

观点与分析

编辑日期: 2026年10月31日

没有回到任何人手里的两小时

> 在一家美国科技公司，人工智能让启动任务变得如此容易，工作开始侵入原本会让它停下来的时间。在丹麦，新任务不断出现，记录中的工时与工资却几乎没有变化。悖论不在于机器没有节省时间，而在于：任务省下的时间，并不能独自决定它最终归谁。

失去休息资格的休息时间

2026 年 2 月 18 日，加州大学伯克利分校哈斯商学院刊出对 Xingqi Maggie Ye 的研究访谈。那不是对十年后就业的预测，而是她与 Aruna Ranganathan 在一家约有 200 名员工的美国科技公司中开展的八个月民族志田野研究。研究者跟踪日常实践，完成 40 多次半结构式访谈，随后发现了一个事先没有假定的现象：在这些观察中，人工智能让任务更容易，并没有使工作收缩，反而扩张了工作的边界。1

Ye 描述了三种模式。人们开始接手过去会交给别人、或根本不会去做的工作。向人工智能模型提交请求变得足够容易，于是会前、午休和晚间都可以顺手发出。多条工作链也开始并行：员工审阅一份文件时，另一项工具正在起草新稿，第三项工具则等他作出决定。研究者认为，什么可以被算作“我的工作”，其边界已经扩大。这里是对其已发表描述之意的中文转述，并非虚构的员工访谈。

两位研究者的分析于同月 9 日发表于《哈佛商业评论》（Harvard Business Review）。这是仍在进行的质性研究结果，不是随机实验，也不是能够代表所有公司的统计调查。它没有告诉我们，所有使用人工智能的人都延长了工时；但它揭示了一种细微机制：任务启动成本越低，拒绝新增任务可能越难；执行环节省下几分钟，也可能消除原本隔开一项工作与下一项工作的短暂停顿。2

这些停顿并非生产计划的漏洞。人在答应之前思考，或在作出决定前恢复注意力，都需要这种时间。启动成本下降，不一定产生空白，也可能产生新的需求。

让一项任务腾出时间，与让一个人免于另一段时间的义务，是两回事。

三本账，以及“我们省下了”三种不同含义

如果智能助手把一份报告的制作时间缩短一半，项目负责人可以说：我们省了两个小时。他对自己测量的对象可能完全诚实。但财务负责人听见的也许是另一句话：劳动成本减少了两个小时的工资。员工听见的则可能是第三个承诺：我的工作时间会少两个小时。

这是三本不同的账，没有证据，不能把一本账上的数字搬到另一本。2026 年 1 月发表于《Management Science》的一项美国研究，根据具有就业人口代表性的调查估算，在 2024 年末一轮调查中，参与者自报因使用生成式人工智能（Generative AI）省下的时间，相当于总工时的 1.4%。这是参与者报告的估值，不是工资记录，也不是对企业现金节余的测量。13

第一本是任务账：相对于明确替代方案，产出一项达到可接受质量的特定结果需要多少时间，例如审阅合同或概括报告。第二本是机构账：现金成本、收入或提供额外服务的能力发生了什么变化。第三本是个人账：实际工时、注意力强度、非工作时间内的劳动，以及个人能否为自己和家人收回时间，究竟发生了什么变化。

任务在它自己的账上可以变得更好，另外两本账却纹丝不动。甚至，机构生产更多材料的能力提高时，个人账还可能恶化。这在算术上并不矛盾；矛盾发生在机构把其中一本账说成全部三本账的时候。

从现金会计看，如果员工薪酬不变，也没有减少付薪工时、合同成本或原本确实会发生的招聘需求，就不能证明现金节余等于“腾出的小时数 × 员工小时成本”。工作质量提高或能够承接其他任务，可能创造经济价值，但需要独立测量。时间不会像硬币一样存进机构金库。

这种区分并非贬低技术成果，恰恰是在公允评价它。让研究者多读一个来源，或让工程师多检验两个方案的技术，可能创造真实价值，即使工资账单一里亚尔也没有减少。若只有伪装成“成本下降”才能让人承认创新价值，反而有失公允。

丹麦：任务在动，工资没动

Anders Humlum 与 Emilie Vestergaard 于 2026 年 3 月 13 日发布研究稿 *Still Waters, Rapid Currents: Early Labor Market Transformation under Generative AI*。他们把大规模调查与丹麦行政记录相连接；2024 年末的主要一轮调查覆盖约 2.5 万名劳动者、7000 个工作场所，涉及 11 种预计会受对话模型影响的职业。3

研究发现，采用工具与员工自报的生产率收益相伴，也出现了内容写作、监督模型输出质量、把工具整合进工作等新任务。但研究者使用双重差分法（Difference-in-Differences）估计时，在观察窗口内，无论劳动者个人还是工作场所层面，都没有发现工资或记录工时具有实际意义的变化；按研究设计，其置信区间排除了平均幅度约高于 2% 的影响。

那么，收益消失了吗？没有。能否证明机构拿走了全部收益？也不能。结果的准确含义是：在测量时段和所检验指标上，这些早期收益没有转化为工资与工时清晰可见的实质变化。吸收过程可能发生在岗位内部：旧工作腾出的位置，立即长出新工作。

一个值得注意的细节是：在没有机构推动计划的环境中，约 8% 的工具使用者称自己接下了因使用工具而生的新任务；在存在有效推动计划的工作场所，这一比例约为 17%。这只是研究数据中的类别比较，不是对推动计划独立因果效应的估计。有些新任务不可或缺，例如隐私与质量审查；有些可能是机构过去无暇提供的有价值服务；还有一些，也许只是因为生产变容易了才得以存在。

两位研究者没有把每一项新任务都叫作浪费。这里必须守住一条科学边界：工作增加本身不能证明技术失败；工资不变本身也不能证明收益分配公平。

这项研究也有必须随结论一同携带的强限制。所引用版本在发布时是尚未完成最终同行评审的讨论论文。研究依赖采用者与未采用者之间的比较假设、采用前趋势的有效性，以及行政记录捕捉有薪工时的能力。记录可能看不见一小时内注意力密度的变化，也看不见未被登记为加班的晚间消息。新技术扩散两年的情况，也不能裁决十年后的影响。

尽管如此，这些数据让我们提出一个比“模型省了多少分钟”更好的问题：旧工作离开以后，哪些新工作进入了岗位说明？

我们能从煤炭学到什么，又有什么不能照搬？

1865 年，英国经济学家 William Stanley Jevons 出版 *The Coal Question*。他指出，蒸汽机效率提高、完成单位工作所需的煤炭减少，不必然使煤炭总消费量下降。可用能源变得更便宜后，工厂、用途与市场可能扩张，需求增长甚至超过每单位消耗的节约。[4]

后来的文献把这种思想称为反弹效应（Rebound Effect），其强形式常被称作杰文斯悖论（Jevons Paradox）。但它并不意味着效率总会增加消费。结果取决于需求弹性、价格、替代品、新投资与市场规模；反弹可能只有一部分，也可能不足以抵消节约。

把这个想法移到人工智能领域时，必须为类比划出清楚边界：人不是一吨煤，一天的时间也不是市场可以增产的资源。不过，任务执行成本下降时，任务需求的确可能扩张，即使个人可用时间仍然有限。

如果审阅 10 个战略假设从一周缩短为一天，管理层也许会要求测试 50 个假设。若新增假设检验的是有价值的问题，这可能很好；若展示文稿的数量取代决策质量，变成思考能力的证据，它也可能只是优雅的忙碌。答案不由工具速度决定，而由新任务的边际价值决定：其收益是否超过它将消耗的时间、资源和注意力？

Daron Acemoglu 与 Pascual Restrepo 对自动化和新任务的研究，提供了更深的经济解释。他们在 2019 年说明，自动化可以把人类劳动从某些任务中挤出，新任务又会重新创造劳动需求。这是劳动经济学框架，并非所有工作都会变得更累的预测；最终变化取决于替代、任务创造、生产率与需求之间的平衡。[5]

杰文斯的世界与办公室的差别在于：市场可以要求再做 1000 份报告，员工却只有一段人生。需求弹性走到这里，就变成了劳动边界问题。

腾出工时账

本文提出一种机构测量工具，称为腾出工时账（Freed-Hour Ledger）。这是本文建议的分析术语，不是独立学术理论，也不是正式会计准则。它的作用是：在追踪任务缩短后那些分钟去了哪里之前，不把任务用时减少直接计作现金节余或休息时间。

假设一项重复工作在质量明确、工作量可比且固定的条件下，每周减少了 S 分钟。员工其余时间发生了什么？可以把关系简化为：

$$\Delta H = -S + \Delta N + \Delta V + \Delta O$$

其中，ΔH 是实际工作总分钟数的变化，S 是可比任务减少的分钟数，ΔN 是过去不存在的新任务所占时间，ΔV 是审阅、纠错和整合人工智能输出的新增时间，ΔO 是其他既有任务所获时间的变化。没有测量，这个等式不能估出任何变量；它只是用于组织分析的恒等式，不是声称这些成分足以解释全部变化的实证模型。

设想一个明确的假设例子：某工具每周为一位员工的任务缩短 120 分钟，其中 50 分钟重新用于检查错误与需要核验的输出，40 分钟用于新项目，30 分钟用于更深入地完成既有工作。一周结束时，总工时一分钟也没有减少。部分产出的质量也许提高了，新项目也可能毫无价值；只有测量结果后，才能知道是哪一种。

另一个假设例子中，任务同样缩短 120 分钟，40 分钟回到其他任务，20 分钟用于审阅，60 分钟则受到保护，变成提前下班或减少非工作时间劳动。这一次，员工真正得到一小时，不是模型替她决定了归属，而是工作设计允许这段时间从任务账转入个人账。

即使如此，计算仍需谨慎。提高任务质量可能让它与昨天的任务不可比；需求的季节变化也可能在与人工智能无关的情况下改变 ΔO。核验时间不能既算进原任务，又在 ΔV 中再算一次。一本好账的首要规则，是不把同一分钟卖两次。

员工感觉省时，计时结果却看不见

2025 年，模型评估与威胁研究组织（METR）发布了一项随机实验，研究熟悉既有开源项目的资深开发者。研究包含 16 名开发者和 246 项任务；部分任务允许使用 2025 年 2 月至 6 月期间可用的人工智能工具，另一些任务则禁止使用。[6]

开始前，参与者预计允许使用工具会把用时缩短 24%。实验结束后，他们估计自己快了约 20%。但对实际完成时间的测量显示，在这一具体情境中，允许使用工具时，完成任务反而多用了 19% 的时间。

这不能证明人工智能普遍拖慢程序员。它非常精确地展示了：感觉自己更快，并不等于测得自己更快。快速模型的互动会制造推进感，额外的分钟却可能消耗在阅读建议、撤销修改，以及让方案适配复杂代码库上。

2026 年 2 月 24 日，同一团队发布重要更新：他们尝试用更新的工具重复测量时，遇到参与者选择偏差，因为有些开发者已不愿在没有人工智能的情况下工作；多智能体参与也让计时变得困难。团队因此称，新的提速信号不足以可靠估计一个具体效应，尽管他们倾向认为 2026 年工具对开发速度的提升，可能高于 2025 年初的工具。7

管理层应该这样对待惊人的数字：既不否认，也不把它永久化。要追问技能、任务类型、工具年代、实验设计与学习效应。不能因为今天的员工感觉工作轻松了，就直接编制明天的工资表。

有时，反对意见理应胜出

时间确实可能回到它的主人手中。不能因为实验证据妨碍“工作总会扩张”的叙事，就把它埋掉。

一项持续六个月、覆盖 66 家公司和 7137 名知识工作者的随机研究，把集成在通信、写作和会议应用中的人工智能工具访问权限分配给部分参与者。根据发表于 *American Economic Review: Insights* 的结果，在实验后半段，实际使用该工具的人——约占获得访问权限组的 80%——每周少花两小时处理电子邮件，常规工时之外的工作也减少了。研究者没有发现，向个人提供工具会明显改变任务数量或结构。8

这里确有一些空间回到了人手中。但差别仍然重要：少花两小时处理电子邮件不必然等于总工时减少两小时，即使非工作时间劳动下降是一个积极信号。还必须明确效应是在哪个群体上测得的，不能把工具实际使用者的平均值归给所有公司员工。

于是，我们看到三组结果：一家美国公司中，研究者观察到工作强化；一个大型丹麦样本中，记录工时没有实质变化；一项随机研究中，工具使用者的邮件时间和非工作时间劳动减少。三者怎么可能同时成立？

因为它们回答的不是同一个问题，测量的也不是同一层面。第一项观察一家机构内工作日的含义与边界；第二项跨多个职业测量工资、记录工时与任务；第三项测量向特定企业人员提供一种工具，对数字行为和非工作时间的影响。指标不同不是方法上的奢侈，而是问题的核心。

有时，机器能够归还时间。但它没有分配时间的权利。

谁继承腾出的工作时间？

经济学家 Gary Becker 在 1965 年论文 *A Theory of the Allocation of Time* 中提出一个核心观点：时间本身是具有机会成本（Opportunity Cost）的资源。人不只用金钱购买事物；还要在劳动、消费、照护和其他活动之间分配时间，而一小时的价值取决于为它放弃了什么。9

把这个观点用于机构中的劳动者，还需要多走一步。工具在有薪工时内腾出的时间，并不会自动成为员工可自由支配的财产；劳动合同、制度与管理决定其大部分配置。同时，机构也不能假定这段时间没有人的成本，因为注意力、疲劳恢复能力与家庭义务都不是无限资源。

那么，谁从技术中获益？客户可能得到更好或更快的服务；机构可能从新增业务获得利润；员工可能减少重复劳动、获得休息或提升能力；也可能在工资不变的情况下承受工作强化，而机构从更多产出中获益。部分结果可以同时出现，不能仅凭购买了人工智能许可就预设其中一种。

Gallup 于 2026 年 10 月 6 日发布的调查中，人工智能影响分析涵盖 7845 名美国员工的回答，他们称自己过去一年曾在工作中使用人工智能。其中 63% 表示工具帮助自己更快完成工作，31% 表示，使用工具促使雇主向自己提出更多职责要求。这些是自报结果，不是测量每位雇主实际要求及新增时间的因果实验；但它让收益分配从经济推导变成一个具体问题。10

管理会议之外的人同样重要。如果任务延伸至晚间，成本就转移到机构系统没有记录的家庭、学习或照护时间。如果员工获得更多灵活性，这也是没有出现在工资表上的真实收益。生活质量不能被工资数额概括，也不应在没有测量时捏造福祉数字。

还有一项相反结果值得正视。2025 年发表于 *Nature Human Behaviour* 的研究跟踪了一项不减薪四天工作制的组织干预，取得六国 141 家机构约 2896 名员工干预前后的数据，并设有对照企业。研究观察到所考察的倦怠、工作满意度和健康指标有所改善，而对照企业没有出现同样模式。11

这不是人工智能实验，也不是可代表全球机构的一般随机实验。它在这里的重要性是：要把效率与组织能力转化为个人时间，需要明确的机构安排——重新设计工作、确定更短工时，并对清晰结果作出承诺。休息不是会议结束后偶然剩下的时间，而是一项必须被选择、保护并测量的产出。

不能建立在假想分钟上的利润表

在经济学中，劳动生产率（Labor Productivity）通常是实际产出与工时之比，而不是劳动者声称自己省下多少分钟。美国劳工统计局（Bureau of Labor Statistics）在其统计范围内也采用这个定义。因此，“任务少用了几分钟”不足以宣布机构生产率增长，更不能据此宣布利润增长。12

设想一家有 100 名员工的机构，平均劳动小时成本为 150 里亚尔。它声称每名员工每周节省两小时，一年按 48 周计算。电子表格会得到：

$$100 \times 2 \times 48 \times 150 = 1{,}440{,}000$$

单位：沙特里亚尔。计算没有错，但结果并不会自动变成进入金库的钱。如果 100 名员工仍领取相同薪酬、工作相同时间，就没有实现这一金额的工资节省。如果时间改用于新增服务，其价值也不能按员工小时成本登记，而应按它实际增加的价值或收入、或可证明避免的成本计算，同时扣除工具、审阅与监督费用。

可以把更简单也更诚实的财务决策标准写成：

净现金影响 = 实际避免的成本 + 已实现的额外贡献毛利 − 工具、验证与变更成本

这是供管理层思考的表达式，不是正式会计准则，也不建议据此登记某个具体财务科目。它要求同一份价值不能重复计算：如果避免新增一个岗位已经计作节余，就不能再把这个假设岗位的工时计为新收入。还必须证明现实基线：没有人工智能，招聘是否真的会发生？新增需求能否实现？完成质量是否可比？

把“节省工时”放大写满演示文稿的管理者，可能以为自己发现了新的金融资产。但一个数字不会因为字号变大就变成现金。

如果最好的效果是让一位员工不再重复录入数据，转而帮助一位困惑的客户，那当然值得肯定；但它应当作为工作质量与客户体验的改善来评价，而不是作为从她工资里虚构出来的折扣。

一项可以证伪的沙特实验

我并不声称沙特已经开展了以下实验。这是一个建议设计：在数据、劳动与隐私规则允许，并取得主管部门批准、保护员工的前提下，银行、服务机构或科技公司都可以实施。

选择一项质量定义清楚的活动，例如为中小企业合同准备一套文件。采用工具前，抽取任务样本，记录准备与审阅时间、输出准确性、后续纠错、员工下班时间和晚间工作。随后通过随机分配单位，或采用有充分理由的准实验设计，比较使用助手的组与沿用常规方法的组。

实验期间，不要只问参与者“你省了多少时间”。可以从工作记录中抽样，并使用保护隐私的自愿日记实际测量时间；然后追踪机构如何使用腾出的能力。审阅增加了吗？出现过去不存在的任务了吗？晚间工作请求减少了吗？原本确实会采购的外部资源需求下降了吗？经过足够时间后，客户体验质量提高了吗？

接着，检验不同管理决策：一组允许任务量随需求扩张；一组把特定比例的时间重新分配给工作质量或培训；第三组在服务性质允许时保护休息或提前下班时间。若某组承接的案例更难，或各组主管存在实质差异，比较就不成立。还应事先约定，服务安全与员工权利是不可拿来和分钟数权衡的约束。

“腾出工时账”的预测可以被证伪：如果任务用时缩短会自动变成工时下降，那么无需分配决策，节省分钟数与实际工时下降之间也应出现稳定关系。如果技术相似，时间去向却随工作规则与激励变化，这便支持一个判断：时间分配既是机构决策，也是技术结果。这不意味着所有公司都应实行更短工作周，或不应扩张服务；而是要求机构有意识地选择能力的去向，并如实公布自己测量的对象。

某些机构最好的投资回报，可能是提高一种无法压缩成小时工资的服务质量；另一家机构可能通过避免昂贵合同或减少付费加班获得最佳回报；第三家机构则可能选择把时间还给员工，让他们收回办公室之外的生活。只要结果符合机构义务，且不冒充另一种结果，这些选择都可以成立。

下次会议之前的那个问题

民族志研究把我们带回原本隔开任务的时刻：会议前、午休时，以及不再拥有明确终点的夜晚。管理者可能从未正式决定延长工时，员工也没有主动要求。只要新增任务先变得容易，再变得习惯，最后变成期待，边界就会改变。

在这种情形下，工具并没有缩短工作失败。它也许成功得让旧问题“这要花多久”不再如新问题有用：“我们为什么要做它，它留下的分钟归谁？”

时间不会因为从屏幕上腾出来，就自动成为现金节余。模型回答更快，也不会自动产生人的休息。还需要另一个决定，而且它可能比购买技术本身更难：如果缩时真的成功，必须明确写下这段时间将如何处理。

如果明天机构告诉你，人工智能每周为员工节省了两个小时，最有用的问题也许不是模型多准确，也不是买了多少许可。

请问它：那两个小时后来出现在哪一本账上?

参考文献

1. 加州大学伯克利分校哈斯商学院，2026 年 2 月 18 日。 AI promised to free up workers' time. UC Berkeley Haas researchers found the opposite。类型：对原研究者的机构访谈。支持：一家约有 200 名员工的科技公司中，持续八个月、包含 40 多次访谈的民族志研究，以及任务范围扩大、工作侵入休息时间的模式。局限：单一公司、仍在发展的质性研究，不是全国因果平均值。
2. Aruna Ranganathan、Xingqi Maggie Ye，2026 年 2 月 9 日。 AI Doesn't Reduce Work—It Intensifies It，Harvard Business Review。类型：原研究者呈现进行中研究结果的研究分析文章；部分内容可能需要订阅。支持：任务变容易时，工作可能自我强化。局限：不是基于广泛概率样本的已发表同行评审论文。
3. Anders Humlum、Emilie Vestergaard，2026 年 3 月 13 日。 Still Waters, Rapid Currents: Early Labor Market Transformation under Generative AI，RFBerlin Discussion Paper 078/26。类型：发布时尚未同行评审的初步工作论文。支持：11 种职业中约 2.5 万名劳动者、7000 个工作场所，工具相关任务增长，以及记录工资与工时未出现清晰实质影响；置信边界排除高于约 2% 的平均影响。局限：依赖因果识别假设的观察性比较，集中于高暴露职业，且时间窗口较早。
4. William Stanley Jevons，1865 年。 The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-Mines。类型：有原文可查的历史经济学著作。支持：煤炭使用效率提高可能通过用途扩张推动总消费增长。局限：把它用于知识工作需求扩张，是有边界的经济类比，不能证明每名员工都会延长工时。
5. Daron Acemoglu、Pascual Restrepo，2019 年春。 Automation and New Tasks: How Technology Displaces and Reinstates Labor，Journal of Economic Perspectives，33(2):3 – 30。类型：提出任务型劳动经济框架的同行评审论文。支持：自动化替代部分任务、新任务又重新创造劳动需求，不对所有市场断言单一结果。
6. Joel Becker 等，2025 年。 Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity，METR。类型：由实施机构作为研究发布的随机实验，包含 16 名开发者和 246 项任务。支持：在参与者自估节省时间的同时，允许使用 2025 年初工具时，实测完成时间增加约 19%。局限：资深开发者处理成熟项目，使用的是特定时期的特定工具。

7. METR 团队，2026 年 2 月 24 日。 We Are Changing Our Developer Productivity Experiment Design。类型：同一研究团队的初步方法更新。支持：参与者选择与多智能体工作计时困难，后续结果不足以可靠估计最新的精确效应。局限：这是方法反思，不是证明某个具体改进幅度的新对照实验。

8. Eleanor Wiseman Dillon、Sonia Jaffe、Nicole Immorlica、Christopher Stanton，2025 年研究，后获接收发表。 Shifting Work Patterns with Generative AI，American Economic Review: Insights；NBER 工作论文版本。类型：覆盖 66 家公司、7137 名知识工作者的随机田野实验。支持：实验后半段，实际工具使用者每周少花约两小时处理电子邮件，非工作时间劳动减少。局限：电子邮件时间不等于总工时；本文也不把该研究作为开篇案例或中心论点重复使用。

9. Gary Becker，1965 年 9 月 1 日。 A Theory of the Allocation of Time，The Economic Journal，75(299):493 – 517。类型：奠基性同行评审经济学研究。支持：时间是一种有机会成本的资源，而不是任务缩短后自动获得的现金资产。局限：个人经济选择层面的时间配置模型，不能独自决定劳动合同的工时安排。

10. Gallup，2026 年 10 月 6 日。 AI Benefits at Work Unevenly Distributed，American Job Quality Study。类型：原始调查结果；2026 年数据采集于 1 月 26 日至 3 月 24 日。支持：7845 名使用人工智能工作的美国员工中，63% 自报工作更快，31% 自报雇主要求额外职责。局限：特定人群的自报估计，不是因果证明。

11. 《Nature Human Behaviour》研究者，2025 年在线发表。 Work Time Reduction via a 4-Day Workweek Finds Improvements in Workers' Well-Being，第 9 卷，2153 – 2168 页。类型：同行评审的纵向组织干预研究，涵盖六国 141 家机构 2896 名员工，并有对照企业。支持：有意缩短工时可以伴随部分福祉指标改善。局限：研究未测试人工智能影响，机构参与也不是一般性随机分配。

12. 美国劳工统计局，方法参考。 What Is Labor Productivity?。类型：对统计概念的官方解释。支持：劳动生产率是实际产出除以工时，不等于把自报节省的时间直接折算为现金成本。

13. Alexander Bick、Adam Blandin、David Deming，2026 年 1 月 20 日在线发表。 The Rapid Adoption of Generative AI，Management Science，72(10):8440 – 8456。类型：美国同行评审调查研究。支持：2024 年末一轮调查中，自报节省时间估值相当于总工时的 1.4%。局限：调查估计，不是对工资或收入的直接测量。