

初中 · 高中共用 | 统一议题 · 赛前阅读材料

能力、判断与社会条件

# 生成式AI的普及 让人更自由，还是更不自由？

心灵哲学 · 伦理学 · 社会科学 · 教育与劳动

**正方** 生成式AI的普及，总体上让人更不自由。

**反方** 生成式AI的普及，总体上让人更自由。

从这个问题开始

## 当工具越来越强大， 我们是否也越来越能做主？

能力增加，不自动等于自由增加；产生依赖，也不自动等于失去自由。争论的关键是：新增的力量，如何变成一个人真正能够支配的生活？

| 阅读路径   | 共同问题与材料              | 页码    |
|--------|----------------------|-------|
| 共同题解   | 讨论范围、双方责任与研究地图       | 2-3   |
| 能力与学习  | 01-04 心灵延伸、密尔、两项教育研究 | 4-11  |
| 自主与说服  | 05-06 影响、操控与AI说服实验   | 12-15 |
| 机会与工作  | 07-09 可行能力、技术路径与工作现场 | 16-21 |
| 创作与多样性 | 10 个人表达与集体新颖性        | 22-23 |
| 补充选读   | A1-A3 杜威、比斯塔、写作伦理    | 24-29 |
| 出处与说明  | 原文入口、许可、翻译与使用说明      | 30-31 |

无需哲学基础。完整译段、独立概要与编者分析分别标示。材料提供比较工具，不提供标准立论，也不把任何作者永久分配给某一方。目录与来源链接可点击。

## 共同题解 · 01

## 我们究竟在比较什么？

本题讨论生成式AI广泛进入学习、工作、创作与日常决策之后，人的自由发生了怎样的整体变化。它不是评选最强模型，也不只讨论作弊、就业或某一种使用方式。双方可以从具体情境进入，但须说明这些情境如何支持总体判断。

**比较对象：**生成式AI广泛普及与未广泛普及的现实情形，而不是“有AI”与“没有任何工具”。讨论立足当下及有依据的近期发展，不预设AI永不出错、风险必然发生，或完美监管已经实现。

### “自由”至少包含三个需要区分的问题

| 观察角度      | 需要解释什么？                                        | 不能直接跳过哪一步？                    |
|-----------|------------------------------------------------|-------------------------------|
| 实际能力与机会   | 我是否获得原本没有、而且真正能够实现的选择？资源与工具如何转化为机会？<br>〔01、07〕 | 功能更多、产出更快，不自动等于我的生活选择更多。      |
| 自主判断      | 我能否理解、反思、修正或拒绝建议？行动是否仍与自己认可的理由相连？〔02、05〕       | 受到外界影响，不自动等于被操控；点击确认也不自动证明自主。 |
| 社会条件与支配关系 | 谁设定工作与使用规则？谁得到收益？个人能否负担拒绝、更换或退出的代价？<br>〔07、08〕 | 个人完成任务更有效率，不自动说明他更有权决定节奏与收益。  |

### 三个角度要相互比较，而不是各讲各的

依靠眼镜、教材或同伴并不天然损害自由；同样，依靠AI也必须分析具体联系。《心灵延伸论》挑战以身体边界排除认知的做法，密尔强调亲自选择对人的成长的意义，罗贝恩斯区分资源与真实机会。它们提出不同的判断工具，并不直接给出本题的胜负。〔01、02、07〕

设想AI让员工半天完成原来一天的任务：他可能获得半天自主时间，也可能被要求完成两倍工作。这是用于分清问题的**假设情境**，不是本包报告的一项研究结果。参赛者需要找出，在现实中哪些条件使其中一种变化更可能发生。

### 共同边界

本题不是决定论意义上的“自由意志”问题。传统推荐算法、数字监控等可以作为类比，但不能不作说明就把它的一切影响算到生成式AI头上。也不把一般AI辅助、专门设计的辅导和整份代做当成同一个处理条件。

**最重要的比较：**AI增加了哪些能力，改变了哪些判断活动，又怎样改变落实选择的条件？只说“更方便”或“更依赖”，双方都还没有完成自由判断。

## 共同题解 · 02

# 双方要证明什么，证据又能走多远？

## 正方 | 总体上更不自由

可以承认AI带来便利与能力，仍论证新的依赖、隐蔽影响或社会约束，削弱了理解、拒绝和自主行动的条件。

**需要补足：**这种依赖为何不只是正常分工？损失影响哪些人、通过什么机制发生，为什么足以超过新增选择？

## 反方 | 总体上更自由

可以承认错误、依赖与不平等，仍论证工具扩展认知、学习支持与发展机会，使更多人能够实现自己的目标。

**需要补足：**能力如何成为现实选择？收益是否由使用者支配，为什么足以超过相应的自主性与条件损失？

双方均须支持自己的方向性判断，不要求证明每个人、每种用途都同向变化。说明了好处或风险“存在”，还要比较其范围、强度、持续性与发生条件。制度与教育方案可以讨论，但不能只凭“可以改进”就假设它必然实现。

## 五项研究：测到的变化，与尚未回答的问题

| 材料与情境                           | 读前需要知道的发现                                            | 不要直接推出                   |
|---------------------------------|------------------------------------------------------|--------------------------|
| <b>03 高中数学</b><br>Bastani等      | 普通AI辅助提升练习表现，却在后续独立测验中出现损失；保护型辅导基本消除该损失，未显示相应独立测验收益。 | 所有AI使用都会使人退化。            |
| <b>04 大学物理</b><br>Kestin等       | 专门设计的AI辅导在所测课程内容上取得更好的学习效果；194名学生纳入分析。               | 普通聊天机器人优于所有教师，或长期自主已获证实。 |
| <b>06 网络辩论</b><br>Salvi等        | 个性化AI相对人类基线更有说服力；2026年更正后，它与非个性化AI的直接比较未达通常显著阈值。     | 被说服就是被操控；个性化额外优势已被确定。    |
| <b>09 客服工作</b><br>Brynjolfsson等 | 企业内分阶段部署显示工作表现改善；系统暂时不可用时的表现也为学习提供线索。新手收益更明显。        | 工资、就业、空闲时间或议价能力必然提高。     |
| <b>10 短篇创作</b><br>Doshi与Hauser  | AI构思使个体作品评价提高，同时不同作品更相似；收益因前测创造力而异。                  | 个人表达收益虚假，或整个社会已经失去文化多样性。 |

## 怎样把理论与实践接起来？

理论说明什么值得计入自由，以及为何重要；研究帮助检验某种变化是否发生。没有一项研究直接测出“人总体上自由了多少”。现有材料来自特定教育、工作和实验情境，不是中国中学生或全社会的代表性调查；将结论迁移到别处，需要说明相同与不同的条件。

**阅读的目标不是选一句支持自己的话。**保留研究的反面结果、限制与未知，也认真回应对方最有力的解释。引用作者不能代替补上从事实到价值、从局部到总体的论证。

核心阅读

01

精选材料

# 克拉克与查默斯：心灵的边界

《心灵延伸论》· The Extended Mind (1998)

理论 · 哲学 | 独立论证概要

当一个人借助外部工具完成思考时，工具只是思考之后的输出渠道，还是可能成为思考过程本身的一部分？先理解作者的边界论证，再判断它能否用于生成式AI。

## 作者的论证 | 不以头骨预先划界

克拉克与查默斯先让读者比较三种完成图形旋转任务的方式：在脑中想象图形转动，借助电脑按钮转动屏幕上的图形，以及使用一个假想的神经植入装置加快旋转。三者实现相近的认知功能。如果我们愿意把植入装置中的操作算作认知，只因为电脑在身体外面就排除它，便把有待论证的边界当成了答案。作者要挑战的不是大脑的重要性，而是“认知必然止于皮肤与头骨”这一先入之见。

他们提出的比较方法常被称为“对等原则”：如果外部世界中的某个过程，假如发生在头脑内部，我们会毫不犹豫地视为认知的一部分，那么在其他相关条件相当时，就不该仅凭它的位置把它排除。用纸笔做复杂运算、移动字母片寻找词语，都可能是帮助形成答案的操作，而不仅是把已经想好的答案记录下来。人的行动改变外部表征，改变后的表征又影响下一步行动。

不过，作者并不是把一切影响人的事物都归入心灵。他们关注的是人和环境中某些要素形成的、相互作用的耦合系统：各部分共同承担一项认知任务。一个偶尔干扰注意力的声音，对思考有因果影响，却不因此成为思考系统的组成部分。区分“影响认知”与“参与构成认知”，是这套理论必须面对的难题。

这一观点也不要求外部过程都进入意识。许多脑内过程同样不为我们直接察觉，因而“我没有意识到它如何计算”不能单独作为排除外部过程的理由。至于外部工具会损坏、遗失、暂时不可用，作者提醒我们：生物记忆也会失效。应当比较实际的连接方式、稳定性和功能，而不是要求外部工具具备脑内过程并不具备的完美可靠性。

### 概念提醒 | 构成，不只是帮助

“使用工具能改善表现”是一种效果判断；“工具成为认知的一部分”是一种关于认知由什么构成的判断。证明前者，不等于已经证明后者；承认后者，也不自动决定怎样的依赖最有教育价值。

### 来源定位

本页与下页为编者依据原文 § 5 2-5 撰写的独立概要，不是原文译段。作者的讨论先于当代生成式AI；下页对AI的联系另行标注。

【01】未核实适用的开放翻译转载许可；本包采用独立概要，不复制原文。

原始来源 ↗ 完整出处（末页）

核心阅读

01

续读

# 一册笔记本，能否属于一个人的心灵？

Otto与Inga · 从外部记忆到可靠联系

理论 · 哲学 | 独立论证概要（续）

心灵延伸并不意味着“网上能查到的知识都是我知道的”。作者通过一个具体例子，考察外部信息与个人认知生活之间需要怎样的联系。

## 作者的论证 | Otto的笔记本

Inga打算去博物馆，想起地址之后便动身。Otto患有记忆障碍，依靠随身携带的笔记本保存信息；他查到其中记录的地址，也动身去同一地点。作者认为，如果笔记在Otto的日常生活中持续承担类似记忆的角色，仅因信息保存在纸上，而不是神经组织里，就否认它可以构成他的信念，理由并不充分。笔记不是每次都要重新证明的外来消息，而是他稳定依靠的一套记忆实践。

作者讨论了这个案例中几项重要特征：笔记本常伴左右，在相关情境中会被调用；其中的信息容易取得；取出的信息通常被直接认可，而不是每次重新核验；信息过去可能经过本人有意识的认可。这些特征并非一个已经解决所有边界争议的机械清单，作者对最后一项的必要性尤其保留疑问，但它们说明了稳定依赖为何不同于偶然接触。

有人可能把笔记看作普通工具，认为没有必要将其归入心灵。作者的回应是，日常解释行为时，我们本来就依靠信念和认知的功能来理解行动；对Otto而言，不承认这些稳定的外部记忆，反而会使解释变得曲折。与此同时，这并不意味着互联网上所有可搜索的信息都已经是某个人的信念。能否在正确时刻可靠调用，以及当事人如何把信息纳入生活，是需要另行分析的条件。

## 联系生成式AI | 以下为编者分析

这套理论为“外包一定削弱人”的说法提出挑战：可靠的工具联系可能使一个人真正拥有更广的行动能力。反方仍需说明这种联系在AI使用中如何形成。相反，AI会生成新的、可能错误的内容，提供者还可能改变功能与使用条件；不能把Otto对自己笔记的通常信任，直接变成学生应当默认相信模型输出的要求。

正方也不能仅凭“能力在脑外”判定人不自由，而应说明哪种依赖降低了检验、调整或拒绝的能力。两边要共同回答：这个人与工具组成的系统，是否仍由这个人有意义地设定目标、纠正错误并调节使用？这是本包提出的规范问题，不是原论文已经给出的AI政策结论。

### 辅助视频 | Inkstone

Episode 4: the Extended Mind · [点击观看](#)

用于帮助理解心灵边界与外部工具；不能替代原论文，也不作为独立实验证据。

【01】 未核实适用的开放翻译转载许可；本包采用独立概要，不复制原文。作者公开全文 § 2-5；重点为主动外在论、耦合系统与Otto案例。

原始来源 ↗ 完整出处（末页）

核心阅读

02

精选材料

## 密尔：选择为什么需要自己练习？

《论自由》第三章 · 1859

理论 · 哲学 | 完整论证段落 · 中文译文

密尔并不要求人拒绝经验与建议。他追问的是：遵从一条好道路，与自己理解、选择一条道路，是否对人的发展具有同样的意义？

### 原文选译 | 第1段

〔1〕然而，不论人们多么不习惯洪堡那样的学说，也不论他们发现个性被赋予如此高的价值时多么惊讶，我们仍不得不认为，问题只能在于程度。没有人关于行为卓越的观念，会要求人们除彼此模仿之外什么也不做。没有人会主张，人们不应在生活方式和处理事务时留下任何属于自己判断或个性特征的印记。另一方面，假装人们应当像在他们来到世上之前什么都不曾被认识过那样生活，仿佛经验尚未告诉我们某种生存或行为方式优于另一种，也同样荒谬。没有人否认，青年应受到教导和训练，从而知道并受益于人类经验中已经确立的成果。但是，当一个人的各种能力已经成熟，以自己的方式使用和解释经验，是他的特权，也是他应有的处境。他必须自己找出，记录下来的经验中哪一部分适合自己的境遇与性格。其他人的传统与习俗，在一定程度上，是他们从经验中学到什么的证据；这是一种推定性的证据，因此值得他尊重。但首先，他们的经验可能过于狭窄，或者他们没有正确解释经验。其次，他们对经验的解释即便正确，也可能不适合他。习俗是为惯常的情形和惯常的性格制定的，而他的境遇或性格可能并不寻常。第三，即使习俗作为习俗是好的，也适合他，仅仅因为它是习俗而遵从它，也不能教育或发展任何人类所特有的品质。人的观察力、判断力、辨别性的感受、心智活动，甚至道德偏好，都只有在作出选择时才得到运用。因为一件事是习俗就去做的人，并没有作出选择。他没有练习辨认或追求最好的东西。心智和道德的能力，如同肌肉力量，只有通过使用才能改善。仅因别人如此而行动，并不能运用这些能力，正如仅因别人相信就相信一样。如果一种意见的根据不能使当事人自己的理性信服，他采纳这种意见就不可能增强其理性，反而可能削弱它；如果一个行动的诱因并不符合他自己的感受和性格——在不涉及感情或他人权利时——那么，这就相应地使他的感受和性格变得惰性而迟钝，而不是活跃而有力量。

### 阅读提示

这是一整个自然段：“青年应向经验学习”与“仅仅服从习俗不能培养判断力”原本就在同一论证中。本包没有把后半段单独截成格言。洪堡是密尔在上文讨论个体发展的思想来源。

〔02〕 公版英文原著；本包自译，不摘用现有商业中译本。

原始来源 ↗ 完整出处（末页）

核心阅读

02

续读

## 好的结果，与成为怎样的人

完整自然段续读 · 第2-3段

理论 · 哲学 | 完整论证段落 · 中文译文

第2段紧接前页，第3段选自本章稍后的位置。原文涉及19世纪的用语与社会判断；保留作者论证，不代表本包赞同其所有历史表述。

### 原文选译 | 两段均完整

〔2〕让世界，或世界中自己所属的那一部分，替自己选择人生计划的人，除了类猿的模仿能力之外，不需要任何其他能力。自己选择人生计划的人，则要运用全部能力。他必须运用观察去看，运用推理和判断去预见，主动搜集作出决定的材料，运用辨别力作决定；决定之后，还要以坚定和自制坚持经过考虑的决定。在他的行为中，依照自己判断和感受决定的部分越多，他就越需要并越能运用这些品质。他可能无需具备这些品质，就被引导到某条良好的道路上，并且免于伤害。但这样一来，他作为一个人的相对价值如何呢？真正重要的，不只是人做了什么，也包括做这些事的人是怎样的人。在人类的作品中，在那些值得人用生命去完善、美化的事物中，最重要的无疑是人自身。假如房屋可以由机器建造，谷物可以由机器种植，战争可以由机器进行，案件可以由机器审理，甚至教堂也可以由人形自动机建造、祷告也可以由它们进行，那么，即便拿今天生活在世界较文明地区、而且确实只是自然能够并将会造就之人的贫乏样本的男女，去换取这些自动机，仍会是一种重大损失。人性不是一部照模型造好、再被安排去完成指定工作的机器，而是一棵树；它需要依照使它成为生命之物的内在力量的趋向，向各个方面生长和发展。

〔3〕人之所以成为一种高贵而美好的观照对象，并不是由于把自己身上一切个别的東西磨成整齐划一，而是在他人权利与利益规定的界限内，培养并发挥这些个别的東西；作品会带有创造者的性格，因此，同一过程也使人类生活变得丰富、多样而富有活力，为高尚思想和高远情感提供更充分的养料，并通过使人类更值得归属于其中，加强个人与全体之间的联系。每个人的个性发展得越充分，他对自己就越有价值，也就越可能对别人有价值。他自身的存在具有更充实的生命，而构成整体的个体生命越充实，整体也越充实。为了防止人性中较强的个体侵犯他人的权利，必要的约束是不能省去的；但即使从人的发展来看，这也能得到充分补偿。个人因不能以伤害他人的方式满足倾向而失去的发展手段，原本主要就是以他人的发展为代价获得的。而对他本人而言，对自私部分的约束，使其社会性的一面得到更好的发展，也构成了充分的补偿。为了他人而遵守严格的正义规则，能够发展以他人利益为对象的感情和能力；但在不影响他人利益的事情上，仅因他人的不悦而受到限制，除了在抵抗约束时可能产生的性格力量外，不会发展任何有价值的东西。如果顺从这种限制，整个人性都会变得迟钝。要使每个人的天性得到公平的发展空间，就必须允许不同的人过不同的生活。任何时代给予这种余地越多，就越值得后人记住。即使专制，只要个性仍存在于其中，就还没有产生最坏的后果；而任何压制个性的东西都是专制，不论它叫什么名字，也不论它声称执行的是上帝的意志还是人的命令。

### 共同追问 | 编者分析

AI可能代替判断，也可能帮助人看清选项。怎样区分“主动组织外部帮助”与“让别人替自己选择人生”？不能只凭最后是否亲手打字作判断。

〔02〕公版英文原著；本包自译，不摘用现有商业中译本。第三章3个完整自然段，段首为“Little, however...”、“He who lets...”、“It is not by wearing...”；第3段与前2段不连续。

原始来源 ↗ 完整出处（末页）

核心阅读

03

精选材料

## Bastani等：练习更顺利，就等于学会了吗？

高中数学现场研究 · PNAS (2025)

实践 · 教育 | 独立研究概要

同一项工具可以改善使用当下的表现，却对随后独立完成任务产生不同影响。这份研究把“现在做得好”与“之后学会了”分开观察。

### 研究对象与比较方式

研究在土耳其一所高中开展，涉及约一千名九至十一年级学生，围绕正常数学课程安排多次课堂活动。学生先接受教师讲解，再完成练习，随后参加不能使用AI的测验。作者比较三种学习条件：没有AI、可以使用一般式GPT辅助（GPT Base），以及使用带教学保护设计的GPT辅导（GPT Tutor）。这不是把最主动寻找AI的人与完全不用AI的人简单对比，而是以班级为实施单元、经过分配与平衡安排的现场实验。

GPT Base能够直接回应学生的问题，较容易给出完整解法。GPT Tutor则结合教师提供的题目解法和常见错误，受到专门指令约束：通过提示引导学生，而不是轻易交付最终答案。因此，二者不只在“有没有同一个模型”上不同；教师的前期设计和工具允许的互动方式，本来就是干预的一部分。不能把Tutor的结果当作任何聊天框都会产生的效果。

三个条件下的练习表现与后续测验分别记录。这个区分很重要：作业可以在工具帮助下完成，但教育目的也可能包括离开某种提示后，学生能否解释和解决问题。如果只观察练习分数，就无法区分答案被提供与学习被支持；如果只看无工具测验，又可能遗漏人与工具共同解决问题本身的价值。研究选择了其中可测量的部分，并没有替所有教育目标排好顺序。

### 主要发现 | 先把两个阶段分开

在可以使用AI的练习阶段，两种AI条件的表现都提高。论文报告，相对于无AI对照，GPT Base与GPT Tutor的练习成绩分别提高约48%与127%。这些是相对于对照水平的变化，不是提高48或127个百分点，也不是对“聪明程度”的测量。

到了不能使用AI的测验，一般式GPT辅助组的成绩比对照低约17%。带教学保护的Tutor基本消除了这项负面差异，但其独立测验成绩并未显示相对于对照的显著提升。不能把“不再观察到明显损害”写成“已经证明学得更多”。

#### 读数提醒

辅助表现、独立测验和长期能力是三个不同概念。本文直接观察前两者；不能把短期课程测验推成“AI让一代人永久失去思考能力”。

【03】 CC BY-NC-ND 4.0；本包为独立研究概要，不发布原文改编译作。

原始来源 ↗ 完整出处（末页）

核心阅读

03

续读

## 同样叫“使用AI”，学习过程可能不同

使用记录、机制解释与适用边界

实践 · 教育 | 独立研究概要（续）

这份研究的关键不只是一个负面数字，而是：即时成绩上升与独立表现下降可以并存，工具的教学约束也会改变这种关系。

### 作者怎样解释结果

作者结合学生与系统的互动记录，讨论学生把AI当作“拐杖”的可能：在一般辅助条件下，学生可以直接取得解题步骤，而不必充分尝试、检查和组织自己的推理。帮助完成了当前任务，却可能减少为了以后独立完成而需要的练习。日志与测验模式支持这种解释，但日志并不是另一个单独操纵“思考投入”的实验，因此不应把每个心理环节都说成已经被独立证实。

教学保护的意义在于改变帮助的形状。先问问题、提供渐进提示、针对常见错误回应，与给出完整答案，会让学生经历不同的活动。结果说明，在该研究中，“有没有AI”不是唯一重要区别。不过，保护型系统不等于完美保证：它需要设计投入，也可能仍给出过多帮助。研究没有显示只要增加一条“请启发学生”的提示，就可以复制全部效果。

学生对工具帮助的主观感受，也不能代替学习测量。任务更顺利、答案更容易得到，可能让人高估自己已经掌握的程度。作者据此提醒，在教育中采用生成式AI，需要关注学习后果而不仅是使用满意度；这是一种关于具体设计与评估的警告，不是宣布任何自动化帮助都违反教育原则。

### 证据能走到哪里？

实验集中在一所学校、特定年级和数学教学单元，并在较短时间范围内测量效果。无法由此断定在写作、语言学习或长期自主选题中也有同样大小的影响。它也没有直接测量“自由”。把结果用于本题，还需要说明：受影响的学习能力为何是实际选择或自主判断的重要条件。

这里的比较对象是没有AI的课堂练习，而不是资源最匮乏、完全没有任何辅导的学生。换到其他教育环境，新增帮助的价值与学习替代的风险可能改变。将研究用于中国中学生时，应说明课程、使用规则和支持条件之间的联系，不把地点差异忽略掉，也不因地点不同便自动否定证据。

### 连接两方 | 编者分析

正方据此检验“当前能力增加”的说法，反方可据此说明工具设计会改变后果。双方都要比较现实中哪些用法更可能普及、保护设计能否获得，而不是各自挑选最好或最差的版本。

【03】 CC BY-NC-ND 4.0；本包为独立研究概要，不发布原文改编译作。依据研究设计、练习与独立考试结果、使用记录及讨论；数据为论文中的特定实验。

原始来源 ↗ 完整出处（末页）

核心阅读

04

精选材料

## Kestin等：AI也可以支持真正的学习吗？

大学物理随机交叉实验 · Scientific Reports (2025)

实践 · 教育 | 独立研究概要

把AI当成答案机器，并不是唯一的教学安排。这项研究检验了一个经过教学设计的AI辅导系统，与主动学习课堂在具体课程内容上的效果。

### 比较的不是“机器人对传统说教”

研究在哈佛大学本科物理课程中进行。课程有233名注册学生，但依据同意参与、完成两种教学条件和相关测验等标准，纳入研究的参与者为194人。材料中应使用后一个数字描述分析样本，而不能把全班注册人数当作实验完成者。

研究采用随机交叉设计：学生被分组，在不同学习单元中分别经历AI辅导和课堂教学，使比较不只是由两批完全不同的学生构成。对照是包含互动、练习和反馈的主动学习课堂，而不是只听教师讲授的低互动课程。学生在学习前后接受测验，研究据此比较所选内容的学习收益，并调查学习感受和投入时间。

AI条件并非让学生随意打开普通聊天机器人。辅导系统围绕课程目标安排材料与问题，通过分步引导、及时反馈和适应性的交流支持学习。研究团队把学习科学的设计要求纳入系统，而不是仅要求模型“做一个好老师”。因此，所比较的是一套具体的教学安排，与另一套具体的教学安排。

### 研究观察到了什么

在研究覆盖的教学内容和短期测验中，AI辅导条件获得了更高的学习收益。学生也报告较积极的参与和学习体验。研究表明，经过设计的AI帮助可以不只是提高交付答案的速度，也可能促进学生在测验中显示出的理解。它为“工具必然使人跳过学习”的说法提供了反例。

学习时间也需要同口径比较：课堂约75分钟中包含前后测，实际教学时间约60分钟；AI辅导的中位学习时间约49分钟。不能直接把49分钟与包含测验的75分钟相比，夸大节约。部分学生使用AI的时间超过60分钟，所以中位数也不意味着每个人都学得更快。

#### 关键区别

随机分配支持的是本研究所比较教学条件的因果判断，而不是“所有AI优于所有教师”。课程设计、反馈方式与使用任务，都包含在被检验的方案内。

【04】 CC BY-NC-ND 4.0；本包为独立研究概要，不发布原文改编译作。

原始来源 ↗ 完整出处（末页）

核心阅读

04

续读

## 为什么这一结果不与上一项研究简单矛盾？

教学设计、迁移与长期自主性

实践 · 教育 | 独立研究概要（续）

理解研究，不是挑选一个支持立场的标题。需要先比较：谁在学、学什么、怎样使用工具，以及结果是在什么条件下测量的。

### 作者解释的重点

作者把较好的学习结果与系统设计联系起来：让学生参与解决问题而不是被动接收，及时回应错误，按可管理的步骤推进，并给予学生对节奏的控制。这些因素可能使AI承担高密度、个别化的教学支持。它们解释了为什么“聊天机器人能生成答案”与“一个AI辅导方案能帮助理解”不是同一命题。

不过，实验没有分别随机操纵系统中的每个设计要素，不能宣布某一个环节单独造成了全部收益。更准确的说法是，经过这一组合设计的AI教学条件，在所测内容上获得了较好的结果。把效果归因到具体组件，还需要进一步拆分比较。

对教学的意义也不是取消教师。课程内容、问题组织、系统设计与结果评价都需要人的工作。研究检验的是某些教学活动能否以不同方式实施，并没有比较一所完全没有教师的学校。把“某个单元可以更有效地辅导”扩展成“教育不再需要人”，远远超出了证据。

### 与Bastani等研究合读

两项研究的学生年龄、学科内容、对照教学、工具约束和测验安排不同。它们共同提醒我们，不应把所有AI使用都合并成一个固定处理。Bastani等重点揭示一般辅助与保护型辅导在独立测验上的差别；Kestin等则展示一种专门设计的AI辅导可以产生学习收益。这两点可以同时成立。

仍然需要谨慎的是长期迁移。短期后测进步，不直接证明数月后仍能保持，也不证明在陌生任务、没有同一系统或生活决策中能够更自主。另一方面，没有测量这些远期结果，也不等于研究已经证明它们不会改善。对未知结果，应明确保留，而不是按持方填空。

#### 连接自由 | 编者分析

反方须说明学到的能力如何形成真实机会，正方须回应有助于能力形成的AI用法，而不能只谈代写。值得比较的是：可获得的教学支持能否让更多人理解、提问和纠错，以及这种支持的条件是否现实。

【04】 CC BY-NC-ND 4.0；本包为独立研究概要，不发布原文改编译作。依据教学设计、随机交叉实验、学习与时间结果、局限；194人为纳入分析的参与者。

原始来源 ↗ 完整出处（末页）

核心阅读

05

精选材料

## Susser等：有影响，就等于有操控吗？

Technology, autonomy, and manipulation (2019)

理论 · 哲学 | 完整论证段落 · 中文译文

这一框架区分理性说服、强制与隐蔽操控。它讨论数字环境中的自主性，但并不是一项针对今天聊天机器人的效果实验。

### 原文选译 | § 1中的两个连续自然段

〔1〕说服——在理性说服的意义上——是试图通过提供对方能够思考和评价的理由，来影响对方。

〔原注10〕强制则是通过限制一个人的选项来影响他，使其唯一理性的行动途径就是强制者所希望的那一种（Wood，2014）。说服与强制带有很不相同、甚至近乎相反的规范含义：说服某人做某事几乎总是可以接受的，而强制几乎总不是如此。然而，两者的共同点在于，它们都是公开的影响形式。当有人试图说服或强制我们时，我们通常知道这一点。相较之下，操控是隐蔽的——我们只有事后才知道有人曾试图引导我们的决定，甚至可能永远不会发现。

〔2〕因此，操控的独特之处在于：当我们知道自己受到操控时，会觉得被人耍弄了。〔原注11〕回过头思考自己为什么那样行动，我们意识到，在作决定的当时，并不理解自己的动机。我们像木偶一样，被操纵木偶的人牵引。操控因而扰乱了我们成为自己人生作者的能力——它僭越地替我们决定应当怎样生活，以及为什么那样生活。如下文所论，这会引出一组特定的伤害。眼下重要的是看清这里涉及哪种影响。说服和强制公开面对其对象，操控则隐蔽地进行。受到强制时，我们通常有理由感到愤怒，但愤怒针对的是施加给我们的限制。相反，受到操控时，我们没有被限制；而是在意识不到的情况下，被引向我们无法辨认的行动理由，以及我们可能希望避开的目标。

### 读懂作者的判准

在这一解释中，问题不只在于我有没有被迫接受某个选项，而在于影响是否有意隐蔽、是否利用了我难以察觉的决策脆弱性。作者的定义是有争论的理论主张，并非所有学者对“操控”的唯一用法。

### 边界检验 | 编者分析

给出清楚的理由，使人重新评价原来的看法，可以是对自主的支持。隐藏自己的利益、利用当事人无法识别的弱点来安排选择，则需要不同的评价。仅凭AI最终改变了一个人的意见，还无法判断是哪一种。

〔05〕 CC BY 3.0 Germany；本包中文翻译，已作语言转换。

原始来源 ↗ 完整出处（末页） 许可 ↗

核心阅读

05

续读

## 自主的人，并不生活在真空中

§ 2：能力、认同与社会条件

理论 · 哲学 | 完整论证段落 · 中文译文

作者既批评操控，也拒绝“完全不受外界影响才算自主”的要求。把这两部分放在一起，才能避免把任何依赖都自动判为不自由。

### 原文选译 | 第3–4段；与前页不连续

〔3〕哲学家有时通过区分自主性的能力条件与真实性条件，把这一概念具体化（J. P. Christman, 2009, 第155页起）。首先，自主意味着具备认知、心理、社会和情感能力，能够思考自己的选择、形成相关意图，并依此行动。其次，这意味着经过批判性反思，一个人认同自己的价值、欲望和目标，真诚地认可它们是自己的。当然，许多人批评这种自主观过于理性主义，要求高得不可信，认为我们很少这样作决定。他们说，我们是有情感、依赖习惯的行动者，在社会化和文化塑造中形成了特定的选择方式，却几乎从不反思或认可这些方式。不过，我们广义地理解自主性：我们所说的审议，不仅包括信念与欲望，也包括情感、确信和经验；批判性反思还可以是反事实意义上的——原则上，我们必须能够批判性地反思并认可自己的行动动机，但不必真的对自己的每一步行动都加以反思。

〔4〕除了拒绝要求过高、过于理性主义的自主观，我们也拒绝过于原子化的自主观。依我们看，自主的人始终处于社会、文化、历史和政治之中。换言之，我们承认“自我与身份认同中，对个人自主及道德、政治能动性而言具有重要性的主体间与社会维度”（Mackenzie与Stoljar, 2000, 第4页）。

〔原注22〕社会情境可能通过塑造刻板信念与行为来限制选择，例如性别化的社会期待；但也是社会情境赋予自主性以价值，教我们什么是独立决定，并向我们提供可供选择的丰富选项。

### 怎样与前一页连起来？

一个人可以依靠他人的知识、在情感与习惯中生活，同时仍保有理解、反思和认可自己行动的能力。作者反对的是扰乱这些条件的影响，而不是世界与个人之间的一切联系。

### 联系生成式AI | 编者分析

正方需要说明具体AI实践如何妨碍认出或检验行动理由，而不能仅说“建议来自外部”。反方需要说明系统如何帮助使用者理解与决定，而不能仅说“用户最后点了确认”。形式上的点击、主观满意以及事后的理由认可，并不是完全相同的指标。

同样，社会支持可以扩大选择，社会压力也可以压缩选择。争议不只在模型内部，而在于谁设定目标、使用者知道多少、能否更换或拒绝，以及拒绝是否要付出不可承受的代价。此处为本包联系，不是作者已经检验了所有生成式AI场景。

【05】 CC BY 3.0 Germany；本包中文翻译，已作语言转换。§ § 1–2的4个完整自然段：“Persuasion...”、“What makes...”、“Philosophers sometimes...”、“In addition to rejecting...”；两组不连续。

原始来源 ↗ 完整出处（末页） 许可 ↗

核心阅读

06

精选材料

## Salvi等：AI可以多大程度改变人的看法？

On the conversational persuasiveness of GPT-4

实践 · 社会心理 | 完整论证段落 · 中文译文

研究比较人在结构化网络辩论中面对人类或GPT-4时的意见变化，并区分对手是否获得个人信息。它研究说服效果，不直接测量“自主性减少了多少”。

### 研究背景 | 编者导读

最终样本为900名美国成年参与者。实验随机安排对手类型、是否提供个人信息以及议题强度，在辩论前后测量对所分配立场的认同。下文选取的是作者解释结果与限制的正文，不是摘要。

### 原文选译 | Discussion；已纳入作者更正

〔1〕我们的研究表明，对个性化与AI说服的担忧是有根据的：它通过展示大语言模型如何借助微定向在网络对话中比人类更有说服力，强化了先前研究的结果〔原文献42、44、45、48〕。我们注意到，尽管“人—AI（个性化）”条件显著优于“人—人”基线，但“人—AI（个性化）”与“人—AI”两种条件的直接比较并未达到统计显著（ $P=0.07$ ），可能是因为成对处理比较的统计功效有限。因此，应谨慎地将观察到的优势完全归于个性化，而不是GPT-4一般的说服能力。尽管如此，考虑到我们收集的个人信息很少——性别、年龄、族裔、教育水平、就业状况和政治倾向——以及要求模型使用这些信息的提示词极为简单，这一效果仍特别值得注意（完整提示词见补充材料第2.5节）。利用个性特征与道德基础等个人心理属性，或通过提示工程、微调、专门领域知识设计更强的提示词，可能得到更强的效果。在这一背景下，有意部署聊天机器人进行大规模虚假信息活动的恶意行动者，可能利用细致的数字痕迹和行为数据，建立能适应个人目标的复杂说服机器。

〔2〕我们主张，网络平台和社交媒体应认真考虑此类威胁，并加大实施措施的力度，以应对AI驱动的说服传播。一条有希望的应对大规模虚假信息活动的路径，可能由大语言模型自身提供：生成同样个性化的反向叙事，帮助那些可能受到欺骗性内容影响的旁观者〔原文献27、58〕。这一方向已有初步尝试，通过与GPT-4对话减少阴谋论信念，得到了有希望的结果〔原文献59〕。我们的分析也初步揭示了大语言模型说服背后的机制。我们发现，GPT-4与人类辩手的写作风格存在明显差异。例如，大语言模型辩手生成的文本更难阅读，也包含更多与逻辑和分析性推理有关的语言标记（图4）。

### 读数提醒

“个性化AI显著优于人类基线”与“个性化显著优于非个性化AI”是两次不同的比较。前者成立，不自动使后者成立。段末更强个性化攻击的描述是作者提出的可能风险，不是本实验已经测得的效果。

〔06〕 CC BY 4.0；中文选译已纳入2026-09-03作者更正。

原始来源 ↗ 完整出处（末页） 许可 ↗ 作者更正 ↗

核心阅读

06

续读

## 说服效果，不等于现实中的操控结论

生态效度、样本边界与2026年更正

实践 · 社会心理 | 完整论证段落 · 中文译文

上一页保留了作者对风险的担忧，也保留了其提出的AI反制用途。这里继续阅读他们对结果外推范围的限制。

### 原文选译 | 第3-4段；非连续选段

〔3〕第二，我们的实验设计要求对话具有预先规定的结构，严格遵循辩论的阶段与规则。我们认为，这一设置捕捉了许多网络互动的基本特征：人们近乎同步地相互回复，或实时回应他人的评论；但它仍是一种人工环境，与自发、难以预测地发展的网络对话，可能存在显著差异。此外，我们平台上的对话完全匿名，不同于人们通常的互动条件。因此，我们承认研究结果的生态效度有限，因为尚不清楚它们能否推广到社交网络和其他网络平台上的自然讨论。基于这一原因，我们的工作应被视为对大语言模型说服能力的一次概念验证，而不是对其在真实开放环境中说服力的现实评估；后一个问题仍需未来研究。

〔4〕第四，我们的实验邀请了从Prolific招募的人类参与者。他们因完成辩论而获得报酬，也知道自己处于受控的实验环境中。虽然先前研究发现，与竞争平台相比，Prolific的数据质量最好，而且基于Prolific的研究经常具有可推广性〔原文献61、62、63〕，但平台活跃工作者的社会人口构成，仍不同于美国总体人口，也不同于其他网络平台与社交媒体的用户群。因此，还需要研究检验：使用一个更具代表性、准确反映人类说服技能整体范围的样本，能否重复这些发现。此外，在比较中加入人类专家，例如从事竞技辩论、政治竞选或公共传播的人，也会很有意义。

### 版本核对 | 2026年9月3日作者更正

原分析中成对比较的参照类别有误。更正后的“个性化AI”与“非个性化AI”直接比较为 $P=0.0678$ ，正文四舍五入为0.07，不再达到通常的0.05阈值。本包第1段采用已更正表述。不能继续沿用旧版本“ $P=0.04$ ”的结论。

这也不等于证明两种AI完全没有差别；“未达到显著”与“证明效果相同”不同。更正并未取消个性化AI相对于人类基线的显著结果。点击核对作者更正。

### 连接自由 | 编者分析

要从这项研究走向自由判断，还需解释意见改变的原因：是理解了更充分的理由，还是难以识别的影响绕过了判断？下一步也要检验，这种变化能否保持、能否转化为行动。该实验没有直接回答这些问题。正方不能把“更能说服”自动写成“更能操控”；反方也不能仅凭参与者自愿讨论，便断言不存在自主性风险。

〔06〕 CC BY 4.0；中文选译已纳入2026-09-03作者更正。Discussion中4个完整自然段：“Our study suggests...”、“We argue...”、“Second, our experimental...”、“Fourth, our experiment...”；非连续。

原始来源 ↗ 完整出处（末页） 许可 ↗ 作者更正 ↗

核心阅读

07

精选材料

## 罗贝恩斯：资源不是自由的终点

Wellbeing, Freedom and Social Justice (2017)

理论 · 哲学 | 完整论证段落 · 中文译文

一个免费账号、一套知识或一辆自行车，都可以是重要资源。但评价人的自由，还需要知道：他是否真的能够把资源转化为自己有理由选择的生活。

### 先分清两个词 | 编者导读

“功能实现”（functioning）指已经做到或成为的状态；“可行能力”（capability）指真正可选择、可实现的机会，不等同于狭义技能。有机会出行而选择留在家，与想出行却没有条件，并不是同一种处境。

### 原文选译 | § 2.6.3第1-2自然段

〔1〕可行能力方法的第三个核心观念是：不同的人，把资源转化为功能实现的能力不同。决定一个人能在多大程度上把资源转化为功能实现的因素，被称为转化因素。这一直是阿马蒂亚·森的可行能力方法中的一个重要观念（Sen, 1992a, 第19-21、26-30、37-38页），对受其著作影响的学者也同样重要。资源具有某些特征，使它们对人们具有吸引力；资源既包括可以交易的商品与服务，也包括非市场经济中产生的商品与服务，例如家庭生产。在森的福利经济学研究中，“资源”这一概念限于物质性和／或可衡量的资源，尤其是金钱或消费品；但我们也可以把转化因素这一概念应用到对资源更宽泛的理解上，例如一个人拥有的学历。

〔2〕自行车经常被用来说明转化因素这一观念。我们对自行车感兴趣，主要不是因为它由某些材料制成，具有特定形状和颜色，而是因为它能把我们带到想去的地方，而且比步行更快。商品的这些特征使一种功能实现成为可能，或有助于它的实现。自行车使移动成为可能，即能够自由地移动自己，并且比步行更快。但由于相关转化因素不同，一个人把这一资源转化为有价值的功能实现的程度，可能与别人不同。例如，一个身体健康、童年时学会骑车的人，在把自行车转化为高效出行能力方面具有较高的转化因素；而有身体障碍的人，或从未学会骑车的人，则具有很低的转化因素。因此，转化因素表示一个人能从某种资源中获得多少功能实现；在这个例子里，就是一个人能够从自行车中获得多少移动能力。

### 回到资源与人之间

作者在未选入的分类段中区分个人、社会与环境因素。本包不把这些因素当成永远不变的命运：下页将保留她关于改变条件的完整论证。

〔07〕 CC BY 4.0；本包中文翻译，已作语言转换。

原始来源 ↗ 完整出处（末页） 许可 ↗

核心阅读

07

续读

## 改变条件，才能扩大真正的选择

转化因素可以改变 · 手段与目的必须区分

理论 · 哲学 | 完整论证段落 · 中文译文

同一套理论可以解释技术为何不能自动带来自由，也可以指出怎样的技术与制度安排能够扩展自由。它并不预先站在乐观或悲观的一边。

### 原文选译 | 第3-4段；非连续选段

〔3〕请注意，许多转化因素并非固定不变或天然给定，而是能够通过我们的政策与选择改变。某个转化因素的效果，也可能取决于一个人拥有的社会与个人资源，以及其他转化因素。例如，如果一个人因身体障碍不能行走，又没有轮椅可用，道路状况糟糕、公共交通工具也不便轮椅使用，那么这种障碍就会严重限制其移动的可行能力。但是，假设建成环境不同：所有行走障碍者都有权获得轮椅，道路对轮椅友好，公共交通便于轮椅使用，社会还具有这样一套规范——人们认为，帮助不能行走的同行者是再自然不过的事。在这一替代的社会状态中，由于资源以及社会、环境转化因素不同，同一个人转化因素——不能行走——就产生了很不相同的影响。总之，为了知道人们能够做什么、能够成为什么，我们需要分析他们的资源和各种转化因素的全貌，或者直接分析其功能实现与可行能力。清楚知道需要哪些资源、需要哪些具体转化因素的好处是：这也为那些希望扩大可行能力集合的人，提供了可以在哪里进行干预的信息。

〔4〕可行能力方法的第四个核心特征，是手段与目的的区分。这一方法强调，在评价某物时，我们应始终弄清：我们是把它作为目的本身来重视，还是把它作为通往有价值目的的手段来重视。对于可行能力方法而言，在对不同人的有利处境作比较时，最终目的，是人们有价值的可行能力（也可能存在其他目的，见第2.6.6节）。这意味着，可行能力方法要求我们根据政策及其他变化对人的可行能力与实际功能实现的影响来评价它们；与此同时，我们需要追问，这些可行能力的前提——手段和有利条件——是否已经具备。我们必须问，人们是否能够健康，以及这种可行能力所需的手段或资源是否存在，例如清洁的水、适当的卫生设施、医疗服务、防护感染与疾病的措施，以及基本健康知识。我们必须问，人们是否营养充足，以及实现这种可行能力所需的手段或条件，例如充足的食物供应和获得食物的权利，是否得到满足。我们必须问，人们能否进入高质量的教育体系，能否真正参与政治，能否参与提供支持、帮助他们应对日常生活困难、促进关怀性友谊的社区活动。因此，我们确实需要把手段纳入考虑，但只有先知道目的是什么，才能这样做。

### 连接生成式AI | 编者分析

反方可以检验：AI是否降低语言、知识或专业支持的门槛，从而改变转化条件？正方可以检验：账号虽多，使用者是否仍缺少核验能力、工作机会或退出条件？双方都要指明具体环节，不能把“人人可用”直接等同于“人人更自由”，也不能把“仍有差异”直接等同于“技术毫无改善”。

〔07〕 CC BY 4.0；本包中文翻译，已作语言转换。 § 2.6.3的第1、2、4自然段及 § 2.6.4首段，共4个完整段；非连续，正文依据原书及授权在线版本。

原始来源 ↗ 完整出处（末页） 许可 ↗

核心阅读

08

精选材料

## 阿西莫格鲁、奥托与约翰逊：效率属于谁？

Can We Have Pro-Worker AI? (2023)

理论 · 社会科学 | 独立论证概要

一种技术可以节省劳动，也可以创造新的劳动任务。作者关心的不是机器是否越来越强，而是技术发展方向如何改变人能做的工作，以及谁获得由此产生的收益。

### 作者的论证 | 自动化不是完整的繁荣机制

这份政策备忘录区分两条并不互斥、但社会后果可能不同的路径。自动化把原本由劳动者承担的任务交给机器；补充人的能力，则帮助劳动者提高质量、处理陌生问题，或承担过去无法完成的新任务。两者都可能提高生产率，但对劳动需求、工资和劳动者地位的作用不必相同。某个软件同时包含两种作用，也需要分析各自的重要性。

作者反对一种过于简单的推论：只要生产率提高，所有劳动者最终都会同等获益。自动化可能减少某些任务对人的需要，即使企业收入或经济总产出增长，失去原岗位的人也未必能立即进入新增岗位。新机会需要的技能、地点与进入条件，可能和受影响劳动者原有条件不匹配。因此，整体收益不能直接抵消每个具体人的损失。

他们也质疑一种看似平等化的说法：如果AI取代了部分高薪专业工作，收入差距就会缩小，从而使劳动者受益。被挤出的劳动者可能转向其他岗位，增加对原有低薪工作的竞争。作者认为，更有价值的平等化路径，是使原本处于较低位置的人能够承担更有价值的工作，而不只是拆掉职业阶梯上较高的台阶。

备忘录把技术变化放在历史分工中理解：自动化与新任务创造的相对速度，会影响劳动者能否分享技术收益。作者据已有经济研究提出，自20世纪后期以来，这种平衡对许多没有大学学位的劳动者不利。这里是基于历史研究形成的理论判断，不是该备忘录新做的一项实验，更不是证明未来所有AI部署必然复制过去。

### 从产量到自由 | 编者提示

若AI让一份工作更快完成，员工是否获得时间、选择权或议价能力，取决于工作要求与收益安排。反过来，不能因为收益分配需要制度条件，就否认效率改进可能提供扩大选择的资源。

### 材料类型说明

政策备忘录提出解释框架和可选择的发展路径。其举例、引用研究与政策建议，应分别阅读，不应统称为作者已经“实验证明”的结论。

【08】 未核实适用的开放翻译转载许可；本包采用独立概要，不复制原文。

原始来源 ↗ 完整出处（末页）

核心阅读

08

续读

## 把机器的力量，变成人的能力

补充路径、组织选择与现实条件

理论 · 社会科学 | 独立论证概要（续）

作者并非主张停止发展AI。他们提出有条件的乐观：技术可以帮助更多劳动者运用知识、解决问题，但这条路径需要被实际选择和建设。

### 作者的论证 | 人机互补的可能

信息多，并不等于在正确时刻得到有用的信息。备忘录指出，生成式AI可能帮助劳动者从大量资料中找出与当前任务相关的内容，补足专门知识，并为决策提供支持。作者举的例子包括：电工面对少见故障时获得诊断线索；受语言限制的熟练移民得到沟通帮助。这里的目标不是把人变成机器的附属，而是使人能够运用和扩展已有专长。

这种安排也可能重新组织职业。部分操作步骤可以自动完成，而人承担更复杂的判断、监督、客户交流和服务设计。作者引用编程、写作和客服研究作为可能性的例证：自动化与增强可以同时发生，关键在于人是否仍运用判断，把AI的中间建议转化为合适成果。本文引用的客服研究与本包材料09属于同一研究项目，不能重复算作两项独立证据。

教育是作者重点讨论的场景。个别化辅导可以帮助准备程度不同的学生，但设计工具、发展课程和维护学习质量，也可能为教师创造新任务。另一方面，如果学校只把AI视为减少人力成本的办法，它也可能被用来削减教师与个别关注。两种路径都涉及同一类技术，其分歧不只由模型能力决定，也由预算、激励和决策方式决定。

作者因此主张，把“有利于劳动者的AI”视为需要引导的发展方向，而不是技术进步自动附带的结果。让工具补充人的技能，鼓励能创造有价值新任务的应用，发展相应训练与公共能力，都是这一方向的组成部分。不能把这些建议写成已经普遍实现的事实；也不能把当前激励的批评改写成技术本身毫无改善可能。

### 让两方在同一个问题上比较

正方可以要求解释：在现实激励下，替代、标准化与控制为何可能超过能力补充；反方可以要求解释：降低知识门槛、提高服务能力和创造任务为何可能成为主要效果。两边都需要说明发生条件、涉及人群与改变的可能性。只说“制度有问题”，或者“技术本身中立”，都还没有完成比较。

进一步需要分清个人与整体：一个员工获得帮助，不必然代表劳动市场中的机会整体增加；某些岗位减少，也不自动代表所有人的选择缩小。社会科学的工作，是把这些关系说明，而不是用宏观词语替代具体机制。

【08】 未核实适用的开放翻译转载许可；本包采用独立概要，不复制原文。围绕自动化、补充人的能力、劳动需求与技术选择；保留作者提出的替代路径与制度条件。

原始来源 ↗ 完整出处（末页）

核心阅读

09

精选材料

## Brynjolfsson等：帮助可以留下能力吗？

Generative AI at Work · QJE (2025)

实践 · 劳动经济 | 完整论证段落 · 中文译文

研究观察AI建议系统在一家企业客服工作中的分阶段部署，涉及5,172名客服人员。整体生产率提高约15%，收益对新手和较低技能员工更明显；这里继续追问这些收益如何产生。

### 研究设计 | 编者导读

这是基于分阶段上线与工作记录的现场研究，不是随机对照试验。作者控制员工、时间与任职经验等因素，并利用系统故障时段检验学习。故障分析主要测量对话时长，不能直接等同于服务质量。

### 原文选译 | § VI.B中3个完整段；非连续

〔1〕截至目前的发现提出了一个关键问题：这些生产率提升和沟通模式变化，反映的是劳动者人力资本的持久变化，还是仅仅反映他们对AI帮助日益增加的依赖？在后一种情形中，引入AI帮助实际上可能导致人力资本受损；那么，我们会预期，接受过AI帮助的劳动者在无法继续取得AI时，更难处理客户的问题。例如，认知科学研究发现，与使用传统地图相比，遵循GPS指示的人对空间导航学到的东西较少（Brügger、Richter与Fabrikant，2019）。

〔2〕在B面板中，我们使用相同的处理前观察值，但把采用AI之后的时段，限制为受到重大故障影响的时段。我们发现，即使在AI系统不能工作的故障期间，接触过AI的客服人员仍然能更快处理咨询，相当于对话时长减少15%至25%。由于AI故障较少，我们的估计噪声较大，也可能反映故障时段与非故障时段所遇对话类型的差异。然而，如果关注估计效果随时间变化的大小，就会出现一个有意思的模式。非故障时段的A面板显示，采用AI后，对话时长立即下降并大体保持稳定；与此不同，B面板显示，在故障时段，接触AI帮助的收益会随时间增加。也就是说，如果故障发生在采用AI一个月后，劳动者处理对话的速度，相对于使用前的基线，并没有快很多。如果故障发生在接触AI建议三个月之后，劳动者处理对话就会更快——尽管这两种情形下，他们都没有正在接收直接的AI帮助。

〔3〕B面板凸显了改善现有员工培训方式的潜在空间。在AI帮助出现之前，培训限于每周简短的辅导：管理者回顾部分对话并提供反馈。受客观条件限制，管理者只能对客服人员进行的少量对话提供反馈。此外，管理者常常时间不足，也可能缺乏培训，因此他们往往只是指出不理想的指标——“你需要减少处理时间”——而不是指出员工怎样能更好地处理问题的策略——“你需要在一开始多问些问题，以便更好地诊断问题”。这种辅导可能无效，甚至不利于员工投入（Berg等，2018）。相比之下，AI帮助向劳动者提供具体、实时、可执行的建议，因此可能弥补传统辅导方式的一项局限。

### 术语与图表指向

“B面板”指原文图VII的故障时段分析；本包不重印该图。“人力资本”在此指劳动者掌握的知识、技能等能力，不是说人只具有经济价值。

〔09〕 CC BY-NC 4.0；本包非商业中文翻译，已作语言转换。

原始来源 ↗ 完整出处（末页） 许可 ↗

核心阅读

09

续读

## 同一项研究，也必须保留它的警告

过度依赖、收益分配与不能回答的问题

实践 · 劳动经济 | 完整论证段落 · 中文译文

能力提升的证据不应被抹去，作者保留的疑问也不应被删除。下面两段分别涉及高技能劳动者的采纳行为，以及绩效收益是否能长期由员工保留。

### 原文选译 | § VI.A末段与 § VIII；非连续

〔4〕在线附录图A.XI与A.XII的结果，与客服人员随时间逐渐重视AI建议的解释一致，尤其是那些最初较为怀疑的人。我们注意到，高技能客服人员采纳建议的增长速度，与低技能同事同样快，尽管他们的生产率收益更小，某些质量指标甚至为负。这提示另一种可能：一些客服人员对AI建议的依赖，可能超过长期最优的程度。尤其是顶尖客服人员，当一条够用的AI建议唾手可得时，可能觉得花时间提供最高质量的输入，并没有多少额外价值。当前高度采纳AI建议，随后可能降低未来用于AI训练的解决方案的质量或多样性。然而，在短期内，我们的分析没有发现模型质量在样本期内下降的证据。在在线附录图A.VIII中，我们显示，较晚获得AI系统、因而使用较新版本的劳动者，在第一个月的处理效果，与较早获得模型版本的劳动者相近。

〔5〕第二，我们报告的是AI部署在局部均衡下的短期至中期影响。虽然我们没有薪酬数据，但我们访谈的管理者认为，劳动者可能因AI帮助而获得了更高的绩效报酬，因为奖金通常与平均处理时间和解决率的目标挂钩。他们提醒说，潜在的奖金收益可能不会持久，因为当达到目标的客服人员太多时，调整绩效目标是常见做法。因此，如果AI帮助使绩效目标不断上调，劳动者最终可能受到一种“棘轮效应”的影响。

### 进一步的证据边界 | 编者概述

作者明确表示，他们不能直接观察长期的一般均衡反应。效率提高后，企业可能用较少工时处理同样数量的问题；更好的服务也可能增加需求，并产生训练、检查模型等新任务。研究不能仅凭局部生产率估计，决定劳动需求、工资和工作时间的总体方向。

第5段关于奖金的说法来自管理者访谈，不是研究取得的工资数据。“棘轮效应”是作者提出的可能机制，也不能写成样本中已被测量证实的普遍结果。同样，第4段对训练数据质量的担心，与“样本期没有观察到模型质量下降”必须同时保留。

### 共同追问

如果新人更快掌握工作技能，却没有更多决定劳动节奏的权力，他在哪些意义上更自由、又在哪些意义上仍受限制？两者的相对重要性需要论证，不能由一个生产率数字代答。

〔09〕 CC BY-NC 4.0；本包非商业中文翻译，已作语言转换。 § VI.B中“A key question...”、“In Panel B...”、“Panel B highlights...”； § VI.A末段及 § VIII以“Second, we report...”开头的段落，共5段，非连续。

原始来源 ↗ 完整出处（末页） 许可 ↗

核心阅读

10

精选材料

## Doshi与Hauser：更多人能够表达

个体创造力与作品评价 · Science Advances (2024)

实践 · 创造力研究 | 完整论证段落 · 中文译文

创作的门槛降低，可能让更多人表达自己。但如果许多人的起点来自相似的机器构思，集体作品会怎样变化？研究把个人收益与作品之间的差异一起测量。

### 研究设计 | 编者导读

实验随机安排作者独立写作、可获取一个AI构思、或最多五个AI构思；任务是写八句话的短篇故事。最终收集293篇故事，由另一个600人的评价者样本评分。作者可以选择是否调用构思；主要分析比较的是被分配到各条件的效果，不只筛选实际使用者。

### 原文选译 | Discussion中两个完整段；非连续

〔1〕我们的结果揭示了生成式AI怎样增强创造力。获得生成式AI帮助，会使故事比作者独自创作时可能达到的水平更加“专业化”。总体效果是，故事更新颖，尤其更有用，而且写得更好、更令人愉快。然而，写出更具创造力故事的收益，并非对所有人都相同：较缺乏创造力的作者获得更大提升，他们的故事在创造力上提高10%至11%，在令人愉快和写作质量上提高22%至26%。

〔2〕不过，我们的研究有局限：创造性任务的长度受到限制，即八个句子；媒介受到限制，即写作；输出类型受到限制，即短篇故事；与大语言模型之间没有互动，提示词也没有变化。这些约束限制了研究的可推广性，以及我们能从中得出的结论。如此限制任务，可能限制参与者表达创造力的程度，也可能无法推广到限制更少的创意任务。如果AI构思的内容不足以引导作者，那么在更长的故事中，它的效果可能减弱。此外，不同媒介中的生成式AI构思，例如图像或音乐，可能以不同方式被吸收，产生不同效果。例如，如果任务是画一幅画，AI构思对几乎没有绘画经验的人可能并不那么有效；写作则不同，大多数人对这项任务已有经验。就此而言，我们注意到，创造力测量中的“有用性”构念是为适合本研究情境而调整的；未来研究应重新考察我们的定义，并确保它能够适用于不同创作领域，以恰当地捕捉创造力的这一方面。同时，我们没有研究或改变现实世界中鼓励创造力的诸多动机。引入经济激励〔原文献10〕、鼓励为问题提出创造性解决办法〔9、11〕，或只是鼓励为了自己的快乐而创作，都可能以不同方式影响人们使用和吸收生成式AI构思。

### 词语提醒

文中的“较缺乏创造力”依照研究使用的前测指标区分，不是对一个人的全部创造潜力作终身判定。“有用性”也按故事任务界定，并非所有领域共用的一把尺。

〔10〕 CC BY 4.0；本包中文翻译，已作语言转换。

原始来源 ↗ 完整出处（末页） 许可 ↗

核心阅读

10

续读

## 个人作品更好，整体也会更多样吗？

互补、群体差异与集体新颖性

实践 · 创造力研究 | 完整论证段落 · 中文译文

这不是“人与机器谁更会写”的简单竞赛。作者关心人获得帮助之后的变化，同时检验不同作品之间是否变得更相似。

### 原文选译 | Discussion的第3-4段

〔3〕已有大量文献讨论AI可能替代人类劳动，例如自动化〔35-37〕，或讨论人与AI构思之间的“竞赛”〔38-40〕。我们关注的是，AI在人类创意生产中可能发挥的互补作用。我们使用的是学术研究中常见、相对“典型”的参与者样本，而这也带来人口代表性上的局限〔41〕；换言之，我们没有研究专业作家或创造力异常突出的人。这些人是重要却研究不足的群体，AI可能以其他方式对他们产生变革性影响，例如带来效率收益或提高执行速度〔6〕。尽管如此，我们的结果表明，生成式AI对创造力较低的个体，可能产生最大的影响。

〔4〕虽然这些结果指向个人创造力的提高，但也存在集体新颖性流失的风险。在一般均衡层面，一个有意思的问题是：经AI增强并启发的故事，能否在由此产生的作品中创造足够的差异？具体而言，如果出版与自助出版行业更多采用AI启发的故事，我们的发现提示，作品作为整体将变得较不独特，彼此更加相似。这种向下的螺旋，与一种正在形成的社会困境相似〔42〕：如果个别作者发现，AI启发的写作被评价为更有创造力，他们以后就有动机更多使用AI；但这样做，又可能进一步降低故事整体的新颖性。简言之，我们的结果提示：虽然生成式AI增强了个人创造力，但如果它在创意任务中被更广泛采用，我们也许需要保留一份警惕。

### 已经发现的，与尚待推断的

研究观察到AI条件下的故事彼此更相似，并使用文本表征的相似性进行比较。但从短篇实验到整个出版行业的长期“螺旋”，仍是作者根据机制提出的推论，不是一个已经完成的行业追踪研究。第4段中的条件语气应当保留。

另一方面，不能因为整体多样性问题存在，就把个体表达的收益当作虚假。原本难以组织故事的人，可能确实写出了更容易被理解和欣赏的作品。个人获得表达手段，与共同文化环境中的差异减少，可以同时成为值得重视的变化。

### 连接自由 | 编者分析

反方需要说明表达支持怎样形成可行机会，而不只停留在作品分数；正方需要说明趋同为何限制真实选择，而不只说“相似就是不好”。选用相似风格可以是自主决定，被无形的默认起点约束也可能是真问题。二者应如何辨别？

〔10〕 CC BY 4.0；本包中文翻译，已作语言转换。Discussion中“‘Our results provide...’”、“‘Regardless, our study...’”、“‘Much has been written...’”、“‘While these results...’” 4个完整段；非连续。

原始来源 ↗ 完整出处（末页） 许可 ↗

补充选读

A1

精选材料

## 杜威：自由不是省去思考

How We Think (1910) · 第五章第二节

补充选读 · 教育哲学 | 完整论证段落 · 中文译文

这份选读进一步探究能力如何形成。杜威区分外在活动顺畅与心智自由，但并不赞成为训练而人为制造无意义的困难。

### 原文选译 | 连续5段的前3段

〔1〕当我们从智力的角度理解训练，把它理解为有效处理问题的习惯性能力时，它就与真正意义上的自由一致了。因为心智的自由，是能够独立运用、不再由别人牵着走的心智能力，而不只是外在行动不受阻碍。当自发性或自然性被理解为短暂冲动或多或少随意的发泄时，教育者就会倾向于提供大量刺激，让自发活动持续下去。各种有趣的材料、设备、工具和活动方式都被提供出来，以免自由的自我表达有所停顿。这种方法忽略了获得真正自由的一些必要条件。

〔2〕（a）冲动倾向的直接、即时发泄或表达，对思考是致命的。只有当冲动受到某种程度的阻止，并转回来作用于自身时，反思才会发生。当然，以为必须从外部强加任意的任务，才能提供作为思考必要开端的困惑与困难，是愚蠢的错误。任何具有一定深度与广度的有生命力的活动，在努力实现自身的过程中，都不可避免地遇到障碍；这使得寻找人为的或外加的问题完全多余。不过，教育者应当珍惜而非减轻经验发展过程中出现的困难，因为它们是反思性探究的自然刺激。自由不在于使外在活动持续不断、不受阻碍，而是一种通过个人反思、克服困难而获得的东西；这些困难原本阻止了活动立即倾泻而出、顺利自发地取得成功。

〔3〕（b）一种方法即使强调心理的和自然的方面，若没有看见在成长的每个阶段，好奇心、推断和检验的愿望构成了自然倾向中何等重要的部分，它仍然不能保证自然的发展。在自然生长中，每个相继的活动阶段，都无意识却充分地为一个阶段的出现准备条件，植物的生长周期就是如此。没有理由认为，“思考”是一种特殊、孤立的自然倾向，只因为各种感觉和运动活动此前得到自由表现，或因为观察、记忆、想象和手工技能此前在没有思考的情况下受到锻炼，它就必然在适当时候开花。只有在运用感官和肌肉、引导和运用观察与动作时不断使用思考，才能为随后更高类型的思考准备道路。

### 读懂“训练”

这里的discipline不是服从性的惩罚，而是有效处理问题的稳定能力。“有困难才会思考”也不意味着所有困难都值得保留：人为的阻碍与真实探究中的问题并不是一回事。

〔A1〕 公版英文原著；本包自译，不摘用现有商业中译本。

原始来源 [↗](#) 完整出处（末页）

补充选读

A1

续读

## 思考能力不会在将来自动出现

同一选段续读 · 当下活动已经形成习惯

补充选读 · 教育哲学 | 完整论证段落 · 中文译文

“先依赖，长大后自然会独立”与“只要没有工具，就在认真思考”，都可能过于简单。杜威把注意力放在当下究竟反复练习了什么。

### 原文选译 | 连续选段第4-5段

〔4〕目前流行一种观念，认为童年几乎完全是不反思的时期，只发展感觉、运动和记忆，而青春期会突然带来思想与理性的表现。但是，青春期并不是魔法的同义词。青年时期无疑应当拓展童年的视野，使人能够回应更大的关切与问题，并以更宽广、更普遍的眼光看待自然和社会生活。这种发展为比以往更全面、更抽象的思考提供了机会。但思考本身始终如一：追踪并检验生活中的事实与事件所提示的结论。当一个婴儿丢失了正在玩的球，开始预见某种尚未存在的可能性——把球找回来，开始设想实现这一可能性的步骤，并通过试验，按照自己的想法引导行动、从而也检验这些想法时，思考就开始了。只有充分利用已在童年经验中活跃着的思考因素，才有希望或根据期待更高的反思能力在青春期或任何后来阶段出现。

〔5〕（c）无论如何，确定的习惯总在形成：如果不是认真察看事物的习惯，就会是匆忙、粗心、不耐烦地掠过表面的习惯；如果不是连贯追踪所出现的想法的习惯，就会是随意、像蚱蜢一样跳来跳去地猜测的习惯；如果不是在通过考察证据检验推断之前暂缓判断的习惯，就会是轻信与轻率怀疑交替出现的习惯，而相信或不相信都以一时兴致、情绪或偶然情况为依据。获得细心、彻底与连贯这些品质——如我们已经看到的，它们是“逻辑的”要素——的唯一方式，就是从一开始便练习这些品质，并确保环境要求人们运用它们。

### 联系AI | 编者分析

工具可能去掉无意义的繁琐，让学习者有余力提出问题、检验证据；也可能在学习者尚未尝试之前就完成了这些活动。因而不能只按“有没有使用AI”评价。同样一项搜索或写作任务，安排不同，学习者承担的判断工作也不同。

杜威要求观察思考的活动，而不仅是最后的作品。如果AI给出了一个答案，学生进一步检查它的前提、寻找反例并修正它，这可能包含反思；如果学生没有理解也没有检验，漂亮的结果不能替他完成这段成长。两种可能究竟在普及过程中占多大比重，仍需现实证据。

### 与核心材料连接

可与材料01的“人—工具共同认知”、材料03和04的学习研究合读。不要用一位哲学家的成长理论代替对实际学习效果的测量，也不要一次测验取消对教育目的的讨论。

〔A1〕 公版英文原著；本包自译，不摘用现有商业中译本。连续5个完整自然段：“When discipline...”、“(a) Direct...”、“(b) The method...”、“At present...”、“(c) In any case...”；去电子版边注及页码。

原始来源 ↗ 完整出处（末页）

补充选读

A2

精选材料

## 比斯塔：教育不只有一种成功

教育目的的三个维度 · 授权意大利文版本（2023）

补充选读 · 教育理论 | 完整论证段落 · 中文转译

“学得更多”“完成得更快”还没有说明学什么、为了什么，以及成为怎样的人。以下保留作者在正文中建立区分的论证，而不是论文摘要。

### 原文选译 | 所选4段中的前2段

〔1〕教育、学校和其他教育机构的一项主要功能，是使儿童、青年和成人具备资格与能力。这涉及向他们提供知识、技能和理解，往往还包括相应的倾向和判断方式，使他们能够“做某件事”。这种“做”可能十分具体，例如为某项工作或职业、某种技能或技术接受训练；也可能宽泛得多，例如理解现代文化或西方文明，或者学习生活技能，等等。资格与能力培养无疑是有组织教育的主要功能之一，也是国家提供资助的首要理由。这些理由主要但不完全是经济性的：它们与教育培养劳动者的作用有关，从而与教育对经济发展和增长的贡献有关。政府、雇主及雇主组织不断讨论教育似乎未能使人充分为工作世界做好准备，表明这种理由的重要性；在英国，这一点经常被称作“技能差距”。然而，资格与能力培养并不限于为工作做准备；向学生提供知识和技能，对其他方面同样重要。例如，政治素养，即行使公民资格所需的知识与技能；或者更广义的文化素养，即通常被认为在社会中生活所需的知识与技能。（当然，是否能够给这种素养下定义，是另一个有争议的问题；例如参见Hirsch，1988；Apple，1993。）

〔2〕不过，这已涉及教育的第二项重要功能，我称之为社会化功能：它涉及我们通过教育成为特定社会、文化与政治“秩序”的成员和一部分的诸多方式。毫无疑问，这是教育的实际效果之一，因为教育从不完全中立，它总是代表某些东西，并以特定方式这样做。有时，教育机构主动追求社会化，例如传递特定的规范和价值、传承特定的文化或宗教传统，或者进行职业社会化。即使社会化不是教育计划和实践的明确目标，它们仍会发挥这种作用；有关“隐性课程”的研究已经说明了这一点。通过社会化功能，教育把个人带入既有的行为方式和存在方式，因此在延续文化与传统方面发挥重要作用，无论这些文化与传统的方面是值得追求的，还是不值得追求的。

### 术语提醒

qualification译为“资格与能力培养”，不只指文凭；socialisation译为“社会化”，不只指善于交朋友，而是进入既有的规范、传统和实践。

### 转译说明

原作者：Gert J. J. Biesta；意大利文译者：Chiara Montà。中文由本包依据《Encyclopaideia》2023年CC BY 4.0意大利文版本转译，非现成中文译本，也非作者审定译本。

【A2】 CC BY 4.0；中文据该意大利文版转译，非现成中文译本或作者审定译本。

原始来源 ↗ 完整出处（末页） 许可 ↗

补充选读

A2

续读

## 能够完成，与能够自主行动

主体化及教育目的之间的张力

补充选读 · 教育理论 | 完整论证段落 · 中文转译

会做某类任务、熟悉一种规范，与能够对自己的判断和行动负责，并不完全相同。比斯塔让我们先分清目的，再讨论效果。

### 原文选译 | 第3–4段；第4段非连续

〔3〕然而，教育不仅促进资格与能力培养和社会化，还会影响我们可以称作个体化的过程；或者按我更喜欢的说法，主体化的过程，即成为一个主体。主体化功能或许最好被理解为社会化功能的对面。这里不是把“新来者”纳入既有秩序，而是指向独立于这些秩序的存在方式；在这种存在方式中，个人不只是更大秩序的一个“样本”。是否所有教育事实上都有助于主体化，是可以争论的。有些人认为并非必然如此，教育的影响可能只限于资格与能力培养和社会化。另一些人认为，教育始终影响个人，因此始终具有个体化的“效果”。但更重要的是主体化的“质量”——在这里，我们必须从教育实际发挥什么功能的问题，转向教育的目标、宗旨和目的问题——也就是特定教育安排与结构使何种主体性或哪些类型的主体性成为可能。正是在这一点上，一些人主张或曾经主张，例如分析传统中的Peters（1966；1976）、Dearden等（1972）及较近的Winch（2005），以及批判传统中的Mollenhauer（1964）、Freire（1970）、Giroux（1981），任何名副其实的教育，都应当促进主体化过程，使受教育者在思考和行动中变得更加自主、更加独立。

〔4〕当我们讨论教育的根本问题，也就是回答什么构成良好教育时，有必要把这三个维度区分开。我们需要明确，自己的回答如何涉及资格与能力培养、社会化和／或主体化。最重要的是意识到这几个维度，意识到它们有不同的前提，也意识到它们之间一方面可能协同，另一方面也存在发生冲突的可能；尤其是，我想指出，一边是资格与能力培养及社会化，另一边是主体化，两者之间可能发生冲突。

### 联系AI | 编者分析

学习使用AI，可能本身属于现代资格与能力的一部分；共同核验并说明使用过程，也可能形成合作与责任规范。与此同时，只把更快生成正确文本视为成功，可能遮蔽学生是否形成了自主回应问题的能力。两种解释都需要观察实际活动，而不能由“技术先进”或“亲手完成”直接推出。

这个框架并不规定三种目的必须永远等重，也不意味着主体化就是拒绝一切外部帮助。参赛者需要解释在具体阶段、任务与教育条件中，哪些目标特别重要，以及为什么某种取舍有理由。此处是本包的应用，不是比斯塔对当代生成式AI的原话。

### 与核心材料连接

可与材料07的手段一目的区分合读：成绩、效率和工具都是可能的重要信息，但不能不经论证，就替所有教育价值作出判断。

〔A2〕 CC BY 4.0；中文据该意大利文版转译，非现成中文译本或作者审定译本。 § 3，第14–15页；资格与能力培养、社会化、主体化及“È necessario separare...”段，共4段；末段与前三段不连续。

原始来源 ↗ 完整出处（末页） 许可 ↗

补充选读

A3

精选材料

## Aylsworth与Castro：为什么应当自己写作？

Should I Use ChatGPT to Write My Papers? (2024)

补充选读 · 伦理学 | 独立论证概要

即使教师允许AI，即使生成的论文合格，学生是否仍有理由自己完成思考？作者把问题从是否违反校规，转到对自身自主能力的责任。

### 作者首先排除哪些不充分的理由？

作者认为，“这是作弊”没有穷尽问题：校规可以改变，教师也可以公开允许使用AI。在这样的情形中，学生是否仍会失去某种值得珍惜的教育机会？要回答它，不能只诉诸现行规定。作者主要讨论大学人文学科论文，不把拼写检查和整篇内容生成混为一谈；“写作”指观点与表达的形成，而不是手指是否亲自输入。

“使用工具会失去一种技能”也不够。技术经常使旧技能不再重要，但我们没有义务保存每一种旧技能。反对者还需说明，某种能力为什么具有特殊价值，以及失去它为什么不能被其他收益补偿。作者同样认为，“写作就是思考”仍过于宽泛：并非所有困难任务或思考活动都具有必须亲自完成的价值。

### 核心理由 | 培养自主性的义务

作者采用康德式论证，把“人性”理解为理性行动者设定并追求目的的能力，也就是自主性。这种能力使我们能够评价欲望、反思继承而来的承诺，并决定怎样生活。作者认为，我们不仅应尊重他人的这种能力，也对维护和培养自己的能力负有义务。人不能因为一时省力，便把决定哪些目标值得追求的能力当成可有可无的工具。

从自主性有价值，到必须写这篇论文，仍需要连接。作者把培养自身的责任理解为有一定余地的“不完全义务”：不是每个机会都必须接受，但也不能不断放弃机会，却仍声称自己认真承诺这一目标。他们使用“如果不是现在，那是什么时候”的检验，追问一个人是否正在无充分理由地错过珍贵的发展机会。

大学人文学科写作被视为一种这样的机会。学生在接近成年、重新考察人生方向时，需要读懂他人的理由，权衡相互冲突的价值，提出自己的判断，再接受教师对思考的反馈。把整篇论文交给模型，可能保住成绩，却错过了这段练习。作者由此主张学生有理由自己写论文，而不仅是交出一篇达到要求的文本。

#### 来源性质

以上为对作者论证的独立概要，不是原文翻译；尤其不能把一项伦理论证写成“实验证明使用AI会损害自主”。

【A3】 CC BY-NC-ND 4.0；本包为独立论证概要，不发布原文改编译作。

原始来源 ↗ 完整出处（末页）

补充选读

A3

续读

## 最强的反对：工具也可能增强自主

作者的回应与仍需检验的连接

补充选读 · 伦理学 | 独立论证概要（续）

这篇文章最值得读的，不是“禁止代写”这一结论，而是作者如何回应：技术经常帮助我们设定和实现目标，为什么这一次要特别谨慎？

### 作者的回应

作者承认，技术能够节省琐事、扩展信息取得，也可能帮助表达或提出反对意见。让机器校正拼写、规划路线或提醒日程，并不必然放弃自主。因此，他们并不要求所有认知劳动都留在脑内，而是把重点放在评估理由、检查价值与形成生活目标等能力上。作者也明确承认，口试、讨论、辅导和展示等形式可能实现相同的教育目的；重点是写作所促进的价值，而不是纸面论文这种形式本身。

他们进一步区分已掌握技能的专家与仍在学习的人。专家可以判断建议、看出问题，在与AI合作中继续提高；学习者若在尚未形成判断基础之前就让AI代做，可能没有机会形成未来合作所需的能力。作者用这一差别回应“人机合作使人更强”的反驳，但这不是说所有新手都不可能从工具中学习；具体学习安排仍是争论的关键。

文章也把使用动机放进教育环境。学生可能原本重视成长、理解与自主，却在成绩、毕业和就业压力下，把教育简化成分数。作者借用“价值捕获”的讨论说明，选择代写不只来自懒惰，也可能来自制度传达的成功标准。教育者因此需要让任务的真正价值显现，不能只要求学生服从一条禁令。

他们认为，培养自主性的机会不应只属于少数精英学校，也讨论论证向高中及其他教育场景延伸的可能。但这种延伸仍需任务分析：数学演算、信息整理、创意写作和人文学科论文，不必以同样方式发展同一能力。文章没有为所有科目和所有AI辅助提供同一个现成结论。

### 还需要怎样的回应？ | 编者分析

最有力的反驳不是否认自主性有价值，而是检验作者的连接：这次写作是否真是难以替代的培养机会？让学生比较模型答案、说明修改理由并接受追问，能否保留甚至增加相同的反思？正方若引用这篇文章，需要说明哪些活动确实被替代；反方若提出合作方案，需要说明不是只保留了形式上的参与。

还应区分个人义务与制度权力。即使一个人有责任培养自己，也不意味着学校可以用任何方式强迫他；反过来，选择使用工具，也不自动证明教育目的已经达成。本题讨论的是普及带来的自由变化，不能把一种课程规范直接当作整个社会的裁决。

#### 选读提示

补充材料用于深化思考，不是默认必备知识。比赛应评价理由、证据和回应，而不是谁引用了更多作者。

【A3】 CC BY-NC-ND 4.0；本包为独立论证概要，不发布原文改编译作。依据 § 2-4：不足的理由、培养自主性的义务、写作的独特机会及反对意见。

原始来源 ↗ 完整出处（末页）

出处 · 授权与翻译说明

## 原始来源 | 核心材料01-07

阅读正文中的〔编号〕对应以下来源。选段位置在各材料末尾；段落编号为本包所加。译文与概要均为中文阅读辅助，不是作者审定中文版本。

### 〔01〕 Andy Clark & David Chalmers (1998)

The Extended Mind. Analysis, 58(1)。

未核实适用的开放翻译转载许可；本包采用独立概要，不复制原文。

原文 / DOI ↗

### 〔02〕 John Stuart Mill (1859)

On Liberty. 第三章：Of Individuality, as One of the Elements of Well-Being。

公版英文原著；本包自译，不摘用现有商业中译本。

原文 / DOI ↗

### 〔03〕 Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı & Rei Mariman (2025)

Generative AI without guardrails can harm learning: Evidence from high school mathematics. PNAS, 122(26), e2422633122。

CC BY-NC-ND 4.0；本包为独立研究概要，不发布原文改编译作。

原文 / DOI ↗ 公开全文 / 作者版本 ↗

### 〔04〕 Greg Kestin, Kelly Miller, Anna Klaes, Timothy Milbourne & Gregorio Ponti (2025)

AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports, 15, 17458。

CC BY-NC-ND 4.0；本包为独立研究概要，不发布原文改编译作。

原文 / DOI ↗

### 〔05〕 Daniel Susser, Beate Roessler & Helen Nissenbaum (2019)

Technology, autonomy, and manipulation. Internet Policy Review, 8(2)。

CC BY 3.0 Germany；本包中文翻译，已作语言转换。

原文 / DOI ↗ 公开全文 / 作者版本 ↗ 许可文本 ↗

### 〔06〕 Francesco Salvi, Manoel Horta Ribeiro, Riccardo Gallotti & Robert West (2025；更正2026)

On the conversational persuasiveness of GPT-4. Nature Human Behaviour。

CC BY 4.0；中文选译已纳入2026-09-03作者更正。

原文 / DOI ↗ 公开全文 / 作者版本 ↗ 许可文本 ↗ 2026作者更正 ↗

### 〔07〕 Ingrid Robeyns (2017)

Wellbeing, Freedom and Social Justice: The Capability Approach Re-Examined. Open Book Publishers；第2章。

CC BY 4.0；本包中文翻译，已作语言转换。

原文 / DOI ↗ 公开全文 / 作者版本 ↗ 许可文本 ↗

### 怎样核对原文？

先按段首提示找到对应自然段，再看前后语境。跨页段落作为一个完整自然段处理；非连续选段已标注。原作者引注与本包〔材料编号〕不是同一套编号，原注、参考文献和图表可由原文链接核对。

材料06采用2026年9月3日更正后的表述，未沿用旧版关于个性化额外优势的统计结论。

出处 · 授权与翻译说明

## 原始来源 | 其余核心与补充选读

### 【08】 Daron Acemoglu, David Autor & Simon Johnson (2023)

Can We Have Pro-Worker AI? Choosing a Path of Machines in Service of Minds. MIT Shaping the Future of Work Initiative · 政策备忘录。

未核实适用的开放翻译转载许可；本包采用独立概要，不复制原文。

[原文](#) / [DOI](#) ↗

### 【09】 Erik Brynjolfsson, Danielle Li & Lindsey Raymond (2025)

Generative AI at Work. The Quarterly Journal of Economics, 140(2), 889–942。

CC BY-NC 4.0；本包非商业中文翻译，已作语言转换。

[原文](#) / [DOI](#) ↗ [公开全文](#) / [作者版本](#) ↗ [许可文本](#) ↗

### 【10】 Anil R. Doshi & Oliver P. Hauser (2024)

Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), eadn5290。

CC BY 4.0；本包中文翻译，已作语言转换。

[原文](#) / [DOI](#) ↗ [公开全文](#) / [作者版本](#) ↗ [许可文本](#) ↗

### 【A1】 John Dewey (1910)

How We Think. 第五章第二节：Discipline and Freedom。

公版英文原著；本包自译，不摘用现有商业中译本。

[原文](#) / [DOI](#) ↗

### 【A2】 Gert J. J. Biesta；意大利文译者：Chiara Montà (2023 (对应英文论文2009))

La buona educazione in un' epoca di misurazioni: sulla necessità di recuperare la questione dello scopo dell' educare. Encyclopaideia, 27(1S), 9–20。

CC BY 4.0；中文据该意大利文版转译，非现成中文译本或作者审定译本。

[原文](#) / [DOI](#) ↗ [公开全文](#) / [作者版本](#) ↗ [许可文本](#) ↗

### 【A3】 Timothy Aylsworth & Clinton Castro (2024)

Should I Use ChatGPT to Write My Papers?. Philosophy & Technology, 37, 117。

CC BY-NC-ND 4.0；本包为独立论证概要，不发布原文改编译作。

[原文](#) / [DOI](#) ↗

### 辅助视频 Inkstone · Episode 4: the Extended Mind

材料01的辅助理解入口；本包不转载视频内容，不将其计为独立实证研究。

[视频页面](#) ↗

#### 使用与编排说明

本包供非商业教育讨论。各来源分别适用其许可；不对整个合集作统一开放再许可。公版认定具有地域性；密尔与杜威依据公版英文自译，不复制现有出版中译本。许可带ND的材料采用独立概要，而非发布改编译文。

核心阅读按整页配置为理论10页、实践10页，哲学8页、社科及教育12页；选读另计。正反资源均衡是编选目标，不表示每项研究都支持相反结论，也不意味着两方每种主张都有同等强度的证据。

编者导读、联系与追问均已另行标示。来源及版本核对日期：2026年9月14日。