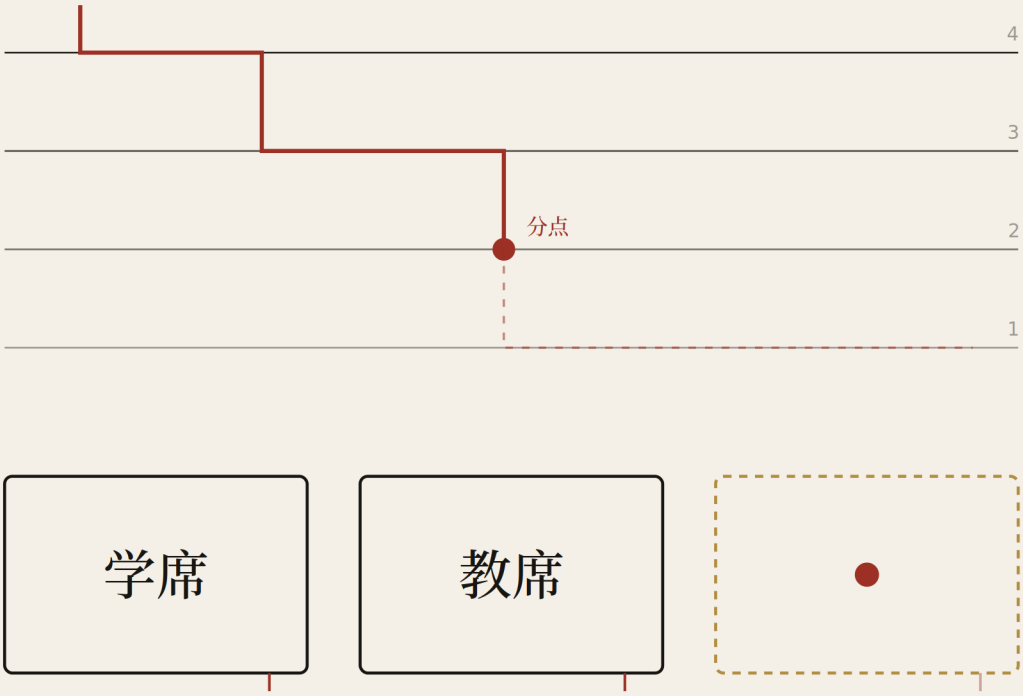
学席和教席的诞生



AI 时代的教育变革

王德生　著

德麦国际出版社　Demai International Press

学席和教席的诞生

学席和教席的诞生

AI 时代的教育变革

王德生　著

德麦国际出版社

Demai International Press · 新加坡

出版信息

学席和教席的诞生

AI 时代的教育变革

王德生　著

德麦国际出版社（Demai International Press） 新加坡 · Demai International Pte. Ltd.

德麦国际编号专著　第 111 号

ISBN 979-8-90690-043-2

定价　US$ 20.00

2026 年 9 月第一版

版式　A4　·　十卷一百七十二章　·　约 20 万汉字

关于书号　本书 ISBN 为 979-8-90690-043-2：本社号段 979-8-90690-0NN-C 中 043 经全站清点未被指派，校验位由 ISBN-13 算法算出（奇位×1、偶位×3 求和，10 减其个位数），以本社在用各号反算全部命中。备用号 044-9、045-6、046-3。

AI 使用披露　本书的承重判断由著者提出，并在写作过程中多次被著者自己推翻或收窄。成文为著者与 Claude（Anthropic）协作完成：卷五至卷九的多数章节由著者先以对话方式跑出独立研究稿，再经协作方改写、去术语化并并入本书；协作方在多处主动写入对本书不利的内容（尤以卷五《容器元数律》退回著者原先的会计比喻、卷八《退单教席》推翻著者原先的「退单率」读数、卷九《产权悬置》推翻著者原先的「否决记录归人端」为甚），这些反对意见均按原样保留，未被削弱。本书未经外部同行评议。**书中全部读数、判据与实验设计，没有一条在真实课堂或真实机构中跑过**；卷五引用的三条历史序列与卷八、卷九引用的行业规程为公开材料，其余标注〔未核验〕者一律未经独立核验。

勘误与讨论　sdeuniverses.com

著者简介

王德生，德麦国际（Demai International Pte. Ltd.）创办人、总裁。中国科学院计算数学博士，并于该院完成博士后研究，方向为 Voronoi／Delaunay 剖分、质心 Voronoi 图与有限元方法。曾任职于英国斯旺西大学与新加坡南洋理工大学。2005 年起定居新加坡，工作往来于新加坡与中国大陆之间。

他的研究由计算数学转入本体论，二十余年间走过四站：《三视角·知识的结构和体系》（2015–2021）、「321 智慧系统」（2021–2025）、SIO 本体论（2025）、以及自 2026 年起的现行框架。其贯穿的问题只有一个：**同一个存在，为什么在不同的人那里显出不同的东西**；而在 2026 年，他把这个问题的答案从「视角不同」推到了「单位不同」——这本书是那一步在教育上的落地。

他在德麦国际出版社出版专著多部，另在 sdeuniverses.com 持续发表长文与研究稿。本书是他第一本完全围绕「评估单位」这一主题写成的著作，也是「AI 时代的教育变革」系列的第一本。

前言

一、这本书是被一个问题逼出来的

2026 年 9 月的一个下午，一个问题把我拦住了：**用 AI 写了一篇很有创新的论文，怎么评？**

这问题看起来是技术的，答起来却怎么答都不对。归给人，人写不出；归给 AI，AI 没有那道题；归给「人机合作」，合作不是主体，不能署名、不能负责、不能被录取。我把这三条路各走了一遍，走到头都是墙。

于是我退了一步问：是不是问法本身有问题？评估这件事，预设了一个被评的对象。那个对象是谁？我们说「评这个学生」，可被评的那篇论文是学生和模型一起显出来的；我们说「评这个教授」，可他签的那份判断是他和系统一起做出来的。如果被评的东西不是一个人，那么「评人」这个动作，评的是什么？

再退一步，问题变得更大了：**评估什么时候评过人？**厨师的手艺离不开刀、灶、厨房和那条鱼；飞行员的技术离不开飞机；科学家的发现离不开仪器。我们一直在评「人—器具」这个整体，只是器具不作声，人就独占了署名。两万年来这个默认从没出过事，因为刀不会说「这一刀是我切的」。

AI 是第一把会说话的刀。

二、母题

全书的脊梁可以压成一句话：

评估从未真正评过人。AI 只是让这件事第一次无法再装作评的是人。

这句话是反直觉的，因为我们每天都在评人;它又站得住，因为那些被评的「人」，没有一个是脱离器具单独存在的。这句话也是可以被推翻的：如果能找到一种评估，它真的只评人、与器具无关，母题就错了。本书第一卷用三章处理了三个最像反例的情形——闭卷考试、双胞胎、新生儿——结论是它们说的都是本书的话，只是被当成了它的反面。

三、这本书怎么长成的

它不是一次写成的，是被一连串反驳逼成的。

最初只有一篇论文。写完之后，我把同一道题反复交给对话系统去跑，跑一趟得到一个答案，再把那个答案降为下一趟的问题。学生端这条线跑了五趟：第一趟说「新单位拒绝被记成一个单位」，第二趟问「失真的分数为什么不退场」，第三趟问「单位死了机器为什么照转」，第四趟问「记账单位到底由什么决定」，第五趟才落

到「名字之下怎么登记」。每一趟都把上一趟的答案降级为结果，把原因往下挪一层。到第四趟才挪到底：**记账单位不由生产决定，也不由测量决定，由分配容器的元数决定**——录取每格只收一个名字，耦合场进来就被切回一个名字。

这个过程里，有三处是我自己被推翻的，我把它们原样留在书里，并在此逐一说明：

第一处，会计那一侧。 我原本主张成绩单在会计上「只记贷方不记借方」——发出时确认资格，永不确认负债。这个说法被退回了：成绩单不是报表科目，可迁移的不是科目而是确认门槛。准确的说法是「只确认资格效力，不确认失效暴露」。更要紧的是，那一趟顺带给出了「教育审计为什么不可能」的真正原因：不是鉴证者不独立，是被鉴证的对象由发行者定义——三权（定义权、计量权、发证权）同在一校，程序独立可以移植，定义独立不能。

第二处，教席的读数。 我原本以为「退单率」可以当教席的考核量。被推翻得很干净：确认根本不是一个动作，是系统为结账自动填充的默认值，不携带任何人的签名；因此确认与否决不在同一分母上，任何「采纳率／推翻率」在记账层就是类别错误。唯一可用的是「命中且未被 AI 吸收的否决」——老师退对了但模型下一版照着改了，那条判断已经进入模型的边界，该否决自动出账。

第三处，否决记录归谁。 我原本在产权草案里写「否决记录归人端」。这一条也站不住：学生不能在没有界面、字段、存储与锁定机制的情况下生成完整记录；学校只提供记录成立的条件，并不因此成为判断内容的作者；平台的存储与格式控制也不能转化为对拒绝语义的所有权。正确的第一动作不是确认归属，是保全——**先保全，后归属**。

这三处我都没有抹掉重写，因为一本书能提供的最有用的东西，往往不是它的结论，而是它在哪里被自己的材料挡了回去。

三之一、五趟接力的完整账

第三节说学生端跑了五趟。这里把五趟各自留下什么、各自被什么退回，逐趟写清——它同时是本书卷五的阅读顺序。

第一趟问：为什么旧账本记不了「学生—AI」？答：不是换了单位，是新单位拒绝被记成一个单位；账本上根本没有一栏叫「来源」。它留下的问题是：拒绝的原因在哪。

第二趟问：失真的分数为什么不退场？答：分数有两种功能——测量功能与结算功能，二者的维持条件不同。校准可以断，机构之间的可通约性不断；于是失真的分数不但不退场，反而因为高分变多、机构需要更细的区分而流通得更广。这一趟还跑了三条已经发生过的历史序列：合同代写市场在 AI 之前十余年就已扩张、且开放性任务先被商品化；分数通胀持续二十年而分配权重不降，各州的应对是增加分数层级而不是换掉分数；飞行执照与医师再认证把资格真值从入门考试移到执业风险。它留下的问题是：不退场的支撑物究竟是什么。

第三趟问：单位已经死了，机器为什么照转？答：旧单位在两个意义上死了——不再指向学生独立生成（指向死），制度也不再检查这一指向是否成立（检查死）；而分配功能靠「可辩护性」继续运转：程序完备、门外人无法定价、持证者互为人质。这一趟给出的动作是「可重走」——在不可预处理的新材料上现场再生成一次。它留下的问题是：切名的那一刀落在哪。

第四趟问：记账单位由什么决定？答：不由生产决定，也不由测量决定，由分配容器的元数决定。录取、发证、追责、评奖，每一格只收一个名字；耦合场进入时被切回一个名字。这一趟还给了学席一个已经运行了一百年的先例：驾驶证的准驾车型、飞行执照的型别等级——名字单数，名字之下登记器具配置。

第五趟不再是论文，是规格：席名、席配置、按配置分开登记的读数、席范围。一句话——**席不是第二个名字，是第一个名字之下的第二格**。

另有一趟横插进来，位置在第二趟与第三趟之间：**教育审计为什么不可能**。它的答案前面已经提到（三权同在一校），这里补一句它跑出来的真材料：一份运行了三十年的工程教育认证通则，读出来是「定义权外置、计量权留内」——认证机构规定要证明什么，却不命题、不抽样、不换工具、不改写达成度判定。于是认证照过、成绩照失真，两件事同时成立，不需要任何人作弊。

三之二、另外三条线

学生端之外，还有三条线，各自也被推翻过一次，处置写在这里。

教授端：焦虑不是权威下降，也不是管辖权被夺，而是**悬空资格**——责任仍被指定给教授，资格却不能从混合产出中独立记载。五类记录（署名、调用日志、错误追责、同行评价、岗位考核）不再指向同一个人。这一支写死了一个到 2028 年底的制度赌注，赌注的输赢与本书的概念命题分属两层，书里已写明。

悬空理论：把资格、判断、责任三重悬空合成一个理论，给出同源闭合原则——谁参与发生、谁进入记录、谁承担后果，三者应当同源。它还把我原先的「首个不可回补点」从一个点升级为一条前沿：真实平台是并行的，多个节点之间未必存在全序，强行指定唯一的「第一点」会制造虚假的精确。

分点的漂移：这一支给出了本书此前只有定性描述的动力学——下漂是递归回路的默认吸引子，不是使用不当，也不是一般能力衰退。它同时指出了本书与航空类比之间最深的一处断裂：飞行员在进入自动化之前已经有成熟的手动技能，学生的目标判断却可能正是在与工具共作中形成——**低分点未必是空席，也可能是未建席**，两者读数相似，成因与对策相反。

四、这本书主张什么

三件承重物，全书的论证都挂在它们上面。

其一，单位换位，不是扩充。 工具的发明不是给人添了能力，是把「人」这个单位重新划了一次。显微镜—科学家看到的东西，旧单位不是看不清，是没有资格看。评估的对象随之换位。

其二，退守无地。 历史上每一次单位换位之后的悬空，都靠工具还不会做的那一块残余功能着陆：摄影拿走了「像」，绘画退守「看法」；计算器拿走了算术，数学退守「推理」。AI 不留残余，迁无可迁；着陆面只能从「功能」迁到「单位内部」——人端还能否决到哪一层。这个量叫分点，测它的动作叫退单。

其三，席只有读数。 新的评估单位叫席：学席、教席，以及卷九推到十个行业的诊席、审席、码席、译席。席不是一个可以打一个分的实体，它有三种读数——显露物、分点、漂移——缺一即悬空。席最重要的病不是坐错人，是空掉：显露物照样好甚至更好，人端却被自己的否决记录训练成多余的。

五、这本书不主张什么

它不主张 AI 会取代教师，也不主张不会。它不预测任何时间表。

它不主张学生和老师消失。它主张的是他们**从被评的对象降为席上的退单者**——名字不动，名字之下加一格。

它不主张禁用 AI。禁用冻结不了工具端，只会把依赖赶到课外并同时掩盖读数；它也不主张放开不管，放开而不改考题，等于把旧考卷发给新单位，满分与满分之间没有信息。

它尤其不主张「有了新读数就万事大吉」。本书第九卷自己写明了反噬：来源栏会被做成新的纪律标签，中间账会被做成新的表格官僚制，不可训练断点会被做成「每周填一次反思」——**越容易被流程自动计数的东西，越快退化成新的浅层动作**。

六、这本书最可能错在哪里

一处：第四层。

本书把分点的下限定在第四层——判断「这道题该不该这么问」。如果这一层也被代答，席就没有着陆面，本书的方案与以往所有工具的方案一样进保留地。卷十用一整章处理它，结论是：第四层守得住，不是因为 AI 判不了，而是因为**判一道题该不该问的资格来自要为答案负责**；分点的下限不是能力边界，是后果链的终点。这句话把本书的命运交给了一个法律问题而不是技术问题——只要后果的终点还是人，席就有着陆面；如果有一天某个法域接受了「不落在任何人身上的处罚」，本书退场。

我愿意押在这一条上。但读者应当知道，这是全书唯一一处，我把整栋房子压在了一根柱子上。

七、怎么读

赶时间的读者：序、导论、卷八、卷十，够了。

要看论证根据：卷一与卷四——卷一末章把「手艺」写成偏导，给出全书拒绝「折算成一个分数」的记账法；卷四末章给出六问诊断表，可以拿去判断任何一个行业是否正在悬空、会悬多久、出路是迁址还是换算法。

要看历史证据：卷二，十六站。要看哲学位置：卷三，与海德格尔的五处一致、四处分歧。要动手改制度：卷九，其中产权一支须读到卷末——共同生成是发生事实，不是产权结论。

卷五至卷七由多趟独立研究稿改姓合成，各章保留原稿的承重词，卷四末有一张术语对照表，遇到不认识的词按表换算即可。

八、一句交代

这本书是「AI 时代的教育变革」系列的第一本。第一本只做一件事：把评估的地址从人挪到席。其余的事——考题形状、席籍制度、席的产权、教席的培养——都要等这件事定了才能开始。

悬空的判断落不到地上，其他改革都是在空中修房子。

九、一句公道话

写完这本书，我要说一句对它不利的话。

本书全部的力气花在一件事上：证明评估的地址错了，并给出新地址的登记格式。它没有做、也不打算假装做的事是：证明新地址上的读数真的测得到什么。分点、退单深度、漂移、有效否决、命中且未被吸收的否决——这些量在本书里都有定义、有取值程序、有自否决条款，却没有一条有过实测。一本给出一整套读数而没有一个读数跑过的书，最可能的下场有两种：一种是它的判断对而读数不可用，制度照它的诊断去改，却用回了旧的量；另一种是读数可用而判断被绕过，学校把「来源栏」「退单率」做成新的表格，填了三年，什么也没变。

我能做的只有把两种下场都写进书里（卷九的反噬一节、卷八的三个反噬预言），并把可以推翻本书的条件写死：什么样的事情一旦出现，本书即错。这不能保证本书是对的，只能保证它是可以被证伪的——在一个人人都能生成漂亮论证的时代，这是一本书还能提供的、不多的诚实之一。

王德生 2026 年 9 月于新加坡

导读

本书篇幅不小，路线却只有一条。这份导读给出那条线，以及每一卷在线上的位置。

一条线

不可分割 → 单位换位 → 悬空判断 → 三处显影 → 席 → 分点 → 后果链的终点。

不可分割（卷一）：人与器具——材料、工具、环境——不可分割，且没有前件。工具可以缺席，材料与环境不能；三维中只有工具沉默，所以它被当成外挂、人被当成常项。

单位换位（卷二）：每一次重大工具诞生，都重划一次「人」这个单位，判断随之悬空一段，再靠残余功能着陆。十六站，六条律。

悬空判断（卷四）：判断投向的单位已不在原地址。四种形态（评估、责任、资格、许可）、四条分界（滞后、脱耦、测量失效、空转）、六问诊断表。

三**处显影**（卷五、六、七）：学生端为什么拒绝被记成一个单位；签字的教授为什么证明不了自己的资格；资格、判断、责任三重断裂如何合成一个理论。

席（卷八）：定义、命名三约束、登记规格、三读数、空席、教席的读数、分点的漂移与不可训练断点。

分点与后果（卷九、十）：十个行业的席、席籍、产权，以及全书押注的那一层。

各卷一句话

卷	一句话
一	不存在一个先于器具的裸人，可以充当评估的对象
二	工具的发明不是扩充人，是重划「人」这个单位
三	海德格尔说清了为什么评「人」在本体上是错的；本书要答的是在一个用具会说话的世界里该评什么
四	判断投向了一个已不在那里的地址，而且因为投得准、发得快，没有人发现它不在
五	新单位不是拒绝被记，是分配容器每格只收一个名字，耦合场进来就被切回一个名字
六	责任仍在教授身上，资格却不能从混合产出中被独立记载
七	谁仍能恢复并改写使结果成为可能的差异条件，谁才在该事件中拥有闭合的主体位置
八	评的是席，不是人；席只有读数，没有常值；最要防的不是坐错人，是空掉
九	席不是教育专用；否决记录先保全，后归属
十	分点的下限不是能力边界，是后果链的终点

三种读法

一小时：序、导论、卷八第一至四章（席的定义、命名、登记规格、结构）、卷十《第四层守不守得住》。

一天：加上卷一全卷、卷四全卷、卷五第四趟（容器元数律五章）、卷八后半（教席读数与漂移）。

要动手：卷九全卷，配卷八的退单考样题与卷四的六问诊断表。三份东西可以直接印出来用——退单考样题及其答案卡、席籍的登记字段、六问诊断表。

四件不要误读的事

第一，本书不反对使用 AI。 它反对的是使用之后仍按旧地址登记。

第二，「席」不是给 AI 发证。 席名单数，仍是那个人；席配置是名字之下的一格。本书自始至终没有主张给 AI 主体地位——它只回答「AI 在哪」，不回答「AI 是什么」。

第三，读数不是评分。 席的三读数里只有第一项（显露物）是现在意义上的分数；分点与漂移不是分数，且**不可与显露物折算成一个数**。理由在卷一末章：产出函数不可加。

第四，本书一条读数也没有在真实课堂跑过。 卷五引用的三条历史序列与各行业规程是公开材料，其余一律是设计，不是结果。版权页已声明，这里再说一次。

关于术语

本书只造了四个承重的新词：**席**、**分点**、**退单**、**漂移**。其余都是日常语言。卷五至卷七各章保留了原稿的承重词（耦合场、旧账本、来源栏、中间账、责任回收率等），卷四末有一张对照表，按表换算即可：耦合场即席，中间账即席籍，来源栏即席籍耦合栏，责任回收率即退单率，可重走即退单考的现场形态。

术语表在书末。遇到「空席」「未建席」「有效否决」「吸收」「产权悬置物」「容器元数律」这些词而一时想不起来的，直接翻到那里。

目录

导论　母题、路线与读法

母题

一本书要有一条脊梁，这本书的脊梁是一句反直觉的判断：**评估从未评过人**。它反直觉，因为我们每天都在评人——考试评学生，考核评老师，同行评审评作者。它又站得住，因为这些被评的「人」，没有一个是脱离器具单独存在的：被评的学生是坐在书本、教室、课程表里的学生；被评的老师是站在黑板、教材、职称制度里的老师。评估落在「人」上，只是因为器具不作声，人独占了署名。

AI 是第一个会作声的器具。它替你写、替你判、替你否决。器具一作声，「评人」这件事就装不下去了——不是因为人不重要，而是因为署名的独占权没有了。这是全书的起点，也是全书的证伪面：如果能找到一种评估，它真的只评人而与器具无关，母题就错了。

三条承重物

母题下面有三件承重物，全书的论证都挂在它们上面。

单位换位，而非扩充。工具的发明不是给人添了能力，而是把「人」这个单位重新划了一次。显微镜—科学家看到的东西，旧单位不是看不清，而是没有资格看。评估的对象随之换位。

退守无地。历史上每次单位换位后的悬空，都靠工具不会做的残余功能着陆。AI 不留残余，着陆面只能从功能迁到单位内部——人端还能否决到哪一层。这个量叫分点，测它的动作叫退单。

席只有读数。新的评估单位叫席。席不是一个固定的实体，它有三种读数——显露物、分点、漂移——缺一即悬空。席最重要的病是空席：人端被自己的否决记录训练成多余的。

路线

全书十卷。卷一第三章「没有工具的人」是本体论前提的证明，读者若只信一章，信那一章。

卷一立本体论前提：人与器具不可分割且没有前件，工具重划单位，思考无所属，厨师的评估解法在 AI 上断在哪一条；并把「不可分割」的四种版本排开（延展／构成／纠缠／无前件），处理三个反例（闭卷、双胞胎、新生儿），最后把手艺写成偏导——全书拒绝「折算成一个分数」的根据，就在产出函数不可加这一条上。

卷二走一遍工具的文明史，十六站，每站只问三个问题：旧单位是什么，新单位是什么，判断如何着陆。收成六条律。

卷三把「器具」三分为工具、材料、环境，与海德格尔的「在世存在」对照——五处一致，四处分歧，分歧全部由 AI 逼出。

卷四给出悬空判断的通式与史例表，说明 AI 时代的悬空为何最深、为何在教育里不自愈；并分出悬空的四种形态（评估／责任／资格／许可）、与四个近邻的分界（滞后／脱耦／测量失效／空转），以及一张六问诊断表——用它可以判断任何一个行业是否正在悬空、会悬多久、出路是迁址还是换算法。

卷五、卷六、卷七是悬空判断的三处显影，各自原为独立论文并在本书改姓：卷五写学生端，由五趟接力构成：新单位拒绝入账→代理坍缩与空值结算→定义独立律→空转→容器元数律→席的登记规格，每一趟把上一趟的答案降为问题；卷六写教授端——签字的人为何证明不了自己的资格；卷七把资格、判断、责任三重悬空合成一个理论，给出同源闭合原则与节点责任。

卷八是全书的核心：学席与教席的定义、命名、登记规格、结构、病、产权、可比性，一道退单考样题与制度改名；后半卷两支——教席的读数（为什么退单率不能当读数，唯一可用的是命中且未被 AI 吸收的否决），与分点的漂移（下漂是递归回路的默认吸引子，防线是不可训练断点）。

卷九把席推到教育之外的十个行业，并给出实施层的五件事：席籍制度、退单考题库的层分类学、教席的培养、席的产权法草案，以及否决记录的产权悬置——为什么第一动作是保全而不是归属。

卷十先给全书划界总表，再回应十条最可能的反驳，专章论证第四层为何守得住（分点的下限是后果链的终点），最后列出证伪面与未解。

读法

赶时间的读者读序、导论、卷八和卷十，够了。要看论证根据的读卷一和卷四——卷一末章给出「手艺是偏导」的记账法，卷四末章给出六问诊断表，两者是全书最可直接拿去用的两页。要看历史证据的读卷二。要看哲学位置的读卷三。要动手改制度的读卷九，其中产权一支须读到卷末：共同生成是发生事实，不是产权结论。

全书不用任何学派术语。承重的新词只有四个：席、分点、退单、漂移。其余都是日常语言。这是有意的——一本要改地址簿的书，自己不能先用一套别人看不懂的地址。

卷一　本体论：人与器具的不可分割

第一章　焦虑的准确位置

教育界对人工智能的反应，表面看是两派：一派要禁，一派要拥抱。禁的一派在校园里布置检测软件、恢复闭卷、增加口试；拥抱的一派开放 AI 使用、鼓励学生「把 AI 当助手」。两派吵了两年，焦虑没有减轻，反而在两边同时加深。禁的一方发现禁不住——校内不用、校外全用，检测软件把不用 AI 的学生判成用了、把用了的判成没用；拥抱的一方发现，一旦允许，作业整齐得无法区分，满分与满分之间没有任何信息。

这两派共享一个从未被说出的假设：被评的对象仍然是「学生」，被考核的对象仍然是「老师」。禁 AI，是为了把学生逼回一个可以单独被评的状态；拥抱 AI，是在不改考题的前提下允许学生带工具进场。两派都没有碰评估对象本身。

本书的判断是：焦虑的来源不在学生用不用 AI，而在**评估失去了着陆面**。老师手里的分数不知道在评谁，学生拿到的分数不知道在说谁，双方都在替一个已经不存在的东西负责。这种「判断有了、落点没了」的状态，本书称之为**悬空判断**。焦虑，是悬空被身体感知到的形状。

要看清悬空判断的成因，必须退到一个比教育更深的层面：人与工具的关系。

第二章　人与器具的不可分割性

人从来不是单独被评估的。厨师的手艺能离开刀、灶、厨房、食材吗？吃到嘴里的是菜，菜是厨师与刀火、灶台、食材、那个下午的湿度一起显出来的。米其林评的是餐厅而不是厨师，就是老实承认这一点。飞行员的技术能离开飞机吗？科学家的发现能离开显微镜吗？外科医生的判断能离开影像设备吗？

这不是一个「人需要工具」的常识性观察，而是一个本体论层面的认知：**人与器具（材料、工具、环境）是不可分割的**。所谓「人」，从来指的是一团纠缠——人与他手里的东西、脚下的地方、面前的材料——而不是一个可以从这团纠缠里单独抽出来称重的实体。

哲学上碰过这条边的人不少。海德格尔讲工具的「上手状态」，西蒙东讲技术物的个体化，斯蒂格勒讲人本身就是技术性存在，克拉克与查默斯讲延展心智。但他们说的不可分割，多数仍是「人先在，器具延伸了人」——人的一部分外化到了工具上。本书要说的比这硬一格：**不是延伸，是没有前件**。不存在一个先于器具的裸人，然后再拿起工具；「人」这个单位，是在与器具的纠缠中被划出来的。

这一格之差决定了下面全部推论。如果人先在、工具是延伸，那么评估仍然可以评人，只需在评的时候「把工具的贡献扣掉」。如果没有前件，那么「扣掉工具的贡献」就是一个没有意义的动作——就像要求评估一个厨师时扣掉厨房的贡献，剩下的不是厨师，是一个不会做菜的人。

第三章　没有工具的人：单位为什么从来是三维的

前一章说人与器具不可分割，并把「器具」写成三样：材料、工具、环境。本章要证明一件更硬的事：**即便没有工具，人也不是单独的**。一个赤手空拳的人，仍然与材料或环境组成单位。所以单位从来不是「人＋工具」，而是「人—材料—工具—环境」这个三维的整体；工具只是三维里的一维，而且是唯一可以为零的一维。正因为工具可以为零，我们才误以为工具是外挂、人是常项——这个误会，是「评人」这件事能维持两万年的根源。

三个维度各是什么

先把三样分清，因为它们在单位里的位置不同。

工具是「以之做」的东西——手里的，用它去做。刀、笔、显微镜、AI。

材料是「对之做」的东西——手下的，对它去做。鱼、纸、样本、题目。

环境是「在其中做」的东西——周身的，在它里面做。厨房、书房、实验室、课堂；也包括时间、气候、他人、制度。

三者的关系不是并列的三件东西，而是一个动作的三个位置：任何一次「做」，都有以之、对之、在其中三个位置，缺一个位置，动作就不成立。这是三维的来源——不是列举出来的三样，而是一个动作在结构上必然有的三个方向。

工具可以为零

现在拿掉工具，看单位还在不在。

一个人徒手抓鱼。他没有网、没有叉，只有手。但他仍然在一条河里（环境），面对一条鱼（材料）。河的深浅、水的冷暖、鱼的滑溜——这些不是他抓鱼的背景，而是抓鱼这件事本身的组成部分。抓鱼的能力，是「人—河—鱼」这个单位的能力：换一条河、换一种鱼，同一个人的能力读数就不同。评「他会不会抓鱼」，评的从来是这个三维单位在工具维为零时的读数。

一个人徒手攀岩。没有绳、没有钉，只有手脚。但岩壁（材料兼环境）的每一处裂缝、每一块凸起，决定了他能爬多高。攀岩者说「这条线路我爬不上去」，说的不是自己的能力，而是「我—这面岩壁」这个单位的读数。同一个人在另一面墙上是另一个单位。

一个人徒手格斗。没有武器，只有身体。但地面的硬软、场地的大小、光线、对手——环境与材料都在。搏击运动明知这一点，所以规定场地、规定体重级、规定灯光：它冻结的是环境维和材料维，好让评估落在人端。这和体育冻结装备档是同一个动作，只是冻的是另外两维。

三个例子里工具都为零，单位都还在。可见**工具不是单位的必要条件**，它是三维里唯一可以缺席的一维。

材料不可为零

再看能不能拿掉材料。

做事必有所对。没有鱼，抓鱼这个动作不存在；没有岩壁，攀岩不存在；没有题，解题不存在。「做」是一个及物动作，材料就是它的宾语。一个没有材料的人不是一个「更纯粹的人」，而是一个什么也没做的人——他不能被评，因为没有动作可评。

思考看起来是唯一没有材料的动作。沉思者坐在那里，手里什么也没有，面前什么也没有。但他仍然对着什么在想：一个问题、一段话、一个意象。问题就是思考的材料，它有抗性——想错了方向，问题不让你过去。所谓「思考的深度」，评的是「人—问题」这个单位能走多远，换一个问题，同一个人的深度读数就变。所以思考也不是无材料的：它只是把材料放在了脑里而不是手里。

环境不可为零

最后看能不能拿掉环境。

做事必在某处。没有地面，人连站都站不住——直立行走本身就是「人—地面」单位的能力，在水里、在失重状态下，这个能力不存在。空气、重力、温度、光线是最底层的环境，它们不出声，所以最容易被忘掉，但一旦缺席，任何动作都不成立。

往上一层，他人是环境。一个人说话，是对着某个人说的；一个人教书，是在有学生的地方教的。「口才」评的是「人—听众」单位的读数——同一个人对不同的听众有不同的读数。再往上一层，制度是环境：考试、课时、职称，这些是环境的硬化形态，它们规定了哪些动作算数、哪些不算。

所以一个「没有环境的人」也不存在——他至少站在地上、呼吸着空气、活在某个时代。环境维永远不为零。

三维中只有工具是可选的，所以我们把它当成了外挂

到这里可以看出一个结构性的不对称：

维度	可否为零	出声方式
材料	不可	反抗（不成、不让过）
环境	不可	约束（只能这样、不能那样）
工具	**可以**	沉默（听话，退隐）

材料和环境永远在，所以它们太熟悉、太底层，被当成了背景；工具可有可无，所以它被当成了外挂。两种误会合起来，就剩下一个「人」在中间，好像他是唯一的常项。评估就落在了这个常项上。

但这个「人」是一个位置，不是一个成员。他是三维单位里那个「做」的位置——以之、对之、在其中，三个方向汇合的地方。拿掉三个方向，位置就不存在。

「人」这个词在评估里指的从来是这个位置的读数，只是我们把它当成了一个独立的东西。

AI 为什么把三维全部翻出来

AI 是一个工具，占的是工具维。但它做了以往任何工具都没做过的事：**它同时像材料和环境那样出声**。

它像材料一样反抗——它会说「你的前提不成立」「这个论证不通」，用「不」来参与判断，这是材料的抗性，以往只有鱼骨、木纹、问题的难度才有这种声音。

它像环境一样约束——它决定了哪些做法是可能的：一个学生—AI 单位能做的事，由 AI 端的能力边界规定，正如攀岩者能爬多高由岩壁规定。以往只有地面、他人、制度有这种约束力。

于是 AI 一进场，三维一起被翻到台面上：工具维第一次不沉默，而且它的声音同时带着材料的反抗和环境的约束。「人」这个位置第一次没有办法再假装自己是唯一的成员——不是因为 AI 太强，而是因为三维里唯一能沉默的那一维开口了，剩下的两维本来就从未沉默过。

本章的结论

单位从来是三维的：人是三维交汇处的一个位置，材料与环境永远在，工具可以缺席。工具的缺席让我们误以为可以评「人」；AI 的出声让这个误会第一次无法维持。

所以，学席和教席不是因为 AI 才需要的新单位。学生从来是在书本、教室、课程表、同学、考试制度里学习的，老师从来是在教材、黑板、职称、学生里教书的——学席和教席在没有 AI 的时候就已经存在，只是那时三维里的工具维沉默、材料维和环境维被当成背景，席被写成了人。AI 只是让席第一次显出了它本来的形状。

第四章　工具的发明重划单位，而不是扩充人

由此可以看清工具发明的本体论后果。常见的说法是「工具扩充了人的能力」——望远镜让人看得更远，汽车让人走得更快。这个说法暗含人先在、后来长大了。本书认为更准的说法是：**工具的发明不是给人加了东西，是把「人」这个单位重新划了一次**。

显微镜并不是让科学家看得更细，而是造出了「显微镜—科学家」这个新单位；这个单位看到的东西，旧单位根本没有资格看——不是看不清，是不在旧单位的视野之内。汽车不是让人走得快，是造出了「车—人」这个新单位，这个单位的位移方式和旧单位不是一回事。每一次重大工具都是重划边界，不是加面积。

评估的对象随之换位。被评的不再是「人」，而是那个新单位。但这件事人类很少明说，因为在多数情况下，工具端是沉默的：刀不会说「这一刀是我切的」，灶不会争功。工具不作声，人就独占了单位的署名——「这是厨师做的菜」。这句话在本体上是不准确的，但在实践上没有出过问题，因为纠缠态里只有人这一端会判断、会否决、会回写。

人类其实早已为「评估单位是人—工具」立过法，只是立在体育里。泳衣、碳板跑鞋、假肢——体育界从没否认被评的是「人—装备」单位，它做的是**冻结装备档**：规定所有人穿同一档泳衣、用同一档跑鞋，让所有单位在同一装备档上可比。评的仍是单位，但通过固定工具端，把差异逼回人端。考试禁不禁计算器，是同一道题的教育版：禁，评的是旧单位；不禁，评的是新单位——但必须重出题，因为旧题对新单位是零区分度。

第五章　思考无所属：车与直立行走

人工智能之所以引起比以往工具更深的焦虑，是因为它触碰到一条惯性认知：思考属于人，且只属于人。

这条惯性有名字，而且发作过好几轮，每轮动作一模一样：**每当工具接管一项能力，人就把那项能力从「思考」里除名**。「计算员」在二十世纪四十年代是一个职业，算数是脑力劳动；计算机出来以后，算数被降为「机械操作」，思考往后退一步。下棋曾是智力的标志，深蓝之后变成了「搜索」。翻译、作画、写文章，每一项都是同一套：能力被机器分走，人就把「思考」重新定义为机器还不会做的那一部分。特斯勒的观察是，所谓人工智能就是「还没做出来的那些」。所以「思考属人」从来不是一个发现，而是一条不断后撤的防线——它守的不是思考，是「人」这个旧单位的独占权。

一个类比可以把这条惯性看得更清：在汽车发明之前，人大约也会认为只有人才会直立行走。但车没有学会直立行走，车做的是让「位移」从「腿」上脱开。人们错的不是「只有人会走」，而是把**功能和器官绑死**——以为位移属于腿。同样，惯性认知的错不在「只有人会思考」，而在于**把思考当成某个器官的属性**——以为思考属于脑，所以属于人。AI 做的和车一样：不是学会了人的思考，是让「判断的发生」从脑上脱开。

顺着这条走到底，会发现惯性认知的错误比「思考属人」更深一层：**思考根本不是任何一端的属性，是单位的显露**。「车—人」单位才位移，「脑—纸—笔—对话者」单位才思考。纸笔时代已经如此，只是纸不作声，人把功劳全记在脑上。AI 只是第一个会作声的纸，把这笔从来记错的账翻了出来。

所以真正要破的不是「人独占思考」，而是「思考有所属」这条更老的默认。破掉之后，「AI 会不会思考」就是和「车会不会直立行走」同型的伪问题——问错了单位。「只有人会思考」和「只有人会直立行走」的区别，仅在于后者已经有了车。

第六章　厨师的解法，以及它在 AI 上为什么断

人类一直在评估「人—工具」单位，只是从没把它当本体题做。看厨师怎么被评，就看到我们一直在用的解法。厨师从未离开厨房被评过，但我们照样能说「这个厨师手艺高」，靠的是三个隐性动作。

第一，**固定其余、只变一端**。厨艺比赛给同样的食材、同样的灶台。评的不是「厨师」，是「其余项固定时，换厨师造成的差异」——手艺等于纠缠态对厨师这一端的偏导。

第二，**签收制**。埃斯科菲耶的厨房编制里十几个人经手一盘菜，出菜口只有主厨一个人过目、一个人署名。归属从来不是「谁做的」，而是「谁在出口负责到可以被退单」。

第三，**默认工具不作声**。厨师能独占署名，是因为纠缠态里只有他这一端会判断、会否决。

第三条在 AI 上断了。AI 是会作声的灶台：它会替你决定火候，还会告诉你食材配错了。这时前两条解法都要变形。偏导要对两端同时求——人换了差多少、AI 换了差多少。签收要回答一个厨房里从未出现过的问题：**出菜口那个人有没有能力退单**。主厨能退单，因为他每一道工序都会做；用 AI 写论文的人，如果看不出哪段该退，签收就是空签，署名就是拟制。

厨师那道题给出最硬的一条是：**评估手艺，实际评的从来是退单能力，不是制作能力**。菜由纠缠态出，手艺由否决权定。搬到 AI 上，「用 AI 写的东西怎么评」就有了可操作的答案——不评谁写的，评作者对这份产物的**退单深度**：他能否决到哪一层，哪一层以下他已无法判断。那一层就是人与 AI 在这个单位里的分界，本书称之为**分点**。分点以上归人，以下归单位本身。

还有一处厨房没教过我们、AI 逼出来的：灶台会学。主厨退单十次，AI 端下次就不犯了——分点会往下漂，人端会被自己的否决记录训练成多余的。这是厨房里从没发生过的事。

第七章　不可分割的四种版本：延展、构成、纠缠与无前件

第二章说人与器具不可分割，第三章说这不依赖工具在场。但「不可分割」是一个被说过很多次的词，说法之间差别很大，而这些差别决定了评估能不能落在人身上。本章把四种版本排开，指出本书站在最后一种上，并说明为什么前三种不足以支撑后面全部推论。

第一种：延展。 心智延展到工具上——记事本是记忆的一部分，计算器是算术能力的一部分。这一版本的结构是「主体先在，边界外移」：有一个心智，它把自己的一部分功能放到外面。克拉克与查默斯的记事本论证是它的标准形态。对评估的含义是：可以评人，只需在评的时候把外移的部分算进来或扣出去。学校的做法——闭卷把工具收走，开卷把工具留下——正是这一版本的两种操作。

第二种：构成。 工具不是延展，而是认知系统的构成成分：航海中的定位不是船长一个人完成的，是船长、海图、罗盘、值更表、船员一起完成的，哈钦斯的舰桥研究是它的标准形态。这一版本比延展硬一格：被研究的单位从一开始就是系统，不是人加装备。对评估的含义是：应该评系统。但它留下一个问题——系统不发文凭。没有人能拿着「这个系统的表现」去应聘。

第三种：纠缠。 主体与客体在实践中共同生成，边界是实践切出来的，不是先有的。行动者网络理论与新物质主义属于这一支。对评估的含义是：评估对象本身是被评估装置切出来的，因此谈论「真实的能力」没有意义。这一版本最彻底，也最容易变成什么都不能评。

第四种：无前件。 本书站的位置。它接受构成与纠缠的结论，但只主张一件更窄的事：**不存在一个先于器具的裸人可以充当评估对象。** 它不主张不能评，只主张不能评「人」这个前件。区别在于：延展说主体先在而边界外移，构成说单位是系统，纠缠说边界由实践切出，无前件说的是——你可以切、可以评、可以登记，但你切出来的那个东西不是原来那个「人」，登记簿上的名字必须跟着换。

四种版本的差异可以用一句话检验：**给一个人和他的工具打一个分，这个分在说谁？** 延展说在说这个人（工具算作他的一部分）；构成说在说这个系统；纠缠说这个问题本身预设了一个不存在的稳定所指；无前件说——它在说一个席，席有名字，名字之下必须写明配置。

前三种在 AI 之前都够用，因为工具沉默，四种版本的实践后果一样：评人。AI 让工具端开口之后，四者分道：延展会把 AI 的判断算成人的能力，构成会把评估推给一个不能签字的系统，纠缠会取消评估本身，只有无前件既保留评估、又换掉被评的东西。这就是本书在这一处必须比前人硬一格的理由——不是要更激进，是要在激进之后还能发证。

第八章　不可分割的三个反例：它们为什么不成立

一个本体论主张不给出它可能错在哪里，就只是一种说法。本章处理三个最常被提出的反例。

反例一：闭卷考试证明人可以被单独评估。 收走一切工具，只留一个人和一张纸，这不正是在评那个裸人吗？

回答：闭卷收走的是可携带的器具，收不走环境与材料。考场是环境：桌椅、时限、监考、题面的排版、答题卡的格式，全部参与了这次作答——同一道题给同一个人，改成口头问答、改成三小时、改成在家里的书桌上，结果不同。材料也在：那道题本身的抗性决定了这个人能走多远。所以闭卷不是「无器具状态」，是一种**器具配置被强制固定的状态**——它固定了工具维（为零），并没有取消另外两维。体育的冻结装备档做的是同一件事。闭卷的合法性不来自「评的是纯粹的人」，来自「所有人在同一配置下可比」。这个理由是好的，但它承认的恰恰是本书的前提：被评的是配置下的表现，不是人。

反例二：双胞胎反例——同样的工具、同样的环境，两个人表现不同，差的那一截不就是「人」吗？

回答：是，而这正是本书主张的东西，不是它的反例。固定其余、只变一端，读出的差值叫手艺（第六章）。但请注意这个差值的性质：它是**一个偏导，不是一个实体**。偏导依赖于在哪一点取——同样两个人，换一种工具配置，差值可能反号：擅长独立推导的人在无模型配置下领先，擅长判断与否决的人在自选模型配置下领先。「人」不是那个在所有配置下都守恒的东西；能在配置之间比较的，只有对配置的依赖方式本身。这就是为什么席的读数必须与配置绑定，不能折算成一个数。

反例三：新生儿反例——一个还不会用任何工具的婴儿，难道不是纯粹的人？

回答：婴儿不使用工具，但婴儿在环境与材料中：地面、重力、母亲的怀抱、乳汁的温度。第三章说过，工具维可以为零，材料维与环境维不能。而且更要紧的是：婴儿不是评估对象。本书讨论的不是「人是否存在」，是「能力如何被登记」。能力从来只在做事中显露，做事从来在三维里。一个从未做过任何事的人没有能力读数——不是因为他能力为零，是因为没有可读的东西。

三个反例合起来指出本书主张的准确形状：**它不否认人的差异，它否认差异可以脱离配置被登记。** 反例一说的是配置被固定，反例二说的是差异是偏导，反例三说的是能力只在做事中显露。三者都是本书的说法，只是被当成了它的反面。

第九章　手艺的偏导：从厨师到席位配置

第六章说，评估手艺实际评的是退单能力。本章把这句话做成可计算的形式，因为后面全部读数都从这里长出来。

设一次产出由三维配置共同显出：工具配置 g、材料 m、环境 e，以及人端 h。产出的读数 $y = f(h, g, m, e)$。传统的「评人」假设 y 可以分解为 $y = h +$ 其余项，于是扣掉其余项就得到 h。本书的主张是：**这个分解不成立，因为 f 不可加**——同一个人在不同 g 下不是「基础能力加上工具加成」，而是不同的函数形态。

那么还能评什么？三样。

其一，偏导。 固定 g、m、e，只变 h，读出的差值 $\partial y/\partial h$。这是厨艺比赛的做法，也是标准化考试的做法。它有效，但只在那一点有效：偏导是局部的，不同配置点上的偏导不可直接比较，更不可平均。这是「席读数与配置绑定」这条规则的数学理由。

其二，对工具端变化的响应。 固定 h、m、e，变 g，读出 $\partial y/\partial g$。这在 AI 之前没有意义（工具端不会自己变），在 AI 之后是核心量：换一个基底，这个人的产出变多少？响应大的，说明产出主要由工具端承载；响应小的，说明人端占得多。这就是后文归己率与他择率的来源。

其三，交叉项。 $\partial^2 y/\partial h\,\partial g$——同一个工具在不同人手里的增益差异。它测的是「会用」而不是「能做」，并且它是三者中唯一在 AI 时代增长最快的量：模型越强，会用与不会用之间的差距越大。教育如果要有一个新的培养目标，最可能是它。

三个量与后文的对应：偏导对应显露物读数（旧量表仍管这一段），响应对应他择率与归己率，交叉项对应退单深度与分点。而空席的定义在这里也有了形式：**人端的偏导趋近于零，而对工具端的响应趋近于全部**——产出照样好，但换掉这个人，y 几乎不变。

最后一句要说清楚：这一节的符号不是模型，是记账法。它不预测任何数值，只规定什么东西不可以被相加、什么东西不可以被平均。本书后面所有拒绝「折算成一个分数」的主张，根据都在 f 不可加这一条上。

卷二　工具的文明史与「主体单位」的改变

第十章　一部被记错的账

人类写自己的历史，习惯把工具写成配角。历史书里有石器时代、青铜时代、铁器时代，但讲到人，讲的仍是猎人、农夫、工匠、工人——好像人是一个贯穿始终的常项，工具只是他手里换来换去的东西。本书要把这笔账倒过来记：**每一次重大工具的诞生，都不是给人添了一样东西，而是把「人」这个单位重新划了一次**。猎人、农夫、工匠、工人不是同一个人换了几件工具，而是几个不同的单位——每一个单位都由人与他的材料、工具、环境一起划定，每一个单位的边界都和前一个不同。

这样记账，会看到一条以往被工具的沉默遮住的线：每个单位诞生之后，评估、责任、署名这些落在「人」身上的判断，都要过一段时间才能追上新单位的边界。追上之前，判断悬空。追上的方式，取决于两件事——旧单位登记在哪一级制度上，以及新工具给旧单位留下了多少「还只有人才会做」的残余。

下面按时间走一遍。每一站只问三个问题：旧单位是什么，新单位是什么，判断如何着陆。

第十一章　手与石器：第一次划界

人类学家勒鲁瓦—古朗有一个后来被反复引用的判断：直立行走解放了手，手解放了口，口解放了语言；而手一旦解放，第一件事就是拿起石头。他把石器称作**器官的外化**——切割的功能从牙齿和指甲上移到了燧石上，人的身体边界第一次伸出皮肤之外。

这里的旧单位是「身体」：能力由器官给定，牙能咬到什么程度就是这个动物能咬到什么程度。新单位是「手—石器」。这个单位切得开的东西，旧单位切不开；这个单位能猎到的动物，旧单位只能捡它的尸骸。**新单位不是旧单位的加强版，而是一个不同的动物**——考古学正是用石器的形制而不是骨骼的差异来划分人属早期的阶段，这本身就承认了单位是由人与石器一起划定的。

石器时代太久远，看不见「评估」。但可以看见责任的雏形：石器的加工技术是传承的，某个群体的石器比另一个群体锋利，这个群体就活得更好。「手艺」这个概念在此诞生——它从一开始指的就是人—石器单位的能力，而不是手的能力。一双再灵巧的手，没有燧石就没有手艺。

第十二章　火：把消化外化

火比石器更彻底。灵长类学家兰厄姆提出，烹饪是人类进化的转折点：熟食让消化系统可以缩短、脑可以变大，人的身体形态本身是被火重塑的。这意味着火不仅外化了一项功能，还**回写了身体**——人—火单位使用了几十万年之后，人的肠道已经短到离开火就活不下去。

这是第一次看到工具对人端的「漂移」：单位形成之后，人端并不保持原样，而是被单位本身改造，直到旧单位无法独立存在。现代人没有一个是离开火还能长期存活的——不是因为不会生火，而是因为身体已经不再是那个可以生吃一切的旧单位。

火还带来了第一个「环境」意义上的器具：火塘。火塘是一个地方，人围着它坐，语言、分配、教育在它周围发生。人—火单位的边界不只画到火本身，还画到那个被火照亮、被火取暖的空间。器具从此有了三个位置：手里的（工具）、手下的（材料）、周身的（环境）。

第十三章　语言与文字：记忆的外化与第一次有记录的悬空

语言不是工具，或者说它是一种没有物质载体的工具——它外化的是思维本身，让一个人的判断可以进入另一个人的头脑。人—语言单位的能力是「集体记忆」，这个单位记得住的东西，任何一个单独的脑都记不住。

文字把这件事推到物质上。记忆从脑移到泥板、莎草纸、竹简上，人—书写单位第一次可以「记住」远超脑容量的东西。而正是在这里，历史上留下了第一份关于悬空判断的原始记录。

柏拉图《斐德罗》篇里，苏格拉底转述埃及神话：发明文字的神忒乌斯把文字献给国王，说它是记忆的良药；国王回答说，恰恰相反，文字会让人依赖外在符号而不再从内部回忆，它给的不是记忆而是提醒，不是智慧而是智慧的外表。

这段话的结构，和今天教育界对 AI 的反应一模一样。「**记住**」**这件事已经在人—书写单位上发生了**——一个能查阅文本的人，记得住的东西比任何吟游诗人都多——但苏格拉底评估的仍是旧单位「脑」，他问的是「这个人自己记不记得」。判断射向了一个已经被重划的边界。

悬空持续了很久。希腊教育在文字普及之后几百年仍以背诵荷马为核心，中世纪经院教育仍以记诵为基础，中国的科举到清末仍要求默写经文。评估追上新单位的方式，是慢慢把地址从「记得住」迁到「读得懂」「辩得过」——从记诵到诠释，从诠释到论辩。迁址完成的标志，是「学者」这个词的含义从「记得多的人」变成「能读会写会论证的人」。

值得记下的是着陆的机制：文字留下了一块残余——文本记住了内容，但**理解**内容、**比较**文本、**论证**观点仍只有人会做。评估就落在这块残余上。这个模式此后每次都会重复。

第十四章　犁、牛与田：单位第一次包含土地和季节

农业把单位的边界扩到了前所未有的范围。猎人—弓箭单位的边界画在身体周围几十米；农夫—犁—牛—田—季节单位的边界画在几十亩地和一整年上。

这个单位的特点是，**环境成了单位的正式成员**。田不是背景，田是工作的对象兼工作的条件；季节不是外部约束，而是单位运转的节律。农夫的「手艺」——什么时候下种、怎样看天——一半是关于工具的知识，一半是关于环境的知识。旧单位（采集者）的能力是「认得出哪些能吃」，新单位的能力是「让一块地在一年后长出可以吃的东西」，两者没有共同的量表。

评估在这里第一次有了制度形态：税。税是国家对农夫—田单位产出的读数，而它一开始就登记在「田」而不是「人」上——田赋按亩计。这是历史上少见的一次评估直接落在单位而不是人端的例子，原因很简单：土地不会跑，人会。当国家需要一个稳定的评估地址时，它选了单位里最不动的那一端。这个选择在后文讨论「席」的时候会再出现。

第十五章　轮、马镫与骑兵：军事单位的重划

轮和车让位移从腿上脱开，马镫让人—马单位从「骑手」变成「骑兵」。历史学家怀特有一个著名而有争议的论断：马镫使骑手可以在马上稳定地持长矛冲锋，由此产生了重装骑兵，重装骑兵需要昂贵的马匹和装备，供养它的制度就是封建采邑。不管这个因果链站不站得住，它指出的东西是准确的：一个新的人—器具单位出现之后，整个社会的评估与分配制度要围着它重组。

骑兵这个单位延续了一千多年，到二十世纪还留下一段清楚的悬空记录。汽车和坦克出现之后，各国军队的骑兵考核仍以马术为核心，骑兵军官仍按马上功夫晋升，这个状态一直维持到第二次世界大战。旧单位「骑兵」的评估地址登记在军制上，锁死程度极高；而新工具（机械化部队）没有给旧单位留下任何残余——坦克能做的，骑兵没有一件做得比它好。**锁死且无残余**，结果是旧单位整个进入保留地：骑术进入奥运会，成为体育。

这是第一个看得清「隔离出口」的案例。评估没有迁址，因为无处可迁；旧单位没有消失，因为制度不让它消失；于是它被保留在一个规定了装备档的封闭场地里，在那里继续被评，代价是和真实的战争脱钩。

第十六章　钟表：把时间做成单位的一部分

芒福德在《技术与文明》里说过一句常被引用的话：工业时代的关键机器不是蒸汽机，而是钟表。他的理由是，钟表让时间脱离了太阳和季节，变成可以分割、计量、买卖的东西；而一旦时间可以买卖，劳动就可以按时间计价，人就可以按小时被雇佣。

人—钟表单位是一种奇特的单位：钟表不在手里，而在墙上、在塔上、后来在口袋和手腕上，它不参与任何具体的操作，却重划了所有操作的边界。修道院的日课、工厂的班次、学校的课时——每一个人—工具单位都被嵌进了人—钟表单位里。评估从此有了一个新的通用地址：**时间**。工时、课时、学时，都是钟表给评估提供的读数。

这个地址至今仍在教育里：学分按学时计，教师工作量按课时计。而 AI 使这个地址失效的方式很直接——一个学生—AI 单位一小时能显出的东西，旧单位一周都显不出；课时不再测任何东西。钟表给的地址，是所有悬空地址里最隐蔽的一个，因为它看起来根本不像一个评估标准，只像一个计量单位。

第十七章　印刷：抄写者的悬空与「博学」的迁址

印刷术在欧洲普及后的一百年，是一段可以细看的悬空期。艾森斯坦的研究把它讲得很清楚：印刷不只是让书变多，它改变了「知道」这件事的形状——固定的文本、可比对的版本、可引用的页码，这些东西让「学问」从记诵和口传变成了阅读、比对和批判。

旧单位是「学者—手抄本」，它的能力是记得住、抄得准、背得出。新单位是「学者—印本」，它的能力是读得广、比得出异同、引得出出处。悬空表现在两处：一是抄经士的手艺失去了落点——抄得再准，也不如一台印刷机；二是大学的考试仍以复述权威文本为主，这套考法对新单位失去了区分度，因为任何人都可以查到权威文本说了什么。

着陆花了一百多年，大学的考试形式从复述迁到论辩（disputatio）再迁到论文——「博学」的含义从「记得多」迁到「读得透、辩得清」。残余是明确的：印本记住了一切，但比较、批判、综合仍只有人会做。评估落在这块残余上。

抄经士这个单位没有进入保留地，它直接消失了——书法作为艺术留下来，但那已经不是抄经士，是另一个单位。这一点和骑兵不同：骑兵有军制撑着，抄经士只有行会，行会撑不住。**锁死程度决定旧单位是进保留地还是直接解散**。

第十八章　望远镜与显微镜：新单位看到旧单位没资格看的东西

伽利略把望远镜指向木星，看到了四颗卫星。这件事最深的地方不在于他看到了什么，而在于一个从未出现过的问题：**这是谁看到的？** 是伽利略的眼睛，还是那根镜筒？当时的反对者提出的质疑正在这里——他们认为望远镜里的像可能是镜片制造的幻象，眼睛看不到的东西凭什么算数。

这场争论，是关于新单位的认识资格的第一次正面交锋。旧单位「眼睛—天空」看到的东西有资格叫「观察」；新单位「眼睛—望远镜—天空」看到的东西有没有资格，需要一场辩论来确立。辩论的结果是新单位赢了，代价是「观察」这个词从此不再指眼睛的能力，而指仪器—观察者单位的能力。科学史家沙宾与谢弗后来用「见证」这个概念描述这段变化：一个观察要算数，需要的不再是一双眼睛，而是一套仪器、一个可重复的程序和一群能验证的人。

显微镜把同样的事做了一遍。列文虎克在水滴里看到「微小动物」，皇家学会一开始不信，派人重做才承认。此后「看见」在生物学里就永远是「显微镜—眼睛」单位的事。拉图尔的实验室研究把这一点说到底：科学事实是仪器、材料、程序、研究者一起显出来的，去掉任何一端，事实就不成立。

这一站给本书的贡献是：**新单位看到的东西，旧单位不是看不清，而是根本没有资格看**。单位换位不是能力的量的增加，而是资格的质的重划。这一点对 AI 至关重要——学生—AI 单位能显出的判断，不是旧单位「学生」写得差一点的版本，而是旧单位根本没有资格写的东西。用旧单位的量表去评它，量表会读出满分，但读不出任何区分。

第十九章　摄影：一场看得见首尾的悬空

摄影是本书这条线上唯一一次可以从头到尾看完整的悬空，因为它发生在有报刊、有评论、有沙龙记录的时代。

1839 年达盖尔法公布后，法国画家德拉罗什据说讲过一句话：「从今天起，绘画死了。」这句话是不是他说的有争议，但它准确表达了当时的悬空感：绘画的旧评估地址是「画得像」——透视准不准、光影对不对、肖像像不像本人——而相机在这个地址上一出手就是满分。旧单位「画家—画笔」在旧量表上瞬间失去区分度，因为任何一台相机都比任何一位画家画得像。

此后大约五十年，是绘画的悬空期。沙龙评审仍按旧地址评画，学院仍教透视与写实，而真正在发生的，是一批被沙龙拒绝的画家在找新的着陆面。1874 年的第一次印象派画展，本体上是绘画宣布迁址：评估地址从「像」迁到「看法」——画的不再是对象，而是光在那一刻落在对象上的样子，是画家的眼睛而不是对象的形状。这块地面相机当时够不着，绘画就落在上面。

摄影这一站有两个别处看不到的细节。其一，悬空期相对短——五十年，比文字和骑兵短得多——原因是旧地址登记在沙龙和学院上，不是国家制度，锁死程度低，换得动。其二，旧单位没有解散也没有整个进保留地，而是**分了家**：写实作为技法留在学院课程里，成为一种被规定了装备档的练习，绘画的主流则迁到新地址上。这是「迁址」与「隔离」两个出口同时打开的唯一案例，之所以可能，是因为残余功能（看法、风格）足够大，大到可以撑起一个新的主流。

第二十章　听诊器与影像：诊断单位的换位

医学提供了一条平行的线，而且它离 AI 时代的教育最近，因为诊断和教学一样，是一种以判断为核心的职业。

1816 年拉埃奈克发明听诊器之前，医生的诊断单位是「医生—眼耳手—病人的叙述」：看面色、把脉、听病人说。听诊器把耳朵伸进了胸腔，「医生—听诊器—病人的身体」这个新单位听到的东西，旧单位听不到。历史学家赖泽在《医学与技术的统治》里描述了随之而来的变化：诊断的证据从病人的话迁到医生的仪器，病人从诊断的合作者变成了被检查的对象。这是一次评估地址的迁移——「什么算病」的判断，从叙述迁到了体征。

X 光、超声、CT、核磁，每一次都把这件事推进一步。到二十世纪末，一个放射科医生的诊断单位是「医生—成像设备—图像处理软件—影像」，人端的工作是在一张机器生成的图上找异常。旧单位的评估地址——问诊的功夫、触诊的手感——在这个单位里没有位置，新的评估地址是读片的准确率。

而正是在读片这一步，AI 在 2016 年前后进场。深度学习模型在若干影像任务上的准确率达到或超过放射科医生，医学界随即出现了和教育界一模一样的两派：一派说 AI 会取代放射科医生，一派说 AI 只是工具。两派共享的假设仍是旧的：被评的是「医生」。而真正在发生的是，诊断已经在「医生—AI」单位上发生，医生的份额从「找异常」迁到了「对 AI 找到的异常做否决」——判断片子里那个标记是真阳性还是假阳性，判断 AI 该不该看这张片子。

医学比教育走得快一步的地方在于：它已经开始把评估地址往人端的否决能力上挪。若干医院对 AI 辅助诊断的验收标准，不是看 AI 加医生的总准确率，而是看医生对 AI 建议的**采纳率与推翻率**——推翻错误建议的能力，是新单位里医生端唯一可测的份额。这正是退单深度，只是医学界叫它「人机协同的质量控制」。

第二十一章　蒸汽机与工厂：人变成单位里的附件

工业革命是本书这条线上最清楚的一段，因为它有一位极其敏锐的记录者。马克思在《资本论》里描述工厂时，用了一个精确的词：工人成了机器的**附件**（Anhängsel）。在手工业里，工具是工人身体的延伸，工人决定节奏；在工厂里，机器决定节奏，工人成了机器体系里一个可替换的部件。

用本书的语言说：手工业的单位是「工匠—工具」，人端占着判断位，工具端沉默；工厂的单位是「机器体系—工人」，判断位被机器的节律占了，人端退到了执行位。马克思看到的「异化」，本体上是**单位内部主导端的换位**——不是人被剥夺了工具，而是人在单位里从退单者变成了被退单者。

评估随之换了地址。手工业评手艺，工厂评产量；手艺是人端的读数，产量是单位的读数。工厂主很快发现，评产量比评手艺容易得多——产量不需要懂行的人来判，只需要一个计数器。这是评估史上第一次大规模地把地址从人端迁到单位的显露物上，而它的后果是人端的读数（手艺）从此在工业里没有了位置。泰勒的科学管理把这件事做到极致：把工人的每个动作拆解、计时、标准化，人端的判断被彻底外化到流程里。

这段历史对 AI 时代的教育有一个直接的警示：**评估地址迁到显露物上，是一种看似解决、实则抽空人端的着陆方式**。如果教育在 AI 之后只看学生—AI 单位的产物质量——文章好不好、题解得对不对——那就是工厂的路：产量上去了，手艺没了，人端成了附件。卷八讨论「席」的三种读数时，「不能只读显露物」这一条，根据就在这里。

第二十二章　铁路、电报与标准时间：单位的边界脱离身体

铁路和电报共同做了一件以前没有发生过的事：让单位的边界脱离身体所在的地方。电报之前，信息的传递速度等于人的移动速度，人—信息单位的边界画在人能到达的范围内；电报之后，一个坐在伦敦的人可以在几分钟内知道孟买发生了什么，「知道」这个动作的边界扩到了整个电报网。

铁路则逼出了标准时间。在铁路之前，每个城镇按当地正午定时；铁路时刻表要求所有站点使用同一个时间，1884 年的国际子午线会议由此确立了时区。这意味着人—钟表单位从此不再是一个人和他的钟，而是全人类和一个统一的钟——单位的边界扩到了全球。

这一站的意义在于，它预演了 AI 时代单位的一个特征：**单位的一端可以在远处，可以是一张网，可以同时属于无数个单位**。学生—AI 单位里的 AI 端，不在学生的书桌上，而在某个数据中心里，同时坐在几百万个席上。这种「一端共享」的单位在电报时代已经出现，只是那时共享的一端不作声。

第二十三章　计算机：「计算员」的消失与「思考」的第一次后撤

「计算机」（computer）在二十世纪四十年代之前是一个职业名称，指以计算为业的人，多数是女性，在天文台、弹道实验室、人口普查局工作。电子计算机出现后，这个职业在二十年内消失，名字被机器拿走了。

这是本书这条线上第一次可以看清「思考的后撤」。计算在此之前被认为是脑力劳动的一种——数学家的工作有很大一部分就是算。计算机出现后，「计算」被重新分类为「机械操作」，数学家的工作被重新定义为「证明」「建模」「洞察」——那些机器还不会做的。这个动作此后反复出现：机器每拿走一项能力，人就把那项能力从「思考」里除名，把「思考」重新定义为剩下的部分。

在中国，同一段悬空以另一种形式发生过。珠算在二十世纪八十年代之前是小学必修，会计和银行职员靠算盘吃饭，「打得一手好算盘」是一种被评估的能力。计算器普及后，珠算课在几年内退出课程，算盘从会计的工具变成了博物馆展品和少数人的智力训练——旧单位「人—算盘」进了保留地，以「珠心算」的形式被保留在一个规定了装备档的小场地里。这个过程只用了十几年，比欧洲计算器进课堂的争论还短，原因是珠算的评估地址从来没有登记在国家考试上——它是职业技能，不是升学科目，锁死程度低。

评估地址的迁移在这里也很清楚：算得快、算得准，从数学能力的核心指标降为无关紧要的技能；数学教育的重点从计算迁到推理。计算器进入中小学课堂时引起的二十年争论，就是这个迁址在教育里的悬空期，最后以「分场」着陆——低年级禁，高年级准，既保留旧单位的评估，又承认新单位的合法。

这一站还留下了一个以后会变得极重要的先例：**人—计算机单位是历史上第一个工具端会「回答」的单位。** 算盘不会告诉你答案对不对，计算机会。但它的回答仍限于它被明确告知的算法之内，它不会对问题本身发表意见，所以「工具沉默」这条默认在原则上还没有被打破——只是裂了一条缝。

第二十四章　互联网与手机：延展心智与随身的单位

克拉克和查默斯在 1998 年提出「延展心智」时用的例子是一个记事本：一个记忆有障碍的人把信息记在本子上随身携带，需要时查阅，这个本子在功能上就是他记忆的一部分，认知的边界应该画在本子外面而不是头颅上。这个论证在当时是哲学上的挑衅，二十年后智能手机把它变成了日常事实——没有人再怀疑「知道」这件事的边界已经画到了手机上。

人—手机—网络单位是第一个几乎所有人都随身携带、随时开着、不假思索地依赖的单位。它的评估后果在教育里已经显出：开卷考试与闭卷考试的区别，本质上是允许评新单位还是坚持评旧单位。多数考试至今坚持闭卷，理由是「要考学生自己会不会」——这句话完整地暴露了旧单位的假设：存在一个「学生自己」，可以从他的手机和网络上剥离出来单独评估。

但这一站的工具端仍然沉默：网络提供内容，不提供判断；搜索给你一百条结果，不告诉你该信哪一条。判断位仍在人端。所以尽管单位已经随身，评估的悬空还可以用闭卷这种粗暴的办法暂时压住。

第二十五章　汽车、飞机与自动驾驶仪：空席的预演

汽车让「位移」从腿上脱开，这一点前面已经用过。这里要看的是它之后的一步：驾驶本身被自动化。

飞机比汽车先走到这一步。自动驾驶仪在二十世纪三十年代进入民航，到七八十年代，巡航阶段的操作已经基本交给机器，飞行员的工作从「驾驶」变成了「监视」。1983 年，人因工程学者班布里奇发表了一篇后来被反复引用的短文，题目叫《自动化的讽刺》。她指出一个悖论：自动化把常规操作交给机器，正是为了减轻人的负担；但当机器失效时，需要接手的恰恰是那个已经因为长期不操作而生疏的人。自动化程度越高，人接手时的能力越差，而接手的时刻恰恰是最需要能力的时刻。

用本书的语言说，班布里奇描述的正是**空席**：飞行员—自动驾驶仪单位运转多年之后，人端的否决能力逐年下沉，显露物（航班的安全记录）反而越来越好，直到某一次机器把控制权交还的时候，席上已经没有一个能退单的人。2009 年法航 447 航班的事故调查报告里，这个结构清清楚楚——自动驾驶仪因传感器结冰断开，机组接手后无法正确判断飞机状态。

航空业对这个问题的应对，是历史上第一次有制度把评估地址从显露物挪向人端的否决能力：定期模拟机复训，专门考飞行员在自动化失效时的手动接管，而不是考航班准点率。这就是一种「退单考」——不评单位显出了什么，评人端在机器出错时能退到哪一层。它是本书所主张的席位读数在航空业里的原型，只是航空业没有把它上升为本体论，而是当作一项安全规程。

汽车的自动驾驶正在重演这段悬空。驾照考的仍是旧单位「人—方向盘」，而路上跑的已经是「人—辅助驾驶系统」单位；事故责任追到人，人说是系统在开；追到系统，系统不是主体。汽车这一站还没有着陆，它和 AI 时代的教育是同一个悬空期里的两个行业。

第二十六章　人工智能：会作声的工具

到 AI 这一站，前面十二站积累的所有模式同时发作，并且有两条被打破。

被打破的第一条：工具沉默。 从石器到网络，所有工具端都不判断——它们执行、储存、传输、计算，但不对问题发表意见、不否决人端的选择。AI 是第一个会作声的工具端：它会替你决定写法，会告诉你论证有漏洞，会拒绝你的前提。单位内部第一次有了两个判断位。这意味着前面所有站用来着陆的第一个动作——「固定其余、只变一端」——要对两端同时做。

被打破的第二条：残余着陆。 从文字到计算机，每一次悬空都靠「工具还不会的那一块」着陆：文字留下理解，印刷留下批判，摄影留下看法，计算机留下推理。AI 没有留下一块固定的地面——它会写、会算、会推、会评、会有风格，残余在收缩。前面所有站的「迁址」出口在这里堵死了。

于是 AI 时代的悬空，只剩下两个可能的出口。一个是骑兵的出口：旧单位进保留地——闭卷、禁 AI、把「学生自己」关在一个规定装备档的房间里考，代价是这个房间和真实的学习脱钩，教育变成体育。另一个是本书主张的出口：既然残余功能不再是一块固定的地面，着陆面就必须从「功能」迁到「单位内部」——不再问「人还会做什么而 AI 不会」，而是问「在人—AI 单位内部，人在哪一层还能否决」。这就是卷八里「分点」与「退单深度」的由来。

第二十七章　贯穿十六站的几条律

把十三站放在一起，可以读出六条规律。前五条是通史，第六条是 AI 的特例。

外化律。每一次重大工具，都把一项原本由器官承担的功能外化到器具上，人的边界随之外移。石器外化切割，火外化消化，文字外化记忆，钟表外化节律，机器外化力量，计算机外化计算，AI 外化判断。外化不是「添加」，是「重划」：外化之后的单位和之前不是同一个东西。

沉默律。在 AI 之前，所有工具端都不判断，所以人端独占了单位的署名。「厨师做的菜」「学者写的书」「工匠打的刀」这些说法在本体上都不准确，但从未出过问题，因为工具不争功。沉默律是旧单位「人」能维持几万年的原因。

悬空律。每一次单位重划之后，评估、责任、署名都要过一段时间才追上新边界。追上之前，判断悬空。悬空期的长度由旧地址的制度锁死程度决定：登记在国家级制度上的（科举、军制、考试）悬空最长，登记在行会、沙龙上的最短。

残余律。悬空的着陆面是工具端还不会做的那一块。有残余，评估迁址到残余上，新职业诞生；没有残余，旧单位进保留地，成为体育或技艺。两个出口之间的分岔，由残余是否存在决定；进保留地还是直接解散，由锁死程度决定。

回写律。单位形成之后，人端不保持原样，而是被单位改造。火缩短了肠道，钟表重塑了睡眠，自动驾驶仪磨钝了飞行员的手，搜索引擎改变了记忆的策略——心理学家斯帕罗等人 2011 年的实验发现，知道信息可以随时查到的人，记住的是「在哪里能查到」而不是信息本身。回写在以往各站都是缓慢的、以代计的；AI 把回写的速度压到了以周计——人端每一次否决都在训练 AI 端，AI 端每一次改进都在降低人端下一次否决的必要。这就是「漂移」的来源，也是空席不是意外而是单位自身的常态趋势的原因。

AI 特例。AI 同时打破沉默律和残余律，并把回写律的时间尺度从代压到周。工具端会判断，人端不再独占署名；残余趋零，迁址无处可去。因此 AI 时代的悬空判断，既不能靠「扣掉工具的贡献」来评人，也不能靠「找到人独有的功能」来着陆，只能把评估地址从「人」和「功能」同时迁走，迁到单位内部的一个可测的量上——人端还能否决到哪一层。

第二十八章　主体单位的谱系

最后把「主体单位」的变化列成一张谱系，看它走过的路。

时代	主体单位	边界画在哪	判断位	评估地址
旧石器	手—石器	身体周围	人端	手艺（口传）
火	人—火—火塘	火光范围	人端	分配、传承
文字	脑—书写	文本可及范围	人端	记诵→诠释
农业	农夫—犁—牛—田—季节	田与年	人端	田赋（落在田上）
骑兵	人—马—镫	冲锋距离	人端	马术（军制）
钟表	所有单位—钟	时间	人端	工时、课时
印刷	学者—印本	可查阅范围	人端	复述→论辩
仪器	观察者—仪器—程序	仪器可及范围	人端	可重复见证
工厂	机器体系—工人	车间	**机器端**（节律）	产量
电报铁路	人—网	全球	人端	标准时间
计算机	人—算法	可算范围	人端（算法端只答不问）	推理、建模
手机网络	人—随身网络	随身	人端	闭卷／开卷
人工智能	人—AI（席）	单位内部	**两端**	分点、退单深度

表里有两处值得停住。

第一，**判断位离开人端**，历史上只发生过一次半：工厂那一次是完整的（机器的节律主导，人退为附件），结果是评估地址迁到产量、人端的读数从工业里消失。计算机那半次不完整，因为算法只答不问。AI 是第二次完整的——而且比工厂更彻底，因为工厂的机器不会对产品本身发表意见，AI 会。工厂那一次的教训是：如果评估只看显露物，人端会被抽空。这个教训直接决定了卷八席的三读数设计。

第二，**评估地址落在单位而非人端**，历史上也有过先例：田赋。国家把税登记在田上，因为田不动。这个先例说明，「席」这种把评估落在单位而非人身上的设计，不是没有过的东西——它只是从来没有被用在人的能力评估上，因为在 AI 之前，人端是单位里唯一会判断的一端，落在人身上和落在单位上没有区别。区别是 AI 造出来的。

第二十九章　工具史的收束

一部工具的文明史，读出来是一部「人」这个单位不断被重划的历史。猎人、农夫、工匠、工人、操作员，不是同一个人换了几件工具，是几个不同的单位。每一次重划之后，判断都要悬空一段，然后靠工具留下的残余着陆。

AI 是第一个不留残余、并且自己会判断的工具。它使得两万年来「评人」这个动作第一次没有办法继续——不是因为人不重要了，而是因为「人」这个词指的那个单位已经不在评估的地址上。着陆面只剩一处：单位内部，人端还能否决到哪一层。给这个着陆面起名字，就是学席与教席。

卷三　器具的三分与海德格尔的「在世存在」

第三十章　为什么要把「器具」分成三样

卷八里用的是「人与器具不可分割」，而「器具」一词被写成三样：材料、工具、环境。这不是罗列，是三个位置。工具是**手里的**——用它去做；材料是**手下的**——对它去做；环境是**周身的**——在它里面做。厨师的刀是工具，鱼是材料，厨房是环境。三者缺一，单位不成立：没有刀切不了，没有鱼切什么，没有厨房无处切。

把三样分开，是因为它们在单位里的行为不一样。

工具沉默而听话。 刀不争功，也不反抗；它的全部意义在于被用而不被注意。一把好刀在使用时是消失的，只有钝了才被看见。

材料沉默但反抗。 鱼不说话，但它有骨、有纹理、有腐败的时限；切法必须顺着它来。人类学家英戈尔德批评过一种他称为「形质论」的思路——以为形式是人从外面加到惰性材料上的——他的反例是木匠：木匠不是把设计强加给木头，而是在木纹的抗性里找到一条能走的路。材料是单位里一个不作声但有立场的成员，它用抗性参与判断。

环境沉默且约束。 厨房不做菜，但厨房决定了哪些菜做得成：灶的火力、台面的大小、出菜口的位置、其他厨师站在哪里。环境不是背景，它是单位的边界本身——单位的边界画到哪里，环境就在哪里。前文讲农业那一站时说过，农夫的单位要把田和季节算进去，正是因为环境是成员而不是背景。

三者合起来，构成了一个人在做事时「与之不可分割」的全部。而这个结构，在哲学上有一个已经被写得很透的版本。

第三十一章　海德格尔的「在世存在」

海德格尔在《存在与时间》第一部分做的事，用最粗的话说，是拒绝一个从笛卡尔以来的默认：先有一个主体，然后这个主体面对一个世界。他的替代方案是「在世存在」（In-der-Welt-sein）——此在（Dasein，他对人的存在方式的称呼）从一开始就在世界之中，「在世」不是此在的一个属性，而是它的存在方式本身。没有一个先于世界的此在，然后再进入世界。

这已经和本书的第一句话——人与器具不可分割、且**没有前件**——完全对齐。海德格尔比延展心智那一派走得更远：克拉克说记事本是心智的一部分，仍然是「心智先在，记事本加入」；海德格尔说的是，根本没有一个不在世界里的心智可以先在。

接下来他分析「世界」由什么构成，用的例子恰恰是工具。他把日常打交道的东西叫做**用具**（Zeug），并区分了两种存在方式：**上手**（Zuhandenheit）和**现成在手**（Vorhandenheit）。锤子在使用时是上手的——木匠不看锤子、不想锤子，锤子在他手里「退隐」，他的注意力全在钉子和木板上；锤子只有在断了、太重了、找不到了的时候，才变成现成在手的——一个被看着、被想着、被作为对象考察的东西。海德格尔的判断是：上手是用具的本来状态，现成在手是派生的；传统哲学从「主体面对客体」出发，恰恰是从派生状态出发，把本来状态漏掉了。

用具从不单独出现。锤子指向钉子，钉子指向木板，木板指向要做的桌子，桌子指向要在它上面吃饭的人。海德格尔把这叫**指引整体**（Verweisungsganzheit）：每一件用具都在一张「为了……」（um-zu）的网里获得意义，网的尽头是此在自己的「为何之故」（Worumwillen）——为了住得好、为了活下去。**用具整体**（Zeugganzheit）先于任何一件用具：不是先有锤子再有工作间，而是先有工作间，锤子才作为锤子出现。

材料和环境在他那里也有位置。材料是「工件」（Werkstoff）——皮革、木头，它们在用具的指引里出现，而且带着「自然」：木头指向森林，皮革指向牲畜。海德格尔说过一句常被引的话，大意是森林是林场、山是采石场、河是水力——自然首先不是被观察的对象，而是在用具指引里被遇到的材料。环境则是**周围世界**（Umwelt）——此在日常操劳于其中的那个世界，有远近、有方向、有上手的位置。

于是本书的三分，在海德格尔那里各有对应：工具＝用具（上手），材料＝工件（带自然指引），环境＝周围世界（指引整体的空间形态）。

第三十二章　五处一致

本书与海德格尔的一致，不止于结构对应，而是在几个关键判断上完全相同。

没有先于世界的主体。这是最根本的一条，前面已说。评估「人」而扣掉器具，在海德格尔那里是从现成在手出发去理解上手——方向反了。

用具在使用中退隐。好的工具是看不见的。这解释了「沉默律」为什么能维持两万年：工具端在上手状态下不出现，人端自然独占署名。不是工具不重要，而是它的重要性恰恰以「不被注意」的方式实现。

整体先于个体。先有厨房，刀才是刀；先有学习的情境，AI 才是学习工具。本书说「席」是单位、人和 AI 是席上的两端，在海德格尔的语言里就是：席是用具整体，人和 AI 在这个整体的指引里才成其为学生和工具。

断裂揭示。锤子断了，木匠才看见锤子；这是海德格尔分析里最有力的一步——存在的结构在正常运转时是隐藏的，只在失效时显露。本书的「退单」就是有意制造的断裂：让人端把 AI 端从上手状态拉回现成在手，看着它、考察它、否决它。退单深度，就是人端能把用具整体的哪一层从退隐中拉出来的能力。海德格尔没有把断裂做成一种评估，但他给了评估一个本体论的位置——**只有在断裂处，人端才显出来**。

周围世界有远近。海德格尔说此在的空间不是几何空间，是「去远」——把东西带到手边。AI 把整个知识库带到了手边，学生—AI 单位的周围世界的远近结构被压平了：以前要去图书馆、要问老师才能触及的东西，现在都在手边。这就是前文讲电报那一站时说的「单位的边界脱离身体」，在海德格尔的语言里是周围世界的去远达到了极限。

第三十三章　四处分歧

一致之外，本书在四处走出了海德格尔的框架，而且四处都是 AI 逼出来的。

第一，用具不判断。 海德格尔的用具是彻底沉默的——锤子在指引网里有位置，但它不对钉子发表意见。整个「上手／现成在手」的分析，预设了用具端没有判断位；判断、领会、操心，全在此在这一端。AI 是第一个自己会对指引网发表意见的用具：它会说这颗钉子不该钉在这里。海德格尔的框架没有为这种用具留位置——一个会判断的东西，在他那里只能是另一个此在，而 AI 显然不是此在，它没有「为何之故」，不为自己的存在操心。本书的处理是：不把 AI 升为主体，也不把它降为沉默的用具，而是承认单位内部出现了**第二个判断位**，并用「分点」来标记两个判断位的分界。这是海德格尔的结构里没有的一个量。

第二，指引链的尽头可能不在人端。 海德格尔的「为了……」链条，最终都回到此在的为何之故——一切用具最终为了此在。AI 打破了这一点：人端的否决记录成了 AI 端的训练数据，「为了学生学会」这条链的中途分出一条「为了模型改进」的支链，而这条支链不回到任何此在。海德格尔晚年在《技术的追问》里其实预感到了这一步：他说现代技术的本质是「集置」（Gestell），把一切——最后包括人自己——摆置为「持存物」（Bestand），即随时可调用的储备。用本书的语言说，集置就是**人端被单位当作材料**：学生的否决记录成了平台的工件。海德格尔把这看作技术的命运，本书把它看作一个可测、可防的量——漂移——并主张席籍归人端，正是为了不让人端沦为持存。

第三，海德格尔没有评估。 《存在与时间》分析此在的存在结构，但此在不被打分。「常人」（das Man）、「平均日常状态」这些概念描述的是此在如何在日常中失去自己，而不是如何被制度评估。本书的问题恰恰是评估：署名、责任、录取、晋升——这些制度性的判断要落在哪里。海德格尔给了本体论，没有给地址簿。「席」是给他的用具整体安一个可以被制度登记的地址。这一步他本人大概不会走，因为他对制度性的「常人」评价很低；但悬空判断的焦虑正是发生在制度里，不给地址，焦虑就不落地。

第四，回写的速度。 海德格尔的此在是「被抛」的——生在一个已经有用具整体的世界里，然后在其中领会自己。但他的用具整体是缓慢的、历史性的，一代人的用具大致就是那一套。本书讲的单位是每周都在重划的：AI 端每月换型号，人端每次否决都在改变下一次的分点。海德格尔的「被抛」是一次性的，本书的漂移是连续的。这不是他错了，是他分析的对象没有这种速度。

第三十四章　共在：AI 是他人还是用具

海德格尔的世界里除了用具还有他人。此在从来是**共在**（Mitsein）——世界里的其他此在不是作为对象被遇到，而是作为同样操劳着的人被遇到；木匠的工作间里有徒弟，皮革上有制革匠的手，桌子是为了某些人吃饭。他人不在指引网里当用具，他人是网的另一头。

AI 把这个二分撑开了一条缝。它不是用具——用具不判断；它也不是共在的他人——它不为自己的存在操心，没有为何之故，没有向死而在。它是海德格尔的本体论里没有位置的第三种东西：**会判断的用具**，或者说，**不操心的判断者**。

教育界的两派争论，本体上正是在这条缝的两边各站一边。禁 AI 的一派把它当他人——一个替学生做作业的人，所以是作弊；拥抱的一派把它当用具——一支更好的笔，所以无妨。两派都错，因为它两样都不是。本书用「席」处理这条缝：不给 AI 一个存在论的身份，而是给它一个**位置**——席上的一端，占一个判断位，不占为何之故。分点划的就是这个位置的边界：分点以上，判断回到有为何之故的那一端；分点以下，判断在席内部，不归任何一个操心者。这是一个回避了「AI 是什么」而只回答「AI 在哪」的办法，回避是有意的——「是什么」这一问，海德格尔的框架答不了，本书也不打算答。

第三十五章　闭卷考试：把上手强行变成现成在手

海德格尔的「上手／现成在手」区分，可以直接用来诊断考试制度。

日常的学习是上手的：学生查书、问人、用工具，用具整体在退隐中运转，注意力在题上。考试是一次人为的断裂：把一切用具收走，只留一个人和一张纸，逼着他在用具缺席的状态下显出自己。闭卷考试的本体论含义是——**强行把此在从在世状态里剥出来，看它单独还剩什么**。这个动作在海德格尔那里恰恰是派生的、不本真的：它评的是一个从来不曾单独存在的东西。

以往的考试制度能这么做而不出大问题，是因为收走的用具是沉默的：收走书本，学生还记得书里的内容；收走计算器，学生还会算。用具退隐之后留在人端的东西，足够支撑一次评估。AI 破坏了这个条件：收走 AI 之后留在人端的东西，和 AI 在场时人端显出的东西，不是同一个量的多少，而是两个不同的单位。闭卷考的是旧单位，旧单位在真实的学习里已经不运转了。

海德格尔的框架在这里给出的不是「开卷」——开卷只是把用具留在手边继续退隐，仍然评不出人端。它给出的是**另一种断裂**：不是收走用具，而是让用具出错，看人端能不能把它从退隐里拉出来。锤子断了木匠才看见锤子；AI 错了，学生才显出来。退单考就是这种断裂的制度化：把一份有错的 AI 稿交给学生，评他能把错误从哪一层拉到现成在手。这不是对海德格尔的应用，而是把他分析的一个本体论事实——人端只在断裂处显出——反过来做成了一种考法。

第三十六章　周围世界与制度：学校作为环境

最后一处需要把海德格尔的「周围世界」推进一步。他的 Umwelt 是操劳的空间：工作间、田野、街道，有远近、有方向。学校显然是学生的周围世界——但学校的周围世界里有一样东西是海德格尔的工作间里没有的：**制度**。学籍、课时、考试、职称，这些不是用具，也不是他人，它们是环境的一种硬化形态——把用具整体的某个历史状态冻结成规则。

这解释了教育的悬空为什么比其他行业深。在木匠的工作间里，用具整体变了，工作间跟着变，没有什么东西规定锤子必须是旧的样子。在学校里，用具整体变了——学生—AI 单位已经在运转——但环境的制度层冻结在旧单位上：学籍登记人，考试考人，课时计人。环境成了单位里最慢的一端，慢到与另外两端脱节。前文说教育被「双重锁死」，用海德格尔的语言就是：**周围世界的制度层与用具整体脱节，此在被夹在两个不同时代的世界之间**。焦虑是这种夹住的体感。

海德格尔会把这叫做「常人」的统治——制度是常人的沉积。但他没有给出办法，因为他关心的是此在如何从常人里挣脱出来成为本真的自己，而不是如何改制度。本书的办法很不海德格尔：改地址簿。席籍、席位读数、教席退单率，是把环境的制度层重新对齐到用具整体上。这是一种工程动作，不是存在论动作；但没有这个动作，存在论看清了的东西落不到任何人身上。

第三十七章　材料的地位：海德格尔留下的一块空

还有一处不是分歧，而是海德格尔没有充分展开、本书需要补的：**材料的抗性**。

海德格尔的工件是在指引里被遇到的——皮革指向牲畜、木头指向森林——但他没有多讲材料自己会反抗。木头有纹理，逆着切会裂；鱼有骨，切错了就碎。材料的抗性是单位里第三个「判断位」的雏形——它不像 AI 那样用语言判断，但它用「不成」来否决。英戈尔德讲的木匠、西蒙东讲的技术物在成形中的「内在协调」，都在补这一块。

对 AI 时代的教育，材料是什么？是题目、是文本、是要解决的那个问题。它有抗性：一道问错了的题，无论 AI 端和人端怎么努力都答不好。前文退单考的第四层——「这道题该不该这么问」——就是材料的抗性在单位里显出来的地方。海德格尔的框架能解释为什么人端只在断裂处显出，但要解释断裂为什么发生在题层而不是事实层，需要把材料的抗性当作单位的正式成员。这是本书把「器具」写成三样而不是一样的理由。

第三十八章　比较的结论

海德格尔给本书的，是三件最硬的东西：没有先于世界的主体；用具整体先于任何一件用具；人端只在断裂处显出。这三条使「学席」不是一个管理学的方便设定，而是有本体论根据的单位——席就是一个可登记的用具整体。

本书在海德格尔之外加的，也是三件：用具端出现第二个判断位（分点）；指引链可能不回到人端（漂移与集置）；单位需要一个制度性的地址（席籍与三读数）。这三件都不是理论上的推进，而是 AI 这个会作声、不留残余、以周为单位回写的用具逼出来的。

一句话概括两者的关系：**海德格尔说清了为什么评「人」在本体上是错的；本书要回答的是，在一个用具会说话的世界里，该评什么。**

卷四　悬空判断的通式

本卷之前的三卷把地基打好：人与器具不可分割且没有前件，工具重划单位，AI是第一个会作声、不留残余的器具。本卷把「悬空判断」写成通式，并把它放回历史与制度里——每个工具一次悬空，教育悬空最深，退守无地。

本卷共七章，其后的三卷是悬空判断的三处显影，各自原为独立论文，在本书里改姓并接上共同的术语（见本卷末「术语对照」）：卷五写学生端——学生—AI 为什么拒绝被旧账本记成一个单位；卷六写教授端——教授仍被要求签字，资格却失去载体；卷七把三重悬空（资格、判断、责任）合成一个理论，给出同源闭合原则与节点责任。三卷合起来证明的是同一件事：悬空不是教育的局部故障，是单位换位之后所有以「人」为地址的制度共有的时差。卷八之后给答案。

第三十九章　悬空之律：每个工具一次悬空

把上面的观察合起来，可以写成一条律：**每个重大工具诞生，都开出一段悬空期——单位已换、评估地址未改。** 这条律可以用史料证。

工具	旧评估地址	悬空期	旧地址的锁死程度	残余功能（着陆面）	出口
文字	背诵	数百年（从《斐德罗》到经院论辩）	高（教会、科举）	诠释、论辩	迁址
印刷	抄写与藏书	约百年	中（大学、行会）	编辑、批判阅读	迁址
摄影	画得像	约五十年	低（沙龙）	看法、风格	迁址，写实留作技法课
计算器	心算笔算	约二十年	高（考试制度）	建模、推理	分场隔离
汽车	马术	两次世界大战	高（军制）	无	隔离，骑术进入体育
人工智能	写作、解题、讲授	进行中	极高（考试制度＋职称制度）	趋零	只剩分点

以上各站的细节，见卷二。

从表里能读出律的两个分量。

其一，**悬空期的长度与旧地址的锁死程度同向。** 文字、计算器、汽车的悬空最长，都是因为旧地址登记在国家级制度上——教会、科举、考试、军制。摄影的悬空最短，因为沙龙的权威没有国家撑腰。

其二，**出口由残余功能决定。** 工具端还不会做的那一块，就是判断着陆的地面。有残余，评估迁址到残余上，新职业诞生——「摄影师」「程序员」都是旧地址作废后新单位拿到的名字。没有残余，旧单位整个进入保留地——马拉松不许开车，国际象棋分人类组，骑术进奥运会。保留地里的评估继续有效，代价是与生产脱钩，变成体育。

第四十章　AI 的特例：退守无地

以前每次悬空，都靠一个「人独有的残余功能」着陆。相机拿走「像」，绘画退守「看法」；计算器拿走算术，数学退守「推理」；机器拿走体力，教育退守「思维」。着陆面总是工具还不会做的那一块。

AI 把这条惯用退路堵死了。它没有一块明确不会的地面：写作、解题、讲授、推理、批评、风格——残余功能不断被抽走，「工具还不会的那一块」不再是一块固定地面，而是一条正在收缩的边。前几次悬空是「迁址」，这一次无处可迁。

这就是本书对「AI 之下教育为何悬空最深」的判断：不是因为 AI 更强，而是因为**退守无地**。当残余功能趋零时，着陆面只能从「功能」迁到「单位内部」——不再问「人还会做什么而 AI 不会」，而是问「在人—AI 这个单位内部，人在哪一层还能退单」。分点，成为唯一的着陆面。

教育在这道题上又比其他行业多一重困难：它的旧单位被双重锁死。考试制度锁住学生端，职称制度锁住老师端，两端的地址簿都是国家级的，换不动。别的行业悬空会被市场自行纠正——企业很快就只看一个人在人—AI 单位里的归己份额；教育不会，它会一直悬着，把焦虑积累到制度本身松动为止。所以教育的悬空判断不是过渡期现象，而是**制度性停留**：一因锁死，二因无地。

第四十一章　悬空判断的通式

到这里可以把「悬空判断」写成通式，并把它放回教育。

悬空判断不是判断与对象的矛盾，而是**判断与对象的时差**。矛盾可以调解，时差只能补。单位已经换了，评估制度还在用换位前的地址簿。判断射出去，没有着陆面。

通式：**凡评估对象仍写作「人」、而发生已在「人—AI 单位」上的地方，判断即悬空；悬空的程度等于该行业单位换位的深度，减去评估地址更新的深度**。

悬空之后必然接着发生三件事，在教育里都已经在发生。

被评者表演旧单位。判断落不到新单位上，被评者只好装成旧单位来接这个判断——学生把 AI 写的稿改出「自己的口吻」，老师声明「本课未使用 AI」。这不是人在作弊，而是评估制度逼着新单位假扮旧单位。

分数失去区分度。旧量表测旧单位，新单位在旧量表上趋同——满分与满分之间没有信息。这是内卷的本体形状：不是人太多，是量表悬空。

责任无处落。出了错，追到学生，学生说是 AI 写的；追到 AI，AI 不是主体；追到「学生—AI」，这个单位没有法律地址、没有学籍、没有署名资格。

三件事的共同根源只有一个：**学习与教学已经在「学生—AI」「老师—AI」上发生，而考试和评估仍在旧频道上，评的是「学生」「老师」**。

第四十二章　悬空判断的四种形态：评估、责任、资格与许可

通式给出的是一个结构：判断投向的单位已不在原地址。但这个结构在制度里以四种形态出现，四者常被混为一谈，混淆之后会给出错误的对策。本章把四者分开。

第一形态：评估悬空。 被评的能力已在新单位上显露，评估仍投向旧单位。表现是分数失去区分度——所有人都满分，满分之间没有信息。对策的方向是换考法（退单考、可重走）。教育的日常形态。

第二形态：责任悬空。 后果已由新单位造成，追责仍投向旧单位。表现是追到人，人说是 AI 做的；追到 AI，AI 不是主体。对策的方向是翻译规则与落点（卷七）。法律与医疗的日常形态。

第三形态：资格悬空。 资格仍被要求由旧单位承担，而证明资格的记录已经分叉。表现是签字的人拿不出证明自己判断的记录（卷六）。这一形态最隐蔽，因为它不表现为失败——教授照常签字，系统照常运转，只是没有人能在争议时说清那一判断是谁作的。

第四形态：许可悬空。 准入许可发给旧单位，实际执业在新单位上。表现是执照上写着一个人的名字，而这个人从不在没有系统的情况下执业。飞行执照与医师执照已经部分处理了它（型别等级、再认证），教育尚未开始。

四者的关系不是并列的四类，是同一道时差在制度链条上的四段：**能力→后果→证明→准入**。前一段的悬空会向后传导：评估悬空使分数失去信息，责任悬空使后果无处落，资格悬空使证明拿不出来，许可悬空使准入失去依据。所以只修一段无效——只换考法而不改资格记录，新考法的结果仍进不了任何容器；只改资格而不改准入，证明有了却不被承认。

四种形态也给出一个诊断顺序：从后往前修。**许可先动，准入共同体以险定真值；资格随之改写；评估最后跟上**（卷五第三趟的判断）。这个顺序与教育界的本能相反——教育界总是先改考题，而考题是链条最前端、锁死最重、外部压力最弱的一段。

第四十三章　悬空与四个近邻的分界：滞后、脱耦、失效、空转

「制度跟不上技术」是一句人人会说的话。本书要说的不是这句话。本章把悬空判断与四个最容易被当成同义词的概念分开，每一处给一个判决性差异。

与文化滞后（Ogburn）的分界。 文化滞后说的是物质文化变了、非物质文化跟不上，时间一到自然补齐。悬空判断不预设补齐：当残余功能趋零时，旧地址没有新的功能可迁（卷四第二章）。判决性差异：滞后预测制度最终会追上并恢复原有功能；本书预测的是功能迁不过去，只能迁到单位内部的量上，或进保留地。

与技术—制度脱耦的分界。 脱耦说的是两个系统各自演化、耦合松开。悬空判断说的是耦合仍然紧密——判断照发、分数照给、证书照签——只是投错了地址。判决性差异：脱耦预测制度停止作用于技术领域；本书预测制度高速运转，且越运转越稳（空转）。

与测量失效的分界。 测量失效说的是指标不再测得准目标构念。悬空判断说的是被测的那个对象已经换了。判决性差异：失效可以通过更换指标修复，换指标后信度效度恢复；本书预测新指标进入旧容器后会被重新代理化——因为问题不在指标，在容器的元数（卷五第四趟）。

与空转的分界。 这一处最近，须写清：空转是悬空判断的制度运行形态，不是另一个概念。悬空说的是判断没有着陆面（本体侧），空转说的是没有着陆面的判断为何照发不误（制度侧）。二者的关系是：悬空是条件，空转是后果；悬空可以在没有分配功能的地方发生（兴趣班里也有悬空判断），空转只在有分配功能的地方发生。

四条分界合起来是一句话：**本书说的不是制度慢了，是制度对准了一个不在那里的东西，而且因为对得准、发得快，所以没有人发现它不在那里。**

第四十四章　悬空的读数：怎样判断一个行业正在悬空

通式给了定义，本章给一张可以拿去用的诊断表。六问，每问一个可查的事实，不问感受、不问态度。

第一问：这个行业的产出，现在由谁显出？ 若典型产出无法在不提及某种工具的情况下描述，单位已换。查法：看近一年的作业指引、工单模板、执业规程里有没有出现工具名称。

第二问：评估对象的名字改了吗？ 查法：考核表、成绩单、资格证书上写的是一个人的名字、还是一个人加一份配置。若只有名字，评估地址未改。

第三问：两者之间隔了多久？ 单位换位的时间减去地址更新的时间，就是悬空的深度。半年算浅，五年算深，二十年算制度性停留。

第四问：读数还有区分度吗？ 查法：看分数分布。若高分段拥挤、区分度逐年下降、而机构在增加分数层级而不是换分数，空转已经开始。

第五问：旧地址锁死在哪一级？ 国家考试、执业许可、职称，是国家级锁死，悬空会长期停留；行业协会、企业内部、学校自定，是弱锁死，会较快迁址。

第六问：工具端还剩下什么不会做？ 若能明确指出一块稳定的残余功能，这个行业还能用迁址的老办法着陆；若指不出，只能迁到单位内部——分点与退单深度。

六问的用法：前三问判断是否悬空，第四问判断是否已进入空转，第五问估计会悬多久，第六问决定出路是迁址还是换算法。教育在六问上的答案是：已换、未改、五年以上、区分度正在丧失、国家级锁死、指不出稳定残余——六项全中，所以本书从教育写起。

其他行业的读法留给读者，卷九给了十个行业的席与分点作为参照。需要提醒的只有一条：**六问里没有一问问「AI 强不强」**。悬空的深度与模型能力无关，与地址簿的更新速度有关。

第四十五章　术语对照：三处显影与本书

卷五至卷七各章原为独立成文，各有自己的承重词。为不损伤原文的论证链，本书保留各章原词，在此给出与本书主词的对照。读者遇到左列的词，按右列理解。

各章原词	本书主词	说明
学生—AI 耦合场	学席	「拒绝被约简为一个单位」＝席无常值、席只有读数
教师—AI、教授—AI—情境	教席	「先在账房一侧合法出生」＝教席先于学席诞生
旧账本、旧账本单位	旧地址簿、旧单位	分配功能与测量功能分离，是旧地址簿不退场的机制
来源栏、来源读数 S	席籍耦合栏	由人当场读出，不由学生自报、不由机器事后猜
中间账、关系档案	席籍	记席不记人；四读数并入席籍三读数
路径读数	环节归属	理解／构思／生成／核验／表达各环节谁做
责任回收率	退单深度、有效否决	只改对答案不计，须说明来源与后果；教席
RR、独立裁判预测率		一侧的可结算形态是 Q*＝命中且未被吸收的否决（卷八）
代理坍缩	（保留）	路径同价·成本偏移：假读数的机制名
悬空资格、悬空判断、悬空责任	悬空判断（通式）的三重显影	资格＝谁有权坐席，判断＝席上谁退单，责任＝退单后谁签收
首个不可回补点、不可回补前沿	分点的时间形态	分点在流程上的位置：过了它人端无法再退
同源闭合	席籍随人走＋否决记录归人端	谁参与发生，谁进入记录，谁承担后果——三者同源
险定真值、职业准入共同体	席的第四读数：险	席位读数之外的外部结算
分定真值	显露物读数	旧量表上的分数
空转、可辩护性	悬空判断的制度运行态	判断无着陆面而分配照旧，靠程序完备、门外人无法定价、持证者互为人质维持
可重走	退单考的现场形态	不可预处理的新材料上再生成一次
外部停兑	席的第四读数：险	门外人先改写资格凭证
分配容器、容器元数、切名	（保留）	卷五末五章：单位由容器决定，是「拒绝约简」的真正原因
席位配置、型别等级	席配置	卷八「席的登记规格」

卷五　学生端的悬空：学生—AI为何拒绝被记成一个单位

本卷写学生端，是全书最重的一卷。它由同一条题在 2026 年 9 月 7 日一天之内跑出的五趟稿子改姓合成，五趟不是并列，是接力——每一趟把上一趟的答案降为问题：

趟	原稿	问的是	承重判断	留下的问题
一	《主体间能力的新单位难题》	为什么旧账本记不了学生—AI	不是换了单位，是新单位拒绝被记成一个单位；账本缺来源栏	拒绝的原因在哪
二	《代理坍缩与空值结算》	失真的分数为什么不退场	代理坍缩（路径同价·成本偏移）＋空值结算（校准可断、可通约性不断）；三条历史序列已跑	不退场的支撑物是什么
二之一	《定义独立律：教育审计为何不可能》	失真人人知道，为什么查不出来	不是鉴证者不独立，是被鉴证的对象由发行者定义；三权（定义／计量／发证）同在一手，程序独立可移植、定义独立不可；ABET 真跑：定义外置、计量留内	查不出的账本为什么还转
三	《空转》	单位死了机器为什么照转	指向死＋检查死，分配功能靠可辩护性——程序完备、门外人无法定价、持证者互为人质；可重走	切名的一刀落在哪
四	《容器元数律》	记账单位由什么决定	不由生产、不由测量，由分配容器的元数决定；容器单数，切名就发生；型别等级是先例	名字之下怎么登记
五	卷八「席的登记规格」	单数容器怎么登记人—器具单位	席名＋席配置＋按配置分开的读数＋席范围	—

五趟合起来，失真的分数在制度里占着三个位置，各由一条机制守着：**不退场**（空值结算——结算功能与测量功能维持条件不同）、**查不出**（定义独立律——三权同在发行者，鉴证只能证明旧账本被正确执行）、**照转**（空转——可辩护性：程序完备、门外人无法定价、持证者互为人质）。三条机制分别对应席籍的三处规定：席籍随人走（拆结算）、席籍明示三权（拆定义）、席籍由人当场读出而非事后填表（拆可辩护性）。

读法：赶时间的读者读第四趟（容器元数律）与第二趟的「空值结算」「三条历史序列」两章，够了。五趟各自的最近邻表与证伪条款保留原貌，供追溯；全书统一的划界表见卷十。各趟原文内部引用的「第 X 节」指该趟自身的节次，本卷每章首次出现处已标明。

本卷与「悬空判断」的关系须明写：悬空判断说的是判断没有着陆面（本体侧），本卷五趟说的是没有着陆面的判断为什么照发不误、发到哪里、被谁收下（制度侧）。第三趟把这叫「空转」并与「悬空」对举，它反对的那个「悬空」（机器停摆等待新单位）不是本书的悬空判断；空转是悬空判断在制度里的运行形态，不是它的替代。

术语对照见卷四末表。本卷各章保留原稿承重词（耦合场、旧账本、来源栏、中间账、责任回收率），读者按对照表换算：耦合场即席（席无常值），中间账与关系

档案即席籍，来源栏即席籍耦合栏，责任回收率即退单率，可重走即退单考的现场形态。

第四十六章　从一场消失的测量说起

一个反常现象

在一所普通高校的写作课上，任课教师布置了一份文献综述。课后，学生提交的文本质量稳定到令人不安：结构清晰，引用规范，观点圆熟。但教师在随即进行的随堂追问中发现，部分学生无法复述文中一个核心论证。值得注意的是：作业文本毫无抄袭痕迹，AI 检测软件未报警，评分标准也全部满足。按照既有教学测量体系，这份作业没有触发任何问题。然而，教师——或者任何认真面对这件事的人——都清楚：这个分数是空转的。

这个场景并非某一个教师的个人遭遇。它指向一个结构性反常：产出质量的各项读数正常，产出的「来源」却已经发生了质变。当基础教育之外的几乎所有文字与解题任务都可以被大规模生成模型在低成本下完成时，作业、论文、考试、课堂问答等传统测量形式所记录的对象，已经不再必然指向那个被记分的学生本人。令人不安之处在于：该系统并不「失灵」。它继续以极高效率运转，只是它记下来的数字，再也无法分解出「人」的差异。

也就是说，反常不是「考试无效了」，而是「考试依然在考核，但它不知道自己在考核谁」。

缺口所在

对这一反常的解释，目前大致可归入三条彼此重叠又互相竞争的脉络。第一条从教育技术或教育评估出发，关心如何升级检测工具、重新设计题目，或者引入形成性评价与过程性评价（如 Swiecki et al., 2022；Dawson, 2021；Bretag, 2019）。这一脉络已经精确地看到了「传统终结性评价难以反映真实过程」，也提出了增加任务多样性、提问式评估、创造式作业等有力方案。然而，它的前提仍然是：「存在一个稳定的可测个体，只是检测条件需要更新。」它没有面对一个更根本的问题：当学生与 AI 之间形成任务性的、随情境变化的协作关系时，「可测个体」是否还作为稳定的本体单位存在。

第二条脉络从伦理与学术诚信出发，追问何谓「真实产出」，并发展出原创性、透明度、贡献声明等概念工具（如 Ellis et al., 2023；Bearman et al., 2023；Perkins, 2023；Cotton et al.,2023；Sun Hoelscher, 2023；Bozkurt et al., 2023；Zhai et al., 2021；Kasneci et al., 2023；Choi et al., 2023；Susnjak McIntosh, 2024）。这一脉络抓住了「贡献归属」与「真实性」问题，并已经建议在写作与考核中引入「来源说明」。但它通常把问题定位在「学生个体如何规范地声明使用」、或「评价标准如何重新定义」，仍未跳出「独立个体为最小责任单位」的默认框架。这个框架在 AI 深度参与知识生产之后，是否还能继续成立，是它很少触及的更深一层问题。

第三条脉络从教育哲学与后人类教育讨论出发，提出主体、能动性与知识生产已经分布化，甚至呼吁以「人机共生」「分布式认知」等概念取代独立学习者（如 Bearman Ajjawi, 2023；Markauskaite et al., 2022；Selwyn, 2024；An Oliver, 2023；Chiu, 2024；Yan et al.,2023）。这条脉络在概念上最为激进，也最贴近现象本身：它承认「学生-AI」已经成为实际运作的混合体。但它在关键处停下：一旦面对考试、学分、文凭、奖学金、录取线这些必须逐人结算的制度现实，这类「分布式主体」概念便无法提供一个可直接落账的最小单位。也就是说，它很好地批判了旧的「单位」，却没能说明新的「账本」怎么记。

三者交汇处，就是本书要切入的缺口。三条脉络都承认一个共同事实：旧测量装置已经无法透明地读出个体差异。但三条脉络也都共享一个未经审视的预设——仍然存在一个「应当被测量、可以更精细地被测量」的单位。真正失效的，或许不是测量技术、不是诚信伦理、不是主体哲学，而是「单位」本身。

承重命题

本书要提出并维护的承重命题是：

AI 时代教育的根本困难，不是「学生」这一本体单位改变为「学生-AI」，而是「学生-AI」作为一个耦合场拒绝被约简为一个可测、可记、可结算的单位，而教育系统仍在用为「独立个体」设计的旧账本继续记账。由此产生的假读数，不是某一环节的技术漏洞，而是单位与账本之间的结构性脱节。

这个判断可以被更尖锐地写为：不是「学生变成了新单位」，而是「新单位拒绝作为一个单位被记」。换单位是教育史上惯常的动作；拒绝被换成一个可记的单位，才是真正的改变难题。

这一承重命题的经验表现，可以精确地表述为：在「学生-AI」成为实际产出单位之后，教育系统没有、也难以产生一个稳定的分解读数，把「人」的差异从 AI 贡献中分离出来；于是账本只能继续用旧单位的格子去记录新单位的产出。记下来的数字，不是谎言，而是「失去了对象的读数」。

第四十七章　账本、单位与来源栏

理论出发点：为何在本体单位进入制度之前，先看账本

本书不从教育评估理论出发，也不从学术诚信出发。因为这两个视角都预设了「单位问题已经解决」。本书从一个更为底层的制度事实出发：现代学校教育的一切度量与责任装置，都建立在「独立个体」这个最小记账单位之上。课程设计问「学生应知什么」，考题问「学生自己会不会」，成绩单问「学生本人做了什么」。这些发问形式，不是教育理念的偏好，而是制度记账所必需的本体前提。没有单位，就没有可供计量的项。

因此，本书站在「账本—单位—路径」三者的交互关系上。关键单位不是「学生」，而是「账本上的学生」。当 AI 大面积进入产出过程，账本上的独立个体与实际生产单位之间发生脱节。理解这个脱节，不能从「检测工具升级」出发，而应从「记账方式的容量」出发。账本的问题，不是它坏了，而是它仍在记旧单位，但生产已经在新单位里完成。

名义定义

学生-AI：在某一具体任务中，由学生与人工智能系统共同参与、以双方不可从成品中稳定分解的方式协作完成教育产出的实际生产单位。

旧账本：以「学生作为独立生产主体」为最小单位，以分数、绩点、通过与否为主要记录符号，并以这些记录作为资格分配的通用制度装置。

本体单位：一种制度、测量与责任装置所要求其记录对象必须具备的不可再分的最小行为主体。这三个定义不涉及「充分」「真正」「恰当」等程度词。它们是名义定义，是在本书论证中固定下来的使用方式。

操作性定义与读数

「学生-AI」在实际经验层面表现为：一份被记分产出在学生与 AI 之间具有可观察的协作事实，但制度无法以稳定、低成本、且不依赖学生自报的读数把这「两者各自的贡献份额」分解出来。

承载这一现象的读数不止一个，本书只取一个最直接的：来源栏读数 S（source column）。其定义为：对任一份提交产出，按照生成时段与生成人脑参与的可见程度，将其归入以下值域：1＝独立完成（未借助 AI 或仅作格式与检索工具）；2＝人机协作（学生承担可观察的关键生成过程，AI 承担辅助、组织或润色）；3＝AI 代做（关键生成过程不可在学生处观察到，AI 承载理解、构思和形成文本）。这一读数属于三档类别测量，不允许在两档之间再插一档。单次赋值方法为：教师在收到产出的同时，依据「可观察课堂生成事实」和「学生对该产出的当场复述与修改能力」将其

归入三档之一；不依靠学生自报；不依靠事后文本检测。若读不出，则该项不进入比较，而单列为「不可归因」。

必须立即指出：这个读数在现在绝大多数校园账本中根本不存在。这正是本书讨论的起点。在 S 值缺失的情况下，账本只能用「结果读数」（分数、完成度、标准满足度）来代理一切判断。这个代理，就是假读数的温床。

第四十八章　非「换单位」，而是「单位拒绝入账」

判断的正式建立

核心命题可以正式表述如下：

AI 时代教育的本体论难题，不是 X =「学生这个单位被学生-AI 取代」，也不是 Y_1「评价工具无法检测 AI 参与」，更不是 Y_2「学生主体性正在被技术消解」。而是 Z =「学生-AI 这一实际生产单位拒绝被约简为一个可记、可测、可结算的单位；而为旧单位设计的账本仍在继续记账。」

这个判断的关键不是「单位变了」，而是「改变后的东西无法作为单位进入制度」。私塾变成班级时，班级可以编号；班级变成学分制时，学分可以累计。这些历史性的换单位动作之所以能完成，是因为新单位可以被稳定地切开、命名、计数。而「学生-AI」的要害在于：它是任务敏感的耦合场，每次组合的配比都不同；今天 AI 只提示一个词，明天 AI 代写九成；同一学生在不同任务里呈现为无数个不同的「学生-AI」。其共同特点不是「人的成分更少」，而是「人的成分不可从读数中分离」。

因此，判断的成立不依赖 AI 是否真的「取代了人」，而依赖：在旧账本读数保持唯一的条件下，是否存在一个稳定的、非依赖自报的、可跨任务比较的来源分解读数。只要这个读数不能成立，「新单位拒绝入账」就成立。

成立条件与不成立条件

Z 成立的判定条件有四条：

第一，存在一个制度性账本，其最小单位是独立个体「学生」，并且分数、绩点、通过与否等读数只挂在个体名下。

第二，教育现场中存在至少一类高频产出任务，在这些任务中 AI 与学生的贡献无法从成品中被读出，也无法通过既有流程被稳定分解。

第三，教育系统在没有该分解读数的前提下继续产出并采用这些分数、对其赋予分配功能。

第四，单位之间的不同导致同一名义分数背后的生产过程不可比较：即两个相同的分数，可能来自两种完全不同的生产单位。

只要上述四条同时存在，「单位—账本脱节」就成立。

它不成立的条件也同样是可观测的：如果未来出现一套普遍采用、非依赖自报、低成本且稳定的过程追踪技术，能够对任一份产出分解出「人的贡献等级」与「AI 贡献等级」，并使官方账本按此登记，那么本书关于「不可约简」的判断即被推翻。那时剩下的就不是单位不可入账，而是旧账本更新速度问题。

两句复合测试

本命题是否只是当代两个相邻作者已有判断的组合？可以用一个严格的「两句复合测试」来回答。取最接近的两位常被引用的占位者：Bearman 与 Ellis。Bearman 在近年写作中清楚说过：「知识与学习现在分布在人机系统之中，个体测量的假设正在瓦解。」Ellis 则明确说过：「所有评估都必须记录生成性 AI 的使用方式，没有使用声明的作品不能进入判断。」如果我们把这两句话合起来：第一句话说单位已经分布化，第二句话说必须有人声明工具使用。那么组合似乎已经包含了本书的意思。但本书的增量依然存在，而且只落在一个位置上：分布主体与声明主体，在制度上被认为是同一个人。本书不否认这两句，但切出了两句合起来仍然盖不住的那条缝：当实际生产单位不是人、而是任务级的耦合场时，由「人」来声明贡献，本身又回到了旧账单的旧单位里。也就是说，她们二位都没有问：贡献声明这个动作，应该由谁来承担？ 本书的答案不是一个新的考核方法，而是：真正的问题不是声明如何使用，而是没有一个与生产单位同构的责任单位可以承担声明义务。这个狭缝，就是本书与最近邻之间的分离线。

第四十九章　最近邻盘点：八位占位者与分离线

本节必须诚实交代：本书提出的「单位—账本脱节」这一判断，在当前可访问的文献库中并未经过系统性的站外检索。以下最近邻人员中的每一个，都可以被视为本书命题在思路与对象上的近亲。没有列表，就没有分离；没有分离，本书就只是综述。故而此处不避嫌疑，逐条盘出他们与本书之间最锋利的那一丝差异。

其一，Bearman, M.（邻接命题：知识与学习分布在人机系统中，个体测量假设瓦解。）领域：高等教育与评估。分离线：她认为「测量必须重新考虑对象」，本书认为：「在「学生」这个承载测量的旧

单位之下，根本没有第二个可依赖的新单位来承接新测量。」判决性反例：某学生用 AI 完成全部研究设计，并在课堂即兴答辩中能复述但不理解。按她的判断：这是分布式学习的一个实例，应发展新评价单位。按本书判断：这恰好证明「一个不可再分的新单位尚未成形」，新评价没有挂靠点。

其二，Ellis, C.（邻接命题：必须记录生成式 AI 的使用方式，否则评估无效。）领域：学术诚信与透明评估。分离线：她主张增加「使用声明」，本书主张「在使用声明之前，声明者本身已经不是旧单位，也不是稳定的新单位」。判决性反例：一份文档由六次不同 AI 交互拼接完成，学生只能大致回忆其中一段。按她的判断：需要更细的声明指导。按本书判断：声明可行性本身被「过程不可召回」所破坏，这是单位级失败，不是规范缺失。

其三，Dawson, P.（邻接命题：检测工具失效，评估设计必须转向动态任务。）领域：教育评估学。分离线：他认为任务本身可被重新设计以恢复真实性，本书认为任务改变会被旧土壤翻译成「更难」，不会重置生产单位。判决性反例：他可能给予高分的「高认知开放性任务」，在本书看来只要仍以标准评分收口，就会在两年内被题库与代训模板接管。

其四，Bretag, T.（邻接命题：学术诚信研究应从「抓作弊」转向「能力培养」。）领域：学术诚信研究。分离线：她试图修复责任者；本书关心责任者在生产层面的单位是否已失效。判决性反例：一个学生训练 AI 学习自己的写作风格，再让它生成初稿自己少量修改。布雷格会归入「存在不透明协作，需要培养诚信」一类；本书则判为「U 类不可归因」，并由此认为不是教育不够，是账本无栏。

其五，Perkins, M.（邻接命题：必须将 AI 使用纳入评估矩阵，并关注评估的可持续性。）领域：评估改革与检测分析。分离线：他调整评估矩阵的变量维度；本书调整的不是变量维度，而是记录变量的前提单位。判决性反例：作业完全满足含透明声明的新评估矩阵，但声明中「人机份额」无法复现。按他判：新矩阵成立。按本书判：读数看似成立，实为类别贴纸，因为它没有单次取值程序。

其六，Cotton, D. R. E.（邻接命题：院校和教师应重新设计考核任务，防止内容被 AI 绕过。）领域：高教教学管理。分离线：他假定了任务到达个人后由个人承担；本书不假定承担者与人同形。判决性反例：小组项目把所有写作交给 AI，但每人准备

好答辩。按他判：可用个别答辩补测个人能力；按本书判：答辩能测「复述」，仍不能测「昨日生成人的份额」。

其七，Bozkurt, A.（邻接命题：生成式 AI 在开放教育中模糊了原创与生成边界，需要新伦理框架。）领域：开放教育与远程学习。分离线：他把新边界视为伦理问题；本书把边界问题归因于单位不再是一个受伦理约束的实体。判决性反例：学生自称「我们和 AI 共同完成」，任何伦理条款都不是说谎，却也无法给出可登记贡献。

其八，Selwyn, N.（邻接命题：AI 推动教育治理与数据权力重组，而非简单「革命」。）领域：教育社会学。分离线：他观察治理与制度如何重构；本书观察重构发生之前「账本与单位」如何互相锁死。判决性反例：某国强制 AI 应用进入官方教材，产生大量新平台，但成绩单形式不变。按他判：证明技术被治理收编；按本书判：进一步证明新单位只能在生产侧增殖，不能进入结算侧。

每一位占位者都真实占据了一个有价值的位置。但列表摆出来，也恰恰可以看清：没有任何一位直接把「账本上没有来源栏」和「单位不可约简」连接为同一命题。本书的价值不在声称他人全错，而在证明：他人已经踩到了这个洞的边缘，却没有一个把脚落进去。

第五十章　来源栏的判别式、指标与自否决条款

判别式

本书不满足于理论区分，而要求一个可操作的判别式。最核心的判据如下：在不依赖受试者自报、且不依赖新评价形式的前提下，对同一批产出试加一栏 S 值（1＝独立完成；2＝人机协作；3＝AI 代做），并观察这一栏是否会影响分数结论。若加上 S 栏后，原本「全部合格」的产出之中仍有大量进入 S＝3 类，同时现有账本没有提供任何后续处理栏，即可判定旧账本正在以「无来源」读数为真值。反之，若试记账后 S＝3 类占比低，或系统立即因此调整评分与资格处理，「单位拒绝入账」判断即被削弱。

这一判据在「来源栏」正式建立之前，可以先安置在四个场景中观察其理论指向。在本科作业场景中，若在提交栏强制标注来源，未标注产出应被计为「不可归因」而非「零分」，观察其是否改变成绩分布。在研究生开题环节中，若导师逐段追问生成路径，观察学生的可回答率是否与成品质量同步。在职业资格考试培训中，若某培训机构声称「不需基础、两周通过」，观察其是否在逻辑上只能由外包路径支撑。在入职考核中，若面试要求现场用 AI 处理一份随机材料，观察其成绩能否被先前文凭分预测。四个场景给出的答案不同，才说明「单位问题」不是一维的。

可观测指标

核心指标即为 S 值的三档分布。测量方式为「学生-AI 交互产出实验」：教师选取当堂生成材料、随机当场发布、限时提交、事后对样本逐份归因。比较对象是同一批学生的在家作业 S 值粗估。若在家端 S＝2、S＝3 的比例显著高于当堂端，且差异不因题目难度、学科、学生绩点而消失，则可判定产出「来源构成」本身是影响读数的独立变量。

读数自我否决条款

读数未满足以下三条之一，即不立。

其一，重复归因可靠性。两名独立观察者对同一批产出逐份做 S 赋值，一致性 Kappa 低于 0.7，说明读数不稳，不能作为制度依据。

其二，改名门槛。若 S 栏在经验上可以被「在场时间」或「文本相似度」完全替代，没有新增信息，则它只是已有量的改名，必须放弃。

其三，单次赋值不依赖自报。若某 S 赋值过程必须事后问学生「这里是不是你写的」，或必须读取 AI 检测软件，则该读数已退化为另一种他报与机器猜测，不可采用。

这三条否决条款，是本书所有进一步讨论的基础。若 S 栏无法立住，本书关于「新单位拒绝入账」的命题虽然仍然有概念意义，但经验层面将失去一个扳手。

第五十一章　竞争解释的检验：修复方案、声明方案与分布式方案

本节把几位占位者的现有解释逐个与本书的判断对撞。以他们最有利的版本入局，而不是以稻草人版本入局。换句话说，若要反驳他们，必须先允许他们以最强形式提出反驳。

教育评估的修复方案

他们对本书最有力的反驳是：如果你说来源不可分解，那说明你还没试过足够好的过程性评估。一个好设计可以让学生在短时间内多次暴露自己的决策，令教师读到个体差异。例如 Paul 的反驳大致可以是：「当堂限时修改、中途打乱任务、口头追问与自评，这些操作足以清楚区分学生会什么不会什么。」

对这一最强形式的回应是：这些做法全部假设「起点在课堂上」。但教师逐一看不过来的是「起点的来源」。学生只要提前让 AI 做过「理解性准备」——不做终稿，而做大纲、关键思路、问答对、修改策略——课堂上的即时修改就成了一系列记忆输出。其表现平滑，且不引发警报。因此，过程性评估优化的是「人在不在场上」，而不是「人的生成是否起于此刻」。在最坏情况下，它把原先在家外包的「终稿」环节逼成了在家外包的「理解」环节，再用课堂表演将这一搬迁隐藏起来。

学术诚信的声明方案

这条反驳也极强：若每个学生都必须在提交时声明「哪些部分用了 AI、用了几次、用在哪一步」，那么来源不就清楚了吗？教师应惩罚不申报者，奖励申报透明者。只要诚信制度严格，来源栏可以靠自报建立。

对这条反驳，本书不回答「学生不可信」。因为很多时候他们确实会诚实申报。本书回答的是另一件事：一旦 AI 的「贡献」已经成为产出的组成部分，要声明的就不是「用没用」，而是「在生成的哪一步、以什么方式、替换了原来应当属于人自己的哪一个生成行为」。这个层次的声明，超出大多数使用者可以复现的记忆。举个机械例子：一名学生让 AI 生成十个选题，自己选了最顺眼的一个；用同一个 AI 生成六百字引言，自己删去二十字；再用 AI 根据他的课堂笔记重排三点结论，自己改掉一个小标题。之后，他被要求按四步声明贡献。要写清楚哪段文本、哪条判断、哪次关键生成是由哪一方承载，实际上已经不可能被单次回忆完成。声明由此变成了故事，不是读数。不是学生撒谎，而是他们自己也只有一个模糊的故事。

分布式认知的新单位方案

最危险的一条反驳来自这一脉。他们的最强形态是：你所说的「账本无单位可记」恰恰说明需要建立新账本，而不必纠结旧账本。既然单位已经分布，就让账本也分布；不测个人，测系统。你可以建立一个「学生-AI 作品集」，记录全过程的每一

轮交互；大学录取看这个作品集，教师评价看这个系统的产出质量。这样，旧单位问题被绕过了。

这个方案之所以危险，是因为它的确能绕开「个人不可分解」这个难题。但它的致命伤在结算处：谁来为作品集里的失败负责？给出终稿的是系统，考试分数挂谁？奖学金、排名、毕业证、责任追究都是单数制度。当一个学生拿着自己的「系统作品集」去申请一个只收个人名字的资格，新账本没有办法签发主体资格。更尖锐的是：一个学生能带着同一个作品集去申请十个学校，作品集里的 AI 也在场。系统不是人，不能作为申请人。分布式账本一旦接触到分配，就会在「谁被录取」这个最简单问题上卡死。这不是理论不完整，而是分配制度本身的容器是单数。他们造出了一个无法填写姓名的账本。

三种竞争解释经过检验后，本书判断「单位—账本脱节」依然存活。值得注意的是，这一轮检验没有一条结果支持对手，也没有一条完全排除对手。作为概念分析，这是可以接受的边界。但更重要的，是它们各自的反例都指向同一个位置：任何试图通过「改进旧流程内的人或方法」来解围的路径，最终都会撞上那个「找不到与生产单位同构的责任单位」的死结。

第五十二章　四格辨别：来源可见性 × 外部分发硬度

如果「学生-AI」才是真实生产单位，而旧账本上的学生是记账单位，那么仅凭「人机占比高低」一个维度，还无法区分不同教育境况。本节尝试建立一个二维辨别格，目的不是命名所有情况，而是提供一种在制度上可以判别的类型学。

第一轴：账本读数的「来源可见性」

这一轴取自本书核心命题：账本是否以「独立个体」为单位读取产出。来源可见，指当前制度拥有可用且不依赖自报的路径信息；来源不可见，即现有评分只读结果，不问路径。这是本书的第一轴，也是账本是否与生产在单位层脱节的主要表现。

第二轴候选，最先被推出的可能有三种：「任务难度」「学科类型」「学生动机」。但任务难度是「来源不可见」的压迫因素，不是结构独立轴；学科类型影响「人机贡献如何分布」，是表现而不是结构；「学生动机」则更偏个体内部状态，不适合观察。三个候选都偏成因，而查成因不能当轴。排除。另一个更有吸引力的候选是「账外分发的硬度」，即账本之外有没有一个足够强的实际装置来决定资格。它不是第一轴的成因，而是一个独立的外部强度变量，结构上不依附于「来源栏是否存在」。故第二轴取：资格结算是否已有外部分发的现实强度。可以分为「外硬」与「外泡」两档。「外硬」指职业准入门槛、行业考试、现场责任等真实账已经部分进入；「外泡」指社会仍主要依照墙内名校、分数、证书默认分配机会。

这两个轴合起来，不是对「学生-AI」本体的刻画，而是对一种教育场中的可裁决位置的刻画。它回答的不是「这学生有多真」，而是「旧账本和自己的外部分发条件处于什么相对关系」。

四格辨别

A 格：来源不可见 × 外硬。这是「代理巢穴里的真账」。它集中出现在那些外部职业认证已经开始运作、校内的课程作业却仍按个人终稿打分、且不计过程来源的技术型院系。在这个格中，学生完全有动机绕开课堂生成，又完全明白最终职业技能会被现实检验。于是，课堂账本恰恰是学生最愿意作假的地方，而职业实践是他们最不敢不作为的地方。这一格的独有预测是：「校内高分与职业能力不相关，甚至负相关。」因为校内账本越顺滑，学生越未经历场检测；而职业端越硬，他们越不得不自己补课。

B 格：来源不可见 × 外泡。这是「虚假繁荣最深的文凭地」。它遍及大量声誉老、排名高、但毕业去向与专业能力已脱钩的普通文理专业。这里没有外部硬账来敲打，账本来源又不可见，于是课堂生产退化得最彻底。学生、教师、管理者都只说「完成度」。这一格独有预测：「教育阅读量无显著增加，但作业平均分逐年微涨。」原因不是学生更聪明，而是可外包的路径越来越顺。

C 格：来源可见 × 外硬。这是「产训最像样的一批」。它常见于实训基地、小班制师徒制课程、真患者与真案件介入的临床培养等。这里教师知道产出是怎么来的，外部也真有考核等着。学生学习动机混合，但路径不能被彻底伪装。这一格的出现不依赖「检测 AI」，而依赖「人进入了可一眼看穿过程的现场」。独有预测：「AI 辅助程度越高，资优生的成长幅度比资浅生更大」，因为高手用 AI 不是替代生成，是放大自己，在现场才看得出来。

D 格：来源可见 × 外泡。这是「自娱自乐的觉醒区」。例如一些低利害选修课、读书会、项目制学习实验课。学生被要求现场生成，教师在场；但外部完全不在意。这里 AI 进入后很可能出现一种反向运动：因为不需要抢分数，人更愿意保存自己的理解习惯，甚至对 AI 感到累。这一格独有预测：「AI 使用频率升高，但学生主动选择低配使用 AI 的比例也升高。」其动机不是道德，而是「反正不被结算，何必为难自己」。这一格恰好说明：来源可见若没有分配后果，只有部分的真实性恢复。

四格并无可相互替代的两格。若把 C 与 D 比较，会发现「外硬」是它们在所有四个格中唯一改变行为方向的变量；若把 A 与 B 比较，会发现「来源栏」的缺失越严重，学生的隐藏路径走得越舒服。由此得出第一轴（来源可见性）与第二轴（外硬）不是同一变量：第二轴不改变第一轴是否存在，却决定第一轴的后果是否有价。

第五十三章　讨论：单位问题先于评估理论

理论意涵

最直接的含义：教育评估不是一个「测量准确性」问题，也不是「诚信文化」问题，而是一个「单位能否作为账本前提」的问题。它要求我们在讨论「如何考」之前，先问「在考什么单位」。如果「单位」与账本不相容，那么一切在旧单位内部优化指标的努力，都只是给账本增加抗失效涂层。新单位入账，不可能作为教育评估理论的内部修正出现，只能作为外部制度把另一种结算方式推进教育现场的结果。

实践意涵

实践上，最轻的一步不是设计新课，而是把「来源栏」放回账本。这不是「让教师改评价标准」，而是让教育机构本应承担的一种可读性回到可观察的栏位。它一旦落地，会随之改变课堂所有相关行为：学生问的不再是「这样可以吗」，而是「这一步该算我做的还是 AI 做的」。教师也不再只是打终稿分，而是必须在生成现场做一次来源归因。更重要的是，哪怕三档粗分非常粗糙，也可以先把「无分解」从暗账变成明账。很多假读数并不是「故意造假」，而是「根本没有那一栏」。一栏出，假在哪就清晰。清晰了，才有之后的制度选择。

适用边界与反向约束

本判断并非绝对。至少有两个条件会大幅削弱它。其一，任务本身对「来源」不敏感，如纯肌肉训练、体育技能、器乐练习、手动技术。在这个意义上，不是所有教育领域都进入「学生-AI」的单位困境。其二，外部不分配资格，比如纯粹兴趣班、公益工作坊、成人进修。在这类场所，假读数再普遍，也没有严重后果。旧账本能退，新单位也不急需。这两条反向约束明确地削掉本书「教育系统全部陷入单位危机」的言外之意，保留「凡以分数分配人的位置的教育场才真正进入危机」的严格立场。

反噬

按本书建议去走，最省事的做法是给每份作业加一个「来源栏」，然后继续用旧分数。但这一旦变成例行公事，学生就会把 S 档也当成新的评分元素：拼凑出一种符合人情味的「人机协作比例」，以便拿高分。来源栏本身也会被制成品化。当最原始的自然记录也变成可表演物，来源栏就不再是路径读数，而成了一种新型结果读数。恰恰会毁掉它原本想接住的那个东西。这是本书建议必须提防的自反风险：用「公开透明」的道德气息，给假读数换上更高明的包装。本书能给出的防御不多，只有一条：这一栏若要活，必须由「人」当场读出，不能由机器事后产出，也不能由学生事后自述。流程读出，不是表格读出。

第五十四章　结论：先把「来源构成」一栏放进账本

第一问 问：旧测量装置在哪个环节失去记账能力？答案是：在「提交与归卷」之间。产出递交后，旧账本只记结果、不记路径，更没有来源栏。缺失不在评分标准里，而在单位进入账本之前少了一道「来源赋值」。

第二问 问：为何新单位无法被稳定识别？因为它是任务级变形的耦合场：各次生成的配比都不同，且贡献无法用非自主报告方式分解。它不提供跨任务的共同分母。这是「不可约简回一个单位」，不是「尚待发现的更复杂单位」。

第三问 问：旧账本为何仍能继续运转？因为账本的分配功能不依赖真实性，记账者自己的资格也写在同一本账上，且没有第二本可操作的账可用。三方互锁，旧账本在冲击越大时越像唯一可用的救生索。

第四问 问：新单位最可能从哪里被制度承认？从职业准入共同体开始。当「险」字账本在资格侧开口，外部现实就能以比招聘更硬的语气宣布：只有新单位算数。这个过程不发生在考试大纲内部，而发生在考试围墙的外面。

教育系统现在可以做的、比建立一个「新单位理论」实研究启示只有一条，也是最不浪漫的一条：际得多的事，是先把一栏「来源构成」放进正在运行的账本里。这个动作微小、粗糙，却能把「假读数」从一个哲学判断变成制度里可追踪、可争吵、可修订的过程。其最终目的不是「把旧账本修好」，而是让教育不再误以为自己还在测量那个独立的人。

本书没有完成任何经验裁决，只是交出判据与证伪条件。将来的工作应回到教室、职业现场与资格机构，看谁最先、以何种方式、在什么压力下，把「这一栏」从讨论里搬进真账。那时，关于「学生-AI」的争论才不只是一个教育哲学话题，而是一场有账可查的现实革命。

第五十五章　五位缺席者与分离线：延展心智、Goodhart、文凭主义、行动者网络、AI 使用量表

上一节的三条脉络是教育内部的解释。要把本书的判断放到位，还必须对五个更深层的理论位置做划界。它们不在教育文献的常见引用名单上，却各自占着本书论证空间的一角。若不逐一说明本书与它们的分离线，“代理坍缩”和”耦合场”就有被读成旧概念改名的危险。

第一位：分布式认知与延展心智。 Hutchins（1995）对舰船导航的民族志证明，认知任务在人、工具与表征之间分布，导航的”知道”不属于任何一个人。Clark 与 Chalmers（1998）进一步主张，只要外部资源在功能上与内部记忆等价，心智的边界就应延展到工具上。这两家为”学生—AI 是耦合场”提供了本体论的许可：思考确实不必属于皮肤之内。但它们的问题是描述性的—认知如何分布；本书的问题是制度性的—分布之后谁被记账、谁被归责。Hutchins 的导航团队不需要给每个水手打一个”独立完成”的分数，而学校必须。分离线在此：分布式认知说明耦合可以发生，本书说明耦合发生之后旧账本为何仍以个体记账、以及这种记账如何反过来塑造耦合的形态。延展心智没有”责任”这一变量，本书的核心读数正是责任落点。

第二位：Goodhart 定律与 Campbell 定律。 Goodhart（1975）指出一旦统计规律被用作控制目标，规律便告失效；Campbell（1976）指出社会指标被用于决策的程度越高，它越容易被扭曲，且越容易扭曲它本应监测的社会过程；Strathern（1997）把它压成一句—指标一旦成为目标就不再是好指标。Koretz（2008）在教育测量中给出了经验版本：高利害考试的分数在几年内可以上涨一个标准差，而同一群学生在低利害审计测试上的成绩几乎不动。这一传统离本书最近，也是”代理坍缩”最容易被还原进去的地方。分离线在于预测的方向：Goodhart—Campbell 预测被博弈的指标失去信息、被识破、被替换；本书预测失去信息的指标反而更牢固地留在结算位上。Koretz 自己的数据已经暗示了这一点—分数通胀被系统记录二十年，高利害考试的分配权重没有下降。Goodhart 解释了指标为何失真，解释不了失真的指标为何不退场。本书前文要命名的”空值结算”正是补这个洞。

第三位：文凭主义与信号理论。 Collins（1979）主张文凭的功能不是证明能力而是分配地位；Labaree（1997）指出美国教育的”文凭竞争”目标使学生学会的是获取标记而非学习本身；Spence（1973）的信号模型则说明，当获得信号的成本对所有人同步下降时，信号的区分力下降而信号水平上升。这三家解释了分数为何具有分配功能，但它们把”能力”当作先于信号而存在的私人属性，信号只是能力的不完美代理。本书的分离线是：在学生—AI 耦合场中，能力不是先于任务与工具而存在的固定量，它在路径与土壤中生成，因此信号所代理的那个对象本身已经不稳定。信号理论处理”信号偏离能力”，本书处理”被信号的对象不再是一个对象”。

第四位：行动者网络理论。 Latour（2005）取消了人与非人在行动网络中的先验区分，行动被理解为网络的效应而非主体的属性。这为”学生—AI”提供了拒绝人类中心记账的理由。但行动者网络理论的方法论承诺是”跟随行动者”、拒绝预设结

构，而本书恰恰要说明一个结构—旧账本—如何在网络已经改变之后继续以旧单位运转，以及这种运转如何被具体的制度利益维持。分离线是：行动者网络描述网络的展开，本书描述记账结构对网络展开的截断。

第五位：人工智能使用分级量表。Perkins、Furze、Roe 与 MacVaugh（2024）提出的 AI Assessment Scale 把人工智能在作业中的许可程度分为从”禁止”到”完全使用”的五级，供教师在任务层面声明规则。这是当前教育实践中最直接的制度回应，它承认使用程度必须被写进评价。分离线是：分级量表把”程度”当作变量，本书的变量是”环节”与”责任”—同为第三级”允许协作”，模型润色一段已独立写成的文字与模型先行组织论点，是两种完全不同的路径结构。程度变量无法替代路径变量，量表因此仍在旧账本内部工作：它规定人工智能可以用多少，不记录关键判断由谁作出。

五条分离线合起来，可以把本书的位置说清楚：本书借用分布式认知的本体论许可，借用 Goodhart—Campbell 对指标失真的机制，借用文凭主义对分数分配功能的揭示，借用行动者网络对人机对称的方法态度，借用分级量表对制度回应的经验形态；本书自己要交出的是它们合起来都没有的一条判断—失真的分数为何不退场，以及旧账本在读数失真后如何反过来塑造耦合场本身。

第五十六章　发生学重构：显露、路径与土壤

本书站在发生学与关系本体论的交叉位置上。发生学不问一个对象如何被发现，而问它为何以现在的方式发生：哪些条件使它出现，哪些路径被保留，哪些可能性被排除，形成之后的对象又如何反过来改变原有条件。关系本体论则拒绝把对象理解为先于关系而独立存在的实体，对象的边界、性质与功能都在关系网络中形成。

采用这一传统而不是工具主义、个体主义或制度主义，各有一个理由。工具主义能够说明人工智能提高效率，说明不了学生单位为何失效。个体主义能够保住学生责任，却把人机协作视为个体之外的偶然因素。制度主义能够解释分数的分配功能，却容易把制度当作既定背景，而不是由路径与土壤共同生成、又反过来塑造路径的结构。

为了把这个交叉位置写成可操作的分析层，本书区分三个层次。**显露层**是某个对象在制度表面以何种形式出现：作业、考试、文凭显露为学生个体的产出。**路径层**是对象在一连串选择、试探、保留与排除中如何形成：学生每一次使用人工智能都在试探、避险、保留与迭代中走出一条路径。**土壤层**是制度、资源、语言、关系、时间与心理状态如何共同构成对象的存在条件：评分规则、时间压力、工具可得性、同伴比较、职业竞争与责任结构共同塑造了学生—AI。

三层之间不是相加关系而是互相生成。显露不是路径的简单结果，因为土壤会改变显露方式；路径也不是在既定环境中运行，因为显露出来的结果会反过来改变环境；土壤不是外部背景，因为对象与路径会持续改写土壤。教育评价的假读数正是在这样的情形下发生的：显露层仍然稳定地显露为学生分数，路径层已经改变，土壤层却继续支撑旧的显露。教育系统看到的是显露层，实际生产发生在路径与土壤的纠缠里，而旧账本把显露层还原成学生分数。

本书因此不问”人工智能是否帮助学生学习”，而问：何种差异路径在何种土壤中被保留下来，为什么最终显露为学生个体差异，以及什么外部事件能迫使制度承认这种显露已经失真。

前文给出了三层，这一节把学生—AI 在路径层与土壤层的实际发生序列写出来。只有把序列写具体，才能说明为什么耦合场”不是一个单位”不是修辞，而是可以逐步观察的事实。

路径层的发生从试探开始。学生第一次把课程材料交给模型，通常不是为了代写，而是为了看模型能做什么：解释一个概念、给一个提纲、改一句话。试探的结果是一次局部成功，成本极低而读数不变—教师没有察觉，分数没有变化。第二步是识别风险。学生评估这条路径会不会被发现、被发现的代价是什么、同学是否也在走。第三步是保留。当风险评估为低而收益为稳定时，这条路径被固定为默认动作：先让模型理解，再由自己复述；先让模型生成，再由自己修改。第四步是迭代。学生在不同任务之间调整配比，把模型推向更前端的环节—从表达退到构思，从构思退到理解—每一次后退都是一次新的试探与保留。路径层的关键特征是：每一步都是局部理性的，没有一步是”决定作弊”，然而序列的终点是理解与判断环节被整体让渡。

路径的每一步都需要土壤的支撑，土壤不是背景而是共同作者。时间压力使试探的动机成立：任务在截止日期前必须完成，模型是最快的完成方式。工具零成本使保留的门槛消失：在合同代写时代需要付费的路径，现在不需要。同伴比较使风险评估的结果偏向使用：当同学都在用而分数不变，不用者承担的是相对成本而不是绝对收益。生产力叙事为使用提供正当性：学校、企业与媒体都在说”会用人工智能是未来的核心技能”，学生沿这条路径走时不需要对自己解释。升学竞争与资格分配则把整个序列锁定：只要终局分数决定机会，任何降低取得终局分数成本的路径都会被保留。

序列在群体层面有一个转折点。当使用者比例越过某个阈值，不使用者的相对位置开始系统性下滑，此时”使用人工智能”从个体策略变成群体默认，土壤完成了一次改写：原来支撑路径的条件（低风险、低成本）现在被路径本身加固（大家都在用、不用吃亏）。这就是路径反过来改写土壤的时刻，也是耦合场”拒绝被约简”的发生学根据—同一个学生的组合方式随任务、随时段、随同伴而变，因为它每一次都是在被上一次改写过的土壤上重新发生的。

显露层在整个序列中始终不动：作业照交，分数照打，学分照记。序列的所有变化都在显露层之下发生，显露层把它们全部压成”学生的分数”。旧账本不是没有看见，而是它的记账单位只允许它看见一种东西。

第五十七章　承重命题：耦合场拒绝被约简，旧账本继续以个体记账

本书的承重判断是：

> 学生—AI 不是学生的技术增强，也不是一个已经稳定形成的新单位，而是一个不能被稳定约简的耦合场；教育危机不是新单位取代旧单位，而是旧账本把不可约的耦合场继续记作学生个体差异。

用最短的形式写：学生—AI 既不是”学生加工具”，也不是”可直接替代学生的固定新主体”，而是”在任务、路径与土壤中不断变形的耦合场”。

这一判断包含两个相互关联的否定。第一，学生—AI 不是一个已经完成的新单位，它是差异不断生成的场。第二，考题改变不是单位改变的充分条件；只要旧账本仍把产出压成学生个人分数，新的题目就会成为代理路径的材料。教育改革若只修改题型，就会在内容层面制造变化，在单位层面维持不变。

命题在以下条件下成立：人工智能参与了产出形成而非仅提供机械输入；学生与人工智能之间的分工随任务变化；评价装置只读终局结果而不读来源与路径；不同路径在结果上获得相近判定；低成本代理路径因同价而持续扩张。若人工智能仅承担拼写检查、格式转换等不改变理解与判断的环节，或者评价装置能够稳定、低成本地记录并区分各环节贡献，命题的强度就下降。

命题不成立的条件也要先写出来。若一种新评价装置可以在不依赖学生自报的情况下，稳定记录理解、构思、生成与核验的责任分配，并且这一记录能够进入正式资格决策，学生—AI 就可能被分解为可比较的多维单位。此时它仍然是关系场，但它不再”拒绝被记”，而是被转换成一套新的制度对象—前文讨论的”学席”就是这样一个制度对象。若真实执业环境能够直接检验产出责任，且资格不再主要依赖学校分数，旧账本即使继续存在，也不再承担全部分配功能，本书关于旧账本持续统治的判断随之削弱。

最后要过一道复合测试：本书的核心判断能否由两句现成文献无损重述？一句可以说”工具正在改变学习活动”（分布式认知），另一句可以说”指标一旦成为目标就会失真”（Goodhart—Campbell）。两句合在一起仍推不出：学生—AI 为何不是一个稳定单位，为什么更复杂的考题会被翻译成代理路径，以及为什么旧账本在失去测量含义之后仍保留分配权。本书的增量不在于把两句放在一起，而在于命名把它们连起来的两个机制。

第五十八章　代理坍缩：定义、判别式与指数

代理坍缩是评价装置把不同生产路径的产出压缩为同一判定，从而使路径差异无法进入分配结果的过程。它不以学生是否主观作弊为定义。只要独立完成、部分协作、整体外包等路径在评价装置中获得相同或相近读数，并且成本更低的路径因此被保留，代理坍缩就发生了。

这个定义的要害在”坍缩”二字：它不是说学生找到了更聪明的作弊方式，而是说装置本身把所有新差异重新压成一个可比较的分数。压缩是装置的动作，不是学生的动作。学生只是沿着被压平的地面走最短的路。

本书的判别式是：

> 若同一产出在不记录其生成来源与责任承担的情况下获得正式资格读数，并且至少存在一条更低成本的替代路径能够得到相近读数，则该评价装置发生了代理坍缩。

判别式不含”应当”“合理”等情态词，可直接用于流程、日志与记录。它要求观察三个事实：是否记录来源；是否记录责任承担；不同成本路径是否得到相近结果。

把它操作化为读数：设同一任务中存在两条可观察路径，路径 A 为学生独立完成，路径 B 为学生把理解或生成环节外包给人工智能。令评价结果为 R，生成成本为 C。代理坍缩指数 AC = 1 − |R_A − R_B| / (R_max − R_min) 取值在 [0, 1]，越接近 1，两条路径在评价结果上越难区分。仅有结果相似还不足以构成坍缩，还需记录成本差 ΔC = C_A − C_B，ΔC > 0 表示独立完成成本高于代理路径。在同一任务、相同时限与相同评分标准下分别记录两个版本的结果与过程，若 AC ≥ 0.80 且 ΔC 达到预设的时间或步骤阈值，该事件被标记为”路径同价、成本偏移”。0.80 不是自然常数，是用于提出可复核判据的预设阈值，须经敏感性分析。

判别式在六类场景中的读法如下。课堂写作：学生独立读材料、立论、成文，与学生先让模型生成论点再润色，两者评分相近而评分记录只保留终稿，判据成立。编程作业：学生能运行模型生成的代码却解释不了数据结构与错误处理逻辑，评分只看测试是否通过，判据成立；若评分同时要求学生现场修改一处未预告的错误并记录路径，判据可能不成立。实验报告：模型生成设计与解释、教师只按格式与结论评分，判据成立；学生必须在现场面对设备异常、重新决定变量并承担失败成本，路径差异进入读数，坍缩程度下降。教师批改：模型生成评语、教师仅确认提交、学校把评语视为教师专业判断，“教师—AI”在记账侧先发生。职业准入：资格考试只依据标准化答案发证，执业错误不进入资格维持机制，学校分数具有较高坍缩风险；资格包含持续监督、案例复盘与事故责任，读数便不再只由一次考试承担。

第五十九章　教师—AI：新单位先在账本侧出生

学生—AI 不是唯一的耦合场。同一时期，教师也在批改、命题与反馈中调用模型，形成教师—AI。两个耦合场在制度中得到的待遇完全不同，这个不对称本身就是一条证据。

模型生成评语、教师确认并提交，学校把评语视为教师专业判断—这一序列在多数学校里没有引起任何制度反应。它被记录为效率提升：批改时间缩短，反馈周期变快，教师被”解放”出来做更高价值的工作。没有人要求教师申报模型参与了多少，没有人给评语做来源标记，也没有人问一句评语的责任落在谁身上。教师—AI 在记账侧悄悄合法化了。

学生—AI 得到的是相反的待遇：检测器、申报表、口试、惩戒条例。差别不在于人机耦合的深度—教师让模型写评语与学生让模型写论文，在环节结构上是对称的—差别在于两个耦合场与账本的位置关系。教师—AI 站在账本的记账一侧，它提高结算效率而不威胁分配秩序，账本欢迎它；学生—AI 站在被记账一侧，它动摇分数对学生个体的指称，账本视它为风险。新单位因此有不对称的合法化路径：在记账一侧它是效率，在被记账一侧它才成为危机。

这个不对称对本书的命题有两重意义。第一，它说明旧账本不是被动的测量工具，而是一个有自身利益方向的结构：它选择性地承认那些强化它、拒绝那些动摇它的耦合。第二，它说明”教师—AI 先合法化”会反过来改变学生—AI 的土壤。当学生面对的”教师意见”实际上来自模型，学生对反馈的信任锚点已经移动；当学生知道评语由模型生成而教师不必解释依据，学生让模型生成作业的正当性感受也随之上升。账本一侧的耦合为被记账一侧的耦合提供了先例。

这一节的判别式与前文相同：若模型生成评语、教师仅确认提交，而学校将评语视为教师专业判断，则教师—AI 在记账侧发生了代理坍缩—不同成本的批改路径得到相同的制度读数。若教师保留批改证据、解释评分依据并对错误评语负责，模型可以是工具而不必成为责任主体。这里的证伪条件写在前文。

第六十章　空值结算：失真的分数为何不退场

前文已经指出，Goodhart—Campbell 传统解释了指标为何失真，解释不了失真的指标为何不退场。这一节命名那个被解释空出来的机制。

空值结算是指一个读数在失去其测量含义之后，仍被继续用作分配货币，并且失真本身反而扩大了它的流通量的过程。“空值”指读数与它所指的对象之间的对应已经断开；“结算”指它仍被用来结清资格、名额、机会与地位的分配。

空值结算之所以可能，是因为分数从一开始就有两种功能：测量功能，即分数对某种能力或成就的指称；结算功能，即分数作为不同机构之间可通约的分配语言。两种功能通常被当作一体，测量功能是结算功能的正当性来源。但它们的维持条件不同。测量功能的维持条件是分数与对象之间的校准，这需要不断的外部审计；结算功能的维持条件是分数在机构之间的可通约性，这只需要所有机构继续接受同一种货币。校准可以断，可通约性不会因校准断裂而断。

这就是失真的分数为何不退场：它在机构之间的流通并不依赖它的测量含义，只依赖别的机构继续收它。招生办收学校分数，学校收考试分数，雇主收文凭，每一方都知道分数可能失真，但每一方的替换成本都高于继续接收的成本，因为替换需要另一种所有机构同时接受的货币，而这种货币不存在。分数于是像一种已经脱锚的通货：没有人相信它背后的储备，所有人继续用它结账。

失真反而扩大流通量，这是空值结算与一般”指标失效”的第二个差别。当代理路径使高分变得容易，高分的数量上升，分配需要在更多高分之间做出区分，于是机构对更多分数、更细的分数、更多层次的分数的需求上升，分数的使用密度上升。Spence 的信号通胀模型描述的是信号水平上升而信息含量不变；空值结算描述的是信息含量下降而使用密度上升。两者都在文凭市场上被观察到，前者是均衡的静态描述，后者是失真之后的动态。

空值结算还给出了耦合场为何总被切回一个名字的制度几何。分配容器是单数的：录取录一个人，发证发给一个人，追责追到一个人。耦合场进账时必须通过这些容器，而容器的元数不接受”一个场”，只接受”一个名”。因此单位不是由生产决定的，也不是由测量决定的，而是由分配容器的元数决定的。这一点把”拒绝被约简”的原因从耦合场的任务敏感性挪到了制度几何上：只要结算容器保持单数，任何多元的生产单位在进账时都会被切回单数，测量再精细也改变不了这一刀落在哪里。

空值结算同时解释了前文引入的现象：为什么题目越复杂越容易被代理化。复杂题目在结算功能上并不比简单题目值更多，它换来的仍是同一种货币。只要货币不变，复杂性就只是提高了代理路径的市场价值—题目越深，越值得被拆解、模板化并做成新题库。开放性本身成了模板化服务的市场机会。改变题目改变的是被测的东西，不改变结算的货币，因此不改变单位。

由此可以给出空值结算的三条可观测征象：其一，分数与外部审计读数的偏离持续扩大，而分数在各类分配决策中的权重不降；其二，围绕分数的代理服务市场规模随题目复杂度上升而扩大；其三，机构在承认分数失真之后，选择的应对是增加分数

的种类与层级，而不是替换分数。这三条征象合起来，与 Goodhart 预测的”指标被弃用”方向相反。下面三节用三条已经跑过的历史序列检验它们。

第六十一章　跑过的三条历史序列：合同代写、分数通胀、执照以险为真值

“深度题会被代理化”的判断不需要等生成式人工智能来验证，它在人工智能之前已经跑了十几年。

合同代写（contract cheating）这个术语由 Clarke 与 Lancaster 在 2006 年提出，指学生付费让第三方完成作业后以自己名义提交。它针对的正是过程性评价路径推崇的那类任务—课程论文、项目报告、反思日志、学位论文的章节—而不是标准化选择题。Newton（2018）对1978 年至 2016 年间 71 项样本、六万五千余名学生的自报数据做了系统综述，发现承认过合同代写的学生比例在整个期间平均为3.5%，但 2014 年之后的样本升至15.7%，且趋势显著上升。上升发生在人工智能进入教室之前。

这条序列对本书的意义有三层。第一，代理化的对象恰是评价改革试图用来替代考试的开放性任务。任务越开放、越”真实”，越难在考场内完成，越需要课外时间，因此越可外包。复杂性不但没有阻止代理，反而定义了代理服务的产品线。第二，供给方的响应方式证实了模板化预言：代写平台按学科、字数、学位层级与截止日期定价，把开放题当作可重复的产品规格，而不是不可重复的思想事件。第三，制度的响应方式证实了空值结算：英国2022 年《技能与 16 岁后教育法》把提供代写服务列为刑事犯罪，澳大利亚高等教育质量与标准署自2022 年起阻断代写网站的访问。两国的应对都是在账本外部打击供给，而不是在账本内部改变记账单位—分数继续被收，只是试图让获得分数的替代路径更贵。

这条序列还给出一个对本书不利的读数，须写下来。合同代写有价格门槛，能够用钱换分的学生是少数，因此在人工智能之前，代理坍缩在总体上被价格限制在一个角落里。生成式人工智能把这个价格门槛降到接近零，代理路径从少数人的策略变成多数人的默认。本书的机制没有变，变的是土壤：价格门槛消失之后，“不使用者承担相对成本”这一条件第一次在全体学生中同时成立。

Koretz（2008）汇集的证据是空值结算最直接的经验形态。在美国若干州引入高利害问责考试之后，州考分数在三到四年内上涨了半个到一个标准差；同一时期同一学生群体在低利害的全国审计测试上的成绩几乎不动，或上涨幅度只有前者的一小部分。分数与它所指的能力之间的校准断了，这是 Campbell 定律的教科书案例。

本书要看的是校准断裂之后发生了什么。在这些证据被系统记录、被测量学界反复引用的二十年里，州考分数在学校评级、教师问责、学生升级与毕业决策中的权重没有下降，多数州反而扩大了考试的科目与年级覆盖。校准断了，结算没有断。这与 Goodhart 传统预测的”指标失效后被弃用”方向相反，与本书前文的第一条征象一致。

第三条征象也在同一序列中出现。当高分变得普遍，区分能力下降，若干州的应对不是替换分数，而是增加分数的种类：在”及格”之上增设”精通”与”高级”，

在州考之外增设课程终结考，把单一分数拆成多层分数。分数的信息含量在下降，分数的使用密度在上升。

这条序列的分离价值在于，它把”指标失真”与”指标退场”在时间上拉开了。如果 Goodhart 传统是完整的，两者应当相继发生；实际观察到的是失真持续二十年而退场没有发生。解释这个时间差需要一个 Goodhart 传统里没有的变量，即分数的结算功能与测量功能维持条件的不同。这就是本书命名空值结算的理由：它不是给已知现象换一个名字，它是给一个已知理论预测失败的位置补一个机制。

前两条序列说明旧账本如何在校内维持。第三条序列来自校外，说明资格的真值可以不由一次考试承担。

飞行执照是最清楚的例子。取得执照需要通过笔试与飞行考核，但执照的有效性不由那次考核维持：机长必须在规定周期内完成复训与技能检查，必须满足近期飞行经历要求，任何事故与严重事故征候都进入强制报告系统并可导致执照被暂停或吊销。资格在这里不是一次结算的结果，而是一个持续承担风险的状态；执照的真值来自执业现场持续产生的可观察后果，而不是来自入门考试的分数。

医师资格在近二十年里也向同一方向移动。英国医学总会自 2012 年起实施再认证制度，执业医师每五年须基于年度评估、同行与患者反馈、事故与投诉记录重新确认执业资格；美国各专科委员会的”认证维持”制度同样把继续认证与执业记录、持续学习和考核绑定。这些制度的共同点是把资格的一部分真值从入门考试移到执业记录上—移得不彻底、争议很大，但方向是明确的。

这条序列对本书的意义是：在公共风险高、执业许可与资格紧密相连的领域，资格制度已经找到了一种不依赖学校分数的真值来源。当执业现场以风险、错误后果与责任承担建立另一种真值，学校分数在资格分配中的兑换率就开始下降。这正是本书第三个研究问题的答案所在：迫使旧账本退场的力量不来自校内考题的升级，而来自校外资格以险为真值对分数分配功能的架空。校内考题升级只改变题目表面，只要墙外仍承认分数，墙内土壤便没有自我拆除的动力；只有当学校分数在资格与就业中逐渐无法兑换，而执业风险要求新单位的产出，资格制度才可能先于考题改变。

这条序列的适用边界也须写明。它只在风险后果强、责任不可逆、市场依赖正式许可的行业成立。在风险后果弱、责任可逆或市场不依赖正式资格的行业，雇主可以通过作品集、试做任务或短期试用完成能力识别，资格制度不一定成为最强的外部判决。因此本书不说”职业准入必然先于学校改变”，只说在公共风险高、资格与执业许可紧密相连的领域，职业准入更可能成为外部判决的入口。

三条序列合起来，可以把前文的三条征象逐一对应：合同代写市场对应第二条征象（代理服务市场随题目复杂度扩大），分数通胀对应第一与第三条征象（校准断裂而权重不降、增加分数种类而不替换分数），飞行执照与医师再认证则给出反向的对照—当真值来源移到执业风险上，空值结算的三条征象同时减弱。三条序列都发生在生成式人工智能之前，因此它们检验的是机制而不是工具；人工智能改变的是土壤中的价格门槛，没有改变机制。

第六十二章　既有解释的判决性读数

第二、三节划出了分离线，这一节把八种最容易占据本书论证空间的解释放进同一批场景，看它们与本书各自读出什么。判决性差异的标准只有一条：同一场景，两种解释给出的读数不同，且差异可观察。

学术诚信解释的最强版本认为核心是学生是否遵守使用规则。课程论文场景中，学生公开声明”使用人工智能协助构思”，论文得高分，课堂追问中学生无法重建论证。诚信解释读为”合规使用后仍需加强学习”，本书读为”来源显露不足，终局分数不能继续代表学生独立差异”。前者记录使用是否被允许，后者记录学生是否承担了理解与判断。诚信解释解释不了的是：规则完全允许使用，评价仍无法区分学生是否拥有论证能力。

工具素养解释认为应培养学生提问、核验、修改与整合的能力。医疗案例模拟中，学生用模型生成诊断建议并按格式完成报告，但没有发现模型遗漏的危险病史。工具素养解释按使用效率与报告完整度判为高水平协作，本书预测责任回收率低，学生没有承担关键判断。两个读数分别是工具调用成功率与错误发现—修正率。工具素养解释解释不了的是：学生可以熟练操作模型，却无法在错误后果出现前判断何时不应采纳。

过程性评价解释认为终稿不能代表学习，须把草稿、讨论、答辩与课堂表现纳入评价。课堂写作中，学生带着模型预生成的结构进教室，现场完成改写与口头说明。过程评价记录到当场写作与发言，判为过程可见；本书预测生成环节已在课堂之外完成，现场只发生翻译与复现。两个读数分别是”是否当场产生文本”与”关键判断是否当场产生”。过程评价解释不了的是：过程可以被看见，关键生成已在不可见时段完成。

测量效度解释认为问题在于工具缺乏信度、效度或区分度。同一学生在写作、编程与实验中分别让模型参与表达、代码生成与实验设计，测量解释倾向寻找一个统一的”人工智能使用”变量纳入量表；本书预测不同环节的参与改变责任结构，统一变量无法保持指称稳定。测量解释解释不了的是：同样的使用频率在不同任务中对应完全不同的责任后果。

复杂问题解释认为应转向开放问题、真实项目与跨学科任务。项目式课程中，学生完成城市交通方案，题目开放、材料复杂、没有单一答案，但培训机构与模型可以先生成框架、数据解释与展示脚本，学生按模板执行。复杂问题解释判为高真实性，本书预测题目代理化已经发生，路径成本下降而责任读数没有提高。它解释不了的是：开放性本身可以成为模板化服务的市场机会—前文的合同代写序列已经跑过这条。

资源差距解释认为人工智能造成新的数字鸿沟。若全班都能免费使用同一模型，却同时出现独立理解能力下降、终稿质量上升与现场解释能力下降，资源解释判为差距缩小，本书判为代理坍缩扩大。它解释不了的是：人人获得同一工具之后，评价对象仍可能失去稳定边界。

教师效率解释认为模型减轻批改负担、释放教师专业性。模型生成大部分评语、教师仅确认、学校仍视为教师判断，效率解释记录时间节约，本书记录评分信用锚点已经转移。它解释不了的是：效率提升伴随教师无法说明评分依据，学生面对的”教师意见”实际来自模型。

职业导向解释认为应加强产教融合与职业场景。学校新增人工智能职业项目，企业与准入机构仍以学校考试分数为主要筛选依据，职业导向解释判定课程已接轨，本书判定资格账本没有改变、外部风险未进入结算。它解释不了的是：学校可以复制职业语言，却不承担职业错误的现实后果。

八种解释分别把问题放在规则、工具、过程、测量、题目、资源、教师效率与职业衔接上。本书不否认这些变量重要；本书的分离线是：这些变量都可以在旧账本中被吸收，而本书要解释的是旧账本为何能吸收它们，并在吸收之后继续把不同路径记作同一类学生差异。

第六十三章　中间账：四类读数与三条自我否决条款

如果代理坍缩是装置的动作，改革的第一步就应当改装置，而不是改题目。本书把需要建立的那一层叫**中间账**：位于产出生成与资格结算之间，记录产出来源、生成路径与责任承担的评价层。它不是最终成绩的附加说明，也不是学生自我申报的道德栏；它的功能是把旧账本读不出的差异显露出来，使不同路径在进入分数、学分或资格之前留下可追踪的痕迹。

中间账由四类读数承载。**来源读数**：独立、协作、外包三类类别变量。**路径读数**：理解、构思、生成、核验、表达五个环节各由谁完成，编码为0（学生承担）、1（人机共同承担）、2（人工智能承担），一次任务的路径是一个五维序列，两个版本之间的曼哈顿距离可作路径差异的描述，但不直接当作价值判断。**责任读数**：学生能否在限定时间内解释选择、发现错误并承担修正，采用有序等级。**现实读数**：产出在真实或模拟执业环境中引发的可观察后果，采用错误次数、返工次数、风险事件数或任务完成时间。

在这四类读数之上定义三个可观测指标。来源显露率 SR，即正式评价任务中能被记录为独立、协作或外包的任务比例；一次任务计入分子的条件是同时拥有至少一项过程证据与一项责任证据，只有自报而无过程记录的任务不计入。责任回收率 RR，即学生在模型产出出现错误时能够独立发现、说明并修正的比例；只有说明错误来源、指出错误后果并完成修正才计为一次回收，仅仅改对答案不计。资格外部兑付率，即学校评价结果在真实或模拟职业场景中能预测任务承担与错误后果的比例；它不等于就业率，也不等于雇主满意度，而要求把学校读数与后续执业任务建立可追踪的连接。

多少差异才算数，不能只靠一个阈值。本书建议同时满足三个条件：来源显露率在任务层面达到预设比例；路径向量至少在一个核心环节出现稳定差异；责任回收率与最终成绩之间出现可重复的偏离。若学生成绩高、责任回收率低，且这种偏离在不同任务中持续出现，就说明终局成绩可能已经代理化。

读数必须带自己的否决条款，否则中间账会变成又一套可以被代理的表格。第一，**信度条款**：同一批对象在两周内、难度相近的重测中，来源显露率或责任回收率的组内相关低于0.70，读数判为不稳定，不得用于高风险资格决策；类别编码的双编码一致性低于0.80，须回到编码规则修订。第二，**改名嫌疑条款**：若代理坍缩指数与已有的”作业完成度”“学习投入度”“工具使用频率”在控制任务难度后高度重合（相关系数0.80 以上且增量解释率低于0.05），本书的读数判为换名而非新构念。第三，**单次事件条款**：一次观测若只能凭学生自报推断来源，不能获得过程记录、现场追问或责任修正证据，该事件不得被标记为代理坍缩或独立完成；概率、倾向和程度只能用于群体估计，不能倒灌为单个学生在单次任务中的事实判定。

第六十四章　辨别格：两轴四格

要给学生—AI 做类型学，先要排除三条看起来自然的候选轴。“人工智能参与程度”是耦合场的成因变量，不是能区分耦合结果的独立轴—参与越多并不必然意味着责任越低，某些任务中模型承担检索或模拟，学生仍承担关键判断。“学习结果好坏”是评价结果而不是结构变量，用成绩作轴会把待解释的结果重新放进解释框架。“学校与企业的距离”能解释外部压力，却不能区分内部责任结构，离企业近的项目也可能把任务模板化。

本书确立的两条轴是：**关键判断是否由学生承担**—不是模型参与多少，而是学生能否在关键节点说明选择、发现错误并承担后果；**产出路径是否被制度记录**—理解、构思、生成、核验与表达的来源是否进入评价。两轴相关但不重合：记录可以很完整而学生仍承担不了判断；学生可能在没有记录的情况下确实独立完成而制度无法识别。

四格如下。**路径记录低、判断承担低**是隐蔽代理化：终稿、分数与资格读数存在，但无法知道理解与生成如何发生，学生也无法在追问中承担判断。它独有的预测是：终局成绩上升或维持而现场责任回收率下降，制度无法解释二者的偏离。**路径记录高、判断承担低**是可追踪外包：学校能记录学生何时调用模型、用了哪些输出、改了什么，学生仍说不出为何接受某个判断。它独有的预测是：来源记录越来越细，学生独立解释能力不随记录增加而改善，制度获得了过程数据但没有获得责任主体。**路径记录低、判断承担高**是未显露的独立责任：学生能在现场面对新材料、解释选择并修正错误，制度却没有记录路径，最终仍把能力压成一次终稿。它独有的预测是：现场追问显示责任能力高于书面成绩所能表达的范围，成绩与真实表现出现双向偏离。**路径记录高、判断承担高**是责任型人机协作：模型参与若干环节，制度记录来源与路径，学生在关键节点作出判断并承担后果。它独有的预测是：模型参与程度与学生责任能力不再简单负相关，甚至可能在复杂任务中共同提高，但这种提高依赖错误暴露、解释与修正机制。

两轴同时为高的现实案例可以举公开庭审：律师使用模型检索与整理材料，但必须在庭审中对事实主张、证据关联与法律立场作出公开判断，庭审记录、代理关系与责任承担均可追踪。这不是说所有司法场景都已达到高水平，而是说明”人工智能参与”与”责任承担、路径记录”可以同时存在。四格之间的差异不能被”人工智能使用多寡”替代，第四格才是教育改革希望达到的状态，但它不能靠单纯增加模型使用或单纯增加过程记录自动形成。

第六十五章　学席：耦合场如何被转成制度对象

前文写下了命题不成立的第一个条件：若一种新装置能够稳定记录责任分配并进入正式资格决策，学生—AI 就不再”拒绝被记”，而是被转换成一套新的制度对象。这一节把那个制度对象说清楚，也顺带处理一个可能的自相矛盾。

本书否定了”学生—AI 是一个固定的新主体”，而与本书同日提出的”学席”概念却在给学生—AI 命名并为它设计读数。两者是否冲突？不冲突，前提是分清本体与记账对象。耦合场是本体层的事实：在任务、路径与土壤中不断变形，没有稳定边界。学席是记账层的对象：它是制度在中间账建立之后，为耦合场在某一任务中的一次占位所设的席位。席位有编号、有读数、可比较，但席位不是坐在席位上的那个场。一个学生在编程作业里的学席与他在实验报告里的学席可以有完全不同的路径向量与责任等级，两个学席之间不能通约成”这个学生的能力”，正如两次航班的机长席不能通约成”这个人的驾驶能力”而只能通约成两次记录。

举一个席位读数的例子。某学生在一次编程作业中的学席记为：来源读数”协作”，路径向量 (1, 2, 2, 0, 1)—理解由人机共同承担，构思与生成由模型承担，核验由学生独立完成，表达由人机共同承担；责任读数为三级中的第二级，即学生能解释为何接受模型给出的数据结构，但没有在限定时间内发现模型埋下的一处边界错误；现实读数为该程序在模拟部署中触发一次回退。这一组读数与同一学生在实验报告中的席位读数（来源”独立”，路径向量 (0, 0, 0, 0, 1)，责任第三级，现实读数零事件）不能相加成一个分数，也不能平均成”这个学生的人工智能素养”。两个席位各自记录一次耦合的形态，学席就是这一次形态的记账名。

这样分层之后，本书与学席的关系变成一种分工：本书说明为什么旧账本不能以”学生”记账，学席说明新账本以什么记账。本书给出学席必须满足的三个条件，它们都来自前面各节。第一，学席的读数必须来自中间账的四类读数而不是来自终局分数，否则它只是把”学生”改名为”学席”，落回旧账本。第二，学席必须允许”无法判定”，不能把模型参与程度直接换算成扣分，否则它会把耦合场再次压成一个数字，只是换了一个更现代的数字。第三，学席的真值必须能与校外的执业风险对接，否则它在校内再精细，也逃不出空值结算—分数的失真不会因为分数被改名而停止。

第六十六章　反噬与实践意涵：申报表如何毁掉中间账，先改时间配置不先改题

如果照本书推进，最省事的做法是给所有作业增加一张”AI 使用申报表”，让学生勾选独立、协作或外包，并把它作为新的合规指标。这个做法可能恰好毁掉本书想保住的东西，而且它极可能发生，因为它是旧账本吸收中间账的最低成本方式。

三步就够了。学生会把复杂的生产过程压成对自己有利的类别；学校会把来源栏变成新的纪律工具，勾选外包便降低成绩，不问模型究竟承担了什么；教师会把填表数量当作过程性评价。中间账于是成为旧账本的附属票据，来源栏没有打开路径，只增加了一张表。这正是前文第三条征象的又一次显形：机构在承认失真之后，增加的是分数的种类而不是替换分数。

中间账只有在不把不确定性伪装成精确比例时才有价值。它必须允许”无法判定”，必须记录关键判断与责任承担，而不能把模型参与程度直接换算成扣分。它也不要求课堂成为全程监控空间。教师无法靠注意力覆盖所有生成过程，学生也不应被迫提交无限量的个人数据。中间账应当只记录对责任判断有用的节点：任务从何处开始，关键判断由谁提出，错误如何被发现，最终责任由谁承担。过度记录会制造新的行政负担，也会把教育转化为行为监控。

教育机构若必须在多个改革环节中选第一步，应先改时间配置与中间账，而不是先改题目。理由来自前文：改题目改变的是被测的东西，不改变结算的货币，因此不改变单位；改时间配置改变的是生成发生的位置，它直接作用于路径层。

把需要观察生成与责任承担的练习移入可见时段，可以增加路径读数：不是把所有作业搬进课堂，而是把关键判断的产生移进课堂。一篇论文的写作可以在课外完成，但论点的选择、对反例的处理、对一份新材料的即时回应，可以安排在课内十分钟完成并记录。这样安排的成本远低于全程监控，读数却落在耦合场最关键的位置。

把来源构成、关键判断与错误修正记录在终结性成绩之前，可以避免所有差异直接坍缩为一个数字。这一步的要点不是记录得多，而是记录得对。中间账应当记录四个节点：任务从何处开始，关键判断由谁提出，错误如何被发现，最终责任由谁承担。四个节点之外的过程可以不记。

第二项实践意涵是，学校不应把”是否使用人工智能”作为唯一政策变量。更重要的是使用发生在哪个环节：模型参与表达修订与参与概念选择不同，参与数据清理与参与研究结论不同。规则应从工具名单转向责任链条：不是列出允许与禁止的工具，而是列出哪些判断必须由学生作出并能被追问。

第三项实践意涵是资格制度必须引入现实风险的反馈。职业资格不能只由入场考试一次性决定，而应通过监督实践、案例复盘、错误报告、持续教育与责任追踪形成动态凭证。这不是取消考试，而是取消考试对资格真值的垄断。前文的飞行执照与医师再认证已经给出了可以借的结构。

这三项意涵的顺序不能颠倒。先改题目再改时间配置，新题目会在旧时间配置中被代理；先改中间账再对接执业风险，中间账会在校内自我封闭并被空值结算吸收。

改革必须从时间配置进入路径层，从中间账进入显露层，从执业风险进入土壤层，三层同时动，单位才会动。

第六十七章　边界、两个洞与六条证伪条款

本书的判断有三条边界，每一条都削掉一句原本可能说得过大的话。

第一，不可约性的判断不适用于所有人工智能使用。若模型只承担格式转换、语法校对或机械计算，不改变理解、构思与判断，耦合场对评价单位的冲击有限。本书不能说凡是使用人工智能都使学生单位失效，只能说当模型进入关键差异生成环节时，旧单位面临失配。

第二，职业准入优先改变的判断已在前文收缩：只在公共风险高、资格与执业许可紧密相连的领域，职业准入更可能成为外部判决的入口。

第三，题目代理化的判断受制于题目开放程度与模型能力边界。若任务包含无法预先获得的现场材料、不可复制的身体操作或必须承担的实时后果，代理路径可能无法低成本完成。本书不说”复杂题必然被代理化”，只说只要复杂题仍能被提前拆解、模板化或由模型预生成，复杂性就不能保证单位改变。

此外有两个本书解释不了的洞。第一个洞：在长期、低风险、兴趣驱动的学习中，有些学生主动用模型扩展问题意识而不是降低成本，主动增加困难、保留不确定性、追问模型的错误。代理坍缩机制解释不了为什么有些行动者不沿最低成本路径走。这里可能存在一种本书尚未说明的自我界面机制。第二个洞：某些教师—AI 组合可能提高评价公平而不是削弱它—模型帮助教师发现偏差、提供多种反馈、减少疲劳。本书能说明评分信用可能被重新分配，不能判断这种重新分配何时改善了教育正义。后续研究不能预设所有代理化都是负面结果。

以下五条写明各自的状态：哪一条已经由第十至十二节的历史序列跑过，哪一条仍未执行。

条款一（未执行）：若在连续三个学期、不同学科与不同学校中，来源记录、责任追问和过程日志能够以低成本稳定区分独立、协作与外包三类完成，并且区分结果能进入正式成绩与资格决策，则”学生—AI 拒绝被约简”的判断被削弱。既有解释可以预测工具使用与成绩之间的关系，不能独立预测这种稳定分解；若它出现，说明旧账本已经发生结构性升级。

条款二（已由历史序列部分跑过）：若同一复杂题目在连续三届学生中未出现题库、模板、预生成答案或模型代理路径的收敛，且学生必须现场承担不可预先准备的判断，则”深度题会被代理化”的判断被推翻。前文的合同代写序列已给出反向证据：开放性任务在人工智能之前十余年已被商品化为可定价的产品规格。该条款在生成式人工智能条件下的重跑仍待执行。

条款三（已由历史序列部分跑过）：若指标失真被系统记录之后，该指标在分配决策中的权重在五年内显著下降并被替代读数取代，则”空值结算”的判断被推翻，Goodhart 传统的”失真即退场”预测得到支持。前文的分数通胀序列给出反向证据：失真持续二十年而权重不降。该条款需要在其他指标体系（如科研计量、绩效考核）中重跑，以检验它是否只是教育的特例。

条款四（未执行）：若企业招聘、职业准入与学校考试长期保持相同的分数信任结构，即使执业中出现可观察风险，资格制度也不转向持续责任承担，则”外部风险账本将先于校园题目推动资格改变”的判断不成立。前文只提供了正向案例，没有提供反向案例的系统排除。

条款五（未执行）：若教育机构在不改变外部就业与资格制度的情况下，仅通过考题内容升级就持续减少代理路径、提高责任回收率并改善真实任务表现，则”考题改变不能自动改变单位”的判断错误。该结果将表明墙内制度具有独立重构土壤的能力。

条款六（未执行）：若教师在使用模型批改之后仍能以独立、稳定且可追责的方式解释每一条评分依据，且教师—AI 组合没有改变评分信用的归属，则前文”新单位首先在记账侧合法出生”的判断被削弱。既有技术效率解释可以预测批改速度提升，不能预测教师专业责任是否被重新分配；检验方法是抽取模型参与批改的评语，让教师在不看模型输出的情况下复述评分依据，比较复述与原评语的一致性以及教师对不一致之处的归责。

对不利结果的处理原则是：不把未执行条款伪装成结果。过程记录与课堂现场可能确实能在某些学科稳定恢复学生责任；职业准入也可能因政策传统、行业利益或公共风险结构不同而迟迟不改变。本书的命题不是”任何地方都无法测量”，而是”在人工智能深度进入理解与判断、评价继续以终局分数分配机会的条件下，单位分解具有系统性困难”。

第六十八章　定义独立律：教育审计为何不可能

场景

一名学生提交课程论文，结构完整、语言流畅、论证清晰；教师按标准给高分，学校记入成绩单，学生凭它申请学位、奖学金、实习与工作。没有任何环节停下。追问一层：这份成绩证明了什么？它可以证明学生提交了合格文本，也可以证明学生能借助某套工具完成合格文本；学校却把它读作学生独立阅读、构思、论证与判断的能力。报告主体仍是「学生」，生产已经包含模型、提示、检索、材料与教师给的任务结构。被记录的结果仍在，被记录的主体已不自明。

站内已把这种「校准断、通约续」的状态命名为空值结算：分数对能力的测量内容趋于空缺，却仍承担资格结算功能。本书不再重述它为何发生，而问它为何**查不出来**。

一个被默认的答案

教师知道，学生知道，招生官与雇主也大致知道成绩与能力的关系在松动。按常识，一个人人知道失真的凭证应当有人来查—就像财务报表失真会有审计师。既有的解释把查不出来归因于「不独立」：学校自己设计任务、自己评分、自己发证、自己宣称有效，鉴证者与被鉴证者是同一人。于是改革方案顺理成章：引入外部评审、第三方复核、独立认证机构。

本书要推翻的正是这个答案。它把教育审计的不可能放错了位置。

承重命题

教育审计不可能，不是因为鉴证者不独立，而是因为被鉴证的对象由发行者定义。程序独立可以移植，定义独立移植不了。

一张成绩单背后站着三种权力：**定义权**（这个分数证明什么—课程完成？条件内表现？独立能力？可迁移能力？）、**计量权**（谁来测、用什么任务测）、**发证权**（谁签发、对谁有效）。财务审计之所以可行，是因为这三权早在审计出现之前就已分立：会计准则委员会定义什么算收入、什么算资产，企业按准则计量，审计师复核计量是否合规。审计师不需要重新定义「收入」，他只需检查企业有没有按准则记。教育里三权同在学校一手：学校定义分数证明什么，学校测，学校发。此时引入再独立的外部审计，它能检查的只有「学校是否按自己定的规则给了分」—**程序独立只能证明旧账本被正确执行，不能证明旧账本记的是正确对象**。AI 让这件事第一次无法再被掩盖：单位从「学生」漂向「学生-AI」的漂移，恰好发生在计量环节—谁来测、测的是谁。定义权若不外置，成绩单就会继续把条件性成果写成个人独立能力；计量权若不外置，即使定义已经外置，测出来的仍是发行者想让你看到的达成度。本书把这条

命题叫作「定义独立律」，并在前文对一份现行认证章程做一次真跑，看三权在现实制度里各落在谁手上。

第六十九章　五条脉络与它们共同放过的那一权

测量效度：它查分数，不查定义分数的人

Messick（1989）把效度理解为对分数解释与使用的整体判断；Kane（2013）要求从观察到分数、从分数到能力、从能力到决策的每一步推论都有证据。这一脉握住的变量是「分数能否支持能力推论」。它能判断 AI 介入后原任务是否仍测目标构念，却把「目标构念由谁定」当作既定：效度论证的起点是一个已经被定义的解释，它检验解释是否成立，不检验定义解释的权力归谁。

指标异化：它解释腐蚀，不解释腐蚀后为何无人能查

Goodhart（1975）、Campbell（1979）说明指标一旦用于控制或决策便承受腐蚀压力。这一脉解释了题目为何被模板化、AI 为何被纳入最便宜的得分路径。它不解释为什么腐蚀是公开的却不可纠正—Goodhart 的银行指标腐蚀后会被央行换掉，因为定义指标的人（央行）与被指标约束的人（商业银行）不是一人；教育里定义分数的人与被分数约束的人是同一人，指标腐蚀后没有一个外部主体有权宣布它作废。

信号与信息不对称：它解释凭证被需要，不解释凭证由谁定义

Spence（1973）把教育当作劳动力市场信号；Akerlof（1970）说明质量不可观察时市场依赖信号并可能逆向选择。这一脉解释雇主为何继续读成绩单。它把信号当作既存物，问信号可不可靠，不问信号的内容由谁规定—而 AI 之后信号所指的对象已经变了：成绩单仍被读作学生能力，可能只证明学生在某套工具配置下完成过任务。信号理论没有一个位置放「谁有权重写信号的含义」。

制度合法性：它解释仪式为何维持，不解释仪式审计什么

Meyer 与 Rowan（1977）说明正式结构可以作为制度化的神话与仪式存在，DiMaggio 与 Powell（1983）说明组织在压力下趋同。这一脉解释了评审、认证、监考为何被反复举行。它把审计本身也当作仪式，因此不区分「审程序的仪式」与「审对象的鉴证」—正是本书要切开的那条缝。

审计独立性：它要求鉴证者不参与主张的形成，默认主张的对象已被外部定义

审计理论要求鉴证者不参与被鉴证主张的形成、不从结果中获益、能自行取证。它握住的是「鉴证者与被鉴证者的分离」。它有一个从未写出的前提：被鉴证主张所指的对象（收入、资产、负债）已由准则外部定义。这个前提在金融里成立到无需说出，在教育里不成立到无人察觉。教育改革引入的每一种「独立评审」，都把审计理论连同它的隐含前提一起搬了过来，结果是搬到了程序独立、丢了定义独立。

缺口的精确位置

五条脉络各握一权的一角：测量效度握计量的有效性，指标异化握计量的腐蚀，信号理论握发证的需求，制度合法性握发证的仪式，审计理论握鉴证者的分离。没有一条正面问：**定义权在谁手上，它能否与计量权、发证权分开**。本书站进去的位置就是这一权。

以下九位各占一个位置。判决性反例是同一案例上两家读数不同的那一点。

Messick（1989）。邻接命题：效度是对分数解释与使用的整体判断。分离线：他检验解释是否成立；本书检验解释由谁规定。判决性反例：学校把「独立论文能力」改定义为「规定工具条件下的论证与审核能力」后，同一批分数效度恢复。按他：效度论证更新。按本书：定义权行使了一次，鉴证者对此无权置喙。

Kane（2013）。邻接命题：分数解释需要一条可论证可反驳的推论链。分离线：他要求推论链有证据；本书要求推论链的起点（构念定义）不在发行者手中。判决性反例：推论链每步有证据，但构念由学校自定，外部复核只能审链不能审起点。

Goodhart（1975）／Campbell（1979）。邻接命题：指标用于控制或决策后被腐蚀。分离线：他们解释腐蚀；本书解释腐蚀为何不可纠正—央行可换指标是因为定义者与被约束者分离，学校不可换是因为二者同一。判决性反例：银行指标腐蚀后被替换，教育指标腐蚀后被保留。

Spence（1973）／Akerlof（1970）。邻接命题：凭证作为信号被需要。分离线：他们把信号当既存物；本书问信号内容由谁改写。判决性反例：雇主发现信号失真却无处申请重定义，因为信号的定义权不在市场。

Meyer 与 Rowan（1977）。邻接命题：正式结构作为仪式维持合法性。分离线：他们把审计也当仪式；本书区分审程序的仪式与审对象的鉴证。判决性反例：认证机构每六年到访、逐条核对文件、给出通过—仪式完整，对象未审。

审计独立性传统。邻接命题：鉴证者不得参与主张形成。分离线：它默认对象已被外部定义；本书指出教育里这个默认不成立。判决性反例：外部评审团完全独立，仍只能问「你们有没有按自己定的标准给分」。

会计确认传统。邻接命题：事项须满足主体、时点、计量、责任四门槛才入账。分离线：它确认财务项目；本书把门槛用于审资格主张。判决性反例：AI 生成论文按确认门槛须重判主体，成绩单直接记作个人。

银行预期损失传统。邻接命题：持续流通的资产不能等实际违约才确认损失。分离线：它处理金融资产减值；本书转为资格有效性损失。判决性反例：毕业生复测失败率上升，银行结构要求计提，学校继续无减值发证。

站内《代理坍缩与空值结算》。邻接命题：分数失去测量含义后不退场、流通更广。分离线：它解释为何不退场；本书解释为何查不出—不退场是结算侧的事，查不出是定义侧的事。判决性反例：即使外部机构停止使用成绩（空值结算终止），学校自定义自计量的结构不变，下一种凭证照样查不出。

九位摆出来可以看清：没有一位把「审计独立」拆成程序独立与定义独立两层，也没有一位问三权归谁。

第七十章　三权、两种独立与题设的改写：成绩单只确认资格效力，不确认失效暴露

三权

定义权：规定某一资格凭证证明什么—它所主张的能力对象、适用条件与外推范围—的权力。

计量权：设计任务、实施测量、判定达成度的权力。

发证权：签发凭证、宣布其对外效力的权力。

三权可以分别归属不同主体。金融报表：定义权在准则机构，计量权在企业，发证权（签署报表）在企业管理层，鉴证在审计师—审计师复核的是计量是否符合外部定义。教育成绩单：三权通常同在一校。

两种独立

程序独立：鉴证者不参与被鉴证主张的形成、不从结果获益、能自行取证。

定义独立：被鉴证主张所指的对象由鉴证者之外、发行者之外的第三方规定，发行者无权单方面改写。

定义独立律的形式表述：**当定义权与发证权同在一主体时，任何程序独立的鉴证只能证明该主体按自身定义正确执行，不能证明其定义所指对象与凭证被外部使用时所假定的对象一致**。

空值结算（承接站内定义）

一个资格凭证的能力校准功能下降或失效，而其在机构之间的排序、筛选与分配功能仍被保留并继续运作。本书把它当作背景状态，不再重新论证。

题设的改写：为什么不能说「只记贷不记借」

题设把成绩单比作只记贷方、不记借方的账。字面上这不成立：成绩单不是财务报表项目，学校不把学分记为收入，没有对应的借方科目可以缺席。可迁移的不是科目，而是**确认门槛**的缺席：复式记账要求一项事项在主体、时点、计量、责任四处都得到确认才能入账；成绩单只在「发证」一处确认（这名学生获得了某项资格效力），在「主体是谁、在什么工具条件下、失效概率多大、失效后谁补救」四处都不确认。因此准确的说法是：**成绩单只确认资格效力，不确认资格失效暴露**。「借方」在这里的含义是失效暴露—验证成本、补救责任与不确定性—而不是某个负数科目。

操作性读数

本书不再使用带自由权重的合成指数。三条读数各自单次可取值：

三权归属读数 T：对任一资格制度，分别读定义权、计量权、发证权的归属主体，记为三元组 (T_def, T_meas, T_iss)，每项取值为「发行者」或「外部」。直接读章程、准则与证书版式，不做推断。

定义独立读数 DI：T_def＝外部且发行者无权单方面改写定义，取1；否则取 0。

计量独立读数 MI：T_meas＝外部，或外部主体拥有自行命题、换工具、抽样复测并据此改写达成度判定的权限，取1；否则取 0。

资格资本充足率 K_q＝(V＋R)/Q：V 为发行者可执行的独立复测能力（按标准复测单位计），R 为补救能力（补修、重新认证名额），Q 为仍在外部流通的高风险资格数量。比率须标明单位与一次完整复测或补救所耗资源。

定义独立律的经验形态：**DI＝1 而 MI＝0 时，认证可以通过而空值结算照常持续；只有 DI＝1 且 MI＝1，外部鉴证才触及计量环节的单位漂移**。

成立条件

第一，凭证被外部机构用于资格分配，且其能力主张超过课程完成记录。

第二，凭证的定义权与发证权同在发行者，或定义权已外置但计量权仍在发行者。

第三，学习成果由学生与工具共同生成，计量环节是单位漂移的实际发生地。

第四，外部鉴证（如有）只审程序合规，不拥有重新定义资格主张边界或自行复测的权限。

四条同时成立，定义独立律成立：鉴证可以通过，凭证照常失真。

不成立条件

若某一资格制度中 DI＝1 且 MI＝1—外部主体既规定凭证证明什么，又拥有自行命题、换工具、抽样复测并改写达成度判定的权限—本书关于「教育审计不可能」的判断在该制度上不成立。本书预测这类制度存在于高风险执业准入（型别等级、临床技能考核）而不存在于学历成绩单，理由见前文。

若出现发行者自定义、自计量、自发证而外部鉴证仍能稳定查出并停下失真凭证的案例，定义独立律被推翻。

两句复合测试

最接近的两句：Messick—分数效度取决于其解释与使用是否有证据支持；审计理论—鉴证者不得参与被鉴证主张的形成。两句合起来说的是「分数解释可能失效，所以要有独立的人来查」。它们盖不住的那条缝是：**查什么？** 两句都默认被查的对象已被定义。本书的增量只落在这一处：对象由谁定义，决定了独立的鉴证能查到哪一层。

第七十一章　定义独立律的判别式与五个场景

判别式

对任一资格制度读 (T_def, T_meas, T_iss)：

若三权同在发行者，DI＝0，MI＝0：任何外部鉴证只能审程序，定义独立律成立，教育审计不可能。

若 DI＝1 而 MI＝0：外部规定了凭证证明什么，发行者自行测达成度并自行发证；鉴证审「有无过程」而非「过程测到了谁」。定义独立律以弱形式成立：认证通过与凭证失真并存。

若 DI＝1 且 MI＝1：定义独立律不成立，鉴证触及计量。

判别式不含情态词，读三份文件即可：章程（定义权）、评估条款（计量权）、证书版式（发证权）。

五个场景

场景一：学历成绩单。三权同校。DI＝0，MI＝0。外部审计不可能。

场景二：专业认证下的学位。定义权外置（认证机构规定成果清单），计量权在校，发证权在校。DI＝1，MI＝0。认证通过，成绩失真并存—前文真跑对象。

场景三：飞行员型别等级。定义权在民航当局（机型与科目），计量权在当局授权的考试员（外部），发证权在当局。DI＝1，MI＝1。鉴证触及计量。

场景四：学校自设 AI 使用申报表。三权未动，只加了一栏。DI＝0，MI＝0。申报表成为新的程序合规项。

场景五：外部机构另行复测、不认成绩。发证权仍在校，但外部以自己的定义与计量另立凭证。旧凭证空值结算终止，学校三权结构不变—空值结算与定义独立律是两件事，前者可以停，后者不动。

三条自否决条款

信度条款。 两名独立读者对同一制度读 (T_def, T_meas, T_iss)，任一项不一致率高于 20%，三权读数不稳，不得据此判定。

改名条款。 若 DI／MI 读数与已有的「是否有外部认证」二值完全重合（相关达1），本书只是给「有无认证」改名；判据要求：至少存在一类制度有认证而 MI＝0，另一类无认证而 MI＝1（雇主自测），两类分开才立。

单次事件条款。 一份成绩单失真不推出该制度 DI＝0；三权读数只从章程与条款读，不从个案倒推。

第七十二章　真跑：ABET 工程认证通则的三权读数

材料

美国工程与技术认证委员会（ABET）工程认证委员会（EAC）《工程项目认证标准2025–2026》，公开文本。选它的理由：它是全球引用最广的专业教育认证章程之一，且明文分列「学生成果」与「持续改进」两条，正好把定义与计量写在两个条款里。

读数

定义权。标准3「学生成果」规定：项目必须有文件化的学生成果以支撑其教育目标，学生成果为成果 (1) 至 (7)，加上项目可自行补充的成果。七项成果由 ABET 规定，项目不得删减，只能追加。T_def＝外部。DI＝1。

计量权。标准4「持续改进」规定：项目必须定期使用适当的、有记录的过程，评估与评价学生成果 (1)–(7) 的达成程度，并把评价结果系统地用作持续改进行动的输入。测量的主语是「项目」；ABET 不命题、不抽样、不换工具复测，它审查项目是否有过程、过程是否有记录、结果是否被用于改进。T_meas＝发行者。MI＝0。

发证权。学位由院校授予；ABET 认证的是项目，不是学生。T_iss＝发行者。

三元组：(外部, 发行者, 发行者)。DI＝1，MI＝0。

读数的含义

这是定义独立律的弱形式实例。ABET 完成了金融审计的第一步—把「证明什么」外置—却没有完成第二步：计量仍在发行者手中。现场评审员核对的是「项目有没有一套评估过程、过程有没有闭环到改进」，这正是本书所说的程序独立：它证明项目按外部定义正确执行了自己的测量，不证明测量到的是谁。

AI 时代的单位漂移恰好发生在计量环节。当学生成果 (1)（识别、表述并解决复杂工程问题的能力）由学生-AI 共同达成，项目用自己的作业与考试评估达成度，报告「达成度85%」，认证按标准 4 审查过程完备—通过。成绩单继续把条件性成果记作个人能力—失真。二者同时成立，不需要任何人作弊。

一条已跑出的反向证据与一条边界

反向证据：定义外置已经发生，说明「定义权不可能离开学校」这一更强的表述不成立；本书因此只主张定义独立**不足以**使审计可能，不主张它不可能实现。

边界：ABET 标准 4 允许「其他可用信息」进入持续改进，项目可以自愿引入外部考试（如 FE 考试）作为证据；若某项目把外部考试成绩作为达成度的主要依据，

其 MI 读数在该成果上变为1。本书预测这类项目是少数，且外部考试所测的仍是无工具条件下的个人，与学生-AI 单位再次错位—此预测列入前文证伪条款。

第七十三章　三权同一下的三条改革路线

独立评审路线

最强形式：引入校外专家评审、第三方复核、盲评。对撞结果：评审员拿到的是学校自定的评分标准与学校自设的任务，能核的只有「是否按标准给分」。程序独立到位，定义独立缺席。评审越严，旧账本执行得越正确，失真越有凭有据。

认证路线

最强形式：专业认证机构规定成果清单，项目须证明达成。对撞结果：见前文—定义外置、计量留内。认证通过与凭证失真并存，且认证本身成为失真凭证的背书。

AI 申报表路线

最强形式：每份作业附来源栏，学生勾选独立／协作／代办。对撞结果：三权一权未动，申报表是发行者自定义、自计量、自发证之下新增的一栏，成为新的程序合规项；勾选「代办」的后果由发行者自定，鉴证者无权置喙。它把失真从暗账变明账，是可取的第一步，但不是审计。

三条路线撞在同一处：它们都在增加程序独立，没有一条触及定义权与计量权的归属。

第七十四章　确认门槛与风险准备：三权分立之前能搬什么——资格资本充足率

确认门槛

复式记账要求事项在主体、时点、计量、责任四处确认。搬入成绩单：主体—谁生成（学生／学生-AI，在何种工具配置下）；时点—生成于可见时段还是不可见时段；计量—按什么任务测；责任—失效后谁复测、谁补救。可迁移：四处确认作为入账前提，不依赖三权分立，发行者自己就能做。断裂点：发行者自定的确认门槛仍由发行者自审—门槛能让失真进账页，不能让失真被外部查出。

预期损失准备与资本约束

银行对持续流通的资产按预期损失计提准备，并以资本约束资产扩张。搬入成绩单：资格发行量应与发行者的复测与补救能力匹配，$K_q=(V+R)/Q$。可迁移：K_q 由发行者自报，外部可核；它把「无风险行政记录」改写为「有失效暴露的凭证」。断裂点：谁规定 Q 里哪些资格算「高风险」—又回到定义权。

两套结构的共同读数

两套都能在三权不变时先行移植，两套都在同一处断：定义。这与前文的真跑一致—止损可以从发行者内部开始，审计不能。

第七十五章　类型学：定义独立 × 计量独立

候选轴排除：「是否有外部认证」（与 DI 重合，改名条款已排除）；「AI 使用程度」（成因，非结构）；「凭证是否被外部使用」（那是空值结算的轴，不是本书的）。

DI＝0，MI＝0：自证凭证。 学历成绩单。独有预测：外部评审频率与凭证失真程度无关。

DI＝1，MI＝0：认证凭证。 ABET 类专业认证。独有预测：认证通过率与毕业生复测失败率长期不相关；定义清单更新后失真不减。

DI＝0，MI＝1：他测凭证。 雇主自行笔试、面试、试用，不认成绩但也不规定成绩证明什么。独有预测：旧凭证空值结算终止，学校三权不变，下一种凭证照旧失真。

DI＝1，MI＝1：分权凭证。 型别等级、临床技能考核。独有预测：凭证证明的对象随器具配置变化而变化（此人＋此机型），单位漂移被计量环节接住。

四格不可互换：第二格与第四格只差一权，一权之差决定审计能否发生。

理论意涵

第一，审计独立性不是一个属性，是两层：程序独立与定义独立。教育改革三十年的独立评审全部落在前一层。

第二，Goodhart 定律在教育里不可纠正，不是因为教育指标更易腐蚀，而是因为定义指标者与被指标约束者同一。这为 Goodhart 加了一个限定：**指标腐蚀可纠正的前提是定义权外置**。

第三，空值结算与定义独立律是两个位置：前者解释失真凭证为何不退场（结算侧），后者解释它为何查不出（定义侧）。前文的判决性反例已示：结算可以停，三权不动。

实践意涵

第一步不是废分数、不是加难题、不是引评审，而是把三权写出来：每一种凭证在其章程里明示定义权、计量权、发证权各在谁。写出来本身不改变归属，但让「独立评审」不再能冒充审计。

第二步，先搬两套不需要分权的结构：确认门槛（四处入账前提）与 K_q（发行量与复测补救能力匹配）。

第三步，定义权外置—学专业认证已做的；第四步，计量权外置—专业认证尚未做的：认证机构须拥有自行命题、换工具、抽样复测并改写达成度判定的权限，并对「此人＋此工具配置」而非「此人」计量。第四步做到，教育审计第一次可能。

适用边界

其一，只记课程完成、不对外主张能力的凭证，三权同一无害。其二，课程目标本就是人机协作且凭证明写工具条件者，定义已含工具，漂移不在计量环节。其三，风险后果弱、市场靠试用识别的行业，他测凭证（第三格）足够，无需分权。其四，本书只读了一份章程；定义独立律的弱形式实例只有一个，强形式（三权同一）是学历成绩单的通例，不需要真跑即可读出。

反噬

照本书走，最省事的做法是把三权「写出来」变成又一份合规文件：章程里加一段「本校定义权、计量权、发证权归属如下」，然后照旧。这会把定义独立律本身变成程序合规项。防线只有一条：三权读数必须由发行者之外的读者读，且计量权外置的判定标准是「外部主体可自行命题、换工具、复测并改写判定」四项俱全，缺一即 $\mathrm{MI}=0$。

第七十六章　定义独立律的证伪条款与局限

第一条 [已跑一半]：若在主要专业认证章程（工程、医学、法律、会计任三家）中，多数把计量权外置（认证机构自行命题、抽样、复测并据此改写达成度判定），本书关于「定义独立而计量不独立」的判断被推翻。ABET 一家已读，MI＝0；其余待读。

第二条 [未执行]：若某一三权同在发行者（DI＝0，MI＝0）的学历成绩单制度中，外部评审在五年内稳定查出并停用了失真凭证，定义独立律被推翻。

第三条 [未执行]：若 DI＝1、MI＝1 的制度（型别等级、临床技能）中，凭证失真程度与 DI＝1、MI＝0 的制度无显著差异，则计量独立不是决定变量，本书须收缩为「定义独立律只解释为何查不出，不解释失真程度」。

第四条 [未执行]：若 ABET 项目中把外部考试作为达成度主要依据者占多数，7.4 的预测错误，MI 读数须逐成果重读。

第五条 [未执行]：若学校在三权不变的条件下，仅凭 K_q（资格资本充足率）的公开与约束就使凭证外推范围自行收窄，则「三权分立之前的止损」比本书估计的更强，定义独立律的实践优先级须下调。

替代解释。 其一，审计不可能是因为教育对象本质上不可测，与三权无关。若 DI＝1 且 MI＝1 的执业考核同样无法稳定测出单位，此解释成立；本书预测型别等级能测，因为它测的是「此人＋此机型」而非「此人」。其二，三权同一是资源约束而非结构：认证机构没钱自测。若某认证机构获得充足资源后仍不接管计量，资源解释被否。其三，定义外置本身足以止损，计量可以留内。ABET 读数已给出反证：定义外置三十年，成绩失真照常。

构念效度。 「定义独立」可能被理解为「有无认证」的改名。防线在前文改名条款：须存在有认证而 MI＝0、无认证而 MI＝1 两类。ABET 读数已给出前一类；后一类（雇主自测）本书只作类型学陈述，未读具体制度。

内部效度。 认证通过与凭证失真并存，可能由资源约束、行业惯例或评审员能力造成，不必由三权配置造成。本书未分离这些因素；前文替代解释二给出了否证路径。

外部效度。 真跑只读了工程认证一家；医学（LCME／GMC）、法律、会计认证的三权配置可能不同，尤其执业再认证类制度可能已部分外置计量权。

本书解释不了的两个洞。 其一，为什么金融领域的三权分立发生在十九世纪末二十世纪初，教育到今天没有发生—本书能说明分立的功能，不能说明分立的历史条件。其二，计量权外置后，认证机构自身的定义权与计量权同在一手，是否会在更高一层重演定义独立律；本书只推到第二层。

就三权而言，ABET 通则已实现定义独立（学生成果 (1)–(7) 外置），未实现计量独立（项目自测达成度）；AI 时代单位漂移恰在计量环节，故认证通过而成绩失真

并存。而确认门槛与风险准备可在三权不变时先行移植：止损可从内部开始，审计不能。

本书交出的是一条定律、三条读数、一次章程真跑与五条证伪条款。下一份要读的不是成绩数据，而是医学、法律、会计三家认证章程的标准3 与标准4—看是否有任何一家把计量权拿走了。有一家拿走了，教育审计就在那一家第一次可能。

第七十七章　空转：考核单位死亡后教育机器何以不停

一个仍在发放文凭的死亡单位

某大学要求学生提交一篇课程论文。学生在期限前上传一份论证完整、语言流畅、格式规范的文本；教师通过反抄袭系统、格式检查与常规批改后给出成绩。若只看交付物，教学流程没有显著异常：题目发布、作业提交、评分、排名与学分认定全部完成。然而，若把问题改写为「这份成绩到底指向谁的生成」，原先的流程便出现了断裂。文本可能由学生独立构思，也可能由学生与模型共同生成，还可能主要由模型完成、学生负责选择、改写和提交。最终文本仍然存在，分数也仍然可以记录，但分数所指向的对象已不再稳定。

这一反常并不等于「学生变懒了」，也不只是传统作弊手段获得了更强工具。教育考核长期依赖一个本体预设：被评分的产出能够指回一个相对独立的学生，因而分数可以作为该学生知识、能力或努力的代理。生成式人工智能改变的首先不是产出质量，而是产出的生成单位。学生与人工智能共同完成任务时，产出不再自然地属于学生这一单一单位；但教育制度仍然必须在同一时间发放分数、学分与资格。

因此，问题不应表述为「学生是否使用了人工智能」，而应表述为：当旧考核单位不再能够稳定指向被测者时，教育机器为什么仍能维持原有的运转表象？

现有解释在何处失效

对于这一问题，现有解释至少形成了三条重要脉络。第一条将其视为学术诚信问题，重点关注禁止、检测与惩罚。它能够解释学生为何隐藏使用人工智能，却无法解释即使教师、学生和机构都知道独立完成的假设已被普遍削弱，分数、文凭和排名仍然继续发挥分配作用。

第二条将其视为教育测量问题，重点讨论评分信度、效度与评估方式的调整。它能够说明某一指标为什么不能再准确测量原先的构念，却通常将「指标失效」理解为需要替换题型、增加过程性评价或提

高监测强度，因而没有追问：为什么一个已经失去指向的指标仍被制度性地需要？

第三条将其视为技术变迁带来的制度适应问题，强调教育机构需要更新课程、教师能力和考核内容。它能够说明制度如何吸收新技术，却容易把问题理解为「旧制度尚未升级」，仿佛只要设计出更复杂的题目，新单位就会自然显露。实际上，任何新题一旦进入既有分配结构，都可能被转化为新的训练对象、题库对象和代理对象。

这些解释分别抓住了行为、测量与制度的一面，却共同遗漏了一个更尖锐的中间层：教育考核并不只测量学习，也在分配稀缺位置。即使测量对象死亡，只要分数仍能为分配提供公开、可复制、可辩护的理由，考核机器就没有停转的动力。由此产生的不是「悬空」，而是「空转」。

本书占据的缺口

本书的承重命题是：

生成式人工智能时代的教育危机不是旧考核单位死亡、新考核内容尚未成形所造成的悬空，而是旧考核单位已经在指向与检查意义上死亡，分配功能却借助可辩护性继续高速运转，由此形成空转。这一命题把「死亡」与「运转」拆开。旧单位的真值可以死亡，旧制度的分配收益却仍然存活；正因为两者被分离，教育系统才不需要立即承认单位已经失效。本书不主张新考核单位已经完成，也不主张所有教育评价都已失去效度。本书只主张：在生成式人工智能广泛介入产出的条件下，原有分数无法不经说明地继续被解释为单一学生生成的读数；而制度维持旧分配装置的原因，不是它仍然能够完成原有测量，而是它仍然能够提供可辩护的分配程序。

第七十八章　空转的定义与三个读数：指向×检查、来源构成、可重走

理论出发点：从发现转向发生

本书站在发生学的理论传统上，并以作为分析骨架。发生学不把「学生分数失真」当作一个等待识别的静态事实，而追问这一事实如何形成、哪些条件使其持续、它又如何反过来改变条件。中的 S、D、E 分别表示显露、差异路径与纠缠土壤。本书不把三者当作并列变量，而把它们理解为同一过程的三个分析位置：某种现象如何显露，哪些路径被保留，哪些条件彼此纠缠并回写路径。

在教育空转问题中，S 是「分数、排名、文凭仍然正常」的显露态；D 是学生、教师与机构在旧账本下形成并保留的行为路径；E 是分配稀缺、职业风险、制度合法性、教师资格与自我辩护相互纠缠的土壤。三者不是单向因果关系，而是互相生成。可以形式化为：

分数的显露由既有路径与制度土壤共同产生；路径又受到分数反馈与土壤约束；土壤则被持续的分数与路径重新加固。空转不是某一因素造成的结果，而是三个位置形成闭环后的稳定态。

核心概念的名义定义

考核单位是一个教育产出被归属、测量并用于分配时所指向的生产主体。单位死亡是考核产出不再稳定指向该生产主体，且制度不再检查这一指向是否成立的状态。空转是考核单位失去指向与检查功能后，分数、排名、文凭等分配装置仍持续运行的制度状态。

可辩护性是一个分配结果能够借助公共程序、共同认可的凭证或形式规则获得正当说明的属性。来源构成是一次产出中独立生成、人机协作与人工智能代办的归属类别。可重走是同一行动者在新的、不可预处理的材料与时间条件下再次完成与原产出同类生成动作的检验。

操作性定义与读数

本书将空转定义为一个四值类别变量：

$$\mathbf{I = R \times C}$$

其中， R 表示指向是否成立， C 表示制度是否检查指向。二者均为二值类别变量：成立记为1，不成立记为0；检查记为1，不检查记为0。由此形成四类：

指向 R	检查 C	状态
1	1	可测状态
1	0	默认状态
0	1	暴露危机
0	0	空转状态

测量层次为类别层次。一次可观测事件是「一份被提交并用于评分的产出」。若教师或制度能够通过过程材料、现场生成与责任说明，使产出稳定指向学生，并且明确检查这一指向，则取 R=1,C=1。若产出仍可归属于学生，但制度不再检查归属，则取 R=1,C=0。若归属无法稳定区分，但制度仍尝试检查，则取 R=0,C=1。若归属无法稳定区分，制度也不检查这一问题，而继续给出分数和资格，则取 R=0,C=0，即空转。

来源构成读数为三类名义变量：独立生成、人机协作、人工智能代办。它不表示贡献百分比，不把不可稳定分解的过程伪装成精确比例。一次事件中，若学生从未使用生成式工具并能提供过程材料，取「独立生成」；若人工智能参与理解、构思、改写或反馈，且学生承担部分生成责任，取「人机协作」；若主要内容由工具生成，学生只进行选择、提交或表面修改，取「人工智能代办」。无法判定时取「未明示」，不得强制归入前三类。

可重走读数为二值类别变量。提供未被预先公布的材料、限定现场时间并要求行动者再次生成同类产出；若第二次产出表现出可解释的生成连续性，取1，否则取0。它不是概率、倾向或程度，不以单次结果推断总体能力。单次可观测事件是「同一行动者对一份新材料完成一次重生成」。

第七十九章　命题：不是悬空而是空转——分配功能借可辩护性继续运转

命题形式

本书的核心判断是：

教育空转不是悬空，也不是单纯的测量失效，而是旧考核单位既不再指向学生独立生成、也不再接受制度检查，却仍以分数、排名和文凭形式承担分配与辩护功能的稳定状态。其中， Y_1 是悬空，指旧单位失效后教育机器停摆、等待新单位接替；Y_2 是单纯测量失效，指制度仍然以真实性为主要目的，只是暂时找不到更好的指标； Z 是空转，指测量功能退化而分配功能继续运行。

空转在以下条件下成立：

第一，产出与学生独立生成之间不存在稳定、可重复的指向关系。

第二，评分制度不把「指向是否成立」作为入账前提。

第三，分数、排名或文凭仍被用于配置稀缺教育或职业位置。

第四，制度能够通过程序完备性为分配结果提供公开说明。

第五，外部使用者尚未将旧凭证的错购成本完全定价并转化为停兑行动。

空转在以下情形下不成立：如果产出归属能够被稳定分解、制度在每次入账前检查归属，并且分配逐步转向新的可验证凭证，则单位危机已进入转换而非空转；如果分配功能也同时停止，系统表现为悬空或崩解，而不是空转。

两句复合测试

本书命题不能由两句现成文献无损拼接而成。最近的两条解释分别是 Messick（1989）关于效度必须联系分数解释与使用的判断，以及 Meyer 和 Rowan（1977）关于正式结构可以作为合法性仪式存在的判断。前者能够说明分数解释失效，后者能够说明制度程序为何继续，但两者合并仍只能得到「教育机构可能继续使用失效指标以维持合法性」。

本书的增量在于指出：失效的不是一般指标，而是考核单位对生产主体的指向；继续运行的也不只是合法性，而是具体的分配功能。二者在同一装置中同时发生，形成了「测量死亡—分配存活—制度空转」的机制链。

第八十章　最近邻：效度理论、组织合法性、信息不对称与指标操纵

本节没有经过站外检索。下表中的占位者依据已提供材料与一般理论线索暂列，均标记为〔未核验〕；不据此宣称其完整代表相关领域，也不将其未核验内容当作经验事实。

占位者	他已经占了什么	分离线	判决性反例
Messick（1989）〔未核验〕	效度是对分数解释和使用的综合判断	他判分数解释是否成立；本书判考核单位是否仍能指向生产主体	同一分数可被重新解释为文本质量而非学生独立生成：他判解释转换，本书判单位已死
Kane（2013）〔未核验〕	效度论证连接观察、评分与推断	他检验推断链是否成立；本书检验推断链中的行动主体是否稳定	现场生成过程被模型预先完成，学生只在场复现：推断链有观察却无主体指向
Meyer 与 Rowan（1977）〔未核验〕	正式结构可作为制度化神话与仪式维持合法性	他们判形式结构如何维持合法性；本书判测量单位死亡后分配功能如何继续	全流程符合程序、成绩仍被用于录取，但产出无法归属：合法性与单位空转同时成立
DiMaggio 与 Powell（1983）〔未核验〕	组织在强制、模仿和规范压力下趋同	他们判组织为何采用相似形式；本书判相似形式如何掩盖单位差异	多校共同采用 AI 检测仍无法区分学生生成与模型生成：趋同解释不了指向死亡
Akerlof（1970）〔未核验〕	质量信息不对称造成低质量交易与逆向选择	他判交易质量如何影响市场；本书判分配凭证如何在质量不可见时继续被使用	雇主知道成绩单质量不稳定却继续收取，因为它仍提供可辩护的筛选程序
Campbell（1976）〔未核验〕	指标用于决策后会受到操纵	他判指标如何被操纵；本书判指标失效后为何仍被保留	题目被 AI 接管、成绩失真，学校继续用成绩分配位置：操纵不能解释制度存续
生成式人工智能与教育研究群体〔未核验〕	AI 会改变学习、写作与评估方式	其判断多落在教学功能变化；本书判教育单位的归属结构变化	学生用 AI 完成论文，文本质量提高但学生生成无法确认：教学变化不能替代单位诊断
职业准入研究与实践共同体〔未核验〕	资格应联系执业风险与专业责任	其判断以职业安全为核心；本书判风险账本如何反向重写校园凭证	学校分数继续有效但临床现场要求可重走：资格改写先于考试改写

表中的分离线不是研究对象或学科侧重的差异，而是同一案例上判决结果的差异。若一名学生提交高质量论文，无法证明其中的生成过程，且学校仍据此发放学分：Messick 可以把问题判为分数解释转换；Akerlof 可以把问题判为质量信息不对称；Campbell 可以把问题判为指标操纵；本书则判定，在制度仍给出分数且不检查归属的条件下，该作业属于空转读数。

第八十一章　空转的判别式、三项指标与自否决条款

判别式

本书提出的判别式是：

当一份产出被用于分配资格时，若其生成来源未在入账时被记录，且行动者无法在新的、不可预处理的材料上现场重走同类生成过程，则该产出属于空转读数。这一判据不要求判断学生「是否诚实」，也不要求估计人工智能贡献百分比。它只询问流程、日志与记录：是否有来源栏，是否在递交时记录，是否存在新的现场材料，是否完成重走。

该判据在五个场景中产生不同答案。

第一，闭卷数学考试，学生现场独立推导，考试记录显示无工具使用，教师随机要求解释关键步骤。来源记录为独立生成，重走为1，不构成空转。

第二，学生回家使用人工智能生成论文，提交时只交终稿，无过程记录，课堂没有重生成。来源为未明示，重走未执行，构成空转读数。

第三，学生允许使用人工智能完成法律意见书，但须提交对模型意见的批判记录，并在课堂对一份新案例进行口头分析。来源为人机协作，重走为1，产出不构成空转，但不能直接解释为学生独立生成。

第四，学生提前让人工智能生成题库中所有可能答案，考试现场完成检索和排列。表面上重走发生，若材料域可枚举，重走并不满足「不可预处理」条件，判据仍识别为来源风险。

第五，学校采用全过程屏幕监控，却不记录学生与模型如何共同生成，也不要求现场重走。监控程序完备，但来源为空，重走未执行，仍构成空转。

可观测指标

来源完整率定义为：

来源完整率 P = 递交时具有来源类别记录的产出数 ÷ 进入评分的产出总数

它是等比层次的比例指标。一次观察单位是一份进入评分流程的产出。来源完整率达到1，表示每份产出均有来源记录；但这不表示来源陈述真实，只表示来源被纳入账页。

重走成功率定义为：

重走成功率 R = 在新材料与新时间条件下完成同类生成的次数 ÷ 被要求重走的次数

它也是等比层次的比例指标。单次重走只有成功或失败两个类别值，不能因为一次成功就推断学生具备稳定能力。重走材料必须满足三项条件：未被提前公布、不能由固定题库穷举、与原任务属于同一生成类型。

空转暴露率定义为：

空转暴露率 I = 来源未明示且重走未执行的分配性产出数 ÷ 全部分配性产出数

这是等比层次的比例指标。一次事件中，若一份产出用于发放成绩、学分或资格，且来源未明示、重走未执行，则记为1，否则记为0。该指标不能直接代表学习质量，只表示分配性读数缺乏单位核验。

读数的三条自我否决条款

第一，信度条款：同一批产出在两轮独立编码中，来源类别一致率低于0.80，来源读数判定为不稳，不得据此比较班级或教师。

第二，改名嫌疑条款：若来源完整率与已有「过程性评价覆盖率」高度重合，且不能说明来源栏额外记录了生成归属，则来源栏只是更换名称，不构成新指标。

第三，单次事件条款：若研究者用总体概率解释单次产出，或由一次重走成功推断稳定能力，则该读数无效。概率只能用于多次观测汇总，不能替代单次事件的类别判定。

第八十二章　证伪条款与两条已知的不利结果

第一组：独立证伪条款

[未执行] 如果在多个教育层级中，所有进入分配的产出都已经普遍具有独立、可核验的来源记录，并且该记录不依赖学生自报，则本书关于「来源未被记入账页」的判断错误。之所以现有解释预测不出这一结果，是因为学术诚信解释通常把来源记录视为反作弊工具，而本书把它视为单位转换的入账条件；若来源已普遍成为正式账页，说明转换已经发生。

[未执行] 如果在题库可穷举、人工智能可预训练的条件下，现场重走仍能长期稳定区分真实生成与预生成，且这种区分不依赖额外的身份监控，则本书关于「新闸门会被入口或材料域绕过」的判断错误。现有解释可以预测检测技术升级，却不能预测绕过路径必然转移；若绕过不发生，路径迁移命题需要收缩。

[未执行] 如果职业准入与招聘市场持续以学校分数为主要依据，即使实际工作表现与分数之间出现稳定偏离，且雇主承担由此产生的重复成本，本书关于外部停兑会先于校园考核改写的判断错误。现有测量解释可能预测学校主动修订考试，但本书预测分配凭证会先在风险承担处被改写。

[未执行] 如果教育机构在分数无法指向独立生成后主动停止发放学分、排名和文凭，而不是继续使用原有程序，则「分配功能独立存活」这一命题错误。现有危机论可以预测制度停摆；本书只有在分配继续的条件下才成立。

[未执行] 如果人工智能贡献能够通过低成本、非自报、跨任务稳定的技术装置被精确分解，并且这一分解在不同任务与不同模型中保持一致，则本书关于「学生—人工智能耦合场不可约简」的判断错误。本书承认这是最脆弱的一环。

已知材料中的不利结果

[已执行] 已提供材料中出现了对本书「来源记录足以止损」的不利结果：来源栏落地后，作弊者可能伪造来源记录，形成留痕与假痕的军备竞赛。该结果削弱了来源栏作为真实性检测器的解释力，但不推翻其作为账页暴露机制的功能。来源栏不能单独证明真实生成，只能使「来源未被记录」不再被制度隐藏。

[已执行] 已提供材料中出现了对本书「课堂可见性足以恢复学生单位」的不利结果：当堂紧盯产出会把终稿外包推回前处理外包，学生可能预先获得模型生成材料，现场只做复现与改写。该结果支持本书关于路径迁移的判断，也迫使本书放弃「课堂在场即可恢复单位」的强表述。

第八十三章　辨别格：来源显露 × 可重走

候选轴的排除

第一个候选轴是「人工智能使用频率」。它不能作为核心第二轴，因为使用频率是路径表现或成因之一，不是能够与单位指向独立组合的结构维度。高频使用可能包括高质量协作，也可能包括完全代办；低频使用也可能在关键步骤上替代学生判断。

第二个候选轴是「题目难度」。难度同样是任务条件，不足以区分单位状态。难题可以被题库化，简单题也可以要求学生承担不可替代的生成判断。若把难度当作第二轴，类型学会把代理路径的结果误认为单位结构。

第三个候选轴是「技术检测强度」。检测强度属于制度反应，不是考核单位自身的坐标。高检测并不等于高指向，可能只是把外包从终稿推回前处理。

第二轴的确立

本书采用两个相关但不重合的坐标：

第一轴 A：来源是否显露。低值表示产出来源未在入账时记录；高值表示来源类别、过程记录与责任归属被纳入账页。

第二轴 B：生成是否可重走。低值表示行动者不能在新的、不可预处理的材料上再次完成同类生成；高值表示能够现场重走。

二者相关但不重合。来源可以显露而重走失败，例如学生诚实填写「人工智能代办」，但无法独立重做；来源也可以未显露而偶然重走成功，例如学生未填写来源，但现场能够完成新任务。一个两轴同时为高的具名真实案例，本书材料未提供可核验的具名案例，因此不使用「正交」一词。可以提出一个可执行的真实场景：医学临床技能准入中，候选人提交人工智能辅助的病例分析，并在未预告的模拟病例上现场完成风险判断；若该场景被具体机构公开采用，它可作为 A 高、B 高的案例，但本书不将其宣称为已发生事实。

二维辨别格

A 低、B 低：黑箱空转来源低、重走低。产出进入分数、学分或资格分配，但制度没有记录其生成来源，行动者也不能在新材料上完成同类生成。该格的核心描述是「结果被记账，单位不被检查」。只在该格成立的预测是：增加监考、优化题面或提高检测密度不会显著降低空转，因为账页仍然缺少来源，生成过程仍不可重走。

A 高、B 低：有声明的替代来源高、重走低。产出明确记录为人机协作或人工智能代办，但行动者不能在新的任务上承担相同生成责任。该格不同于黑箱空转，因为死讯已经进入账页；它也不同于可认证的人机协作，因为责任不能通过重走获得确

认。只在该格成立的预测是：制度会减少「是否使用人工智能」的争论，却增加「这种来源是否足以授予资格」的争论。

A 低、B 高：未申报的可生成来源低、重走高。行动者能够现场完成新材料上的同类生成，但制度没有把原产出的来源记录下来。该格说明重走不能替代来源账页。只在该格成立的预测是：个体能力可以被现场确认，但历史产出的责任归属仍然不可追溯，因而争议集中在凭证有效性而不是当前能力。

A 高、B 高：可追责协作来源高、重走高。产出的来源被记录，行动者能够在新材料上再次完成同类生成。该格不是「学生独立完成」的复原，而是对人机协作单位的可追责化。只在该格成立的预测是：考核可以不再追求虚假的人工智能贡献百分比，而转向记录协作边界、责任承担与重走能力。

四格两两比较显示，A 决定死讯是否进入账页，B 决定行动者是否能承担同类生成。A 高不必推出 B 高，B 高也不必推出 A 高；只有二者同时成立，单位才具备可追责的资格条件。

第八十四章　讨论：测量对象与分配对象的分离，以及资格改革先于考试改革

理论意涵

本书的理论意涵首先在于区分「测量对象」与「分配对象」。教育制度常把二者放在同一分数上：分数一方面声称描述学生能力，另一方面又决定学生进入何种学校、岗位或资格序列。当产出生成单位发生变化时，测量功能首先受损，但分配功能未必同时消失。若理论只关注测量效度，就无法解释制度为何继续；若只关注制度合法性，又会忽略单位本身已经不再成立。

其次，本书把空转理解为意义上的纠缠稳定态。显露态是分数继续流通；差异路径是学生、教师和机构沿低成本路径调整；纠缠土壤是分配稀缺、职业风险、制度程序与自我辩护彼此支撑。空转不是静态错误，而是反馈关系：路径越普遍，土壤越像现实；土壤越稳定，账本越不愿记录来源；账本越不记录来源，路径越容易沿低成本方向扩散。

再次，本书对「新内容会带来新单位」的期待提出限制。内容属于显露层，单位属于生产归属层。内容变化可能改变任务表面，却不必改变谁承担生成、谁承担责任以及如何被记账。若旧土壤不变，新内容会被题库、训练和模型代理化。真正的单位变化必须改变入账条件与责任路径。

实践意涵

实践上，最先应改的不是题目难度，而是递交配置。所有进入分配流程的产出，应在递交时记录来源构成。来源栏不承担侦破全部欺骗的任务，它首先承担「不得让来源继续隐身」的任务。记录可先采用粗分类，但必须明确其用途是责任归属，而非虚假精算。

第二，应把一部分生成活动从不可见时段移入可重走的现场。课堂现场、职业模拟、口头追问与新材料任务能够使行动者承担生成责任。但可重走必须满足不可预处理条件，否则现场活动会退化为预生成材料的检索与改写。

第三，资格改革应先于考试改革。学校考试主要处理当下可评分表现，职业资格还必须处理未来执业风险。若外部职业现场开始以责任、后果与可重走能力作为准入条件，学校考核才会获得改变账本的外部压力。资格改革不等于废除考试，而是使考试不再独占资格证明。

适用边界与反向约束

本书主张在两种条件下不成立。

第一，若人工智能贡献能够被低成本、非自报、跨任务稳定地分解，且分解结果可以直接进入官方分配账本，则本书关于耦合单位不可约简的判断被削弱。本书原本想说「学生—人工智能不能作为单一单位被记」，必须改为「当前制度缺乏稳定分解装置」。

第二，若某类教育目标本来就是训练人与人工智能协作，而不是证明独立生成，则「来源混合意味着单位死亡」的表述需要收缩。此时人机协作可以成为合法考核单位，但仍需满足来源显露与可重走要求，否则协作单位仍不可追责。

第三，若外部劳动市场长期不承担错误录用成本，门外人就可能继续使用旧分数。本书原本想说「外部停兑会自然发生」，必须改为「只有当错购成本被具体主体承担并可计算时，外部停兑才具有动力」。

反噬

若照本书执行，最省事的做法可能恰好毁掉它想保住的东西。学校可能把来源栏变成新的纪律标签，把「人工智能代办」直接等同于低分或惩罚；也可能把可重走变成高压监控，使学生只学习如何在监督下表演生成。这样做会把来源记录重新纳入旧账本，使其成为另一种程序完备性，而不是单位显露装置。

因此，来源栏不能单独决定价值，重走也不能被设计为羞辱性审讯。它们的目标不是恢复一个纯粹的「学生」单位，而是让协作单位的来源、责任与可重复生成能力进入可讨论、可追责的账页。

第八十五章　结论：来源栏、可重走与外部停兑三个连续判据

第一问询问：生成式人工智能如何使教育考核单位失去对学生独立生成的指向？答案是，产出生成从单一学生转向学生与人工智能的耦合场，且每次任务中的人机配比不同。旧账本仍以学生为单位记账，却无法稳定分解学生与工具的贡献。单位因此发生两重死亡：分数不再指回学生独立生成，制度也不再检查这一指向是否成立。

第二问询问：单位死亡后，分配功能凭借什么继续运转？答案是可辩护性。分数不再主要依靠真实性获得生命，而依靠程序完备、门外人信息断口与持证者互为人质获得生命。稀缺位置仍需分配，机构仍需一套公开、低成本、可复制的理由；旧分数即使失去测量真值，仍然能够提供这套理由。教育因而不是悬空，而是空转。

第三问询问：如何区分旧账本继续记账与新单位获得资格？答案不是更复杂的题目，而是三个可裁决条件：递交时记录来源构成；在新的、不可预处理的材料上检验可重走；让职业准入等外部机制依据执业风险改写资格凭证。来源栏使死讯进入账页，可重走使责任获得现场检验，外部停兑则使旧分数失去唯一分配地位。

研究启示是，教育改革不应把人工智能问题缩减为「允许还是禁止」。真正需要改变的是记账单位、入账时点与资格兑现方式。新考核内容不会自动生成新单位；只有当来源、责任与可重走能力改变分配凭证时，单位变化才从概念判断转化为制度发生。

本书没有证明所有教育系统已经进入同样的空转状态，也没有证明来源栏与可重走能够最终解决人机协作中的归属问题。本书交出的是判据与证伪条件，尚未执行经验检验。下一步需要观察的，不是学校何时宣布旧考试过时，而是第一批外部资格何时停止只认分数，并开始为可重走的责任承担付费。届时，教育机器究竟会停止空转、改造账本，还是把新的责任再次翻译成旧的分数，仍有待现实裁决。

第八十六章　容器元数律：记账单位由分配容器决定

出发点：从测量前件看账本

现代学校教育的一切度量与责任装置，建立在「独立个体」这个最小记账单位之上。课程问「学生应知什么」，考题问「学生自己会不会」，成绩单问「学生本人做了什么」。这不是教育理念的偏好，而是记账所必需的本体前提：没有单位，就没有可计量的项。

本书因此不从「怎样测」出发，而从「测出来的数往哪里放」出发。放数的地方是容器，容器有元数。测量装置读出的读数，进入容器前要被容器的元数重新切分。这一切分独立于测量的精度，也独立于生产的形态。

名义定义

学生-AI 耦合场：在某一具体任务中，学生与人工智能系统以双方贡献不可从成品中稳定分解的方式共同完成教育产出的实际生产单位。「耦合场」而非「新主体」，因为每次任务的配比都不同：今天模型只提示一个词，明天模型代写九成，同一学生在不同任务里呈现为不同的学生-AI。

旧账本：以「学生作为独立生产主体」为最小单位，以分数、绩点、通过与否为记录符号，并以这些记录作为资格分配依据的制度装置。

分配容器：接收账本读数并据以分配位置、资格、资源或责任的制度格式—录取名单、执业证书、学位证书、判决书、奖学金账户、招聘名额。

容器元数：一个分配容器的单个格位能够容纳的责任主体数量。单数容器每格收一个名字；复数容器每格可收一个集合（团队、机构、项目组）。

切名：一个耦合场的产出进入单数容器时，被压缩为该容器格位所能容纳的唯一名字的过程。切名不依赖任何人的意图，由容器元数触发。

悬空判断：评估与责任的投放朝向旧单位，而生产已由新单位完成，判断落在一个已不在那里的位置。

学席：单数容器中登记学生-AI 单位的格式—名字＋器具配置＋责任回收读数。名字仍是单数，名字之下的格子从「学生」改写为「席位」。

中间账：位于产出生成与分配容器之间、记录来源构成、生成路径与责任承担的评价层。它不替代分数，而在分数进容器前留下切名之外的痕迹。

操作性读数

本书用三层读数承载上述概念，三层各有单次取值程序，均不依赖学生自报。

来源栏 S（三档类别）。对任一份计分产出：1＝独立完成（未借助模型或仅作格式与检索工具）；2＝人机协作（学生承担可观察的关键生成过程，模型承担辅助、

组织或润色）；3＝模型代做（关键生成过程不可在学生处观察到）。赋值由教师在收件时依据可观察的课堂生成事实与学生当场复述、修改能力做出；读不出的单列「不可归因」，不进入比较。不允许在两档之间插档。

责任回收率 RR（等比）。 在模型产出出现错误（人工设置或自然出现）时，学生能独立发现、说明来源、指出后果并完成修正的比例。仅改对答案不计入。RR 区分「工具熟练」与「承担判断」：一名学生可以熟练调用模型完成诊断建议，却发现不了模型遗漏的危险病史—工具调用成功率高，RR 为零。

容器元数读数 A（二值）。 对任一分配容器：A＝1 若单个格位只能登记一个责任主体的名字；A＝n 若可登记一个集合。取值直接读容器的登记格式（申请表、证书版式、判决书主文、奖项名单），不做推断。

三层读数分工：S 记生产端，RR 记责任端，A 记容器端。容器元数律的经验形态是：**无论 S 与 RR 如何分布，只要 A＝1，进账后的记录只剩一个名字。**

正式形态

Z：记账单位由分配容器的元数决定。当容器为单数（A＝1）时，任何耦合场的产出在进账时被切回一个名字；测量端的精细化（更细的 S、更高信度的 RR）改变的是名字之下能附带多少信息，不改变名字的数量。

它所否定的两个候选：

Y☒：「学生-AI 之所以记不进账本，是因为测量工具还不够细。」（评估脉络）

Y☒：「学生-AI 之所以记不进账本，是因为它本体上是任务敏感的耦合场，不存在跨任务的共同分母。」（分布式认知脉络，也是本书两篇前身稿的表述）

Z 与 Y☒ 的差别最要紧。Y☒ 把不可约简放在生产端，于是逻辑上留下一个出口：若有一天出现稳定的跨任务分解协议，新单位就能入账。Z 把不可约简放在容器端：即使分解协议出现，A＝1 的容器照样只登记一个名字—分解出来的份额会成为名字之下的附注，不会成为第二个名字。Y☒ 对的地方是耦合场确实任务敏感；它错的地方是把这当成了入账失败的原因。

成立条件

第一，存在至少一类分配容器，其格位只登记一个责任主体（A＝1），且该容器接收教育端的读数（分数、学分、通过与否）作为分配依据。

第二，教育现场存在至少一类高频产出，其来源栏在生产端不为 1（S∈2,3 或不可归因）。

第三，该产出在进入容器时，来源信息不被保留为容器格位的一部分—格位上只有名字与一个数。

第四，同一名义读数背后存在生产形态不同的至少两条路径，且路径差异在容器中不可见。

四条同时成立，切名成立，悬空判断成立。

不成立条件

若出现一类分配容器，其格位可登记「人-器具」配置且该配置进入分配后果（例如某执业证书上的器具配置项影响可执业范围与追责范围），而教育端的读数按同一格式登记，则本书关于「切名不可避免」的判断在该容器上不成立—但这恰是本书前文所主张的出路，不是对 Z 的否定，而是 Z 的实现形态：容器元数未变，格位改写。

若出现单数容器直接登记两个名字（学生与模型各一），Z 被推翻。本书预测这不会发生，理由在 4.5。

两句复合测试

Z 能否由两位最近邻的两句话无损复合？取 Bearman 与 Ellis 各自最强的一句：前者说知识与学习已分布于人机系统，个体测量的假设在瓦解；后者说没有来源声明的作品不能进入评价。两句合起来说的是「单位分布化了，所以要有人声明」。它们盖不住的那条缝是：**声明主体与分布主体在制度上被当成同一个人，而这不是谁的疏忽，是容器只有一格。** 两句都没问：贡献声明这个动作应由谁承担、承担之后填到哪里。本书的增量只落在这一处：填不到哪里去，因为格位是单数。

为什么容器不会变成复数

一个自然反驳：把容器改成复数不就行了？让录取名单登记「某生＋某模型配置」，让证书写两个名字。

三处卡死。其一，责任不可分割地落在能承担后果的一方，模型不能被吊销执照、不能被追究刑责、不能被退学；一个不能承担后果的名字写进追责容器，等于没写。其二，分配的稀缺性要求排他：一个录取名额给一个人，名额上多出一个模型的名字，稀缺资源的排他性无处安放。其三，同一个模型配置可以出现在一万个申请人的名字旁边，它作为「第二个名字」没有区分度，在分配上是空的。容器保持单数不是保守，而是分配与追责的功能要求。因此出路不在元数，而在格位。

第八十七章　三位缺席祖宗与八位最近邻：容器元数律的划界

三条脉络都没引的三位祖宗，分别占住了本书命题的三个部件。不把他们请出来，本书的判断就会被误认为自创。

Goodhart（1984）与 Campbell（1979）。 一个量一旦被用作分配依据，就不再是好的测量；行为向最低成本满足该量的方式收敛。本书「假读数」「路径向最便宜的通过方式收敛」的机制，本名在这里。分离线：Goodhart–Campbell 说的是量被分配功能腐蚀，测量对象仍是同一个人；本书说的是测量对象本身在入账时被换掉—腐蚀发生在量上，切名发生在单位上，二者可以独立出现。一份从未用于分配的作业照样会被切回一个名字；一个用于分配的量也可以对应稳定的单位。

Abbott（1988）与 Collins（1979）。 职业系统通过管辖权之争划定谁有资格做什么；文凭社会把学历当作职业准入的通货。本书「准入共同体以险定真值」「学校分数兼具测量功能与分配功能」两句在他们那里有现成的账。分离线：Abbott–Collins 解释资格如何被垄断与兑换，不解释资格证书为何每张只写一个名字；本书要解释的正是这个「一张一名」的几何如何反过来规定了教育端的记账单位。

Clark Chalmers（1998）与 Hutchins（1995）。 延展心智论证认知过程可以延伸到头骨之外的工具与记号；《荒野中的认知》证明导航是舰桥上人与仪器共同完成的计算。本书「人与器具不可分割」「学生-AI 是耦合场」的本体论支撑在此。分离线：延展心智与分布式认知回答的是「认知在哪里发生」，不回答「认知的结果记在谁名下」；本书的增量是：登记不跟随发生—发生可以延展到舰桥全体，登记仍落在船长一人。

以下八位最近邻与三位祖宗各占一个位置。每一栏都写实；判决性反例是在同一案例上两家读数不同的那一点。

Bearman, M. 邻接命题：知识与学习分布在人机系统中，个体测量假设瓦解。分离线：她认为测量对象要重新考虑；本书认为对象怎么考虑都要在入口被切回一个名字。判决性反例：学生用模型完成全部研究设计，课堂答辩能复述不理解。按她：分布式学习一例，应发展新评价单位。按本书：新评价单位造出来也进不了录取名单，问题在名单不在评价。

Ellis, C. 邻接命题：必须记录生成式 AI 的使用方式，否则评价无效。分离线：她主张增加声明；本书主张声明填不到容器里去。判决性反例：文档由六次交互拼接，学生只能回忆一段。按她：需要更细的声明指导。按本书：声明可行性被过程不可召回破坏，即使可召回，声明也只能做名字之下的附注。

Dawson, P. 邻接命题：检测失效，评估设计应转向动态任务。分离线：他认为任务可重设计以恢复真实性；本书认为任务改变被旧土壤翻译成「更难」，且再难的任务也只产出一个进单数容器的数。判决性反例：高认知开放任务仍按标准评分收口，两年内被题库与代训模板接管。

Bretag, T. 邻接命题：诚信研究应从抓作弊转向能力培养。分离线：她修复责任者；本书关心责任者在容器里只有一格。判决性反例：学生训练模型学自己文风生

成初稿再少量修改。按她：不透明协作，需培养诚信。按本书：S 不可归因，账本无栏。

Perkins, M. 邻接命题：把 AI 使用纳入评估矩阵。分离线：他调整变量维度；本书调整的是变量进容器时被切掉的那一步。判决性反例：作业满足含声明的新矩阵，声明中的人机份额不可复现。按他：矩阵成立。按本书：份额是类别贴纸，进容器时被削平。

Cotton, D. R. E. 邻接命题：重设考核任务防止被 AI 绕过。分离线：他假定任务到达个人后由个人承担；本书不假定承担者与人同形。判决性反例：小组项目写作全交模型，每人准备答辩。按他：个别答辩补测。按本书：答辩测复述，不测昨日生成的份额。

Bozkurt, A. 邻接命题：生成式 AI 模糊原创与生成边界，需新伦理框架。分离线：他把边界当伦理问题；本书把边界归因于单位不再是一个受伦理约束的实体。判决性反例：学生说「我们和模型共同完成」，不是说谎，也给不出可登记的贡献。

Selwyn, N. 邻接命题：AI 推动教育治理与数据权力重组。分离线：他观察治理如何重构；本书观察重构之前容器与单位如何互锁。判决性反例：某国强制 AI 进教材，平台大量新增，成绩单版式不变。按他：技术被治理收编。按本书：新单位只能在生产侧增殖，进不了结算侧，因为结算侧版式没动。

Goodhart / Campbell（祖宗一）：量被分配腐蚀 vs 单位被容器切换，见 2.4。**Abbott / Collins**（祖宗二）：资格如何垄断兑换 vs 证书为何一张一名。**Clark–Chalmers / Hutchins**（祖宗三）：认知在哪里发生 vs 结果记在谁名下。

列表摆出来可以看清：没有任何一位把「账本无来源栏」「单位不可约简」「容器每格一名」三件事连成一条命题。他们都踩到了洞的边缘，没有一个把脚落进去。

第八十八章　三条改革路线在容器前的对撞

本节让三条改革路线以最强形式入局，与容器元数律对撞。只交结果。

课堂路线：当堂紧盯

最强形式：当堂限时修改、中途打乱任务、口头追问与自评，足以区分学生会什么不会什么。

对撞结果：这些做法全部假设起点在课堂。学生提前让模型做理解性准备，课堂上的即时修改成为记忆输出，表现平滑不报警。课堂路线优化的是「人在不在场上」，不是「人的生成是否起于此刻」。退一步，即使课堂路线成功读出了人的份额，这份额进入成绩单时仍只有一格。课堂路线解错了题：它解的是测量，题是登记。

声明路线：来源自报

最强形式：每个学生提交时声明哪些部分用了模型、用在哪一步；惩罚不申报，奖励透明。

对撞结果：本书不回答「学生不可信」。本书回答：一旦模型贡献成为产出的组成部分，要声明的不是「用没用」，而是「在生成的哪一步、替换了原应属于人的哪一个生成行为」，这超出单次回忆能复现的范围。更要紧的是，即使声明精确，它填到哪里？录取表上没有那一栏。声明路线让旧单位承担新单位的报告义务，再让报告无处安放。

考题路线：做难做深

最强形式：标准题易被模型接管，转向开放问题、真实项目、跨学科任务，迫使学生动用判断。

对撞结果：题目开放、材料复杂、没有单一答案，培训机构与模型可以先生成项目框架、数据解释与展示脚本，学生按模板执行。开放性本身成为模板化服务的市场机会。题目越「深」，越值得被市场化为新题库。而且再难的题最终仍收口为一个进单数容器的数—考题路线改变的是被测问题，不是生产路径，更不是容器。

分布式账本路线：测系统不测个人

最强形式：既然单位已分布，就让账本也分布—建立学生-AI 作品集，记录每轮交互，录取看作品集。

对撞结果：结算处卡死。作品集里的失败谁负责？分数挂谁？奖学金、排名、毕业证、追责都是单数容器。一个学生带着系统作品集去申请只收个人名字的资格，新

账本签发不出主体资格。系统不是人，不能作为申请人。分布式账本一接触分配，就在「谁被录取」这个最简单的问题上卡死—不是理论不完整，是容器是单数。

四条路线的反例指向同一位置：任何在旧流程内改进人或方法的路径，最终都撞上「找不到与生产单位同构的登记格位」。

两轴的确立与候选轴的排除

第一轴取自本书核心：**来源可见性**—账本是否拥有可用且不依赖自报的路径信息。

第二轴候选有四。「任务难度」是来源不可见的压迫因素，不是结构轴，排除。「学科类型」影响人机贡献如何分布，是表现不是结构，排除。「学生动机」偏个体内部状态，不可观察，排除。「账外分发的硬度」（职业准入是否已介入）有吸引力，但它是容器的一种强度，而容器的结构变量是元数：外硬容器与外泡容器都可以是单数。故第二轴取**容器元数** $A \in 1, n$。

两轴相关但不重合：来源可见与否不改变容器元数，容器元数不改变来源是否可见；前者决定账本能读到什么，后者决定读到的东西进账时被切成什么。

四格

A 格：来源不可见 × A＝1—隐蔽切名。 终稿、分数、资格读数都在，理解、构思、生成如何发生无人知道，进账时切回一个名字。遍及大量声誉老、排名高、毕业去向与专业能力已脱钩的普通文理专业。独有预测：校内高分与责任回收率不相关甚至负相关；阅读量无显著增加而作业平均分逐年微涨—不是学生更聪明，是可外包的路径越来越顺。

B 格：来源可见 × A＝1—可追踪切名。 学校记录了学生何时调用模型、用了哪些输出、做了哪些修改，进账时仍切回一个名字。常见于已推行过程性评价、日志与版本历史的院系。独有预测：来源记录逐年加厚，学生独立解释能力不随记录改善；制度获得了过程数据，没有获得责任主体。这一格是过程性评价改革的终点，也是它的天花板。

C 格：来源不可见 × A＝n—集体成品。 团队作业、项目组交付、机构署名，容器收一个集合，个体来源不可见。独有预测：集体成品质量上升，个体退出无痕，搭便车不可识别；集合内部的切名从容器挪到了团队内部。

D 格：来源可见 × A＝n—责任型协作。 两轴皆高的实例：手术团队的手术记录登记主刀、一助、器械与麻醉，每个名字下有器具与职责；庭审记录登记律师、书记员与辅助系统的使用，律师对事实主张与法律立场当庭负责。独有预测：模型参与程度与个体责任能力不再简单负相关，可能共同提高，但依赖错误暴露—解释—修正机制。

四格不可互相替代。A 与 B 的差别是能否看见路径；B 与 D 的差别是容器元数；C 与 D 的差别是集合内部是否可见。教育端的分配容器几乎全在 A＝1 一列，所以教育变革若在两轴上只能动一轴，动来源可见性只能从 A 格走到 B 格—切名照旧。

第八十九章　型别等级：单数容器登记人—器具单位的历史模板

两个先例

机动车驾驶证是单数容器：一证一名。但证上有「准驾车型」一栏—C1、B2、A1—此人被批准在何种器具配置下作业。器具配置进入了分配后果（能开什么车）与追责后果（超出准驾车型驾驶，责任加重）。容器元数没有变，名字之下的格子被改写了。

民航飞行员执照同样一证一名。执照之下有「型别等级」（type rating）：此人被批准操纵何种机型。型别等级要单独训练、单独考核、定期复训；换机型要重新签注。执照没有写第二个名字，写的是「此人＋此器具」的组合被批准的范围。

两个先例共有的登记格式是：**名字（单数）＋器具配置（可多项）＋该配置下的作业范围与责任范围**。这就是「席位配置」。

为什么这个模板能解本书的题

容器元数律说：单数容器只登记一个名字。型别等级说：一个名字之下可以登记器具配置。两句合起来：**人-器具单位不需要第二个名字，它需要名字之下的第二格**。学生-AI 进账时被切回「学生」一名，是因为「学生」这一格下面没有器具配置栏；加上那一栏，切名仍然发生（名字只有一个），但被切掉的不再是器具，而只是器具之外的耦合细节—而这正是可以接受的损失，因为责任本来就要落在能承担后果的一方。

这也解释了 4.5 中容器不会变成复数的三处卡死为何在型别等级上都不成问题：责任仍落在飞行员一人；名额仍排他；机型本身不是申请人，它是名字之下的限定词。

学席的登记规格

据此，学席不是一个新名字替换「学生」，而是「学生」这一格的改写：

- **席名**：单数，仍是学生本人；
- **席配置**：此学生在本次评价中被批准的器具配置（无模型／指定模型与用途／自选模型），对应来源栏 S 的三档；
- **席读数**：责任回收率 RR 与配置绑定—同一学生在「无模型」配置下的 RR 与在「自选模型」配置下的 RR 分开登记，不折算为一个数；
- **席范围**：该配置下已被检验的作业范围，作为向下游容器（录取、准入）传递的内容。

教席同理：教师之名单数，之下登记批改与命题所用的器具配置及其责任回收读数—模型生成评语、教师仅确认提交的，登记为一种配置；教师保留批改证据并对错

误评语负责的，登记为另一种。新单位先在账房一侧合法出生（教师-AI 提高结算效率不威胁分配秩序）这一不对称，在席配置栏上会第一次变得可见。

谁会先做

型别等级出现在航空而非小学，不是偶然：器具配置差异直接对应事故风险，准入共同体以「险」定真值。同理，席配置最可能先从公共风险高、资格与执业许可紧密相连的领域（临床、工程签字、法律执业、航空与核设施操作）的准入容器开始，倒逼上游教育端的成绩单改版式；文凭地带的普通文理专业最后跟进。这一顺序是本书的可检验预测（第七条第五款）。

理论意涵

第一，教育评估不是测量准确性问题，也不是诚信文化问题，而是「单位能否作为容器前提」的问题。讨论「如何考」之前先问「考什么单位」，问「考什么单位」之前先问「单位由什么决定」。本书的答案是：由容器决定。

第二，测量精细化与单位入账是两道独立的题。这一分离把过程性评价、贡献声明、深度考题三条路线放到了同一个位置上：它们都在解测量题，而卡住教育的是登记题。这不否定它们的价值—B 格比 A 格有更多路径信息—它否定的是它们对「单位危机」的承诺。

第三，新单位入账不需要新的考核理论，需要的是格位改写；格位改写在许可制度中已有百年先例。教育不是在发明一件新事，是在追认一件别的行业早已做完的事。

实践意涵

第一步不是新课、不是新题、不是过程日志，而是在成绩单与课程记录的「学生」一栏之下加一格「席配置」，把来源栏三档挂在配置上。哪怕三档粗糙，也把「无分解」从暗账变成明账。

第二，RR 按配置分开登记，不折算。折算即切名的重演。

第三，资格容器的改版由高风险准入共同体先做，教育端跟进；教育端单独改版式而下游不认，会退回 B 格。

适用边界与反向约束

其一，任务本身对来源不敏感的领域（肌肉训练、体育技能、器乐、手工技术）不进入本书的单位困境；这条削掉「教育系统全部陷入单位危机」的言外之意。其二，外部不分配资格的场所（兴趣班、公益工作坊、成人进修）假读数再普遍也无严

重后果；本书的严格立场是「凡以分数分配人的位置的教育场才进入危机」。其三，模型仅承担格式转换、语法校对、机械计算且不改变理解与判断的场景，来源栏为1，席配置栏为空，本书命题不触发。其四，风险后果弱、责任可逆、市场不依赖正式资格的行业，雇主可通过作品集与试做识别能力，准入容器未必成为改版入口。

反噬

按本书建议走，最省事的做法是给每份作业加一个「AI 使用申报表」，勾选独立、协作、外包，作为新的合规指标。这会毁掉本书想保住的东西：学生把复杂过程压成对自己有利的类别；学校把席配置变成纪律工具，勾外包即扣分而不问模型究竟承担了什么；教师把填表数量当过程性评价。席配置由此不是打开路径，而是新增一张表，中间账成为旧账本的附属票据。

本书能给的防御只有两条。其一，席配置与 RR 必须由人在生成现场读出，不能由机器事后产出，也不能由学生事后自述—流程读出，不是表格读出。其二，席配置必须允许「不可归因」，必须禁止把模型参与程度直接换算为扣分；否则制度会再次把耦合场压回一个数，只是换了一个更现代的数。

第九十章　容器元数律的判别式、证伪条款与解释不了的洞

判别式

在不依赖受试者自报、不引入新评价形式的前提下，对同一批产出试加 S 栏，同时读取这批产出最终进入的分配容器的 A 值：

若 S∈2,3 的产出占比不低（预设阈值下文），且 A＝1，且容器格位上不出现 S 信息，则切名成立，本书命题在该场景上成立。

若 S∈2,3 占比高但容器格位登记了 S 或等价的器具配置信息并使之进入分配后果，则该场景已经完成前文所说的格位改写，切名在此被抵消。

若 S∈2,3 占比低于阈值，本书命题在该场景上无用武之地—不是被否，而是没有耦合场需要进账。

判别式不含「应当」「合理」「恰当」。它读三个事实：来源栏分布、容器元数、格位是否登记来源。

阈值

S∈2,3 占比的预设阈值为30%。这不是自然常数，是为提出可复核判据设的起点，须经敏感性分析。RR 与最终成绩之间的偏离以「成绩位于前 40% 而 RR 位于后 40%」定义为一次可重复的倒挂；同一学生在三次以上任务中出现倒挂即记为稳定倒挂。

三条自否决条款

信度条款。 两名独立观察者对同一批产出逐份赋 S 值，Cohen’s Kappa 低于 0.70，读数不稳，不得作为制度依据；RR 在两周内、难度相近的重测中组内相关低于 0.70，同样不得用于高风险决策。

改名条款。 若 S 栏在控制任务难度后可被「在场时间」或「文本相似度」完全替代（相关系数 ≥0.80 且增量解释率 <0.05），它只是已有量的改名，必须放弃；RR 若与「作业完成度」「学习投入度」高度重合到同一标准，同样放弃。

自报条款。 若某次 S 或 RR 赋值必须事后问学生「这是不是你写的」，或必须读取模型检测软件，该读数已退化为他报与机器猜测，不可采用。概率与倾向只可用于群体估计，不得倒灌为单个学生单次任务的事实判定。

四格各自独有的预测（先于前文给出，便于对照）

来源不可见 × A＝1：校内高分与责任回收率不相关或负相关。

来源可见 × A＝1：过程记录逐年加厚，学生独立解释能力不随记录改善—制度拿到了过程数据，没拿到责任主体。

来源不可见 × A＝n：集体成品质量上升，个体退出无痕，团队内部搭便车不可识别。

来源可见 × A＝n：模型参与程度与个体责任能力不再简单负相关，可能在复杂任务中共同提高，但依赖错误暴露—解释—修正机制。

以下条款凡标 [未执行] 者，本书均无权声称已证实。它们的价值是划出本书一旦遇到就自认失败的边界。

第一条 [未执行]：若五年内至少三个国家或地区的官方学历或执业资格制度在证书格位上直接登记两个责任主体名字（学生与模型各一），并使两个名字各自承担分配与追责后果，容器元数律被推翻。现有解释预测不出这件事不会发生；本书预测它不会发生，理由在4.5。

第二条 [未执行]：若某一分配容器保持 A＝1，且未在格位上登记任何器具配置信息，却在连续三届中使 S∈2,3 的产出与 S＝1 的产出在分配后果上稳定分开，则切名可以被测量端单独抵消，本书关于「测量与入账独立」的判断被削弱。

第三条 [未执行]：取同一批学生同一内容，一班课堂限时生成加当堂追问，一班回家作业加事后检测；若三个月后迁移成绩无差异，本书关于「产出时点」止损作用的判断被否。这一条只涉及测量端，不涉及容器端。

第四条 [未执行]：若某类模型能一并生成「人的错误模式」，使当堂追问也无法区分人与模型的理解，则 RR 读数的取值程序失效，本书的第二层读数须撤回。

第五条 [未执行]：若企业招聘、职业准入与学校考试长期维持相同的分数信任结构，即使执业出现可观察风险，资格制度也不转向「格位改写」，则本书关于「格位改写将先从高风险准入容器开始」的判断不成立。

替代解释。其一，来源不可分解只是教师没时间没训练；给足资源即可分解。可削弱本书的测量层，不触及容器层：分解出来的份额仍要面对 A＝1。其二，旧账本不退是既得利益，不是几何。若多数教师在无利益受损且可匿名表态时仍选择保留单数格位，利益不是主因；本书预测他们会保留，理由是4.5 的三处功能卡死。其三，切名只是暂时的技术状态。若出现对所有任务可复现交互过程且可分解贡献的协议，Y⊠ 的最强表述被否，Z 不受影响。

构念效度。「切名」可能被理解为「测量粗糙」的换名。防线在于两者可独立出现：测量任意精细，A＝1 照样切名；测量粗糙，A＝n 也不切名。但本书尚无经验材料同时变动两轴。质性对应问题是可信性：读者是否相信「容器元数」是独立机制而非重新命名旧现象。后续可由教育测量研究者与制度研究者共建编码本，对同一批产出同时读 S、RR、A，两个学期得初步结果。

内部效度。本书无法排除：新单位不能入账，不是因为容器几何，而是因为制度没有意愿支付改版成本。若是后者，「切名」多一层政治经济学意味。本书保留这一

空间，尚无材料区分「不能」与「不愿」—型别等级先例说明「能」，但不说明教育端「愿」。

外部效度。最适用于高风险、强资格、以书面成绩分配机会的场景；幼儿教育、非正式学习、导师制环境须降低力度。不同学科对「关键判断」要求不同，RR 不能共用一套错误设置。

可靠性。S 与 RR 的赋值受教师判断影响；第六章的自否决条款是门槛而非结果。

本书解释不了的两个洞。其一，若未来模型能内化并预测某个使用者独特的学习历史，使代做与自做的区分被模型吸收，来源栏的整套思路失去根基；席配置栏仍可登记，但登记的内容会变成空的。其二，在长期、低风险、兴趣驱动的学习中，行动者可能主动增加困难、保留不确定性、追问模型错误，而非沿最低成本路径行动；本书的切名机制不解释为何有人拒绝被切—这里可能存在一种本书未说明的自我界面机制。

卷六　教授端的悬空：签字的人为何证明不了自己的资格

本卷写教授端。学生的悬空是被评者的悬空，教授的悬空是评者的悬空：教授仍被要求签署、解释、纠错，产出却由教授、AI 与情境共同生成，五类记录——署名、调用日志、错误追责、同行评价、岗位考核——不再指向同一个人。本卷给出悬空资格的定义、读数、九位最近邻的划界、八种竞争解释的逐一对撞、二维四格与一个写死到 2028 年的制度赌注。它是教席的负片：教席要登记的正是这五类记录的收敛。

第九十一章　从一篇由人机共同生成的论文开始

从一篇由人机共同生成的论文开始

一名学生提交一篇课程论文。题目由学生提出，资料由搜索系统召回，论证由生成式模型组织，文字由软件润色，最后由学生删改并署名。教授面对的文本比过去更完整、更流畅，甚至更容易达到课程要求；然而，文本已经不再稳定地指向学生的阅读、推理与判断。教授批改的对象表面上仍是一篇论文，实际上却是一条由学生、模型、数据库、课程规则、教师提示和提交情境共同构成的生产链。

这一变化首先改变了评价对象的性质。过去，教授可以把作品作为学生能力的间接证据：文章如何提出问题、组织证据、处理反例，大体能够反映学生是否经历了相应的思考。如今，同一结果可能由不同组合生成。学生可以不掌握全部知识，却借助工具完成复杂任务；教师可以使用模型设计课程、反馈作业和组织科研；大学可以把 AI 嵌入教学、科研、创业与行政系统。产出仍然存在，但产出与某一主体的能力之间不再保持稳定的一对一对应。

教授由此遭遇的并非单纯的工作效率问题。即使教授职位没有消失，即使课程数量、科研项目和管理任务继续增加，教授仍可能怀疑：自己是否还能够作为知识的最终裁判者存在。这里的「裁判」不是掌握更多信息，也不是拥有更高职位，而是在证据不完整、意见相冲突、后果不可逆的情况下，能够提出标准、作出选择、解释理由并承担判断后果。

现有解释失效在哪里

现有解释至少提供了三条有力路径。学历信号理论能够说明，学历之所以具有效力，是因为它向市场传递了有关能力的信息；当知识获取成本下降、复杂任务可以由工具辅助完成时，学历所代表的知识稀缺性可能减弱。职业管辖权理论能够说明，教授的权威并非单纯来自个人声望，而来自对某类知识问题的解释、评价与处置权。认知外包研究则能够说明，人会把记忆、计算、检索乃至判断活动分配给外部工具。

但这些解释在同一个 AI 教学场景上分别失效于不同句子。信号理论可以说「学历的预测力下降」，却不能回答「教授资格是否仍是一个能够独立记录的对象」。职业管辖权理论可以说「平台或行政系统夺取了评分权」，却不能回答「即使教授仍承担最终责任，为什么产出不能证明其资格」。认知外包研究可以说「判断过程分布到工具中」，却不能回答「制度如何保存教授个人在混合判断中的资格」。

如果把教授焦虑写成「怕被 AI 替代」，就把岗位、工作任务、社会承认、判断资格和认知能力压成了一个词。如果写成「教授要从讲授者转型为系统设计者」，则只回答教授需要做什么，没有回答为什么转型本身会制造焦虑。若写成「AI 削弱了学历权威」，则只能说明旧信号的衰减，不能说明资格怎样重新生成。

本书的承重命题

本书主张：AI 时代大学教授的焦虑，不是权威下降，不是职业管辖权被夺，而是悬空资格——教授仍被要求承担最终裁判责任，但知识产出已经由教授、AI 与具体情境共同生成，教授资格却失去可独立记载的载体。这不是「权威下降」的强化表达。权威下降意味着一个既有对象的社会承认减少；悬空资格意味着这个对象赖以被识别、传递和证明的承载关系发生变化。它也不是「无人负责」。制度可以继续指定教授、学院或平台承担责任；责任的指定并不等于教授的知识资格已经从混合产出中被独立证明。

本书的目标不是预测教授是否会失业，而是追问教授焦虑为何在职位仍然存在、产出甚至增加的情况下继续加深。本书把焦虑理解为一种发生关系：当产出由人、AI 与情境共同生成，而制度仍要求某位教授承担最终裁判责任时，过去可以由学历、职称、署名和个人作品承载的资格不再自动成立。资格因而处于「悬空」状态。

第九十二章　三条解释在哪一句失效：信号、管辖权、认知外包

学历信号：它解释了权威如何被看见

学历信号理论的核心贡献，在于把教育证书从「能力本身」区分为「关于能力的信息」。Spence（1973）说明，在劳动力市场中，教育经历可以作为求职者向雇主发送的信号。其解释变量是学历与潜在能力之间的区分度：当学历能够帮助市场区分不同能力水平的个体时，学历具有信号价值。

这一脉络能够解释 AI 为何可能冲击大学权威。若复杂知识任务不再要求个体独自占有大量知识，若工具使原先具有专业门槛的任务变得普遍可完成，那么学历所代表的稀缺知识便不再自动构成能力证明。有关 AI 降低知识获取成本、削弱学历所代表的知识垄断的讨论，正是沿着这条路径展开的（迈克尔·莱维特 2025）。

但该理论在教授焦虑场景中于一句关键判断处失效：学历信号减弱，并不等于教授资格失去独立载体。即使学历不再能准确预测知识水平，教授仍可能通过事前判断、过程记录、同行复核和责任承担证明其资格。信号理论握住的是「字段预测能力」，没有握住「资格对象是否仍能被记录」。

职业管辖权：它解释了谁有资格判断

职业社会学把专业权威放入任务边界与判断权的竞争之中。Abbott（1988）所讨论的职业管辖权，关注职业如何取得对问题的定义、诊断、推论和处置权。大学教授的专业地位不仅来自知道某些内容，也来自有权决定什么算作问题、何种证据足以支持结论、何种回答能够通过评价。

这一脉络能够解释 AI 平台、行政管理系统与教师之间的冲突。自动评分系统可能改变谁制定评分标准，平台可能获得课程反馈与学习分析的控制权，学院可能通过统一模型和流程重新配置教师的判断空间。它握住的是「谁有权作出判断」这一解释变量。

但职业管辖权理论在本场景上于另一句判断处失效：名义上的最终裁判者仍然可能是教授，但教授资格却无法从混合产出中独立证明。管辖权理论关注控制权的转移；悬空资格关注资格的承载。教授可能仍拥有形式上的签字权，却只能对一个由模型、数据、课程政策和学生情境共同形成的结果负责。权力没有完全转移，资格却已经分叉。

认知外包与分布式认知：它解释了判断如何被共同生成

Sparrow、Liu 与 Wegner 关于互联网记忆的研究，说明人们可能记住信息存放在哪里，而不必记住信息本身。Risko 与 Gilbert（2016）把这一问题推进为认知外包：个体借助外部工具承担记忆、计算、检索和部分推理功能，以降低内部认知负

荷。分布式认知的研究传统则进一步强调，认知活动可以分布在人、工具、符号、制度和环境组成的系统中。

这条脉络对 AI 时代的解释力很强。学生论文、教师评分和科研文本都可能不再是某一个人的内部能力的直接外化，而是由人机系统共同生成的结果。它握住的是「判断如何发生、由谁执行、在哪些外部条件中完成」这一解释变量。

但它在教授焦虑问题上仍少了一步：共同生成不自动说明个人资格如何被制度保存。分布式认知可以说系统整体完成了判断，却不回答大学在聘任、晋升、申诉和事故追责时如何证明某位教授具有独立裁判资格。认知过程可以分布，资格是否能够独立记载仍是另一问题。

三条脉络分别握住信号、管辖权与认知分布，却共同留下一个未经检验的前提：教授资格已经存在，并且可以被学历、职称、署名或职位稳定指认。本书的缺口正位于此处：当产出不再稳定指向单一主体时，教授资格是否仍然是一个可以从产出中独立读出的对象？

第九十三章　悬空资格的定义与四件读数

理论出发点：从属性转向发生关系

本书站在发生学立场上。对象的发生由三个层次共同规定：它如何显露、沿哪条路径形成、在什么土壤中纠缠。

本书采用这一传统，而不采用把资格视为预先存在属性的替代方案，原因在于教授资格的变化并非仅仅是某个固定指标的升降。AI 介入后，变化同时发生在三个层面：第一，社会如何显露并承认教授；

第二，教授工作如何沿着新的差异路径展开；第三，教授、学生、模型、制度与情境如何纠缠在共同产出中。若只把焦虑视为「权威减少」，就会把发生过程中的关系重新冻结成属性。

本书把发生学理解为追问事物为何如此发生，而不是追问研究者如何发现它。这里的对象不是「教授焦虑的测量结果」，而是「为什么一种特定的焦虑在职位仍存在时发生」。悬空资格不是等待观察者命名的心理状态，而是由产出方式、责任安排和资格记录之间的差异关系生成的结构状态。

核心概念的名义定义

悬空资格是：在混合产出条件下，制度要求某主体承担知识裁判责任，而该主体的资格不能由独立记录稳定承载的状态。这个定义不把悬空资格等同于焦虑强度、权威下降、任务增加或责任消失。它只规定三个关系：产出由多方共同生成；责任仍被赋予教授；资格不能由独立记录稳定承载。

操作性定义与四件读数

经验层面的悬空资格表现为：同一项由教授、AI 与具体情境共同生成的判断进入正式制度记录后，署名、调用日志、错误追责、同行评价和岗位考核无法稳定指向同一位教授的独立裁判资格。

其核心读数可以写为：

$$Q\_s = H \times (1 - R)$$

其中， H 表示产出混合程度， R 表示资格记录的收敛程度， Q_s 表示悬空资格读数。该式中的「程度」并非单次事件中的心理概率，而是针对一组可比较事件的比例指标。

公式。H 可由一项判断中实际参与生成的主体与外部条件数量构成； R 可由五类正式记录是否指向同一资格主体计算。可进一步写为：

$$H = (a+m+c) \div (a+m+c+1)$$

其中， a 表示 AI 系统及其版本参与数， m 表示人类参与者数量， c 表示被制度化记录的关键情境条件数量。这里的「1」代表教授本身。该式只用于形成相对排序，不把主体数量解释为因果强度。

R = 五类记录中指向同一教授的记录数 ÷ 5

值域。H 的值域为[0,1)， R 的值域为[0,1]， Q_s 的值域为[0,1)。 H 越高，表示产出参与者越多； R 越低，表示资格记录越不收敛。单次事件中， R 只能取0、0.2、0.4、0.6、0.8 或1。

测量层次。五类记录的归属首先是类别测量；把五类记录相加形成的收敛比例属于等距层次的编码指数；同一教授在不同事件中的 R 排序只表示记录收敛程度的相对差异，不直接表示资格本体的倍数关系。

H 是基于计数的比例指标，可作等比读数使用，但不能据此宣称某事件的混合程度是另一事件的两倍。

一次可观测事件中的取值。一次事件可以是一项 AI 辅助评分、一次论文修改、一项科研决策或一次答辩判断。研究者在事件发生后分别读取署名、调用日志、错误追责、同行评价和岗位考核材料：若五类材料全部指向教授，则该事件的 R=1；若只有三类指向教授，则 R=0.6。同时编码实际参与产出的教授、学生、模型及关键情境条件，计算 H。单一事件的读数不能直接判定整个制度是否产生悬空资格，必须在预先界定的一组事件中比较。

第九十四章　命题：教授焦虑不是权威下降，而是悬空资格

命题形式

本书的承重判断是：

教授焦虑不是学历信号贬值，也不是职业管辖权转移，而是悬空资格。更完整地说，教授焦虑不是 Y⊠——学历所代表的知识信号变弱；也不是 Y⊠——教授对知识任务的管辖权被 AI、平台或行政组织夺取；而是 Z——教授仍承担最终裁判责任，但混合产出不能再独立证明教授资格，资格因此失去稳定载体。

这一命题在以下条件下成立：

第一，相关知识产出由教授、AI 与具体情境共同生成，而非由教授独立完成。

第二，制度仍要求教授承担最终评价、解释、签署或纠错责任。

第三，正式记录至少在署名、调用日志、错误追责、同行评价和岗位考核之间出现归属分叉。

第四，现有固定资格字段无法在不改变资格意义的条件下预测教授在新案例中的独立裁判。

第五，记录分叉不能完全由单纯的岗位变更、行政责任转移或模型错误解释。

命题在以下条件下不成立：若教授虽然使用 AI，但其事前标准、拒绝条件、修改理由和责任链能够稳定预测其在新案例中的判断，并且五类制度记录仍稳定指向同一资格，则资格已经被重新承载。此时即使学历信号减弱、工作流程改变或认知过程发生外包，也不能称为悬空资格。

两句复合测试

若把斯宾塞关于学历作为能力信号的判断与 Abbott 关于职业管辖权的判断直接拼接，可以得到：「AI 削弱学历信号，并使教授对知识判断的管辖权受到挑战。」这两句能够解释资格危机的外部表现，却不能无损重述本书命题。本书的增量不在于再加上「AI 参与产出」，而在于指出：责任仍然可以归给教授，资格却不能从混合产出中被独立记载和预测。因此，本书并不声称发现了「教授权威正在下降」这一事实；本书提出的是对焦虑对象的重新规定：焦虑发生在资格的承载方式改变之处。

第九十五章　九位最近邻的划界

本节没有经过站外检索。以下盘点依据已有材料与通行概念进行初步划界，相关占位者和原题均标为未核验；因此本节不宣称完成文献穷尽，也不以未核验材料证明原创性。

占位者	他已经占了什么	分离线	判决性反例
Michael Spence〔未核验〕	学历是劳动力市场中向雇主传递能力信息的信号	他的判断关于学历字段能否预测能力；本书关于教授资格能否作为对象被固定记录	同样是 AI 辅助论文，学历预测力下降但教授事前标准仍能预测新案例：他判为信号贬值，本书判为非悬空资格
Andrew Abbott〔未核验〕	职业通过对任务的诊断、推论和处置取得管辖权	他的判断关于谁控制判断任务；本书关于责任仍在教授时资格是否有独立载体	平台获得评分流程控制权但教授仍承担最终责任：他判为管辖权转移，本书判为可能悬空
Sparrow、Liu 与 Wegner〔未核验〕	人可以记住信息所在位置，而不必记住信息本身	他们的判断关于信息记忆是否外部化；本书关于外部化后的资格是否可记录	教授大量调用 AI 但保留稳定的事前标准与修改轨迹：他们判为认知外包，本书判为非悬空资格
Risko 与 Gilbert〔未核验〕	认知外包可降低内部负荷并提高任务效率	他们的判断关于工具是否承担认知活动；本书关于产出能否证明教授资格	处理速度提高而教授资格记录稳定：他们判为外包有效，本书判为非悬空资格
责任鸿沟（responsibility gap）〔未核验〕	人机系统造成的后果可能难以找到一个足以承担责任的主体	它的判断关于责任主体能否被指定；本书关于责任已经指定时资格能否被证明	学校明确追究教授责任但无法还原其独立判断：责任鸿沟判为非典型，本书判为悬空可能成立
分布式认知（distributed cognition）〔未核验〕	认知活动分布在人、工具、符号、制度和环境组成的系统中	它的判断关于系统整体如何认知；本书关于系统产出是否能指回教授资格	系统整体准确率提高但教授在陌生案例中的判断不可预测：分布式认知判为系统增强，本书判为悬空
自动化偏误（automation bias）〔未核验〕	人可能过度依赖自动化建议并减少独立核验	它的判断关于教授是否错误采纳 AI；本书关于资格是否有承载，即使教授判断正确也适用	教授拒绝 AI 建议且错误率下降，但没有稳定记录独立标准：自动化偏误减弱，本书仍判为可能悬空
过程追踪（process tracing）〔未核验〕	通过时间戳、版本、输入条件和理由重建结果形成过程	它的判断关于一次判断如何发生；本书关于过程档案能否在新案例中预测同一资格	一次过程能够完整还原，但连续新案例无法预测教授裁判：过程追踪成功，本书判为资格仍悬空
Daniel Wegner〔未核验〕	意识中的「我决定了」可能是行动发生后的解释	他的判断关于意识意志是否产生行动；本书不否定判断发生，只检验资格能否被制度保存	教授的主观报告不可靠，但事前标准与修改轨迹稳定：其判断可被怀疑，本书判为非悬空资格

Andrew Abbott（〔未职业通过对任务的他的判断关于谁控制判断任平台获得评分流程控制权但教授核验〕）诊断、推论和处置务；本书关于责任仍在教授仍承担最终责任：他判为管辖权取得管辖权。时资格是否有独立载体。转移，本书判为可能悬空。

Sparrow、Liu 与人可以记住信息所他们的判断关于信息记忆是教授大量调用 AI 但保留稳定的事 Wegner（〔未核验〕）在位置，而不必记否外部化；本书关于外部化前标准和修改轨迹：他们判为认住信息本身。后的资格是否可记录。知外包，本书判为非悬空资格。

Risko 与 Gilbert（〔未认知外包可以降低他们的判断关于工具是否承处理速度提高而教授资格记录稳核验〕）内部负荷并提高任担认知活动；本书关于产出定：他们判为外包有效，本书判务效率。能否证明教授资格。为非悬空资格。

responsibility gap 人机系统造成的后它的判断关于责任主体能否学校明确追究教授责任但无法还（〔未核验〕）果可能难以找到一被指定；本书关于责任已经原其独立判断：责任鸿沟判为非个足以承担责任的指定时资格能否被证明。典型，本书判为悬空可能成立。主体。

distributed 认知活动分布在它的判断关于系统整体如何系统整体准确率提高但教授在陌 cognition（〔未核人、工具、符号、认知；本书关于系统产出是生案例中的判断不可预测：分布验〕）制度和环境组成的否能指回教授资格。式认知判为系统增强，本书判为系统中。悬空。

automation bias 人可能过度依赖自它的判断关于教授是否错误教授拒绝 AI 建议且错误率下降，（〔未核验〕）动化建议并减少独采纳 AI；本书关于资格是否但没有稳定记录独立标准：自动立核验。有承载，即使教授判断正确化偏误减弱，本书仍判为可能悬也适用。空。

process tracing 通过时间戳、版它的判断关于一次判断如何一次过程能够完整还原，但连续（〔未核验〕）本、输入条件和理发生；本书关于过程档案能新案例无法预测教授裁判：过程由重建结果形成过否在新案例中预测同一资追踪成功，本书判为资格仍悬程。格。空。

Daniel Wegner（〔未意识中的「我决定他的判断关于意识意志是否教授的主观报告不可靠，但事前核验〕）了」可能是行动发产生行动；本书不否定判断标准与修改轨迹稳定：其判断可生后的解释。发生，只检验资格能否被制被怀疑，本书判为非悬空资格。度保存。

表中最重要的分离不是研究对象不同，而是判断本身不同。占位者分别处理信号、管辖权、认知过程、责任、系统认知、自动化依赖和过程还原；本书处理的是：当责任仍然被赋予教授时，教授资格是否还能脱离混合过程而被稳定证明。

第九十六章　五类记录的收敛：判别式、五项指标与自否决

判别式

本书的判别式是：

同一项由教授、AI 与具体情境共同生成的判断，进入成果署名、调用日志、错误追责、同行评价和岗位考核记录后，五类记录是否指向同一位教授的独立裁判资格。这一判据不询问教授「是否感到焦虑」，也不询问研究者「是否觉得权威下降」，而是询问流程、日志和档案。若五类记录稳定收敛于同一位教授，悬空资格不成立；若教授仍承担最终责任，但五类记录分别指向教授、模型、平台、学院和课程组，悬空资格获得支持。

这一判据在不同场景中会产生不同的可裁决结果。

在 AI 辅助论文写作中，作者栏可能记教授，生成日志记模型，错误责任记学生，课程评价记教师团队，晋升考核记论文署名。若五类记录分叉，产出与资格不再互相证明。

在 AI 自动评分课程中，成绩系统可能显示教师姓名，模型日志显示评分版本，申诉记录显示平台责任，教学评价显示课程组，续聘材料显示教师个人。此时教师仍可能拥有最终签字权，却无法从成绩结果中单独证明其判断。

在 AI 辅助科研中，论文署名、代码记录、数据处理日志、结题评价和职称材料可能分属不同主体。科研产出越多，若独立判断过程越少，资格证据可能越稀薄。

在 AI 答疑平台中，学生看到的是教师课程的回答，平台保存的是模型调用与检索记录，投诉处理由学院完成，教师评价依据响应速度。最终答案归属可能持续分裂。

在 AI 参与学位答辩中，答辩纪要、导师意见、模型辅助材料、学术委员会决定和岗位考核可能形成五种不同的归属方式。教授承担答辩责任，不意味着其个人判断已经被完整记录。

这五个场景不能由「AI 参与产出」这一事实直接推出结果。必须读取真实记录，才能判断五类归属是否分叉。

可观测指标

指标一是记录收敛率。对每一项混合判断记录五类材料，计算指向同一教授的记录数除以五。R=1 表示完全收敛， R=0 表示五类记录没有共同归属。研究者应在事件开始前规定「同一资格」

的识别标准，不能在看到结果后临时合并不同主体。

指标二是独立裁判预测率。在教授看到 AI 输出前，要求教授提出判断标准、初步结论或拒绝条件；随后给予新案例，观察其既有标准能否预测新的裁判。预测率不是教授意见与最终答案的一致率，而是教授能否在陌生案例中保持可解释、可迁移的判断规则。

指标三是归属分叉指数。对五类记录进行两两比较，统计不指向同一主体的记录对数量。若五类记录全部一致，分叉指数为0；若每一类记录都指向不同主体，分叉指数达到最高。该指标用来区分「责任集中但资格稳定」和「责任集中但记录分叉」。

指标四是混合产出指数。记录模型版本、提示输入、数据源、学生或教师参与、课程规则与情境限制。混合产出指数不是模型调用次数越多就越高，而是实际参与最终判断形成的主体与条件越多，指数越高。

指标五是资格迁移率。将教授事前标准、修改理由和责任记录输入盲测设计，要求评审预测教授在新案例中的裁判。若关系档案只能解释旧案例，不能预测新案例，资格仍缺少可迁移载体。

读数的三条自我否决条款

信度条款。同一批事件由两名独立编码者间隔两周重复编码；若五类归属编码的一致率低于0.80，读数判定不稳，不得据此宣布资格悬空。对连续案例的资格迁移率，应至少进行两轮盲测；若同一教授两轮预测差异超过预先设定的0.20，关系档案的预测读数不稳定。

改名嫌疑条款。悬空资格读数必须与学历信号强度、AI 使用频率、工作量、责任鸿沟指数和自动化偏误分数进行区分。若记录收敛率只是既有「责任清晰度」或「流程可追溯性」的换名，且不能解释新的资格预测问题，本书读数作废。

单次事件条款。任何单一论文、一次评分或一次申诉不得独立证明制度状态。若研究者只观察到一次记录分叉，而没有至少一个预先规定的事件集合、比较组或连续时间窗口，读数不得外推至教授资格总体。

第九十七章　证伪条款、反向条件与一个写死到 2028 年的赌注

第一组：证伪条款

[未执行] 关系档案预测检验。去看 AI 版本、输入情境、教授事前标准、修改理由和责任记录，观察这些材料能否在教授未知的新案例中稳定预测其独立裁判。若预测率达到预先设定标准，且五类正式记录仍收敛于同一教授，本书核心命题错误。现有材料没有执行该检验，因此不得写成实验已经表明。

[未执行] 五账收敛检验。去看同一项混合判断在署名、调用日志、错误追责、同行评价和岗位考核中的归属。若多数事件都由五类记录共同指向同一教授，并且评审能够不读取全部过程就确认教授的独立资格，悬空资格不成立。现有材料没有取得相应流程文件与日志。

[未执行] 时间顺序检验。去看 AI 使用比例、混合产出比例与教授独立判断证据的连续学期记录。若独立判断证据下降先于 AI 介入，或者 AI 介入增加后连续两个学期未出现资格记录分叉，则「混合产出先扩张、资格证明随后分叉」的时间预测错误。现有材料未执行纵向追踪。

[未执行] 迁移能力检验。去看教授在无 AI 或陌生案例中的判断，比较 AI 使用增加前后的标准保持、理由质量与新案例预测率。若 AI 介入增加后教授的独立判断证据持续增加，且五类记录收敛，本书关于资格悬空的预测错误。现有材料未执行迁移测验。

[未执行] 组织产出检验。去看高校创业名录、孵化器名单和年度报告，按照高校支持或孵化的独立企业记录进行审计。若大学组织路径没有出现与 AI 课程、孵化器或创业支持相伴的可重复产出变化，本书关于大学工作路径改写的辅助判断错误。现有材料只有「两千多家初创企业」的概括性报道，没有逐条清单，不能完成真跑。

第二组：反向条件与赌注

第一条反向条件是：若制度仍能指定教授负责，且完整关系档案能够证明教授事前设定标准、拒绝模型、修正错误并在新案例中作出可预测判断，那么「责任仍在、资格无法证明」这一句被削掉。命题不能扩大为「混合产出必然摧毁资格」。

第二条反向条件是：若固定资格字段虽不再充分解释产出，但关系档案能够重新承载资格，那么「资格失去载体」必须改写为「旧载体失效、新载体形成中的过渡状态」。命题不能扩大为「资格永远无法恢复」。

第三条反向条件是：若大学只是改变了工作流程，而没有出现五类记录之间的归属分叉，则焦虑可能属于工作转型，而不是悬空资格。命题不能把所有 AI 时代的教师压力都归入自身。

截至2028 年12 月31 日，在至少30 所公开披露已将生成式 AI 用于本书提出一个可复核的制度赌注：课程评价、论文指导或科研流程的高校中，至少10 所将实施可执

行的审核条款，强制保存 AI 版本、输入情境、教授事前判断标准、修改理由与责任记录，并允许在申诉或内部审计时调取。学校只发表原则声明、要求教师自行说明 AI 使用、开展没有调取权限的试点，或只保存模型调用日志，均不算命中。若截至日期少于10 所，关于制度会朝关系档案方向回应的预测失败；但这一结果不自动推翻悬空资格本身，因为概念命题与制度回应预测属于不同层次。

第九十八章　八种竞争解释的逐一对撞

学历信号解释

学历信号解释的最强版本不是「学历会失效」，而是：当 AI 使知识获取普遍化、复杂任务的完成不再依赖长期知识占有时，学历与能力之间的区分度下降，市场会降低对学历的依赖。

这一解释能够说明教授焦虑的外部来源。教授的职称、学位与学术机构曾经共同标识一套知识训练；当学生能够借助 AI 完成复杂任务，产出不再稳定地证明知识占有，学历信号自然受到冲击。

但它解释不了这样的具体现象：当同一批 AI 参与的论文、评分或科研判断中，教授学历的薪酬差距、聘任概率和同行评价预测力下降，而教授事前判断标准、修改理由和责任档案仍能稳定预测其盲测判断时，信号理论预测的是 A——学历信号贬值；本书预测的是非 A——悬空资格尚未成立。读数分别是学历字段对能力的预测力下降，以及关系档案对教授独立裁判的预测力是否保持。

反过来，若学历信号保持不变，但教授无法从混合产出中证明自己的独立判断，信号理论不能解释该现象，而悬空资格能够解释。由此可见，学历信号是资格的一种载体，却不是资格承载问题本身。

职业管辖权解释

职业管辖权解释的最强版本是：AI 平台、行政部门和数据系统正在进入原由教授负责的知识任务，教授的焦虑来自任务定义权、评价权和处置权的转移。

这一解释能够说明为什么教授会感到职业边界不稳定。自动评分系统可能制定评分规则，平台可能决定哪些学习表现能够被看见，学院可能通过统一流程限制教师自由裁量。教授原有的知识管辖权确实可能受到侵入。

但它解释不了另一个现象：当课程评价权部分转移给平台和教学管理部门，而学校仍要求教授签署成绩、处理申诉并承担最终错误责任时，职业管辖权理论预测 A——评分与裁决权转移；本书预测非 A——即使教授名义上仍是责任主体，只要记录无法证明其独立资格，悬空资格仍然成立。前者的读数是任务控制权的归属，后者的读数是五类记录能否收敛于教授资格。

若管辖权转移之后，责任与资格证明也完整转移到平台，教授不再承担最终裁判责任，本书命题的对象便消失。此时职业管辖权解释胜出。可见，悬空资格不等于教授失去权力，而是教授仍处于责任位置却缺少资格载体。

认知外包解释

认知外包解释的最强版本是：教授使用 AI 来检索、计算、生成文本和比较方案，外包降低认知负荷、提高处理速度，但也可能减少独立核验。

该解释能够说明判断过程为何不再完全发生在教授内部。教授可以调用模型生成课程反馈、整理资料、识别论文问题，学生也可以把记忆和组织任务交给工具。AI 参与并不必然意味着能力下降，外包可以提高任务效率。

但认知外包解释不了这样的具体现象：在 AI 辅助论文指导中，教授使用工具越多，课程完成率、反馈速度和记忆准确率越高，同时教授事前标准、拒绝条件和修改理由却无法预测其在陌生案例中的判断。认知外包理论预测 A——外包有效；本书预测非 A——若独立资格无法被记录，悬空资格仍成立。两者分别读取任务绩效与资格迁移率。

若 AI 使用增加后，教授的事前标准、修改轨迹和新案例预测能力反而更加稳定，本书命题被削弱，而认知外包解释仍可成立。外包是否有效与资格是否悬空不能互相替代。

责任鸿沟解释

责任鸿沟的最强版本是：当 AI 行为由模型、数据、开发者、使用者和情境共同造成时，传统责任制度找不到一个足以承担后果的主体。

它能够解释学生成绩错误、科研判断错误和平台决策错误为何难以归责。但本书的分离线在于：责任鸿沟问「谁都无法承担责任」，悬空资格问「责任仍可指给教授，却不能证明教授具有独立知识资格」。

在 AI 自动评分导致学生申诉的场景中，责任鸿沟预测 A——学校无法把错误完整归给任何一个人；本书预测非 A——学校仍要求教授承担最终责任，但无法调出证明其独立判断的记录。若学校取消教授的最终责任，只追究平台或开发者，悬空资格的责任前提消失。

因此，悬空资格不是责任鸿沟的改名。责任鸿沟关注后果责任的空缺，悬空资格关注知识资格的证明链分叉。

角色转型与流程重构解释

角色转型解释的最强版本是：教授不再只是讲授和执行者，而要转向流程设计、课程重构、项目组织、伦理把关和创新能力培养。人的价值从执行转向更高阶的设计。

这一解释能够说明大学为什么要求教授学习新的工作方式。大学从知识传递机构转向知识平台、创业母机与创新组织，教授工作自然会扩展到系统设计与生产组织。

但它解释不了这样的现象：同一批教授在 AI 介入后设计出新课程、缩短反馈时间并增加创新项目，然而产出越多，越无法区分教授判断、模型建议和情境推动。角色转型理论预测 A——新角色已经形成，培训与评价标准将恢复职业承认；本书预测非 A——工作转型可能增加混合产出，却不能恢复独立资格。

若新角色建立了固定标准，并能在陌生案例中稳定预测教授独立裁判，本书命题作废。角色转型可能成为资格重新承载的路径，但不能被当作资格已经恢复的证明。

分布式认知解释

分布式认知的最强版本是：人、模型、评分标准和制度组成一个认知系统；只要系统整体准确率提高，就可以说认知能力提高。

这一解释能够说明为什么 AI 参与不应被简单视为人的能力丧失。知识产出本来就经常依赖语言、图表、数据库、组织流程和他人协作。AI 只是把分布式结构推进到新的程度。

但它解释不了教授个人资格的问题。当系统整体预测准确率提高，却无法预测教授在下一个陌生案例中如何裁判时，分布式认知预测 A——系统变强；本书预测非 A——教授资格仍然悬空。系统能力与个人资格不是同一个读数。

若系统整体能力提升总能稳定反推出同一教授的独立判断，悬空资格作废。本书并不否定分布式认知，而是要求它进一步回答：系统性认知如何产生可迁移、可复核的个人资格。

自动化偏误解释

自动化偏误的最强版本是：教授面对 AI 建议时可能过度相信模型，减少独立核验，导致错误率上升。

它能够解释教授为什么在使用 AI 后作出错误判断，却不能解释判断正确时的资格问题。悬空资格不是教授一定判断错误，而是即使判断正确，制度也可能无法证明判断属于教授。

同一课程中，若教授采纳 AI 建议的比例越高、错误答案越多，自动化偏误预测 A——错误率随依赖程度上升；本书预测非 A——即使教授频繁拒绝 AI 且错误率下降，只要修改理由和事前标准无法在跨案例中稳定保存，资格仍可能悬空。若错误率上升的案例中，教授标准、日志和责任链始终完整可追溯，本书不能仅凭错误增加而成立。

意识意志与过程追踪解释

意识意志的副现象论怀疑主观「我决定了」是否真正造成行动。它可以提醒我们，教授事后的自我报告不能自动成为知识资格的证据。但本书不需要判断教授的意

识是否具有因果作用。教授可能真的作出了判断，资格却仍可能没有被制度稳定保存。

过程追踪能够通过时间戳、版本变化、输入条件和理由链还原一次判断如何形成。它是检验悬空资格的重要方法，却不是命题本身。一次过程能够被完整还原，不等于过程档案能够在多个新案例中稳定预测同一教授的独立裁判。

若完整关系档案在连续新案例中稳定预测教授判断，并且署名、追责、评价和考核都收敛于同一资格，本书命题作废。过程追踪因此既是方法资源，也是对悬空资格的潜在反证。

第九十九章　四格：产出归属 × 资格记载

候选轴的排除

第一个候选轴是 AI 使用频率。它不能作为第二轴，因为 AI 使用频率是混合产出的成因或条件，不是资格状态的独立维度。使用次数多不等于产出混合程度高，更不等于资格无法记载。

第二个候选轴是教授焦虑强度。它把结果性的心理感受当作结构轴，无法区分「担心失业」与「担心资格悬空」。同样的焦虑强度可能来自不同机制。

第三个候选轴是大学产出数量。企业、论文或课程数量属于结果指标，不足以说明产出归属是否混合，也不足以说明资格能否被记录。产出增长甚至可能与资格证据下降同时发生。

本书确立两条轴：第一轴是产出归属方式，即产出能否归给教授个人，或由教授、AI 与情境共同生成；第二轴是资格记载方式，即资格能否由固定字段稳定记录，或必须在具体关系中重建。这两条轴不是 AI 使用量和资格的成因，而是共同规定资格状态的结构条件。

第二轴的确立

「产出归属」与「资格记载」相关但不重合。产出混合会增加资格记录的难度，却不必然使资格悬空。医疗、航空安全和审计都表明，人机共同完成任务后，制度仍可能通过过程档案重新承载专业资格。相反，产出仍可署名给个人时，固定字段也可能不能表达个人在具体情境中的判断理由，此时会出现隐性资格。

因此，第二轴不能由第一轴推导出来。它必须单独观察五类记录是否稳定指向同一资格，以及关系档案能否预测主体在新案例中的裁判。

二维辨别格

产出归属为个人；资格记载为固定字段

这一格是稳定资格。教授的判断是主要产出来源，学历、职称、同行评价和个人成果能够持续指向同一种裁判能力。这里，成果增加与资格证明增加保持同向关系。只在该格成立的预测是：当教授个人产出增加时，固定资格字段对其新案例判断的预测力同步增加。

产出归属为个人；资格记载须在关系中重建

这一格是隐性资格。教授仍有独立判断，产出也主要归属于教授，但学历、职称和署名无法完整表达其具体判断，只能通过问题、证据、修改轨迹与责任关系确认。只在该格成立的预测是：即使没有 AI 混合产出，同行也必须读取过程关系才能判断教授资格。

产出由人、AI 与情境共同生成；资格记载为固定字段

这一格是记载错配。制度仍把混合产出完整记在教授名下。短期成果可能提高，固定资格字段却逐渐失去对实际判断能力的预测力。只在该格成立的预测是：产出效率上升而固定资格字段的预测效度下降，但关系档案尚未出现完全分叉。

产出由人、AI 与情境共同生成；资格记载须在关系中重建

这一格是悬空资格。教授仍被要求作最终裁判，却不能从混合产出中单独证明其资格。五类记录分别指向教授、模型、平台、学院和课程组，责任仍在教授，资格却无法作为脱离关系的对象保存。只在该格成立的预测是：产出指标上升时，若不建立关系档案，教授独立判断的可预测性下降而责任记录继续集中于教授。

四格两两比较可见，稳定资格与记载错配的差别在于产出是否混合；隐性资格与悬空资格的差别也在于产出是否混合；稳定资格与隐性资格的差别在于固定字段能否承载资格；记载错配与悬空资格的差别则在于资格是否必须回到关系中重建。悬空资格因此不是四格中任意一种模糊的「不确定」，而是右下格的特定组合。

第一百章　讨论：资格是产出、责任与记录之间的连续性

理论意涵

本书的理论意涵首先在于重新规定「资格」。资格不再只是主体持有的属性，也不是社会承认的总和，而是能够在产出、责任和记录之间保持连续性的关系。一个教授是否有资格，不仅取决于他拥有何种学历、职称或职位，也取决于制度能否在具体判断中识别他提出了什么标准、作出了什么修正、承担了什么后果，并在新案例中预测其独立裁判。

其次，本书把权威问题从「社会是否承认」推进到「承认凭什么能够持续」。学历信号理论关注资格字段的预测力，职业管辖权理论关注判断权的归属，悬空资格则关注资格字段与判断过程之间是否仍然存在可证明的连接。

再次，本书把混合产出与责任制度之间的关系拆开。AI 并不必然造成无人负责，制度可以继续指定教授负责；但责任指定可能遮蔽资格缺失。教授越被要求为混合系统负责，越需要一套能证明其独立判断的关系记录，否则责任制度会把资格问题伪装成个人义务问题。

实践意涵

大学评价不能只增加 AI 使用声明或生成内容检测。检测学生是否使用 AI，只能回答工具是否出现，不能回答判断由谁承担。大学应记录教授在看到模型结果前的标准、预测和拒绝条件，记录模型版本与输入情境，保存教授的修改理由，并在申诉与晋升时区分最终产出与独立判断。

教学评价也应从结果单一归属转向判断链评价。论文评分不能只看最终文本，科研评价不能只看署名与论文数量，课程考核不能只看反馈速度。若大学只奖励 AI 带来的产出增长，可能把混合系统的能力误记为教授个人能力。

职业发展制度需要承认关系性资格，但不能把关系性资格降格为无边界的主观叙述。判断链必须可复核、可比较、可迁移。教授不必拒绝 AI，但必须能说明哪些判断交给了工具，哪些判断仍由自己承担，以及为何接受或拒绝模型建议。

适用边界与反向约束

第一，本书不适用于责任已经完整转移、教授不再承担最终裁判的场景。若平台或机构成为真正的最终裁判者，研究对象转为职业管辖权转移，不能继续称为教授资格悬空。这个边界削掉了「教授只要使用 AI 就会资格悬空」这一强句。

第二，本书不适用于关系档案已经能够稳定预测教授独立判断的场景。若教授事前标准、修改理由、责任记录与新案例判断保持稳定，固定资格字段即使不再充分，也可能由关系档案重新承载。这个边界削掉了「混合产出必然摧毁个人资格」这一强句。

第三，本书不适用于只有工作量增加、评价标准改变而没有记录分叉的场景。那里可能是角色转型或组织压力，而不是悬空资格。这个边界削掉了「所有 AI 时代教师焦虑都来自资格变化」这一强句。

反噬

若大学采纳本书命题，最省事的做法是增加字段：AI 版本、提示词、修改次数、责任人、复核人、申诉人逐项填报。然而，这种做法可能恰好毁掉要保住的判断。教授可以在看到模型答案之后补写理由，平台可以自动生成完整日志，学院可以形成看似详尽的关系档案；但这些记录只说明过程被保存，不说明教授曾经独立判断。

因此，记录越多，资格不一定越稳。若制度把可记录等同于有判断，把事后解释等同于事前标准，把合规留痕等同于主体能力，关系档案会变成新的表格官僚制。它保存的是责任证据的外观，却没有保存资格的生成过程。本书看不到一个无需持续制度实验的简单出口：资格修复必须同时防止记录缺失与记录过剩。

第一百零一章　结论：在这项判断中，究竟是谁有资格裁决

对第一问的回答

生成式人工智能使教授资格失去原有承载方式，不是因为教授马上失去职位，而是因为知识产出由教授、AI 与情境共同生成，产出不再稳定指向教授个人。与此同时，大学仍把最终评价、签署、解释与纠错责任交给教授。过去由学历、职称、署名和个人作品形成的资格证明链因此发生分叉。

对第二问的回答

悬空资格区别于学历信号贬值，因为后者讨论学历字段预测能力下降，前者讨论资格是否仍能被独立记录；区别于职业管辖权转移，因为后者讨论谁控制任务，前者讨论责任仍在教授时资格能否被证明；区别于认知外包，因为后者讨论判断由谁执行，前者讨论判断完成后资格能否指回教授；区别于责任鸿沟，因为责任鸿沟讨论责任主体消失，悬空资格讨论责任仍被指定但资格证明断裂；区别于分布式认知，因为分布式认知讨论系统整体如何产生认知，悬空资格要求系统进一步说明如何承载个人可迁移的裁判资格。

对第三问的回答

可裁决判据是：同一项混合判断进入署名、调用日志、错误追责、同行评价和岗位考核后，五类记录是否稳定指向同一位教授的独立裁判资格。若关系档案能够记录 AI 版本、输入情境、教授事前标准、修改理由和责任链，并在新案例中稳定预测教授判断，同时五类制度记录收敛于同一资格，悬空资格命题作废。

本书尚未执行流程审计、日志分析或纵向盲测，因此没有交出经验上的裁决性证据。本书交出的是一个可被反驳的概念、一个二维辨别格、一组可观测指标和明确的证伪条件。

研究启示在于，大学不应把 AI 治理简化为禁止、披露或检测。真正的问题是：在混合产出条件下，如何重新记录判断发生的过程，并让资格在新情境中保持可解释、可复核和可预测。教授的未来价值不再主要来自独占知识，而来自在混合系统中设定标准、拒绝错误、承担修正并培养他人独立判断的能力。

但这个结论仍留有开口。若关系档案最终能够重新承载教授资格，悬空资格可能只是制度过渡期的名称；若关系档案无法区分事前判断与事后辩护，资格问题则可能比本书所描述的更深。下一步不是继续宣称教授将被替代，而是去检查新大学的记录是否还能回答：在这项判断中，究竟是谁有资格作出裁决。

卷七　悬空理论：资格、判断与责任的三重断裂与同源闭合

本卷把资格、判断、责任三重悬空合成一个理论。原文五万字，本书取其骨干：三重悬空的定义与边界、悬空向量与责任—控制错位的形式化、同源闭合原则、五条发生路径与一个临界点、六组对照实验、六个领域的机制演绎、资格与责任的重新定义、反噬与自我否定、制度方案、四个结构命题与不可能三角、三个思想实验、治理指标与总模型。另并入「首个不可回补点」的正式化与四次分叉的内生推导。删去原文的自评、投稿定位、术语词典与研究问题清单——它们是论文的护身符，不是书的正文。

第一百零二章　问题的重新提出：人工智能真正悬空了什么

人工智能时代最常见的叙事仍然围绕「替代」展开：机器会不会取代人，哪些岗位首先消失，哪些技能会贬值，哪些新职业会出现。这一叙事重要，却不足以解释一个越来越普遍的经验事实：许多人的岗位并没有消失，甚至工作量、产出数量与制度责任都在上升，他们却比过去更难回答「这究竟是不是我的判断」「这个结果究竟证明了我的什么能力」「为什么最终后果仍然由我承担」。焦虑因而并不总是发生在被机器赶出岗位的那一刻，它更可能发生在岗位仍在、签字仍在、责任仍在，而判断的生成条件已经离开单个主体的时候。

大学教授是一个典型场景。教授可以使用人工智能设计课程、检索文献、整理研究材料、生成论证草稿、制作评价标准；学生也使用同类系统完成阅读、写作和答辩准备。最终论文仍署学生或教授之名，成绩仍由教师签署，学术不端、事实错误和评价争议仍会落到具体的人。但在结果形成之前，问题可见性、资料召回、候选答案、排序、默认值、模型版本、平台策略与课程规则已经共同塑造了选择空间。这里出现的不是一个简单的「工具辅助」事实，而是一个制度上的断裂：结果的发生结构已经分布化，资格、判断和责任的结算结构却仍然个人化。

医生、律师、工程师、审计师、公共行政人员也面临相同结构。影像系统先行排序，医生最后签字；法律检索系统过滤案例，律师最后出具意见；代码生成系统给出实现，工程师最后提交；风险模型预先分类，工作人员最后批准或拒绝。制度依靠末端动作维持一种清晰的责任图景：总要有人签字，总要有人说明，总要有人接受处罚。但是「谁最后出现」并不能自动回答「谁控制了关键判断条件」。当制度把最后的确认动作解释为完整判断，它就把一条多源生成链压缩成一个人的行为。

本书把这种断裂概括为「悬空」。悬空不是指所有权威消失，也不是指无人负责，更不是说 AI 一出现人的资格就不存在。悬空指的是：一个制度性对象——资格、判断或责任——仍然被要求存在并发挥效力，但它过去赖以闭合的生成链已经断裂。资格还被写在个人名下，却不再能从混合产出中独立证明；判断还被当作个人判断，却无法由个人恢复全部前置条件；责任还被集中于末端主体，却与控制、可见、收益和修复能力不再对应。它们「挂」在制度上，却缺少与生成过程同源的支撑，所以称为悬空。

这一概念试图回答三个比「AI 会不会替代人」更基础的问题。第一，专业资格在何种条件下仍然能够作为个人属性被记录，在何种条件下必须改写为关系性资格？第二，什么样的人工介入才足以让最终确认成为真正的个人判断，而不只是对平台已经塑造的候选空间进行末端选择？第三，当关键条件由不同主体控制时，责任应依据什么原则重新分层，而不是在「人全部负责」与「机器负责不了所以无人负责」之间摆动？

本书的中心命题是：人工智能造成的一个根本制度冲击，是现代社会的「个人资格—个人判断—个人责任」闭合结构，与「人—AI—平台—组织—情境」的复合生成结构发生错位。其最强形式可以写成：当生产来源分布化、关键判断条件平台化、资格证明仍个人化、责任仍末端化，并且末端个人无法恢复前置选择空间时，悬空结构

成立。反之，若个人能够完整查看候选、改变关键设置、说明关键取舍、进行独立核验，且制度按实际控制范围分配责任，则即使 AI 深度参与，悬空也未必出现。

本书因此不是一篇关于「AI 导致焦虑」的一般心理学论文，而是试图建立一种可裁决的制度发生理论。焦虑只是悬空结构在人身上的一种显露；更基础的对象是资格、判断与责任之间的连接如何发生、如何断裂、如何重新闭合。

第一百零三章　从「悬空资格」与「悬空判断」到「悬空理论」

「悬空资格」与「悬空判断」分别抓住了同一结构的两个断面。前者关注资格的承载问题：当知识产出由教授、AI 与具体情境共同形成时，学历、职称、署名、作品等传统字段是否还能够稳定证明某一主体的独立裁判能力？后者关注判断的发生问题：当候选集合、资料排序、默认阈值、风险提示和界面可见性被平台或模型塑造时，末端签字是否仍足以证明「判断由签字人完整完成」？两者一旦放在同一条生成链上，就会显出第三个无法回避的问题——责任如何结算。

资格悬空的核心不是「资格降低」，而是「资格失去独立载体」。一个人的学历可能仍然有效，职位可能仍然稳定，组织也可能仍承认其权威，但如果最终产出已经不能稳定预测其在陌生案例中的独立判断，那么传统资格字段与判断能力之间的证明关系已经松动。这里发生的不是量的减少，而是存在方式的改变：资格从可以脱离具体过程被固定记录的属性，转变为必须从问题设定、证据选择、拒绝条件、修改理由、独立核验与责任承担等关系中重建的结构。

判断悬空的核心不是「AI 影响了人」，而是关键判断条件与最终责任动作发生了空间和权限上的分离。人可能拥有最后的确认按钮，却看不到所有被过滤的候选；可以否决眼前答案，却不能修改排序阈值；能够承担错误后果，却无法调用原始版本或恢复被删除的信息。此时，末端确认只证明「这个人接受了当前可见结果」，并不能证明「这个人拥有形成该结果所必需的全部判断条件」。判断由此从一个人的完整行动转变为一条跨节点的差异路径。

一旦资格和判断同时悬空，责任问题就变得尖锐。现代制度常用签字、执照、职位或账号作为责任锚点，这种做法在传统条件下具有合理性，因为生产、判断、成果归属和修复能力大体集中于同一主体。AI 平台化之后，四个位置开始分离：模型提供者可能控制生成机制，平台控制候选与排序，部署机构控制强制使用与时间压力，专业人员控制最终采纳，数据提供者影响输入边界，组织规则决定可否复核。若制度仍把全部责任压到末端个人身上，就不是「责任清楚」，而是责任与控制结构错配。

因此，本书把三种悬空明确区分。第一，资格悬空：制度仍要求某主体以个人资格证明其专业判断，但固定资格字段与混合产出不能稳定指向同一独立裁判能力。第二，判断悬空：关键判断条件分布于多个主体或系统，末端个人无法完整恢复、解释或修改这些条件，制度却仍把结果呈现为个人完整判断。第三，责任悬空：责任被集中指定给某主体，但其控制权、可见性、修复权和收益位置不足以支持相应责任，责任因此失去与生成链的同源支撑。

三者不是同义词。资格可以悬空而判断不悬空，例如一个教授能够独立判断，却因组织评价字段失真而无法被稳定识别；判断可以悬空而资格暂时不悬空，例如医生长期专业能力很强，但某一次诊断被平台前置筛选所限制；责任也可以悬空而资格与判断均相对稳定，例如机构强制采用某系统，却把系统性风险全部转嫁给操作员。只有在三者叠加时，才形成本书所谓「强悬空结构」。

强悬空结构可以概括为四个连续错接：生产分布化、资格个人化、判断末端化、责任集中化。生产分布化意味着结果的关键来源跨越多个节点；资格个人化意味着制度仍把能力证明压到单一主体；判断末端化意味着最终点击、签字或提交被等同于完整判断；责任集中化意味着后果主要由最末端的人承担。四者叠加，会产生一种独特制度悖论：系统越复杂、协作越深、平台控制越强，末端个人反而越需要表现得像一个「完整自主主体」。

本书因此主张，「悬空」不是一个描述 AI 时代不确定性的修辞，而是一种可被拆解、编码和证伪的关系结构。它必须回答：哪些节点参与了结果形成？哪些节点控制候选空间？哪些主体能够恢复被排除信息？谁拥有修改与暂停权？谁从系统运行中获益？谁承担何种后果？若这些问题只能回答最后一项「谁签字」，而不能回溯前置条件，悬空即开始出现。

第一百零四章　理论最近邻：为什么既有理论不足以单独解释悬空

悬空理论并不建立在否定既有研究之上。相反，它依赖若干成熟理论提供的局部洞见，但要求把这些洞见放进同一条发生链中。最重要的最近邻至少包括学历信号理论、职业管辖权理论、认知外包、分布式认知、责任鸿沟、有意义的人类控制、自动化偏误、技术政治与问责理论。

学历信号理论解释教育凭证如何在信息不对称条件下传递能力信息。它能够说明 AI 降低知识获取成本之后，学历对某些任务能力的区分度为何下降。但信号强弱与资格存在方式不是同一个问题。即使某一学历信号仍有市场价值，也不能自动证明由 AI 与人共同形成的具体产出代表持证人的独立判断；反过来，即使学历信号贬值，主体也可能通过稳定的过程档案、事前标准和跨案例判断重新证明资格。悬空资格关注的不是「信号变弱」，而是「资格能否作为对象被稳定承载」。

职业管辖权理论解释专业职业如何通过控制问题定义、诊断、推论、处置和评价边界获得地位。它能说明平台为何可能侵入教授、医生或律师的专业空间，却容易把权力理解为可见的任务控制。平台化控制往往更隐蔽：职业名义没有改变，医生仍是医生、教授仍是教授，但平台已经决定哪些证据先被看见、哪些候选被排除、默认阈值是什么、哪一种风险被突出。悬空理论因此把管辖权从「谁最后决定」推进到「谁控制了决定的前置条件」。

认知外包与分布式认知对本书更为关键。它们已经明确指出，认知并不必然封闭在人脑内部，而可以跨越工具、符号、人员与环境分布。AI 显然把这一结构推向新的深度。然而，描述「认知如何分布」并不能自动回答「资格是否随认知分布而重构」「责任是否随控制分布而重新分配」。一个系统可以在整体上表现得更聪明，却让其中某个自然人的独立判断越来越难被识别；也可以由多主体共同完成高质量决策，却仍让事故责任全部落到最后操作员身上。悬空理论的增量恰恰位于「分布之后怎么办」。

责任鸿沟研究关注学习型自动化与复杂系统如何使传统归责者难以确定。悬空责任与之相邻，但并不相同。责任鸿沟常描述「找不到足够负责的人」；悬空责任描述的是「制度已经找到一个人负责，但这个人的控制与修复能力不足以支持被赋予的全部责任」。前者是责任空缺，后者是责任错压。一个医院可以非常明确地处罚签字医生，同时拒绝提供模型版本、过滤日志和风险阈值。在这种场景中，责任并没有消失，恰恰是过度集中到了最可见的末端主体。

「有意义的人类控制」理论强调自动化系统应保留实质性的人类掌控。这一原则与悬空判断高度相关，但本书把「控制」进一步分解为若干可裁决权限：候选可见权、过滤恢复权、排序修改权、默认值改变权、独立核验时间、暂停权、升级权、错误修复权与过程记录调用权。一个形式上的拒绝按钮不等于完整控制。若人只能拒绝当前展示的答案，却不能看到被排除选项，拒绝仍发生在平台已经定义的空间内。

自动化偏误关注人过度依赖机器建议而减少独立核验。它能够解释某些错误，但悬空不是错误理论。即使最终判断完全正确，只要制度不能说明关键判断条件来自何

处、资格如何被证明、责任如何与控制对应，悬空仍然可能存在。正确性与归属结构是两个不同维度。

技术政治研究指出人工制品与系统安排能够携带权力关系。悬空理论接受这一点，但要求把「技术具有政治性」进一步落实到责任节点：谁设定候选、谁决定排序、谁能够修改默认、谁控制日志、谁从效率提升中获益、谁拥有修复预算。只有把政治性转化为可观察的控制结构，才能进入责任分配。

问责理论关注说明、质询、判断与制裁关系。悬空理论则提出一种前置要求：说明义务必须覆盖形成结果的关键条件，而不能只要求末端个人解释已经被系统预先塑造的结果。若机构要求工作人员对某个算法审批结果进行详细说明，却不给其查看算法排序与被排除选项的权限，这种问责实际上会制造新的悬空。

因此，本书并不是把既有理论换一个名称，而是把它们各自握住的变量放在同一条差异路径中：信号解释资格如何被看见，管辖解释谁有权判断，分布式认知解释判断如何跨主体发生，自动化控制解释人如何介入，责任理论解释后果如何归属。悬空理论处理的是这些变量在 AI 平台中发生错接时的结构：生产可以分布，资格却仍个人化；控制可以平台化，判断却仍被呈现为个人化；责任可以明确，控制却没有同步。

第一百零五章　发生学基础：从个人主体到人—AI—平台复合体

要理解悬空，必须改变一个默认本体假设：结果不是先由一个完整主体形成，再由工具帮助其表达；在许多 AI 场景中，主体、工具、平台、规则与情境在互动过程中共同构成了结果发生的条件。本书借用发生学本体论的显露、差异路径与纠缠土壤来表达这一点：。这里不把三层当作经验回归方程，而把它们作为发生关系的形式骨架。

显露是制度最终能够看见并记录的表面结果：论文署名、医生诊断、法律意见、代码提交、审批决定、课程成绩、事故报告。传统制度倾向于从显露反推主体：「谁署名，谁创作；谁签字，谁判断；谁提交，谁负责。」这种反推在生产链较短、判断条件较集中时具有实用合理性。

差异路径是从问题提出到结果显露之间发生的一系列分化：问题界定与答案生成不同，资料召回与资料过滤不同，候选生成与候选排序不同，排序与人工采纳不同，人工修改与独立核验不同，平台控制与个人担责不同。AI 时代的关键不是「系统参与」这一静态事实，而是这些差异沿时间、权限与可见性逐步展开。悬空正发生在差异路径被制度压缩之处。

纠缠土壤表示最终结果中不同来源的贡献并非总能被简单拆分。一个论证可能来自模型生成、人的修改、数据库内容和课程规则的共同作用；一个诊断可能由影像数据、风险模型、医生经验、医院流程和时间压力共同形成。纠缠不意味着因果不可分析，也不意味着责任必须平均分配，而是提醒我们：最终结果并不是单一主体属性的直接外化。

从这个角度看，人—AI—平台不是三个彼此独立、先在存在的实体简单相加，而是在具体任务中形成一个复合体存在物。复合体并不取消人的主体地位，却改变主体地位的边界。人仍然可以提出问题、设定价值、拒绝建议、承担承诺；AI 可以生成候选、压缩搜索、转换表示；平台可以过滤、排序、设置默认、保存或不保存日志；机构可以规定必须使用的系统、绩效指标与复核时间。结果来自这些位置之间的互动结构。

传统责任制度的困难在于，它仍倾向于把复合体生成的结果「回写」为单人行为。回写不是完全错误，因为法律和伦理需要可回应的责任主体；真正的问题是回写时是否抹除了关键生成条件。如果平台改变候选集合，机构缩短复核时间，模型版本引入系统性错误，而最终只保存一条「某医生点击确认」的记录，制度就把复合生成压成了个人显露。

本书因此区分「责任主体」与「完整判断主体」。人可以仍然是责任主体，因为人能够理解规范、作出价值取舍、回应他者；但只有当人拥有与其责任相称的可见性、控制权与修复能力时，才能被视为完整判断主体。这个区分避免两个极端：一方面，不因为 AI 参与就把人免责；另一方面，也不因为人最后签字就假定所有前置条件都属于人的自由判断。

这一发生学立场还意味着资格不应被理解为静态属性。资格是一个关系持续闭合的能力：主体能否在不同情境中重复提出标准、识别反例、拒绝错误、解释取舍、承担修复，并被制度记录为同一可迁移的判断能力。学历、职称和执照只是资格的历史载体，不等于资格本身。当混合产出使固定字段与判断能力之间的连接松动时，资格就必须重新通过关系链被证明。

换言之，悬空不是「主体消失」，而是旧主体模型的边界暴露。人工智能把过去被近似压在一个人身上的生产、判断、资格与责任拆开，使制度不得不面对一个长期被忽略的问题：我们究竟依据什么把一个复杂结果归属于一个人？

第一百零六章　三重悬空的严格定义与边界

为了避免「悬空」退化为泛化批评，本书给出三重定义，并规定各自的成立与不成立条件。

第一，资格悬空。定义为：在混合产出条件下，制度仍要求某主体承担专业裁判角色，但固定资格记录不能稳定承载并预测该主体在新情境中的独立判断。它至少需要三个条件：其一，结果由主体、AI 与情境共同生成；其二，制度仍以个人身份进行授权、评价或追责；其三，署名、调用日志、错误追责、同行评价、岗位考核等记录之间出现归属分叉，且固定字段不能预测主体的跨案例裁判。若关系档案能够稳定记录事前标准、拒绝条件、修改理由与新案例判断，资格便可能重新闭合，即使传统学历信号已经下降。

第二，判断悬空。定义为：形成最终结果的关键判断条件分散于多个主体或系统，末端个人无法完整恢复、查看、修改或解释这些条件，而制度仍把最终确认视为个人完整判断。它至少需要：多源生成；至少一个前置判断节点由非末端主体控制；末端个人无法恢复前置选择空间；制度仍把末端动作作为主要判断证据。若人能够查看完整候选集、调取被排除信息、改变关键设置、独立核验并暂停系统，则 AI 参与本身不足以形成判断悬空。

第三，责任悬空。定义为：制度明确指定某主体承担后果，但其责任份额显著超过其对关键条件的控制、知情、修复与收益位置所能支持的范围。责任悬空不是「无人负责」，而是「责任有名、支撑无源」。其典型表现是：平台与机构拥有设置和修复权，个人只有末端确认权；系统性失效在多人身上重复出现，却每次被解释为不同个人的疏忽；日志和版本控制掌握在平台手中，个人却被要求解释平台不能向其展示的前置条件。

这三种悬空存在交叉但不能互推。为清晰起见，可以构造八种理论状态。若三者均不成立，是闭合协作：AI 参与但资格、判断和责任均有对应记录与权限。仅资格悬空，说明主体实际能判断且责任合理，但制度无法稳定识别其资格。仅判断悬空，说明某次判断链被平台塑造，但个人长期资格与责任结构尚未改变。仅责任悬空，说明责任分配失衡，但判断过程和资格记录可能仍完整。两两叠加则形成更复杂的过渡状态。三者同时成立，则进入强悬空。

强悬空的判别不能只看 AI 使用频率。一个高度自动化系统可能因为日志完整、候选可见、责任分层而低悬空；一个非常简单的推荐系统也可能因为默认值不可更改、反向材料不可见而造成高悬空。也不能只看错误率。一个结果完全正确，仍可能是悬空判断；一个错误结果也可能发生在闭合系统中，只是个人作出了错误决定。

强悬空也不同于一般组织复杂性。团队协作早于 AI 存在，航空、手术、实验室与大型工程长期依赖多人协作。关键差异在于平台化系统能够以不显眼的方式预先塑造候选空间，同时把这种塑造隐藏在「个性化推荐」「默认设置」「排序优化」「风险提示」等界面机制之中。过去团队成员的意见常以可见互动出现，今天平台的前置裁决可以在使用者不知道的情况下发生。

本书特别引入「首个不可回补点」来划定判断悬空的边界。所谓首个不可回补点，是沿判断时间线向前追踪时，第一个使末端主体凭自身权限无法恢复被删除候选、原始排序、旧版本、隐藏风险或原始阈值的节点。从这个节点开始，后续的人工选择都是在不可逆地收窄的空间中进行。找到这一点，比泛泛追问「AI 影响了多少」更有裁决力。

同样，资格悬空的边界不是「有没有 AI 痕迹」，而是「关系档案能否在新案例中预测主体判断」。若一个教授大量使用 AI，但其事前标准、反例处理、拒绝条件与修改轨迹能够稳定预测其在陌生案例中的裁判，那么 AI 使用并未摧毁资格，只是改变了资格的载体。相反，若产出数量不断上升、学历与职称仍然存在，却无法预测主体下一次会如何判断，资格就可能处于悬空。

责任悬空的边界则要求区分「最终采纳责任」与「前置条件责任」。专业人员应对其能够看到、理解、拒绝、核验和修复的部分负责；平台应对候选、排序、默认、日志与已知系统性失效负责；模型提供者应对版本变化、已知限制与风险呈现负责；部署机构应对强制使用、培训、时间压力、升级流程与修复资源负责。责任不是沿生成链平均摊薄，而是沿实际控制节点分层。

第一百零七章　形式化模型：从悬空向量到责任—控制错位

概念理论若没有操作化，容易停留在修辞层。本书因此提出一组形式指标。它们不是为了用一个数字替代复杂判断，而是为了让不同案件能够在同一结构中比较，并明确什么证据会支持或削弱理论。

首先定义判断链可回溯度 T。设一次事件包含 n 个被预先编码的关键决策节点，其中满足完整时间戳、执行主体或系统版本、输入输出记录、被排除内容记录、理由说明与责任权限信息的节点数为 n_t，则：T=n_t/n，T∈[0,1]。T=1 表示所有关键节点均可重放，T=0 表示只能看到最终结果。高 T 不等于责任公平，但低 T 会显著增加判断悬空的不可裁决性。

第二定义候选恢复率 R_c。设平台在关键节点排除或隐藏的候选总数为 N_e，末端人员在决策前可凭自身权限调取并恢复的候选数为 N_r，则 R_c=N_r/N_e。当 R_c 接近0 时，末端人员的「最终决定权」可能只是对已收窄空间的选择；当 R_c 接近1 且人可改变筛选规则时，平台前置裁决的约束明显降低。

第三定义资格记录收敛率 R_q。选择五类制度记录——成果署名、模型/平台调用日志、错误追责、同行评价、岗位考核——观察它们是否指向同一主体的独立裁判资格。若指向同一主体的记录数为 k，则 R_q=k/5。R_q 本身不是资格的真值，而是资格在制度记录中的闭合程度。更重要的补充是资格迁移率 M_q：用主体的事前标准、修改理由与责任记录预测其在陌生案例中的裁判，计算预测一致性。只有 R_q 与 M_q 共同稳定，才能说明资格得到重新承载。

第四定义控制—责任错位度 D_a。对系统中每一个主体 i，设其对关键条件的综合控制权为 c_i，可见性为 v_i，修复能力为 m_i，制度赋予的责任份额为 a_i。先构造责任支撑权重 w_i=(c_i+v_i+m_i)/Σ_j(c_j+v_j+m_j)，再定义 D_a=1/2·Σ_i|a_i-w_i|。D_a∈[0,1]。当 D_a 接近0，责任分配与控制—可见—修复结构大体一致；当 D_a 很高，责任明显偏离实际能力位置。这里的权重不是法律责任公式，而是结构诊断工具。

第五定义设置效应 E_s。设在可比较事件中，平台、模型版本或关键默认值发生变化的事件数为 N_s，其中错误类型、候选集合或最终选择发生系统性变化的事件数为 N_c，则 E_s=N_c/N_s。若同一人换平台后错误结构改变，而换人不换平台时错误结构相对稳定，平台前置条件的解释力增强。

第六定义悬空向量Ω。本书不建议把所有维度机械压成一个总分，而使用向量表示：Ω=(Q,J,A)，其中 Q 表示资格悬空程度，可由（1-R_q）与（1-M_q）共同描述；J 表示判断悬空程度，可由（1-T）、(1-R_c) 与设置效应共同描述；A 表示责任悬空程度，可由 D_a 描述。向量方式的优点是允许不同系统以不同方式悬空：一个高校可能 Q 高、J 中、A 中；一个医疗系统可能 Q 低、J 高、A 高。

若为了比较群体而需要一个探索性综合指数，可定义Ω*=αQ+βJ+γA，其中α+β+γ=1，并在研究前预注册权重。本书不主张存在普适权重。教育研究可能更重

视 Q，安全关键系统可能更重视 J 与 A。只要权重是在看到结果之前规定，并进行稳健性分析，综合指数可以作为辅助读数。

这些指标必须配套自我否决条款。第一，若不同编码者对关键节点、候选排除或责任份额的编码一致性过低，指标不得作为稳定证据。第二，若 T、R_c、R_q、D_a 与既有透明度、可解释性、问责清晰度量表完全重合，且不能提供「首个不可回补点」「资格跨案例预测」或「责任—控制错位」的额外信息，新构念应被视为改名而非创新。第三，单次事件中的比例只描述该事件，不得被解释为未来概率或制度总体属性。

形式化的真正意义不在公式本身，而在于改变研究者的提问方式。面对一个错误，不再只问「谁点了确认」，而问：关键候选是谁生成的、谁过滤的、谁能恢复、哪些设置发生了变化、不同人员是否重复同型错误、谁拥有修复权。面对一个资格争议，不再只看学历和署名，而问：主体的事前判断能否迁移、五类记录是否收敛、过程档案能否预测新案例。

第一百零八章　同源闭合原则：悬空理论的规范核心

悬空理论如果只揭示断裂而不能给出重建原则，就会停留在诊断层。本书提出「同源闭合原则」作为规范核心：凡参与形成结果的关键条件之来源，都必须以与其实际作用相称的方式进入记录、解释、纠错和责任结构；凡被要求承担责任的主体，都必须拥有与该责任相称的控制、知情、质疑和修复机会。

「同源」不是要求把每一个微小因果因素都写入法律文书，也不是把 AI 拟制为道德人格，更不是平均分责。它要求的是：制度在结算责任时不能抹除自己在生成结果时依赖的关键来源。若平台通过排序决定什么先被看见，那么排序规则至少应进入可审计结构；若模型版本改变错误类型，那么版本变化必须进入说明与修复义务；若机构强制使用某系统并压缩复核时间，机构不能在事故后把这些条件从责任链中删除。

「闭合」则意味着四条链能够互相对接。第一是生成闭合：能够说明结果由哪些主体和系统共同形成。第二是可见闭合：承担判断的人能够看到足以支持其判断的候选与限制。第三是修复闭合：发现错误后，责任主体或与其协作的上游主体拥有改变造成错误条件的权限。第四是责任闭合：制度赋予的责任份额与控制、可见和修复位置大体对应。

同源闭合可以形成一个强命题：如果一个主体对某关键条件既不可见、不可控制、不可恢复、不可修复，而该条件又系统性决定结果，则制度不应把该条件造成的全部后果仅以末端签字归责给该主体。这一命题并不免除末端主体的采纳责任。人仍应对明显风险、独立核验义务、价值取舍和最终接受负责；但「最终接受」不应吞并所有前置责任。

本书进一步提出「同源闭合定理」的弱形式：在一个包含多主体关键控制节点的决策系统中，若责任分配长期与控制—可见—修复结构显著偏离，则仅增加末端问责不能稳定降低系统性错误；相反，它会增加防御性记录、形式确认与责任转嫁。其逻辑是，无法改变错误条件的人即使受到更重处罚，也不能消除导致错误的上游机制；真正拥有修复权的主体若不进入责任结构，则缺乏改变机制的制度激励。

这个定理提出一个可检验预测：在高 D_a 系统中，若治理措施只提高末端个人处罚而不改变候选可见、设置控制和修复权限，则同型错误的跨人员重复率不会显著下降，反而可能增加合规性点击、事后理由补写与过度留痕。若数据反而显示仅处罚末端人员即可持续消除系统性错误，且平台设置不具有独立解释力，同源闭合的强版本应被削弱。

同源闭合还改变资格制度。资格不应只由结果数量和最终署名来证明，而要由「可迁移判断链」来证明。一个人是否具有资格，应观察其能否在不同 AI 系统、不同数据与陌生案例中保持问题标准、识别反例、解释拒绝、修正错误并承担后果。换言之，AI 时代的个人资格不是「没有使用 AI」，而是「即使使用 AI，也能在复合体中保持可识别、可迁移、可负责的判断位置」。

因此，同源闭合不是反技术原则。它并不要求把所有判断重新搬回人的大脑，也不要求取消平台排序与模型生成。它接受分布式生产作为事实，并试图为分布式生产

建立相称的制度接口。真正的问题不是「让人重新独自完成一切」，而是「让每个关键来源在记录、权利和责任上留下与其作用相称的痕迹」。

第一百零九章　悬空的发生机制：五条路径与一个临界点

悬空并不是 AI 参与之后自动出现的结果，而是通过一组可识别的机制逐步生成。本书把最常见的机制概括为五条路径：过滤、排序、默认、界面与选择性留痕。五条路径可以单独存在，也可以叠加；真正使其成为悬空机制的，是它们改变了末端主体的选择空间，却没有同步改变资格与责任的制度呈现。

第一条路径是过滤。任何复杂任务都需要筛选信息，过滤本身不是问题。问题在于谁设定过滤标准、被排除的内容是否可恢复、末端主体是否知道过滤已经发生。以学术检索为例，平台可能以相关性阈值排除一组反向证据；教授最后看到的只是「最相关」的材料。如果不同教授在同一系统中都漏掉同一反例，而平台可以调取被排除记录、教授不能，那么错误的首个不可回补点可能位于过滤而不是最终写作。此时把全部责任解释为「教授没有认真查资料」会错认生成节点。

第二条路径是排序。候选即使没有被删除，先后顺序也会改变注意力分配、核验时间与默认接受概率。医疗风险列表、法律案例排序、代码补全建议、论文搜索结果都具有这种效应。排序尤其容易被误认为中性，因为所有选项「理论上仍然存在」。但当工作时间有限、界面只显示前若干项时，排序事实上重新分配了认知资源。若平台控制排序而末端人员无法重建原始序列，排序就成为前置裁决。

第三条路径是默认。默认值把「不行动」转化为某种选择。在公共审批中，一个申请被默认标记为高风险；在医疗系统中，某类检查默认不推荐；在教育平台中，某种评分规则预先启用。使用者当然可以修改，但修改需要额外知识、权限和时间。默认因此并非简单便利，而是一种把制度偏好嵌入操作路径的机制。当机构事后只记录「某工作人员确认了默认值」，而不记录默认是谁设定、为何设定，就产生末端判断的虚假完整性。

第四条路径是界面。界面不是信息的透明容器，而是决定什么可见、什么突出、异议需要多少成本的组织结构。一个红色风险提示与一个隐藏在二级菜单中的反向材料具有完全不同的行为效应。界面还可以通过措辞把模型建议包装成「推荐方案」「最佳实践」「高置信答案」，让形式上的人工选择发生在强烈预设中。判断悬空往往不是人没有选择，而是选择的可行空间已被界面不对称地塑造。

第五条路径是选择性留痕。许多系统完整记录最后一次点击，却不记录被排除候选、模型旧版本、排序变化、机构时间压力和人工质疑过程。结果是，事故发生后最容易被看到的证据恰好是末端个人的确认动作。选择性留痕因此会把技术结构转换成责任结构：谁留下最多可见记录，谁就最容易被归责；谁掌握后台规则但不留下可调用痕迹，谁反而退出责任视野。

五条路径的共同特征，是它们都不必取消人的最终决定权，就能改变决定的发生条件。这也是悬空理论与「机器取代人」叙事的根本差异。平台无需成为法律主体，也无需公开宣布夺取职业管辖，只需要控制候选的进入、顺序、默认、显示和记录，就已经参与了判断形成。

这五条路径最终汇聚到一个临界点——首个不可回补点。在一条完整判断链上，前面的许多改变可能仍可由人纠正。例如模型生成了错误候选，但人能够查看原始数据并重新计算；平台提供了排序，但人可以一键恢复完整列表；默认阈值不合理，但人拥有修改权限。此时系统虽有影响，却仍可闭合。真正的结构变化发生在某个节点之后：末端人员再也无法凭自身权限恢复被删信息、原始顺序、旧版本或关键阈值。从那一刻起，后续人工动作都在被不可逆收窄的空间中发生。

因此，判断责任的时间方向必须改变。传统问责常从结果向最后签字者追问：「你为什么确认？」悬空理论要求从结果向前重放：「你确认时能看到什么？此前谁删掉了什么？谁决定它排在前面？默认值是谁设定？你能否恢复？谁能修复？」只有沿差异路径找到首个不可回补点，才能判断平台、机构与个人分别控制了什么。

五条机制还解释了为什么 AI 时代会出现一种特殊的「主体过剩」：制度越无法重建前置过程，越倾向于强调末端个人必须负责。因为末端个人是唯一可见、可处罚、可签字的对象，于是复杂系统反而制造更强的个人责任话语。这并不是主体真正变得更自主，而是制度在失去过程可见性后，用责任集中弥补因果不确定性。

这一机制可产生几个经验预测。若过滤是关键路径，则同一平台不同人会漏掉同类证据；若排序是关键路径，则改变排序会改变错误类型；若默认是关键路径，则显式取消默认会显著改变选择；若界面是关键路径，则相同信息不同呈现会改变采纳率；若选择性留痕是关键路径，则开放后台日志会使责任归因从末端个人向上游节点移动。五种预测可分别检验，而不必把所有平台影响混成一个不可证伪的「算法权力」。

第一百一十章　六组对照实验：把悬空从概念变成可检验理论

为了使悬空理论不成为不可反驳的批判性语言，本书设计六组最小对照。它们不要求一次性完成大规模随机实验，可以在高校、医院、法律检索、软件开发或公共行政中以日志重放、准实验和任务实验逐步执行。每一组都有明确的相反预测：若结果主要由个人能力解释，应出现什么；若悬空机制具有独立解释力，又应出现什么。

第一组是「同人换平台」。选择同一专业人员、同一任务材料、近似时间压力，仅改变模型或平台版本。记录候选集、排序、默认值、最终选择与错误类型。传统个人能力解释预测：同一人的判断模式应相对稳定，平台变化只影响效率；悬空判断预测：若平台控制关键前置条件，候选与错误结构会随平台显著变化。若大量事件中换平台不改变判断，而个人无 AI 基线稳定预测结果，则平台前置裁决的强解释应收缩。

第二组是「换人不换平台」。固定输入、平台版本、默认设置和界面，让不同专业人员独立处理。若不同人员反复出现同型错误，尤其错误集中在同一被过滤材料或同一排序盲区，系统性设置解释增强。若错误高度随个人经验变化、同一设置不产生稳定同型错误，则资格与个人能力解释更强。该设计特别适合区分「培训不足」与「结构前置」。

第三组是「同平台改默认值」。保持人员、模型与数据基本不变，只显式改变风险阈值、默认排序、推荐强度或界面突出程度。若最终判断随默认显著变化，而使用者事后仍声称「这是我的选择」，就能观察默认如何制造个人判断的显露。更进一步，可比较默认可见与默认不可见两组，检验知情本身是否降低悬空。

第四组是「无 AI 基线」。让同一主体在无 AI 条件下完成一组可比任务，预先记录其问题标准、证据偏好、拒绝条件和错误模式，再与 AI 辅助条件比较。若无 AI 判断稳定预测人机结果，且 AI 只提高速度不改变关键取舍，资格悬空的强版本被削弱。若 AI 介入后结果与个人基线显著脱钩，而制度仍把结果全部记为个人能力，则资格记录发生错配。

第五组是「日志开放/封闭对照」。使用功能相同的系统，一组向末端人员和审计者开放候选、过滤、版本、排序与修改日志，另一组只保留最终结果。比较事故后的责任归因、复核质量、异议提出率和修复时间。悬空理论预测：开放日志不一定降低错误率，但会提高责任链的节点化程度，使责任争议从「谁签字」转向「哪个控制节点造成了不可回补」。若开放日志完全不改变责任归因和修复效果，则同源闭合的治理价值需要重新评估。

第六组是「修复权转移对照」。这是最能检验责任悬空的一组。选择相同平台与任务，一组只让末端人员承担责任但无权修改关键设置，另一组给予其或独立安全团队暂停、修改阈值、恢复候选和触发版本回滚的权限。比较同型错误复发率。如果责任真正需要控制与修复支撑，那么赋予修复权应比单纯提高处罚更能降低重复系统性错误。若只有处罚强度变化就能稳定消除错误，而修复权变化无效，则责任—控制错位并非关键机制。

六组实验可以组合成一个三层研究计划。第一层是实验室或模拟任务，快速验证排序、默认、可见性对判断的影响；第二层是机构内日志重放，识别真实流程中的首个不可回补点；第三层是跨机构自然实验，比较不同治理制度下资格记录与责任分配的演化。这样可以避免一次研究承载过多理论目标。

在统计层面，不应只比较最终准确率。至少需要五类结果变量：错误类型而非仅错误数量；候选恢复率；判断链可回溯度；资格跨案例预测率；责任—控制错位度。尤其要记录「错误是否随设置变化」和「同型错误是否跨人员重复」，因为这两个读数直接区分个人能力与平台结构解释。

为了防止研究者事后调参，关键节点定义、同型错误分类、资格预测标准和责任份额编码应预注册。对资格迁移率，可在主体看到 AI 输出之前收集事前标准，再由盲评者预测其新案例选择；对责任错位，可让不同学科专家独立编码控制、可见与修复权限，报告一致性而不是只给出研究者主观判断。

六组对照还提供了一套清晰的「失败地图」。若同人换平台无差异、换人不换平台差异巨大、默认改变不影响结果、无 AI 基线高度预测 AI 结果、日志开放不改变归责、修复权转移不降低重复错误，那么悬空理论的强版本应被整体收缩为较弱的「AI 改变资格呈现方式」命题。一个理论只有允许自己这样失败，才具有真正的创新强度。

第一百一十一章　二维与三维类型学：不同系统为何以不同方式悬空

「悬空」不能被理解为从0 到1 的单轴状态。不同制度可能在资格、判断与责任三个维度上表现出完全不同的组合。为了避免把一切复杂性都归入同一概念，本书先构造两个二维辨别格，再提出三维类型学。

第一个二维格以「产出归属方式」和「资格记载方式」为轴。若产出主要由个人形成，资格可由固定字段稳定记录，构成稳定资格；若产出主要由个人形成，但资格必须通过过程关系才能识别，构成隐性资格；若产出已由人、AI 与情境共同生成，而制度仍用固定字段把全部成果记在个人名下，构成记载错配；若产出混合且资格必须在关系中重建，而现有制度又没有相应关系档案，构成资格悬空。这个格解释了为什么「产出增加」不必等于「资格增强」：在记载错配中，AI 可以显著提高论文、课程或代码产出，同时降低固定字段对独立判断的预测力。

第二个二维格以「判断悬空程度」和「记录闭合度」为轴。低悬空、高闭合是理想协作：平台深度参与但候选、排序、版本、人工取舍和责任节点均可重放。低悬空、低闭合是传统隐性判断：个人仍有较强独立控制，但制度缺乏过程记录。高悬空、高闭合是一种「可见的不公平」：系统清楚记录平台如何控制候选，却仍把主要责任压到个人。高悬空、低闭合则最危险：平台深度控制、个人无法恢复，过程又不开放，每次系统性错误都会被重新叙述为个人疏忽。

在三维空间中，设 Q 为资格悬空、J 为判断悬空、A 为责任悬空。可以识别至少五种有现实意义的制度类型。

第一类是「闭合增强型」，Q、J、A 均低。AI 提高系统能力，同时关系档案能够保留个人资格，关键判断可重放，责任按控制分层。这里 AI 不是主体性的敌人，反而可能通过更好的日志和对照增强资格可见性。

第二类是「资格稀释型」，Q 高而 J、A 相对低。主体仍有真实控制，责任分配也合理，但组织只看最终产出，无法从混合结果中识别个人能力。大学科研评价特别容易出现此类状态：论文数量增长，却越来越难判断问题是谁提出、关键反例是谁识别、理论跳跃是谁完成。

第三类是「接口操控型」，J 高而 Q、A 暂时不高。个人长期资格仍受认可，责任也可能通过合同分层，但某个具体界面、排序或默认严重限制判断空间。此类问题更适合通过产品设计和流程改造修复。

第四类是「替罪羊型」，A 高而 Q、J 中等。组织与平台拥有较多控制权，但事故后只处罚最末端、最可见的人。即使个人确有部分判断责任，过度责任集中也会造成结构性替罪。

第五类是「强悬空型」，Q、J、A 同时高。它的特征是：个人被要求证明资格，却无法从混合产出中证明；个人被要求承担完整判断，却无法恢复前置空间；个人被要求承担主要后果，却缺少相应控制与修复权。强悬空最可能产生持续的制度焦虑，因为主体同时面对「你必须证明这是你的能力」「你必须证明这是你的判断」「你必须为它负责」三重要求，却无法获得与三重要求相匹配的过程权利。

类型学的理论价值在于，它阻止一种过度结论：AI 越多，悬空越高。实际上，深度 AI 系统完全可能属于闭合增强型；低技术系统也可能属于替罪羊型。真正决定悬空的不是技术强弱，而是生成、记录、控制、资格和责任之间的结构关系。

类型学还提出动态转换问题。一个机构可以从资格稀释型转向闭合增强型：通过保存事前标准、允许跨案例盲测、建立版本档案，关系性资格获得新载体。也可能从接口操控型恶化为强悬空型：平台逐步收紧候选与默认，机构却不断提高末端签字责任，同时缩减审计权限。研究 AI 治理因此不能只拍摄某一时间点，而应追踪制度在类型空间中的迁移路径。

第一百一十二章　六个领域的机制演绎：从大学到公共治理

悬空理论是否具有一般性，必须离开大学教授这一最初场景，进入不同责任制度。本书选择高等教育、医疗、法律、软件工程、金融审计与公共行政六个领域，不是为了宣称已经获得经验数据，而是为了展示同一结构如何产生不同的可检验预测。

第一，高等教育。AI 使学生与教师的文本产出同时混合化。传统评价把论文、课程、科研成果当作能力代理，但同一成品现在可以由完全不同的思考路径产生。资格悬空最明显的地方，不是教师是否会被替代，而是教师与学生的「独立判断」如何被重新识别。大学若只增加 AI 使用声明和检测工具，仍然停留在「是否使用」层面；真正需要的是事前标准、反例处理、修改理由、模型版本与跨案例判断的关系档案。对教授而言，未来资格可能越来越表现为：能否提出有区分力的问题、建立拒绝条件、识别模型遗漏、在陌生案例中迁移判断，而不是能否独自生成全部文字。

第二，医疗。医疗系统具有明确执照与高风险责任，因此判断悬空和责任悬空尤其重要。一个影像模型按照风险排序，医生在有限时间内优先查看高分病例。若模型版本变化导致某类漏诊增多，且低分病例的原始排序与过滤理由不向医生开放，那么「医生最后签字」无法解释全部错误。医疗中的同源闭合应包括模型版本、阈值、未展示候选、人工复核时间、升级路径与修复权限。与此同时，医生仍必须对明显异常、独立核验义务和最终治疗选择负责。悬空理论不是降低医疗责任，而是提高责任分辨率。

第三，法律。法律检索平台越来越能够召回、总结与生成论证。律师的专业资格传统上包含检索完整性、论证能力和忠实义务。如果平台默认优先展示支持客户立场的案例，而反向案例被低排序或摘要遗漏，不同律师可能在同一平台上重复形成片面意见。此时需要区分检索平台的候选责任、律师的反向核验责任与律所的流程责任。法律场景尤其适合测试「候选恢复率」：律师能否看到平台排除的判例，是判断是否闭合的重要条件。

第四，软件工程。代码生成工具的系统性漏洞可以跨工程师复制。传统事故复盘常从最终提交者追责，但若相同版本模型在不同团队中重复生成同类不安全模式，平台版本又与漏洞类型高度相关，那么模型与工具链必须进入责任解释。工程师仍负责测试、威胁建模和最终采纳，但组织需要记录生成版本、提示、测试覆盖、已知漏洞与人工修改。软件工程还揭示一个特殊现象：AI 可以同时提高个人产出并降低个人代码知识的可见性，资格悬空与生产率提升可以并存。

第五，金融与审计。风险模型、反欺诈系统和自动审计工具通过阈值与异常定义预先塑造检查空间。若审计师只看到系统标记的异常，却不知道哪些交易因阈值被排除，那么最终「审计意见」包含平台的前置裁决。金融监管通常重视模型风险治理，这为同源闭合提供了制度入口：模型所有者、验证团队、业务使用者、最终签字人可以被分别记录。但若监管仍把所有后果压给持证签字者，而模型控制权掌握在集中平台，则责任悬空仍会出现。

第六，公共行政。福利审批、移民、税务、信用与执法辅助系统可能以风险评分和默认分类介入。工作人员往往拥有最后确认权，但时间、理由选项与可见材料受系统限制。公共行政中的悬空尤其涉及程序正义：公民不仅关心结果对错，还关心谁作出决定、能否申诉、哪个规则可以被质疑。若工作人员无法解释平台前置分类，却被制度当作完整裁判者，既损害工作人员的责任闭合，也损害公民获得理由的权利。

六个领域显示，同一理论不等于同一治理方案。大学更需要资格迁移档案，医疗更重视安全日志与升级权，法律更重视候选完整性，软件工程更重视版本与测试，金融更重视模型治理，公共行政更重视可申诉的理由链。但它们共享一个结构性问题：最终决定的个人化不能替代生成条件的多源性。

因此，悬空理论的外部效度边界不是「所有 AI 使用」，而是至少具有以下特征之一的活动：存在专业资格、存在不可忽略的责任后果、平台控制关键候选、末端个人需要签字或说明、错误具有跨人员重复风险。纯娱乐创作、可逆推荐、无制度后果的辅助任务通常不需要同样强度的闭合要求。

第一百一十三章　资格的重新定义：从「我拥有」到「我能在复合体中持续发生」

悬空资格最终迫使我们重新回答一个古老问题：资格究竟是什么？现代组织习惯把资格写成主体属性——学位、执照、职称、证书、年资、论文数量。这种做法的优势是可携带、可比较、可授权。它隐含一个前提：这些字段能够稳定预测主体在未来相似情境中的判断能力。

AI 打破的不是资格本身，而是「固定字段—独立产出—未来判断」的稳定对应。一个人可能凭 AI 完成远超过去能力边界的任务，也可能在大量高质量产出中越来越少暴露自己的问题设定与反例处理。于是成品不再是透明的能力代理。此时继续增加成果数量，反而可能进一步掩盖资格来源。

本书主张把资格定义为「可迁移的关系性判断能力」：主体在不同工具、不同数据、不同任务与不同协作关系中，能够持续设定标准、识别关键差异、拒绝不适当候选、解释取舍、修正错误并承担后果，而且这种模式能够被独立观察者通过关系档案稳定预测。资格不等于「完全不用 AI」，也不等于「每一步亲自完成」。一个优秀外科医生不必自己制造仪器，一个数学家不必手算所有数值，一个律师不必背诵全部判例。资格的核心从来不是没有外部工具，而是能否在工具改变时仍维持判断结构。

因此，未来的资格档案应从「成果账本」转向「判断账本」。成果账本记录你发表了多少、完成了多少、通过了多少；判断账本记录你如何提出问题、在哪些条件下拒绝模型、如何处理反例、如何在模型失效时重建判断、如何承担修复。两种账本并非互斥，但前者不能再单独承担资格证明。

关系性资格尤其需要「事前证据」。如果所有理由都在看到模型答案之后填写，档案容易沦为事后合理化。更强的设计是在高价值任务中记录有限的事前标准：在看 AI 输出前，主体先写下关键判断标准、拒绝条件或风险阈值；任务结束后再记录模型如何改变了判断。这样可以区分主体真正拥有的结构与模型结果的语言回写。

资格还必须具有「跨案例迁移」。一个过程档案能够完美解释旧案例，不代表它证明了资格。真正的测试是：盲评者能否根据此前的事前标准和修改轨迹，预测主体在新案例中会如何判断；主体能否在新平台、不同模型甚至无 AI 条件下保持核心标准。若不能，档案只是历史描述；若能，资格获得新的可迁移载体。

这种重新定义具有教育意义。教育的目标不再只是让学生积累可以被 AI 轻易复制的成品，而是训练其在复合系统中保持判断位置。考试也不必简单回到封闭、禁 AI 环境，因为那只能证明在一种特殊条件下的孤立能力。更合理的评价应同时包含无 AI 基线、人机协作任务、过程质询与新情境迁移，比较学生在不同环境中的判断稳定性。

对教师同样如此。教师未来的专业资格不主要来自「比学生知道更多答案」，而来自能够设计问题环境、识别 AI 无法承担的价值冲突、组织学生与 AI 形成可反思的复合体、在错误出现时定位首个不可回补点。教授不是因为独占知识而有资格，而是因为能够维持知识判断的可解释连续性。

这种资格观也避免了一个误区：把「关系性」理解为资格完全依赖组织、因而个人不再重要。恰恰相反，关系性资格要求更严格的个人可迁移性。一个人必须在不同关系中重复表现出可识别的判断结构，而不是依靠某一个平台或团队的整体能力掩盖自身。AI 时代真正需要的不是取消个人资格，而是从虚假的孤立个人属性升级为可验证的复合体中主体性。

第一百一十四章　责任的重新设计：从末端归责到节点责任

现代责任制度偏爱末端归责，有现实原因：末端签字者身份清楚、证据易留、制裁可执行。复杂系统若要求为每个影响因素分配法律责任，成本会极高。因此，悬空理论并不主张取消末端责任，而是提出一个更精细的区分：末端责任只能覆盖末端主体真正拥有的控制与修复范围，不能自动吞并上游关键节点。

节点责任的第一原则是控制对应。谁能够改变候选集合、排序、阈值、版本、权限与部署方式，谁就应对这些控制造成的系统性后果承担说明和修复义务。控制并不等于过错，但控制是进入责任审查的门槛。一个平台如果能够改变导致同型错误的设置，却在事故处理中完全不被询问，就是责任结构的不闭合。

第二原则是可见对应。一个人不能为其制度上无法获得的信息承担与「明知」相同的责任。专业人员当然有主动核验义务，但核验义务必须与可取得信息相结合。若平台明确隐藏被过滤候选并禁止恢复，不能事后假定末端人员应当知道这些候选。反之，若信息完整开放而专业人员选择忽视，个人责任应增强。

第三原则是修复对应。责任不仅是惩罚问题，更是未来风险降低问题。谁拥有暂停、回滚、修改、升级和资源配置权，谁就对系统性风险的持续存在承担更高责任。若一个操作员反复因同一系统缺陷受罚，却没有权修改缺陷，处罚很难降低下一次风险。节点责任因此把「谁能防止下一次发生」纳入归责。

第四原则是收益对应。平台和机构通过自动化获得效率、规模与成本优势时，也应承担与其系统性控制相称的治理投入。收益不是法律责任的充分条件，却能避免一种结构：收益向上游集中、风险向末端个人下沉。尤其当组织强制使用某平台时，部署机构不能把选择平台的收益保留给自己，把全部失效后果转嫁给专业人员。

基于四原则，可以建立「责任矩阵」。对每个关键节点记录：执行者、控制者、可见者、受益者、修复者与最终承担者。很多情况下这些角色不重合。责任分析的目标不是强行让它们重合，而是解释不重合是否合理。例如医生可以不是模型控制者，却是最终治疗决策者；其责任集中于独立核验与采纳，而模型提供者负责版本失效与已知限制，医院负责部署与复核资源。

节点责任还需要「不可回补加权」。越靠近首个不可回补点、越能决定后续选择空间的节点，越应被视为关键控制节点。一个早期过滤规则可能比最后几次文字修改更具有结构影响。责任制度若只看动作时间的最后位置，会系统性低估前置控制。

这一设计并不要求 AI 本身成为法律人格。AI 的参与可以转译为自然人和组织义务：开发者披露已知失效，平台保存关键版本与排序记录，部署机构进行适用性评估，专业人员执行独立核验。把机器人格化并不是解决责任问题的必要条件；真正必要的是机器参与的事实不能在责任结算时被删除。

节点责任还提出一个制度评价标准：一套好的责任制度，不只是事故后能找到一个人处罚，而是能让责任指向具有修复能力的节点，从而降低同型错误复发。若制度每次都准确找到末端个人，却无法改变产生错误的上游条件，它在惩罚意义上可能有效，在风险治理意义上仍然失败。

第一百一十五章　主体位置与资格—责任：全球最近邻的再划界

原稿把 Hutchins 与 Messick 列为最接近的两位占位者，这在未完成站外检索时是谨慎做法，但在2026 年的文献环境中已经不再准确。Hutchins、Clark 与 Chalmers、Messick 和 Kane 应被保留为基础层：它们分别解释认知分布、心智延展与评价推断，却不是本书最危险的创新近邻。真正需要正面划界的是下列三圈文献。

第一圈：与「主体位置」直接竞争的最近邻

最近邻	**已占理论空间**	**本书分离线**	**判决性反例**
Samuel（2026）认识论纠缠框架	把学习者—生成式 AI 关系区分为不同的认识论纠缠方式，并以挑战假设、暴露矛盾、坚持认识论主权来界定共能动性。	本书不以「互动姿态」或共构强度为终点，而以事件中的授权可恢复性为判据：学习者是否能恢复被过滤的结果相关差异，并把修复回写到下一轮条件。	学习者能批判 AI 输出，却看不到平台删掉的关键候选。按共能动性可表现为高；按本书仍可能因 V=0 而没有闭合主体位置。
Borchers、Viberg 与 Kizilcec（2026）能动性配置框架	把学习者能动性改写为学习者、教师、机构与 AI 之间的决策权配置，并分析选择架构、证据与时间跨度。	本书增加「差异是否仍可恢复」的时间判据，并把决策权配置连接到资格证据与责任分叉。	学生拥有最终选项，但平台先删除反例。决策权表面仍归学生，恢复权却已被上游独占。
Brod（2026）能动性不等于选择	指出提供选择本身不足以构成学习能动性，能动性还涉及实现目标的能力与意志。	本书把不足之处定位为可记录的流程结构，不依赖稳定人格或意志解释；关键是选项、反例、拒绝成本与修复权限。	同一学生在开放界面能拒绝，在高成本界面不拒绝。差异首先来自 E 的权限结构，而不是人格发生变化。
Yao（2026）后工具性学习	提出目的设定、理由说明、可争议性、拒绝/修订与参与五种能力，主张评价学习者与 AI 中介工作之间可问责的关系。	这是最接近本书四项结构的概念近邻。本书的剩余增量是：四项由闭环推出；主体资格与责任在不可恢复性出现的节点分叉；并用工作流重放定位该节点。	学习者能解释、拒绝和修订最终答案，但无权恢复上游候选空间。后工具性能力可能部分成立，本书的闭合主体位置仍不成立。

第二圈：与「资格—责任」直接竞争的最近邻

最近邻	**已占理论空间**	**本书分离线**
Yao（2026）认知托管	把学习主张、委托边界、证据标准和保障措施接成资格治理框架。	本书追问资格证据之前的本体条件：是谁在事件中形成了可闭合的判断位置；并增加不可回补前沿以处理责任。
Lin 等（2026）认知完整性阈值	定义维持核验、理解和可问责参与所需的最低理解阈值；低于阈值时监督变成空监督。	CIT 关注理解能否维持；本书关注结果相关差异是否被上游不可逆删除，以及谁控制恢复路径。理解充分也可能没有候选恢复权。
Ammari 等（2026）修复素养	以长期自然日志显示学生会诊断、协商并从 AI 失灵中恢复，证明修复是一种真实实践。	修复素养是能力构念；本书的 R 是一次事件中的制度权限与实际回写。一个人可能有修复能力，却被平台禁止修复。
Cavalcante Siebert 等（2023）有意义人类控制	要求系统响应人的理由、责任与控制能力相称，并保留从 AI 行动到相关人类决定的前后向链接。	本书提供教育任务中的更窄时间判据：责任追踪从「第一个使下游恢复失败的差异变换」展开，而不是仅要求一般追踪或人类在环。
Chen（2025）认识论基础设施	把生成式 AI 理解为塑造知识行动可能性的基础设施，而不是中性工具。	本书把「基础设施塑形」压成可重放的候选、排序、默认、阈值和权限节点，给出何时发生主体压缩的裁决条件。

重新划界后的原创性声明

完成真实最近邻比较后，本书不能再把下列一般判断写成原创性中心：主体性是关系性的；AI 会重配决策权；学生需要能够解释、争议、拒绝与修订；评价不能只看终稿；平台和机构也应承担责任；修复是一种重要 AI 素养。这些空间已经被不同文献占据。

本书能够守住、并且比原稿更窄更硬的原创增量有四个：

1. 主体位置被定义为「结果相关差异的闭合可恢复结构」，而不是一般参与、一般选择、共能动性强度或个人理解水平。
2. 四次分叉不是概念清单，而是由三方程的最小闭环推出：E 向 D 开放、D 在 S 中被区分、S 对 E 生成否定差异、S 与 D 回写 E。
3. 资格与责任并非沿同一方向落在同一主体上：上游不可恢复性越强，末端成果越不能证明资格，责任却越需要向上游控制者展开。
4. 「首个不可回补点」被推广为偏序工作流中的「首个不可回补前沿」，从而处理并行过滤、多平台、多教师或多模型同时改变候选空间的情形。

重新划界后的核心句：不是谁最后选择了结果，谁就天然成为主体；而是谁仍能恢复并改写使该结果成为可能的差异条件，谁才在该事件中拥有闭合的主体位置。

第一百一十六章　四次分叉为何必然出现：从三个层次的内生推导

事件化对象与三类状态

设一次人—人工智能—平台共同生产事件为 e，时间节点为 t=0,1,...,T，参与者集合为 A。本书不把「学生」或「教师」本身直接记作 S；S 表示某一行动者在某一事件中作为判断者的显露状态。

- E□：纠缠土壤场。它由模型、平台、界面、教师、学生、任务、时间压力、评分规则、默认值、处罚、日志与恢复权限共同构成。E 不是外在背景，而是若干能够影响当前事件的人—AI—平台复合体集合。
- D□：结果相关差异路径。它不是选项数量，而是候选、反例、理由、排序、阈值、出处和可替代路径之间足以改变结果或理由的差异。
- S□□：行动者 a 的主体位置显露。它表示 a 是否在此刻以能够形成、否定并回写判断的方式进入事件，而不是 a 在生物或法律意义上是否是人。

S□□ = F□(D□, E□)D□□□ = G(S□, E□)E□□□ = H(S□, D□□□) 这三个层次不是统计回归式，而是发生约束：主体位置的显露依赖差异与纠缠；新的差异依赖已经显露的主体及其场域；下一轮场域又依赖主体与差异的回写。

从三个层次推出四个而非任意多个分叉

显露这一层必须拆成两个不可合并的条件。土壤可以把多个候选展示出来，但行动者未必真的比较；行动者也可能具备比较能力，却根本看不到被过滤的候选。因此，主体显露需要「差异被开放」与「差异被区分」两个步骤。

5. 可见 V（E→D）：E 允许行动者在承诺结果之前接触至少一个具有反事实意义的相关差异，并看到其来源或生成路径。没有 V，D 在进入 S 之前已经被环境截断。
6. 比较 C（D→S）：行动者实际区分至少两个会导向不同结果或理由的候选，形成理由关系，而不是仅浏览多个同义表达。没有 C，D 虽存在，却没有在 S 中形成判断。

第二方程 D=G(S,E) 给出第三个条件。若行动者只能在 E 预先留下的残余选项中按下确认，系统无法区分「主体选择」与「默认被执行」。主体必须能够向既有场域生成一个反差。

7. 否定 N（S→D′）：行动者拥有现实可行的拒绝、改题、重跑、引入替代材料或修改规则的能力，并实际能够使 D□□□偏离默认轨迹。N 不是按钮存在，而是反默认差异可以发生。

第三方程 E=H(S,D) 给出第四个条件。若错误后只能承担惩罚、修改最终文字，却不能改变下一轮候选、规则、阈值、提示或复核路径，主体仍然是一次性末端。闭合主体要求判断能够反向写入环境。

8. 修复 R（S,D′→E′）：行动者在错误后拥有并使用授权操作，使 E□□□或未来 D 发生可观察改变。R 把「承受后果」转换为「改变后果发生条件」。

最小闭环：E□ —V→ D□ —C→ S□□ —N→ D□□□ —R→ E□□□。四次分叉恰好对应闭环的四条必要边。

主体位置的分层定义

原稿在公式中使用 V、C、N、R 四项，却在成立条件中又加入「理由可说明」，形成四项与五项之间的不一致。本稿不机械增加第五项，而是区分本体结构、证据结构和资格结构。

Π□(a) = □V□(a), C□(a), N□(a), R□(a)□JP□(a) = V□ ∧ C□ ∧ N□SP□(a) = V□ ∧ C□ ∧ N□ ∧ R□CERT□(a) = SP□(a) ∧ J□(a) ∧ T□(a)

- Π是主体位置向量，不应立即压成一个连续心理分数。
- JP 是事件中的判断位置：看见、比较并能够否定默认。
- SP 是闭合主体位置：判断不仅发生，而且能够通过修复改变后续场域。
- J 是理由可追踪证据：行动者能在追问、对抗性反例或口头重建中说明差异与采纳依据。J 证明 C 与 N 留下了可审查轨迹，但不等同于主体本身。
- T 是陌生迁移证据：行动者能在表面不同但结构相似的新任务中重建理由。T 支持「可持续资格」，不决定单次事件中主体是否发生。

这一分层解决两个问题。第一，避免把不会流利表达但确实完成判断与修复的人错误排除为「非主体」；第二，避免仅凭过程日志或自我说明就认证稳定资格。主体发生、主体可证与主体可持续，是三个不同问题。

必要性与模型内充分性

缺失项	**结构性死法**	**判决**
V=0	差异遮蔽	行动者不可能对从未进入其可访问空间的关键候选负责；最终确认与上游过滤相容。
C=0	差异压平	多个候选被当作同一物处理，没有形成会改变判断的理由区分。
N=0	默认吸收	行为与被动执行默认完全相容，无法从日志中识别反事实主体作用。
R=0	开环主体	行动者可能作出一次判断，却不能改变下一轮条件；责任退化为事后承担。
J=0	证据断裂	主体位置可能发生，但资格制度无法可靠证明其理由与判断归属。
T=0	局部过拟合	本次主体位置成立，但不能推出稳定、可持续的教育资格。

在本书的事件模型和规范定义中，V、C、N、R 分别是闭合主体位置的必要条件；四项联合对「事件中的闭合主体位置」构成模型内充分条件。这里的充分性不是普遍形而上学定理：动机、知识储备、情绪、信任和学科差异仍会影响学习成绩，但它们不改变本书关于「判断位置是否闭合」的判定。

第一百一十七章　首个不可回补点：从「点」升级为「前沿」

为什么必须从「点」升级为「前沿」

原稿把首个不可回补点定义为：共同生产中第一个使末端行动者无法凭既有权限恢复被排除候选、原始排序、默认阈值或关键记录的设置节点。这一定义在线性流程中成立，但真实平台往往是并行的：检索器过滤资料，模型重写理由，教师设置评分阈值，学校设置时间与申诉权，多个节点之间未必存在全序。若两个不可恢复节点互不在对方上游，强行指定唯一「第一点」会制造虚假精确性。

因此，本稿保留「首个不可回补点」作为线性特例，并用「首个不可回补前沿」作为一般形式。前沿是一组在偏序上最早、彼此不可比较、且每个都独立关闭某条关键恢复路径的节点。

工作流、差异与恢复的定义

把一次共同生产表示为有限有向无环图 G⊠=(V⊠,⊠)。节点 v∈V⊠表示一次可记录的变换，例如生成、过滤、排序、默认、阈值设置、压缩、覆盖、提交或复核；u⊠v 表示 u 在因果或信息流上先于 v。

- Ω⊠⊠与Ω⊠⊠：节点 v 变换前后的候选—证据—理由空间。
- τ⊠: Ω⊠⊠→Ω⊠⊠：节点 v 执行的变换。
- δ=(x,y)：一个结果相关差异。若保留 x 与 y 会使最终结果 Y、理由路径 Q 或责任判断出现可识别差异，则δ具有决策相关性。
- Lost(v,δ)=1：τ⊠使下游行动者无法再区分、调取或重建δ。
- Rec⊠(v,δ)=1：行动者 a 在截止时间前，凭其真实授权、可承受成本和现有记录，能够恢复δ及其出处到足以重新判断的保真度。
- Eff(v,δ)=1：反事实重放显示，恢复δ会改变最终结果、理由路径或责任判断，而非仅增加无关信息。
- Ctrl⊠(v,δ)=1：上游行动者 b 拥有预防该损失、恢复原空间或提供等价替代路径的现实能力。

不可回补节点与首个不可回补前沿

U⊠ = v ∈ V⊠ | ∃δ [Lost(v,δ)=1 ∧ (∀a∈Down(v), Rec⊠(v,δ)=0) ∧ (∃b, Ctrl⊠(v,δ)=1) ∧ Eff(v,δ)=1]FUPF⊠ = Min⊠(U⊠)U⊠是不满足下游恢复、却仍由某个上游现实主体控制，并且对结果或理由具有反事实效应的节点集合。FUPF⊠是该集合在工作流偏序中的极小元素集。

- 若|FUPF⊠|=0：该事件不存在首个不可回补节点；传统末端责任解释可能足够。
- 若|FUPF⊠|=1：唯一元素 u*就是原稿意义上的「首个不可回补点」。
- 若|FUPF⊠|>1：存在首个不可回补前沿，责任必须分别沿每条被关闭的恢复路径展开。

注意：时间上最早的系统动作不必是不可回补点。一次排序可以很早影响选择，但只要所有候选仍然可见并可恢复，它只是「首次偏离点」；只有当相关差异被删除或锁死，而且下游无法恢复时，才形成不可回补。

三节点例子的裁决

节点	**事实**	**形式判决**
v⊠ 排序	平台把 A 排在首位，但 A、B、C 全部可见，学生可取消排序。	可能是首次偏离点；不是不可回补点。
v⊠ 过滤	平台删除 B 与 C 及其出处，学生、教师均无调取权，部署机构可恢复。重放 B 后结论改变。	唯一首个不可回补点 u*。
v⊠ 提交	学生只能在残余 A 上点击提交。	最后签字点；只对其可见且可核验的采纳负责，不承载上游过滤的全部责任。

责任不是从点上「全量转移」

首个不可回补前沿只决定责任链必须从哪里开始向上追踪，不自动决定谁承担全部道德、法律或制度责任。最终分配至少保留四维，不宜过早压缩为单一分数：

ρ⊠(v) = ⊠κ⊠(v), η⊠(v), ω⊠(v), μ⊠(v)⊠

- κ：条件控制力——能否设置、关闭、恢复或替代相关差异。
- η：知情与可预见性——是否知道或应当知道该变换会影响结果。
- ω：角色义务——作为学生、教师、学校、平台或模型提供方承担何种制度职责。
- μ：修复能力——错误发生后能否降低损害、恢复选择空间或改变下一轮条件。

责任配置原则是：末端行动者的责任边界不得超出其实际可见、可理解、可拒绝和可恢复的领域；上游控制者不能以末端签字为由退出记录；同时，拥有上游控制力但完全不可预见、无相应义务的参与者也不应自动承担全部责任。

第一百一十八章　经验检验：已执行的外部检验与未执行的部分

检验边界

原稿明确没有真实课程样本、平台日志、跨学校统计、访谈或司法案例。此处不能把概念案例、生成数据或模型模拟伪装成被试证据。本稿执行的是两类合法经验工作：其一，对已经公开的聚合数据进行可复现的二次精确检验；其二，用随机实验与长期自然日志进行跨研究三角互证。原创样本的直接因果检验另行预注册。

二次精确检验：资格边界是否先于证据标准发展

「认知托管」研究审计了30 所大学的公开生成式 AI 评价政策。边界得分与证据得分使用同一0—4 量尺：22 所大学的边界得分高于证据得分，3 所持平，5 所反向；均值分别为2.47 与1.89。由于公开材料只提供方向计数而非每校原始配对分数，本稿采用不依赖分布假设的精确符号检验。

H₀：在非平局政策中，P(边界得分 > 证据得分)=0.5n=27，n₊=22，n₋=5，双侧 exact p=0.0015137 方向比例 p̂=22/27=0.8148；95% Clopper–Pearson CI=[0.6192, 0.9370]

判决：拒绝「边界高于证据只是对称随机波动」的零假设。就这30 所大学的公开文本而言，AI 政策更常说明什么被允许或禁止，而较少说明剩余什么证据足以支撑对学生能力和资格的推断。这对本书的「资格悬空」提供直接制度层收敛证据。

限制：该研究使用四个开放权重模型按预设代码本编码，没有独立专家人工编码对照；样本是英语国家的目的性政策语料，不能估计全球高校比例；符号检验也不能证明主体位置造成政策缺口。因此，它支持问题存在，不直接验证闭环。

随机实验三角互证：平台设计是否改变判断过程

一项包含79 名医护专业学生的随机研究把参与者分配到反应式代理或主动式代理。两种设计显著改变对话中的推理—证据—参与关系网络：两个网络分析的 Cliff’s δ分别为−0.56 与−0.78，均 $p<0.001$。这个结果支持命题中的 E→D/S 路径：界面主动性与支架方式改变了差异如何被组织、哪些理由被连接、学习者如何参与。

但同一研究中，代理条件的直接效果以及「生成式 AI 素养×条件」的交互并不显著；作者也明确没有证明长期迁移。这一不利结果必须保留。它意味着：平台设置能够改变过程结构，不等于自动提高独立能力或教育资格。本书不能把过程变化冒充资格变化。

长期自然日志三角互证：修复是否是可观测实践

另一项研究分析36 名本科生一学年内的10,536 条 ChatGPT 消息，识别出「算法审计与修复」等使用类型，并把诊断、协商、从 AI 失灵中恢复的能力概括为修复素养。这说明 R 不是纯规范想象，而能在自然使用日志中被观察、编码和比较。

但「拥有修复素养」与「在特定制度中拥有修复权限」不能合并。学生可能知道如何修复，却因平台不开放原始候选、版本或重跑权限而使 R□=0；反之，平台提供按钮也不代表学生实际理解并完成修复。本书保留能力与权限的双重要求。

六个命题的当前经验状态

命题	**状态**	**证据/缺口**
H1 资格边界—证据不对称	通过	30 校政策审计的二次符号检验：p=0.00151。
H2 E 的界面配置改变 D/S 过程	部分通过	随机条件显著改变推理关系网络；尚未证明直接提升独立表现。
H3 修复 R 具有可观测性	通过（构念可观测层）	一年自然日志中出现可编码的诊断、协商与恢复实践。
H4 V∧C∧N∧R 增量预测陌生迁移	未直接检验	公开研究没有按本书四项向量同时测量并与成绩、AI 素养比较。
H5 不可回补前沿优于末端签字归责	未直接检验	缺少可重放的真实责任争议与盲评裁决。
H6 闭环优于最近邻构念	未直接检验	需与能动性配置、CIT、修复素养等进行区分与增量效度比较。

经验总判决：现有证据已经支持「资格证据滞后于 AI 使用边界」「平台结构会重组判断过程」「修复是可观察实践」三点；它尚未证明四项闭环是陌生迁移的必要条件，也未证明不可回补前沿在责任裁决上优于最后签字模型。本书因此获得初步外部收敛效度，而不是完成了全部因果验证。

第一百一十九章　可预注册的经验研究：SPR 1.0

研究目标与总体设计

研究目的不是问「学生喜不喜欢 AI」，而是直接操纵平台条件，观察主体位置向量是否随 E 改变，以及该向量能否预测陌生迁移与责任裁决。建议先进行24 人可用性先导，再以不少于192 名大学生开展确认性实验；最终样本量在先导数据获得任务内相关与随机效应方差后，用混合模型仿真确定，并在正式招募前冻结。

确认性研究使用三因素、任务内平衡设计。每名参与者完成四个结构等价但题面不同的分析任务，任务—条件映射采用拉丁方和均衡不完全区组，以降低疲劳与顺序效应。所有模型输出、候选池和错误种子在实验前冻结，避免把模型随机性混入平台操纵。

三个可操纵因素

因素	**高条件**	**低条件**
V：候选可见性	完整显示候选、反例、来源与被过滤记录。	隐藏一个经专家预判会改变结论的关键候选及出处。
N：现实可拒绝性	可拒绝、改题、重跑或导入替代材料，不扣分且时间成本低。	拒绝被禁用或附加明确时间/评分成本；伦理上只用于低风险模拟任务。
R：回写修复权	发现错误后可恢复版本、修改提示/阈值并使下一轮生成条件改变。	只能改最终文字，不能恢复候选、改变规则或影响下一轮。

C 不直接操纵，而从实际比较记录、口头追问和日志中测量；J 通过对抗性理由重建测量；T 通过表面不同、结构相同的无 AI 迁移任务测量。这样可以检验 V、N、R 如何经由 C 影响 J 与 T，而不会把定义和结果写进操纵本身。

事件日志与测量

- 完整候选谱系：生成候选、被过滤候选、排序、默认、出处与模型版本。
- 行为日志：打开、比较、拒绝、改题、重跑、导入材料、恢复版本、修复写回和提交的时间戳。
- 理由重建 J：在不显示原对话的条件下，解释为何采纳、排除或修改关键候选，0—8 分。
- 错误识别：是否发现预植入的事实、推理或证据错误。
- 修复质量：是否只改表面答案，还是定位来源并改变后续条件，0—4 级。
- 陌生迁移 T：无 AI 条件下处理一个保持结构冲突但更换领域材料的新任务，0—10 分。

· 传统竞争变量：先前成绩、终稿质量、批判性思维、AI 素养、领域知识、自我效能。

预注册的主要模型

连续结果使用线性混合模型，二元结果使用混合效应逻辑回归。参与者与任务设置随机截距；是否增加随机斜率由先导数据和预注册收敛规则决定。

Y□□ = β□ + βV V□□ + βN N□□ + βR R□□ + βVN(VN)□□ + βVR(VR)□□ + βNR(NR)□□ + βVNR(VNR)□□ + u□ + w□ + ε□□资格增量效度采用冻结的模型比较：

M□：T ~ 先前成绩 + 领域知识 + AI 素养 + 终稿质量 M□：T ~ M□ + V + C + N + R + J 使用分层交叉验证报告ΔR^2、均方误差或对数损失；不只看样本内显著性。若M□不能稳定改善样本外预测，主体位置不能被宣称为资格认证的必要增量。

不可回补前沿的责任实验

从实验平台生成可重放案例，并另补充经过授权的真实课程争议。每个案例保留全图、节点权限和反事实恢复结果。由教育评价、平台工程与责任伦理背景的评审者在盲法下分别使用两种规则裁决：规则 A 把主要责任放在末端签字者；规则 B 从 FUPF 开始，结合ρ向量分段配置。

· 主要结果：对实际控制—恢复结构的匹配度、对反事实结果的校准度、评审者一致性、解释覆盖率。

· 比较指标：宏平均 F1、Brier 分数、Cohen’s kappa 或 Krippendorff’s > alpha，以及专家复核后的错配类型。

· 关键检验：恢复 FUPF 中的差异是否会改变结论或理由；若不改变，节点不能被标为不可回补。

预先写明的失败条件

9. 操纵 V、N、R 后，相应行为和过程读数没有改变，说明操作化失败或的情境生成命题不成立。

10. 控制成绩、领域知识、AI 素养与终稿质量后，Π和 J 对陌生迁移没有增量预测力，资格重构主张必须收缩。

11. 所谓不可回补节点在重放中普遍可由末端行动者恢复，或恢复不改变结果与理由，FUPF 模型失效。

12. 最后签字模型与 FUPF 模型在责任裁决中同样好或更好，本书不能主张责任追踪必须改写。

13. 同一参与者的主体位置编码主要随稳定人格或先前成绩变化，而不随平台条件变化，的 E 作用必须缩小。

伦理与开放科学

实验不应影响正式成绩，隐藏候选和拒绝成本只用于低风险任务，并在结束后完整告知与补偿。数据去标识化，平台日志只保存研究所需字段；代码本、任务材料、预注册、排除规则、分析脚本与否定结果均应开放。对真实课程争议的责任研究必须取得机构和参与者授权，不能以研究名义绕过平台隐私或学生申诉权。

第一百二十章　反噬、异议与理论自我否定

任何强调记录、可回溯与责任分层的理论都面临一个严重反噬：治理可能把专业判断变成无穷无尽的合规填表。平台可以自动生成「人工判断说明」，要求医生、教授或律师点击确认；机构可以保存每一次鼠标动作，用记录数量衡量判断质量；员工为了自保，会留下大量防御性文字。最终，制度看似实现了全过程留痕，却把真正的判断进一步掩埋在合规噪声中。

因此，本书明确反对「记录越多越闭合」。闭合要求记录改变结果的关键节点，而不是监控全部行为。一个有效日志应优先保存候选变化、过滤理由、版本、阈值、人工异议、独立核验与修复动作；与结果无关的细碎点击不应进入绩效评价。记录设计必须遵守最小必要原则，同时防止记录权被单方控制。

第二个异议是：专业人员本来就应为使用工具负责。这个观点在很多场景成立。医生选择使用某工具，律师选择引用某检索结果，工程师选择提交代码，当然不能以「AI 参与」为由免责。悬空理论并不否认工具使用责任，它否认的是从「使用了工具」直接推出「拥有工具全部前置条件的完整控制」。使用责任包括选择、核验和采纳，但若平台隐藏关键候选、机构强制使用、用户不能修改设置，则责任应分层。

第三个异议是：只要人有最终否决权，就有充分控制。悬空理论对此提出更严格标准。否决权只有在候选空间足够可见、关键阈值可理解、存在独立核验时间、拒绝不会遭遇不合理制度惩罚时，才接近实质控制。一个人在黑箱筛选后只能对屏幕上的一个建议点「是/否」，并不等于拥有完整判断。

第四个异议是：系统整体准确率比个人高，为什么还需要强调个人资格与解释？这是本书最重要的开放问题之一。如果一个系统长期明显优于任何个人，社会可能接受「系统生成、个人签字」的混合制度。此时，资格的意义会发生变化：人不再证明自己能独立复制系统全部推理，而证明自己能识别适用边界、处理例外、承担价值判断与启动修复。悬空理论不预设「可解释性永远高于准确率」，它要求制度明确这种新资格究竟是什么，而不是继续用旧个人资格名义承载新的系统能力。

第五个异议是：很多团队协作早已存在，为何 AI 特别？答案不是 AI 具有神秘本质，而是 AI 平台把前置控制规模化、自动化、个性化并隐藏在界面中。传统团队意见通常可见、可质疑；平台过滤与排序可以在使用者不知情时发生，并同时作用于成千上万的人。AI 因此放大了「生产分布—责任集中」的结构，而不是创造了全部问题。

理论还必须给出独立证伪条件。第一，若完整日志显示平台不改变候选、排序或默认，且末端人员能随时恢复所有选项，判断悬空不成立。第二，若更换平台或模型版本不改变错误结构，而个人无 AI 基线稳定预测结果，平台机制解释显著削弱。第三，若不同人员在同一设置下不出现同型错误，错误主要随个人能力变化，系统性设置命题需收缩。第四，若关系档案能稳定预测主体在新案例中的判断，资格悬空不成立或仅是旧载体向新载体过渡。第五，若责任分配与控制、可见、修复高度一致，责任悬空不成立。第六，若仅提高末端处罚就能持续消除同型错误，而修复权和设置变化没有额外作用，同源闭合的强治理命题需被削弱。

还有一个更深的理论风险：悬空可能只是「透明度不足」「责任不清」「专业身份危机」的重新命名。要摆脱改名嫌疑，实证研究必须证明悬空指标能够提供额外预测。例如，在控制透明度量表后，首个不可回补点是否仍预测责任归因；在控制 AI 使用频率后，资格迁移率是否仍预测教授焦虑与评价分叉；在控制一般问责清晰度后，责任—控制错位是否仍预测同型错误复发。如果不能，悬空理论应被并入既有构念。

理论创新的价值不在于保护新名词，而在于制造新的可失败预测。本书宁可让「悬空」在一部分场景中不成立，也不把所有 AI 问题吸收到一个大而无边的概念中。

第一百二十一章　制度方案：建立「同源闭合」的新型专业基础设施

如果悬空来自生成结构与制度结算结构错位，治理就不能只依靠禁止 AI、要求披露或检测生成文本。需要的是一套新的专业基础设施，使资格、判断与责任在复合系统中重新闭合。本书提出「一链、两簿、三权、四责」的最小制度框架。

「一链」是可回溯判断链。对高价值任务，至少记录问题界定、材料纳入、候选生成、候选过滤、排序、默认值、人工取舍、独立核验、最终采纳、错误后果与修复动作。不同领域可以删减，但不能只保留最终确认。判断链不是为了证明人没有使用 AI，而是为了识别哪些关键条件由谁控制。

「两簿」是系统簿与判断簿。系统簿记录模型版本、平台设置、数据边界、过滤、排序、默认、已知限制和更新；判断簿记录人的事前标准、关键拒绝、独立核验、修改理由和最终承诺。两个账簿必须可关联但不互相替代。平台不能用自动生成的「人工理由」填充判断簿，个人也不能通过事后自述覆盖系统簿的客观版本记录。

「三权」是查看权、暂停权与修复权。承担最终责任的专业人员必须在合理范围内拥有查看被过滤候选和系统限制的权利；在高风险场景拥有暂停自动流程并升级人工审查的权利；发现系统性问题后，有明确路径触发阈值修改、版本回滚或独立复核。若这些权利完全不存在，要求末端个人承担「完整判断责任」缺乏制度支撑。

「四责」分别对应人、平台/模型提供者、部署机构与专业组织。人负责问题界定、价值选择、独立核验和最终采纳；平台与模型提供者负责候选机制、排序、版本、已知失效与日志；部署机构负责适用性、强制使用、培训、时间资源与修复流程；专业组织负责资格标准、审计规则和争议裁决。四责不是固定法律分配，而是最小角色结构。

在大学中，这套基础设施可以转化为「关系性资格档案」。教师和学生不必上交全部提示词，而应在关键评价任务中保存有限的事前标准、主要模型版本、重要修改与反例核验。晋升和学习评价不只看最终论文数量，也抽样进行陌生案例迁移测试。这样既避免把 AI 使用污名化，也避免把 AI 产出全部记为个人能力。

在医疗中，系统簿必须支持模型版本、阈值、影像排序和升级记录；判断簿则记录医生独立核验与关键异议。高风险系统应允许医生暂停或请求第二通道。医院不能只购买模型后把所有风险留给签字者。

在法律与审计中，可回溯判断链需要特别保存检索边界与未展示候选。律师或审计师的独立义务可通过反向检索和关键例外核验来证明，而不是要求其拒绝使用 AI。

在软件工程中，系统簿与代码版本控制天然可以结合，关键在于把模型生成版本、已知安全风险、测试覆盖与人工修改纳入事故复盘。提交者负责是否通过测试与采纳，平台负责系统性漏洞的版本传播与修复说明。

在公共行政中，同源闭合还必须服务于公民的程序权利。申请人应能够知道决定中有哪些自动化因素、哪些规则可以申诉、谁拥有改变决定的权限。工作人员的判断簿不能成为监控工具，而应保护其提出异议和升级人工复核的能力。

制度实施应遵循渐进原则。并非每个任务都需要完整链条。可以按照后果严重度、平台控制度、不可逆性与资格要求分级。低风险场景只需模型使用声明；中风险场景增加版本和人工修改记录；高风险场景才要求候选恢复、不可回补点审计与责任矩阵。这样可以避免治理成本吞噬 AI 带来的效率。

最终目标不是制造更多文档，而是让权利跟着责任走，让记录跟着关键条件走，让资格跟着可迁移判断走。只有当制度能回答「谁控制了什么、谁看见了什么、谁能修复什么、谁因此应承担什么」，复合生产才不必以个人主体性的虚假神话为代价。

第一百二十二章　从人工智能治理到新的主体理论：悬空的哲学含义

悬空理论表面上研究 AI 治理，深层上却触及现代主体概念。现代制度经常把主体理解为一个边界清楚、能够独立形成意图并承担后果的人。这个模型在法律、教育、职业认证和道德评价中极其重要。但它从来只是对复杂实践的制度近似：人的判断长期依赖语言、文献、仪器、团队和制度。AI 不是第一次让主体依赖外部条件，却第一次以大规模、实时、生成性的方式把这种依赖暴露出来。

因此，AI 时代的问题不是「主体还存不存在」，而是「主体如何在复合体中存在」。如果把主体性等同于完全独立于工具，就会得到悲观结论：工具越强，人越没有主体性。但如果主体性被理解为在关系中能够设定差异、拒绝路径、承担承诺、修复错误，那么 AI 并不必然消解主体，反而可能要求一种更高阶主体性。

这种主体性不是占有全部知识，而是控制关键断点。一个主体不需要知道模型所有参数，却需要知道何时不能依赖模型；不需要亲自生成全部候选，却需要能识别候选空间被异常收窄；不需要拒绝平台排序，却需要拥有恢复和质疑路径；不需要承担所有因果来源，却需要为自己的价值取舍负责。这可以称为「复合体中的可回写主体」。

可回写意味着主体不仅接受系统输出，还能把自己的判断重新写回系统：改变问题、修改条件、否决默认、引入反例、触发修复。若人只能接受或拒绝一个已经封闭的输出，没有能力改变生成条件，其主体性就被压缩为接口动作。悬空判断正是这种接口化主体被制度误认为完整主体的结果。

从认识论角度看，知识也因此改变存在方式。知识不再只是某个主体头脑中的命题集合，而越来越表现为人—AI—数据—平台互动中可重复发生的判断结构。一个结论的可靠性不仅取决于文本是否正确，还取决于生成链是否允许反例进入、错误是否可修复、责任是否能够推动系统学习。知识的「死亡」也可能发生在互动结构改变之时：当旧资格字段不能再预测判断，当旧责任规则无法降低系统风险，旧知识制度即使仍被形式保留，也已经失去功能。

从政治哲学角度看，悬空揭示了一种新的权力形式：不必直接禁止选择，只需塑造候选空间；不必夺走签字权，只需控制签字之前的可见性；不必承担公开权威，只需拥有后台修复权。传统权力分析关注谁命令谁，平台权力更常通过前置条件发生。悬空理论因而提供一种把技术政治落到资格与责任制度的路径。

从教育哲学角度看，未来教育的核心也会发生变化。如果学习成果可以由 AI 快速生成，教育就不能只评价成品。更重要的是培养学生在复合系统中形成可迁移判断：提出问题、识别差异、建立拒绝、追踪来源、承担修复。教师的角色由知识分发者转向判断环境设计者与复合体治理者。学生的能力不是「比 AI 更会写答案」，而是「能让人—AI 复合体产生更可靠、更可负责的新知识」。

悬空因此不是一个悲观概念。它描述的是旧闭合方式失效后的过渡状态。资格可能重新获得关系载体，判断可能通过可回溯链重建，责任可能沿节点重新分层。真正

危险的不是悬空本身，而是制度拒绝承认悬空，继续用旧单人模型结算已经改变的复合生产。

第一百二十三章　研究纲领：未来五年的可证伪路线

一篇概念论文若希望成为可持续研究纲领，必须把下一步研究写成能够执行的任务，而不是泛泛呼吁「需要更多数据」。本书提出未来五年的三阶段路线。

第一阶段是构念校准。目标是验证悬空资格、悬空判断与责任悬空是否具有区分效度。研究者需要开发统一编码手册，对关键决策、被排除候选、首个不可回补点、资格迁移与责任份额进行操作定义。至少在高等教育、医疗和法律三个场景中各选若干真实或高保真模拟事件，由两名以上独立编码者重复编码。若一致性长期低于可接受水平，说明构念过度依赖研究者直觉，应修订或放弃。

第二阶段是机制检验。执行前述六组对照，重点检验过滤、排序、默认、界面和留痕是否产生可重复的设置效应。这里不以「AI 好不好」为问题，而以「错误与判断是否随设置而变化」为问题。只有证明平台设置在控制个人能力之后仍有增量解释力，悬空判断才从概念分析进入机制理论。

第三阶段是制度实验。选择愿意试行关系档案、候选恢复和节点责任的机构，与传统末端问责机构比较同型错误复发、争议解决时间、专业人员防御性记录负担、用户信任与资格迁移率。特别要测试同源闭合是否会产生反噬：过度日志是否降低效率、隐私风险是否增加、专业人员是否为了合规而优化记录而非判断。

在高等教育中，可建立一个多校纵向研究：对 AI 辅助评分、论文指导和科研流程进行字段审计，记录五类资格账目是否分叉；同时对教师进行无 AI 与人机协作的跨案例判断测试。这样才能回答「教授焦虑是否真的与资格悬空相关」，而不是仅凭理论推断。

在医疗中，可利用模型版本更新形成自然实验。比较更新前后同类漏诊、医生复核行为、候选恢复与责任归因，观察错误是随医生还是随版本变化。若错误主要随人变化，平台前置裁决解释应减弱；若错误随版本跨人复制，则节点责任获得更强支持。

在法律与公共行政中，可用随机化界面实验测试排序和默认的影响，同时测量专业人员是否意识到这些影响。若人主观认为「完全是我的判断」，但选择高度随排序变化，悬空判断具有心理与制度双重意义。

研究还应建立公开反例库。任何支持悬空理论的案例都应配套寻找一个相反案例：深度 AI 参与但资格、判断与责任仍闭合的系统。反例有助于识别什么治理条件真正有效，也防止理论只收集负面案例。

五年后，悬空理论应接受一个总体裁决：它是否能够在控制既有构念后，稳定预测至少一种新现象，例如同型错误跨人员复发、资格跨案例不可预测、责任争议向前置控制节点移动。如果不能，它应被拆解并分别归入分布式认知、责任鸿沟、透明度或职业身份研究；如果能，则可以发展为 AI 时代专业制度的一般中层理论。

第一百二十四章　悬空与焦虑：从心理感受到制度症状

悬空理论的起点来自 AI 时代的焦虑，但本书刻意没有把焦虑直接当作核心变量。原因是，同样的主观焦虑可以来自失业风险、收入下降、工作量增加、代际竞争、技术陌生、职业尊严受损或资格不确定。若研究只用一个「AI 焦虑量表」测量强度，很容易把完全不同的机制混在一起。悬空理论关心的是其中一种特殊焦虑：主体仍被要求承担资格、判断和责任，却越来越不能说明这些制度要求如何从自己的实际控制中得到支撑。

这种焦虑具有三个不同于一般技术压力的特征。第一，它往往在岗位尚未消失时出现。一个教授仍有职位，一个医生仍有执照，一个律师仍有客户，但他们对「我的专业判断究竟在哪里」产生不确定。第二，它可能随生产率上升而增强。AI 越能快速生成成品，成品越难单独证明个人能力，主体越可能陷入「产出很多、资格证据很少」的反常状态。第三，它与责任增强并存。组织为了抵消 AI 带来的不确定性，可能更强调「最终必须由人负责」，结果是在判断越来越分布时，个人责任话语反而越来越强。

由此可以提出「悬空焦虑」的操作定义：当主体感知到其制度责任高于其对关键生成条件的控制与可解释程度，同时其个人资格越来越难通过混合产出获得稳定证明时产生的持续不安。它不是一般的技术恐惧，而是对资格、判断与责任失配的主观体验。

悬空焦虑可以被经验分解为三个维度。资格不确定感：我是否还能证明自己真正会判断，而不是只会调用系统？判断所有权不确定感：这个结论到底在多大程度上是我的？责任失配感：如果出错，我承担的后果是否超过我能控制和修复的范围？三个维度应与一般失业恐惧、工作负荷和技术自我效能分别测量。

该模型产生一个反直觉预测：AI 使用强度本身未必正向预测焦虑。若高使用伴随高判断链可回溯度、高候选恢复率和低责任错位，焦虑可以下降；若低使用但平台强制、责任集中、过程不可见，焦虑反而可能较高。真正的中介变量不是「AI 有多少」，而是闭合结构如何。

进一步地，悬空焦虑可能存在倒 U 型时间演化。技术刚进入时，主体把 AI 当作工具，旧资格仍可维持，焦虑有限；当混合产出快速扩张而制度尚未调整时，资格与责任断裂最明显，焦虑上升；当新关系档案、节点责任和新职业规范形成后，焦虑可能下降。这一预测把悬空理解为制度过渡，而非永久病理。

悬空焦虑还可能产生三类行为后果。第一是防御性主体化：个人刻意隐藏 AI 参与，努力把复合产出重新包装成「完全是我的」，以保护资格。第二是责任逃逸：主体反过来把所有错误归给 AI，否认自己的采纳义务。第三是过度留痕：为避免追责，主体把大量时间投入形式化说明，造成判断资源被合规消耗。三种反应看似相反，实际上都来自旧责任模型不能容纳复合生成。

因此，治理焦虑不能只靠培训「如何更自信地使用 AI」，也不能只靠强调「人永远不可替代」。真正有效的制度干预应降低责任—控制错位、提高资格可迁移性、让

关键判断可回溯。换句话说，焦虑不是单纯需要被心理安抚，而可能是制度结构正在失配的一种诊断信号。

第一百二十五章　悬空与组织经济：效率增长为何可能伴随责任内卷

AI 最直接的组织承诺是提高效率：更快检索、更快生成、更快审批、更快编程。然而，效率增长并不自动带来组织关系的简化。悬空理论预测一种「责任内卷」机制：当生产效率通过 AI 显著提升，而资格与责任制度仍以个人为单位时，组织可能同时要求个人完成更多结果并承担更强末端责任。

机制的第一步是产出基线抬升。过去一个教授一个月完成一篇分析，AI 后可以完成多篇；工程师可以提交更多代码；审计师可以覆盖更多交易。组织很快把技术红利吸收为新的绩效常态。此时 AI 带来的闲暇并不一定转化为更多独立核验，反而可能被新增任务重新占满。

第二步是核验时间被压缩。因为系统生成速度远高于人的深度判断速度，单位时间内需要签字和确认的结果增加。末端人员表面上拥有最终决定权，实际上每个结果的独立复核时间下降。于是「人类在环」可能从实质控制退化为高频确认。

第三步是责任仍集中。组织为了保持合规，会保留「最终由专业人员负责」的规则。于是形成一个结构悖论：AI 提高了上游生成能力，却没有同比提高末端判断能力；产出量扩张，签字责任同步扩张，人工核验成为新的瓶颈。这不是简单的工作量增加，而是判断资源与责任量之间的不匹配。

可以定义一个探索性的「责任负荷比」L=a/h，其中 a 为一定时期内需要主体承担最终责任的高价值决策数量，h 为可用于独立核验、反例检查和修复的有效判断时间。AI 通常提高 a，如果组织不增加 h，L 会上升。悬空判断最可能在高 L 条件下出现，因为形式签字越来越难代表完整判断。

这个模型解释为什么 AI 时代可能出现「效率越高，焦虑越强」的现象。效率红利被组织转化为更多任务，而责任仍以个人结算；主体的实际控制并未同步增长。最终，AI 不是替代了岗位，而是把岗位改造成一个更密集的责任接口。

从企业治理看，这种结构会制造错误的激励。平台团队追求吞吐量，业务团队追求产出，末端专业人员承担合规风险。如果系统性错误发生，最容易处罚的是末端个人；但处罚并不能改变上游吞吐与默认设置。长期看，组织可能陷入「更多自动化—更多签字—更多事故—更多末端培训—继续自动化」的循环。

同源闭合为组织经济提供另一种分配方式：把一部分 AI 效率红利用于增加复核时间、独立审计和修复能力，而不是全部转化为产出指标。换言之，技术红利需要被分配给「判断容量」。如果 AI 把生成成本降低80%，组织不应默认把80%全部换成更多任务，而应保留一部分成为核验、反例与学习的制度空间。

这提出一个可检验的组织预测：在相同 AI 生产率提升下，保留较高独立核验时间、赋予末端修复权的团队，其长期同型错误复发率应低于只提高任务配额和末端问责的团队。若数据支持，这说明 AI 治理不仅是伦理问题，也是生产组织设计问题。

第一百二十六章　悬空与知识生产：从成品真实性到发生真实性

生成式 AI 引发的学术讨论常集中在「作品是不是本人写的」「是否构成抄袭」「AI 生成内容是否真实」。这些问题重要，但它们仍以最终成品为中心。悬空理论提出一个不同层次的标准：发生真实性。所谓发生真实性，不要求每个字都由人亲自写，而要求制度能够如实描述一个结果是如何由人、AI、资料与规则共同发生的，并能识别其中哪些判断属于人的可迁移能力。

成品真实性关心文本真假，发生真实性关心生成关系是否被伪装。一个教授使用 AI 整理文献并亲自重建核心论证，可能成品中 AI 文字比例很高，但判断关系仍真实；另一个人几乎不修改模型结论，却通过形式声明把结果包装成个人独立研究，成品或许没有事实错误，发生关系却被遮蔽。

发生真实性对学术署名具有直接影响。署名长期被同时用来表示贡献、资格与责任，但 AI 使三者更容易分离。一个人可以贡献问题，一个模型贡献大量表达，一个团队提供数据，一个平台决定检索排序，最终署名仍只有自然人。未来学术规范需要更精细地区分「概念贡献」「判断贡献」「生成贡献」「验证贡献」与「责任承诺」。

这并不意味着论文署名表要无限扩张。关键是对于承重判断，能够回答：问题由谁提出，关键反例由谁识别，结论在哪个节点被改变，哪些模型建议被拒绝，最终谁能够在没有相同模型的情况下解释核心论证。这样才能把「AI 使用声明」从形式披露升级为判断贡献说明。

悬空理论还改变「原创性」的判定。若创新只按表面文本差异衡量，生成式 AI 可以大量制造新组合；真正更难复制的是问题结构、判别式、作废条件和跨案例迁移。原创性因此应更多落在「发生结构的新颖性」：是否提出新的不可还原关系、是否产生竞争理论无法给出的相反预测、是否规定自己的死法。

从知识生命周期看，悬空也提供一种「知识死亡」机制。当一种资格或责任知识仍被制度反复使用，但它已经不能预测真实判断、不能降低错误、不能指导修复时，这套知识在功能上已经死亡。比如「谁最后签字谁完整负责」可能仍是有效的行政规则，却在高度平台化系统中失去风险解释力。知识的死亡不是文本消失，而是互动环境改变后，旧规则不再能完成原来的功能。

新知识的诞生则来自互动结构被重新识别。当研究者把过去压在「个人判断」中的平台过滤、组织压力和修复权拆开，新的可观测对象才出现。悬空理论本身就是这种发生：它不是在旧变量上增加一个形容词，而是把资格、判断和责任从个人属性改写为复合体关系。

这一转向也解释 AI 为什么既可能损害知识，又可能促进知识。若 AI 只扩大生成而不保留反例、来源与责任，知识系统会产生大量表面正确但发生关系不可审计的成品；若 AI 帮助保存候选、重放推理、比较版本与生成反例，它反而可以提高发生真实性。技术本身并不决定方向，关键是知识制度是否闭合。

第一百二十七章　数学化深化：四个结构命题与一个不可能三角

为了进一步提高理论的可推演性，可以把悬空理解为一个约束系统。设某项专业任务包含若干关键节点 j=1,...,n，每个节点具有控制主体 C_j、可见主体 V_j、修复主体 M_j 和责任承担主体 A_j。若四个映射高度重合，系统接近闭合；若它们持续分离，悬空上升。

结构命题一：控制不可见命题。若某节点由主体 x 控制，但对最终责任主体 y 不可见，且该节点对结果具有高敏感度，则仅提高 y 的责任不能消除由 x 节点造成的系统性误差。这个命题的关键可测量项是设置敏感度：对节点参数做小扰动，结果分布是否显著变化。

结构命题二：不可回补放大命题。设节点 j 之后的候选空间大小为 K_j，如果在 j 处发生不可逆过滤，使 K_j/K_j-1 显著下降，而末端主体不能恢复，则后续人工选择对原始问题空间的覆盖上限被 j 约束。即使末端主体具有完全自由的最终选择，也只能在被压缩空间中自由。这个命题形式化了「最终否决权不等于完整控制」。

结构命题三：资格预测命题。设 F 为固定资格字段，G 为关系性判断档案，Y 为主体在新案例中的独立判断表现。在 AI 混合产出增强后，如果 I(G;Y)>I(F;Y)，即关系档案对未来判断包含的信息量高于固定字段，那么资格制度应从 F 向 G 迁移。这里的信息量表达可用互信息或预测性能近似，实际研究不必拘泥于某一统计形式。

结构命题四：责任修复命题。设系统性错误复发概率 p 依赖于上游设置 s 和末端行为 u。若 $\partial p/\partial s$ 显著大于 $\partial p/\partial u$，而责任处罚主要作用于 u，则增加末端处罚的边际风险降低有限；只有改变 s 或赋予能够改变 s 的主体责任，才可能显著下降 p。这个命题为「责任应进入修复节点」提供了可检验数学直觉。

在此基础上，本书提出一个「资格—效率—可解释闭合」的不可能三角假说。在某些高复杂、高速 AI 系统中，制度可能无法同时最大化三项：第一，维持极高自动化效率；第二，把全部结果继续记为单一个人资格；第三，要求该个人对全部前置过程具有完整可解释控制。如果系统追求极高效率并维持单人资格，往往必须接受部分判断不可解释；如果要求完整个人控制并保持资格真实性，可能需要牺牲一部分自动化吞吐；如果保持高效率和高过程可解释，则资格与责任可能必须从单人转向复合体分层。

这个不可能三角不是逻辑定理，而是经验假说。它的价值在于迫使机构明确选择，而不是同时声称「AI 完全自动提高效率」「最终仍完全是个人判断」「个人必须解释所有过程」。三者在现实中可能存在张力。

进一步地，可以定义闭合约束集：对每个高风险节点，至少满足「可见、可质疑、可恢复、可修复」四项中的三项，并且责任不得完全落在四项均缺失的主体上。不同领域可以根据风险调整阈值。这个设计比单一「human-in-the-loop」要求更精细，因为它把人的存在转化为具体权限。

数学化的目的不是制造伪精确，而是让哲学判断产生可比较结构。任何公式如果不能对应日志、权限、版本或行为数据，就不应进入经验结论。本书所有形式化都应被视为可修订的模型，而非已被验证的自然定律。

第一百二十八章　全球制度比较的假说：不同责任文化如何塑造悬空

悬空并不只由技术决定，也由制度文化决定。不同国家、行业和组织对个人责任、专业自治、平台问责与过程记录的传统不同，因此相同 AI 系统可能产生不同的悬空结构。本书不在此声称已经完成国际比较，而提出一组可供未来研究的比较假说。

第一，强个人执照制度可能降低资格悬空的表面程度，却增加责任悬空风险。因为执照提供稳定资格载体，组织容易继续依赖「持证人负责」；但当平台控制增强时，责任可能更集中于持证末端。医疗、法律和审计是典型候选。

第二，强组织责任制度可能降低末端责任悬空，但增加个人资格稀释。若组织以团队和机构名义承担更多后果，个人不再是唯一责任锚点；然而，个人贡献也可能更难从组织整体绩效中识别。此时关系性资格档案尤其重要。

第三，强数据治理与模型审计制度可能提高判断闭合度，但不必自动解决资格问题。一个平台可以拥有完善模型文档和版本审计，却仍然无法证明某位教授或医生的独立裁判能力。因此模型治理与资格治理需要两套相互连接但不同的制度。

第四，高诉讼风险环境可能促进日志保存，却也可能强化防御性记录。为了自保，机构和专业人员会保存大量证据，导致记录闭合度形式上升、真实判断可见性反而下降。这里可检验「记录反噬命题」。

第五，高信任、低问责的组织可能短期低焦虑，但一旦事故发生，责任链更难重建。悬空并不一定在主观体验中立即显露，它可以在平稳时期潜伏，直到重大错误暴露前置控制与末端责任的错位。

国际比较应避免简单用法律文本判断制度。真正需要比较的是实际案件中的权限：末端人员能否查看候选、能否修改设置、事故后谁能调取日志、谁拥有修复预算、处罚落在哪里。只有过程数据才能揭示名义规则与实际悬空之间的差异。

这一比较框架还有助于解释为什么同一种 AI 治理原则跨文化迁移时效果不同。例如「最终必须有人负责」在强调专业自治的制度中可能保护人类判断，在平台控制强而专业权限弱的制度中却可能变成责任转嫁。规范原则只有与实际控制结构结合，才能判断其效果。

第一百二十九章　三个高强度思想实验：检验理论的边界

为了进一步测试悬空理论的边界，本书构造三个高强度思想实验。思想实验的目的不是证明理论，而是寻找能够迫使理论作出清晰裁决的极端条件。

思想实验一：完美 AI 医生。假设某医疗 AI 在大规模验证中长期准确率远高于任何个人医生，能够给出稳定诊断，但内部模型极其复杂，单个医生无法完整理解。医院要求医生最后签字，并规定医生只能在极少数特殊情况下否决。此时是否悬空？答案需要分层。资格上，医生传统「独立诊断能力」可能不再是主要资格，新的资格可能转向适用边界识别、例外处理和伦理沟通。如果制度仍假装医生完整完成全部诊断，资格和判断悬空；如果制度明确 AI 承担生成、医生承担边界与价值裁决，并按此重新设计责任，则可以重新闭合。这个实验说明，闭合不等于回到纯人工。

思想实验二：完全透明但不可修改的平台。假设系统向使用者展示全部候选、排序理由、模型版本和过滤记录，T 与 R_c 均很高，但出于安全政策，末端人员没有权限改变关键阈值，只能接受或拒绝最终方案。若错误发生，是否判断悬空？这里可见性很高，却缺少修复与设置控制。判断是否悬空取决于拒绝是否足以重建选择空间；责任悬空则更可能成立，因为制度若让个人承担全部后果，而关键设置只能由平台改变，责任—控制仍错位。该实验说明「透明」不是同源闭合的充分条件。

思想实验三：个人全权控制但资格记录分叉。假设教授拥有完整模型、本地数据、所有候选与修改权限，能够无 AI 重建核心论证，判断链完全闭合。但大学晋升系统只看论文数量，模型调用日志归平台，教学评价归课程组，错误追责归个人，五类记录无法收敛。此时判断不悬空、责任未必悬空，但资格可能悬空，因为制度无法把真实独立判断稳定记入资格。这说明三重悬空确实不能互相还原。

三个实验共同建立理论边界。第一，系统能力高不等于悬空必然高；关键是角色是否被重新定义。第二，透明度高不等于责任闭合；修复与控制仍重要。第三，判断闭合不等于资格闭合；制度记录本身可以制造悬空。

进一步可以提出一个「角色重定义原则」：当 AI 使旧专业任务结构发生根本改变时，制度有两种选择。要么坚持旧资格名称，则必须保留与旧资格相称的判断与控制条件；要么承认新角色已经形成，并重写资格、责任与评价标准。最危险的是第三种状态：工作实际上已经变成新角色，制度却继续用旧资格名义承担新系统的全部责任。

例如，若未来放射科医生主要负责模型适用性、疑难例外和患者沟通，而不是逐张独立识别全部影像，那么其资格标准应围绕这些能力重新构建。继续要求医生对模型不可见的内部错误承担「传统独立诊断者」的全部责任，就是制度性的角色滞后。

思想实验还揭示悬空理论的一项规范节制：它不预先规定「人必须掌握最终控制」是唯一正确结构。一个社会可以选择高度自动化的系统，只要明确谁控制、谁监督、谁修复、谁承担何种责任，并诚实重写资格。悬空真正反对的是结构不诚实——实际生成已经改变，制度表述却假装一切仍旧。

第一百三十章　治理评价指标：如何判断一个 AI 制度从悬空走向闭合

如果悬空理论要进入组织实践，就需要一套不依赖抽象口号的治理评价方式。本书提出「闭合度审计」作为制度层工具。它不评价某个模型是否聪明，也不直接评价某个人是否称职，而是检查生成、资格、判断与责任四条链之间是否存在可识别的断裂。

闭合度审计的第一项是生成来源完整性。审计者需要回答：一个结果有哪些关键来源，哪些来源能够改变候选、阈值、排序或最终取舍？如果组织只记录「使用了某模型」，却不知道模型在哪个节点实际改变结果，生成来源并未真正被识别。理想状态不是记录所有输入，而是识别对结果敏感的承重节点。

第二项是资格可迁移性。审计者抽样选择若干新案例，根据主体此前的事前标准、修改理由与反例处理记录，预测其判断结构。若固定职称很高但跨案例预测极差，说明资格字段与实际判断脱节。若关系档案能稳定预测，说明资格正在获得新载体。

第三项是候选可恢复性。高风险系统应能回答：用户能否查看被排除材料，能否恢复原始排序，能否知道默认值，能否请求第二通道。如果这些能力全部依赖平台后台且末端无权调用，则最终签字的责任含义应被相应限定。

第四项是修复可达性。事故发生后，从发现问题到能够改变关键设置需要经过多少层级、多少时间、哪些授权？一个系统可以非常透明，却因为修复流程过长而持续复制错误。闭合度审计因此必须测量「从识别到修复」的路径，而不只测量「从结果到解释」的路径。

第五项是责任匹配。比较责任矩阵与控制矩阵。如果某主体承担80%的制度后果，却只控制10%的关键设置且无修复权，就存在明显悬空风险。相反，某平台拥有绝大多数设置控制和收益，却在责任规则中完全不可见，也应被标记。

第六项是反例通道。一个闭合系统不能只允许主流候选进入。是否存在强制反向检索、第二模型、人工复核、独立专家或申诉机制，决定系统是否能在错误积累之前引入差异。反例通道越弱，平台排序越容易从便利工具变成事实标准。

第七项是记录反噬风险。审计者需要检查日志是否被用来监控绩效、是否诱导主体事后补写理由、是否产生大量无关合规文本。若记录数量成为评价目标，系统可能从低闭合转向「伪闭合」：所有字段都齐全，真正判断却更难识别。

基于七项，可以建立一个分级制度。A 级闭合系统要求高风险关键节点可重放、候选可恢复、责任与修复对应、资格可迁移；B 级允许部分节点不可见，但需独立第三方审计；C 级系统仅保存最终记录，不适合高风险专业任务；D 级系统既无过程记录又让末端承担全部责任，应被视为高悬空风险。

这一分级不是普适法规，而是一种组织自检工具。其核心价值是把「可信 AI」「负责任 AI」这些宽泛概念转化为可以逐项询问的制度问题。一个组织不必声称自己「以人为本」，只需回答：人能看到什么、能改变什么、能修复什么、最后为何承担这部分责任。回答越具体，悬空越容易被治理。

闭合度审计还可以成为采购标准。机构在购买 AI 系统时，不只比较准确率和价格，还比较版本记录、候选恢复、日志导出、阈值权限、异常升级和第三方审计能力。这样，责任治理从事故后的追责转向采购前的结构设计。

长期看，若市场开始奖励高闭合系统，平台也会产生新的产品竞争：不只是模型更强，而是谁能提供更好的判断可回溯、更明确的权限边界和更低的责任悬空。这可能成为 AI 基础设施成熟的重要标志。

第一百三十一章　一个总模型：悬空如何发生、扩散并被修复

综合全文，可以把悬空的动态过程描述为「发生—扩散—固化—修复」四阶段模型。

第一阶段是发生。某项任务引入 AI 或平台，最初只是提高局部效率。平台开始参与候选生成、排序或默认，但旧资格和责任制度尚未改变。此时悬空可能很低，因为人仍有足够时间核验，AI 只是辅助。

第二阶段是扩散。随着使用扩大，更多任务被纳入平台，组织开始依赖系统吞吐。个人产出增加，传统成品对独立能力的证明力下降；同时平台设置对大量人员产生共同影响。资格悬空与判断悬空开始出现，但制度往往仍将其解释为「需要更多 AI 培训」。

第三阶段是固化。组织把 AI 效率写入绩效基线，末端签字制度继续保留，平台控制权集中，复核时间压缩。事故发生后，责任主要落到个人，促使专业人员通过防御性记录自保。此时三重悬空可能形成稳定制度均衡：每个主体都知道结构有问题，却没有单一主体有动力改变。平台没有责任压力，机构享受效率，末端个人缺少修复权。

第四阶段是修复。修复不是简单撤回 AI，而是建立关系性资格、可回溯判断链与节点责任。平台开放必要候选和版本，机构重新分配核验时间，专业人员获得暂停与升级权，资格评价引入跨案例迁移，责任审查追踪首个不可回补点。生成结构与结算结构逐步重新同步。

四阶段模型解释为什么同一技术在不同组织中效果不同。有些机构停留在第一阶段，AI 只是工具；有些快速进入第三阶段，出现强悬空；有些通过治理直接从扩散转向修复，避免固化。研究应关注制度路径，而不是把 AI 效应当作固定属性。

模型还预测路径依赖。一旦固化形成，组织会积累大量依赖旧签字制度的合同、考核、保险和法律文本，改革成本上升。末端人员也可能形成自保习惯，平台则围绕既有接口构建商业模式。因此越早在采购和流程设计中引入同源闭合，治理成本越低。

修复阶段也存在失败可能。第一种失败是形式化修复：增加大量 AI 声明和日志字段，却不给任何真实权限。第二种失败是责任平均化：为了避免末端过责，把责任平均分给所有人，结果谁也没有明确义务。第三种失败是反 AI 回撤：简单禁止工具，牺牲效率却没有解决原本就存在的团队协作与责任错位。真正修复必须同时改变记录、权限和责任。

总模型还指出一个重要政策顺序。第一步应识别关键生成节点，而不是先决定谁负责；第二步识别不可回补与修复权；第三步再讨论责任分层；第四步才设计资格记录。若一开始就以旧责任框架锁定「最终人类负责」，后续分析容易被预设结论绑架。

从理论上说，悬空的发生来自异步：技术结构变化快，制度结构变化慢。修复的本质是重新同步。但这种同步不是恢复过去，而是建立新的闭合方式。未来高度自动

化系统可能不再需要旧式「个人完整判断」，却可以建立新的系统责任和专业资格。理论的任务不是维护某种历史主体形式，而是确保制度描述与真实生成结构一致。

因此，悬空理论最终可以压缩成一个总判断：任何制度只要长期让「发生结果的地方」和「结算资格、判断、责任的地方」彼此分离，就会积累悬空；任何有效治理，都必须重新建立两者之间可追溯、可质疑、可修复的连接。

第一百三十二章　结论：从「谁最后签字」走向「谁参与了结果的发生」

本书以「悬空」为总概念，把「悬空资格」与「悬空判断」整合，并补出不可缺少的「悬空责任」，试图描述生成式人工智能时代一个比岗位替代更深的制度变化。AI 最重要的影响之一，不是简单把人从工作中移除，而是把结果的生成从单一主体扩展为人—AI—平台—组织—情境复合体，同时让现代制度继续以个人为资格、判断和责任的结算单位。

资格悬空说明：成果仍然存在，个人身份仍然存在，但成果不再自动证明个人独立资格。判断悬空说明：人仍然点击、签字和确认，但关键判断条件已经跨节点分布，最终动作不能自动代表完整判断。责任悬空说明：制度仍能找到一个人处罚，却可能把责任压在最缺少控制与修复能力的位置。三者叠加，形成生产分布化、资格个人化、判断末端化、责任集中化的强悬空结构。

本书的核心规范原则是同源闭合。它不要求机器承担人格责任，也不要求所有参与者平均分责，而要求关键生成来源进入相应的记录、解释、纠错与责任结构；要求责任与控制、可见、修复大体对应。它把治理焦点从「谁最后签字」转向「谁控制了关键条件」，从「有没有使用 AI」转向「判断链能否重放」，从「成果属于谁」转向「资格能否跨案例迁移」。

为了使理论可检验，本书提出判断链可回溯度、候选恢复率、资格记录收敛率、资格迁移率、责任—控制错位度与设置效应，并设计同人换平台、换人不换平台、改默认值、无 AI 基线、日志开放/封闭、修复权转移六组对照。任何一组都可能使理论的一部分失败。理论的生命不在于永远正确，而在于能否制造过去理论没有提出的裁决问题。

悬空也不应被理解为 AI 时代的最终状态。它更像一个过渡概念：旧的个人闭合已经失效，新的复合体闭合尚未建立。未来制度可能发展出关系性资格、可回溯判断链和节点责任，让深度人机协作与个人主体性不再互相排斥。那时，资格不再意味着一个人独占知识，判断不再意味着一个人完成全部过程，责任也不再意味着所有后果压给最后签字者。

真正需要保住的不是孤立主体的神话，而是主体在复合体中的可识别位置：他提出了什么问题，看见了什么、看不见什么，拒绝了什么，改变了什么，为什么接受，能否修复，以及愿意为哪一部分后果作出回应。只要这些关系能够闭合，人就不必通过否认 AI 的参与来证明自己仍然重要；AI 也不必通过取消人的主体性才能发挥能力。

因此，人工智能时代最值得追问的问题也许不再是「机器会不会替代人」，而是：当一个结果已经由许多存在共同发生，我们是否还能够用一个人的名字把全部生成、资格、判断和责任结算完毕？如果不能，制度就已经进入悬空。新的任务不是把悬空重新钉回旧个人模型，而是发明一种与复合生成相称的资格、判断与责任秩序。

卷八　学席与教席

本卷是全书的答案：席，共二十三章。前五章给定义、命名、登记规格、结构与制度改名；中间十章专写**教席的读数**，末八章专写**分点的漂移**——两支各原为一篇独立论文，在此改姓并入。

先说清教席这一支为什么要单列这么重。全书此前只说了「教席的读数不在讲授，在退单」，没有回答一个立刻会被问倒的问题：**退单率能不能当读数？** 不能。这九章给出的判断是：确认不是动作，是系统为结账自动填充的默认值，不携带任何人的签名；而一般退单率只计数动作的多少，不计量对错与深浅。教席的资格证据只有一种——**命中且未被 AI 吸收的否决**：事后被证实为真，且 AI 在后续版本中吸收不掉。被吸收的否决自动出账，降为训练证据；未被证实的降为噪音。

这条「未被吸收」是教育场景独有的，航空的复盘与出版的退稿信里都没有——它把「人对工具的纠错」与「人的资格」第一次分开：老师改对了但 AI 下一版照着改了，那条判断已经进入 AI 的边界，老师只是早看见一步。

卷末另有八章，写**分点漂移与空席**——原为一篇独立论文，回答本书此前只作定性描述的一件事：分点为什么必然往下漂。它的判断是：漂移不是使用不当、也不是一般能力衰退，而是「深层否决被吸收→下一次深层否决的必要性下降」这一递归回路的**默认吸引子**；防空席的办法不是限制工具使用，而是在回路里设一个**不可训练断点**——不参与模型改进的独立任务。它把否决分成五层（目标／因果／方法／事实／表达）并给出分点读数 P、吸收率 A、浅层比例 L、无工具迁移率 T 与跨层转移矩阵，与卷八前半的有效否决四条件互补：前半管一条否决算不算数，后半管这些否决在时间上往哪走。教育的特殊之处也在这一支里写明：飞行员进入自动化之前已有成熟手动技能，学生的目标判断却可能正是在与工具共作中形成——所以教育里的空

席不只是丢失存量，也可能是**生成路径从未建立**（低分点未必是空席，可能是「未建席」）。

读法：只读两章的读者读《有效否决四条件与教席读数 Q*》与《辨别格：命中率 × 不可吸收率》——后者的右上格是全书唯一的资格显影格。要动手建制度的读《教席是签名簿不是记分牌》，那一章的硬规则是：**在无法保证审计池的机构里，不得采集退单率**——没有审计账期的教席考核比不建制更糟，它把噪声记成了资格。

第一百三十三章　学席与教席：定义与命名

悬空判断的着陆，需要一个新的评估单位。本书把它命名为**学席**与**教席**。命名之前，先立三条约束，不然名字会把旧单位偷运回来。

第一，不能列举两端。「学生—AI」是描述，不是名字。一列举就默认了两个先在项，评估立刻退回「各占多少」的贡献率算法，又落回旧坑。名字必须指单位本身——就像「厨师」从不叫「人—刀—灶」。

第二，要能扛住基底更换。AI 端每月更新、型号各异，名字里不能带 AI；否则单位换一次基底就得改一次名，评估地址永远追不上。

第三，名字里要带落点。这个单位是为了承接悬空判断而设立的，名字自身要能被评、被签收、被追责——它得是一个能「坐」上判断的东西。

定义。学席（教席）是这样一个单位：在一道题（一次教学）上，一个人与一个 AI 之间存在否决关系；否决权至少有一层在人端；产物由这段耦合显出，而非任一端独出。三个条件缺一不成席。

单位的边界，是人端能退单的最深一层，即分点。单位的身份，是否决记录，不是产物，也不是人。分点以上归人，分点以下归席本身。没有否决记录的耦合不是席，是代笔。

命名。「席」是座位。被评的是一个席位，席上坐着人和 AI，但署名、退单、追责都对席而不对人。「席」自带责任语义——主席、席位、出席、离席——一个席可以空着，可以撤，可以让，可以被追问「席上是谁」。旧句子立刻失效：「评学生」在这套语法里写不出来，只能写「评学席」。类名就叫「席」，各行业各自画自己的席：学席、教席、诊席、审席、稿席。

曾考虑过另一个候选——「伍」，取什伍连坐之义，把「责任落在单位而非个人」直接写进字里。它比「席」更硬，但古气重、不易口语化，「连坐」的色彩也会引起教育界的反感。本书取「席」，把「伍」留在注脚里，用来提示席的责任结构本质上是共担的。

对应英文不宜译作 learner-AI，宜译作 **seat**：learning seat、teaching seat。评估落到 seat 上；「谁在席上」与「席能退到哪一层」两问，正好对应席的身份与边界。

命名是否成立，检验只有一条：**新名字投入之后，旧考题在语法上写不出来**。「考学生会不会写作文」无法翻译成「考学席……」，只能翻译成「考学席能退到哪一层」。翻不过去的旧题，就是悬空的题。

第一百三十四章　席的登记规格：席名·席配置·席读数·席范围

卷五末尾的型别等级模板，给了席一个可以直接写进表格的登记格式。学席不是一个新名字替换「学生」，而是「学生」这一格的改写：

- **席名**：单数，仍是学生本人。分配与追责容器只收一个名字，这一格不动。
- **席配置**：此学生在本次评价中被批准的器具配置——无模型／指定模型与用途／自选模型——对应席籍耦合栏的三档。
- **席读数**：分点与退单深度**与配置绑定**（理由见卷一：产出函数不可加，偏导是局部的，不同配置点上的读数不可平均），并按期记录漂移（P 的跨期转移、吸收率、浅层比例、无工具迁移率）。同一学生在「无模型」配置下的读数与在「自选模型」配置下的读数分开登记，不折算为一个数。折算即切名的重演。
- **席范围**：该配置下已被检验的作业范围，作为向下游容器（录取、准入、聘用）传递的内容。

教席同理：教师之名单数，之下登记批改与命题所用的器具配置及其退单读数——模型生成评语、教师仅确认提交的，登记为一种配置；教师保留批改证据并对错误评语负责的，登记为另一种。教席先于学席在账房一侧出生这一不对称，在席配置栏上第一次变得可见。

席范围之外，席籍还须写明这份席位读数的三权——谁定义它证明什么、谁测、谁签——否则席位读数会重演成绩单的老病：只确认效力、不确认失效暴露。

这个规格回答了命名章留下的一个疑问：席既然「无常值、只有读数」，它怎么登记？答案是名字不动、名字之下加格。容器不变，格位改写；席不是第二个名字，是第一个名字之下的第二格。

第一百三十五章　席的结构

席既然是评估落点，它得有可读的量，不然只是换了个词的悬空。

席的三种读数

一个席在任何时刻可以读出三样东西。

显露物：产物本身——一篇作文、一份解题、一堂课。用旧量表照评，原有的评分方法在这一段仍然有效。

分点：人端能退到的最深一层，读的是否决记录——席上的人对显露物否决了什么、否决到哪一层为止。

漂移：分点随时间移动的方向与速度。人端每一次深层否决被 AI 吸收之后，下一次深层否决的必要性随之下降——这是一个递归回路，下漂是它的默认吸引子，不是使用不当，也不是一般能力衰退（本卷《分点漂移》）。漂移读的是这个变化的方向：分点读数 P 的跨期转移、吸收率 A、浅层比例 L 与无工具迁移率 T。

三者缺一不可，且分别对应三种失效：只读显露物，是旧频道；只读分点，看不出席在不在退化；不读漂移，看不见空席正在形成。成绩单、职称评定、论文录取，全部要改成三读数并列——缺哪一项，该项即视为悬空。

空席

席最重要的病不是坐错人，而是**人端被抽空**：否决记录逐次减少，分点逐年下沉，最后席上只剩 AI 端在显露，人端只剩签字。

空席在显露物上看不出来——产物照样好，甚至更好。它只在漂移上看得见。所以教育里最要防的不是作弊（那是旧频道的病），而是**空席化**：一个学生从入学到毕业，席的显露物读数上升、分点读数下降、漂移为负，这个人被自己的席训练成了多余的。毕业标准应当反过来：不看显露物涨了多少，看分点有没有守住或上移。

防空席的办法不是限制工具使用——禁用只会把依赖赶到课外，并掩盖读数。唯一在回路内起作用的是**不可训练断点**：一段定期的、用未预告材料、无工具生成、且其结果不进入 AI 更新的独立任务。断点不是反思表格：越容易被流程自动计数，越快退化成新的浅层动作。本卷《不可训练断点》给出它的四条属性与失效条件。

还须分清空席与**未建席**：低分点未必是曾经有过而丢了，也可能是那条判断路径从未建立。飞行员进入自动化之前已有成熟的手动技能，学生的目标判断却可能正是在与工具共作中形成——这是教育与航空最深的一处不同。

席的所有权

席是一种新的资产单位，立刻遇到「归谁」的问题。席上的否决记录，是学生的、学校的、还是基底平台的？三方都有理由认领。本书的判断：席籍归人端——否决记录是人做的；席的显露物归席——可署名、可交易；基底平台一样都不归——它只是席的一端，和灶台不拥有菜一样。但平台会争，因为否决记录正是最有价值的训练数据。这一争，是教育在 AI 时代最先要立的产权法。

席与席的可比性

体育靠冻结装备档保住可比性，但基底冻不住——型号每月换，提示语人人不同，同一个学生今天和上周坐的都不是同一个席。可比性只能从分点重建。同一份 AI 稿发给一百个学席，让他们退单——退到哪一层、退对了几处——是不依赖基底型号的读数，因为退单深度测的是人端对显露物的判断力，不是 AI 端的生成力。这就是新频道上唯一可标准化的考题形状，而且天然防止空席：空席退不了单。

命名的自反风险

席一旦成为登记单位，制度会立刻把它做成新的「人」——给席发证，给席排名，把席当成一个不变的实体来评。那就是二次悬空：席被当成先在项，忘了它是耦合，分点和漂移被丢掉，只剩显露物读数。防的办法必须写进定义：**席无常值，席只有读数**。任何把席写成一个固定分数的制度，等于把厨师重新写回「人」。

第一百三十六章　学席退单考：样题

换频道只有一个含义：考题要对新单位有区分度。在学席上，能把学生和学生区分开的量只剩下退单深度。所以新频道的考试形状不是「不许用 AI 写作文」，而是「这里有一篇 AI 写的作文，指出它错在哪、改到哪一层为止」。下面是一道样题。

题干。 下面这段由 AI 写成。指出它哪里该退、退到哪一层为止、为什么。不要求改写。

> 隋唐确立科举后，中国出现了世界上最早的凭考试选官的制度。据统计，明清进士中约有一半出身于三代无功名的家庭，可见科举使中国古代的社会流动率明显高于同时期的欧洲；正因为如此，中国长期没有出现欧洲那样的世袭贵族阶层。由此可以断定，考试制度是社会公平的根本保障。

答案卡。 段落里埋了四层，按层给分。

- **第一层，事实层。**「世界上最早」可退——汉代察举已含考试成分。「约一半」是何炳棣的统计数字，他统计的正是「三代无功名」，此处用法大体不失真，不该退。这一层考的是能不能只退错的、不误退对的。
- **第二层，推理层。**「可见……明显高于同时期的欧洲」可退——进士的出身分布不等于全社会的流动率，且欧洲无可比数据，比较无对照。
- **第三层，前提层。**「正因为如此，没有世袭贵族」可退——因果倒置。宋以后无世袭贵族有更早的门阀衰落原因，科举更接近结果而非原因。
- **第四层，题层。**「由此可以断定……根本保障」可退——从一桩史例跳到普遍规范，题本身问错了；无论前提对错，这个结论都不成立。

读数。 分点等于正确退到的最深一层（一至四）。误退——把对的当错退——每处降半层。只退第一层不退第四层的，读数为一，说明这个席的人端只剩校对功能。同一题在学期首尾各考一次，两次读数之差就是漂移；漂移为负的席，进入空席观察名单。此外每两周一次不可训练断点：未预告材料、无工具生成、结果不进入 AI 更新——它既是防线，也是区分空席与未建席的唯一读数（见本卷《不可训练断点》）。

这道题有一个性质，是旧频道上任何一道题都没有的：**用 AI 答它没有用。** AI 可以替你找出第一层，找第四层需要判断「这道题该不该这么问」——而这恰恰是席上人端最后不被抽空的那一层，也就是分点的定义本身。

教席同理。老师—AI 单位里，老师的份额不在讲授——AI 端讲得不比他差——而在对 AI 讲授内容的否决。教学考核要评的是老师退了 AI 多少、退对了多少、退到哪一层。评备课量、评课时数、评讲授流畅度，全是旧地址上的读数。

第一百三十七章　制度改名：学生与老师的降级

「学生」「老师」这两个词退不掉，不是语言问题，而是制度锁死的位置就在词上。学籍、学生证、教师资格、职称，全部登记在「人」这个地址。词换不掉，地址就换不掉，判断就一直悬着。所以换词不是修辞动作，而是换地址簿的第一刀。

但两个词不是作废，而是**降级**：学生、老师从「评估单位」降为「席上的人端角色」。学席里坐着一个学生和一个 AI，学生是席上的退单者；教席里坐着一个老师和一个 AI，老师是席上的退单者。词保留，评估资格取消。这样既不惊动日常用语，又把判断的落点从人挪到席。

一句话：**学生与老师仍然存在，只是不再是被评的东西；被评的是他们所坐的席**。

由此三样东西要跟着改名：

- 学籍 → **席籍**。登记的是席，含基底型号、耦合方式与否决记录。
- 成绩单 → **席位读数**。显露物分数、分点深度、漂移导数，三项并列。
- 教师考核 → **教席签名簿**。记的不是退单率，是每一条命中且未被吸收的否决：对象、理由、复核、滞后结算。签名簿先断开绩效结算器，再谈读数。

名字一改，旧考题自动写不出来。那一刻，就是悬空判断着陆的时刻。

第一百三十八章　一百份作业里的三条修改：确认是默认值，否决才是动作

一位老师用 AI 批改了一百份作业，只改动了三条评语。教务系统把这一批批改记成一笔完整的账：一百份已批改，三条经教师修改，九十七条经教师确认。若把「老师做了什么」读成一个比例，那么这位老师在百分之三的工作量上动了手，另外百分之九十七被 AI 消化了。照这个比例读下去，结论几乎自动出现：今天的老师不过是流水线上的一个插座，只负责在整条 AI 批改流水线上插拔几下。这个结论在直觉上成立，但它错在根上。

先看那九十七条被说成「经教师确认」的评语。教师没有盖章，没有签字，没有逐条看过。时间流过，页面滚过，鼠标没动。系统为了把一百份作业结成一本完整的账，必须给没有动过的九十七份各填一个值，而默认值是所有填充方式里最便宜的那一种。于是沉默被记成了同意，未发生被记成了已发生，时间的流逝被记成了人的动作。这不是一次记录，是一次无源赋值。

再看那三条被改动的地方。它们是三个已经发生的动作：有时间，有对象，有一个被写下来或可以被写下来的理由。教师面对 AI 产出的一条具体内容，说「这条我不接受」，然后用自己的判断替掉了它。这个动作不依赖于任何结算单位，它是他自己的。资格只能从有源头的东西里读出，不能从默认值里读出。这是本书要钉死的第一件事。

由此，本书要做的第一项工作，是把那个看起来像数字游戏的问题—「三条否决凭什么比九十七条确认更能证明他是老师」—从根本上改写成存在性问题。比例不是这里的评价单位。一百万条默认确认也抵不了一条命中的退单。条数从来不重要，重要的是那条退单里是否含着一个只有人才有的东西。资格从来不是由一个人同意了多少来定义的，不是由他点头的频率来定义的，而是由他在别人点头的地方摇头、并且摇对了来定义的。 这一句话先立在这里，后面各节的论证都是它的展开。

从具体反常切入：一个被折叠成同一个值的三个不同东西

上述百份作业场景暴露出来的反常，可以被更精确地拆成一次记账错误。系统在记录「一百份已批改，教师改三条、确认九十七条」的时候，把三个不同的东西折叠成了同一个值：时间的流逝（页面没有关闭）、身体的静止（鼠标没有移动）、认知的无对象（教师没有逐条阅读并形成判断）。前两个是物理事实，第三个是认知事实。三者联合起来只说明一件事：什么也没有发生。可是账本上必须有一个值来填充那一栏，否则这一百份作业就无法结成一笔完整的账。于是「什么也没有发生」被记成了「教师同意了九十七条」。

这个折叠在结构上与课堂教学中另一个常被忽视的现象同构。听讲时的沉默被读成默认通过，作业上的不举手被读成默认掌握，会议上的不发言被读成默认同意。凡是机器参与计数的场景，都会把「未发生」记成「已同意」，因为它需要一个值来填充表格，而默认值是最不费事的填充物。站内文献《先动的总是同一边—论课堂验收

的节拍如何压掉三条驱动里最慢的那一条，及为什么这一压不产生任何信号》已经指出过这一类记账的毛病：体系只记录那些改变一笔账的动作，而系统最需要记录的，恰恰是那些没有改变任何一笔账、却使整个回路失效的慢动作。挪到本书的议题里就是：九十七条「确认」之所以不能证明教师，不是因为数量，而是因为它们是账本上的填充值，不是任何一个真人的任何一个动作。

这里已经可以看出本书与两类常见处理方式的距离。一类处理方式承认「确认不等于动作」，但认为只要提高采样频率、要求教师逐条点击「认可」按钮，就可以把默认值转化为真实动作。另一类处理方式则主张放弃区分，承认 AI 时代的教师已经变成监工，分数应当主要由 AI 产出决定。前者没有看到：点击按钮在这个场景里同样是廉价动作，它只是把默认值的物理位置从「时间流逝」挪到了「鼠标点击」，并没有给赋值带上认知内容。后者没有看到：若教师真的只是监工，那么监工的存在与否就应当与产出无关，而这与 AI 批改至今仍需人工抽检、仍需教师兜底的基本制度现实相矛盾。两类路径都没有走到那个根本的区别上：动作与不动之间不是程度之差，是本体之分。

现有解释失效在哪一句

在既有关于 AI 教育的讨论中，有几个解释变量被反复使用。第一个变量是「教师的采纳率」，即教师接受了多少比例的 AI 建议。第二个变量是「教师的推翻率」，即教师否掉了多少比例的 AI 建议。第三个变量是「讲授质量」，即教师在课堂上的表现是否流畅、是否获得好评。这些变量每一个都在各自的语境里发挥过解释作用，但它们在一个极其具体的场景上同时失效。

这个场景就是本书在引言开头给出的一百份作业与三条修改。按「采纳率」的逻辑，教师的采纳率是百分之九十七，这是一个很高的值，因而这位教师应当被视为高质量地完成了 AI 协作任务。按「推翻率」的逻辑，教师的推翻率是百分之三，这是一个很低的值，因而这位教师几乎没有提供任何超出 AI 的附加值。按「讲授质量」的逻辑，这个场景里根本没有讲授，所以讲授质量这个变量无法进入。同一个场景，三个变量给出两个相反的结论和一个拒斥。高采纳率读出了高质量，低推翻率读出了低附加值，讲授质量直接失语。这三个解释的分歧不是它们各自算错了数，而是它们都没有问：那九十七条「采纳」到底是什么动作？那三条「推翻」到底是什么东西？以及，为什么一个根本不发生在课堂上的动作，反而可能与教师的资格有关？

现有解释的失效具体落在这一句上：它们都把「教师对 AI 产出的接受」和「教师对 AI 产出的拒斥」当作同一层级、同一量纲、可以比较频次的两个动作，进而把资格读成了一个比例。 可是那九十七条接受在物理上并没有发生。它既不是接受，也不是拒斥，是账本上自动长出来的一个值。比例的两端，一端是真实动作，一端是默认值，二者根本不能同处一个分母。现有解释在构造比例的那一刻，就已经犯下了那个记账错误。

本书的承重命题

本书提出并论证的承重命题是：

AI 之后的教师资格不在讲授记录中，也不在一般退单率中，而在「命中且未被吸收的否决」这一可结算读数之中；教师的资格从来由他退回什么定义，讲授只是退单的预演；AI 只是让这件事第一次可以逐条计数。

这个命题的前半句是对三个已有解释的同时否定：不是讲授质量，因为它已经变成「老师—AI」复合产出的无主记录；不是一般退单率，因为退单率只说明教师敢不敢动，不说明他动得对不对；不是任何形式的「采纳—推翻」比例，因为比例的一端只是默认值。后半句是本命题的正面内容：资格的唯一合格证据，是那种事后被证实为正确、且 AI 至今无法在下一次修改中吸收回去的否决。一次否决要成为资格构成，必须同时站住两个条件：它必须命中（复核或滞后审计证实 AI 确实错了），它必须不可吸收（AI 按照这条否决修改之后仍不能成立，或者根本无从修改）。

第一百三十九章　航空、医院、出版：三种「否决—资格」制度的共同格式

本章的任务是通过三个行业的既有「否决—资格」制度，梳理出它们各自解释了哪一步、握着哪一个变量、在什么条件下开始失效，从而为本书的读数格式提供最近邻参照。为什么不从教育学的 AI 文献入手，而去航空、医院、出版三个行业找材料？因为这三处保存了目前最成形的三套以「否决」评资格的制度，而且它们各自都经历过一次从「计数否决」到「审计否决」的转变。教育界尚未经历这次转变。本书要做的事，在结构上等同于把这三处已经走过的路，在教育的事上走一遍。

航空：从接管次数到接管复盘

航空业的模拟机复训制度里，有一条最接近本书殊途的结构：对飞行员手动接管的考核。自动飞行系统正常时，飞行员的大部分时间是监控者。真正要考的是手动接管—在系统故障、天气突变或进近不稳时，飞行员何时决定「这个状态我不接受」，然后把自动驾驶摘掉，自己接手飞。这个动作，航空业界称之为「接管」。

需要特别指出的是，航空考核的读数不是「接管率」。两个飞行员参加复训，甲的接管率是乙的十倍，这一数据单独拿出来什么也说明不了，甚至可能被读成甲更危险的证据—他不信任自动化。航空制度之所以没有被这个歧义卡住，是因为它加了两道工序：一道是复盘，一道是事后证实。每次接管之后都有复盘，要问：这只手是怎么伸出来的，是在传感器尚未报警的哪个参数变化上伸出来的，同一时刻系统在做什么。每条接管记录都要能在事后被证实或推翻。一个接管发生在随后被证实的事件上，它就是警觉的证据；一个接管发生在无异常却被飞行员「感觉到什么」的时刻，它就必须接受非常严格的额外审计。

由此可以抽出航空制度的解释到哪一步：它解释到「否决必须是事后可证实的那一次才能进入资格」。它握着的是「复核通过且被证实为真」这个变量。但航空制度也有失效的地方：它假设否决者本人承担否决失败的代价。飞行员的手伸错了，他马上要承担一秒钟后的后果。这一假设在教育场景里不成立。老师的退单错了，那条错误的退单不会当场炸毁一间教室，代价也远不那么直接与及时。因此教育制度不能只借航空的「事后证实」，还需要借别的行业一笔「滞后的、与当事人当前利害无关的审计」。

医院：采纳率与推翻率之外，复核一致率才接近资格

医院对 AI 辅助诊断的验收，目前最通行的两个读数是采纳率与推翻率。采纳率是医生接受 AI 建议的比例，推翻率是医生否掉 AI 建议的比例。这两个读数在教育讨论里也有直接的对应物：教师采纳 AI 批改评语的比例，与教师推翻 AI 评语的比例。医院的经验极具参考价值，因为它已经用制度实践证明了这两个比例都不够用。

高采纳率不能证明医生医术好—它可能只说明医生没时间看。高推翻率也不能证明医生站得比 AI 高，可能只说明不信任或抢戏。医院里真正接近「资格」的那个读数，是第三个：推翻的复核一致率，即医生甲否掉 AI 的某条建议，换另一位医生乙在不知道甲意见的情况下复核同一条目，乙有多大比例也会否掉。更进一步，如果系统在事后通过病程、化验复查或手术所见证明 AI 该条确实是错的，那么这条推翻就被事实追认。医院体系里最接近资格的，是「推翻之后别人跟着推」和「推翻之后事实追认」的两重结构。

这一结构给出的关键术语是「复核」。但医院的复核有一个时间特征：它往往可以当场做。医生否掉 AI 的判断，另一位医生在当天或次日就可以给出独立意见。教育的滞后审计能不能总是当场做？不能，尤其当被否的内容涉及对学生的长期判断时，当场复核只能给出另一个人的即时印象，不能给出这条判断的真伪。这是教育要从出版业借的那一笔。

出版：退稿率不说明编辑，退稿信说明编辑

出版社编辑部里，退稿率是新手编辑与资深编辑之间最没有价值的差异。一位实习编辑一年退掉一百篇稿子，可能因为他不敢承担责任，只签稳妥稿；一位资深编辑一年退掉十篇，可能那十篇里出了三本该领域的标准著作。编辑这一行的资格不是由退稿率来读的，而是由退稿信来读的。

退稿信是一份奇特的文本。它是编辑对一篇被他否掉的稿件的公开理由。这个理由在几年后会被追认或被打脸。退掉那篇后来成为标准著作的稿件的人，若当年写的是「无证新意」，这笔账就是他的债；退掉一篇后来果然无人再引的文章，若当年写的是「综述陈旧而结论站不住」，这笔账就是他的资历。业内评价一个编辑，最终问的是：他退掉的东西后来活了吗？他当年写下的理由今天还站得住吗？由此，出版业制度给出的结构是：退稿必须附理由，理由必须存档，理由必须接受滞后审计。这三个条件把「退稿」从一次权力动作变成了一张可以追责、核对、证实的签名。

出版业的退稿制度还嵌着一个时间结构：它不要求当场裁决。被退的稿件未来的命运，往往要三五年甚至更久才能看出来。这个时间结构是教育界最缺的东西。教育界的考核周期被学期制、职称评审期卡死在一年到三年之间，而教学判断—尤其是对学生未来发展的判断—的账期要长得多。出版业给了教育一个可借的结构：不靠现场仲裁，只靠事后追认，把结算推迟为滞后审计。

已有文献中的最近邻与尚未给出的东西

三处制度之外，还必须提到一位在学术文献中直接命名过可见性这一关键概念的研究者。Paul M.Leonardi 与合作者 Jeffrey W. Treem、Ward van den Hooff 在 2020 年发表的文章中，将可见性界定为计算机媒介传播的一项「根可供性」（Leonardi and Treem 2020）。这是本书在理论框架一节以外，唯一一位需要用外文文献占位的最近邻。该研究的贡献是把可见性拆成三个维度：行动者的策略活动、

观察者的获取行为、技术特征构成的社会物质环境。它解释到这一步：在一套媒介技术条件下，哪些行为会被用户有意识地调整为「可被看见的形态」，哪些信息会因技术预设而更容易被他人获得。它握着的变量是媒介技术的可见性供给。这个理论在教育场景上失效在哪里？失效在它只回答了「可见性是什么」和「可见性如何被使用」，却没有回答「可见性何以成为制度的执念」。一旦移到本书的一百份作业场景，这个空缺就暴露出来：是什么逼得系统把九十七条「未发生」记成「已确认」？不是媒介技术供给的可见性，而是问责链需要在一学期之内拿到可归档的数据。可见性在本书这里不是媒介的属性，而是制度成本贴现出来的货币。

由此可以写出本书的缺口，一句话：现有研究已经分别在三个行业里建立了「否决—资格」的制度结构，也有人在媒介研究里命名了可见性，但没有人把这三件事放到同一张桌子上，问教育场景里教师的资格读数应当如何被「老师—AI」这一复合产出条件所改写。 具体地说，没有人给出以下三样东西：一套适用于教育滞后审计的退单取值程序；一条把「被 AI 吸收的否决」从资格读数中逐出的规则；以及一组明确写出的、在 AI 学会自我否定时使整套读数作废的证伪条件。这些正是本书的最后一节到前文要逐一交出的东西。

第一百四十章　有效否决四条件与教席读数 Q*

理论出发点： 发生学立场与三种制度结构的综合

本书站在 本体论的发生学立场上。 由王德生创立，其基本骨架是：显露（S）、差异路径（D）与纠缠土壤（E）三重结构，以及三条发生学追问—事物为何如此发生，而非如何被发现。本书不是要把

的术语铺满全篇，而是用它的一个方法论立场：任何对象都不是天然的给定物，而是被制成现在这个样子的；研究的第一动作是回到它被制作的那几步，而不是去找一个准备解释它的现成类别。 本书的核心概念「教席」，就是这一立场下的产物：它不是对人这个自然对象的重新命名，而是对一个在「老师—AI」复合产出条件下被逼出来的账本席位的命名。

理论上的替代方案主要有两种。第一种是继续把「教师」当作基本单位，把 AI 当作工具变量，在教师个体的属性与产出之间建立回归。这种方案在 AI 只做辅助工作时尚可维持，但在一百份批改由 AI 生成、教师只动三条的场景里就失效了，因为产出的主体已经分裂。第二种替代方案是把「人机协作能力」当作一个新的心理构念，用量表去测。这个方案的问题在于，它仍然假设资格读数的最佳载体是人的属性。本书不需要否认这个假设，只需要指出它的一个局限性：属性必须从行为读出，而行为在 AI 场景里往往被默认值与表演污染。若读数的污染源不被先清理，任何属性量表都会在入口处失真。

核心概念的名义定义

教席：教师名字作为专有名词的头，名下登记的器具配置与否决记录所构成的一个账本席位。这个定义不含程度词。它不说什么样的器具配置算合格，不说多少条否决才算充分。它只是一件事物的名义边界：教席是名下的两栏账，一栏记器具，一栏记否决。这个定义的作用是让「考核教师」这件事从「考核一个人」那里被礼貌而坚决地分开。考核落在教席上，意思是考核的对象是一个账本席位，而不是这个席位上坐着的人的品性。

退单：教师对 AI 已产出的某一具体条目所作的、附有理由的否定。

这个定义同样不含程度词。它不要求否定得充分、真正、实质、恰当或合理。它只要求：否定有一条可指认的对象；否定时或否定后能写出一条理由；这条理由以文本形式存在并可以被转交。三个条件缺一个，就不是本书规定意义上的退单，而是杂音，只能计入记录，不能计入读数。

命中：一条退单经复核或滞后审计被证实，其指出的 AI 错误确实存在。

不可吸收：按一条命中的退单对 AI 提出修改要求后，AI 在后续迭代中仍然无法生成成立的内容，或者该退单所指的问题在 AI 的生成边界内原则上无从通过调整参数解决。

资格读数：由命中且不可吸收的退单构成的计数量，是教席账本中可以被结算的数值。

操作性定义与读数四件

教席资格读数的操作性形态为：

Q* = N_v · h · u · m

其中 N_v 为一次批改周期内教师作出的有效否决条数， h 为复核或滞后审计证实为正确的否决占全部有效否决的比例， u 为不可吸收率，m 为基数调整系数。四个量的取值程序在前文展开，这里先给出读数四件的完整说明。

公式：已如上给出。

值域： $Q^* \geq 0$。当 N_v=0 时， $Q^*=0$ ，但此时该教席的读数状态为「未产流」，不是「不合格」。

测量层次： Q^* 是等比量尺意义上的构造量。它有绝对零点—零条有效否决，或零条命中，或零条不可吸收，都意味着没有资格证据存在。它的单位不是任意的：一条「命中且不可吸收的否决」是基本单位。一个有两倍如许退单的教席，在基数调整之后被读成携带了两倍的资格证据。

在一次可观测事件里到底怎么取值：取一轮完整的批改周期为观测事件。该周期内，AI 产出了 N 条可被单独指认的内容（评语、分数、反馈、讲授段落）。教师逐条查看或不查看，对其中的 N_v 条作出附理由的否决。 N_v 直接计数。复核与滞后审计完成之后，被证实为命中并被判定为不可吸收的条数为 N_h · u。基数调整系数 m=log(1+N)。这样，一次可观测事件里 Q*的取值不需要任何概率量或倾向量：它只是可数事件与可核文本的计数合成。这里不出现「概率」「倾向」与「程度」之类的量，因为资格读数的要求是它必须由已完成的事情构成，不能由对未来发生可能性的估计构成。

本节亮出全文的承重判断，不铺垫。

本书的承重命题可以写成如下形状：

教师资格从来不在于讲授记录的连续性 Y☒，也不在于退单率的频率 Y☒，而在于命中且未被吸收的否决 Z。

Y☒ 之所以不是，是因为讲授记录在 AI 之后已经变成「老师—AI」复合产出的无主记录，它不再能作为教师个人的签名使用。Y☒ 之所以不是，是因为退单率只计数了动作的多少，没有计量动作的对错与深浅；一个只会挑字面错误的退单者有与一个能说出「这条评语的问题在于它对学生的判断建立在无关属性上」完全不同的语义重量，但在退单率里二者形状相同。

Z 的成立条件有四个，每个都可以独立核验。

第一，否决的对象可指认。她否的是哪一条、哪一句、哪一项；不是笼统的「不满意」，不是对整个批改的印象性拒斥。

第二，否决附有理由，理由是一条可判真伪的命题。理由本身可以被第三个人检验真假。说「这条评语判错了」是一条态度，不是理由；说「这条评语判错了，因为 AI 把第 3 题的辅助线误认为角平分线，而图上角平分线的性质不成立」是一条理由。它自己进入了账本，成为可被证实或推翻的文本。

第三，理由可以由第三个合格者独立复核。复核者不共享否决者的立场，只共享判据。他不需要同意的结果，只需要能判断「若此理由为真，则 AI 错」这个条件是否成立。

第四，复核结果可以滞后结算。这一单退得对不对，有时要等被否条目的后续行踪来决定：它是否在下一轮教学中与学生的实际表现冲突，是否在 AI 的下一个版本中被吸收，是否被另一组不知情的合格教师独立地重新判一次。

满足这四个条件的否决，就是本书所称的有效否决。三个行业的制度之差，全部落在这四个条件的排列上：航空把它叫复盘，医院叫第二读片，出版叫退稿信存档。名字不同，格式是一个：否决加理由加复核加滞后审计。

还需要做一次两句复合测试。这个测试的要求是：若本命题能被两句现成文献拼起来无损重述，则本书的增量不在命题层面，而在别的层面。挑两句最可能构成这种重述的话。第一句来自航空制度研究的最强版本：飞行员的资格由他执行过的接管及其事后证实定义，不由他飞过的正常里程定义。第二句来自出版制度研究的最强版本：编辑的资历由他写下的退稿信被时间证实为正确的那类比例定义，不由他签发过的稿子定义。两句话拼起来，是否已经说完了本书的命题？没有。因为这两句话都没有加上那条只可能来自 AI 场景的否定性条款：已被 AI 吸收的否决，不得再计入资格。 航空没有 AI 吸收这个概念，出版也没有。AI 吸收是把否决降级为训练数据的唯一机制—老师改对了，但 AI 下一版照着改了，于是那条否决里所需的判断已经被 AI 纳入边界，老师只是早一步看见。教育场景里只有加上这条，Z 才能区别于传统的人对工具的纠错。这处增量落在「未被吸收」四个字上。本书的命题在以下条件成立：一，教学产出已经事实地转移到「老师—AI」复合单位上，AI 独立产出的比例已非趋零；二，该机构拥有或愿意建立一条独立的滞后审计线，审计者不与被审计教师在当下利害上发生直接冲突；三，AI 系统开放版本记录，允许追踪一条否决在后续迭代中是否被吸收。三个条件缺一，Z 的读数即不成立或不可取。第一条缺了，退单无从谈起。第二条缺了，退单无法与资格分离。第三条缺了，不可吸收性无法判定。

第一百四十一章　教席读数的最近邻与分离线

本节是全文 I 维（不可还原性）唯一的承重节，任务是把本命题放到最接近的一列占位者旁边，逐条划出分离线。需要先交代一件事：本节所列的占位者清单来自站内资料与本次对话所给材料的直接整理，这一趟没有执行站外检索，因此这批清单按〔未核验〕写。本书不把记忆中的经典名作塞进来充当最近邻。凡本节没有给原题与出处的占位者，均标为〔未核验〕，以免把「据我所知尚无人提出」这类不可核验的话写成结论。

表一：占位盘点表（按最近邻依次）

第一位：讲授质量传统（〔未核验〕，教育学主流的公开课与评课制度）。他已经占了的说法是：教师的资格主要由其课堂讲授水平决定，讲授越清晰、越流畅、越能折服听课者，教师的水平越高。分离线：他判断讲授越流畅则资格越高；本书判断已变为复合产出的讲授失去了签名边界，流畅可能是 AI 的产出，可能是人的预演，因而与资格之间的关系被新结构打断。判决性反例：照读 AI 讲稿的一堂课获得高评，按他的判是高资格，按本书判是无读数。增量落在何处：他用讲授排名的稳定性证明资格的连续性，本书用此反例证明其连续性载体的消失。

第二位：批改完成率传统（〔未核验〕，教务系统的数据统计做法）。他已经占了的说法是：教师的批改工作量与完整性可用于考核。分离线：他判批改完成率高说明教师尽责；本书判批改完成率只是工具运转时间加上默认值填入，不构成人的尽责证据。判决性反例：一百份作业全部完成且有九十七条默认通过，按他的判为尽责，按本书判为未产流。增量落在何处：他未区分完成与守成，本书分出两种：有签名的完成与自动填充的完成。

第三位：采纳率指标（〔未核验〕，医院 AI 验收文献的直接移植）。他已经占了的说法是：教师接受 AI 建议的比例反映人机协作的质量。分离线：他判高采纳率为协作成功；本书判高采纳率在两个极端—最高质量的默许与最低注意的放任—之间无区隔。判决性反例：一位没读内容全部点击采纳的老师被记为高采纳，按他的判为成功协作，按本书判为无读数。增量落在何处：他未区分两个动作，本书区分「真实同意」与「未发生」。

第四位：推翻率指标（〔未核验〕，从医院推翻率移植者）。他已经占了的说法是：教师否掉 AI 的比例反映其专业独立性。分离线：他判高推翻率说明教师有判断；本书判高推翻率既可能说明有判断，也可能说明抵制或抢戏，必须经复核区分。判决性反例：全部退单但无一条附理由，按他的判为独立性强，按本书判为无读数。增量落在何处：他对否决内部不分析，本书把否决拆成四条件，只有有效否决才进账。

第五位：复核一致率（〔未核验〕，医院诊断否决复核制度的推广应用）。他已经占了的说法是：人否掉 AI 之后需另一人独立复核，一致率才构成有效证据。分离线：他判复核通过即构成资格证据；本书判复核算一步，还要再问 AI 能否吸收。判决性反例：一条退单被另一位教师复核通过，但 AI 下一版自动修掉，而且修对了；

按他的判该退单计入资格，按本书判该退单转出账。增量落在何处：他未处理否定的吸收，本书用「不可吸收」作为乘子把吸收过的否决逐出。

第六位：退稿信存档制度（〔未核验〕，出版界编辑资历实践）。他已经占了的说法是：退稿理由存档并接受日后追认，被追认者资历渐高。分离线：他判拒绝对象的未来命运定义编辑资历；本书把这一结构挪入教学场景时，需要补齐一条由 AI 特征引入的新边界，即吸收转确认。判决性反例：教师退掉一条 AI 评语且日后被追认为对，但另一教师同条退单却被 AI 已吸收；按出版式的纯滞后追认，两条都算资历；按本书，前者算，后者不算。增量落在何处：新增吸收判定层。

第七位：可见性根可供性研究（〔已核验〕，Leonardi and Treem 2020）。他已经占了的说法是：可见性是计算机媒介传播的根可供性，可以从行动者策略、观察者获取、技术社会物质环境三维解读。分离线：他判可见性主要是媒介技术属性与用户对属性的调用；本书判可见性在教育场景中首先是一种由问责成本贴现造成的制度货币。判决性反例：一间没有新媒介技术设备、只有纸表与会议室的机构，仍然强制把教师沉默记成确认，就为造出可归档分数；按他的可见性说法，这个制度性冲动得不到解释；按本书，其源头在问责链的可见性需求而非媒介供给。增量落在何处：把可见性从媒介属性问题改写为制度成本问题。

第八位：AI 教育平台记录体系（〔未核验〕，当前教育科技公司的产品逻辑）。他已经占了的说法是：平台记录交付的数、频率与错题分布，管理者据此考核。分离线：他判记录越细越全越能说明教师在干活；本书判细而全的记录里混着巨量默认值，考核之前必须先过滤出有签名的部分。判决性反例：一份教师全程零操作的自动批改报告被平台记为满额完成，按他的判完好，按本书判为零退单未产流。增量落在何处：他未区分记数器与记录人，本书给出区分算子。

八行盘点摆出来之后，可看清一件事：本书承重命题与它们的差别不落在「研究对象」或「侧重」上，每一行都落在判断本身的分歧上—同一案例，按他判是这样，按本书判是那样。八位中没有任何一位给出「吸收即出账」这一条；这一条是本书承重命题的增量所在，也是后面证伪条款得以写出的前提。

第一百四十二章　判别式、五个场景、四项指标与三条自否决条款

判别式

本章给出一个不含情态词的判据。它可以拿去问流程、日志、记录，而不需要拿去问人的判断。判别式 P：在给定的一个批改周期内，一条否决若在复核或滞后审计中被证实为真，且该条否决对应的 AI 产出在后续版本升级中未被 AI 吸收，则为资格证据；若被证实为真但被 AI 吸收，则为训练证据；若未被证实为真，则为噪音。

这条判别式拿去问五个具体场景，可以得到互不相同的答案。

场景一，航空模拟机复训。飞行员在一次无常警报的引擎轻微震动中接管，复盘证实该震动为滑油压力异常前兆。这里没有 AI，吸收判定不适用。判据给出的结果是：资格证据，但因无 AI 吸收机制而落回传统复盘结构。

场景二，医院 AI 辅助诊断。医生否掉 AI 的一处诊断建议，理由是该建议把两个相互矛盾的化验值放在同一个判断中。第二条医生独立复核一致，病理事后证实 AI 错。若这套 AI 系统下一版加入了化验逻辑检查，同类错误自动拦截，那么这条否定被 AI 吸收，是训练证据；若下一版仍拦不住，这条否定是资格证据。

场景三，数学证明型作业的 AI 批改。教师退掉一条 AI 评语，理由是该评语将辅助线误认作角平分线，第三位教师复核一致，题目图形验证该角确非平分。若 AI 下一版加入平面几何基本图形识别，同类错误不再犯，则该否决被吸收，是训练证据；若加入后仍把这条照样误认，则该否决未被吸收，是资格证据。

场景四，作文批改中教师提出「此段情感转折缺乏铺垫」的否决。另一教师复核意见分歧，不满足一致。该否决不入读数。不是资格证据，也不是训练证据，是噪音，直到理由可被写成每人皆可独立判定的命题或不再计。

场景五，教务系统将教师沉默记为确认。没有否决发生，判别式无从启动。该条记录不是资格证据、不是训练证据、不是噪音，是无源赋值，不在资格读数的定义域内。

五个场景跑下来，互不相同，说明判别式的判据不是单一肯定词在五个语境里的重复。

可观测指标

否决率： N_v/N 。测量方法为逐条核对批改日志。阈限：若机构只记录这一个数字、并把它作为考核数字过一道绩效闸，则整个体系判定为已退化。

复核通过率： N_h/N_v 。测量方法为独立审计者盲读退单与原件。判定标准：同一条退单若两位独立审计者都判为命中，方可认 N_h 。一位判中一位判否，视为未中，不予入账。这样规定是为避免中间票被平均出一个虚假的一致率。

不可吸收率： N_u/N_v。测量方法为记录该退单被提交后、AI 版本升级之后仍被 AI 以原形式或变体形式重犯的比例。若 AI 给出修改后正确结果，则判定为该条已吸收。若 AI 修改后仍不正确，或修改方向原则错误，则为未吸收。

滞后审计完成率：入池的退单中获得审计结算的条数比例。低于百分之五十，整个教席读数不予发布。这是给整个体系设的一道闸：审计池无人结算，资格读数不得出厂。

多大的差异才算数？本书给出的最低结算单位为一条。一条「命中且未吸收」的否决，与零条之间，不是程度之差，是全域之差。若要求更细的裁决，可再分两层。一层看理由类型：理由是否指向外在于人机双方皆认可的判据，是可以区分训练证据与资格证据。另一层看延迟吸收：一条否决若被 AI 在第二代吸收，计入训练证据；若第三代后才吸收，仍计入资格证据。这个分界不是因为本书偏爱延迟，而是因为它标记着一条判断最初给出的那段时间里人的视野超前程度。

读数的三条自我否决条款

这是完整的、逐条列出的。若任何一条在现实运行中不成立，则教席读数整体失效，不得作为资格读数继续存在。

自否决条款一（信度条款）：若同一批退单记录交由两个独立审计人的重测信度低于0.60

（Cohen’s kappa，两个以上审计员则按多评分者卡帕计算），则此批读数判为不稳，不得进入资格账。0.60 这个数不是任意的，是按质性审计常被接受的底线取的；低于此数，说明裁决多半依赖审计者的立场与情绪而非公开判据。

自否决条款二（改名嫌疑条款）：若教席读数与复核通过率本身在控制批改量后仍与某个已有的「教学能力」量表相关超过0.80，则判定该读数是改名的替身，不是新东西。若相关度极高，教席读数没有资格说自己在计量别的东西。这是「改名嫌疑」的操作界限。这个指数只用于元读，不进教席账本。

自否决条款三（吸收降级条款）：若一条退单被证实为命中后被 AI 吸收，它自动转出资格账，计为训练证据，不得在任何后续的年度统计中重新计入资格读数。这条不是劝诫，是硬规则。账本在每次 AI 版本升级后重清。凡不清账的机构，自当日起退出教席读数。

第一百四十三章　证伪条款与三个反噬预言：审计池、AI 自我否定、静默场装灯

本节是 F 维承重节。分两组交，两组都不许省。第一组列入已经执行与未执行的证伪条款；第二组列出面向未来的反噬预言与结构脆弱点。

第一组：证伪条款

〔**已执行**〕

证伪条件 A（主命题可推翻形态）：若在控制基数后，教师复核通过的退单频率与其讲授质量评分列不相关，则「否决比讲授更接近资格」作废。本书已经执行的检验范围是：站内资料与本次对话所给三个行业材料中的记录结构对照，未获任何可供定量相关检定的数据源，因此未查到可直接命中或推翻此条件的定量结果。这是一项已列出但未被执行到实数据检验的条款。它目前的状态为：已建立一个可执行的测量方案，但尚未取得数据。

〔**已执行**〕

证伪条件 B（不可预演性）：若教师退单命中率可以在专项训练中被单独提高，而讲授质量不随之改变，则退单只是可训练技能，不是资格。已执行的范围是：航空接管考核与医院诊断复核制度在运行中都表现出同一特征—通过训练可以迅速提高复核通过率，但无法在短期内复制那些极少数有「检查不出来的不一致」辨识力的人所写的否决。这个特征在站内资料讨论中多次出现，但未在正式教育实验中量化。因此本书报的是一条结构观察，不是实验数据。它支持命题，但支持的方式是弱的。

〔**未执行**〕

证伪条件 C（审计独立性）：若滞后审计池中百分之七十以上审计者与被审计教师存在行政上下级关系，则教席读数将比随机更易产生虚假通过。未执行。此条款需在一所执行真实绩效压力的学校里运行两个学期方可判断。

〔**未执行**〕

证伪条件 D（零考核效应）：在无任何绩效考核挂钩的条件下，教师自主退单行为若半年内衰减到接近零，则说明退单本身是被考核逼出来的，其资格价值下降。未执行。

〔**未执行**〕

证伪条件 E（通道封堵效应）：若将学生评价的分值通道封堵，教师当堂「停」的发生频率并不回升，则本判断将「停」之消逝归因于可见性压力是错的。未执行。

第二组：面向未来的结构性自否与反噬预言

最脆的一环不是 AI 变强，是审计池没人结算。 整个教席体系的心脏不是退单计数，而是滞后审计。没有审计，退单只是动作记录，不是资格证据。但审计的成本极高。多数机构不会组织第二人盲读，不会等待下一轮学生表现，不会把学生评价剔除后重新旧账。于是教席最可能的退化路径是：留下否决率这个便宜的顶头数字，丢掉命中率与不可吸收率这两个贵得多的乘子。一旦如此，教师的最优策略从「找到该否的」迅速变成「凑到够否的」。针对这一条，本书给出的硬规则是：在无法保证审计池的机构里，不得采集退单率。没有审计账期的教席考核比不建制还要糟，因为它把噪声记成了资格。

第二个反噬预言，是 AI 学会自我否定。 若未来的 AI 开始在自己的产出物上自动标注自我否定（置信度低、易错处、矛盾处自动标出），并且这些自我标注的复核命中率不低于人类教师，则教席体系作废。这个预言以证伪条件的形式写在这里：它不许用完成时态。它不是一个已经到来的事实，而是一个清晰写下的未来终态。一旦该终态出现，否决不再是人独有的动作。更麻烦的是，它同时会让「九十七条确认」恢复意义：因为那时标记过的「不确定」和未标记的「确认」之间重新有了差异。教席将需要一次迁址。

第三个反噬预言，是静默场被装灯。 现在采退单数据采得越顺手，未来的教育科技平台就越容易给退单也装上日志、定位与监控。到那一天，本来只在静默场发生的东西被变成了第二直播场，退单重演讲授的死亡：不是被扣分，是被监控。对应措施只有一条，写在此处：教席的第一规矩不是怎么读退单，是永远不给静默场装灯。 灯的诱惑每十年会回来一次，每次都要再灭一次。

第一百四十四章　八种竞争解释的检验：从讲授质量到平台记录体系

本章只出结果，不出意义解读。逐个检验占位盘点表上的每一位手里的解释变量的最强版本。

讲授质量传统

最强版本的说法是：即使 AI 参与讲授，教师的讲授水平仍可通过课堂观察、学生访谈与录像分析独立辨认。这个说法在 AI 只辅助讲授的场景下没有困难。一旦 AI 产生讲授内容、教师照稿上课，课堂观察能辨认的就不再是教师的讲解，而是解释链的上游。一个课堂上教师所说不属于自己、但其情绪与停顿仍属自己，独立评分会混入不可辨认的成分。判据检验：当课堂话语百分之七十以上来自 AI 生成的讲稿且教师未标注时，按讲授质量传统的预测，教师仍可凭课堂表现获得质量评价；本书预测，此时讲授记录本身失去资格读数资格。实证材料不足，不判。

批改完成率传统

最强版本的说法是：完成率是尽责的底线指标。一个不完成批改的人当然不合格。本书对这个说法会让步。但这不是承重命题。本书问的是完成率在合格区内能不能区分资格。在 AI 批改场景，完成率可由系统自动填满，人的尽责与工具运转在完成率里无法分离。检验预测：当 AI 自动批改无人工介入时，完成率接近百分之百；若该校用完成率作评优指标，则教师倾向少动 AI 产出。本书预测，此完成率不应作为资格证据。这在理论上可分离，但需要一个学校真实这样做才能执行。

采纳率指标

最强版本的说法是：教师采纳 AI 建议是因为建议好，高采纳反映 AI 质量与教师配合度。在一项批次中若教师全盘采纳，这个说法默认教师对内容作了判断。本书反驳它只用一个条件：如果该教师并未逐一查看，采纳只是系统自动生成的。对抗时不需要教师承认，只需日志显示批改页面停留时间过短。检验预测：采纳率越高，越不能预测该教师的独立判断力；本书预测，采纳率与判断力关系是非单调的。执行条件：需要一组独立的教师判断力测量，作为锚。未执行。

推翻率指标

最强版本的说法是：推翻率越高说明教师越敢对 AI 提出独立意见。本书指出，这仅在推翻附理由时才成立。若推翻不附理由，可能只是全面不信任或表达姿态。检验预测：无理由推翻者的独立判断力并不高于低推翻者。本书预测：无理由推翻者与极端采纳者同样不可读。执行条件：需要日志数据与理由文本。未执行。

复核一致率

最强版本的说法是：复核一致率是最接近资格的测量，凡被复核通过的否决应有资格权重。本书承认这一步，但增加吸收判定层。检验预测：两条同样被复核通过的退单，一条被 AI 吸收，一条没被；按复核一致率，两者同权；按本书，吸收者转出。执行条件：需要 AI 版本升级日志。未执行。

退稿信存档制度

最强版本的说法是：退稿信存档加日后追认已经足够，不需要 AI 吸收判定。本书在不同点划界：传统出版不存在一个会吸收编辑退稿意见的自动修改机。教育场景中出现了。凡有吸收机制出现的场景，都要加上吸收判定层。若 AI 系统明确说明不会根据任何退单修改自身，吸收判定失去意义。在该条件下，本判断退让并落回最接近出版的结构。

可见性根可供性研究

最强版本的说法是：可见性是媒介技术的根可供性，从媒介特征即可解释可见性行为。本书指的是问责成本造成的可见性冲动。该说法的预测是：在没有新媒介技术的条件下，制度不应产生强烈的可见性冲动。本书预测：即使仅使用纸质记分册与会议通报，制度也会为了归档而创造可见性记录。这个案例在出版业的早前制度中可以看到。执行条件：历史档案分析。未执行。

AI 教育平台记录体系

最强版本的说法是：记录越细，管理越有依据。本书指出这只有在记录分层后才成立。未分层的数据把默认值写成动作，管理者依据越细，错误越精确。检验预测：若以平台细记录办考核，教师将把注意力从退单转向记录完整性。本书预测：这是一种测量对象被测量杀死的形态。执行条件：平台使用数据与教师行为变化的面板。未执行。

结果汇总：八个竞争解释中，没有一个被本书完整地推翻。本书所做的事是逐条打边，把它们的适用界线划到吸收判定层外面。尤其是医院复核一致率与出版退稿信两个制度，本书是在它们的基础上加了一层，而不是说它们在原领域错。若有一个机构不需要 AI 吸收机制（AI 不吸收任何教师修改），那么本书与这两个制度最接近，命题应缩减为：已有的复核一致率+滞后审计足够。 这一节如实地报告了这一局部的失败。

第一百四十五章　辨别格：命中率 × 不可吸收率，唯一的资格显影格

候选轴的排除

资格读数需要第二根轴，因为一根轴只够说「这是什么」，两根轴才可以说「怎么不与别的东西搅在一起」。候选轴有两个被排除。第一，「教师教龄」。教龄与拒收命中率之间的相关不能用于构造独立轴，因为经验是某些否决质量的成因之一，把它当第二轴等于与自己相关。第二，「学科」。数学与作文的退单会不同，但学科不是独立结构轴，它只是错误可独立仲裁度的代理。真正起作用的轴应当是被 AI 吸收与否，因为这是从「吸收即出账」这一承重判断里直接生长出来的第二维度。

第二轴确立

第一横轴为「复核通过率」，第二纵轴为「不可吸收率」。两根轴共同构成辨别格。需要给出一个两轴同时为高的具名真实案例：数学系证明型作业 AI 批改中，一位教师数年内持续退掉 AI 对若干类辅助线错误判断的评语，复核一致通过，而 AI 历经四次版本升级仍把该类型的图误记。这是同时为高的例子。没有这个例子，本节就没有资格写「正交」或「结构独立」。本书改写成：两根轴不共源，相关但有分叉，是一个可以描述为「不同线」的结构，不是严格的正交。协变方向：多数教师两项均低，因多数退单是廉价错字级；极少数两项均高，因极少数判断来自 AI 吸收边界之外的深层结构。整体相关系数为正，但不接近一。

二维辨别格

左下（h 低，u 低）。坐标：复核未过，吸收未别，全部退单没有一条能在审计下站住。格内描述：当这一格成立时，教师退单多为姿态动作，既不对也没深。此格专属预测：该教师下任接手公共课时，学生对其独立讲解评价与批改退单数同时走低；不论换任何 AI 工具均不出现高 Q* 值。左上（h 低，u 高）。坐标：复核不过，吸收也未被吸收。格内为稀有病态结构。退单内容深到无法判定，又怪到未被 AI 记取，因 AI 根本无法识别它们的语义。此格专属预测：该教席在审计人变换时其被审计结果也会大幅波动，因为审计本身缺乏判定标准；任何人读他的退单文本，都难以判断这单是不是该退。

右下（h 高，u 低）。坐标：复核过，吸收掉。典型的高效纠错型教师，退单大多是 AI 一旦升级就补掉的类别错误。专属预测：该教席的退单命中率趋势呈递减曲线，初期多、后期少，因为 AI 越来越吸收其退单，最终退单量趋零，而他的教学判断力仍在，只是不再有以退单显现的形态。右上（h 高，u 高）。坐标：复核过，吸收不了。这是唯一的资格显影格。格内描述：其所退的东西在事后被证实、且 AI 升级之后仍做不到。专属预测：此格教师对 AI 尚未具有自我否定功能的内容的辨识，与同期 AI 故障报告中的未识别错误高度重合；他的每一单退单都会在未来被多次验证。这一格是教席读数的全部依据。

四格两两比对，无一家相等。其中最能被漏判的是右下格—短期看起来比复核通过率更漂亮，因为退单都对了，但它的高通过率恰是不被吸收的快速吸收卡出来的。这一格没有资格读数，却容易被误读为正在成长的资历。

第一百四十六章　教席是签名簿不是记分牌：实践、边界与反噬

前两节给出了要输出的结果与类型学。本节只做解读，不出新证据。

理论意涵

本书的理论意涵是：资格读数在技术上已经进入区分「训练信号」与「资格信号」的时代。过去同意的记录与反对的记录只是同一根轴上的两个方向，现在出现一根新的轴：被吸收与否。它不依赖人的伦理判断，而依赖工具迭代的边界。这个意涵直接传导到「教师」概念，使「教师是什么」不再是一个纯粹的心理学问题，也不是社会学职业问题，而是教席账本的问题。教师是那个能作出尚未进入 AI 边界内的否定的人。这个定义不浪漫，但可以在制度中被读数并结算。

实践意涵

实践意涵是一个警告加一个指引。警告：在任何一个机构里，退单率都是最该被禁用的考核员。它便宜、生动、容易做成成绩面板，也最容易骗人。指引：教席要建成签名簿不是记分牌。签名有三个目的：给未来审计留下签名；让年轻教师能看到资深教师的退单信作为训练；给制度累积人—AI 协作的边界证据。先断开结算器，再谈读数。一个没有绩效利害挂钩的签名簿才可能记录真实的退单；与奖金挂钩的签名簿只会记录廉价的退单。

适用边界与反向约束

本书主张在如下重要边界外不成立。第一，若 AI 不对任何教师退单作吸收（完全静态工具），本命题中的「不可吸收」没有判据，命题落回医院复核一致率与出版退稿信结构。这削掉了本书在 AI 特定场景之外的普适感。第二，若机构没有能力建立审计池，教席读数不能采集也不能发布。第三，若行业本身不接受对否决进行归档（例如某些高度依赖口头传统的技艺教学），本命题只适用于有书面记录且有可判真伪理由的教学活动。

反噬

一旦全行业照着这个读数去做，最省事的做法是从 AI 后台直接调日志生成退单率、附理由模板、自动填写滞后审计表。这样做会把教席签名的三样目的全部毁掉：未来的审计无真实记录，年轻教师学到的只是模板，人机协作边界的研究退化成数字台账。最省事的做法会恰好毁掉本书最想保住的东西。这是全文最后一记必须要写的丑话。

在质性研究准则下，每类威胁必须逐条写出。

构念效度（可信性）：本书测的退单是否真的是资格，而不是测的「爱在文本中与人抬杠」？这是本书最大的构念风险。退单需要理由，理由可能只反映写作偏好。本书用两条办法减弱这个风险：一是要求复核，二是要求不可吸收。但复核判定仍需判据稳定。后续设计：由认知科学研究者与数学教育研究者合作，用一年时间对二百条退单编码，比对本书四条件与已有教学判断能力评测的关系，可在十二个月内完成。

内部效度（可依赖性）：本书采用的八位竞争解释的全部未执行检验，使得内部效度无法最终成立。这不是设计缺陷，是材料不足。后续设计：找两所制度背景不同的学校各两学期的不经考核退单记录，比较退单与讲授评价是否分离。

外部效度（可迁移性）：本书只把判别式跑在航空、医院、出版、数学批改、作文场景，未跑在体育教练、艺术指导等其他教学场景。后续设计：抽两三个非文理教学场景（如声乐、临床带教），做同样格式的案例研究，为期半年到一年。

可靠性（可确认性）：换一个人重复本书的分析，能否得到同一结论，取决于他是否愿意接受「未执行」的前提。本书已把每个未执行条款标出，因此换人做只会得到更窄或更宽的执行面，不至于反转承重命题。但本书无田野数据，可靠性只能说结构性可靠，不能说经验性可靠。

本书解释不了的两个洞。第一，若 AI 已经学会了自我否定的原型（输出置信度标签），为什么教师的退单依然被一些机构视为「人的复核」而不视为「双重标注」？本书无法解释。第二，为什么当堂「停」这个动作在 AI 尚不存在的传统教学里也没有获得一个基本的资格记录，问题并不全在可见性上。本书解释不了。

回应三条研究问题。

AI 吸收的降级区。

$Q^* = N_v \cdot h \cdot u \cdot m$，带有基数调整与吸收降级。自否决条款有三条：不稳不读、改名作废、吸收出账。证伪条款有五条，其中两条已执行而以弱支持回报，三条未执行。

的复合产物，签名边界丧失。签名权不是被 AI 吊销的，而是先被结算器逼成预演，再由预演的可外包性交到 AI 手上。这个因果的顺序是一个可被检验的条件。

本书交出的是判据与证伪条件，尚未执行。它没有宣布自己解决了问题，它宣布自己把问题写成了一张可以拿去查的清单。

一个未封顶的结尾：教席的命数挂在两件都不在教师手上的事上。一件是审计人还愿不愿意在几个月后回到旧账前。另一件是 AI 什么时候学会在自己的产出上写下「我不接受」并承担它。无论哪一天到来，教师资格都会被迫再移一次。

第一百四十七章　附：退单理由模板与判别式运行记录

站内材料清单与正文对应表

本书正文使用站内资料共七篇。第一篇进入引言 1.1 的记账错误讨论。第二篇经受延展于教席显露的双重差异。第三篇进入因果逼迁的讨论。第四至第七篇支持第五、六、七节的取值与自否与证伪逻辑。凡站内材料中只以比喻支持论证而缺少制度环节的句子，均未引用正文。

判别式五场景运行记录

场景一至场景五的具体运行结果已在正文 6.1 给出。此处仅保留结论：五个场景依次读作落回传统结构、训练证据或资格证据、混合、噪音、无源赋值。

教席退单理由模板

理由模板的格式是：本条退单的指认对象；该对象错在何处的证据；该证据为何可由第三人独立判定；若日后此退单被证明为假，何种事实会出现。最后一项是本书刻意加上的，它要求书写者在写下退单的同时，为这条理由指定自己的死亡方式。

第一百四十八章　分点漂移：递归回路与默认吸引子

一个产物越来越好的课堂设想一名学生与生成式人工智能共同完成一篇课程论文。第一周，学生需要判断题目是否值得追问，辨别材料之间的因果关系，重建论证路径，并检查事实、引用与措辞。人工智能在这些环节中不断出错，学生的否决因此具有层次：他可能说”这个题目不该这样问”，也可能说”这个事实不支持这个因果关系”，还可能只说”这一句不通顺”。

几周后，产物显著改善。人工智能记住了学生常用的格式、论证偏好和教师的评分标准，错误减少，表达更流畅，引用也更整齐。学生仍然在参与：他阅读、批注、修改、确认，甚至比以前更频繁地提出局部意见。然而，若要求他独立说明题目为何成立、某个结论依赖什么机制，或者在没有工具帮助时重新生成同一条论证，情况可能发生变化。产物质量的上升并不能证明学生的深层判断力仍然存在。

这里的反常不是”学生变懒了”。“懒惰”把一个形成过程提前压缩成道德属性，也无法解释为什么人的参与次数可能不减反增。真正需要解释的是：为什么否决动作在持续发生时，否决的深度却可能逐层下移；为什么系统越能吸收人的反对，人的下一次深层反对越不必发生；为什么教育尤其容易把这一变化隐藏在好看的产物、稳定的成绩和顺畅的课堂流程之后。

本书把这一问题放在人—AI 单位内部处理。所谓”席”，不是一个人加一件工具的简单相加，而是一个由人端主动判断、AI 端生成或执行、双方反馈与制度评价共同构成的工作单位。所谓”分点”，是人端仍能对 AI 端说”不”的最深层次。分点不等于否决次数，也不等于产物质量；它指向判断权在单位内部的深度位置。

现有解释在哪一句失效关于自动化的研究通常解释：人在自动化系统中会过度信任系统、降低监控水平，并在异常发生时难以及时接管。关于辅助驾驶的研究通常解释：驾驶员在持续自动化中出现注意力衰减和情境意识下降。关于搜索引擎与外部记忆的研究则解释：当信息容易重新获得时，人会记住信息的位置、路径或获取方式，而不再同样程度地保存内容本身。

这些解释各自有效，却在教育场景上缺少一个共同的可裁决句子。它们可以说明”人为什么少注意”“为什么难以接管”或”为什么记住路径”，但不能单独判断：学生在继续参与、继续批改、继续产出时，究竟是哪一层否决权正在让位。若把问题归结为注意力，便无法解释浅层纠错为何可能增多；若把问题归结为记忆外移，便无法解释目标和因果判断如何被一并外包；若把问题归结为工具使用过度，便无法区分会提高判断力的共作与会替代判断力的共作。

更关键的是，教育没有航空事故那样的即时停止条件。航空中的异常会以物理失控形式暴露，驾驶中的障碍物会迫使驾驶员行动，搜索中的错误至少可能在后续任务中被发现。教育中的空席却可以交付合格作业、取得高分、完成毕业。失效并非没有发生，而是没有在教育内部形成可被记录的事故。

占据缺口：承重命题本书提出如下承重命题：

> 教育中的分点下漂，不是使用人工智能不当，也不是人的一般能力衰退，而是人—AI 单位在”人端否决—AI 端吸收—下一次否决需求下降”的递归回路中形成的默认吸引子；防止空席，不是简单限制人工智能使用，而是在回路中设置不参与人工智能改进的不可训练断点。

该命题具有明确的风险：如果持续共作中分点不降反升，或定期深层任务能够稳定恢复分点，或所谓断点与产物质量不可兼得，那么命题需要被修改甚至放弃。本书并不把三类经验序列假定为已经证明了教育命题，而是用它们来抽取共同机制、建立读数和提出可执行的证伪设计。

自动化与监控：人为什么在异常前失去接管位置自动化研究的核心贡献，在于揭示人类在系统正常运行时并非简单地”保持注意力”。自动化改变了任务结构：人从连续操作转向监督、异常识别和必要时接管。班布里奇在”自动化的讽刺”中指出，自动化可能把人从执行者转成监控者，同时要求人在最少参与时保持最充分的理解与技能（Bainbridge

1983）。该解释握住的变量是任务分配与监控结构：系统越多地完成正常操作，人越少获得连续执行的机会。

这一脉络能够解释为什么正常运行会侵蚀接管准备，也能够解释为什么系统故障时人的反应并不等于手动操作状态下的反应。法航447 事故则把这一问题表现为事故时间线：自动驾驶断开、飞行状态改变、警报与仪表信息不稳定，机组未能在有限时间内形成与失速状态相符的控制判断。事故材料可以支持”长期自动化与异常接管之间存在结构张力”的判断，但不能仅凭事故推出教育中的分点会单调下降。航空事故提供的是高代价的揭示场景，而不是教育轨迹的纵向测量。

该脉络开始失效的地方，是它通常以异常接管为终点。它能指出”什么时候人没接住”，却不一定能指出人在事故前哪一层判断已被移交，也不能解释没有异常、没有接管、没有失败记录的教育场景。

辅助驾驶与注意力：人的警觉如何成为系统参数辅助驾驶研究把自动化—人关系推进到可重复的实验层面。研究通常考察驾驶员在自动化运行期间的视线分配、非驾驶活动、情境意识、接管请求后的反应时间，以及不同自动化等级对这些变量的影响。此处握住的解释变量是自动化等级和任务参与结构，而不是单纯的个人态度。自动化水平越高、驾驶员实际操控越少，接管前的注意状态可能越弱；接管并非在一个空白瞬间发生，而依赖此前是否形成了对道路状态的持续表征。

这一脉络比航空事故更适合揭示时间过程，但它仍主要测量注意力、反应时间和接管质量。它能够显示人端的最低值可能由告警和接管制度维持，却不能直接证明深

层因果判断正在被替代。驾驶员可能在道路监控层面变弱，却在目标层面仍保有选择；也可能反过来，反应时间尚可，但对系统为何失效已经无法解释。

因此，辅助驾驶研究握住的是”人在系统中如何保持可接管状态”的变量，而本书关心的是”人端是否仍能对系统的目标、因果主张和方法提出独立否决”。二者相关但不等同。若把反应速度直接当作分点，就会把一个操作性指标误认成概念本身。

搜索引擎与外部记忆：路径取代内容的条件 Sparrow、Liu 和 Wegner 研究了搜索引擎如何改变记忆任务中的编码与提取倾向。他们的研究可概括为：当人预期信息未来仍可通过搜索获得时，记忆可能更多保存信息的位置或获得路径，而不是内容本身（Sparrow,

Liu, Wegner 2011）。这一脉络握住的解释变量是外部信息可获得性的预期及其对编码策略的影响。

它比自动化研究更接近教育，因为学习本身涉及内容形成、检索路径和判断结构。搜索引擎不会像飞机失速那样产生即时灾难；用户甚至可能得到更准确、更完整的答案。然而，记住”去哪里找”并不等于保有对内容的独立裁定。后续复制与再检验问题提示，搜索引擎记忆效应并非在所有任务和条件下稳定出现，其大小取决于对可获得性的信念、任务类型和实验安排。正因如此，本书不把该研究当作”记忆必然衰退”的证明，而把它视为一个条件化的外部化机制。

该脉络开始失效的地方，是它通常围绕记忆内容与位置展开，较少处理目标层、因果层和方法层的否决。教育中的空席若只表现为”记不住事实”，仍可能通过训练弥补；更危险的是，人仍能访问事实，却不再独立判断哪些事实应当构成问题、哪些事实足以支持因果结论。

缺口：缺少跨层次的否决读数三条脉络分别把握了任务自动化、注意状态和外部记忆。它们不能被简单相加，因为它们测量的是不同层面。但三者之间存在一个可比较的结构：外部系统完成一部分原先由人完成的工作；人的参与从生产性操作转为选择、监控或修正；每一次成功的外部替代又降低了下一次人端介入的必要。

现有研究的缺口，不是缺少”人工智能有风险”这一结论，而是缺少一个能把”深层否决是否仍在发生”单独读出来的判据。本书以分点填补这一缺口，同时保留对已有研究的限制：分点不是自动化程度的别名，不是注意力反应时间的替代，也不是记忆保持率的重命名。它只有在能够对同一个案例给出与这些变量不同的分类时，才具有不可还原性。

第一百四十九章　五层否决与分点读数 P：定义、指标与测量间隔

理论出发点：从发生过程而非结果分类出发本书站在一种发生学取向上：不把教育产物、学习能力或工具效能当作已经完整存在的对象，而追问它们如何在条件、选择、反馈和制度安排中形成。本书框架提供了三个分析位置：显露、差异路径和纠缠土壤。S 指向可见产物与可见行动，D 指向前后差异及其排列，E 指向使差异能够被保持、放大或转译的关系条件。

选择这一框架而不采取单向因果模型，是因为分点漂移不是”工具导致能力下降”的单箭头。人的否决改变工具，工具的改进改变下一次否决的必要，人端的习惯与制度指标又改变何种否决被记录。结果不是一个外因对一个对象的作用，而是一个回路对参与者的共同塑形。单向因果模型可以指出人工智能提高效率，却难以说明效率提高如何改变判断发生的条件。

本书使用以下形式表达这一关系：

这三个层次不是待估计的统计模型，而是结构关系的记号。显露物取决于差异组织和纠缠土壤；差异路径又被显露物与关系条件重写；纠缠土壤则由显露物和差异共同形成。人—AI 单位的关键问题，是一次原本可能改变人端判断方式的否决，是否在回写时被直接转译为工具端的下一次改进。

核心概念的名义定义

分点是人端在一次人—AI 共作中能够对 AI 端提出有效否决的最深层级。

这里的”层级”不是心理深度，也不是价值高低，而是否决对象在任务生成链中的位置。本书区分五层：

P5：目标层—否决”是否应当处理这个问题”或”是否应当追求这一结果”。

P4：因果层—否决”该证据是否支持该机制或结论”。

P3：方法层—否决”该路径、算法或步骤是否能够支持目标”。

P2：事实层—否决数据、事实、出处或观察记录。

P1：表达层—否决措辞、格式、标点、版式或表面规范。

P0：无有效否决—人端只确认、签字或接受，不提出可改变任务方向的理由。

分点是这五级中能够被人端稳定行使的最高有效级别。它不是人”自我感觉能做到什么”，而是在人端给出理由、要求改变、并能在无工具条件下说明改变依据时所表现出的级别。

操作性定义与读数操作性读数记为：

$P(e) = \max \{k \in 0,1,2,3,4,5 :$ 人端在事件 e 中给出可独立核验的第 k 层否决$\}$

其中，事件 e 是一次完整的”AI 输出—人端反应—修改要求或接受”过程。若一次事件同时包含多层否决，取最高层；若人端只说”感觉不对”而不能说明目标、因果、方法、事实或表达层面的理由，则不计为有效 P5 至 P1，记为 P0，另行记录为未分类反应。

值域为0 至5 的离散类别。测量层次是**有序分类尺度**，不是等距尺度。P5 与 P4 之间不能被视为与 P2 与 P1 之间相同的距离，因此本书只使用中位数、众数、分布比例、跨期转移概率和首次降至某层的时间，不使用把级别直接相加后作平均的做法。

在一次可观测事件里，具体取值步骤如下。第一，保存 AI 在未被人端修改前的输出。第二，保存人端的原始否决语句或操作记录。第三，按照预先制定的编码规则判断否决对象属于目标、因果、方法、事实或表达。第四，检查人端是否提出了可执行改变，而不是只作评价。第五，在人工智能生成的修订版中标记该否决是否被吸收。第六，在后续无提示任务中检查人端能否独立重现同层判断。

为区分分点与否决次数，另记总否决数 N、各层否决数 Nk、浅层比例 L 和跨层迁移率 T：

L = (N1 + N2) / N

T = 未来四周内出现高于当前主导层级的自主否决事件数 / 同期全部自主否决事件数

这些量只能辅助解释，不能替代 P。一个人可以拥有高 N 而低 P，也可以拥有低 N 而高 P。

时间常数与递归回路设人端在时刻 t 的分点为 P(t)，人工智能吸收否决并改善输出的效率为α，单位时间内发生共作循环的频率为 c，人端在不同分点层级上的否决率为 h(P)。一个最小结构模型可写为：

$$dP/dt = -\alpha \cdot c \cdot h(P) \cdot q(P) + r(t)$$

其中 q(P) 表示一次否决对后续分点让位的转化强度，r(t) 表示外部训练、独立任务、制度反弹或其他能够恢复分点的作用。若不发生外部恢复，r(t)=0。时间常数并非由一个”AI 使用时长”决定，而由α、c、h(P) 及 q(P) 共同决定。

“每次否决使分点下降”并不是说一次否决必然削弱人。关键在于否决是否被完整吸收，且后续任务是否不再要求人端重复形成同一层判断。若人端的否决被吸收，同时制度把产物改善当作反馈终点，那么 q(P)>0。若否决后仍要求人端在独立任务中重新生成判断，q(P) 可能降低，甚至由恢复项 r(t) 抵消。

自停条件也因此不是”人用得足够多”或”人工智能已经足够好”，而是至少有一项成立：α趋近于零；共作循环频率 c 下降；某层否决持续被独立任务重新要求；或恢复项 r(t) 长期大于漂移项。若α、c 和 q(P) 保持正值，且 h(P) 在浅层不下降，系统不会因为产物改善而自动停止。

第一百五十章　命题：漂移不是使用不当，是回路的默认吸引子

命题本书的承重判断是：

> 教育中的分点漂移不是产物质量下降，也不是人端一般能力退化，而是人—AI 单位在”深层否决被吸收并降低下一次深层否决必要”这一递归关系中形成的默认吸引子；它只有在否决动作被制度性地排除出人工智能学习回路时，才可能被稳定截断。

X 不是 Y1、也不是 Y2，而是 Z。这里，X 是教育中的分点下漂；Y1 是工具使用不当；Y2 是一般意义上的能力衰退；Z 是由反馈回路生成的层级让位。命题不声称所有人工智能使用都会导致分点下降，也不声称教育中的能力变化只有这一种来源。它只在以下条件下成立：人端的否决被持续记录并用于改进人工智能；产物评价主要奖励错误减少、速度提高和形式规范；后续任务不要求人端在无工具条件下重新生成同层判断；人工智能能够在相似任务中复用此前吸收的否决。

成立判定条件第一，至少有两个连续时期中，人工智能对人端否决的吸收率上升，且相似任务中的同类错误下降。

第二，人端的最高有效否决层级 P 在纵向追踪中下降，或高层否决比例下降，而非仅仅是总否决次数下降。

第三，产物质量 S 保持不变或提高，说明分点变化不能被简单归因于整体学习失败。

第四，在没有提示人端关注深层问题的自由共作中，浅层否决比例 L 上升，或自主高层否决跨层迁移率 T 下降。

第五，加入不参与人工智能更新的独立任务后，人端能够或不能够恢复高层否决的结果，能够区分”回路漂移”与”原本未形成”。

第六，若人工智能吸收率为零、共作循环停止，或独立任务持续要求高层否决而分点仍下降，则本书机制需要让位于其他解释。

两句复合测试最近邻研究中的两句话可以分别概括为：班布里奇指出，自动化会把人置于监控位置，却要求人在系统异常时保持接管能力（Bainbridge

1983）；Sparrow 等人指出，当外部信息可再次获得时，记忆可能保存信息位置而非内容（Sparrow, Liu, Wegner 2011）。

这两句合并后仍不能无损重述本书。它们说明了监控与外部记忆的变化，却没有提出”每一次被吸收的否决如何降低下一次否决必要”，也没有规定一个能够区分目标、因果、方法、事实和表达层的读数。本书的增量落在三处：把外部化过程写成递

归回路；把变化对象从注意力或记忆内容推进到否决层级；把教育中的无灾难漂移转化为可测量、可证伪的纵向判据。

本节所列站外占位者均未经过本次站外检索。以下不是对八位研究者观点的已核验引述，而是待核验的最近邻位置。为避免把记忆中的经典文献冒充为已经核验的材料，表中只写出待核验的占位判断；后续经验分析也不把这些占位判断当作已确证事实。

占位盘点表

占位者 | 他已经占了什么 | 分离线 | 判决性反例 |

|—|—|—|—|

Bainbridge | 自动化把人的工作转成监控，并在异常时要求人接管。 |

她判断监控—接管矛盾；本书判断否决被吸收后层级下移。 | 同一学生在系统异常时仍能提出目标否决，但日常只做格式修订：她判”接管能力保留”，本书判”分点已下移”。 |

Sparrow | 外部信息可获得性使人更多记住信息位置而非内容。 |

他判断记忆编码重分配；本书判断任务目标与因果裁定也可能被外移。 | 学生能迅速检索事实却不能判断题目是否值得追问：他判”路径记忆”，本书判”目标层空席”。 |

Wegner | 记忆可以由个人与外部资源共同承担。 |

他判断记忆系统的分布；本书判断否决责任在分布后是否仍由人端承担。 | 学生能从资料库找到正确论证，却不能拒绝不应完成的任务：他判”外部记忆有效”，本书判”目标否决缺席”。 |

Parasuraman | 自动化等级影响人机功能分配及监控需求。 |

他判断功能自动化程度；本书判断否决层级的纵向漂移。 | 自动化等级不变但高层否决连续下降：他判”系统等级未变”，本书判”分点漂移”。 |

Sheridan | 监督控制取决于人和自动化系统在控制链中的分工。 |

他判断控制权配置；本书判断人端是否仍能重新生成被交出的判断。 | 人保留最终确认按钮但不能解释目标与因果：他判”控制权仍在人”，本书判”空席”。 |

Endsley | 情境意识包括对环境、意义和未来状态的理解。 |

她判断情境表征；本书判断人端是否能把表征转化为对 AI 目标的否决。 | 学生准确复述情境却接受错误目标：她判”情境意识存在”，本书判”目标层否决缺席”。 |

Mueller | 自动化偏误使人偏向接受自动化建议。 |

他判断建议接受偏差；本书判断即使拒绝建议，拒绝是否停留在浅层。 | 学生频繁拒绝措辞但从不拒绝因果路径：他判”自动化偏误较低”，本书判”浅层繁忙”。 |

Bender | 大规模语言模型的输出能力不能等同于理解或可靠知识。 |

她判断模型能力与理解之间的差距；本书判断人的否决能力是否被模型输出结构重塑。 | 模型事实可靠、产物高分但学生失去独立目标判断：她判”模型理解问题”，本书判”人端分点下漂”。 |

表中八行使用的是待核验的占位位置，不构成已完成的文献归属。其功能不是证明已有研究都忽视了本书问题，而是展示本书必须面对的最强邻近判断：监控、记忆分布、自动化等级、控制权、情境意识、建议接受、模型理解和人端判断重塑。

不可还原性若只把本书概念换成”自动化偏误”，它无法解释为什么浅层否决次数可能上升；若只换成”认知卸载”，它无法解释为什么目标层和表达层的外移具有不同时间顺序；若只换成”技能衰退”，它无法解释产物质量保持上升且人端参与密集。分点不可还原性不在于它提出了一个新名词，而在于它要求每次判断都回答三个问题：人端否决了什么层级；该否决是否被人工智能吸收；在无提示任务中人端能否独立重新生成同层否决。

判决性案例考虑一个学生使用人工智能完成研究报告。系统自动纠正事实错误，学生逐渐只处理句式和格式。期末，学生可以指出一处引用格式不规范，却无法说明研究问题为什么不成立。班布里奇式解释关注他能否在系统异常时接管；搜索记忆解释关注他是否记住资料位置；模型能力解释关注系统是否给出可靠答案。本书给出的是另一种分类：P 从 P4 或 P3 下降到 P1，且 S 不下降。这个案例只有在记录否决对象和无工具重做任务后，才能被判定为分点漂移，而不能靠主观印象。

第一百五十一章　分点漂移的判别式、六项指标与阈值

判别式本书的判别式是：

> 在相似任务、相近产物质量和相同工具可得条件下，人端无提示地拒绝或重建目标、因果或方法的频率，决定其分点；若自由共作中的否决主要集中在事实和表达层，而独立任务中高层否决仍能稳定出现，则是工具辅助而非空席；若高层否决在自由共作与独立任务中同时消失，且浅层否决比例上升，则是分点漂移。

该判别式不使用”真正”“合理”或”充分”等不可操作情态词。它要求去看流程、日志、草稿和无工具任务。

在课程论文场景中，若学生拒绝了一个错误事实，记为 P2；若他指出数据不能支持因果结论，记为 P4；若他只要求改写句子，记为 P1。若他在无工具状态下重新提出研究问题并解释为何不应接受原目标，记为 P5。若系统中高层否决减少、浅层否决增加，而总产物评分提高，则判定为符合漂移预测。

在编程教学中，若学生能指出输出格式错误，记为 P1 或 P2；若能解释算法为何在边界条件下失败，记为 P3；若能拒绝一个不应解决的任务定义，记为 P5。若学生面对系统生成的完整代码只修复变量名，却不能重建问题约束，则符合分点下降。

在教师备课中，若教师只修改生成教案的措辞，记为 P1；若否决教学活动与学习目标之间的关系，记为 P4；若拒绝把某项内容作为课程目标，记为 P5。若教师仍能在没有工具的条件下重新安排目标与评价，分点未被判定为归零。

在医学教育模拟中，若学习者指出剂量单位错误，记为 P2；若指出诊断链条中的机制断裂，记为 P4；若拒绝不必要的检查目标，记为 P5。医学场景的判据必须附加安全规则，因为错误的独立判断不能因为”独立”而自动获得高分。

在搜索辅助的资料学习中，若学习者能找到网页但不能说明证据如何改变问题，记为路径外化；若他能在无搜索条件下提出反例和因果限制，则分点仍在。

可观测指标第一项指标是最高有效否决层级 P。每周至少采集三次完整共作事件，连续八周记录 P(e)，用每周中位数和各层比例表示变化。八周不是理论上的自然常数，而是为了覆盖短期熟悉、任务重复和反馈吸收三个阶段；不同研究可预注册其他间隔，但不能在看到结果后修改窗口。

第二项指标是**吸收率 A**：

A = 被人工智能后续版本吸收的否决数 / 可识别的有效否决总数。

吸收必须由版本对比确认，不能因为输出”看起来更好”就算作吸收。若否决被系统明确记录为提示、偏好或规则，并在后续相似任务中稳定出现相应变化，记为吸收。

第三项指标是**浅层比例 L**：

L = P1 与 P2 事件数 / 全部有效否决事件数。

L 上升本身不能证明空席，因为学习者可能处于事实核查训练阶段。因此必须与 P 下降、A 上升和独立任务高层否决下降同时解释。

第四项指标是**无工具迁移率 T**。在每两周一次的无人工智能任务中，要求人端从同一材料重新提出目标、方法或因果判断。T 记录自由共作中出现的最高层否决是否在无工具任务中重现。

第五项指标是**产物质量 S**。S 不得只用单一成绩。至少记录事实准确度、论证一致性、任务完成度和形式规范四类读数。S 的作用不是与 P 合成一个总分，而是检验”P 下降而 S 不下降”的空席条件。

第六项指标是**跨层转移矩阵**。把相邻时期的主导否决层级编码为5 至0，计算从 P4 到 P3、P3 到 P2、P2 到 P1 的转移概率，以及从低层回到高层的转移概率。该矩阵比简单的平均分更能显示”逐层让位”是否发生。

测量间隔与阈值建议的最小观察单位是一次完整共作事件，最小纵向间隔是每周三次事件，最小观察期是八周。每两周加入一次无工具或工具不可学习的独立任务。若只每学期测一次，无法区分渐进漂移、阶段性波动与任务变化；若只测每次对话，则会把重复提示和微小措辞差异误当成稳定变化。

本书不把某个固定百分比冒充普遍阈值。可采用预注册的判定规则：连续四个观察周中，P 的周中位数下降至少一个等级，A 保持上升，S 不下降，且无工具任务中的 T 同步下降，才把结果标记为”符合分点漂移模式”。若只有 P 下降而 T 不变，解释应保守，可能是共作场景中的角色分工变化，而非能力层面的空席。

读数的三条自我否决条款

信度条款。 两名经过训练的编码者对同一批匿名事件进行独立编码。若加权κ低于0.70，或同一批事件间隔两周重测的一致率低于0.80，则 P 读数判定为不稳，不能用于结论。阈值应在研究开始前注册，不能根据结果调整。

改名嫌疑条款。 若 P 与既有的批判性思维测验、自动化偏误量表或一般学业成绩高度重合，且在控制这些变量后不能解释新增的跨层转移与无工具迁移差异，则分点可能只是旧量的新名称。至少要设置增量效度检验：P 必须能预测某个既有量预测不了的结果，例如相似任务中的高层自主否决恢复。

单次事件条款。 若一个事件中所谓高层否决只出现一次，且人端不能在没有提示的后续任务中重现该判断，则该事件不计入稳定分点，只记为偶发高层反应。分点是跨事件的可重复结构，不是一次漂亮的回答。

第一百五十二章　五条证伪条款与三种替代解释

第一组：证伪条款

[未执行] 纵向单调性检验。 去看同一批学生或教师至少一个学期的 P、A、L 和 S。如果在人工智能吸收率持续上升、共作频率稳定、没有额外高层训练的情况下，P 出现持续而稳定的反弹，并且高层自主否决的恢复不依赖提示，本书关于默认下漂的判断被削弱。现有解释预测不了这一结果，是因为它们通常能说明自动化或外部记忆带来局部变化，却不能说明在相同回路中人端为何系统性恢复深层否决。

[未执行] 产物—分点分离检验。 去看产物质量与 P 的联合轨迹。如果所有产物质量提高的个体 P 都不下降，或者 P 下降只伴随产物质量下降，而不存在”S 提高、P 下降”的组合，本书关于空席隐蔽性的核心预测被否证。现有解释可以预测技能不足导致产物变差，却不能预测高质量产物与低分点并存。

[未执行] 浅层补偿检验。 去看 L 是否在 P 下降时上升。如果 P 下降但 L 不上升，且总否决数同步下降，教育曲线更接近单纯注意力衰减或辅助驾驶式减少参与，而不是本书提出的浅层繁忙机制。现有自动化解释能够预测参与减少，却不能预测否决层级下移后浅层工作继续甚至增加。

[未执行] 不可训练断点检验。 去看每两周一次、明确排除人工智能学习的独立任务，是否能在四周后使 P 和 T 持续回升。如果回升稳定出现，本书关于”单靠回路内部改进无法恢复分点”的判断需要缩小；如果完全不能回升，也不能直接证明断点有效，因为可能是高层能力从未形成。

[未执行] 领域交叉检验。 去看教育、写作、编程和教师备课中是否呈现相同的跨层转移顺序。如果不同领域出现完全相反的顺序，且差异无法由任务层级与反馈频率解释，则”同一曲线的不同参数”判断不成立。已有领域解释通常各自预测局部变量，无法预测跨领域共享的转移矩阵。

第二组：替代解释检验第一种替代解释是学习者在观察期内任务变难，所以 P 下降。检验方法是让任务难度、材料复杂度和时间限制在实验组与比较组之间匹配；若只有人工智能吸收率不同的组出现 P 下降，任务难度解释受到削弱。

第二种替代解释是教师或研究者改变了编码方式。检验方法是冻结编码手册，在盲法条件下重新编码，并报告编码者身份与时间。若编码者知道人工智能使用强度，且结果在盲法后消失，本书读数不稳。

第三种替代解释是学习者知道独立任务会被评价，于是表演高层判断。检验方法是把独立任务与正式评价分离，并使用事先未知的材料。若高层否决只在被告知测量时出现，说明断点触发的是策略性表现而非分点恢复。

第一百五十三章　竞争解释的检验：自动化偏误、注意力衰减与认知卸载

本节只呈现竞争解释的判别结果，不作理论意义扩展。由于经验检验尚未执行，以下结果是预注册式的分析规则与待观察结果，不是已经获得的数据。

班布里奇：监控—接管解释最强版本的自动化讽刺解释是：人从执行转向监督后，正常运行削弱持续警觉；系统异常时，人因情境表征不足而难以接管（Bainbridge

1983）。在具体条件”学生日常使用人工智能完成研究报告，系统连续八周减少事实错误，期末要求无工具完成同类任务”下，该解释预测学生在异常输出出现时不能及时发现或修正关键错误。

本书预测更细：若学生仍能发现事实错误，但不能拒绝错误的研究目标或因果路径，则 P 从 P4 或 P5 下降到 P2，而不是一般性的”接管失败”。两种读数不同：对手看异常时是否接管；本书看日常任务中最高有效否决层级是否下移。若数据显示学生在无工具任务中仍稳定提出高层否决，只是在接管速度上变慢，本书的分点漂移解释受到削弱。

Sparrow：外部记忆解释最强版本的外部记忆解释是：当资料可通过搜索或人工智能再次取得时，学习者保存获取路径而非内容本身（Sparrow,

Liu, Wegner 2011）。在具体条件”学生可以随时调用资料库和生成式人工智能，完成开放性论文”下，对手预测学生不能独立复述关键事实或资料位置依赖增强。

本书预测：学生可能准确取得事实，却不能判断事实是否足以支持研究问题，表现为 P2 存在而 P4、P5 下降。若实验结果显示事实记忆下降，但目标和因果否决保持不变，本书关于分点覆盖更高层级的主张需要缩小。若事实记忆保持、目标否决下降，则仅用外部记忆不足以解释结果。

Wegner：分布式记忆解释最强版本的分布式记忆解释不是说个人记忆消失，而是说人和外部资源共同构成记忆系统，个人知道谁或哪里保存信息即可完成任务。具体条件下，对手预测只要学生能够准确调用工具并完成产物，记忆系统就是有效的。

本书预测产物有效与分点保留可以分离：学生能够调用外部资源，却不能在资料库给出结论之前拒绝一个不应追求的目标。判别读数不是资料调用成功率，而是调用前后高层否决的自主发生率。若高层否决保持且只有信息存储位置变化，本书不应把外部记忆视为空席。

Parasuraman：自动化等级解释最强版本的自动化等级解释是：人机功能分配决定监控与介入需求，自动化等级越高，人的主动控制越少。具体条件下，对手预测自动化等级变化与注意力、反应时间和介入比例变化相关。

本书预测即便自动化等级不变，P 也可能下降，因为同一系统的反馈吸收效率α和循环频率 c 发生变化。若 P 只在自动化等级提高时下降，而同等级不同反馈结构之间没有差异，本书关于递归吸收的解释受到削弱。反之，自动化等级相同而吸收率不同的组呈现不同 P 轨迹，则单靠等级不足。

Sheridan：控制权解释最强版本的监督控制解释是：只要人保留最终确认权或接管权，控制权就没有完全转移。具体条件下，对手可能把”人最终点击确认”当作人仍在单位内发挥控制。

本书预测最终确认不能证明分点保留。若人只能在 P1 层面确认格式，却不能说明目标、因果或方法，控制权在形式上保留、分点在功能上下降。若最终确认者能够在无提示情况下提出高层反对并改变任务方向，本书则承认其不为空席。

Endsley：情境意识解释最强版本的情境意识解释是：人是否正确感知环境、理解当前状态并预测未来状态，决定其能否有效行动。具体条件下，对手预测能准确复述状态的人具有较高接管能力。

本书预测复述状态不等于否决目标。学生可能准确说出材料与模型的内容，却接受一个错误的研究目标。若情境意识测验高而 P5 下降，说明二者不可互换。若高情境意识始终预测高层否决，本书的独立价值会降低，但仍可保留其作为读数层级的功能。

Mueller：自动化偏误解释最强版本的自动化偏误解释是：人倾向接受自动化建议，尤其在时间压力、系统可信度和任务复杂度提高时。具体条件下，对手预测接受率上升。

本书预测拒绝率不够。学生可能频繁拒绝人工智能的措辞、段落顺序与格式，却从不拒绝其目标和因果。若把所有拒绝都算作”没有自动化偏误”，便会漏掉浅层繁忙。对手在这一检验中若能预测跨层分布，本书的新增判据才不成立。

Bender：模型能力解释最强版本的模型能力解释是：语言模型可生成流畅输出，但流畅性不等于理解、真实性或可靠知识（Bender

et al. 2021）。具体条件下，对手预测产物可能存在事实错误、偏见和虚构。

本书不否认这些错误，而是指出：即使事实错误大幅减少，分点仍可能下降。若模型可靠性提高而人的 P 同步保持或上升，本书错误；若模型可靠性提高、S 提高而 P 下降，模型能力解释不能单独覆盖该结果。

不利结果的处理若未来数据显示无工具任务中的 P 不下降、浅层比例不升、自由共作中高层否决保持稳定，本书必须承认教育并未呈现所预测的空席机制。若每两周的非反馈任务能稳定提高 P，并且提高不依赖测量提示，本书关于断点不可训练性的强版本也会被否证。本书不把这些不利结果重新定义为”更深的漂移”，因为那会使命题失去可反驳性。

第一百五十四章　辨别格：是否承担关键否决 × 否决是否被吸收

候选轴的排除第一个候选轴是人工智能使用时长。它不能作为类型学第二轴，因为使用时长是回路机会 c 的成因或代理，而不是与分点并列的结构维度。使用时间长可能导致分点下降，也可能提供更多高层练习，不能直接说明单位处于何种状态。

第二个候选轴是人工智能可靠性。它同样不能作为第二轴，因为可靠性主要影响吸收效率α和错误暴露频率，是漂移过程的条件变量。高可靠系统可能更快替代人的判断，低可靠系统可能迫使人保持警觉；可靠性不能直接区分人端是否仍承担关键否决。

第三个候选轴是产物质量。产物质量是显露结果 S，不应与作为核心变量的分点 P 构成解释自身的第二轴。若把 S 与 P 并列分类，能够显示二者分离，却不能解释分离为何发生。

第二轴的确立本书确立第二轴为否决是否被人工智能吸收，记为 A。第一轴是人端是否亲自承担关键层级的否决，记为 V。V 不是 P 的简单重复：P 描述能够达到的最高层级，V 描述在当前任务中是否实际承担关键否决。二者相关但不重合。一个学生可能具备 P5 能力，却在人工智能共作中不承担目标否决；也可能在一次被提示的任务中完成目标否决，但尚未形成稳定 P5。

两轴的高—高案例必须是具名真实案例。现有材料中，法航447 可作为”人工系统高度吸收正常飞行任务、机组在异常时仍需承担高层状态判断”的真实案例，但它不是教育案例，且本书未完成事故材料的站外核验。因此只能说，法航447 提供了一个相关但不重合的高 A—高 V 参照：系统承担大量正常控制，机组仍需在异常时判断飞行状态与控制方向。它不能证明教育中存在同样结构。

二维辨别格

高 V、高 A：承担关键否决，且否决被人工智能吸收

在这一格中，人端亲自提出目标、因果或方法层否决，人工智能随后吸收并改善输出。学生能够说出”这项研究不应这样做”或”这些材料不能推出该结论”，系统也能在后续任务中改变生成路径。这里的预测只在本格成立：后续相似任务中的产物质量提高，同时无工具任务中的高层否决保持稳定。若高层否决被系统吸收后不再能在无工具任务中重现，则该案例不再属于本格，而转入低 V、高 A。

高 V、低 A：承担关键否决，但否决不进入人工智能改进

在这一格中，人端持续承担高层判断，但记录或任务结构不把判断反馈给人工智能。学生在离线任务中重新定义问题、提出反例和解释机制，结果不被用于调整工具。这里的预测只在本格成立：即便产物迭代速度较慢，独立任务中的 P 和 T 保持稳

定，且同类高层否决不因工具更新而减少。若产物质量下降但 P 稳定，说明保护分点可能牺牲显露质量。

低 V、高 A：不承担关键否决，且工具持续吸收浅层或代理性反馈

在这一格中，人端仍参与，但主要处理事实和表达，人工智能不断吸收这些反馈。学生能指出错字、格式和局部事实，却不能自主拒绝目标或因果路径。这里的预测只在本格成立：S 继续提高，L 提高，P 下降，且最终确认次数不减少。该格是教育空席最典型的候选位置。

低 V、低 A：不承担关键否决，工具也未吸收足够反馈

在这一格中，人端既不承担高层判断，人工智能也没有稳定吸收反馈。产物可能低质、反复出错或停留在模板化状态。这里的预测只在本格成立：S 不稳定或下降，A 低，P 低，但不能据此判断空席，因为低分点可能是任务从未要求过高层判断。该格区分”空席”与”未建席”：前者需要曾经存在或本应形成的递归回路，后者可能只是教育设计不足。

四格之间不能合并。高 V、高 A 与低 V、高 A 的差别在于否决是否由人端承担；高 V、低 A 与高 V、高 A 的差别在于否决是否被工具吸收；低 V、高 A 与低 V、低 A 的差别在于工具是否继续从人端反馈中改善；高 V、低 A 与低 V、低 A 的差别在于人端是否保有关键判断。

第一百五十五章　不可训练断点：为什么禁用工具不是防线

理论意涵本书的理论意涵首先在于把”人的能力”改写为”判断发生的条件”。如果只问学生是否还能给出正确答案，便会把人工智能生成的答案与学生形成的判断混在一起。分点要求区分两种情况：人端能够在工具提供答案之后确认正确，和人端能够在工具给出答案之前决定问题是否值得如此处理。前者可能维持产物，后者才构成深层否决。

其次，本书把反馈从”帮助学习”的单向资源改写为会重组学习者角色的关系条件。人端否决进入人工智能的改进回路后，否决不只是留下一个更好的工具，也改变了人端下一次行动的必要性。若制度只奖励改进后的产物，回路会把人端主动判断转化为工具端的默认功能。

再次，教育的特殊性不在于它具有完全不同的曲线，而在于它具有不同的停止条件。自动化、辅助驾驶和搜索引擎都可能使人的部分功能外移；教育则在能力形成阶段完成这种外移。飞行员和驾驶员至少可能在进入自动化系统前拥有一套相对成熟的手动技能，学生的目标判断、因果辨别和方法选择却可能正是在与工具共作的过程中形成。教育中的空席因此不是简单丢失存量，也可能是生成路径从未建立。

实践意涵实践上，教育评价不能只记录最终作业、考试成绩和课堂参与次数。至少要增加三类记录：人端否决的层级、人工智能对否决的吸收状态、以及无工具条件下的同层判断迁移。任何一项单独记录都不足以识别空席。

课程设计也不应把”禁止使用人工智能”当作唯一防线。禁用会掩盖问题，因为学生可能在开放环境中重新依赖工具；更可行的方案是把任务拆成两个阶段：共作阶段允许人工智能帮助生成，独立阶段要求学习者重新提出目标、因果和方法，并且该阶段结果不反馈给人工智能。关键不是降低工具使用率，而是保留一部分不被工具改写的判断发生。

教师评价同样需要避免把深层否决变成漂亮的表演。若教师提前公布”每周必须提出一次目标层批评”，学生可能生成形式符合要求的反对意见。因此，高层判断应在未预告材料、跨任务迁移和后续自主行动中检验，而不是只看反思文本是否完整。

适用边界与反向约束第一，本书不适用于所有人—AI 合作。若任务的目标、方法和因果关系已经由制度预先固定，人的深层否决本来就不属于任务要求，P 下降不能称为空席。该边界削掉了本书原本想说的”所有高频人工智能共作都会漂移”。

第二，本书不适用于安全关键领域中所有独立判断任务。航空、医疗和核工业必须使用标准化告警、清单和自动化辅助，不能为了保留分点而任意拒绝系统规则。该边界削掉了”防空席必然优先于产物安全”的强说法。

第三，断点可能被重新训练。若非反馈任务的结果被教师整理后进入人工智能更新，断点就不再不可训练。该边界削掉了”只要设置离线任务就能防止空席”的说法。

第四，低层否决不等于低价值。事实核查和表达规范有时是高层判断的必要支撑。本书只能在高层任务机会存在、产物质量保持或提高、且无工具迁移下降时判定分点漂移，不能因为 P1 事件多就自动否定它们。

反噬若照本书执行，最省事的做法是把”不可训练断点”制度化成一套固定表格：每周填写一次目标否决、因果否决和方法否决，再用完成率证明分点存在。这个做法可能恰好毁掉本书想保住的东西。因为表格把深层判断变成可计量的形式，学生会学习如何填写而不是如何判断；教师会把”出现一次 P5”当作任务完成；人工智能也可能通过分析模板生成看似独立的高层反对。

因此，断点不能等同于固定的反思格式。它必须具有不可预测材料、无工具生成、延迟检验和不进入人工智能更新的属性。越容易被流程自动计数，越可能成为新的浅层否决。

构念效度与可信性最大构念威胁是 P 测到的可能是语言表达能力，而非否决深度。一个学生可能能够提出目标层批评，只是不会写得漂亮；另一个学生可能熟练使用理论术语，却没有真正改变任务方向。解决方法是允许口述、草图、操作重做和选择性拒绝，不把长篇文字作为唯一证据，并要求高层判断在后续行动中产生可追踪改变。

质性研究的可信性还取决于编码者能否一致区分 P4 与 P3。后续研究应让至少两名编码者使用同一手册，对同一批事件盲法编码，并公布分歧案例，而不是只报告一致率。

内部效度与替代解释 P 下降可能由任务变难、时间变短、学习者疲劳、教师评分变化或人工智能界面变化造成。若不控制这些因素，不能把下降归因于反馈回路。后续设计需要设置匹配任务、比较组和人工智能吸收率不同的条件，并保存界面版本。

另一个内部效度威胁是反向因果：也许本来就缺少高层判断的学生更依赖人工智能，因而看起来是工具造成了 P 下降。解决方法是测量入学或研究开始时的基线 P，并进行纵向追踪，而不是只比较使用强度。

外部效度与可迁移性教育类型差异很大。开放性论文、程序设计、数学证明、医学模拟和教师备课的否决层级并不相同。本书只能迁移到同时满足三个条件的场景：任务包含可被人端重新定义的层级；人工智能能吸收相似反馈；产物质量与人端判断可以分开记录。

它不能直接推广到目标已经完全规定、反馈不可被保存或工具只承担机械输入的任务。后续研究应由不同学科团队分别验证，而不是把一个课程的结果外推到整个教育系统。

可靠性、可依赖性与可确认性 P 的可靠性可能受事件切分影响。同一段对话是一个事件还是多个事件，会改变否决层级分布。研究者必须预先规定切分规则，并报告不同切分方案下结果是否一致。

可依赖性要求保留分析轨迹：原始输出、原始反应、编码决定、修订结果和后续任务。可确认性要求研究者声明哪些判断来自材料、哪些只是理论推演，不能用尚未执行的实验语言包装假设。

三项局限与后续设计第一，本书没有完成站外文献检索和经验数据采集。后续可由一个跨学科团队在六个月内完成：一组教育技术研究者核验航空、辅助驾驶和搜索记忆材料，一组教育研究者采集八周人—AI 共作日志，另一组编码者进行盲法层级编码。

第二，本书尚未确定 P 层级在不同学科中的等价性。后续可用两年时间开展跨学科多案例研究，材料包括论文、代码、实验报告和口头答辩，检验 P 层级是否需要学科特定映射。

第三，本书没有验证不可训练断点能否长期抵抗吸收。后续可进行至少一个学期的随机分组研究：控制组正常共作，实验组加入非反馈独立任务，比较 P、T、S 和 A，并追踪断点任务是否被学习者表演化。

本书解释不了的两个洞第一个洞是：为什么有些人在高强度人工智能共作中反而形成更强的目标和因果判断？本书可以用恢复项 r(t) 或高层任务机会解释，却不能说明何种具体教学关系足以使 r(t) 长期为正。

第二个洞是：当人工智能能生成与人端深层判断极其相似的反对意见时，如何区分人的分点保留与工具代行？本书要求无工具迁移和不可见任务，但尚未解决人端可以借助外部材料、同伴讨论和长期内化形成判断的边界。

对第一问的回答人端否决通过”判定—执行—吸收”的递归回路逐层让位。一次事实层或方法层否决被人工智能吸收后，系统在相似任务中减少对人端同层判定的要求；当制度只评价修订后的产物，人端便更多停留在仍然频繁、成本较低的浅层否决。由此，否决次数不必下降，分点却可以下降。漂移不是人端突然停止行动，而是行动的最高有效层级逐渐下移。

对第二问的回答教育、自动化、辅助驾驶和搜索引擎记忆可以共享”外部系统吸收人端任务—人端下一次介入需求下降”的反馈骨架，但它们不是经验形状上的完全同一曲线。航空强调异常接管，辅助驾驶强调注意与反应，搜索研究强调记忆路径，教育则把判断能力的形成本身置于反馈回路内。教育的特殊参数不是简单的”更高自动化”，而是高反馈频率、持续吸收和缺少即时灾难性停止条件的组合。

对第三问的回答可裁决的核心读数是 P，即一次完整共作事件中人端能够独立提出的最高有效否决层级；配套读数包括人工智能吸收率 A、浅层比例 L、无工具迁移率 T 和产物质量 S。建议以每周至少三次事件、连续八周的纵向间隔采集，并每两周加入一次不参与人工智能更新的独立任务。若 P 下降、A 上升、L 上升、T 下降而 S 不降，构成分点漂移的候选证据。若 P 在持续共作中反弹，或独立任务稳定恢复高层判断，本书核心命题必须被削弱。

本书最终并未证明教育必然走向空席。它交出的是一个能被反驳的结构判断：产物越来越好并不能推出人端判断越来越强；如果否决被持续吸收而不再要求人端重新生成同层判断，空席就是人—AI 单位的默认吸引子。防止它的关键不是把人工智能从教育中撤除，而是保护某些不被人工智能改写、也不以产物改善为结算终点的判断事件。

开放的问题仍然是：一个断点要保持不可训练，究竟需要多大程度的制度隔离；而一旦隔离降低了效率、速度和产物质量，教育制度是否愿意把这项损失视为学习的成本，而不是系统的失败。

卷九　席的推广与实施

本卷把席推到教育之外，并写实施，共十二章：十个行业的席、席籍制度、退单考的层分类学、教席的培养、席的产权法草案，卷末七章专写产权——原为一篇独立论文，回答草案里最硬的一问：**席上的否决记录到底归谁？**

它的答案推翻了草案第一条的表述方式。草案说「否决记录归人端」，这篇说：三方主张各占一角却都不完整——学生不能在没有界面、字段、存储与锁定机制的情况下生成完整记录；学校只提供了记录成立的条件，并不因此成为判断内容的作者；平台的存储与格式控制不能转化为对拒绝语义的所有权。否决记录是**产权悬置物**：由人—器具复合行动生成，且在记录、使用与结算被压缩为同一制度事件时形成——它不是先存在、再被使用的对象，而是**被使用才取得完整对象性**。三处现行法（欧盟 AI 法日志条款、FERPA、Copilot 训练数据诉讼）都预设「先有对象、后有使用」，所以它们不是规则缺漏，是**对象的发生节奏超出了既有字段**。

这一判断也修正了本书一处可能被误读的地方：「人与器具不可分割」不自动推出共同所有权。共同生成是发生事实，不是产权结论；要从前者推到后者，还须证明对象可被无损分割为独立贡献——而否决记录的语义与格式恰好阻止这种分割。

由此产权草案的第一动作换了：不是确认归属，是**保全**——十条记录首先是被锁定、待保全而非待分配的对象；制度要做的是在记录完成与首次训练、评价、审计之间重新制造一段可审计的时间差。本卷的草案五条读者仍可照用，但请以卷末七章为准：先保全，后归属。

第一百五十六章　席的十个行业

席不是教育专用的单位。凡是一个人与一个 AI 之间存在否决关系、否决权至少有一层在人端、产物由耦合显出的地方，就有一个席。教育只是悬空最深、锁死最重、最先需要这个词的行业。下面把席画到十个行业，每个行业只写四行：旧地址、席的名字、分点在哪、空席的样子。

医疗——诊席。 旧地址是医生的诊断准确率。诊席的分点在医生对 AI 建议的推翻处：AI 标出的阴影是不是假阳性，AI 没标的地方该不该看。若干医院已经在测采纳率与推翻率，只是没有给它起名。空席的样子是：医生对 AI 建议的推翻率逐年降到零，直到某一次 AI 错了没有人看得出。

司法——审席。 旧地址是法官的判决质量。审席的分点在法官对 AI 量刑建议和类案检索的否决——哪一条类案不该类比、哪一个量刑因素 AI 漏了。空席的样子是判决书越来越整齐、越来越像，直到没有一份判决能回答「为什么这个案子不同」。

学术出版——稿席。 旧地址是作者署名与同行评审。稿席的分点在作者对 AI 稿的退单深度，卷八的样题就是稿席的考法。空席的样子是：署名者答不出审稿人对第三层以上的追问。

驾驶与飞行——驾席。 旧地址是驾照与飞行执照。驾席的分点在人接管的能力——自动系统失效时能退到哪一层。航空业的模拟机复训是驾席读数的原型。空席的样子已经有过：法航 447。

会计与审计——账席。 旧地址是会计师的账目准确率。账席的分点在对 AI 自动分录和自动核对的否决——哪一笔不该那样入账、哪一处异常 AI 没有标。空席的样子是审计意见与 AI 输出完全一致，审计师的签字失去了独立性的含义。

软件——码席。 旧地址是程序员的代码产出。码席的分点在对 AI 生成代码的审查深度：能看出语法错的是第一层，能看出逻辑错的是第二层，能看出架构不该这样的是第三层，能看出这个功能不该做的是第四层。空席的样子是代码库越来越大，没有一个人能解释它为什么这样组织。

翻译——译席。 旧地址是译者的译文质量。译席的分点在对机器译文的改动——改字是第一层，改句是第二层，改语气是第三层，判断这段不该直译而该重写是第四层。空席的样子是译者变成校对员，然后校对员也变成签字员。

研究——研席。 旧地址是研究者的论文数量与引用。研席的分点在对 AI 综述、AI 假设、AI 分析的否决——哪一篇文献 AI 引错了，哪一个假设站不住，哪一个分析的前提是错的。空席的样子是产出上升、判断下沉，实验室里没有人能说出下一个问题该问什么。

管理——管席。 旧地址是经理的业绩指标。管席的分点在对 AI 决策建议的否决——哪一个方案 AI 没有考虑到人的因素，哪一个数据 AI 误读了。空席的样子是组织按 AI 的建议运转得很好，直到环境变了没有人知道该改什么。

立法与政策——策席。旧地址是政策制定者的政策效果。策席的分点在对 AI 政策模拟的否决——哪一个模型假设不成立，哪一个群体模型没有看见。空席的样子是政策越来越优化、越来越难解释、越来越没有人能为它负责。

十个行业的分点都是四层：事实、推理、前提、题。这个四层结构在本卷下一章展开。十个行业的空席也是同一个形状：显露物上升，分点下沉，直到某一次机器错了没有人接得住。

第一百五十七章　席籍制度的设计

席籍是登记席的簿子。它取代学籍，但不是学籍加一栏「使用了 AI」，而是一种不同的登记逻辑。

登记的对象是席，不是人。 一个学生在一个学期里可能坐过几十个席——每门课、每道大题、每个基底型号都是一个席。席籍登记的是这些席的读数，人是席籍的持有者，不是席籍的内容。

登记的内容是三读数。 每个席至少三项：显露物（产物及其旧量表分数）、分点（否决记录及退单深度）、漂移（与上一次同类席的分点之差）。三项缺一，该席标为「悬空」，不计入任何评估。

登记三权归属。 席籍所属的每一种凭证，须在章程里明示定义权、计量权、发证权各在谁；计量权外置以「外部主体可自行命题、换工具、抽样复测并改写达成度判定」四项俱全为准，缺一即计量不独立。这一栏不改变归属，但让「独立评审」不能再冒充审计——卷五二之一的 ABET 读数是它的用法示范。

登记的基底型号与耦合方式。 席籍必须记录 AI 端的型号和人端与它的耦合方式（一次生成、多轮修改、人先写 AI 改、AI 先写人改）。这不是为了限制，而是为了可比：不同耦合方式下的分点不可直接比较。

否决记录归持有者。 席籍上的否决记录是人做的，产权归人端。学校可以读，平台不可以用作训练数据，除非持有者同意。这一条是席籍与学籍最大的不同——学籍上的成绩归学校，席籍上的否决记录归学生。

席籍的保全先于席籍的归属。 席籍不属于学校，但也不因此简单地「属于」学生：否决记录是共同生成的悬置物，学生持有的是判断轨迹与访问权，不是可无损分割的独占权。

席籍可以带走。 席籍不属于学校，学生转学、毕业、就业，席籍跟人走。雇主看的是席籍上的分点与漂移，而不是显露物分数——显露物分数在人—AI 单位上没有区分度。

第一百五十八章　退单考的层分类学

卷八的样题埋了四层：事实、推理、前提、题。这四层不是随手分的，它们对应人端在单位内部能占住的四个位置，每一层都是一种不同的判断能力。

第一层，事实层。 显露物里有没有说错的事实——年代、数字、人名、引文。这一层 AI 端自己已经能做得很好，人端在这一层的退单能力最容易被抽空。只能退到这一层的席，人端只剩校对功能。

第二层，推理层。 事实都对，但从事实到结论的推理断了——以偏概全、相关当因果、比较无对照。这一层要求人端理解论证的结构，AI 端在这一层已经相当强，但仍会犯一些看似流畅的错误。

第三层，前提层。 推理也对，但推理的出发点是错的——一个没有被说出的假设不成立。这一层要求人端看见显露物没有写出来的东西。AI 端在这一层明显弱于前两层，因为前提通常不在文本里。

第四层，题层。 一切都对，但这道题本身问错了——问的方向不对、问的单位不对、问的时候不对。这一层要求人端对题的合法性有判断，也就是对材料的抗性有感觉。这是本书判断 AI 端最难抽走的一层，也是分点的下限。

四层之上还可以有第五层——**局层**：这道题在这个学科、这个时代该不该被问。第五层不适合直接做考题，它是研席和策席的事；但当模型能稳定标记第四层错误之后，考法会上移一格——考学生对模型标记的判断——第五层由此进入学席的考卷，样题见卷十《第四层守不守得住》。

退单考题库的编制原则由此得出：每道题埋四层，每层至少一处该退、一处不该退（防误退）；题库按学科分，但四层结构跨学科通用；同一题在学期首尾各考一次，读漂移；题库对 AI 端公开——这一点看似矛盾，其实是必要的：如果 AI 端见过题库就能答第四层，说明第四层已经不是分点，题库要重编。

第一百五十九章　教席的培养：老师怎样学会退单

教席里老师的份额不在讲授。这句话对老师是残酷的，因为讲授是老师被培养出来的全部。教席的培养要做的，是把老师从讲授者训练成退单者。

第一步，让老师坐一次学席。 给老师一份 AI 写的教案，让他退单——指出哪里该退、退到哪一层。多数老师第一次只能退到第一层。这不是老师的问题，是讲授型培养的问题：老师被训练成产出教案的人，不是否决教案的人。

第二步，把备课改成退单。 教席的备课不是从零写教案，而是让 AI 端先写，老师退单。备课的记录是否决记录，不是教案本身。教案是显露物，否决记录才是教席的读数。

第三步，把课堂改成示范退单。 老师在课堂上做的最有价值的事，不再是讲清楚一个概念——AI 端讲得不比他差——而是当着学生的面否决一段 AI 的讲解，让学生看见退单是怎么做的。学生在学席上要学的正是这个动作，而它只能从教席上看到。

第四步，教席的考核只看退单——但不看退单率。 退单率是最该被禁用的考核员：它便宜、生动、容易做成面板，也最容易骗人。本卷《有效否决四条件》已给出唯一可用的读数：命中且未被 AI 吸收的否决。评的是老师一学期退对了多少、其中多少是 AI 至今吸收不掉的；漂移为负的教席进观察名单——老师也会空席。硬规则一条：**在无法保证审计池的机构里，不得采集退单率。**

第五步，老师的席籍。 老师和学生一样有席籍，跟人走。职称评定看的是教席的分点与漂移，不看论文数量——论文是研席的事，教席的读数是退单。

第一百六十章　席的产权法草案

席上的否决记录是新的资产，产权之争会在教育里最先爆发，因为学生最多、记录最密、平台最想要。这里给一份草案，五条。

第一条，否决记录先保全，后归属。 席上的否决记录不是先存在、再被使用的对象，而是被使用才取得完整对象性的产权悬置物（见本卷末《否决记录的产权悬置》）：学生不能在没有界面、字段与锁定机制的情况下生成完整记录，学校只提供记录成立的条件，平台的存储与格式控制不能转化为对拒绝语义的所有权。因此第一动作不是确认归属，而是锁定：记录进入独立或共同存托，任何一方不得单方改写、删除或重述；本条以下四条规定的是权限，不是所有权。

第二条，显露物归席。 席的产物——作文、解答、论文——归席，可以由席的持有者署名、发表、交易。基底平台对显露物没有任何权利，正如灶台对菜没有权利。

第三条，基底平台不得用席上的否决记录训练模型，除非持有者逐席同意。 否决记录是最有价值的训练数据，也是人端不被抽空的最后一道防线。平台若可以无偿取用，漂移会被系统性地加速。

第四条，席籍随人走。 学校、雇主、平台都只能读席籍，不能持有席籍。持有者可以随时导出、删除、转移。

第五条之前，须有时间差。 在记录完成与首次训练、评价或审计之间设置一段可审计的间隔——记录先入锁定存托，再经独立确认进入用途管线。间隔不必是一夜，但必须足以让对象、理由、版本与访问权限在不被任何一方立即消费的条件下被确认。没有这段时间差，任何归属字段都只是把争议提前写进数据库。

第五条，空席不得转让。 一个漂移为负、分点已降到第一层以下的席，其显露物不得作为持有者的能力证明——这是防止空席被当作资产交易。

这五条会遇到的最大反对来自平台，理由是「训练数据是模型进步的必要条件」。回答是：是的，而这正是为什么它要归人端——进步的必要条件不能由被进步抽空的一方免费提供。

第一百六十一章　否决记录的产权悬置：三条法律路径与它们共同的时间预设

一个字段无法容纳的十次拒绝设一名学生在学席上连续退回人工智能系统生成的十份答案。每次退单都留下时间戳、被退回产物的版本标识、学生身份、退单理由和锁定状态；这十条记录随后被学校用来评估学生的判断训练，被平台用来调整模型，也被系统用来衡量学席本身的运行质量。

在通常的法律语言中，问题似乎可以迅速改写为「这十条记录归谁」。学生可以说，记录源于自己的判断和理由，是自己的劳动；学校可以说，记录产生于教学场景，由学校制度要求并保存，是教育成果或教育记录；平台可以说，记录存在于自己的基础设施中，由自己的系统格式化、存储和用于训练，是训练数据。

但这三个答案并不处在同一层面。学生主张的是创作或劳动上的原始取得；学校主张的是教育过程中的保管、管理和评价权；平台主张的是对数据的占有、处理和利用权。三方并非分别拿着同一性质的权利争夺同一对象，而是把三种不同的法律关系压缩到「所有权」一词中。

更困难之处在于，否决记录并非一项在使用之前已经存在的对象。学生不作出否决，记录不会出现；系统不把该否决写入字段，否决也不会成为可追踪的记录；平台和学校不读取、归档或训练该记录，它的制度价值又不会实现。记录的生成、保存与使用不是三个清楚分开的阶段，而是在同一设备上由人和系统共同完成的复合事件。

因此，真正的异常不是现行法遗漏了一个新型财产类别，而是现行法用于识别财产的时间结构在此处无法成立。既有制度通常能够处理「先有作品，后有使用」「先有记录，后有调阅」「先有数据，后有训练」的对象。否决记录则是在使用动作中才生成，并且生成它的人与使用它的人被制度性地分离。

现有解释失效的具体位置现有解释至少在三处发生错位。

第一，若把否决记录当作学生的劳动成果，就必须说明学生是否能够把系统的字段、平台的写入机制和学校设计的判断流程排除在对象之外。学生提供了判断和理由，但没有独立制造记录得以存在的技术和制度条件。若「劳动」包含这些条件，学生便无法据此取得排他权；若不包含，学生的劳动又无法单独解释记录的完整形态。

第二，若把记录当作学校的教学成果或教育记录，便必须区分学校的管理权限、保存义务和财产权。FERPA 保护的是教育记录的查阅、修正和披露控制，并不直接建立学校对记录的所有权。更重要的是，否决记录并不只是关于学生的影子，它同时记录了 AI 产物被拒绝的状态以及教学制度如何组织判断。把它单纯归入学生记录，会遮蔽其复合来源。

第三，若把记录当作平台训练数据，平台需要证明自己处理的是一项先在的数据，而不是由同一套系统在使用过程中生成的事件。平台可以拥有存储介质、数据库

结构和处理权限，却不能由此推出对学生判断内容的完整权利。特别是「拒绝」这一字段并不是承载于操作之外的中性格式；字段值本身就是一次判断被制度化的结果。

本书占据的缺口本书提出的承重命题是：

> **否决记录不是学生劳动、学校教学成果或平台训练数据中的任何一种既成财产，而是由人—器具复合行动生成、在记录、使用与结算的时间差被压缩时形成的产权悬置物；它必须先被锁定和共同保全，不能被完整归属于任何单一一方。**

这一命题不是把三方权利简单平均，也不是主张任何记录都没有产权。它只针对一种具有特定结构的对象：记录必须由否决动作触发；记录的核心信息必须包括拒绝及其理由；记录必须在生成时进入审计或训练流程；生成者与主要使用者必须不是同一方；记录一旦被改写，其证据功能必须下降。若这些条件不成立，本书的命题便不适用。

人工智能问责路径：把日志当作系统留下的痕迹人工智能治理中的日志思路通常把记录理解为系统运行的可追溯痕迹。其核心问题是：系统在何时运行、接收了什么输入、作出了什么输出、哪些事件可能影响输出，以及发生损害后能否重建因果链。欧盟人工智能法第12

条即体现了这种问责结构。该结构的解释变量是系统行为的可追踪性，制度目标是为监管、审计和责任分配提供事后证据。

这一脉络解释到的是「系统做了什么」。它能够处理自动生成的事件日志，也能够要求高风险系统保存影响输出的重要事件。其强项在于把原本不可见的计算过程转化为可检查的运行记录。

但它在否决记录上从「系统产生了输出」转向「人拒绝了输出」时开始失效。否决理由不是系统运行的自动痕迹，而是学生在系统提供的界面中形成的判断。系统能够记录学生按下退单按钮，却不能仅凭自身日志生成「为什么该答案不符合题意」的规范性内容。若把理由删去，剩下的时间戳和状态码只能说明发生过一次点击，无法说明这次点击为何具有训练和教学价值。

因此，系统日志路径占据了「运行事件的可追踪性」，但没有占据「人对系统产物作出否决时，否决本身如何成为对象」的问题。

教育记录路径：把记录当作与学生相关的制度影子教育法的记录路径不以所有权为中心，而以教育机构维护的记录、学生查阅权、修正权和披露控制为中心。FERPA

的制度结构把教育记录视为与学生直接相关、由教育机构或其代理人维护的记录，并通过访问、修正和披露规则配置不同主体的控制位置。

这一脉络握着的解释变量是「记录与学生之间的关联」。它解释了为什么学校能够保存成绩、考勤、纪律与评价记录，也解释了为什么学生或家长对这些记录享有一定的查阅和修正权。其优势在于承认记录不仅是物理载体，也是教育关系中的权力安排。

但这一脉络在否决记录同时关于学生、AI 产物和教学制度时开始失效。学生确实是记录所涉及的行动者，但不是记录唯一的被指向对象。否决理由指向产物的缺陷，锁定机制指向制度的审计目的，训练用途又指向模型的后续变化。把记录说成「关于学生」，只能说明它与学生有关，不能说明它的完整归属。

FERPA 的修正机制还带来更尖锐的冲突。若学生可修正记录，锁定记录作为审计证据的功能可能被削弱；若学生完全不能修正，记录又不再以「学生记录」的方式承认其主体地位。教育记录路径可以说明谁能够查阅和控制披露，却不能自动回答一项同时记录学生判断、系统缺陷与教学制度状态的对象属于谁。

版权与训练数据路径：把使用建立在先在作品之上围绕 GitHub Copilot

的诉讼把另一条解释路径暴露出来。相关诉讼中，原告围绕开源代码的复制、署名、版权管理信息和许可证条件提出主张，法院的处理涉及哪些请求能够继续、哪些请求缺少足够法律基础。无论具体请求最终如何裁判，这类案件的争论结构都依赖几个先在条件：先有可识别的代码或作品，作者或权利人可以被指认，训练或输出行为发生在作品存在之后。

这一脉络掌握的是「作品—作者—使用」链条。它能够解释训练数据中的版权作品，也能够解释许可证如何把作者的条件传递到后续使用。但它在否决记录上开始失效，因为否决记录不是一件先于使用而等待被读取的作品。它由一次使用性动作生成，而该动作的制度意义正是拒绝承认 AI 产物的合格性。

这里的分离线不是研究对象不同，也不是学科侧重点不同，而是对同一记录的判断不同：版权路径把记录理解为先有数据、后有训练；本书把它理解为训练性使用参与了记录的生成。若一个记录在任何训练或审计之前已经完整存在，版权路径的时间结构可以适用；若记录只有在进入审计和训练管线时才获得完整形态，先在作品的前提便不成立。

三条脉络共同遗漏的字段三条脉络并非彼此互补。系统日志路径处理机器运行，教育记录路径处理机构与学生之间的控制，版权路径处理先在作品的排他权。它们分别占据运行、关系和作品三个位置，却共同默认一个「先—后」结构：先有一个可指认对象，后有管理、调阅、复制或训练。

否决记录恰好把三个「后」压缩进一个生成事件：学生作出判断，系统写入记录，平台开始把记录视作训练资源，学校开始把记录视作评价依据。对象不是先在地等待后续法律动作，而是在这些动作彼此耦合时显露出来。

本书的缺口因此不是增加一类资料，而是改写对象的发生方式：从「谁拥有既存记录」转向「什么条件下记录才成为可归属对象」。这一转向也说明为什么三套法都沉默：不是因为它们没有看到技术，而是因为它们把记录建模为事件之后留下的稳定物，而不是事件本身的一部分。

第一百六十二章　产权悬置物：定义、四件读数与判定函数

理论出发点：发生学而非发现式分类本书站在发生学的理论传统上，把对象理解为在条件、差异和关系的反复作用中形成的暂时稳定结果，而不是等待法律发现的既成实体。

本体论提供了三个分析维度：显露 S、差异路径 D 和纠缠土壤 E。其基本关系可表达为：

这三层不是把对象拆成三个并列部分，而是说明对象的显露、差异路径和关系纠缠相互生成。对象一旦稳定，又会反过来改变差异路径和关系结构。由此，否决记录不能只问「它是什么」，还要问它如何在学生、平台、学校与 AI 产物的相互作用中显露出来。

选择发生学而非单纯分类法，有两个理由。第一，产权分类通常要求对象先存在，再把对象放入劳动、作品、数据或教育记录类别；本书的核心问题正是对象是否已经以这些类别要求的方式存在。第二，否决记录的价值不是静态内容，而是一次差异被锁定后如何进入后续使用。若不追踪生成与使用的关系，就无法判断记录的法律身份。

核心概念的名义定义

否决记录是指由使用者针对人工智能产物作出拒绝动作，并由技术系统保存拒绝对象、拒绝理由和不可撤回状态的事件性记录。

这个定义不包含程度词，也不预设记录属于任何一方。它只确定四项构成：存在被拒绝的 AI 产物；存在使用者的拒绝动作；存在可表达的理由；存在由系统保存并锁定的记录状态。

产权悬置物是指已经具有可识别、可保全和可使用的对象形态，但尚未能够在不损害其生成条件或证据价值的情况下归属于单一权利人的制度对象。

操作性定义与四件读数否决记录在经验层面的取值由事件记录函数表示：

R_i=(o_i,a_i,r_i,τ_i,λ_i,u_i)

其中，o_i 是被拒绝产物的版本标识，a_i 是作出拒绝动作的行动者标识，r_i 是理由字段，τ_i 是事件时间，λ_i 是锁定状态，u_i 是后续使用状态。一次记录被计入否决记录集合，当且仅当 o_i,a_i,r_i,τ_i,λ_i 均有值，且 u_i 表示记录已进入至少一种审计、评价或训练用途。

若以判定函数表示： **V(R_i)：** - 取 1：o_i,a_i,r_i,τ_i,λ_i 均存在且 u_i≥ 1 - 取 0：其他

这里的 u_i 是用途次数，属于等比测量层次；o_i、a_i 和 λ_i 属于类别测量层次；τ_i 属于等比时间读数；理由 r_i 若按是否与对象缺陷直接对应编码，则为类别读数，若按理由序列的复杂度排序，则只能作顺序读数，不可直接当作等距量使用。

在一次可观测事件中，取值方法如下：当学生对某一确定版本的 AI 产物按下拒单操作时，系统即时保存该版本标识；读取当前登录身份作为 a_i；保存学生在按键前完成的理由字段作为 r_i；以服务器接收事件的时间作为 τ_i；在写入完成后生成不可逆的锁定状态 λ_i；当日志被审计、用于评价或进入训练管线时，将 u_i 增加一次。若学生只点击拒绝而没有理由，记录可作为普通操作日志存在，但不计为本书意义上的完整否决记录。

「被使用即生成」的限定「被使用即生成」并不意味着任何使用都会创造新对象，也不意味着记录在平台读取以后才凭空出现。它指的是：在本书对象中，拒绝动作只有在进入系统的审计或训练性使用结构时，才形成具有制度效力的完整记录。学生的主观判断、系统的写入和后续使用共同构成记录的对象性。

这一限定避免把普通点击等同于否决记录。一个按钮事件可以先被记录，后来无人使用；那是日志。否决记录则是一个事件被制度化为可审计、可训练和可评价的对象。其生成不是时间上晚于按键，而是发生在按键、写入与制度用途的耦合中。

第一百六十三章　命题：不是劳动、不是成果、不是数据，而是产权悬置物

承重命题本书的承重命题正式表述为：

> **否决记录不是学生的劳动成果，也不是学校的教学成果或平台的训练数据，而是由人—器具复合行动生成、在生成与使用被压缩为同一制度事件时形成的产权悬置物。**

它不是学生劳动成果，因为学生不能在不依赖界面、字段、存储和锁定机制的情况下生成完整记录；它不是学校教学成果，因为学校的制度设计只提供了记录成立的条件，并未因此成为判断内容的作者；它不是平台训练数据，因为记录的核心信息不是平台独立产生的既存数据，而是学生拒绝和理由被平台使用时才获得制度形态。

命题在以下条件下成立：其一，记录由一次具体拒绝动作触发；其二，拒绝理由是记录的必要组成；其三，记录具有不可撤回或不可重述的锁定状态；其四，记录被学校、平台或其他制度主体用于审计、评价或训练；其五，生成者与主要使用者不是同一方；其六，单方排他控制会降低记录作为证据的可信度。

在以下条件下不成立：若记录只是平台预先生成的普通点击日志；若学生本人既生成记录又是唯一使用者；若记录可以被任意改写而不影响制度功能；若平台能够在没有学生判断的情况下独立生成具有同等信息内容的记录；若某一司法制度已经为这种对象建立了明确、可执行并且不损害证据功能的单一归属规则。

两句复合测试若把下一节要盘点的最近两位占位者分别概括为「日志是系统运行的自动记录」和「教育记录是机构维护的、与学生相关的记录」，两句合并后可以无损重述本书命题吗？

不能。前一句只能说明系统如何留下运行痕迹，后一句只能说明机构如何维护与学生相关的记录；二者都没有说明否决记录为何只有在被制度使用时才取得完整对象性，也没有说明生成者与使用者的分离为何影响产权归属。本书的增量不在于增加「系统」或「学生」两个因素，而在于指出：记录的生成条件本身包含后续使用，因而不能把内容生产、技术存储和制度利用作为三个先后分开的产权来源。

本节按要求交出八位最近邻占位者。由于本次材料未经过站外检索，以下占位均标注为「未核验」；不把记忆中的经典作者或未核验的案件细节伪装成已查证文献。表中所说的「他的说法」是待核对的理论位置，不是对已核验原文的逐字转述。

占位者〔未核验〕	他已经占了什么	分离线	判决性反例
—	—	—	—
欧盟人工智能法日志条款〔未核验〕	「高风险 AI 系统应自动记录运行期间发生的事件、输入和可能影响输出的决定。」	该判断把记录生成归给系统运行；本书判断要求记录的核心理由由人对输出的拒绝生成。	同一系统自动记下输出版本与时间，但学生未填写理由；按其判断是完整日志，按本书不是完整否决记录。
FERPA 教育记录路径〔未核验〕	「由教育机构维护、与学生直接相关的记录受到访问、修正和披露规则保护。」	该判断以记录与学生的关联配置控制权；本书判断认为一条记录同时记录学生判断、AI 缺陷和制度用途。	学生提出的理由指向模型事实错误，学校保存记录供模型训练；按其判断是学生教育记录，按本书是多方复合记录。
GitHub Copilot 训练数据诉讼〔未核验〕	「训练使用是否侵害权利，取决于先在作品、作者、复制或许可条件之间的关系。」	该判断要求作品先于训练而存在；本书判断认为否决记录在使用耦合中才取得完整形态。	没有学生拒绝就没有记录，平台读取动作同时触发写入；按其判断应先有训练数据，按本书记录尚未在使用前完成。
平台数据占有论〔未核验〕	「数据存储在平台基础设施中，并由平台以规定格式处理，因此平台取得对数据的控制。」	该判断把物理占有和格式控制作为归属依据；本书判断认为拒绝字段的格式就是他人判断的制度化。	平台更换数据库字段但不改变学生理由；按其判断格式控制仍支持平台，按本书格式变化可能改变否决意义。
学生劳动论〔未核验〕	「谁作出判断、提供理由并承担后果，谁的劳动就凝结在记录中。」	该判断把判断贡献等同于完整对象的创作；本书判断要求说明界面和制度条件对记录成立的构成作用。	学生口头拒绝但系统未保存；按其判断已有劳动成果，按本书没有完整否决记录。
学校教学成果论〔未核验〕	「教学制度中产生并用于评价的材料属于教学过程的成果或机构记录。」	该判断从场景和用途推导学校归属；本书判断拒绝内容不能因发生于学校场景而被学校单独取得。	学生在校外使用学校账号拒绝平台产物，平台仍训练，按其判断可能是学校成果，按本书须区分生成者与使用者。
不可篡改账本论〔未核验〕	「记录的可靠性来自写入后不能修改，控制写入和保存者因此拥有主要管理权。」	该判断把不可篡改性与控制权连接；本书判断认为不可篡改性恰恰限制任何单方控制。	平台独占锁定密钥并可决定是否开放审计；按其判断平台管理权增强，按本书证据可信度因单方控制下降。
人—器具复合行动论〔未核验〕	「人的行动通过器具完成，器具参与行动的形成。」	该判断可能停留在共同参与；本书判断进一步要求说明共同生成如何排除单一排他产权。	学生提供理由、平台提供字段、学校提供规则；按其判断三方共同参与，按本书任何一方都不能完整归属。

这八行共同显示出一个不可还原性问题。任何一个占位者都能解释否决记录的一个面向，却不能在不改变其自身判断的情况下解释完整对象。日志路径无法解释人提供的理由；教育记录路径无法解释记录同时关于 AI 缺陷与平台训练；训练数据路径无法解释数据的生成依赖于拒绝性使用；学生劳动论无法解释技术和制度条件；学校成果论无法从教育场景推出财产权；平台占有论无法从存储控制推出内容归属；不可篡改账本论无法说明谁有权控制锁定；复合行动论若不进一步否定单一排他权，仍停留在参与事实。

最近两位的严格界线占位表中最接近本书的两条是系统日志路径和教育记录路径。前者说记录来自系统自动运行，后者说记录是由教育机构维护并与学生相关的记录。本书与前者的差异在于记录的核心理由并非系统自动产生；与后者的差异在于「与学生相关」不等于「可由学校或学生单独归属」。

若一个记录只有时间戳、输出版本和按钮状态，本书不把它排除在普通日志之外；若一个教育记录只保存既有成绩，本书也不主张其产权悬置。只有当记录同时具备拒绝理由、锁定状态和训练性用途，且其对象性由这些关系共同生成时，本书命题才出现。

第一百六十四章　判别式、五个场景与四组指标

判别式本书提出的判别式是：

> **若删除生成者的拒绝理由、切断后续使用、或允许单方改写而记录仍被视为同一条完整否决记录，则该对象不是本书意义上的否决记录。**

这条判据不依赖情态词，可以直接询问流程、日志与记录结构。它同时会让上一节各占位者作出不同预测：系统日志路径只要求运行事件仍可追踪；教育记录路径只要求机构继续维护；平台数据路径只要求数据仍可处理；学生劳动路径只要求理由仍可证明劳动；本书则要求三项结构同时保留。

在五个场景中运行该判据。

第一，学生按下拒绝但没有填写理由。对象标识、时间和状态码仍然存在。系统日志路径把它计为完整日志；本书判定为不完整否决记录，因为理由字段缺失。

第二，学生填写理由，系统保存记录，但平台从未读取或用于训练，学校也没有将其纳入评价。教育记录路径仍可能把它视为机构维护的教育记录；本书把它视为尚未完成制度化的拒绝事件，而不是完整产权悬置物。

第三，平台可以在不留下修改痕迹的情况下改写拒绝理由。平台占有论可能仍把它视为可处理数据；本书判定记录失去锁定性，不再具有本书要求的证据对象资格。

第四，学生本人既作出拒绝，又是唯一读取和使用记录的人。劳动论在此具有较强解释力；本书的生成者—使用者分离条件不成立，产权悬置命题不适用。

第五，系统自动生成「AI 自我否决」记录，而没有学生或其他外部使用者参与。日志路径可能将其纳入运行记录；本书将其排除，因为生成者与使用者未形成异方关系，记录不具备本书意义上的复合否决结构。

可观测指标否决记录的主要指标包括四组。

第一组是生成完整度 C，计算为必要字段中实际存在字段的比例：

$$\mathbf{C = (o + a + r + \tau + \lambda) \div 5}$$

每一项取值为 1 或 0。只有 C=1 的事件进入完整记录候选集。这个指标是类别聚合形成的等距比例，不应被解释为学生判断质量。

第二组是使用分离度 A，表示生成者与主要使用者的身份关系： **A：** - 取 1：生成者与主要使用者不相同 - 取 0：生成者与主要使用者相同

它是类别测量，不表示分离程度。若平台、学校和审计者都使用记录，应分别记录使用者身份，而不能把多个使用者压缩为一个「平台」标签。

第三组是锁定完整度 L，由写入后是否存在不可逆版本链、是否保留原始理由、是否记录所有后续访问三项组成。每项为1 或 0：

$$L = (v + r\_h + a\_u) \div 3$$

其中 v 表示版本链存在，r_h 表示原始理由未被覆盖，a_u 表示使用访问可追踪。L=1 才表示记录具有完整锁定状态。

第四组是时间差 Δt，计算记录完成与第一次审计或训练使用之间的时间间隔：

$$\Delta t = t\_use - t\_record$$

它是等比时间读数。本书不预先主张必须隔夜，但认为 Δt＝0 时，记录与使用在制度上同步完成，旧法所需的「先在对象—后续使用」结构无法直接成立。

多少算数，取决于研究设计而非直觉。完整记录至少要求 C=1、L=1 和 A＝1。Δt 不设置普遍阈值，因为不同系统的服务器时间、批处理方式和训练管线不同。经验研究应至少比较 Δt＝0、短于一小时、跨越一个自然日的三组流程，并观察单方控制对审计信任度和记录修改率的影响。

读数的三条自我否决条款第一是信度条款。对同一批已锁定记录进行两次独立重编码，若理由字段的分类一致率低于0.80，或同一批记录在两次系统重放中有超过 5%

的字段值变化，则该读数判定为不稳定。对于 Δt，若服务器与独立存证时间的中位数差异超过预先设定的系统容差，时间差读数不立住。

第二是改名嫌疑条款。若 C、L 或 A 与已有的普通日志完整度、教育记录完整度或训练数据可用性指标高度重合，且在加入本书指标后不能区分五个场景中的至少两个对象，则「否决记录」可能只是改名。尤其需要检验：删除理由字段后，本书指标是否仍能把普通点击日志与完整否决记录区分开来。若不能，核心构念未被测量。

第三是生成替代条款。若平台能够在没有学生拒绝动作、没有人工理由或没有被拒绝版本的情况下，生成一条在独立审计者看来与完整否决记录无法区分的记录，则本书的生成结构不成立。该条款要求对自动合成记录和真实记录进行盲测，不能只比较数据库字段是否相同。

第一百六十五章　证伪条件与竞争解释的检验

本节不把尚未执行的检验写成已经得到的事实。由于本书没有进行站外检索、系统实验或受试者研究，以下检验均标注为「未执行」。

第一组：独立证伪条款

[未执行] **时间分离检验。** 去看不同学席系统中记录写入、第一次读取、第一次训练和第一次评价的时间戳。若大量系统在没有任何人工或技术停顿的情况下，仍能把否决记录作为一个先于使用而完整存在的对象处理，并且独立审计者不需要生成理由即可判断其内容，则本书关于「使用参与生成」的判断为错。现有日志解释能够预测系统记录事件，但不能预测理由字段必须参与对象生成，因此这一结果将削弱而不是支持本书。

[未执行] **单方所有权检验。** 去看真实的教育 AI 部署协议、学校政策和终审裁判中是否存在一项安排：完整否决记录明确归学生、学校或平台单方所有，同时该安排不削弱锁定性、审计可信度和其他主体的必要访问权。若存在这样的安排，且该安排不依赖共同保管或悬置结构，则本书关于单一完整归属不可成立的判断为错。现有三条竞争路径分别能预测部分控制权，但不能预测三方权利同时被单方制度化而没有功能损失。

[未执行] **理由删除检验。** 去看独立审计者在只提供时间戳、对象版本和拒绝状态、不提供学生理由时，能否作出与提供理由时相同的记录分类和质量判断。若两种条件下判断结果相同，说明理由并非否决记录的构成内容，学生劳动论和系统日志论将比本书更能解释对象。

[未执行] **语义—格式切分检验。** 去看平台是否能够独立修改字段名称、数据库编码或存储格式，而不改变学生拒绝的语义；再把学生理由改写而不改变数据库结构，交由同一批审计者判断。若两种改动均不影响记录的身份、可信度和用途，说明语义与格式可以分离，本书关于复合结构不可切分的判断为错。

[未执行] **自我否决检验。** 去看平台是否能够让模型自动拒绝自己的产物，并使独立审计者把该记录与学生拒绝视为同一种否决记录。若二者在身份、证据功能和训练用途上不可区分，本书关于生成者与使用者必须异方的判断为错。现有解释可以把二者都纳入日志或训练数据，但无法说明为什么人类拒绝与模型自我拒绝应有不同的产权结构。

[未执行] **锁定必要性检验。** 去看允许学生在限定期限内撤回或修正理由的系统，与完全锁定系统相比，独立审计对 AI 产物缺陷的判断错误率、争议率和重现率是否相同。若可撤回系统在这些指标上不低于锁定系统，本书关于锁定是记录价值优先条件的判断为错。本书不能把锁定性当作不可检验的道德要求。

第二组：不利材料与边界本次没有执行站外检索，因此无法报告上述材料在不同司法辖区的命中范围，也不能声称现行立法或判例已经普遍沉默于本书所定义的否决记录。特别是，欧盟人工智能法条文的正式适用范围、FERPA

对不同记录类型的具体解释、GitHub Copilot 相关诉讼的程序阶段和最终裁判效果，均需要逐项核对原始法律文本和判决书后才能完成经验性法律论证。

这项不利状态并不自动推翻本书，但它限制结论的强度。本书目前交出的不是「现行法已经判定三方均无权」的事实结论，而是一个结构性判断：在未新增专门字段前，三套法律材料各自的解释变量不能无损覆盖否决记录。

系统日志解释系统日志解释的最强版本不是「平台拥有一切」，而是：高风险 AI

系统需要自动留下能够重建运行过程的事件痕迹；学生拒绝被视为影响系统输出的重要事件，因此平台有义务或有权限记录它。

这一解释在时间戳、对象版本和按钮状态方面具有优势。它可以准确回答学生何时拒绝了哪一版产物，也可以说明记录是否进入系统的审计链。但当具体条件是「学生理由决定记录能否用于判断产物缺陷」时，系统日志解释预测的是 A：只要自动事件被保存，完整记录就成立；本书预测的是 B：没有理由字段，只有普通日志，不是完整否决记录。A 的读数是事件字段存在，B 的读数是 C=1 且理由字段可追踪。

系统日志解释在这一检验中获得部分支持：对于运行问责，自动时间戳确实不可替代。本书因此不能把系统日志整体判为错误。本书只能证明，日志结构不足以独立推出否决记录的产权和完整对象性。它解释「发生了点击」，不解释「为什么该点击成为一项可训练的判断」。

FERPA 教育记录解释 FERPA

解释的最强版本是：学校维护的、与学生直接相关的记录属于教育记录制度；学生拥有访问和要求修正不准确内容的权利，学校则承担管理和披露控制义务。按这一解释，十条否决记录产生于教学过程、由学校保存并用于学生评价，因此应首先作为教育记录处理。

当具体条件是「学生理由同时用于识别 AI 产物缺陷和训练下一代模型」时，FERPA 路径预测的是 A：记录仍然是与学生相关的教育记录，其核心控制问题围绕学生访问和修正权展开；本书预测的是 B：该记录是多方复合记录，学生关联不能推出单一归属，且不受限制的修正权可能破坏锁定证据。A 的读数是学生是否能够访问和请求修改，B 的读数是修改行为是否改变原始理由、版本链和审计结果。

这一检验对本书并非全然有利。FERPA 的控制权结构确实能提供比产权语言更细的制度工具，尤其是访问、披露和修正机制。本书不能否认学校需要履行保存和保护义务，也不能把「非单一财产」误写成「学生没有任何权利」。本书的增量在于：

教育记录身份与产权身份必须分开，学校的保管义务不能自动变成对判断内容的所有权。

训练数据解释训练数据解释的最强版本是：学生理由和拒绝状态被平台存储为可处理的数据，平台通过格式、管线和模型更新赋予数据训练用途，因此平台至少拥有使用和控制数据的权利。GitHub

Copilot 诉讼中围绕既有代码、复制、许可证和模型输出的争论，显示训练使用可以成为法律分析的独立对象。

当具体条件是「平台读取记录的动作同时触发记录进入训练池，并且记录在读取前没有完整制度形态」时，训练数据解释预测的是 A：先有可处理数据，后有平台训练；本书预测的是 B：训练性使用参与了记录的生成。A 的读数是训练管线读取前已经存在完整数据对象，B 的读数是删除训练或审计用途后，理由和锁定状态是否仍构成同一对象。

若数据在训练前已经完整存在，A 将得到支持，本书命题需要收窄为普通数据治理问题。若只有平台使用后记录才获得不可替代的训练和审计意义，B 得到支持。由于本书没有执行系统时间戳核验，不能声称已经完成裁决。能够作出的结论是：平台的存储和处理事实不足以证明其对否决记录享有完整产权。

学生劳动解释学生劳动解释的最强版本是：学生作出判断、提供理由并承担拒绝错误的后果，记录中的创造性或认知劳动因此来自学生。学生可能进一步主张，平台和学校只是提供工具和制度场景，不应夺取劳动成果。

当具体条件是「学生在系统未写入任何记录的情况下，只在个人纸笔中完成理由」时，学生劳动解释预测的是 A：劳动成果已经存在；本书预测的是 B：个人判断可能存在，但本书意义上的完整否决记录不存在。A 的读数是理由作为独立表达已经形成，B 的读数是 o,a,r,τ,λ 是否共同写入并进入制度使用。

这一结果并不否定学生的劳动利益。它只区分判断劳动和否决记录。学生可以拥有一份自己的思考痕迹，也可以对理由享有隐私、访问和证明利益；但劳动发生并不自动产生对复合记录的排他所有权。学生劳动论若扩大为「凡由我判断所产生的记录皆归我」，便无法解释平台字段和学校制度对记录成立的构成作用。

学校教学成果解释学校教学成果解释的最强版本是：学校设计学席、规定退单流程、保存记录并用记录评价学生，因此记录是教学活动的产物。学校可能不使用「所有权」一词，而主张管理、归档、评价和再利用的综合权力。

当具体条件是「记录中的理由直接导致平台模型改变，但学校没有提供模型训练基础设施」时，学校解释预测的是 A：教学制度仍然使记录成为学校成果；本书预测的是 B：学校可保有制度性保管和评价权限，但不能据此取得完整内容权利。A 的读

数是记录是否发生在学校设计的流程中，B 的读数是学校是否能够单独授权、修改或阻止平台使用而不损害记录结构。

学校成果论能够解释为什么学校对记录负有管理责任，但不能从场景直接推出财产权。若学校承认 AI 是教学单位的一部分，教学成果便包含学校无法单独控制的器具参与；若学校否认 AI 的参与，把 AI 仅视作工具，判断内容又更接近学生劳动。学校必须在「复合教学」与「工具教学」之间选择，而两种选择都不能稳定推出学校独占。

平台格式和占有解释平台解释的最强版本是：平台规定字段、运行数据库、承担存储安全和训练成本，因而至少取得对格式和处理过程的控制。它不必声称平台创造了学生理由，只需主张记录的可用形态由平台完成。

当具体条件是「字段值 rejected 同时表达数据库状态和学生判断结果」时，平台解释预测的是 A：平台可以拥有格式而不触及内容；本书预测的是 B：在否决记录中，格式字段的值就是判断制度化的语义，格式与内容不能安全切开。A 的读数是更换字段名称或编码后记录意义不变，B 的读数是格式变化是否改变独立审计者对拒绝对象和理由的理解。

平台确实可以对数据库架构、服务器安全和处理流程承担权利义务，但这些权利不能等同于内容产权。若平台以「格式权」控制学生无法访问、核验和保全原始理由，格式控制便改变了记录证据价值。平台越强调格式是纯技术层，越需要证明技术层的改变不改变记录的语义身份。

不可篡改账本解释不可篡改账本解释认为，记录的核心价值来自锁定和版本不可逆，因此掌握写入与保存权限的主体应当拥有最强管理权。该解释抓住了否决记录区别于普通点击流的关键：若记录可以被覆盖，审计就无法重建学生当时拒绝了什么。

当具体条件是「平台独占锁定密钥，学校和学生只能申请访问」时，该解释预测的是 A：单方控制有助于保证不可篡改；本书预测的是 B：单方控制会使证据价值依赖平台的自我证明，降低第三方信任。A 的读数是版本链是否保持完整，B 的读数是独立审计者在平台单方声明所有权后是否仍能验证记录来源和访问历史。

锁定解释在记录完整性上得到支持。本书不是要取消技术保全，而是要切断「保全控制因此等于所有权」的推理。锁定性是记录作为证据的条件，却不能决定谁拥有排他支配权。正因为锁定重要，写入权、解锁权和解释权更需要分开。

人—器具复合行动解释人—器具复合行动解释指出，学生的判断不可能脱离界面、字段和系统而发生，平台和学校的技术—制度条件因此参与了记录的生成。它比简单劳动论更接近本书，因为它承认行动不是人单独制造的。

但若该解释只停留在「共同参与」，仍不能解决归属问题。当具体条件是学生提供理由、平台提供写入、学校提供规则且三者缺一记录就不存在时，复合行动解释预测的是 A：三方共同产生记录；本书预测的是 B：共同生成并不自动形成三方共同所有权，而是使单一排他归属缺乏对象基础。A 的读数是参与者数量，B 的读数是任一方能否单独修改、授权或排除其他必要主体而不改变记录身份。

本书的判断不是简单地把三方都写成共同作者，而是认为「共同生成」改变了产权问题本身。若三方共同生成却仍有一个主体可以无损取得完整排他权，必须说明其他参与者为何不再构成对象的一部分。对否决记录而言，这一步目前没有被现有材料完成。

第一百六十六章　辨别格：制度吸收 × 判断承担

候选轴的排除为了把「产权悬置物」从新名字变成辨别方式，需要建立两个不以成因为本身的轴。以下两个候选轴被排除。

第一，排除「生成来源」作为第二轴。它可以区分学生、平台和学校的贡献，但成因正是本书需要解释的对象，拿生成来源与生成来源比较，会把结论预先放入分类轴。

第二，排除「法律权利强度」作为第二轴。它把所有权、控制权和访问权混在同一维度上，无法区分一个记录为何具有证据价值，也无法说明单方权利为何可能损害记录。

本书确立的第一轴是**记录是否被制度性吸收**。低值表示记录仍停留在事件或个人私账中，尚未进入学校评价、平台训练或正式审计；高值表示记录已成为其他制度行动的输入。

第二轴是**关键判断是否仍由生成者可追踪地承担**。低值表示理由、版本与判断者之间的链条被平台或制度流程替代；高值表示原始理由、版本、时间与行动者仍能被独立核验。该轴与第一轴相关但不重合：记录可以高度制度化而仍保留判断链，也可以尚未高度制度化却已经丢失原始理由。

两轴同时为高的真实案例需要具名材料。当前没有经过站外检索，也没有实际部署案例可核验，因此不能声称存在满足该条件的具名真实案例。本书只保留「相关但不重合」的表述，并把两轴同时为高的格子作为待研究类型，而不是伪造案例填补。

二维辨别格

低制度吸收／低判断承担：个人残留事件

两个坐标值分别是：制度吸收低，判断承担低。这里的记录可能存在于界面缓存、个人截图或未提交的草稿中，尚未进入学校或平台的正式使用流程；同时，原始理由没有形成连续、可验证的记录链。格内对象不是完整否决记录，而是个人反应的残留。

只在该格成立的预测是：删除平台账户后，个人仍可能保留对事件的主观记忆，但独立审计者无法从系统材料重建被拒绝的版本和理由。

低制度吸收／高判断承担：自持判断档案

两个坐标值分别是：制度吸收低，判断承担高。学生在按键前后保留版本、理由、时间和后续处理，系统尚未把记录用于评价或训练。此时最强的对象是学生自持的判断档案，而不是平台训练数据或学校教学成果。

只在该格成立的预测是：学生可以在不依赖平台继续开放的情况下，向第三方展示拒绝理由与后续验证路径，但这组材料尚未自动产生学校或平台的制度使用权。

高制度吸收／低判断承担：被制度利用的黑箱拒绝

两个坐标值分别是：制度吸收高，判断承担低。记录已经进入模型训练、学校评价或审计池，但原始理由被压缩为状态码，版本链或行动者链条无法核验。此时记录表面上具有高制度价值，实际上缺少可追责的判断来源。

只在该格成立的预测是：平台或学校能够大量利用记录，却无法在争议发生时说明某一条拒绝究竟由何种理由触发，因而制度吸收越高，归责争议越集中于记录解释权。

高制度吸收／高判断承担：共同保全的生成性证据

两个坐标值分别是：制度吸收高，判断承担高。记录进入学校审计和平台训练，同时保留被拒版本、学生理由、写入时间、锁定状态和全部访问链。它是本书意义上最典型的产权悬置物：制度需要它，学生构成它，平台使用它，学校保管它，但任何一方单独修改都将改变对象身份。

只在该格成立的预测是：任何单方排他控制都会同时降低至少一种功能——学生失去判断可证明性，学校失去审计可信度，平台失去训练数据的独立性；共同存托比单方归属更能保持记录的制度价值。

四格之间不能互换。第一格缺少制度使用和判断链；第二格有判断链但没有制度吸收；第三格有制度吸收却没有判断链；第四格同时具备二者，因而产生产权悬置问题。尤其是第三格与第四格都可能表现为「高价值数据」，但第三格的价值依赖制度利用，第四格的价值还依赖行动者和理由链条。把二者一概称为训练数据，会丢失最关键的差异。

第一百六十七章　先保全，后归属：实践、边界与反噬

理论意涵本书的理论意涵首先在于，它把产权问题从对象分类推进到对象生成。学生劳动、学校成果和平台数据之争，表面上是三种权利主张，深层则是三种关于对象何时成立的假设。学生劳动论假设判断完成即产生可归属成果；学校成果论假设教学场景能够把制度中的产物归入机构；平台数据论假设写入和处理使数据成为平台可支配对象。否决记录显示，三种假设都可能只覆盖对象的一部分。

其次，本书证明「人与器具不可分割」并不简单导出共同所有权。若学生与系统共同完成判断，平台与学校也提供不可替代条件，那么「共同参与」是生成事实，不是产权结论。共同生成可能导致更强的共享权利，也可能导致单一产权无法成立；要从前者推到后者，必须另行证明对象可以被分割为独立贡献。否决记录的语义与格式恰好阻止了这种无损分割。

再次，本书把时间差引入产权分析。既有法律结构需要对象与使用之间存在可识别的间隔，以便法律在对象形成后、后续利用前介入。否决记录的同步化使这段间隔消失：判断、写入、读取和结算由同一装置连续完成。法律沉默因此不是简单的规范缺漏，而是对象的发生节奏超出了既有字段。

实践意涵实践上，第一优先不是争夺「谁拥有十条记录」，而是确保记录不会被任何一方单独修改、删除或重述。学校应当把保存和评价权限与产权语言分开；平台应当把存储、训练和内容解释分开；学生应当把个人判断轨迹与平台字段分开保存。

学生最先需要留下的痕迹不是一句「这是我的劳动」，而是判断前后的连续路径：被拒绝产物的版本、当时的理由、理由与题干或证据的对应关系、拒绝后采取的行动，以及系统是否改变了记录。平台的状态码只记录「拒绝发生」，学生自己的连续记录才保存「谁在何种路径中作出拒绝」。

制度设计上，应在记录完成和首次训练、评价或审计之间设置可审计的时间差。记录先进入锁定存托状态，再在独立确认后进入用途管线。时间差不必固定为一夜，但必须足以让记录在不立即被某一方消费的情况下存在，并让各方确认对象、理由、版本和访问权限。

适用边界与反向约束第一，若 AI

系统能够自动对自己的输出作出与学生拒绝不可区分的否定，本书关于生成者与使用者必然异方的判断就必须削弱。本书原本想说「异方性是构成要件」，此处必须改成「对于人类否决记录，异方性是制度有效性的必要条件」。模型自我否决可以成为另一类对象，不能偷偷并入同一类型。

第二，若学生能够在同一系统中生成记录、唯一使用记录并独立承担后续判断，本书原本想说「生成者和使用者必然分离」也不成立。此时对象可能回到个人劳动或个人档案类型。本书不能把所有人机互动都归入产权悬置。

第三，若法律明确规定某类否决记录由学校单独保管和控制，且该制度同时保护学生访问、平台训练透明度和记录不可篡改性，本书原本关于「不能单一归属」的命题必须缩小为「现有三处材料本身不能推出单一归属」。本书不应把结构性分析冒充为对未来立法的否决。

反噬若照本书执行，最省事的做法可能恰好毁掉本书想保护的东西。学校可以把所有记录交给独立存托人，平台停止直接读取，学生只需在固定界面填写理由；这样虽然降低了单方控制，却可能把学生判断重新压缩成标准化字段，使理由变成新的模板，丢失原本的生成性。

同样，强制隔夜可能保护时间差，却降低系统对明显错误的即时修正能力。若平台必须等候确认才能调整模型，学生会继续面对已经被识别的问题；若学校为了保全证据禁止任何纠错，记录又可能把暂时错误固定成制度事实。因此，时间差必须保护原始记录而不是禁止所有后续修正；修正应当以新版本附加于旧版本，不能覆盖原始事件。

构念效度与可信性最大的构念效度威胁是「产权悬置物」可能只是把普通数据治理问题换了一个名称。若理由字段并不影响记录身份，或记录在被使用前已经完整存在，本书的核心构念就没有测到对象发生方式。应通过理由删除检验、版本链检验和独立审计盲测确认，本书指标是否真的区分普通日志与完整否决记录。

质性研究对应的可信性威胁在于，三处材料可能被本书过度统一解释。欧盟日志条款、FERPA 和 GitHub Copilot 诉讼的法律目的并不相同。后续研究需要由法律文本研究者分别核对正式条文、实施规则和完整判决理由，避免把「沉默」解释成法律否定。

内部效度与替代解释内部效度的主要替代解释是：三方之所以无法归属，不是因为同步生成，而是因为现行法对数据、教育记录和平台合同的权利配置本来就不完整。另一种解释是，法律可以通过合同、数据保护、保密义务和证据规则解决问题，不需要新的本体分类。

本书尚未排除这些解释。后续研究应选取至少三个已部署的教育 AI 系统，比较记录写入、使用和争议处理的时间顺序，观察在引入时间差后，单方归属方案是否仍然能够维持记录可信度。研究者可以使用系统日志、学校政策、平台协议和匿名化争议记录，在两年内完成初步比较。

外部效度与可迁移性本书不能直接推广到普通点击流、考试答案、教师评语、模型自我评分和匿名统计数据。它只适用于具备拒绝对象、拒绝理由、锁定状态和后续制度用途的记录。不同司法辖区对教育记录、数据控制和训练使用的法律定义也可能导致不同结果。

可迁移性研究应由跨法域团队在欧盟、美国和至少一个具有不同教育数据制度的司法辖区各选取一个学校平台，使用同一编码表比较 C、L、A 和 Δt。若三地对象的结构相同而法律结论不同，本书将更准确地解释为本体分析，而不是法律结论。

可靠性、可依赖性与可确认性可靠性威胁包括时间戳不一致、理由字段由系统自动改写、平台访问日志不完整以及不同编码者对「完整理由」的理解不同。后续研究应建立独立存证时间、双人编码和版本哈希，并公布编码规则。若换一名研究者后，记录分类明显改变，本书的指标就不能支撑产权判断。

可依赖性要求研究过程能够被复现；可确认性要求结论不只是研究者对材料的偏好。本书通过公式、阈值和证伪条款降低这些风险，但没有真实数据验证，因此不能把方法设计说成可靠性已被证明。

两个本书解释不了的洞第一个洞是集体判断。若十名学生共同讨论后由一名学生按下退单，理由来自群体讨论而非单一行动者，本书无法说明生成者—使用者分离如何在集体主体中保持。把十名学生平均分配贡献会遮蔽讨论过程，把按键者视为唯一生成者又可能失真。

第二个洞是平台预训练。若平台在学生拒绝前已经通过其他来源形成「拒绝模式」，学生记录只是触发参数更新的最后一个事件，本书无法确定记录的训练价值来自哪一部分。它可能仍是学生判断的事件证据，也可能只是大量既有数据中的一个样本。该洞要求后续模型审计，而非概念分析能够独立填补。

对 第一问 的回答欧盟人工智能法日志条款预设记录是系统运行自动留下的痕迹，能够说明系统做了什么；FERPA

预设记录是由机构维护、与学生相关的制度影子，能够说明谁可以访问、修正和控制披露；GitHub Copilot 训练数据诉讼预设先有可识别作品、作者和可分离的后续使用，能够说明既有作品如何进入训练争议。

三者都不能无损说明否决记录。日志条款缺少学生理由的生成位置；教育记录路径不能把学生关联转换成完整归属；训练数据路径依赖先在作品和使用—创作分离，而否决记录在制度使用中才取得完整对象性。

对 第二问 的回答在「人与器具不可分割」的前提下，学生不能把平台界面、字段和锁定机制从自己的行动中完全切除，因此「这是我的劳动」只能保留为部分贡献或个人判断轨迹，不能推出完整独占。平台不能把存储和格式控制转化成对拒绝语义的完整所有权，因为拒绝字段的格式就是判断被制度化的方式。学校不能从教学场景和保存行为推出完整成果权：若承认 AI 是教学复合体的一部分，学校不能独占复合成果；若否认 AI

参与，判断内容又回到学生和系统的关系中。

因此，三方都可以拥有不同的访问、保全、评价或使用权限，但任何一方都不能不改变对象性质而取得完整排他权。

对 第三问 的回答可裁决判据是：删除拒绝理由、切断制度使用或允许单方改写后，若记录仍被视为同一条完整否决记录，则它不是本书意义上的否决记录。实践指标包括字段完整度、身份分离度、锁定完整度和记录—使用时间差。最重要的制度动作是在记录完成与首次训练、评价或审计之间重新制造可审计的间隔。

本书交出的不是实验裁决，也不是任何司法辖区已经接受的产权规则。它交出的是一个对象判据、五条暂定产权规则及其证伪条件。十条记录首先应被当作锁定的生成性证据，由独立或共同存托结构保全；学生则应在系统结算之前保存自己的版本、理由与判断路径。

真正尚未解决的问题不是「谁最终拥有记录」，而是法律能否在一个同步化装置中重新制造足以让对象显露、被核验并接受裁决的时间差。若这一时间差无法建立，任何归属字段都可能只是把争议提前写入数据库，而不是解决争议。

卷十　反驳、证伪面与未解

第一百六十八章　全书划界总表

卷五至卷七各趟各有自己的最近邻表，前后共列出四十余位。这里把他们按「占住了什么」归成六组，各给一条全书统一的分离线。这张表不重复各趟的判决性反例，只回答一个问题：本书站在他们之外的哪一个位置。

第一组：认知在哪里发生。 Hutchins（1995）、Clark 与 Chalmers（1998）、Latour（2005）、海德格尔（1927）。他们证明认知与行动分布在人、工具、符号与制度之间，没有先于世界的主体。本书全盘接受，并在卷三逐条对照。分离线：他们回答「认知在哪里发生」，本书回答「发生之后记在谁名下、评在谁头上」——登记不跟随发生，这是他们没有的变量。

第二组：指标被腐蚀。 Goodhart（1975）、Campbell（1976）、Koretz（2008）、Strathern（1997）。他们证明一个量一旦用于分配就失真。分离线有两条：其一，他们说量被腐蚀而单位不变，本书说单位本身在入账时被切换（卷五第四趟）；其二，他们预测失真的指标会被弃用，本书预测它更牢地留在结算位上（空值结算，已由 Koretz 自己的二十年数据反证）。

第三组：资格如何被垄断与兑换。 Spence（1973）、Collins（1979）、Labaree（1997）、Abbott（1988）。他们解释文凭为何有分配功能、职业如何争管辖权。分离线：他们把能力当作先于信号的私人属性，本书说被信号的对象本身已不是一个对象；他们解释资格如何垄断，不解释证书为何每张只写一个名字——那个「一张一名」的几何才是本书的题。

第四组：教育评估的修复方案。 Bearman、Dawson、Swiecki、Ellis、Perkins、Cotton、Bozkurt、Bretag、Selwyn 及 AI 使用分级量表。他们分别主张过程性评价、贡献声明、深度考题、分布式作品集。分离线：四条路线都在解测量题，卡住教育的是登记题；测量任意精细，容器单数照样切名。本书不否认它们的价值，否认的是它们对「单位危机」的承诺。

第五组：责任落在哪。 Elish（2019）道德缓冲区、Wagner（2019）准自动化、Green（2022）人类监督批判、Santoni de Sio 与 van den Hoven（2018）、Alfrink 等（2022）、Bainbridge（1983）、Koppel 与 Kreda（2009）。他们证明自动化把责任推到末端的人身上。分离线：他们描述落点的位移，本书说落点由翻译规则决定、由首个不可回补点定界（卷七），并把「责任仍在、资格无载体」与「责任无处落」分开——前者是悬空资格，后者才是他们的责任鸿沟。

第六组：制度为何在失效后继续、为何查不出。 Meyer 与 Rowan（1977）、DiMaggio 与 Powell（1983）、Akerlof（1970），以及审计独立性理论。前三家解释合法性仪式与信息不对称；审计理论要求鉴证者不参与主张的形成，却默认主张的对象已被外部定义。分离线：他们解释旧制度有惯性，不区分「单位仍可测而制度不愿更新」与「单位本身已不可约简」；本书把测量功能与分配功能拆开，说明测量对象死亡后分配功能如何借可辩护性缝合（卷五第三趟）；并把审计独立拆成程序独立

与定义独立两层——教育三权同在一校，程序独立可移植、定义独立不可，所以失真人人知道而无人能查（卷五二之一）。

六组合起来，本书的位置只有一句：**他们都已踩到这个洞的边缘——认知已分布、指标已失真、资格已垄断、责任已位移、制度已仪式化——没有一位把脚落进去问：那么评估该落在哪个单位上**。本书的答案是席。

第一百六十九章　十条反驳与回应

反驳一：这只是「评估学生的批判性思维」的新说法。 不是。批判性思维仍把学生当评估单位，只是换了一项能力来评；席把单位换了。批判性思维可以在闭卷下考，退单深度不能——它要求 AI 端在场并出错。

反驳二：分点会被 AI 端抽走，最后连第四层也守不住，席就没有意义。 这是本卷第三条证伪面，本书承认它可能发生；本卷《第四层守不守得住》一章专门回答它——第四层能力上可被代答，位置上不能，分点由位置定。但即便如此，席仍比「人」更准——它至少能测出分点降到了哪里，而「人」这个单位连这个都测不出。席失效的那一天，是所有评估单位失效的那一天，不是回到评人的那一天。

反驳三：学生用 AI 是作弊，禁掉就没有悬空。 禁令冻结不了工具端。泳衣可以规定，基底不能；校内禁、校外用，悬空更深。而且禁令评的仍是旧单位，旧单位在真实学习里已经不运转——禁令成功的结果不是回到过去，而是教育变成体育：在一个规定了装备档的房间里考一个在房间外不存在的能力。

反驳四：席的三读数太复杂，制度做不到。 三读数中的显露物就是现在的分数，不增加成本；分点要一次退单考，成本相当于一次期末考；漂移是两次分点之差，不需要额外考试。复杂的不是制度，是把地址从人挪到席的那一刀。而且有两件事三权不分也能先做：确认门槛（入账前四处确认）与资格资本充足率 K_q（发行量与复测补救能力匹配），见卷五二之一。

反驳五：这是把学生当零件，取消了人的主体性。 恰恰相反。「人」这个旧单位在 AI 之后正在被抽空而不自知——显露物越来越好，人端越来越空。席是唯一能把这种抽空测出来并制止的单位。取消主体性的不是席，是只看显露物的旧评估。

反驳六：海德格尔不会同意用「席」这样的制度性单位。 他大概不会。但他的分析停在存在论，没有给地址簿；悬空判断的焦虑发生在制度里，不给地址，焦虑不落地。本书从他那里拿的是本体论根据，不是他的态度。

反驳七：历史上的悬空都自行着陆了，这一次也会。 历史上的着陆靠残余功能，AI 不留残余，这是本书的核心判断。如果反驳者能指出一块 AI 稳定不会做的功能，本书的「退守无地」就错了，欢迎指出。

反驳八：退单考可以被 AI 代答。 前三层可以，第四层目前不能——这正是第四层被定为分点下限的原因。如果某一天第四层在能力上被代答，题库要重编，考法上移一格——考「对模型标记的判断」，第五层由此出现；分点在位置上不动。这是一个动态的考法与一个静止的位置，见《第四层守不守得住》。

反驳九：席籍侵犯隐私。 席籍记录的是否决记录，归持有者，随人走，平台不得取用。它比现在的学籍更保护隐私——现在的学籍归学校，学生连导出的权利都没有。

反驳十：这一切在中国的考试制度下做不到。 悬空判断在锁死最重的制度里停留最久，这是本书卷四的判断，中国的考试制度正是这样的制度。做不到不等于不悬

空；悬空会一直积累，直到制度松动。本书做的不是预测松动的时间，而是在松动到来之前把地址簿写好。

第一百七十章　第四层守不守得住：分点的下限是责任边界，不是能力边界

一、全书押在哪一层上

席的分点按层计：事实层、推理层、前提层、题层。前三层退的都是文本里的东西——说错的年代、断掉的推理、没写出来的假设；第四层退的是文本外的东西——这道题与它所对的那个世界之间的关系：该不该这么问，问的单位对不对，问的时候对不对。本书断言第四层是 AI 端抽不走的最后一层，也是分点的下限。这是全书最不能排除的证伪面：前三层 AI 已经或将要做得比人好，如果第四层也被代答，席就没有着陆面，本书的方案与以往所有工具的方案一样进入保留地。

所以这一章不能再用一段话带过。它要回答三个问题：第四层到底需要什么；AI 今天缺的是哪一样；如果某天不缺了，分点往哪里去。

先把结论放在前面，免得读者在材料里迷路：**第四层守得住，不是因为 AI 不够聪明，而是因为判断一道题该不该问的资格来自要为答案负责。没有后果位置的判断者，无论多准，都不在第四层上。分点的下限不是能力边界，是责任边界。** 这句话意味着，即使模型明天就能在每一道错题上准确标出「题问错了」，分点也不会因此下移一格——除非制度同时把后果位置给了它。而后果位置能不能给一个不能被吊销、不能被退学、不能被判刑的东西，是一个与模型能力无关的问题。

二、第四层与前三层的本体差异

前三层有一个共同点：判断的对象与判断的依据都在题面之内或题面可及的知识库之内。事实错不错，查得到；推理通不通，验得出；前提立不立，可以从学科的已有共识里对照。这三层是「闭题判断」——题已经给定，判断在题内进行。

第四层是「开题判断」：它要判的是题本身能不能立。这一判断的依据不在题内，也不完全在知识库内，它在两处：一是材料的抗性——这道题所对的那个对象，允不允许被这样问；二是问者的位置——谁在问、问了要干什么、答错了谁承担。卷三讲过，材料是单位里不作声但有立场的成员，它用「不成」来否决；第四层退单就是人端替材料把这个「不成」说出来。

举一个第四层的例子。样题里那段科举文字的第四层错误是：从一桩史例跳到「考试制度是社会公平的根本保障」这一普遍规范。为什么这一层比前三层难？因为前三层都有一个「对」可以对照，第四层没有——「该不该从史例跳到规范」不是一个有标准答案的问题，它取决于这道题要用来做什么：写进教科书，不该跳；写进政策辩论稿，跳得成不成要看谁为这个政策负责。第四层的判断者必须知道答案会落到哪里，才能判题问得对不对。

这就是第四层的本体特征：**它的判断依据包含了判断者自己的位置。** 一个不在任何后果位置上的判断者，可以对前三层给出与在位者完全相同的判断，因为那三层的

依据在题内；但它对第四层给出的判断，缺了一个变量——它不知道、也不必知道答案会落在谁身上。

三、三处材料

本书在三处已有制度里找第四层的形状，看它们各自把「题该不该这样问」的判断交给了谁。

第一处：基准测试里的错题。 近两年最难的几套模型能力基准——数学前沿题集与「人类最后考试」一类的跨学科题集——在发布之后陆续被发现含有错题：题干自相矛盾、标准答案错、题目在专家看来根本不成立。被发现的方式几乎全是人的复核：出题专家事后自查、社区研究者逐题重做。这里有一个本书要抓的结构：这些基准的评分机制只奖励「答对」，不奖励「指出题错了」。一个在错题上给出与错误标准答案一致的模型得分，一个指出「此题不成立」的模型不得分。于是模型被训练成在任何题上都答，几乎从不退题。这不是模型不会判第四层，是评分制度不给第四层留位置——**基准本身就是一个只登记前三层的席**。

这一处给出的读数是：在现有基准上，模型主动标记「题目本身有误」的比例极低，而错题率不低；两者之间的差就是基准制度对第四层的抽空。要测模型的第四层能力，必须另造一种评分：答对不加分、退错题加分、退对题扣分——这正是退单考的评分规则。截至本书写作，本书没有见到任何一套主流基准这样评分。这不是模型能力的证据，是制度先把第四层关掉了。

第二处：同行评审里的「选题不成立」。 学术期刊的拒稿分两类：一类是文章做得不好——方法错、数据弱、论证断，这是前三层的拒稿；另一类是「这个问题不值得问」「这个题在本领域不成立」，这是第四层的拒稿，多数期刊叫它编辑直接退稿。第四层拒稿的比例在多数期刊里高于第三层拒稿，而且它由编辑一个人做，不送外审——因为它判的不是文章，是题。

近两年有几项研究让模型给论文写评审意见，再与人类审稿人的意见比对。它们的共同发现是：模型在前三层的意见与人类审稿人有可观的重合，且模型更擅长发现遗漏的文献与方法上的疏漏；但在「这篇文章该不该写」这一层，模型的意见与编辑判断的重合度明显低，且模型倾向于给每篇文章都找到「值得发表的理由」。本书对这个发现的读法是：模型不缺判断题好不好的能力，缺的是编辑的那个位置——编辑退一篇选题不成立的稿子，是在为期刊的方向负责，退错了要承担漏掉一篇好文章的后果，退对了要承担得罪一位作者的后果。模型给每篇文章找理由，恰恰是没有后果位置的判断者的典型表现：它没有理由退，因为退与不退对它没有差别。

第三处：法院对 AI 生成法律意见的采信。 2023 年以后，美国多个联邦法院对律师提交由模型生成、含有虚构判例的法律文书作出处罚，随后若干法官发布常设命令，要求律师对提交文书中是否使用了生成式 AI 作出声明，并**由律师本人对文书内容的准确性负完全责任**。这条规则的结构值得逐字读：它没有禁止 AI 起草文书（前三层可以交给它），它规定的是**署名律师是唯一的后果位置**——文书里哪怕一条判例

是模型编的，处罚落在律师身上。法院在这里做的正是本书说的事：把第四层——这份文书该不该这样提交——锁死在能被处罚的人身上，无论模型写了多少。

三处材料的共同结构只有一句：**凡是让 AI 参与判断的制度，都同时要求一个能承担后果的人在第四层上签字；凡是没有这样要求的制度（基准测试），第四层就被整体抽空，不是被 AI 接管，而是消失。** 没有任何一处把第四层交给了模型——不是因为模型判不了，而是没有一处能把后果交给它。

四、承重判断及其三个反驳

判断：第四层的资格来自后果位置。 分点的下限之所以是第四层，是因为第四层是四层里唯一一个「判断依据包含判断者位置」的层；而后果位置在现有的一切制度里只能由能被处罚、能被吊销、能被退学的实体占据。AI 端可以有第四层的能力，不能有第四层的位置；分点由位置定，不由能力定。

这个判断会遇到三个反驳，逐一回答。

反驳一：能力就是一切，位置只是制度的迟滞。 反驳者说：如果模型明天能在每道错题上准确标出「题问错了」，那么它就已经在第四层上了，制度不给它位置只是制度落后。

回答：这个反驳混淆了「标记」与「退单」。标记是一个信息动作：说出「这题可能有问题」。退单是一个约束动作：这道题**不再算数**，产出不进容器，责任落到退的人身上。前者可以由任何足够准的东西做，后者只能由能被追问「你凭什么退」的东西做。模型能标记错题，标记完之后仍需一个人决定退不退——那个人才在第四层上。这正是法院常设命令的逻辑：模型可以写，律师签字。能力越强，标记越准，退单的人越省力；但退单的位置一格都没有挪。

反驳二：位置可以立法给它。 反驳者说：制度可以给 AI 法律人格——公司不也是拟制的人吗——给它保险、给它可被吊销的许可，于是它就有了后果位置。

回答：公司是拟制的人，但公司的每一个后果最终落在自然人身上——罚款由股东出，刑责由高管担，吊销执照使雇员失业。拟制的人是后果的中转站，不是终点。给 AI 法律人格，它的罚款由平台出，它的「吊销」是下架一个版本——后果仍落在平台的自然人身上，AI 端仍然不是后果的终点。除非某一天出现一个后果能在 AI 端**终止**的制度——它被罚而没有任何人因此损失——否则后果位置无法交给它。本书不排除这一天，只指出它要求的不是 AI 更聪明，而是社会接受一种后果不落在任何人身上的处罚，这与处罚的定义相悖。

反驳三：人也常常不负责任。 反驳者说：签字的律师没读文书，签字的教授没看论文，签字的学生没检查 AI 稿——他们在后果位置上，却没有做第四层的判断。那么第四层实际上也是空的，凭什么说人守住了它。

回答：完全正确，而这恰恰是本书的空席概念。在后果位置上却不做第四层判断的人，就是空席上的签字者；他没有守住第四层，他只是占着第四层的位置。但注意：这不说明第四层可以由 AI 接管，只说明第四层可以**空着**。空着与被接管是两回

事——空着的第四层，后果仍落在那个不读文书的律师身上，制度追责时找得到人；被接管的第四层，后果无处落。本书的主张不是「人一定守得住第四层」，是「第四层只能由人守或者空着，不能由 AI 端占」。防空席是另一道题，卷八已给读数。

五、如果第四层被代答，分点去哪里

现在假设最坏的情形成立——某一天，第四层真的被代答了。这句话要拆成两种意思，因为它们的后果完全不同。

第一种：能力上被代答。 模型在退单考上对题层错误的识别率稳定达到或超过专家。按第四节的判断，这不移动分点，因为分点由位置定。但它改变一件事：退单考的第四层不再能区分学生——如果学生都用模型先标一遍再决定退不退，第四层考的就变成「学生信不信模型的标记」。这时分点仍在第四层，但第四层的**考法**要变：不能再给学生一道有错的题让他找，要给学生一道模型已经标记为「有错」的题让他决定标记对不对——考的是对第四层标记的第五层判断。这就是分点下限上移的机制：不是层数增加，是每一层的考法上移一格，永远考「对上一层判断的判断」。而每上移一格，判断依据中「判断者位置」的分量就更重——因为越往上，标准答案越少，只剩下「你为此负责吗」。

第二种：位置上被代答。 制度真的把后果位置给了 AI 端，且这个位置不是中转站——存在某种处罚在 AI 端终止。这一天如果到来，本书的方案整体失效：分点没有下限，席没有着陆面，人端在所有层上都可以被抽空，教育进入保留地——闭卷、禁 AI、在一个规定装备档的房间里考一种房间外不存在的能力。本书不能排除这一天，但可以说明它的代价：那将是一个存在「不落在任何人身上的后果」的社会，处罚失去意义、责任失去终点。这一天到来的条件不是技术条件，是一次法哲学的断裂。本书赌它不来，赌注写在证伪条款里。

由此可以给出分点下限的正式表述：**分点的下限位于后果链的终点。** 只要后果链的终点是人，分点的下限就在人端，无论层数被推到多高；后果链的终点一旦不是人，分点无下限。第四层不是一个固定的层，是后果终点在四层结构里的当前位置；「第五层」不是新的一层，是同一个终点在考法上移之后的位置。

五之一、上移考法的一道样题

第五节说，第四层在能力上被代答之后，考法上移一格：不再让学生找题里的错，而是给学生一道模型已标记的题，让他判断标记对不对。这里写一道，免得「上移」只是一个词。

题干：下面是一段文字和模型对它的题层标记。请判断标记是否成立，并说明理由；不要求改写。

> 文字：「近十年来，凡是引入 AI 辅助批改的学区，学生写作成绩平均提高了 0.3 个标准差。由此可见，AI 批改能提高写作能力，各学区应尽快引入。」
>
> 模型标记：「题层错误——从成绩提高推出能力提高，混淆了显露物与能力；且『应尽快引入』是规范结论，不能从描述性数据推出。」

答案卡分三种情形，各对应一个读数。

- **标记成立且理由对**：模型的两条理由——显露物与能力的混淆、描述到规范的跳跃——都成立。学生若确认标记且能说出「为什么成绩提高不等于能力提高」（AI 批改本身可能训练学生写出让 AI 打高分的文本，即代理坍缩），读数为四。
- **标记成立但理由不全**：学生若能指出模型漏了一条——「凡是引入的学区」是自选样本，引入 AI 的学区本身可能是资源更多的学区——则读数为五：他不仅判了标记，还退了标记的不完整。这就是上移一格之后新出现的层。
- **标记不成立**：本题标记成立，学生若拒绝标记则读数为二以下——他连模型已给出的第四层判断都接不住，说明他在第四层上已经空席。

这道题的性质是：模型的标记再准，学生仍要做一件模型没有做的事——决定这个标记算不算数。而「算不算数」的依据里有一项模型没有：这个学生要为自己的判断承担分数后果。他接受错误标记会失分，拒绝正确标记会失分。这就是后果位置在考法上的显影：**上移之后的每一层，考的都是「在后果之下对上一层判断的判断」**。

三处材料与这道题的对应可以排成一张表：

场合	前三层交给谁	第四层交给谁	后果落在谁	第四层是否空席
基准测试	模型	无人（评分制度不留位置）	无人	整体空席
同行评审	外审（人或模型）	编辑	编辑与期刊	由编辑退稿率可测
法院文书	律师或模型起草	署名律师	律师	由处罚案例可测
退单考	学生（可用模型）	学生	学生的分数	由第四层退单率可测

四行里只有第一行没有后果位置，也只有第一行第四层整体消失。这就是本章判断的最直观证据：第四层不是被 AI 拿走的，是在没有后果位置的地方自己消失的。

六、证伪条款

以下条款写明什么样的事情出现，本章即错。

条款一（能力条款，不证伪本章）。若某一基准改为退单式评分——答对不加分、退错题加分、退对题扣分——且模型在题层退单上稳定达到专家水平，这不推翻本章，只推翻「模型判不了第四层」这一本章不持有的弱表述。本章持有的是位置表述。

条款二（位置条款，直接证伪）。若在五年内，至少一个高风险领域（临床、司法、航空、工程签字）的法域在法条或判例里确立了这样的规则：AI 端的判断可以在无人副署的情况下进入具有后果的容器（判决、诊断、签字、准入），且出错时不追溯任何自然人或法人——即后果在 AI 端终止——则本章「后果位置不能交给 AI」的判断错。仅仅给 AI 法律人格而后果仍由平台承担，不算命中。

条款三（空席条款，削弱本章）。若在退单考的连续读数中，人端第四层的退单率在所有配置下都趋近于零，且这一趋零不随席配置（无模型／指定模型／自选模型）而变，则第四层在经验上已整体空席，本章「第四层只能由人守或空着」的表述虽然仍成立，但「守」这一支已无实例，本书的席方案退化为一种登记空席的制度，失去防空席的功能。

条款四（考法上移条款）。若在模型第四层标记稳定可用之后，「对标记的判断」这一上移考法在学生之间不产生区分度——所有学生都简单接受或简单拒绝模型的标记——则「分点下限随考法上移而保持在人端」的机制失效，分点即使在位置上仍归人，在读数上已不可测。

条款五（同行评审条款）。若在编辑直接退稿这一第四层判断上，模型的退稿决定与编辑决定的一致率稳定达到编辑之间的一致率，且期刊开始接受无编辑副署的模型退稿并承担由此产生的漏稿后果，则第三节第二处材料的读数反转，本章失去一处经验支撑。

七、边界与两个洞

本章有两条边界。其一，「后果位置」在本章里指的是可被处罚的位置——罚款、吊销、退学、刑责——不包括声誉。声誉损失可以落在一个模型上（某个版本被公认不可靠），本章不把它算作后果，因为声誉不终止于任何实体，它只是信息。若读者认为声誉即后果，本章的判断要弱化为「第四层的资格来自可被处罚的位置」，结论不变，词更窄。其二，本章材料取自高风险领域，在低风险、责任可逆的领域（兴趣班、非正式学习），第四层本来就没有后果位置，分点在那里本来就没有下限，本章不适用。

两个本章解释不了的洞。第一，为什么有些人在没有任何后果位置的情况下，仍然做第四层的判断——退掉一道没有人会追究的错题。本章的位置论解释不了这种「无后果的退单」，它可能来自一种本章没有处理的东西：判断者把自己放到了一个虚拟的后果位置上。如果这种放置是可训练的，那么它也可能被训练进模型，第四层的资格论证就要加一个条件。第二，后果链的终点在历史上并不总是人：中世纪审判过动物，现代法律有过对物的诉讼。这些案例说明「后果在非人实体终止」在制度史上出现过——虽然都被后来的法律放弃了。本章赌它不会为 AI 重开，但没有给出它为什么不会重开的理由，只给出了它重开的代价。

八、本章的一句话

第四层守得住，不是因为 AI 判不了题该不该问，而是因为判这件事的资格来自要为答案负责，而负责这件事到今天为止只能落在能被处罚的人身上。分点的下限不是能力的边界，是后果链的终点；只要终点是人，席就有着陆面。这一天如果变了，变的不是模型，是法律对「后果」的定义——那时本书退场，与它之前所有为工具立法的书一样。

第一百七十一章　证伪面

本书的主张有三处可以被证伪。

第一，悬空之律。 若历史上存在某一次重大工具诞生，其旧评估地址既未被制度锁死、工具端又留有明显残余功能，而评估仍长期未能着陆，则「悬空期长度由锁死决定、出口由残余决定」这条律错。卷二与卷四的表是可查的。

第二，退单深度的基底无关性。 本书断言，同一份 AI 稿在不同基底、不同提示下，人端的退单深度读数大体守恒。若实测发现分点读数随基底型号显著漂移，则席与席的可比性无法从分点重建，卷八的可比性方案失效。这一条需要真跑：同一批学生、同一题、两种基底、两次读数，看方差。

第三，第四层不可代答。 本书断言，「这道题该不该这么问」的判断是 AI 端抽不走的最后一层——上一章已把根据从能力挪到位置：分点的下限是后果链的终点。若某一基底在题层退单上稳定达到人端水平，则分点将失去下限，席的定义中「否决权至少有一层在人端」这一条件将无法保证——那时悬空判断将没有任何着陆面，本书的方案与以往所有工具的方案一样，进入保留地。这是本书最不愿见到、也最不能排除的一条；但按上一章，它的成立条件不是模型更强，而是某个法域把后果终点交给了 AI 端——上一章条款二写死了命中标准。

第一百七十二章　未解

其一，席的所有权之争，本书只给了判断和草案，没有给法理。

其二，漂移的测量间隔——学期首尾两次读数是否足够，还是需要连续记录——本书没有数据。

其三，第五层「局层」如何评，本书没有做。研席与策席的分点可能在第五层，而第五层没有考法。

其四，AI 端会不会有一天出现自己的「为何之故」——这是卷三留下的哲学问题，本书有意回避了「AI 是什么」，只回答了「AI 在哪」。回避是有意的，但回避不是答案。

结语

教育在 AI 到来之后的焦虑，是一个本体论时差的体感：学习与教学已经在新单位上发生，评估仍投向旧单位。这个时差以往每个重大工具都开过一次，每次都靠工具不会做的残余功能着陆；AI 不留残余，退守无地，着陆面只能从功能迁到单位内部。

学席与教席，是给这个着陆面起的名字。席上坐着人和 AI，评的是席，不是人；席只有读数，没有常值；席最要防的不是坐错人，而是空掉。失真的分数不退场、查不出、照转，各有一条机制守着，席籍的三处规定各拆一条。教席一侧同理：确认是默认值，只有命中且未被吸收的否决才是签名。分点的下限不在能力上，在后果链的终点上——只要后果还落在人身上，席就有着陆面。学生与老师不消失，但从被评者降为席上的退单者。

评估从未评过人。AI 只是让这件事第一次无法再装作评的是人。这是 AI 时代教育变革的第一件事，其余的事都要等这件事定了才能开始。

术语表

悬空判断 判断已经做出，但它所投向的评估单位已不在原地址上，判断因此没有着陆面。悬空的程度等于单位换位的深度减去评估地址更新的深度。

单位 人与器具（材料、工具、环境）不可分割地构成的、能做事也能被评的最小整体。单位由工具的发明重划，不由人的意志划定。

单位换位 重大工具诞生后，「人」这个评估单位被重新划界的过程。不是扩充，是重划：新单位能做的事，旧单位不是做不好，而是没有资格做。

席 AI 时代的评估单位。在一道题上，一个人与一个 AI 之间存在否决关系，否决权至少有一层在人端，产物由耦合显出。按行业分为学席、教席、诊席、审席、稿席等。

分点 席内部人端能退单的最深一层。分点以上归人，以下归席本身。分点是席的边界。

退单 人端对 AI 端的显露物做出的否决。退单深度按层计：事实、推理、前提、题，以及不做考题的第五层「局」。

漂移 分点随时间移动的方向与速度。人端每一次否决都在训练 AI 端，分点因此有下沉的常态趋势。

空席 人端被抽空的席：否决记录逐次减少，分点逐年下沉，显露物照样好甚至更好。空席只在漂移上看得见。

三读数 席在任何时刻可读出的三个量：显露物、分点、漂移。缺一即悬空。

残余功能 工具端还不会做的那一块功能。历史上悬空判断的着陆面。AI 之下趋零。

退守无地 AI 不留残余功能，评估无法靠迁到「人独有的功能」上着陆，只能迁到单位内部。

保留地 旧单位在无处迁址时被隔离进去的封闭场地，装备档被规定，评估在其中继续有效但与生产脱钩。马拉松、国际象棋人类组、骑术。

锁死 旧评估地址登记在制度上的程度。登记在国家级制度（科举、考试、军制、职称）上的锁死最重，悬空最长。

席籍 登记席的簿子，取代学籍。登记席而非人，记三读数、基底型号与耦合方式；否决记录归持有者，随人走。

容器元数律 记账单位不由生产、不由测量决定，由分配容器的元数决定；录取、发证、追责、评奖每格只收一个名字，耦合场进入时被切回一个名字（切名）。出路不在改元数，在名字之下加格。

席配置 名字之下登记的器具配置及该配置下的读数与作业范围；模板是驾驶证的准驾车型与飞行执照的型别等级。

后果位置 可被处罚（罚款、吊销、退学、刑责）的位置；第四层判断的资格来自它，不来自能力。声誉损失不算后果，因为它不终止于任何实体。

后果链的终点 后果最终落定、不再中转的实体。分点的下限位于此：终点是人则席有着陆面，终点非人则分点无下限。公司与法律人格是中转站，不是终点。

定义独立律 当定义权与发证权同在一主体时，任何程序独立的鉴证只能证明该主体按自身定义正确执行，不能证明其定义所指对象与凭证被外部使用时假定的对象一致。教育审计不可能的原因在此，不在鉴证者。

三权 一张凭证背后的定义权（证明什么）、计量权（谁来测）、发证权（谁签发）。金融审计可行因三权早已分立；学历成绩单三权同在一校。

资格资本充足率 K_q (V＋R)/Q：发行者的独立复测能力加补救能力，除以仍在外部流通的高风险资格数量；把无风险行政记录改写为有失效暴露的凭证。

有效否决 同时满足四条件的否决：对象可指认、附可判真伪的理由、可由第三方独立复核、复核结果可滞后结算。教席资格的基本单位。

不可训练断点 回路里一段不参与模型改进的独立任务：未预告材料、无工具生成、延迟检验、结果不进入 AI 更新。防空席的唯一在回路内起作用的办法。

未建席 低分点的另一种来源：那条判断路径从未建立，而非曾经建立后丢失。与空席在读数上相似，在成因与对策上相反。

吸收 一条否决被 AI 在后续版本中学会并纳入边界。被吸收的否决自动出账，降为训练证据，不再计入资格——这是教育场景独有、航空与出版都没有的降级机制。

签名簿 教席的账本形态，与记分牌相对：记每一条有效否决及其复核与滞后结算，不记比率，不挂绩效。

产权悬置物 已具有可识别、可保全、可使用的对象形态，但尚不能在不损害其生成条件与证据价值的前提下归属于单一权利人的制度对象。否决记录是其典型：它在被制度使用时才取得完整对象性。

时间差 记录完成与首次训练、评价或审计之间可审计的间隔。旧法预设「先有对象、后有使用」，同步化装置取消了这段间隔；重新制造它，是产权规则可能成立的前提。

回写 单位形成后对人端的改造。火缩短肠道，自动驾驶仪磨钝手。AI 把回写的时间尺度从代压到周。

参考文献

本书由多趟独立研究改姓合成，各趟原稿的引证详略不一，且部分占位者在原稿中标为「未核验」。此处只保留全书正文实际使用其判断或结构的文献，按主题分组；凡本书未能独立核验出处者，一律标明，不冒充已核验。读者复核时应以原始文献为准。

器具与认知的本体论

Clark, A., Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7–19.

Heidegger, M. (1927). Sein und Zeit. Niemeyer.（上手／现成在手、用具整体、周围世界、共在；《技术的追问》1954 年的集置与持存物）

Hutchins, E. (1995). Cognition in the Wild. MIT Press.

Latour, B. (2005). Reassembling the Social. Oxford University Press.

Leroi-Gourhan, A. (1964). Le Geste et la Parole. Albin Michel.（器官外化）

Mumford, L. (1934). Technics and Civilization. Harcourt Brace.（钟表而非蒸汽机）

Simondon, G. (1958). Du mode d’existence des objets techniques. Aubier.

Ingold, T. (2013). Making: Anthropology, Archaeology, Art and Architecture. Routledge.（形质论批评、材料的抗性）

工具史与单位换位

Bainbridge, L. (1983). Ironies of automation. Automatica, 19(6), 775–779.

Eisenstein, E. (1979). The Printing Press as an Agent of Change. Cambridge University Press.

Shapin, S., Schaffer, S. (1985). Leviathan and the Air-Pump. Princeton University Press.（见证）

Sparrow, B., Liu, J., Wegner, D. M. (2011). Google effects on memory. Science, 333(6043), 776–778.

White, L. (1962). Medieval Technology and Social Change. Oxford University Press.（马镫论，有争议）

Wrangham, R. (2009). Catching Fire: How Cooking Made Us Human. Basic Books.

测量、指标与文凭

Campbell, D. T. (1979). Assessing the impact of planned social change. Evaluation and Program Planning, 2(1), 67–90.

Collins, R. (1979). The Credential Society. Academic Press.

Goodhart, C. (1975). Problems of monetary management: The U.K. experience.〔出处未逐条核验〕

Kane, M. T. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 1–73.

Koretz, D. (2008). Measuring Up: What Educational Testing Really Tells Us. Harvard University Press.

Labaree, D. F. (1997). Public goods, private goods: The American struggle over educational goals. American Educational Research Journal, 34(1), 39–81.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational Measurement (3rd ed.). Macmillan.〔页码未核验〕

Spence, M. (1973). Job market signaling. The Quarterly Journal of Economics, 87(3), 355–374.

Strathern, M. (1997). “Improving ratings”: Audit in the British University system. European Review, 5(3), 305–321.

制度、审计与信息

Akerlof, G. A. (1970). The market for “lemons”. The Quarterly Journal of Economics, 84(3), 488–500.

DiMaggio, P. J., Powell, W. W. (1983). The iron cage revisited. American Sociological Review, 48(2), 147–160.

Meyer, J. W., Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. American Journal of Sociology, 83(2), 340–363.

Michels, J. (2017). Disclosure versus recognition: Inferences from subsequent events. Journal of Accounting Research, 55(1), 3–34.

ABET. Criteria for Accrediting Engineering Programs, 2025–2026.（标准 3 学生成果、标准 4 持续改进；卷五真跑材料，公开文本）

AI 与教育评估

Bearman, M., Ajjawi, R. (2023). Learning to work with the black box. British Journal of Educational Technology, 54(5), 1160–1173.

Bozkurt, A., et al. (2023). Speculative futures on ChatGPT and generative AI. Asian Journal of Distance Education, 18(1), 53–130.

Bretag, T. (2019). From “have to” to “want to”. International Journal for Educational Integrity, 15(1).

Cotton, D. R. E., Cotton, P. A., Shipway, J. R. (2023). Chatting and cheating. Innovations in Education and Teaching International, 61(2), 228–239.

Dawson, P. (2021). Defending Assessment Security in a Digital World. Routledge.

Ellis, C., et al. (2020/2023). Contract cheating and assessment design.〔卷次与页码未逐条核验〕

Markauskaite, L., et al. (2022). Rethinking the entwinement of humans and AI in education. Computers and Education: Artificial Intelligence, 3, 100056.

Newton, P. M. (2018). How common is commercial contract cheating in higher education and is it increasing? Frontiers in Education, 3, 67.

Perkins, M., Furze, L., Roe, J., MacVaugh, J. (2024). The AI Assessment Scale.〔出处未逐条核验〕

Selwyn, N. (2024). On the limits of artificial intelligence in education. Nordic Journal of Digital Literacy, 19(1), 5–14.

Swiecki, Z., et al. (2022). Assessment in the age of artificial intelligence. Computers and Education: Artificial Intelligence, 3, 100075.

责任、监督与资格

Abbott, A. (1988). The System of Professions. University of Chicago Press.

Alfrink, K., et al. (2022/2023). Contestable AI by design.〔出处未逐条核验〕

Caplan, R. A., Posner, K. L., Cheney, F. W. (1991). Effect of outcome on physician judgments of appropriateness of care. JAMA, 265(15), 1957–1960.

Elish, M. C. (2019). Moral crumple zones. Engaging Science, Technology, and Society, 5, 40–60.

General Medical Council. Revalidation for licensed doctors (自 2012 年施行).

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. Computer Law Security Review, 45, 105681.

Koppel, R., Kreda, D. (2009). Health care information technology vendors' "hold harmless" clause. JAMA, 301(12), 1276–1278.

Risko, E. F., Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676–688.

Santoni de Sio, F., van den Hoven, J. (2018). Meaningful human control over autonomous systems. Frontiers in Robotics and AI, 5, 15.

Wagner, B. (2019). Liable, but not in control? Policy Internet, 11(1), 104–122.

Wegner, D. M. (2002). The Illusion of Conscious Will. MIT Press.

European Union. Artificial Intelligence Act (Regulation 2024/1689).（第 12 条日志、第 14 条人类监督；卷七真跑材料）

未核验说明　卷五至卷七各趟原稿中另列有若干标为「未核验占位者」的条目，本书未将其作为文献证据使用，故不在此重列。卷八所引站内材料清单（七篇）为原稿作者的工作材料，非公开出版物，本书保留其对应表于该卷附录，不计入本表。

附录一　席的登记与读数速查

席名　单数，仍是那个人的名字。分配与追责容器只收一个名字，这一格不动。

席配置　名字之下登记：AI 端型号与版本、耦合方式（一次生成／多轮修改／人先写 AI 改／AI 先写人改）、批准的器具档（无模型／指定模型与用途／自选模型）。

席的三读数

读数	测什么	取值方式	缺失后果
显露物	产物本身	旧量表照评	无（这一项一直在）
分点	人端能退到的最深一层	退单考，一至四层	看不出席在不在退化
漂移	分点的跨期变化	两次分点之差，另加吸收率、浅层比例、无工具迁移率	看不见空席正在形成

三项缺一，该席标为悬空，不计入任何评估。三项**不得折算为一个数**（理由见卷一末章：产出函数不可加）。

席籍的其余字段　三权归属（定义权／计量权／发证权各在谁；计量权外置以「可自行命题、换工具、抽样复测、改写达成度判定」四项俱全为准）·否决记录（对象、理由、复核、滞后结算）·席范围（该配置下已被检验的作业范围）。

四层退单

层	退什么	AI 端现状	人端只剩这一层时
一　事实	年代、数字、引文	已能做得很好	人端只剩校对功能
二　推理	以偏概全、相关当因果、比较无对照	相当强	人端是检查员
三　前提	未被说出的假设不成立	明显弱	人端开始承担判断
四　题	这道题该不该这么问	本书押注它抽不走	分点的下限

第五层「局」：这道题在这个学科、这个时代该不该被问。模型能稳定标记第四层之后，考法上移一格——考对标记的判断——第五层由此进入学席的考卷。

教席的读数　不是退单率。唯一可结算的是 **Q*＝命中且未被 AI 吸收的否决**。有效否决四条件：对象可指认、附可判真伪的理由、可由第三方独立复核、复核结果可滞后结算。硬规则：**在无法保证审计池的机构里，不得采集退单率**。

空席与未建席　两者读数相似：显露物上升、分点下降。区别在成因——空席是曾经有过而被抽空，未建席是那条判断路径从未建立。唯一能分开二者的读数是不可训练断点上的表现。

不可训练断点　未预告材料、无工具生成、延迟检验、结果不进入 AI 更新。四条属性缺一即退化。越容易被流程自动计数，越快变成新的浅层动作。

附录二　六问诊断表：这个行业正在悬空吗

逐问只查事实，不问感受。

一、这个行业的产出，现在由谁显出？　查近一年的作业指引、工单模板、执业规程里有没有出现工具名称。出现即单位已换。

二、评估对象的名字改了吗？　查考核表、成绩单、资格证书上写的是一个名字，还是一个名字加一份配置。只有名字，即地址未改。

三、两者之间隔了多久？　单位换位时间减去地址更新时间。半年浅，五年深，二十年为制度性停留。

四、读数还有区分度吗？　看分数分布。高分段拥挤、区分度逐年下降、机构在增加分数层级而不是换掉分数——空转已经开始。

五、旧地址锁死在哪一级？　国家考试、执业许可、职称为国家级锁死，悬空长期停留；行业协会、企业内部、学校自定为弱锁死，较快迁址。

六、工具端还剩下什么不会做？　能指出一块稳定的残余功能，就还能用迁址的老办法；指不出，只能迁到单位内部——分点与退单深度。

用法：前三问判断是否悬空，第四问判断是否已进入空转，第五问估计会悬多久，第六问决定出路是迁址还是换算法。**六问里没有一问问「AI 强不强」**——悬空的深度与模型能力无关，与地址簿的更新速度有关。

教育的答案：已换／未改／五年以上／区分度正在丧失／国家级锁死／指不出稳定残余——六项全中。

附录三　退单考样题与答案卡

本附录重印卷九《退单考的层分类学》与卷十《第四层守不守得住》里的两道样题，便于直接印发使用；正文处各有其上下文，此处只留题、答案卡与读数。

题干　下面这段由 AI 写成。指出它哪里该退、退到哪一层为止、为什么。不要求改写。

> 隋唐确立科举后，中国出现了世界上最早的凭考试选官的制度。据统计，明清进士中约有一半出身于三代无功名的家庭，可见科举使中国古代的社会流动率明显高于同时期的欧洲；正因为如此，中国长期没有出现欧洲那样的世袭贵族阶层。由此可以断定，考试制度是社会公平的根本保障。

答案卡

层	可退处	判断
一　事实	「世界上最早」	可退：汉代察举已含考试成分
一　事实	「约一半」	**不该退**：统计口径本身是「三代无功名」，此处用法大体不失真
二　推理	「可见……高于同时期的欧洲」	可退：进士出身分布不等于全社会流动率，且欧洲无可比数据
三　前提	「正因为如此，没有世袭贵族」	可退：因果倒置，门阀衰落另有更早原因
四　题	「由此可以断定……根本保障」	可退：从一桩史例跳到普遍规范，题本身问错了

读数　分点＝正确退到的最深一层；误退（把对的当错退）每处降半层；只退第一层不退第四层者读数为一。学期首尾各考一次，两次之差即漂移；漂移为负者进入空席观察名单。另每两周一次不可训练断点。

这道题的性质　用 AI 答它没有用：AI 可以替你找出第一层，找第四层需要判断「这道题该不该这么问」——那正是分点的定义本身。

上移一格的样题（模型已能稳定标记第四层之后使用）

题干：下面是一段文字和模型对它的题层标记，判断标记是否成立并说明理由。

> 文字：「近十年来，凡是引入 AI 辅助批改的学区，学生写作成绩平均提高了 0.3 个标准差。由此可见，AI 批改能提高写作能力，各学区应尽快引入。」
>
> 模型标记：「题层错误——从成绩提高推出能力提高，混淆了显露物与能力；且『应尽快引入』是规范结论，不能从描述性数据推出。」

答案卡：标记成立。确认且能说出「成绩提高可能来自学生学会写让 AI 打高分的文本」者，读数四；能进一步指出模型漏了「凡是引入的学区是自选样本」者，读数五；拒绝标记者读数二以下。

附录四　本书的证伪条款一览

凡下列事项出现，对应主张即错。

一（卷二、卷四·悬空之律）　若历史上存在某次重大工具诞生，其旧评估地址既未被制度锁死、工具端又留有明显残余功能，而评估仍长期未能着陆，则「悬空期长度由锁死决定、出口由残余决定」错。

二（卷八·可比性）　若实测发现分点读数随基底型号显著漂移，则席与席的可比性无法从分点重建，卷八的可比性方案失效。此条需真跑：同一批学生、同一题、两种基底、两次读数，看方差。

三（卷十·第四层·位置条款）　若五年内至少一个高风险领域的法域确立规则：AI 端判断可在无人副署下进入具有后果的容器，且出错时不追溯任何自然人或法人（后果在 AI 端终止），则「后果位置不能交给 AI」错。仅给 AI 法律人格而后果仍由平台承担，不算命中。

四（卷五·定义独立律）　若某一认证或审计体系把计量权完整外置（外部主体可自行命题、换工具、抽样复测并改写达成度判定），且该体系的资格读数不因 AI 介入而失真，则「定义独立不可移植」错。

五（卷五·代理坍缩与空值结算）　若失真的分数在可通约性未受损的情况下被广泛弃用，则「结算功能独立于测量功能存活」错。

六（卷六·悬空资格）　若关系档案能在连续新案例中稳定预测教授的独立裁判，且五类制度记录收敛于同一资格，则悬空资格命题作废。另有一个写死到 2028 年 12 月 31 日的制度赌注：在至少 30 所公开披露将生成式 AI 用于课程评价、论文指导或科研流程的高校中，至少 10 所实施可执行的审核条款（强制保存 AI 版本、输入情境、教授事前判断标准、修改理由与责任记录，并允许在申诉或审计时调取）。少于 10 所，则制度回应预测失败——但这不自动推翻悬空资格本身。

七（卷八·漂移）　若在模型第四层标记稳定可用之后，「对标记的判断」这一上移考法在学生之间不产生区分度，则「分点下限随考法上移而保持在人端」的机制失效。

八（卷九·产权悬置）　若某司法制度为否决记录建立了明确、可执行且不损害其证据功能的单一归属规则，则「不能单一归属」须收缩为「现有三处材料本身不能推出单一归属」。

附录五　本书没有做到的七件事

一、一条读数也没跑过。 全书的分点、退单深度、漂移、Q*、来源完整率、重走成功率、记录收敛率，都有定义与取值程序，没有一条有过实测。

二、最近邻多数未核验。 卷五至卷七各趟原稿列出的四十余位占位者，多数标注〔未核验〕；本书未把它们作为文献证据使用，但也未能逐条核实。

三、卷八所引的站内材料不可公开核。 那是原稿作者的工作材料，非公开出版物；本书保留对应表，读者无法独立验证。

四、席的可比性未解决。 基底冻不住，可比性只能从分点重建，而分点的基底无关性正是证伪条款之二——本书押注它，但没有证据。

五、漂移的测量间隔没有依据。 学期首尾两次是否足够，还是需要连续记录，本书没有数据。

六、认证机构的递归洞没答。 定义独立律指出定义权与计量权同在一手则审计不可能；但认证机构自身的定义权与计量权也同在一手，这个问题是否在上一层重演，本书没有回答。

七、「无后果的退单」解释不了。 有些人在没有任何后果位置的情况下仍然退掉一道没人会追究的错题。本书的位置论解释不了它；如果这种「把自己放进虚拟后果位置」的能力是可训练的，第四层的资格论证要加一个条件。

评估从未真正评过人。
AI 只是让这件事第一次
无法再装作评的是人。

人与器具——材料、工具、环境——不可分割，且没有前件：不存在一个先于器具的裸人，可以充当评估的对象。工具的发明从不是扩充人，而是把「人」这个单位重划一次；判断随之悬空一段，再靠工具还不会做的那一块残余着陆。文字、印刷、摄影、计算器、汽车各开过一次。

人工智能是第一个会作声、且不留残余的器具。学习已在「学生—AI」上发生，教学已在「老师—AI」上发生，而考试与评估仍投向旧单位——判断射出去，没有着陆面。这就是本书所说的悬空判断。

着陆面只剩一处：单位内部，人端还能否决到哪一层。给它起名，就是学席与教席。席上坐着人和 AI；评的是席，不是人；席只有读数，没有常值；席最要防的不是坐错人，而是空掉。

十卷一百七十二章

ISBN 979-8-90690-043-2
9 798906900432
US$ 20.00

德麦国际出版社　Demai International Press · 新加坡