生成式人工智能如何外化研究判断：教材研究中的元认知监控机制

How Generative AI Externalizes Research Judgments: Metacognitive Monitoring in Textbook Research

1 胡月宝 新加坡南洋理工大学国立教育学院
2 黄龙翔 新加坡南洋理工大学国立教育学院

*guatpoh.aw@nie.edu.sg

【摘要】生成式人工智能已成为学术写作的“第二工作流”，但其如何影响研究者的元认知监控仍缺乏系统分析。本文提出“JUDGE 智判-判断外化框架”，认为生成式人工智能通过外化写作中自动完成的研究判断，从而触发元认知监控与调节。以教材研究为情境，本文采质性主导的混合研究设计，对研究写作中的人机互动进行分析。研究数据包含 113 条人机互动轮次，并识别出 10 个判断事件。编码与统计分析结果显示，责任回收与判断显影意识最为常见，并形成稳定的行为配置与序列路径。研究表明，生成式人工智能并非替代研究判断，而是通过外化判断节点激活元认知监控。

【关键词】生成式人工智能；元认知监控；判断外化；人机共研；教材研究

Abstract: Generative artificial intelligence (AI) has become a “second workflow” in academic writing, yet its influence on researchers’ metacognitive monitoring remains underexplored. This study proposes a Judgment Externalization (JUDGE) framework, arguing that generative AI promotes metacognitive regulation by externalizing research judgments that are typically completed automatically during writing. Using textbook research as the context, a qualitative-dominant mixed-method design was employed to analyze human–AI interactions in research writing. The dataset includes 113 interaction turns grouped into 10 judgment episodes. Coding and statistical analyses show that responsibility reclamation and judgment awareness are the most frequent responses, forming stable behavioral configurations and sequential patterns. The findings suggest that generative AI activates metacognitive monitoring by externalizing judgment points rather than replacing scholarly reasoning.

Keywords: generative artificial intelligence, metacognition, judgment externalization, human–AI co-research, textbook research

1. 前言

近年来，生成式人工智能（Generative Artificial Intelligence, AI）迅速进入学术写作与研究实践，并逐渐形成与阅读、书写和分析并行运作的“第二工作流”（second workflow）。在这一情境下，研究者与人工智能之间的互动不再只是工具使用，而逐渐发展为一种人机共研（human–AI co-research）的研究过程。既有研究多关注 AI 对写作效率或文本质量的影响，却较少探讨其如何改变研究者进行判断、监控与反思的认知结构（Luckin, 2018; Holstein, McLaren, & Aleven, 2019）。尤其在语言与人文学科研究中，概念命名、解释路径选择、证据采纳与价值外推等关键判断，往往在写作过程中以内隐且自动化的方式完成，从而绕过显性的元认知监控。基于此，本文从判断外化（judgment externalization）的视角出发，关注生成式人工智能如何通过外化研究判断，使其进入可被比较、质疑与修订的分析过

程，并在人机共研情境中触发元认知监控与调节。为探讨这一机制，本文以教材研究作为经验场域，对人机互动过程中判断外化所引发的元认知行为进行分析，并提出以下研究问题：
RQ1：生成式人工智能如何通过判断外化，使学术写作中的研究判断变得可被察觉、比较与调节？（机制层）
RQ2：在教材研究情境中，判断外化之后会呈现出怎样的元认知行为分布、配置与序列模式？（行为模式层）

2.文献综述：从元认知困境到判断外化

元认知通常被界定为个体对自身认知过程的觉察与调节能力，其核心机制包括认知监控（monitoring）与认知控制（control）（Flavell, 1979；Nelson & Narens, 1990）。相关研究指出，元认知并非一种稳定能力，而是在具体任务中动态发生，其有效性取决于关键判断能否被察觉并进入调节循环。然而，在真实的学术写作中，大量研究判断往往以自动化方式完成，从而绕过监控层级。由此，元认知受限的关键并非反思意愿不足，而在于判断缺乏被暂停与被检视的结构条件（Dunlosky & Metcalfe, 2009）。

基于这一困境，本文提出的“JUDGE 智判-判断外化框架“并非替代经典元认知理论，而是在生成式人工智能介入研究情境下，对元认知调节发生条件的进一步细化。Flavell（1979）将元认知区分为元认知知识（metacognitive knowledge）与元认知调节（metacognitive regulation），Nelson 与 Narens（1990）则进一步提出监控与控制的双层结构模型。JUDGE 智判框架关注的不是元认知构成要素本身，而是这些要素在研究实践中如何被触发、显影与观察。其核心命题是：判断外化是元认知调节得以发生的结构性前提。当研究判断被外化为可比较、可质疑与可延缓的对象时，研究者才能对其进行监控与调节。

教育技术研究早已表明，外显提示与结构化反馈可促进学习者的监控与调节行为，说明元认知高度依赖外部结构触发。进一步的认知研究提出“认知外化” 概念，指出工具可将内隐、瞬时的认知活动转化为可观察对象，从而改变认知调节的发生条件（Kirsh, 2013）。与以往技术主要停留在步骤引导或策略提示不同，生成式人工智能能够在开放、复杂的学术写作任务中持续生成替代表述、解释路径与论证方案，从而使原本自动完成的研究判断显性化。

从元认知视角看，生成式人工智能的关键优势不在于替代认知，而在于外化并延缓判断。AI 通过替代表述与不同论证路径打断熟练任务中的自动化决策，使“此处存在判断”这一事实显性化，从而为监控与调节创造条件。与此同时，AI 不具备学术责任，其输出不可直接署名，这种责任不对称性反而强化了研究者对判断承担的元认知意识。需要注意的是，流畅的 AI 输出也可能制造“伪元认知感“，使判断责任在不被察觉的情况下转移给技术系统。因此，AI 是否真正促进元认知，并不取决于输出质量，而取决于研究者是否将其识别为需要回应的判断信号。

3. JUDGE 元认知视角下的人机共研框架

3.1. 框架核心

基于语言与人文学科研究写作中元认知监控易被自动化判断绕过的困境，本文提出智判（JUDGE）元认知视角，将元认知理解为一种以判断外化为前提、在具体写作情境中被触发的认知调节过程。与将元认知视为稳定能力的传统观点不同，本文提出智判（JUDGE）框架强调：当判断以内隐方式迅速完成并直接固化为文本结果时，监控与调节难以发生；唯有当判断被外化为可比较、可质疑与可延缓的对象时，元认知监控与调节才具备发生条件。

在此意义上，生成式人工智能并不承担判断责任，而是作为一种“判断外化的触发者”（judgment externalization trigger），通过替代表述、结构重组与不同推论路径，使原本自然完成的判断转化为需要被回应的对象。研究者则承担监控、调节与最终认识论决策的责任。

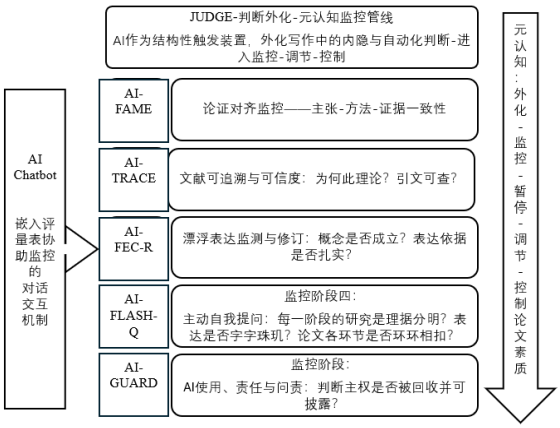
3.2. 五类判断外化机制

在 JUDGE 视角下，人机共研写作中的认知调节可被理解为一条由判断外化通向元认知控制的调节管线（见图 1）。本文将其中的判断外化机制概括为五类：

1. **AI-FAME**：关注论证是否在事实、方法、证据与主张之间保持对齐。
2. **AI-TRACE**：关注解释与推理链是否真实、可追溯。
3. **AI-FEC-R**：关注概念命名是否承担分析责任，避免语言成立但意义悬浮。
4. **AI-FLASH-Q**：关注判断外化后是否引发主动自我提问与形成性调节。
5. **AI-GUARD**：关注判断完成后责任是否被明确回收至人类研究者。也就是指研究者对 AI 提出的概念、解释或论断进行显性区分、修订与归责，并将最终判断明确置于作者可署名、可追责的认识论责任之下。

五类机制共同构成从“论证成立”走向“认识论负责”的元认知控制路径（图 1）。

图 1 *判断外化—元认知监控管线*



4. 研究方法

4.1. 研究设计与案例来源

本研究采用质性主导、结构化统计辅助的混合研究设计，旨在分析生成式人工智能在教材研究写作过程中如何外化研究判断，并触发研究者的元认知调节行为。教材研究被选为经验分析场域，主要基于两点考虑：第一，教材研究高度依赖研究者在解释层级、概念命名与比较前提等关键节点上的判断，因此适合观察判断外化与元认知调节之间的关系；第二，教材文本具有制度化与规范化生产的特点，既承载知识结构，也体现社会文化价值取向，为生成式人工智能生成替代表述、解释路径与比较框架提供了空间。

为检验 JUDGE 智判—判断外化框架 在不同研究范式中的解释力，本文选取两个教材研究案例。案例 A 为质性话语分析取向的教材文化表征研究，案例 B 为量化文本可读性比较研究。两个案例分别侧重理论建构与结构验证，从而形成互补研究路径。

案例 A《语言、文化与认同的教育化转译——新加坡双语教育语境下小学华文教材“华人性”建构研究》以新加坡“华人性”相关中英文文献及小学华文教材《欢乐伙伴 1.0》1–6 年级课文与活动本（20 册）为数据来源，共涉及 14 篇文献与 491 条语篇单元，最终形成 62 个概念与 5 个分析维度。案例 B《从语言生态到人才建构——核心圈与中圈母语教材可

读性结构比较研究》以中国部编版《语文》教材与新加坡《欢乐伙伴》《华文伴我行》为研究对象，共 36 册教材，并通过可读性公式与文本结构指标进行比较分析。

两个案例的基础研究与数据分析工作均已完成。在论文修改阶段，研究者引入生成式人工智能开展人机共研写作过程分析。相关互动记录完整导出，并作为本研究的分析语料。两个案例的人机互动过程分别持续约一个月；本研究语料分析与写作过程则为二个月。

4.2. 人机交互与编码方式

本研究将生成式人工智能触发的判断外化（Judgement externalization）操作化为自变量，将研究者在互动过程中表现出的元认知行为作为因变量。元认知行为包括六类：M1 判断中断、M 2 解释回退、M 3 命名审查、M 4 责任回收、M 5 判断显影意识、T 判断追踪。（示例见附录 A）。

本研究中的人机互动基于 OpenAI 开发的大语言模型 GPT-4 架构的生成式人工智能系统。所有互动均在同一模型版本与同一对话环境中连续进行，并完整导出对话记录作为分析材料。AI 在研究过程中被用作辅助分析工具，用以生成替代表述、解释路径与概念命名建议，而不被视为研究结论的直接生产者。

在具体操作中，研究主要使用三类提示策略：概念提示、解释提示与论证提示，分别用于生成概念命名或分类方式、提供不同解释路径以及检验论证与证据之间的逻辑关系（示例见附录 B）。

本研究的分析单位为判断事件（judgement episode）。判断事件并非单条对话，而是围绕同一研究判断所形成的一组连续互动过程；同一判断事件通常包含多条人机互动条目（interaction turns）。本文将人机互动条目界定为研究者向 AI 提出一次分析请求并获得相应回复的完整轮次。在全部语料中，共记录 113 条人机互动条目，其中案例 A 为 62 条，案例 B 为 51 条。在这些互动过程中，共识别出 10 个判断事件。这些事件分别体现为 JUDGE -智判外化框架下五类判断外化机制在两个案例中的具体实例。所有互动条目随后被归入相应判断事件，并进一步编码其触发的元认知行为。表 1 为研究设计与方法总览：

表 1 *研究设计与方法总览（JUDGE 视角）*

方法维度	内容说明
研究取向	质性主导、结构化统计辅助的混合研究设计（qualitative-dominant mixed-methods design）
研究目标	分析生成式人工智能如何通过判断外化（judgement externalization）触发研究者的元认知调节行为
理论视角	JUDGE 元认知框架：将 AI 触发的判断外化视为研究核心机制
研究案例	两个教材研究案例：案例 A（质性扎根理论研究）与案例 B（量化文本可读性比较研究）
数据来源	案例 A：14 篇文献 +491 条教材语篇单元；案例 B：36 册中、新教材
人机互动数据	共 113 条 interaction turns，其中案例 A 62 条，案例 B 51 条
分析单位	判断事件（judgement episode）：围绕同一研究判断形成的一组连续人机互动
判断事件数量	共识别 10 个判断事件（两个案例各 5 个）
自变量	五类判断外化机制：AI-FAME、AI-TRACE、AI-FEC-R、AI-FLASH-Q、AI-GUARD
因变量	六类元认知行为：M1 判断中断、M2 解释回退、M3 命名审查、M4 责任回收、M5 判断显影意识、T 判断追踪

分析策略 1	分布型分析：统计各类元认知行为出现频次
分析策略 2	配置型分析：分析元认知行为在判断事件中的共现组合
分析策略 3	序列型分析：分析元认知行为的稳定展开顺序
信度检验	Cohen’s Kappa = 0.83（substantial agreement）

本研究邀请两名具有教育研究与质性分析背景的研究助理参与编码。编码分为编码体系建立、试编码与校准、正式编码三个阶段。10 个判断事件都被用于一致性检验，采用 Cohen’s Kappa 系数进行评估。总体结果为 Cohen’s Kappa = 0.83（M1=0.82、M2=0.79、M3=0.86、M4=0.88、M5=0.80、T=0.83），属于较高一致性，说明编码体系具有良好可靠性。

5. 结果

从分布型结果来看，各类元认知行为均在两个案例中出现（表 2）。其中，责任回收（M4）与判断显影意识（M5）的出现频次最高（各 10 次），表明当生成式人工智能外化研究判断时，研究者最常见的反应并非直接接受 AI 生成内容，而是意识到相关结论属于解释性判断，并重新承担相应的研究责任。相比之下，判断追踪（T）的出现频次相对较低（5 次），但通常与责任回收行为同时出现，反映出研究者在承担判断责任之后，会通过文本修订或结构调整留下可追踪的研究痕迹。

从配置型结果来看，元认知行为往往以稳定组合的形式出现。其中，最典型的行为配置包括“命名审查 + 责任回收”（M3 + M4） 与“解释回退 + 判断显影 + 责任回收”（M2 + M5 + M4）。这一结果表明，当 AI 提供概括性命名或顺写式解释时，研究者通常不会直接采纳或否定，而是通过回退解释层级、审查概念命名，并最终显性回收研究判断的责任。

从序列型分析来看，元认知行为呈现出相对稳定的展开路径。多数判断事件首先由 判断中断（M1） 引发解释回退（M2），随后进入判断显影意识（M5） 或命名审查（M3），并最终指向 责任回收（M4） 与判断追踪（T）。其中，“M4 → T” 的路径显示，在研究者重新承担判断责任之后，往往会通过文本修订、图表更新或结构调整等方式留下可追踪的研究痕迹，从而使研究判断在方法层面具有可审计性。

综合分布、配置与序列分析结果可以发现，M1–M5 与 T 六类元认知行为在判断外化触发下呈现出一种相对稳定的判断调节循环（Judgment regulation cycle）。在这一循环中，研究者首先对外化的研究判断产生中断与解释回退，随后对概念命名或解释路径进行审查，并在意识到相关结论属于解释性判断而非数据直接结果时形成判断显影意识。最终，通过责任回收与判断追踪，研究者将相关判断重新纳入作者可署名、可追责的研究实践之中，使研究判断从自动完成的认知过程转变为可监控、可修订与可追踪的研究行动。

表 2：*元认知行为的分布—配置—序列综合统计表*

元认知行为标记	元认知行为说明	案例 A 出现次数	案例 B 出现次数	合计出现次数	典型配置（共现）	稳定后续路径（序列）
M1 判断中断	暂停原有分析路径，对 AI 输出或当前研究判断提出质疑	4	3	7	M1 + M2	M1 → M2

M2 解释回退	将已形成的解释回退至方法、统计或分析前提层级	4	4	8	M1 + M2；M2 + M5	M2 → M5；M2 → M3
M3 命名审查	对核心概念或范畴命名进行比较、质疑或延迟采纳	5	4	9	M3 + M4；M3 + M4 + T	M3 → M4
M4 责任回收	明确区分 AI 表述与研究者判断，回收认识论责任	5	5	10	M3 + M4；M2 + M5 + M4	M4 → T
M5 判断显影意识	显性意识到当前结论属于判断而非数据直接结果，并区分分析层级	5	5	10	M2 + M5；M2 + M5 + M4	—
T 判断追踪	通过文本、图表或结构修订留下可追踪的判断调节痕迹	3	2	5	M3 + M4 + T	—

综述之，RQ1 关注生成式人工智能如何通过判断外化使研究判断进入元认知监控过程。分析结果显示，当 AI 提供替代表述、解释路径或概念命名建议时，研究判断往往被显性化为可被比较与质疑的对象。在这一过程中，研究者首先产生判断中断与解释回退（M1、M2），随后通过命名审查或判断显影意识（M3、M5）重新审视解释层级，并最终通过责任回收（M4）明确区分 AI 表述与研究者判断。这一过程表明，生成式人工智能通过外化研究判断，为元认知监控与调节提供了结构性触发条件。RQ2 关注在教材研究情境中判断外化所呈现的元认知行为模式。综合分布、配置与序列分析结果可以发现，M1–M5 与 T 六类元认知行为在判断外化触发下呈现出一种相对稳定的 判断调节循环。在这一循环中，研究者首先对外化的研究判断产生中断与解释回退，随后对概念命名或解释路径进行审查，并在意识到相关结论属于解释性判断而非数据直接结果时形成判断显影意识。最终，通过责任回收与判断追踪，研究者将相关判断重新纳入作者可署名、可追责的研究实践之中，使研究判断从自动完成的认知过程转变为 可监控、可修订与可追踪的研究行动。

6. 讨论与结论

本研究为理解学术写作中的人机共研模式提供了一种新认知视角。与将 AI 视为文本生成工具的传统观点不同，研究结果显示，其更重要的功能在于通过判断外化使研究判断进入可被监控与调节的元认知过程。从机制层面（RQ 1）看，生成式人工智能主要通过三个方面影响研究判断的调节过程：判断显影、责任回收以及可追踪修订痕迹的形成。AI 通过替代表述与不同解释路径，使原本在写作过程中自动完成的研究判断显性化，从而迫使研究者对这些判断进行再次检视。这使元认知监控不再依赖事后反思，而成为写作过程中持续发生的调节活动。在行为模式层上（RQ2），尽管案例 A 与案例 B 分别属于质性教材分析与量化可读性比较研究，两者在元认知调节结构上不仅表现出较为一致的倾向，也出现了稳定的判断调节循环。这一结果表明，JUDGE—智判外化框架在不同研究范式情境中具有一定的

解释潜力。但需指出的是，本研究所观察到的行为模式亦可能在一定程度上反映研究者自身的元认知监控倾向，因此仍有待在更多研究者与研究情境中进一步检验。

从方法论角度看，研究责任不再仅体现于最终结论的署名，而逐渐扩展为贯穿研究全过程的可追踪判断实践。生成式人工智能通过持续外化判断节点，使研究者能够在写作过程中不断进行元认知监控与责任回收，从而重新界定人文学科研究中“判断”的方法论地位。

同时，生成式人工智能在促进判断外化与元认知调节的同时，也可能带来若干认识论风险。第一，AI 生成的多种解释路径可能产生“伪元认知感”，使研究者误以为自己已经进行了充分反思；第二，AI 输出的高度流畅性可能导致判断责任转移，使研究者在无意识中将解释责任让渡给技术系统；第三，频繁使用 AI 生成概念命名与解释框架，可能导致研究者形成对 AI 的认识论依赖，而未充分反思其背后的理论假设。因此，在 AI 支持的研究实践中，有必要通过明确的元认知监控机制，确保研究判断仍由人类研究者最终承担。JUDGE 智判外化框架所强调的“责任回收”，正是为了避免判断责任在不被察觉的情况下被技术系统所取代，并在促进判断外化的同时维持研究判断的主体性。

本研究仍存在一定局限。作为一项探索性过程研究，本文主要揭示生成式人工智能在研究写作中通过判断外化触发元认知调节的可能机制，相关行为模式亦可能部分反映研究者自身的元认知监控倾向。未来研究可在更多研究者与不同学科情境中进一步检验这一机制。

7. 理论贡献

本研究从判断外化的视角重新理解生成式人工智能在学术写作中的作用，指出 AI 的关键功能并非替代研究判断，而是通过外化原本自动完成的判断，使其进入可被监控与调节的元认知过程。通过对人机互动过程的编码分析，研究识别出由判断中断、解释回退、命名审查、判断显影意识、责任回收与判断追踪构成的判断调节循环（judgment regulation cycle）。此外，在质性教材分析与量化可读性比较研究两个不同研究范式中的案例结果显示，该机制具有一定的跨范式解释潜力。

参考文献

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage Publications.

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Article 316, 1–14.

Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction, 20*(1), Article 3.

Luckin, R. (2018). *Machine learning and human intelligence: The future of education for the 21st century*. UCL IOE Press.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). Academic Press.

生成式人工智能使用声明

本研究在写作过程中有限使用本研究所研发的“JUDGE 元认知视角下的人机共研论证监控框架”以监控研究素质，但人工智能未参与研究判断或结论生成。所有分析、解释与认识论责任均由作者本人承担，人工智能仅作为判断外化的辅助工具使用。

附录 A 典型判断外化与元认知行为序列示例

序列编号	对话片段节选	AI 判断外化情境	元认知行为标记	判断过程编码（0–3）	判断痕迹	原对话轮（18 条来源）
P1	数据统计方式不一样，如何解释？	AI 给出不一致统计结果	M1 判断中断	1（被察觉）	暂停当前分析并质疑统计口径	E1, E2
P2	三组数据属于三个不同层级的分析结果。	AI 解释为层级差异	M2 解释回退	2（被调节）	将解释回退至分析层级区分	E3, E5, E16
P3	不能把三表直接放在一起比？	AI 区分统计类型	M5 判断显影意识	1（被察觉）	形成方法层面的元话语说明	E4, E6, E7, E10
P4	核心范畴这样命名会不会太宽泛？	AI 提供命名方案	M3 命名审查	2（被调节）	延迟概念命名并比较不同方案	E11, E12, E13
P5	结论是怎么下的？	AI 协助总结结论	M4 责任回收	3（被责任化）	明确研究判断的责任主体	E8, E9, E14, E15
P6	要不要把三级编码画出来？	AI 建议方法可视化	T 判断追踪	3（被责任化）	形成可追踪的研究修订痕迹	E17, E18

注 1：0＝判断自动完成且未引发任何元认知行为；1＝判断被察觉并引发判断显影意识（M5）；2＝判断被中断、回退或审查（M1–M3）；3＝判断被明确归责并纳入作者署名责任（M4）。Trace（T）用于记录判断追踪路径与证据链来源。注 2：Trace（T）用于记录判断追踪路径与证据链来源。其中，S-n 指向源材料或编码规则的回溯（source-based tracing），L-n 指向判断在分析层级中的逻辑回退或位置核查（logic/level tracing），V-n 指向解释责任与作者署名的确认过程（validation/voice tracing）。

附录 B AI 提示示例与判断外化机制

框架层级	提示示例	触发的元认知行为	判断调节功能
AI-FAME（论证对齐）	检查这段研究结论是否与前文方法和数据分析保持一致，是否存在推论超出证据范围。	重新审查结论与证据之间的关系	暴露论证边界

AI-TRACE（推理链追溯）	列出从研究结果到当前解释结论之间可能隐含的推理步骤，并指出哪些步骤缺乏证据支持。	外化隐含推理步骤	补充解释链
AI-FEC-R（漂浮性审查）	文中将“文化呈现”解释为“儒家价值观体现”。分析这概念等同是否由数据直接支持。	区分文本描述与研究者解释	防止概念外推
AI-FLASH-Q（写作监测）	针对该研究结论，请提出三条可能的反向问题或替代解释。	形成主动自我提问	触发形成性调节
AI-GUARD（责任回收）	标记这段论证中哪些判断必须由研究者承担解释责任。	明确判断归属	回收认识论责任