

# AAAR：AI 加速学术研究的方法论与经验

Zhang Yu

2026-01-20

# Table of contents

Welcome	4
Buy Today!	4
快速导览	4
重点提示	4
关键原则（有序列表）	4
工具与流程（无序列表）	4
表格示例	4
图片示例	5
代码块示例	5
引用与引用框	6
超链接与行内代码	6
公式示例	6
任务列表	6
注脚与脚注引用	6
引用文献与参考文献	6
折叠块	7
分隔线	7
公式编号与引用	7
图注与交叉引用	7
代码与数据	7
<b>1 序章   为什么写这本书</b>	<b>9</b>
1.1 一、从一次失败的文献综述说起	9
1.2 二、为什么现有的“AI 科研指南”让我不满意	9
1.3 三、这本书想要提供什么	10
1.3.1 3.1 一个可持续的思维框架	10
1.3.2 3.2 一套可复现的工作流	10
1.3.3 3.3 必要的批判性	11
1.4 四、这本书写给谁	11
1.4.1 4.1 不同阶段的读者	11
1.4.2 4.2 不同学科的读者	12
1.4.3 4.3 不同技术背景的读者	12
1.5 五、这本书不是什么	12
1.6 六、本书的结构与阅读建议	13
1.6.1 6.1 整体结构	13
1.6.2 6.2 阅读路径建议	13
1.6.3 6.3 每章的结构	14
1.7 七、一些重要的声明	14
1.7.1 7.1 关于时效性	14
1.7.2 7.2 关于局限性	14

1.7.3	7.3 关于 AI 辅助写作	15
1.7.4	7.4 关于争议性观点	15
1.8	八、为什么选择开源	15
1.8.1	8.1 可复现与可验证	15
1.8.2	8.2 持续更新	16
1.8.3	8.3 社区协作	16
1.8.4	8.4 价值观的表达	16
1.9	九、我们正处于什么样的时刻	16
1.9.1	9.1 炒作周期的高峰与低谷	17
1.9.2	9.2 学术界的特殊处境	17
1.9.3	9.3 一场正在发生的范式转移	17
1.9.4	9.4 个人选择的重要性	18
1.10	十、致谢	18
1.11	十一、写在最后	18
2	<b>第 1 章 AI 能做什么，不能做什么</b>	19
3	<b>第 2 章幻觉、偏差、泄露：三类核心风险</b>	20
4	<b>第 3 章提示词不是方法论</b>	21
5	<b>第 4 章信息压缩与文献整理</b>	22
6	<b>第 5 章结构化表达与写作</b>	23
7	<b>第 6 章代码与计算能力</b>	24
8	<b>第 7 章知识管理与协作</b>	25
9	<b>第 8 章选题与研究问题</b>	26
10	<b>第 9 章文献与理论构建</b>	27
11	<b>第 10 章数据获取与构造</b>	28
12	<b>第 11 章数据清洗与分析</b>	29
13	<b>第 12 章写作、投稿与传播</b>	30
14	<b>第 13 章 RA Level：工具层</b>	31
15	<b>第 14 章 Supervisor Level：认知协作层</b>	32
16	<b>第 15 章 Domain Expert Level：推理与建模层</b>	33
17	<b>第 16 章 Agent Level：AI 作为行动者</b>	34
18	<b>第 17 章 Governance Level：制度与治理层</b>	35
19	<b>终章   AI 时代研究者的新能力结构</b>	36
20	<b>附录（可选）</b>	37

# Welcome

这是一个用于展示排版和视觉效果封面页。我在这里补充了多种 Markdown 元素，方便你快速判断整体的排版质感与细节风格。

## Buy Today!

**AAAR: AI 加速学术研究的方法论与经验**

Open access • Print version coming soon.

[Buy from Amazon](#) [Buy from Publisher](#) [Download PDF](#)

## 快速导览

本书面向从本科生到教授的读者。你可以按“能力模块”阅读，也可以按“科研工作流”顺序阅读。

## 重点提示

AI 擅长生成，但不擅长负责。使用 AI 的关键在于验证与责任分配。

## 关键原则（有序列表）

1. 先定义问题，再让 AI 参与。
2. 先验证证据，再相信输出。
3. 先记录流程，再扩展规模。

## 工具与流程（无序列表）

- 文献：综述不是堆摘要，而是证据链。
- 数据：清洗的透明性决定可信度。
- 写作：表达是结果，逻辑是核心。

## 表格示例

模块	典型任务	主要风险
信息压缩	文献综述	幻觉引用
数据处理	清洗/结构化	规则不透明
写作表达	初稿/润色	逻辑空洞

## 图片示例



Figure 1: 封面示例

## 代码块示例

你是我的研究助理。请将下面的研究问题拆解成可检验的子问题，并给出潜在的数据来源与方法路径。输出为列表。

## 引用与引用框

“AI 擅长生成，但不擅长负责。”  
——写作时请把证据链放在第一位。

### i 提示

把 AI 当作“加速器”，不要当作“裁判”。

### ⚠ 警告

不要把未发表的数据或敏感信息直接交给模型。

## 超链接与行内代码

- 官网链接：<https://yuzhangsjtu.github.io/AAAR/>
- 仓库链接：<https://github.com/yuzhangsjtu/AAAR>
- 行内代码示例：使用 `quarto render` 重新生成站点

## 公式示例

这是一条行内公式： $E[X] = \sum_i x_i p_i$ ，以及一个块级公式：

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^N (Y_i^1 - Y_i^0)$$

## 任务列表

- ☒ 完成大纲
- ☒ 生成示例页面
- ☐ 补充案例
- ☐ 进一步润色正文

## 注脚与脚注引用

这是一个脚注示例。<sup>1</sup>

## 引用文献与参考文献

这是一条文献引用示例 (Knuth 1984)，引用会自动生成参考文献列表。

---

<sup>1</sup>这是脚注内容，用于测试排版与字号。

## 折叠块

### 💡 折叠内容

这里是折叠块的内容，用来展示 Anthropic 风格下的折叠区域样式。

## 分隔线

---

## 公式编号与引用

带编号的公式如下：

$$\tau = E[Y(1) - Y(0)] \quad (0.1)$$

在正文中引用该公式：见公式 Equation 0.1。

## 图注与交叉引用

在正文中引用该图：见图 Figure 2。

## 代码与数据

本书所有内容均开源：

- 代码仓库：<https://github.com/yuzhangsjtu/AAAR>
- 在线阅读：<https://yuzhangsjtu.github.io/AAAR/>



Figure 2: 示意图：AI 加速研究的简单流程



# 1 序章 | 为什么写这本书

## 1.1 一、从一次失败的文献综述说起

2023年初的某个深夜，我坐在办公室里盯着屏幕，面前是一份用 ChatGPT 生成的文献综述初稿。那时候，GPT-4 刚刚发布不久，整个学术圈都在讨论“AI 会不会取代研究者”。我也不能免俗，决定亲自试一试。

那篇综述的主题是关于社会网络对信息传播的影响——一个我研究了好几年的领域。我给模型提供了详细的背景、明确的问题、甚至列出了几篇我认为重要的文献，然后让它“帮我写一篇文献综述”。

结果出乎意料地“好”。结构清晰，语言流畅，甚至引用格式都很规范。如果只看表面，这完全可以当作一份合格的课程作业。但当我开始逐条核对引用时，问题出现了：十二条引用中，有四条是完全捏造的——作者名对不上，期刊名不存在，DOI 链接指向 404。更糟糕的是，另外三条虽然文献确实存在，但模型对它们的概括与原文观点完全相反。

那一刻我意识到，这不仅仅是“工具有 bug”的问题。模型并没有“理解”什么是文献综述，它只是在模仿文献综述的形式——用流畅的语言把看起来相关的东西串起来。它不知道证据链是什么，不知道学术引用的意义在于可追溯和可验证，更不知道一条错误的引用可能导致整个论证链条的崩塌。

这次经历让我开始思考一个更深层的问题：AI 究竟能在学术研究中扮演什么角色？它显然可以提高效率，但效率的提高如果建立在可靠性的下降之上，那种“加速”的代价是什么？

## 1.2 二、为什么现有的“AI 科研指南”让我不满意

在那次失败之后，我开始系统地阅读市面上关于“如何在学术研究中使用 AI”的书籍、文章和教程。令我失望的是，绝大多数内容都存在以下几个问题。

**第一，工具堆砌，缺乏方法论。**很多指南的逻辑是：这里有一个 AI 工具，它能做 x，所以你应该用它做 x。这种“功能清单”式的写作完全忽略了一个核心问题——为什么你需要做 x？x 在你的研究流程中扮演什么角色？用 AI 做 x 和用传统方法做 x 有什么本质区别？没有方法论的工具介绍，最多只能培养“熟练的操作员”，而不是“有判断力的研究者”。

**第二，回避风险，夸大效用。**或许是出于推广的目的，很多内容只讲 AI 能做什么，不讲它不能做什么；只讲成功案例，不讲失败教训。幻觉（hallucination）被轻描淡写地提一句，然后就被“所以记得核实”这种空洞的建议带过。但问题是，核实需要成本，而且很多时候你根本不知道哪些地方需要核实。如果一个工具需要你“处处核实”才能安全使用，那它到底是帮你节省时间，还是把风险转嫁给你？

**第三，混淆“更快”与“更好”。**我见过太多这样的句式：“以前需要三天的工作，现在只需要三小时。”这种表述隐含了一个危险的假设——原来三天的工作

和现在三小时的工作是等价的。但真的等价吗？三天的文献阅读和三小时的模型总结，产出的是同样的理解吗？三天的数据清洗和三小时的自动化处理，留下的是同样的审计痕迹吗？效率的提升往往伴随着某些东西的损失，但这些损失很少被讨论。

**第四，对“责任”的回避。**当 AI 参与学术生产，责任如何分配？如果模型生成的内容包含错误，责任在使用者还是开发者？如果审稿人无法分辨 AI 生成的文本，学术诚信的边界在哪里？这些问题极少被正面回应。很多指南的态度是“工具是中性的，看你怎么用”——这当然是正确的废话，但它回避了真正的难题：在实践中，我们如何具体地、可操作地分配责任？

**第五，缺乏批判性反思。**AI 的广泛使用可能对学术生态产生什么长期影响？如果所有人都用同样的模型来“加速”研究，学术多样性会不会下降？如果 AI 让写作变得更容易，会不会有更多的论文但更少的发现？如果模型的训练数据主要来自英文文献，非英语学术传统会不会被进一步边缘化？这些问题几乎没有人讨论。

正是这些不满意，推动我写这本书。我想提供一种不同的视角——不是“AI 能帮你做什么”，而是“你应该如何与 AI 协作，同时保持研究的严谨性”。

## 1.3 三、这本书想要提供什么

基于上述反思，我给这本书设定了三个核心目标。

### 1.3.1 3.1 一个可持续的思维框架

第一个目标是提供一个可持续的思维框架。所谓“可持续”，是指这个框架不会因为某个具体工具的更新而过时。GPT-5 出来了，这个框架依然适用；换一个模型，框架依然成立。

这个框架的核心是三个问题：

1. **AI 在这个任务中的能力边界在哪里？**它擅长什么？不擅长什么？会犯什么类型的错误？
2. **这个任务的质量标准是什么？**我们用什么指标判断任务完成得好不好？这些指标中哪些是 AI 容易满足的，哪些是困难的？
3. **人机协作的最优分工是什么？**哪些环节交给 AI 更高效？哪些环节必须由人类来做？交接点在哪里？如何确保交接不丢失关键信息？

这三个问题不是一次性回答的，而是在每一个具体任务中反复追问的。我希望读者在读完这本书之后，能够在面对任何新工具、新场景时，自动地问出这三个问题，并且有能力给出自己的回答。

### 1.3.2 3.2 一套可复现的工作流

第二个目标是提供一套可复现的工作流。“可复现”是学术研究的基石之一，但在 AI 辅助的研究中，可复现性面临新的挑战。

传统研究的可复现性主要依赖于方法描述的精确性：你告诉别人你用了什么数据、什么模型、什么参数，别人就能重复你的分析。但当 AI 参与之后，问题变得复杂：

- 模型的版本会更新，同样的提示词在不同版本下可能产出完全不同的结果。
- 模型的输出有随机性，即使版本相同、提示词相同，两次运行的结果也可能不一样。
- 很多 AI 工具是黑箱，你无法精确描述它内部做了什么。

面对这些挑战，我在本书中提出的 workflow 强调三个原则：

1. **记录一切**。不仅记录结果，还记录过程：你用了什么提示词？模型返回了什么？你做了哪些筛选和修改？为什么做这些修改？
2. **保留原始数据**。AI 处理过的数据应该与原始数据分开存放，确保任何时候都能回溯到 AI 介入之前的状态。
3. **人工复核关键节点**。识别 workflow 中的“高风险节点”——一旦出错，代价很高的环节——并在这些节点强制进行人工复核。

这套 workflow 会贯穿全书，在每一章具体任务的讨论中都会体现。

### 1.3.3 3.3 必要的批判性

第三个目标是保持必要的批判性。这不是说我要“反对 AI”——恰恰相反，我是 AI 的重度用户，我的日常研究工作中大量使用各种 AI 工具。但正因为我是用户，我更清楚它的问题在哪里。

批判性体现在几个层面：

**对工具本身的批判**。大语言模型不是魔法，它有非常具体的能力边界和失败模式。理解这些边界和模式，是安全使用的前提。

**对使用方式的批判**。很多“最佳实践”其实是未经检验的假设。比如“提示词越详细越好”——真的吗？在什么条件下这个结论成立？有没有反例？

**对系统效应的批判**。当一项技术被广泛采用，它会改变整个系统的激励结构。AI 让写作更容易，但这会导致更多好研究，还是更多坏研究？投稿数量上升，审稿负担加重，审稿质量会不会下降？这些系统层面的问题，虽然个体研究者难以改变，但至少应该意识到。

## 1.4 四、这本书写给谁

在动笔之前，我认真思考过目标读者是谁。最终的定位是：**所有在学术工作中与 AI 打交道的人**。

这个范围听起来很宽，但实际上是经过取舍的。

### 1.4.1 4.1 不同阶段的读者

本科生和硕士研究生可能是最直接的受益者。你们正处于学术训练的关键阶段，既要学习传统的研究方法，又要适应新工具的冲击。这本书希望帮你建立

一个平衡的视角：AI 是有用的工具，但学术能力的核心——提出好问题、设计好研究、做出好判断——不会被工具取代。

**博士研究生和青年研究者**面临的挑战更具体：你们需要在有限的时间内产出高质量的研究，同时还要应对发表压力、资金压力、职业压力。AI 看起来像是缓解压力的捷径，但如果使用不当，它可能成为新的陷阱——你可能产出更多，但质量下降；你可能写得更快，但想得更浅。这本书希望帮你找到真正能提高研究质量的使用方式，而不是只追求表面的效率。

**成熟的研究者和教授**可能对 AI 工具本身已经有了自己的判断，但你们面临另一个问题：如何指导学生？如何制定团队的使用规范？如何评估 AI 辅助产出的研究成果？这本书的“五层次框架”和“团队协作”章节可能对你们特别有用。

**研究团队的管理者**——无论是实验室主任、项目负责人还是研究机构的领导——需要从更宏观的层面考虑 AI 的影响：如何制定政策？如何分配责任？如何平衡效率与风险？本书后半部分的讨论会涉及这些议题。

## 1.4.2 4.2 不同学科的读者

虽然我的主场是社会科学，但本书并不局限于社科读者。

**社会科学研究者**会发现这本书最贴近你的工作场景。从选题到文献综述，从数据收集到统计分析，从写作到投稿，每个环节都有具体的讨论。社科研究的特殊挑战——比如概念的模糊性、理论的多元性、数据的复杂性——也会被专门讨论。

**人文学科研究者**可能会觉得某些章节（比如统计分析）与你的工作距离较远，但其他章节——比如文献整理、概念辨析、写作表达——应该同样适用。AI 在人文研究中的角色可能与社科有所不同，但核心问题是相似的：如何利用工具而不被工具绑架？如何保持批判性而不是盲目接受？

**自然科学和工程学科的研究者**可能会发现，虽然具体的例子和场景与你的领域有差异，但底层的方法论讨论是通用的。特别是关于可复现性、数据管理、代码规范的讨论，对任何经验研究都有参考价值。

## 1.4.3 4.3 不同技术背景的读者

**对 AI 技术有深入了解的读者**可能会觉得某些技术解释过于基础。我的建议是跳过那些你已经熟悉的部分，直接进入方法讨论。这本书的价值不在于教你“如何使用 ChatGPT”，而在于探讨“在学术研究中应该如何使用 AI”。

**对 AI 技术了解有限的读者**不用担心。本书不假设任何编程知识或技术背景。所有技术概念都会用通俗的语言解释，所有操作都会给出具体的步骤。如果你能用 Word 写论文、用浏览器搜文献，你就能跟上本书的节奏。

## 1.5 五、这本书不是什么

在说明这本书是什么的同时，我想明确它不是什么。

**这不是工具手册。**我不会逐一介绍市面上的 AI 工具，不会比较哪个工具“更好用”，也不会提供“N 个提示词模板让你效率翻倍”。工具更新太快，今天的教

程明天就可能过时。我更关心的是那些不会过时的东西：思维方式、方法论、判断标准。

**这不是提示词大全。**提示词（prompt）确实很重要，但它被过度神秘化了。很多所谓的“高级提示词技巧”其实只是常识的重新包装。真正决定输出质量的，是你对问题的理解、对任务的分解、对结果的评估——这些都发生在提示词之外。

**这不是“AI 取代研究者”的宣言。**我不认为 AI 会取代研究者，至少在可预见的将来不会。但我也不认为 AI 只是“更快的搜索引擎”或“更智能的拼写检查器”。它是一种新型的工具，需要新型的使用方式。本书探讨的就是这种新型使用方式应该是什么样的。

**这不是中立的技术介绍。**我有自己的立场和偏好。我认为研究的核心是发现而不是写作，我认为可复现性比效率更重要，我认为批判性思考不能外包给机器。这些立场会影响本书的内容和论调。如果你不同意这些立场，你可能不会喜欢这本书——但我依然希望你读完它，因为不同意也是一种对话。

## 1.6 六、本书的结构与阅读建议

本书采用“多入口”设计，不同读者可以根据自己的需求选择不同的阅读路径。

### 1.6.1 6.1 整体结构

全书分为四个部分：

**第一部分（第 1-3 章）：基础认知。**这部分回答最基本的问题：AI 是什么？它能做什么、不能做什么？使用它有什么风险？如何建立正确的心理预期？建议所有读者都从这部分开始，即使你觉得自己已经“很了解 AI 了”——很多“了解”其实是误解。

**第二部分（第 4-7 章）：能力模块。**这部分按照“能力”来组织内容：信息压缩、结构化表达、代码与计算、知识管理。每种能力都是学术研究需要的，每种能力也都可以借助 AI 来增强。如果你对某种能力特别感兴趣，可以直接跳到对应章节。

**第三部分（第 8-12 章）：研究工作流。**这部分按照研究的“阶段”来组织内容：选题、文献、数据、分析、写作。如果你正处于研究的某个阶段，想知道“在这个阶段可以怎么用 AI”，可以直接跳到对应章节。

**第四部分（第 13-18 章）：层次与反思。**这部分提出“五层次框架”，讨论从个人使用到团队协作再到制度设计的不同层面。最后一章是全书的总结与反思。如果你是团队负责人或对 AI 政策感兴趣，这部分值得仔细阅读。

### 1.6.2 6.2 阅读路径建议

**如果你是刚开始使用 AI 的新手：**建议按顺序阅读，不要跳章。第一部分的基础认知特别重要——很多后面的讨论都建立在这些基础概念之上。

**如果你已经有丰富的 AI 使用经验：**可以快速浏览第一部分，然后根据自己的需求选择第二部分或第三部分的具体章节。第四部分的“五层次框架”可能

会给你新的视角。

**如果你是带团队的负责人：**建议重点阅读第四部分，然后根据团队的具体需求选读其他章节。你可能需要的不是自己学会所有技巧，而是建立一套团队可以遵循的规范和标准。

**如果你只是想快速了解本书的核心观点：**可以只读第 1 章（基本原则）、第 3 章（提示词不是方法论）和第 18 章（终章）。这三章浓缩了全书最重要的思想。

### 1.6.3 6.3 每章的结构

每一章都采用相似的结构：

1. **问题引入：**这一章要解决什么问题？为什么这个问题重要？
2. **核心讨论：**理论分析、方法讨论、常见误区剖析
3. **实践指南：**具体的操作步骤、工作流示例、提示词参考
4. **案例分析：**真实或仿真的案例，展示如何在实践中应用
5. **反思与延伸：**这一章的局限性、开放问题、进一步阅读建议

你不必每次都按顺序读完一整章。如果你时间有限，可以先读“问题引入”和“反思与延伸”，快速把握要点；如果你需要立刻开始工作，可以直接跳到“实践指南”；如果你想深入思考，可以专注于“核心讨论”。

## 1.7 七、一些重要的声明

在正式开始之前，我想做几个声明。

### 1.7.1 7.1 关于时效性

AI 领域变化极快。这本书写作于 2024 年末到 2025 年初，其中提到的具体工具、功能、政策都可能已经更新。我会尽量确保书中的信息在出版时是准确的，但无法保证你阅读时依然准确。

不过，这也是我强调“方法论”而非“工具介绍”的原因。具体工具会变，但思考问题的方式不会轻易过时。如何评估工具的可靠性、如何设计人机协作的流程、如何在效率与质量之间取得平衡——这些问题在任何时候都是相关的。

### 1.7.2 7.2 关于局限性

我不是 AI 专家，不是计算机科学家，也不是技术哲学家。我的背景是社会科学研究。这意味着本书的视角必然有局限：

- 技术细节可能不够深入。如果你想了解大语言模型的内部原理，应该去读专业的技术文献。
- 社科以外的学科可能覆盖不够。我会尽量涵盖不同学科的例子，但我对某些领域的了解确实有限。
- 文化语境可能有偏差。我主要在中文和英文学术环境中工作，对其他语言 and 文化的学术传统了解有限。

我认为诚实地承认局限性是学术写作的基本态度。如果我假装自己什么都懂，读者反而无法判断哪些内容可信、哪些需要进一步验证。

### 1.7.3 7.3 关于 AI 辅助写作

这是一个无法回避的问题：这本书本身是不是用 AI 写的？

答案是：部分是。

更准确地说，我在写作过程中使用 AI 来做以下事情：

- 初稿的部分段落由 AI 生成，然后我进行大幅修改、重组、补充。
- 一些技术概念的解释参考了 AI 的输出，但核心观点和论证是我自己的。
- 文字润色和错别字检查使用了 AI 辅助。
- 某些案例的框架由 AI 生成，但具体细节是我补充或编造的。

我不认为这是值得隐瞒的事情。首先，这是我在书中倡导的“人机协作”的实践——如果我自己都不用 AI 来写作，这本书的建议就缺乏实践基础。其次，这也是对可复现性承诺的延伸——我愿意披露我的写作过程，让读者知道他们在读什么。

但我想强调两点：第一，本书的核心论点、整体结构、关键判断都是我的，不是 AI 的。AI 可以帮助生成文字，但不能帮我决定“什么值得说”。第二，所有重要的事实陈述和引用都经过我的人工核实。我前面讲的那个“失败的文献综述”的教训，我没有忘记。

### 1.7.4 7.4 关于争议性观点

本书包含一些可能引发争议的观点，比如：

- 很多所谓的“提示词工程”是被过度包装的简单技巧
- “效率提升”不应该是使用 AI 的主要理由
- 学术界对 AI 的某些恐惧是合理的，不应该被嘲笑为“守旧”
- 当前的很多“AI 科研工具”在商业利益驱动下夸大了自己的能力

这些观点不是随意发表的情绪，而是基于我对文献、实践和逻辑的综合判断。但判断可能是错的。如果你不同意某个观点，我欢迎你通过本书的 GitHub 仓库提出 issue，我们可以进行公开讨论。

## 1.8 八、为什么选择开源

这本书从一开始就是以开源项目的形式写作的。这个选择是刻意的，原因有几个。

### 1.8.1 8.1 可复现与可验证

学术研究的核心价值之一是可复现性：别人可以重复你的工作，验证你的结论。传统的学术出版物在这方面做得还可以——你可以看到论文的方法描述、

数据来源、分析代码。但书籍出版往往是不透明的：你看到的只是最终成品，不知道作者是如何得出那些结论的。

开源改变了这一点。在本书的 GitHub 仓库里，你可以看到：

- 每一章的写作历史和修改记录
- 我使用的提示词和 AI 输出的原文
- 读者提出的质疑和我的回应
- 错误的勘误和更正

这不仅是对读者的尊重，也是对自己的约束。当你知道自己的一切都会被公开审视，你会更加谨慎。

## 1.8.2 8.2 持续更新

AI 领域变化太快，一本传统出版的书籍在出版时可能已经过时。开源允许持续更新：当新工具出现、旧工具下架、某个建议被证明有问题，我可以及时修改。读者永远可以访问最新版本。

当然，这也带来版本管理的问题。我会维护清晰的版本号和更新日志，确保读者知道自己读的是哪个版本，以及这个版本与之前版本有什么区别。

## 1.8.3 8.3 社区协作

一个人的知识和经验总是有限的。开源允许社区协作：如果你发现某个错误、有更好的案例、想补充某个领域的视角，你可以直接提交贡献。这本书不仅仅我写的，也是所有贡献者共同创作的。

我承诺会认真对待每一个贡献，无论是批评还是建议。当然，最终的编辑权在我手里——这是为了保持全书的一致性和质量——但贡献者会在书中得到致谢。

## 1.8.4 8.4 价值观的表达

最后，选择开源也是一种价值观的表达。我相信知识应该是开放的，尤其是关于如何更好地生产知识。如果我写一本关于“AI 如何让学术研究更透明、更可复现”的书，然后把它锁在付费墙后面，这在逻辑上是自相矛盾的。

当然，开源不等于“没有成本”。维护这个项目需要时间和精力，我也需要生存。如果你觉得这本书对你有价值，可以通过购买实体书（如果出版的话）、赞助 GitHub 项目、或者简单地在社交媒体上分享来支持我。

# 1.9 九、我们正处于什么样的时刻

在结束这篇序章之前，我想谈谈我对当前时刻的理解。这不是技术分析，而是一种历史感——一种对“我们站在哪里”的判断。



### 1.9.1 9.1 炒作周期的高峰与低谷

如果你熟悉 Gartner 的技术炒作周期 (Hype Cycle)，你会知道每一项新技术都会经历类似的过程：先是过度乐观的炒作期，然后是幻灭期的低谷，最后才是稳定的应用期。大语言模型显然还处于炒作期的尾声——铺天盖地的报道开始减少，“AI 会取代一切”的论调也不如以前那么响亮。

但这并不意味着 AI 不重要了。恰恰相反，真正的影响往往在炒作退潮之后才开始显现。互联网的炒作期在 2000 年泡沫破灭后结束，但互联网对社会的真正改变发生在之后的二十年。我预感大语言模型也会是类似的轨迹：当媒体不再每天报道它，当人们不再把它当作神奇的新玩具，它才会真正融入我们的工作流程，产生深远而隐秘的影响。

这也是我选择在这个时候写这本书的原因。炒作期的内容往往过于狂热或过于恐惧，都难以提供冷静的分析。而在炒作退潮、实际应用逐渐铺开的阶段，才是最需要方法论指导的时候。

### 1.9.2 9.2 学术界的特殊处境

学术界对 AI 的态度一直很矛盾。一方面，很多研究者是 AI 技术的直接创造者和研究者——如果没有学术界的贡献，今天的大语言模型根本不会存在。另一方面，学术界又是最可能被 AI 冲击的领域之一——阅读、写作、分析这些 AI 擅长的任务，恰恰是学术工作的核心。

这种矛盾导致了两种极端的反应。一些人完全拥抱 AI，把它当作解放生产力的神器，迫不及待地用它来加速每一个环节。另一些人则极度警惕，视 AI 为学术诚信的威胁，呼吁全面禁止。

我认为这两种极端都是有问题的。完全拥抱忽视了风险，而完全禁止则忽视了现实——学生和 researcher 已经在用 AI 了，禁止只会把使用推向地下，让问题更难管理。我们需要的是第三条路：承认 AI 会被使用，然后认真思考“如何使用才是负责的”。

### 1.9.3 9.3 一场正在发生的范式转移

回顾科学史，每一次重大工具的引入都会带来研究范式的转移。显微镜让我们看到了微观世界，计算机让我们能处理海量数据，互联网让我们能即时获取全球文献。每一次转移都伴随着担忧和争议，但最终都被吸收进了“正常科学”的实践中。

大语言模型可能正在引发又一次这样的转移。它不只是“更快的工具”，而是一种新型的认知伙伴——它能理解（某种意义上的）语言，能生成（某种意义上的）知识，能与研究者进行（某种意义上的）对话。这是之前任何工具都做不到的。

当然，这种“理解”“知识”“对话”都需要打引号。模型是否真的理解任何东西，是一个哲学上有争议的问题。但无论如何，它的行为——它能做什么——是实实在在的。而正是这些行为，正在改变学术研究的实践。

这本书的写作，正是处于这场范式转移的早期阶段。我不能告诉你转移的终点在哪里，但我希望能帮你更清醒地认识到：转移正在发生，而你需要为此做好准备。

## 1.9.4 9.4 个人选择的重要性

面对技术变革，个体往往感到无力：趋势是不可阻挡的，个人能做什么呢？

但我想强调的是，个人选择依然重要。技术不是自动生效的，它需要被具体的人在具体的场景中采用。你选择如何使用 AI，会影响到你的研究质量；你的选择汇聚起来，会影响到整个学术社区的规范。

如果每个人都选择“快速产出”而忽视“仔细核实”，学术界就会充斥着低质量的研究。如果每个人都选择“隐藏使用”而不是“透明披露”，学术诚信的边界就会变得模糊。但反过来，如果足够多的人选择负责任的使用方式，并且公开倡导这种方式，新的规范就有可能形成。

这本书，从某种意义上说，就是我的选择——我选择花时间思考这些问题，并且把思考的结果分享出来。我希望它能帮助你做出你自己的选择。

## 1.10 十、致谢

这本书的写作得到了很多人的帮助，我想在这里表达感谢。

感谢我的同事和学生，他们在日常工作中与我分享了各种 AI 使用的经验和困惑。很多章节的灵感直接来自于我们的对话。

感谢那些在 GitHub 上提交 issue 和建议的读者（你们的名字会出现在贡献者列表中）。开源写作的意义就在于集体智慧的汇聚。

感谢那些写过“AI 科研指南”的前辈。虽然我在本章中批评了很多现有内容的不足，但我并不否认它们的价值——正是因为有了它们，我才能知道还缺什么，才能尝试填补空白。

感谢我的家人，他们容忍了我在电脑前度过的无数个夜晚。

最后，感谢你，读者。一本书只有在被阅读的时候才真正存在。你选择花时间读这本书，就是对我工作的最大肯定。

## 1.11 十一、写在最后

动笔之前，我犹豫了很久。

犹豫不是因为不知道写什么——相反，我有太多想说的。犹豫是因为担心自己没有资格说。我不是 AI 领域的权威，不是有几十年经验的老教授，甚至不是一个特别成功的学者。我凭什么写这本书？

后来我想通了：正因为我不是权威，我才可能写一本真正有用的书。权威往往离实践太远，他们的建议虽然“正确”，但普通研究者很难执行。而我，就是一个普通研究者。我每天都在用 AI 做研究，每天都在踩坑和填坑。我写的不是“应该怎么做”，而是“我是怎么做的，以及我从错误中学到了什么”。

这本书不会解决所有问题。AI 与学术研究的关系还在快速演变，很多问题现在还没有答案。但我希望它能提供一个起点：一个思考的起点，一个对话的起点，一个改进的起点。

如果你读完这本书，产生了一些想法——无论是同意、反对还是困惑——我希望你能告诉我。这本书是开源的，意味着它永远是未完成的。你的反馈会让它变得更好。

感谢你选择这本书。让我们开始吧。

## 2 第 1 章 AI 能做什么，不能做什么

AI 最擅长的不是“正确”，而是“像”。它能在语言上高度拟合学术表达的样子，能快速压缩文本、重写段落、生成结构化输出，但这并不等于它理解或验证了事实。研究者要先看清这一点：AI 是生成式系统，不是知识库、也不是裁判。

一个实用的判断方式是把 AI 能力分成四类：压缩、改写、重组、推理。压缩指摘要与信息提取；改写指语言润色与风格转换；重组指把散乱信息结构化；推理则是它最容易被高估的能力，因为模型擅长“构造合理解释”，却未必能承担“正确推导”的责任。

因此，AI“很强但不可靠”的根源在于：它更像一个高水平写作者，而不是可靠的事实校验器。你可以让它生成思路、提出假设、搭建框架，但关键环节必须回到人类研究者的验证流程里。把 AI 当成“高效助手”，而不是“权威来源”，这是这本书的首要原则。

### 3 第 2 章幻觉、偏差、泄露：三类核心风险

AI 带来的最大问题不是“效率”，而是“错误的效率”。三类风险贯穿所有研究流程：幻觉、偏差、泄露。它们不只是技术问题，更是方法论和伦理问题。

幻觉是模型在不确定时仍然给出流畅答案的倾向。它在科研中的危害不在于“出现错误”，而在于“错误看起来合理”。这会侵蚀研究者的判断力，尤其在文献综述、理论归纳、背景介绍等环节，幻觉常以“自信口吻”呈现，导致难以识别。

偏差来自模型训练数据与对齐机制，会在不知不觉中渗入研究。比如对某些群体的刻板印象、对某些研究范式的偏好，都会影响输出。研究者必须意识到：AI 的“中立语气”并不意味着“中立立场”。

泄露则是更现实的风险。把未发表的研究数据、敏感信息或受限数据直接输入模型，会带来合规与伦理问题。即便使用本地模型，也要考虑数据治理与权限边界。使用 AI 前先回答一个问题：这份数据能否被任何人看到？如果答案是否定的，就不应该直接交给模型处理。

## 4 第 3 章提示词不是方法论

提示词能提升输出质量，但它解决不了“研究结构”的问题。很多 AI 教程把提示词当作核心技巧，却忽视了研究的真正难点：问题的清晰性、证据的可信度、推理链条的完整性。这些并不会因为一个漂亮的提示词而自动出现。

更有效的思路是把提示词看成“流程中的一个步骤”。你需要的是可重复的研究系统：输入什么材料、让模型做什么任务、输出如何校验、结果如何记录。一个好系统比一千个提示词更重要，因为它能把 AI 稳定地嵌入你的研究实践中。

因此，本书不提供“万能提示词”，而提供“工作流原则”：任务拆解、层级输出、验证机制、版本记录。提示词只是表层，方法论才是底层。

## 5 第 4 章信息压缩与文献整理

AI 最“值钱”的能力之一是信息压缩：把大量材料转化为可读的结构。它能快速梳理文献主题、生成摘要、构建初步的研究地图。但这一步不是终点，只是起点。研究者应把 AI 输出当成“索引”，而不是“结论”。

在文献筛选上，AI 适合做粗筛与主题聚类，帮助你确认“这个领域的主问题是什么”。但在确定核心文献和理论链条时，需要人工判断。尤其在社科领域，关键争论往往不在摘要里，而在方法与证据的细节中。

长上下文模型让“一次性阅读几十篇论文”成为可能，但前提是你提供了高质量输入，并能设计输出格式。更重要的是校验引用：任何 AI 生成的文献综述都必须回到原文核对。可靠的综述不是“写得好”，而是“证据链可追溯”。

## 6 第 5 章结构化表达与写作

AI 能快速生成提纲、段落与结构，这在写作初期非常有用。你可以用它来整理研究框架、提出章节结构、梳理论证顺序。但要记住：结构不是思想，结构只是容器，关键是你要把真正的论证放进去。

“学术语言”与“AI 语言”之间的差别常常被忽视。AI 写出来的文字通常流畅、完整，但容易缺乏学术写作所需的精确性与责任感。它倾向于“把话说满”，而研究写作更强调“留出不确定性”。因此，AI 生成文本需要二次加工：删去过度自信的表达，补上证据与限定条件。

在翻译与润色上，AI 的优势是速度，但风险是术语与语境的偏差。最好的策略是：让 AI 做“第一遍粗修”，再由研究者做“关键点审校”。尤其是方法与结果部分，不能完全依赖 AI 完成。

## 7 第 6 章代码与计算能力

AI 对代码的帮助不在于“自动生成”，而在于“降低门槛”。它可以帮助你模糊想法转成可运行的脚本，减少重复劳动，并在遇到错误时快速定位问题。但这并不意味着你可以不理解代码。任何研究级分析都要求你能解释每一步处理逻辑。

在数据清洗、格式转换、批量处理等任务上，AI 的效率优势非常明显。它能快速给出脚本框架和可复用函数，适合用作“草稿生成器”。你应当把它当作代码搭子：让它产出、让你审核、共同迭代。

在统计分析与可视化上，AI 可以协助你构建模型、输出图表，但不能替代对模型假设的理解。模型选择、变量定义、结果解释都属于研究者的责任。AI 提供的是加速，而不是合法性。



## 8 第 7 章知识管理与协作

AI 的最大价值之一是“对话式启发”，但对话如果不被整理，就会迅速消失。研究需要记录、归档与版本管理。你需要把 AI 对话转成可检索的研究材料，而不是散落的聊天记录。

一个可用的流程是：对话 → 摘要 → 标签 → 归档。每次关键对话都应输出结构化总结，并注明时间、模型、提示词与用途。这些记录构成研究的“隐形材料”，决定了你能否复现自己的思路。

在团队协作中，AI 既能提升效率，也会模糊责任边界。谁写的、谁验证的、谁承担错误？这些问题必须在协作中明确。开源和透明不只是技术选择，也是团队协作的伦理底线。

## 9 第 8 章选题与研究问题

选题阶段是 AI 最容易“帮倒忙”的环节。它能快速生成许多漂亮的问题，但这些问题往往缺乏可检验性、现实可行性或理论价值。研究者必须先明确研究问题的“可研究性”，再使用 AI 辅助拆解。

一个有效的方法是先由人写出核心问题，再让 AI 做“多角度压力测试”：是否存在可用数据？是否已有大量研究？是否存在明确的识别路径？AI 擅长从多个角度给出提醒，但这些提醒必须由研究者来做最终判断。

最常见的风险是“问题被做空”：AI 倾向于把复杂问题简化为泛化叙述，导致研究失去锋利度。研究者要主动保持问题的边界感，避免被 AI“优化成毫无棱角的宏大命题”。

## 10 第 9 章文献与理论构建

AI 可以帮你搭建“文献地图”，但搭建“理论链条”仍然需要研究者的判断力。文献地图强调覆盖面，理论链条强调逻辑性。二者不是同一件事。

在文献整理阶段，AI 适合做主题聚合与脉络描述：哪些研究关注同一问题、哪些方法常被使用、哪些争议存在。但理论构建需要你明确变量关系与因果机制，AI 只能辅助解释，不能替你承担理论责任。

所谓“AI 综述陷阱”是指：输出看起来完整，但引用无法核对，或者概念混杂不清。避免陷阱的办法是把 AI 综述当成“初稿目录”，逐条回到原文校验，并用“证据链”替换“叙述链”。

## 11 第 10 章 数据获取与构造

数据获取是 AI 加速最明显的环节之一。无论是公开数据、爬虫还是 API 接口，AI 都能帮助你快速写出脚本并完成批量抓取。但“获取速度”不等于“数据质量”，任何数据都需要清晰的采集记录与合法性说明。

AI 生成数据是更具争议的方向。它可以用于模拟、训练或思维实验，但不能轻易替代真实样本。研究者必须区分“生成数据用于方法测试”和“生成数据用于经验结论”这两种完全不同的用途。

“硅样本”的方法论争议在于：AI 是否能代表人类行为？如果 AI 是研究对象，它本身就不是“人”。使用硅样本必须明确其理论定位，否则会把“模型输出”误当成“社会事实”。

## 12 第 11 章数据清洗与分析

数据清洗是科研中最耗时也最容易被忽略的环节。AI 的优势在于快速生成清洗脚本和规则，尤其适用于文本数据的结构化处理。但清洗规则背后往往包含方法论选择，这些选择必须透明化、可解释、可复现。

在统计分析阶段，AI 可以帮助搭建模型、生成图表、解释结果，但研究者必须对模型假设负责。AI 擅长“讲故事”，但不擅长“守住边界”。如果你让 AI 解释结果，它可能会给出看似合理但并不严谨的因果叙述。

因此，AI 的角色应该是“分析助手”，而不是“结论生成器”。任何结论性表述必须回到数据与方法本身，这是学术研究的底线。

## 13 第 12 章写作、投稿与传播

AI 最受欢迎的应用场景是写作。你可以用它生成初稿、润色语言、调整结构，但这并不意味着它能替你“完成论文”。论文的核心价值来自研究设计与证据，而不是表达方式。

在投稿流程中，AI 可以帮助你整理摘要、润色投稿信、模拟审稿人视角。但“反 AI 检测”不应该成为研究者的目标。真正合理的策略是：确保内容真实、证据充分、逻辑清晰。所谓“AI 味”，往往来自缺乏具体细节与可验证信息，而不是语言风格本身。

在答辩与传播阶段，AI 适合帮助你做摘要、演讲稿与受众调整，但关键观点必须是你自己的。AI 可以把话说漂亮，但不能替你承担学术立场。

## 14 第 13 章 RA Level：工具层

在工具层，AI 等同于一个高效率的研究助理。它能完成大量重复性工作：整理数据、生成初稿、清洗文本、批量翻译。这个层级的核心价值是节省时间和成本。

但工具层的风险在于“过度依赖”。当 AI 替代了研究者的基础劳动，研究者也可能丧失对材料的直觉理解。你必须确保自己仍然掌握数据细节和研究逻辑，否则 AI 的效率会把你带向“看似完成但无法解释”的境地。

最重要的问题是：谁在控制流程？如果 AI 的输出驱动了你的研究方向，那么你已经从“使用工具”变成“被工具牵引”。工具层的原则是：AI 做脏活累活，人保留判断权。

## 15 第 14 章 Supervisor Level： 认知协作层

在认知协作层，AI 不只是执行者，而是“对话式导师”。它能帮助你提出假设、模拟审稿人、补足背景知识，甚至推动你跨学科思考。这一层是 AI 最有创造性的应用。

但同样，风险也在于“依赖与懒惰”。当你习惯用 AI 生成思路，可能会丧失独立构建问题的能力。AI 会不断给出“合理建议”，但合理并不等于有价值。研究者必须保持自己的问题意识与判断标准。

认知协作层的关键是：让 AI 成为“思维刺激器”，而不是“思维替代品”。你可以借助 AI 扩展视野，但不能放弃自己的判断。



## 16 第 15 章 Domain Expert Level: 推理与建模层

在推理与建模层，AI 能够模拟“领域专家”的思维方式，帮助你理解模型、构建算法或补齐技术短板。对跨学科研究者而言，这一层极具价值，因为它能缩短学习曲线。

但必须明确：推理不等于理论。AI 可以生成“看似合理”的模型，却不一定满足学科内部的规范与逻辑要求。理论构建需要领域知识、文献脉络与方法论约束，而这些不是 AI 能自动承担的。

这一层的正确使用方式是：让 AI 提供“草稿与解释”，由研究者完成“理论合法性与方法论判断”。领域知识的不可替代性，恰恰是学术研究者的核心价值。

## 17 第 16 章 Agent Level: AI 作为行动者

当 AI 开始模拟人类、参与互动或自动决策时，它不再只是工具，而成为“行动者”。这对社会科学提出新的本体论问题：研究对象是否还仅限于“人”？当 AI 参与社会系统，它是否也构成社会事实的一部分？

在这一层，AI 被用来构建虚拟社会、模拟行为、生成硅样本。这些方法可以帮助研究者探索复杂系统，但也必须承认其局限性：模型输出并不等于真实社会。虚拟样本的价值在于“实验性启发”，而不是“经验结论”。

Agent 层提醒我们：AI 不仅改变研究工具，也改变研究对象。研究者必须明确自己是在研究“人”、还是研究“人—AI 互动”，或是研究“AI 本体”。这些区分决定了理论框架与解释边界。

## 18 第 17 章 Governance Level: 制度与治理 层

在治理层，问题已经超越“如何使用 AI”，而是“AI 如何改变学术制度”。当 AI 可以大规模生成文本，学术评价体系会发生什么变化？当研究流程被自动化，学术劳动的价值如何被重新定义？

这一层的核心是结构性影响：论文产出可能变得更快，但“学术质量”可能被稀释；评审机制可能更依赖表面表达，忽视真实研究价值。AI 带来的不是简单的效率提升，而是制度逻辑的重构。

研究者需要在治理层承担新的责任：明确 AI 使用边界、推动透明化、参与学术伦理讨论。真正的挑战不是“AI 能不能用”，而是“学术共同体如何定义可信”。

## 19 终章 | AI 时代研究者的新能力结构

AI 让研究更快，这是事实。但更快不等于更深。真正的问题是：在 AI 时代，研究者的不可替代能力是什么？答案不是“会用工具”，而是“能够提出问题、建立证据链、守住方法论边界”。

未来的研究者需要新的能力结构：一是系统化的问题意识；二是对证据的严格要求；三是对 AI 输出的批判性判断。AI 可以帮助你节省时间，但它不会替你承担学术责任。

开源在这个时代变得更重要。它不仅是分享资源，更是对学术可信度的回应。面对 AI 可能带来的“空洞高产”，开源与可复现是一种必要的抵抗：把研究拉回事实、证据与共同体的监督之中。

## 20 附录（可选）

本书后续版本将提供以下配套材料，全部开源更新：

- AI 科研工具箱清单（按任务分类）
- 提示词模板与工作流脚本（可复用）
- 案例复现材料与数据链接（含版本号）
- AI 使用伦理与声明模板（投稿与答辩场景）

附录的目的不是堆工具，而是让方法论落地。读者可以根据自己的研究领域替换工具，但不应跳过验证与记录步骤。

Knuth, Donald E. 1984. “Literate Programming.” *Comput. J.* 27 (2): 97–111. <https://doi.org/10.1093/comjnl/27.2.97>.