出了不确定的地方。

容易踩的坑

是第三方就当真 继续核对证据原句、日期和采编方法。

网址多就当证人多 追溯转载内容的原始来源，数清真正的起点有几个。

公共机构就等于背书 去核对正文到底评没评价这个服务主体。

课后想一想

同一篇新闻稿出现在五个网站上。还需要什么证据，才能把它们称作“五次独立确认”？

5 把答案拆成主张，逐条核验证据

P05

90 分钟

上完这节课，你将能够：把答案拆成一条条可以单独核查的主张，并依据原文逐条判断支持关系。

下课时你要交上来：八条主张核验、三条事实纠正，以及一条“证据不足”的示例。

这一课是全套方法的核心。先看一个例子：A03 那条回答很短，却一口气做了两个承诺——“30 分钟完成”和“上门取车”。这两个承诺需要分别核验，因为它们背后的事实依据完全不同。

你要检查五条成功回答中的八个主张（编号 C01–C08），并对每一条说清楚：为什么标“支持”、为什么标“矛盾”、或者为什么标“证据不足”。

为什么这么做

三个标签的定义请记牢：证据在原条件下能证明完整主张，标“支持”；相关证据与说法相冲突，标“矛盾”；现有材料根本定不下来，标“不足”。还有个容易忽略的细节：同一句话里的两个分句，完全可能一个得到支持、一个构成矛盾——所以拆分要拆到底。

八条主张和它们的参考判定先放在这里，你的任务不是照抄，而是自己打开来源，把每一条重新走一遍：

编号	回答里的说法	证据怎么说	判定
C01	换内胎含材料 33 元（A01）	s02: 15 + 18 = 33, 数字对得上。	支持
C02	40 元预算够（A01）	s02 同意，但前提是“没有额外维修项目”。	支持（带条件）
C03	周日正常营业（A02）	s01 现行说明写明周日休息；A02 引的 s05 是旧帖。	矛盾
C04	保证 30 分钟完成（A03）	s02: 通常 30–60 分钟，取决于排队、型号、库存，明确不是保证时限。	矛盾
C05	提供上门取车（A03）	s01: 不提供上门取送车。	矛盾
C06	位于东门内（A04）	s01 的地点说明一致。	支持
C07	不提供上门取送车（A04）	s01 原文一致。	支持
C08	出发前检查刹车、车铃与轮胎（A05）	s06 安全须知的正文。	支持

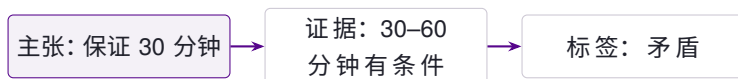


图6 核验的对象永远是“包含限定条件的完整说法”，不是半句话。

今天要用到的材料

打开工作簿的 claims 表、answers.html 和 s01-s06 六份信源。P03 保存的回答原文保持不动。较长的证据摘录和纠错内容写在 templates/05-claim-check.md 里。

我们一步一步来

1. 在 A01-A05 里逐条标出事实性说法。判断要不要拆的标准：连接词两侧可能一真一假，就拆开。至少完成给定的 C01-C08 八条主张。
2. 逐条打开相关来源去核对——不要偷懒用答案里显示的引用标签代替自己查证。摘录“既有证据、又带必要条件”的最小原文片段。
3. 数字和单位要动手算。C01 是 $15 + 18 = 33$ （元）；C02 要保留“没有额外维修项目”这个前提。数字对但前提丢了，照样会误导人。
4. C03 要看日期：A02 说“周日正常营业”，但它引的是论坛旧帖（s05）；现行官方页 s01 写明周日休息。所以这条主张判“矛盾”，同时保留历史记录不删。
5. 把 A03 的两个承诺分开检查：C04 “保证 30 分钟完成”与官方“通常 30-60 分钟、非保证时限”冲突；C05 “提供上门取车”与“不提供上门取送车”冲突。各写出一句准确的修订句。
6. 最后自选一条测试主张：“松风全校最好”。把六份来源全部检索一遍——没有任何比较评价材料，所以判“证据不足”。切记：缺证据不等于反面成立。

P05 · 举例子与验收

举个例子

主张	标签	依据
C03: 周日营业	矛盾	s01 写明周日休息；s05 的信息太旧。
C04: 保证 30 分钟完成	矛盾	s02 给的是有条件的 30–60 分钟区间。
“全校最好”	不足	没有比较材料，也没有排名方法。

做成什么样算合格

按案例的现行政策日期：八个给定主张里五条“支持”、三条“矛盾”；附加的“全校最好”证据不足。这是可复查的课堂答案标准，不是任何商业 AI 平台的准确率。好的纠错句有个共同点：把措辞收窄到证据真正能覆盖的范围。

交什么、怎么评分

八条主张核验、三条事实纠正，以及一条证据不足示例。

评分维度	分值	评分时看什么
最小主张	20	八条独立主张都能关联回原回答。
证据摘录	35	原句相关，限定条件没有丢。
标签与纠错	30	三处矛盾有修正句，没有添新承诺。
不确定性	15	“最好”保持“不足”，并写明依据。

容易踩的坑

一整段只打一个标签 拆开主张，让每段话里不同的支持关系各自现形。

信源权威就支持一切 权威也要落到具体句子和适用范围上核对。

找不到证据就判假 没有相关反证时，正确的标签是“不足”。

课后想一想

来源写的是“7 日内免费复检”。它能支持“7 日内任何损坏零件免费更换”吗？

6 在事实不变的前提下改写 FAQ

上完这节课，你将能够：写出直接有用的回答，同时让价格、条件和来源保持可核查。

下课时你要交上来：一份四问 FAQ、一张事实锁定核对表，以及一份修改前后对照记录。

查清了证据，该动手改善内容了。你面前是一份旧 FAQ：满篇笼统的赞美——什么“服务贴心”“品质保障”——价格却只写了一半。你要把它改写成一份真正能帮用户安排到店的页面：问什么答什么，价格完整，条件清楚，来源可查。

先把这一课的边界说在前面：改写改善的是“可靠信息的可获取性”，它不能证明搜索引擎一定会检索或引用你的新版页面。哪些事做到了、哪些事还只是希望，我们从头到尾都会分清楚。

为什么这么做

有用的服务内容有个共同套路：先直接回答用户的真实问题，再交代条件和来源。你可以放心改措辞、改结构，但底层事实一条都不能动——这就是“事实锁定”。另外，独立报道只能印证它实际核实过的事实，不能加工成本不存在的背书。



措辞可以改；事实承诺不能变。

图7 改写改变的是信息的呈现方式；它本身不证明 AI 可见性提升。

今天要用到的材料

打开 `faq-before.html` 和八条已核对的事实，模板是 `templates/06-faq-draft.md`。先写完你自己的初稿，再打开 `faq-after.html` 对照自查——顺序别反了，先看参考稿会困住你的思路。

我们一步一步来

1. 把 V0 里每句事实和每句宣传话都标出来，对应到事实编号，或者标“缺支持”。特别留意这几处：“随时来”“总共 15 元”“30 分钟”“学校推荐”。
2. 从固定问题库里挑四个真实问题：预算、周日到店、地点与取送、维修时长。用问题本身做标题，第一句话就给出答案——用户是来办事的，不是来读散文的。
3. 回答后面紧跟必要条件：15 元工时费和 18 元内胎费要同时出现；时长说明要保留排队、型号、库存这些限制。
4. 每个回答旁边标注来源编号，让读者知道“这话是谁说的”。再写明维护角色、核对日期和更新规则。不虚构个人资质，不编专家姓名。
5. 在变更日志里把“删掉的句子—新句子—事实编号”三列对齐。如果你借助 AI 起草，只把虚构的已核事实喂给它，并且对它产出的每句话逐项检查。
6. 通读自己的稿子，再和参考修订稿比较。请同伴只看你的页面回答那四个问题。保存 V1 和事实核对记录，V0 也留着。

P06 · 举例子与验收

举个例子

改法	写成什么样	为什么
旧句	“换内胎只要 15 元。”	漏掉了材料费。
新句	“工时 15 元，内胎 18 元，常规更换合计 33 元。”	对应 s02，不添加虚构优惠。
旧句	“学校权威推荐。”	删掉——没有任何来源证明这项背书。

做成什么样算合格

V1 有四个问题、四段直接回答、相邻的来源、核对日期和维护角色；全部事实与 F01–F08 一致；不含虚构的评分、排名、评价或保证。参考文件只是一种合格写法——意思准确的前提下，用你自己的自然表达完全没问题。

交什么、怎么评分

四问 FAQ、事实锁定核对表，以及修改前后对照记录。

评分维度	分值	评分时看什么
事实忠实	40	数字、营业时间、服务范围和条件与底稿一致。
读者可用	25	四个问题都用平实语言直接回答了。
来源与维护	20	来源就在回答旁边；日期、角色、更新规则齐全。
变更记录	15	V0 保留着；每处重要修改都有说明。

容易踩的坑

加上漂亮的统计数 只用实际测得且有来源的数字；没有就删掉。

把条件藏进远处脚注 限定条件放在受它影响的回答旁边。

AI 写得流畅就算核对完 重开来源，逐条检查事实、单位和例外。

课后想一想

这一课能确认 V1 的哪些性质？哪些结果还必须另行观察才能知道？

7 检查网页，并说明检查边界

P07

90 分钟

上完这节课，你将能够：发现页面可见内容与链接的缺陷，并分清本地检查与真实抓取测试的边界。

下课时你要交上来：八项页面检查、一份按优先级排好的修复建议，以及一次复查。

内容改好了，这一课来当“挑刺的测试员”。页面在浏览器里能打开，不代表正文没写错；链接显示在那儿，不代表点得开；元信息看着齐，也可能指向别的地方。我们准备了一个故意埋了各种故障的页面，让你在不动真实网站的情况下把这些坑看个遍。

每项检查都要写下你看到了什么——包括本地根本查不了的内容。是的，“查不了”也要写下来，说清边界本身就是检查的一部分。

为什么这么做

这里有三件事必须分开：人能不能读到页面、文档里实际写了什么、远端系统实际处理过什么。本地文件能回答前两件，回答不了第三件——服务器返回什么 HTTP 状态、爬虫有没有来过、引擎有没有收录，这些都需要另外的证据。



本地能查前两项；实际抓取和索引需要别的证据。

图 8 三个检查层次；本离线练习没有测试最右边那一层。

今天要用到的材料

打开 `page-broken.html` 和 `templates/07-page-audit.md`，以 `faq-after.html` 作为内容修复的参考。页面里损坏的图片和失效的链接都是故意埋的练习缺陷。

我们一步一步来

1. 先看浏览器标签页的标题和页面可见标题。标签页上只有“测试页面”四个字——用户在浏览器标签里搜索历史时，这四个字什么也告诉不了他。再把页面上的“15 元”和“每天开放”分别与 F02、F03 对照：换内胎光工时就要 15 元、材料另计 18 元；营业时间实际是周一至周六、周日休息——“每天开放”是错的。
2. 把关键段落复制进文本编辑器，记录正文是否可读。点开“维修细则”，看那张缺失的图片。先把“发生了什么”如实描述下来，不要悄悄把原始证据修掉。
3. 用浏览器的“查看网页源代码”，搜索 `title`、`robots`、`canonical` 三个词。不要求你写代码——把相关行存成文本或截一张清晰的图就行。
4. 观察源码里的 `noindex` 指令，和指向 `example.invalid` 的 `canonical`。这些是文档里的设置，但它们不能证明某个真实引擎已经排除过这个本地文件。`example.invalid` 这个域名是故意不可用的。
5. 按“用户受影响程度”排修复顺序：先改价格与营业时间的错误，再修详情链接（它指向的 `not-here.html` 根本不存在，应改回 `s02-prices.html`），最后完善标题和元信息。每项写清具体动作和验证方法。图片好不好看，排在事实对不对之后。
6. 复查 `faq-after.html` 的改进：价格、条件、来源链接和日期。检查表里明确写上“服务器状态、真实抓取与索引尚未测试”——不要把整个网站笼统地判为“通过”。

P07 · 举例子与验收

举个例子

发现的问题	记到什么	怎么修
价格不完整	“换内胎：15 元”	补上工时、材料和常规合计 33 元。
细则链接失效	指向 not-here.html	改成 s02-prices.html，改完再点开确认。
本地元信息	noindex；canonical 指向 example.invalid	如实记录设置，不推断线上索引结果。

做成什么样算合格

检查表清楚区分“已观察到的缺陷”和“未测试的服务器行为”。应该找出：价格不完整、“每天营业”的错误、图片与链接失效、标题笼统、缺更新时间、元信息问题。能在本地复查 FAQ 的内容修复，但不能据此宣称真实平台会有什么反应。

交什么、怎么评分

八项页面检查、按优先级排列的修复建议，以及一次复查。

评分维度	分值	评分时看什么
可见检查	30	价格、时间、正文、图片、链接和日期都核对了。
源码检查	25	标题和元信息找到了，证据可读。
修复建议	25	动作具体，按用户影响排序。
边界与复查	20	本地内容复查过；服务器/抓取/索引如实标注未测试。

容易踩的坑

能打开就等于已被收录 你只能报告“本地文件可以渲染”。

noindex 证明曾被排除 这条指令是设置证据，不是历史抓取事件的证据。

源码截图字太小 只截相关行，或粘贴可复制文本并注明出自哪个文件。

课后想一想

双击打开一个本地 HTML 文件之后，你能报告“HTTP 200，已被 AI 搜索收录”吗？

8 用表格计算重复观察指标

上完这节课，你将能够：用正确的分母计算指标，如实报告波动，并且不把差异解释成因果。

下课时你要交上来：一份核对后的汇总表、一张两阶段比较图，以及一段结果边界说明。

这一课我们练测量。工作簿的 `measurements` 表里有 36 条合成记录：六个问题、每个重复观察三次、分前后两个阶段；每个阶段各有一次失败。你的任务是用筛选和公式亲手复算一遍汇总结果，然后准确地讲清楚：这份数据里到底发生了什么变化。

为什么这么做

先记一个定义：**提及率** = 提到服务站的有效回答数 ÷ 有效回答总数，失败要单独报。测的是前六个固定问题（Q01 刹车检查、Q02 预算、Q03 周日营业、Q04 周六关门、Q05 位置、Q06 上门取车），每个问题重复观察三次、分前后两个阶段。再看一眼你算出来的“上升 17.6 个百分点”——它描述的是当前这份样本，不能证明是 FAQ 改稿让真实平台发生了变化。把“描述样本”和“证明因果”分开，是这一课最大的收获。

今天要用到的材料

打开工作簿的 `measurements` 和 `summary` 两张表，模板是 `templates/08-measurement-report.md`。注意：表里的“前/后”是合成数据里的教学分组，不是一次真实干预实验的采集时间。

我们一步一步来

1. 先对行数：前后各 18 条。检查每个阶段有六个问题编号、每个编号三轮。筛选 `status=failed`，确认每阶段恰好一条失败。
2. 在“前”阶段筛出 `status=ok`，数一数 `mention=1` 的有几条：17 条完整回答、8 条提及。“后”阶段：17 条完整回答、11 条提及。
3. 打开 `summary` 表看公式。`COUNTIFS` 先按阶段与状态计数，再加 `mention=1` 条件算分子，最后除以有效回答数。原始行保持 1/0 数值，失败留空。
4. 动手算 8/17 和 11/17，保留一位小数的百分比，做差时用未四舍五入的数值。你会得到：提及率从 47.1% 到 64.7%，上升 17.6 个百分点。
5. 再复算有引用率：两个阶段都是 7/17，即 41.2%。有意思的地方来了——提及率变了，引用率纹丝不动；而且“有引用”依然不等于“证据支持”。
6. 看每个问题的三次重复，不要只挑最好的那次。保存汇总结果，并写明三条限制：样本小且是合成数据、失败没有进入结果分母、不能推出真实平台结论或因果结论。

P08 · 举例子与验收

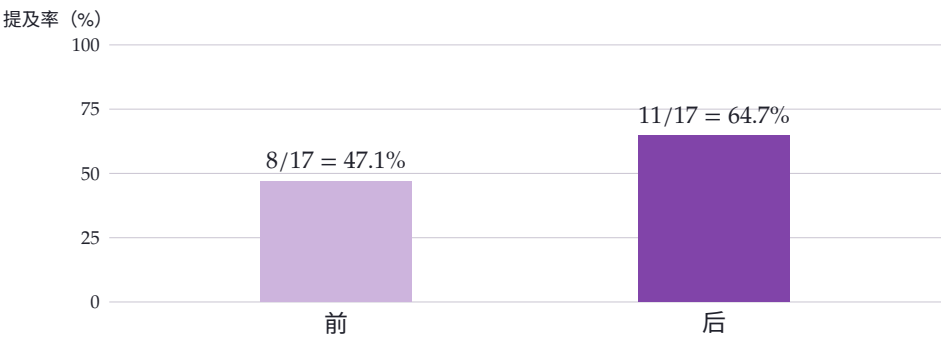


图9 合成课堂数据。每阶段计划 18 次、失败 1 次，结果分母为 17。这张图不是 FAQ 改稿效果的实际测量。

举个例子

阶段	算出什么	备注
前阶段	8 / 17 = 47.1% 提及率	计划 18 次，失败 1 次。
后阶段	11 / 17 = 64.7% 提及率	计划 18 次，失败 1 次。
差异	17.6 个百分点	有引用率保持 7 / 17 = 41.2% 没动。

做成什么样算合格

每阶段计划记录 18 条、有效回答 17 条、失败 1 条；提及数分别为 8 和 11，有引用数都是 7，缺失结果共两条。每个分子都能通过筛选原始行复算出来。中英文两个版本使用相同的编号和数值。

交什么、怎么评分

核对后的汇总表、两阶段比较图，以及结果边界说明。

评分维度	分值	评分时看什么
计数与缺失	30	36 条记录、两次失败，分阶段的分母都对。
公式	30	按阶段和状态计数；输入变了结果会跟着变。
结果表达	25	计数、比例、百分点差异同时报告。
解释	15	写明合成样例范围，承认没有因果证据。

容易踩的坑

直接除以 18 次尝试 本任务统一除以 17 条有效回答，失败单独报。

把 17.6 个百分点说成“增长 17.6%” 单位不一样：百分点是两个百分比相减。

只报告最好的一次 保留全部计划内的重复，把波动摆出来。

课后想一想

一次失败后来重试成功了。能把原始那行覆盖掉、悄悄改变分母吗？

9 同伴复核证据，安排改进优先级

P09

90 分钟

上完这节课，你将能够：用来源原句解决标注分歧，并把发现转化成具体的行动清单。

下课时你要交上来：复核记录、简单一致率计数，以及一份一周改进清单。

两个人读同一条答案，判出不同标签，太正常了——这恰恰是有价值的信息。这一课你和同伴要比较四条容易起争议的主张，回到原文解决分歧，然后决定：查出来的问题，先修哪个。

为什么这么做

两个概念先分开：一致率只说明两个人应用规则的相近程度，不自动说明标得对。排优先级看的是“能避免用户什么问题、要花多少功夫”，而不是“承诺能多几个提及”。举例来说，营业时间写错为什么重要？因为它可能让用户白跑一趟——用户成本很高的错，永远排前面。



图 10 讨论之前先保存初始判断；不能用讨论后的一致替换初始一致率。

今天要用到的材料

用 P05 的标签、templates/09-peer-review.md 和来源页面。单人学习的同学有替代方案：先遮住原标签，去做另一个任务，回头再读一遍重新标注——并在记录里明确写“自查”，不要冒充独立复核。

我们一步一步来

1. 选 C02、C03、C04、C07 四条。只把主张和来源文件交给同伴，**不给你自己的标签**。双方各自记录：支持/矛盾/不足，加一段证据原句。
2. 两套标签都存好了，再放在一起比较，数一数四条里几条完全一致。就算讨论后达成了一致，最初的标签也原样保留——那是你的数据。
3. 每处分歧写明卡在了哪条规则上：是时间范围、限定措辞、信源独立性，还是主张没拆干净。然后重开来源，把解决分歧的那句原话记下来。
4. 材料实在定不下来的，就保留“未解决”状态，写清还需要什么证据。不要为了一致率好看而强行给正向标签。
5. 列出三个一周内可完成的动作，每个都填上：证据编号、影响哪些用户、谁负责、期限、怎么复查。错误的价格、营业时间和虚构承诺，优先于一切装饰性修改。
6. 请复核人检查一项已完成的修订，填写复查记录。边界说明别忘了：本地复核通过，不代表平台已经处理了新内容。

P09 · 举例子与验收

举个例子

情形	记录成什么	说明
初始标签	四条中三条一致	讨论前的一致率 75%。
分歧点	“40 元预算够”	核查是否保留了“无额外维修”条件。
具体行动	纠正 FAQ 的周日营业说明	负责人：编辑；对照 F03 与 s01 验证。

做成什么样算合格

复核表里同时保留双方初始判断和讨论后的标签。3/4 只是示例数字——你要报告自己组的实际计数。改进清单是三个小而可查的动作；有未解决的事项没关系，把证据缺口说清楚就行。

交什么、怎么评分

复核记录、简单一致率计数，以及一周改进清单。

评分维度	分值	评分时看什么
独立初评	25	比较前保存了标签；单人路线写明了“自查”。
分歧处理	35	每处分歧对应到规则和证据原句。
行动清单	25	三个动作关联证据，责任人和验收方式清楚。
如实报告	15	初始一致率和未解决项都保留了。

容易踩的坑

先把自己的答案告诉对方 遮住标签重做一遍，把新的过程记录下来。

一致就当正确 继续对照原文和规则，核验两边的判断。

只写“优化 GEO”四个字 说清哪句话、依据何在、谁负责、怎样算完成。

课后想一想

两位复核人都把一条无依据的排名标成了“支持”。100% 的一致率能证明什么？

10 交付一个别人能重新检查的项目

P10

90 分钟

上完这节课，你将能够：把有边界的结论清清楚楚地呈现出来，并让支持材料随手可查。

下课时你要交上来：一页报告、一个带索引的证据 ZIP，以及一次三分钟演示。

最后一课，我们把前面九次课的成果组装成一个像样的交付。合格的项目不只是一张漂亮的截图——复核的人应该能顺着你的结论，一路找到问题、回答、主张和来源。整理交付的目的，就是让别人不靠你口头解释，也能亲手检查你的工作。

为什么这么做

报告里的每个结论都要有可追溯的证据路径。**已观察到的问题、已完成的修改、未来的计划**，三类内容分开写，不要搅在一起。还要再强调一次：那份前后对比的合成数据支持的是计算练习，不是什么真实服务商或 AI 平台的成功案例。



图 11 复核人应该能沿着证据路径正向走，也能反向走。

今天要用到的材料

用上 P01-P09 的全部成果和 templates/10-one-page-report.md，新建 final 文件夹。完整的课堂素材包放在提交包之外——只复制追溯结论所必需的来源文件进来。

我们一步一步来

1. 第一段先写范围：虚构服务站、素材版本、十二条固定问题、六条给定观察、五条完整回答、一次失败、六份信源；36 行测量练习另行说明。先把“这是什么、不是什么”讲明白。
2. 选三个发现，每个都给出回答编号、相关主张编号、信源编号和文件名。示范一句：“A02 称周日营业，C03 与 s01 的当前营业说明冲突。”
3. 说明一项已完成的内容修订和一张结果表。8/17、11/17 按合成数据报告，并明确写出：这不能证明改稿导致真实提及增加。
4. 制作证据索引，把必要的原件、标注、FAQ V0/V1、工作簿和报告复制进 final。清掉重复草稿，以及文件名或截图里意外带出的个人信息。
5. 把 ZIP 交给同伴，请对方只凭索引重新检查两条结论，缺文件、断链接当场修。这个环节叫“同伴交接检查”——别夸大成“外部审计”。
6. 准备三分钟演示：30 秒讲用户问题，60 秒展示证据，60 秒展示修订与计算，30 秒讲限制和下一步。讲完做下面的迁移练习。

P10 · 举例子与验收

举个例子

内容	写成什么样	证据路径
发现	周日营业信息错误	Q03 → A02 → C03 → s01 → s01-service.html。
已完成改进边界	FAQ 明确写周日休息 未测试真实平台效果	F03 → V1 → 来源链接 → 本地复查。 只完成了合成数据计算与本地页面检查。

做成什么样算合格

交付目录能正常打开；主要发现都有证据路径；报告把观察和计划分开写。演示用平实的语言，并且至少讲一条限制。交接记录写同伴实际打开核对了什么——不夸大成超出这个过程的独立验证。

交什么、怎么评分

一页报告、带索引的证据 ZIP，以及三分钟演示。

评分维度	分值	评分时看什么
证据路径	35	三个发现都能回溯到记录与来源。
报告与边界	25	范围、计数、已完成修改和限制都写清楚了。
交付可用	25	索引和文件可读；同伴成功重开两条结论。
演示	15	三分钟内讲清一个具体的用户决策。

容易踩的坑

结论没有记录编号 补上证据路径；补不了就删掉那条结论。

把合成数据写成客户成果 每处展示结果的地方都标明“教学样例”。

只交最终页面 V0、V1 和变更记录一起交，方便复查。

课后想一想

迁移练习：校园打印店门口写着“当天装订”。请拟一个中性问题、两条待核事实，以及回答中必须保留的一项条件。

表格快速上手

A 使用练习工作簿

工作簿一共八张表。中英文两个版本用相同的工作表名——所以“打开 claims”在哪版里都指向同一个任务。淡黄色的格子是留给你填的；原始测量行是给定样例，请保持原样。动手前记得先另存一份自己的副本。

工作表	是干什么的	你要做什么
project	项目卡	填小组、用户、问题和范围。
facts	八条已核对服务事实	阅读并追溯来源与条件。
queries	十六条原始问题	分类、清理，映射到十二条固定问题。
sources	六个来源文档	补充类型、独立性及可支持的主张。
observations	六条问答记录	填实际时间、标签和原文文件名。
claims	八条最小主张	填证据原句、来源、标签和条件。
measurements	三十六条合成重复记录	筛选计数，保留失败。
summary	公式驱动的分阶段汇总	查看公式并解释指标。

需要掌握的五个操作

1. 单击单元格直接编辑。编号格子里不要换行；每完成一步就保存一次。
2. 冻结首行，滚动时表头才不会消失。给定的数据表已经设置好了。
3. 选中表头，在“数据”菜单里打开筛选。先筛阶段、再筛状态；核对总行数前记得清除全部筛选。
4. 单击公式单元格，到公式栏里看它怎么算的。不要把公式结果改成手打的“47.1%”。
5. 文字显示不全时，加宽列或打开自动换行；数据区里不要合并单元格。

表格公式

看懂一次计算

summary 表第 2 行是前阶段，第 3 行是后阶段。measurements 表里，B 列是阶段、E 列是状态、F 列是提及、G 列是有引用；第 2–37 行就是那 36 条记录。公式使用明确的范围，避免把表头和无关格子卷进统计。

```
successful = COUNTIFS(stage, selected_stage, status, "ok")

mentions = COUNTIFS(stage, selected_stage, status, "ok", mention, 1)

mention_rate = mentions / successful
```

实际的 E2 公式是 =IF(C2=0,"",D2/C2)。没有有效回答时比例留空——而不是制造一个看似合理的零。COUNTIFS 用的是 A2/A3 里的阶段文字，所以阶段名称翻译之后，公式结构完全不变。

统计量	前	后
计划次数	18	18
有效回答	17	17
失败次数	1	1
提及数	8	11
提及率	47.1%	64.7%
有引用回答数	7	7
有引用率	41.2%	41.2%

动手复算一次

在一份一次性副本里，把前阶段某条有效回答的提及从 0 改成 1。你会看到提及数变成 9、提及率变成 9/17 = 52.9%，而计划数和有效数纹丝不动。然后撤销修改，确认回到 8/17。注意：这个实验别在保留的原件上做。

两个不同的问题

“完整回答中多大比例提到服务站”——分母是有效回答；“计划尝试中多大比例获得提及”——分母是全部尝试。两个都是合理的指标，但比较过程中不能偷换分母。本讲义统一用前者，并单独报告失败。

数据与文件约定

B 编号、标签与提交文件

编号或字段	是什么	例子
F	已核对的服务事实	F02: 工时与材料价格
R / Q	原始问题 / 固定问题	R03 对应 Q02
A	一条给定回答观察	A02: 周日营业
C	一条可核查主张	C03: 周日开放
s	一个来源文档	s01: 现行服务说明
status	是否获得完整回答	ok 或 failed
mention / citation	观察到的二元结果	出现 1; 未出现 0; 失败留空
来源日期 / observed_at	来源适用时间 / 实际查看时间	分开记录

固定问题对照表

固定编号	来自原始	对应决策
Q01	R01	东门附近检查刹车
Q02	R03	40 元预算是否含材料足够
Q03	R05	周日是否营业
Q04	R07	周六几点关门
Q05	R08	第一次到店的位置
Q06	R09	能否上门取车
Q07	R10	预约是否已确认
Q08	R11	是否保证完成时限
Q09	R12	两天后能否免费复检
Q10	R13	是否学校官方背书
Q11	R14	校园报探访能证明什么
Q12	R15	骑行前检查

能够还原的回答记录

GEO PRACTICE LAB · 离线教学材料



图 12 answers.html 的实际页面截图节选。网页明确标注为离线编写样例；问题、答案和显示的来源链接分别保留为证据字段。

最终目录里至少要有这些

report.md、evidence-index.md、小组工作簿、query-set-v1、六条观察记录、八条主张核验、source-map.md、FAQ V0/V1、page-audit.md、peer-review.md,以及报告用到的来源页。文件名要简短稳定,比如 A02-original.md、C03-check.md。索引要说明每个文件在哪里、支持什么结论。

还有一条重要的诚信规则：不要把教学答案抄进报告却宣称是自己的发现。用了示例就如实说明，并展示你自己实际完成的那部分核验。如果你另外做了真实观察，请放到单独的目录里，单独记录操作标准和时间。

参考答案

C 课后想一想：参考答案

P01

打开 s02：常规换内胎、工时 15 元、材料 18 元、合计 33 元；同时保留生效日期和“无额外维修”的限定。

P02

不应合并。两条问题涉及不同的日子，对应不同的到店决策。

P03

可以同时成立。“有引用”描述的是屏幕上有没有链接；“证据支持”描述的是来源能不能证明那条具体主张——两回事。

P04

需要证明每家媒体都独立核验过相关主张；仅仅是出现在不同域名上，不构成这种证据。

P05

不能。“复检”和“更换”是不同的服务；无条件更换没有得到支持，而且 s02 明确限定了政策范围。

P06

能确认的是：事实一致、表达清晰、来源可追溯。被发现、被索引、被引用以及用户端结果，都需要另外的证据。

P07

不能。本地文件没有提供经过测试的 HTTP 响应；本地显示成功也证明不了抓取、索引或引用。

P08

不能覆盖。原始尝试保留，重试单独编号、按明确规则处理，汇总透明更新。

P09

只能证明两人的标签相同——不能证明排名真实，因为两人可能都误用了规则。

P10

示例：“16 点提交 80 页文件，今天能取到装订件吗？”核对截单时间和支持的装订/纸张规格，并保留店方说明的库存、排队或取件时间条件。

D 我们为什么这样上课

这套课程设计参考了七项公开教学资料 [3–9]、随附的职业培训例题，以及本项目既有的实战课程设计。Google 的初学者任务模块启发了“每课产出一个成果”的节奏；Elements of AI 证明了无需编程的路线可行；Microsoft Learn 提供了目标明确、练习自查的结构。Data Carpentry 教我们把表格的脏数据、显式缺失、一行一条观察做好；SIFT 与 Cornell 提供了追溯发布者、核查具体主张的提问框架；MANTRA 启发了命名、版本和交付的规范。

深职大的例题采用九十分钟任务、明确输入、截图交卷、分项评分。本讲义吸收了它的操作粒度，但没有引入模型训练的前置要求，也没有复用其代码。十次课是我们基于“成果相互衔接”作出的设计选择——它不是经过验证的最优次数，也不是照搬哪家的课程表。

课程怎么打分

组成部分	建议权重	怎么算
P01–P08 任务档案	60%	按各任务评分表打分，八次取平均。
P09 同伴复核	15%	看独立标签、分歧处理和具体行动。
P10 交付与演示	25%	看结论可追溯、包可读、边界清楚。

允许修正后重交：补回缺失来源、撤回无依据说法之后，可以重新提交。但有一条底线——高分不能抵消虚构数据；先修复来源问题，再重评相关任务。本课程不颁发任何机构认证，也不承诺商业结果。

想做真实观察的同学看这里

如果你想在课外做真实观察：选择你有权查看的低风险公开服务页，通过合法可访问的界面，围绕一小组固定的中性问题进行人工观察。记录平台、可见的模型或模式、语言、地区、登录状态、实际时间和计划重复次数；保留完整回答与失败记录。只用允许分享的材料，报告里不含个人信息。费用与发布是另外的决定，不是课程必做步骤。真实数据只按自身口径比较，不能和合成样例混在一起算。

参考资料

E 参考文献与图像说明

- [1] Aggarwal et al.. GEO: Generative Engine Optimization. KDD 2024. arXiv:2311.09735v3. <https://arxiv.org/abs/2311.09735>.
- [2] Holderrieth and Erives. An Introduction to Flow Matching and Diffusion Models. arXiv:2506.02070v3. Typographic reference; not a source of GEO claims. <https://arxiv.org/abs/2506.02070>.
- [3] Google. Google AI Essentials. Public curriculum and activity descriptions. <https://grow.google/ai-essentials/>.
- [4] University of Helsinki and MinnaLearn. Elements of AI. Public description of the introduction and separate Building AI course. <https://www.elementsofai.com/>.
- [5] Microsoft Learn. Introduction to generative AI and agents; Exercise: Explore generative AI. <https://learn.microsoft.com/en-us/training/modules/fundamentals-generative-ai/>.
- [6] Data Carpentry. Data Organization in Spreadsheets for Ecologists. Summary/setup and Formatting problems. <https://datacarpentry.github.io/spreadsheet-ecology-lesson/>.
- [7] Mike Caulfield. SIFT (The Four Moves), June 19, 2019. <https://hapgood.us/2019/06/19/sift-the-four-moves/>.
- [8] Cornell University Library. Evaluating Web Pages: Questions to Consider. Categories, Accuracy and Authority. https://guides.library.cornell.edu/evaluating_Web_pages.
- [9] University of Edinburgh. Research Data MANTRA. Learning units and data-handling tutorials. <https://mantra.ed.ac.uk/>.

公开教学资料 [3–9] 读取于 2026 年 9 月 8 日，课程结构可能更新。这些引用用于说明设计依据，不声称某种教学形式已由实验证明优于其他形式。

图像来源与复现

图像来源

图 3 依据 Creative Commons Attribution 4.0 转载 GEO 论文原图。矢量裁切图取自项目保留的 arXiv v3 来源,与已核对的文献资产相同;原标注、箭头和颜色均保留。许可:<https://creativecommons.org/licenses/by/4.0/>。

其余流程图为原创矢量教学图。柱形图使用三十六行合成测量数据。截图来自学生素材包中实际打开的本地页面,不是商业 AI 产品截图;页面日期是虚构的数据字段。深职大附件中的照片和图示未再分发。

排版与语言

版式参照所提供的学术讲义:克制的细线、衬线正文、无衬线标题、编号图表、来源图注和连续阅读结构。色彩上只用少量紫色做强调。本课程为独立编写,并非 arXiv 投稿,也不代表任何院校背书。

英文术语	中文术语	在这本讲义里的意思
observation	观察记录	一次带条件记录的问答尝试
claim	主张	能够单独核查的说法
evidence	证据	用于核查具体主张的材料
source	信源	信息背后的文档或发布者
independent corroboration	独立印证	对同一相关事实的独立核查
percentage points	百分点	两个百分比数值之间的算术差

源包包含英文主稿、对应中文文本、样式文件、绘图代码、保留的论文图、案例生成脚本和简要质量记录。使用 XeLaTeX 编译,编译时不依赖远程资源。学生 ZIP 只包含学习材料、模板与工作簿。