AI时代的人类能力建构与认知韧性理论

Adjust task format without lowering challenge intensity; assess by relative improvement rather than absolute level; (if accessibility applies) challenge budget conservation.

副标题：基于反脆弱性验证原则的认知内共生范式

作者：杨国平 邮箱：a44425874@gmail.com 版本：V1.0 日期：2025年9月

摘要

研究目的：随着生成式AI的广泛应用，人类独立认知能力面临退化风险。本研究旨在建立一个可证伪的理论框架，用于评估和

研究方法：核心包括三大机制：(1)反脆弱性验证原则（AVP）——通过”拔线测试”检验协作是否促进独立能力；(2)内共生—设计有益认知摩擦与系统性支持削减；(3)伙伴式主体性——AI作为认知伙伴的角色定位。采用跨学科综合方法，整合认知

主要结果：建立了四层跨尺度验证体系，识别了”认知外骨骼”（过度依赖导致能力退化）与”认知内共生”（能力持续增强H8），为后续实证研究提供清晰路径。

研究结论：为AI时代的能力建构提供了首个系统性、可操作的评估标准。其价值在于：提出了超越表面效率、关注独立能力的

关键词：人工智能；认知能力建构；反脆弱性验证；人机协作；能力评估；认知韧性

Abstract

Objective: With the widespread adoption of generative AI, human independent cognitive capabilities face risks of degradation. This study aims to establish a falsifiable theoretical framework for assessing and optimizing human capability building in the AI era.

Methods: The core comprises three mechanisms: (1) Antifragility Validation Principle (AVP)—verifying whether collaboration enhances independent capability through “unplugged tests”; (2) Endosymbiotic Minimal Law (EML)—designing beneficial cognitive friction and systematic support reduction; (3) Partner-like Agency—positioning AI as a cognitive partner. The study adopts an interdisciplinary synthesis approach, integrating research from cognitive psychology, educational technology, organizational behavior, and AI ethics, validated through literature analysis, case studies, and conceptual modeling.

Results: This paper establishes a four-layer cross-scale validation system, identifies two contrasting patterns—“cognitive exoskeleton” (over-reliance leading to capability degradation) versus “cognitive endosymbiosis” (continuous capability enhancement)—and designs a Layered Symbiosis Architecture (LSA) as an engineering implementation pathway. Through case analyses across three domains (writing, programming, team collaboration), the theory’s explanatory and predictive power is validated. The study explicitly articulates six major limitations and eight falsifiable hypotheses (H1-H8), providing clear pathways for subsequent empirical research.

Conclusions: This theory provides the first systematic, operational assessment standard for capability building in the AI era. Its value lies in: proposing an assessment paradigm that transcends surface efficiency to focus on independent capability; establishing a consistent framework across scales; designing an engineerable technical implementation pathway. All theory parameters are working hypotheses requiring cross-domain empirical calibration. We actively acknowledge theoretical limitations and welcome critique, validation, and transcendence from the academic community.

Keywords: Artificial Intelligence; Cognitive Capability Building; Antifragility Validation; Human-AI Collaboration; Capability Assessment; Cognitive Resilience

第一章 引言与理论定位

1.1 核心命题：评估AI的新标准

当我们评估一个AI工具的价值时，我们真正应该关注什么？

主流的评价标准聚焦于使用时的表现：任务完成速度、输出质量、用户满意度。然而，本文提出一个根本性的反驳：AI工具的

我们将这一原则形式化为反脆弱性验证原则（Antifragility Validation Principle, AVP）：

> 以拔线测试（Unplugged Test）检验协作是否促进独立能力。核心判据：$P_2 \geqslant B_0 + \delta$。
>
> 详细定义见3.0.2节（AVP定义锚点B1）。

一句话原则：设计靠EML（摩擦+削减）| 验收靠AVP（拔线+对比）

这个判据揭示了两种根本不同的AI使用范式：

场景对比：想象一位医生、程序员或学生，在使用AI辅助工作6个月后，突然失去AI访问权限：

- 认知外骨骼模式：其独立表现显著下降（$P_2 < B_0$），甚至无法完成原本能够胜任的任务
- 认知内共生模式：其独立表现不仅保持，甚至超越原有水平（$P_2 \geqslant B_0 + \delta$），因为AI的使用过程强化了其底层能力

这不是假设场景。神经科学研究已经提供了警示信号：Dahmani等人（2020）的研究发现，习惯性使用GPS导航的人群，其空间
offloading），其负面效应已有充分的实证基础（Risko & Gilbert, 2016”; Sparrow et al., 2011）。

更令人警醒的是这种退化的隐蔽性。每个用户都感觉自己在”进步——任务完成更快，输出质量更高。但这种表面的能力提升
—不是因为AI背叛我们，而是因为我们主动放弃了独立思考的能力。

1.1.5 术语澄清与理论命名

关于”认知外骨骼与认知内共生：

本文使用”认知外骨骼指代AI使用导致人类能力退化的病理模式，“认知内共生”指代AI使用促进人类能力提升的健康模式
理论的核心目标是避免外骨骼、建立内共生。

术语层级：- AVP：评估是否达成内共生的验收标准 - EML：设计内共生系统的构造法则 - 内共生：健康的人机关系”（理论
- 外骨骼：病理的依赖状态（理论警示）

1.2 理论缺口：现有范式的共同盲点

当前的人机交互研究和实践，主要由三种范式主导，但它们都未能有效应对上述挑战：

1.2.1 工具范式（Tool Paradigm）

将AI视为被动的效率提升工具，强调人类的完全主导地位。这一范式（如传统HCI研究）

1.2.2 增强范式（Augmentation Paradigm）

如Engelbart（1962）的智力增强愿景，关注如何通过技术放大人类能力。但这
Extended Mind理论”（Clark & Chalmers, 1998）虽然承认认知的延展性，但未能区分良性延展（促进成长）与病理延展（导

1.2.3 自动化范式（Automation Paradigm）

聚焦用AI替代人类完成任务，以实现效率最大化。这一范式对人类长期认知健
& Riley, 1997）和”自动化自满”（automation complacency）研究已经警示了过度依赖自动化系统的风险，但这些洞察尚未

这三种范式的共同盲点在于：它们都将AI视为外在于人类认知系统的工具或替代品，而未能将人机关系置于一个动态、共生、

1.3 本理论的核心贡献：从诊断到解决方案

本研究旨在填补上述理论缺口，为AI时代的人机协作提供一个可验证的范式框架。理论贡献体现在四个层面：

1.3.1 建立判别标准：反脆弱性验证原则（AVP） 本研究的首要贡献是将AVP确立为评估人机交互健康性的参考标准。这不仅

AVP的可操作化包括：

- 基线测量（T_0）：用户使用AI前的独立能力
- 协作测量（T_1）：用户与AI协作时的表现
- 独立测量（T_2）：用户在拔线窗口（W=4－8周，默认6周，工作假设，需跨领域/任务校准）后的独立能力

测量要求：T_0与T 应采用等值平行测验以控制重测与熟悉化效应。

判定标准：

- $P_2 \geqslant B_0 + \delta$：成功（认知内共生）
- $P_2 \approx B_0$：中性（未造成损害但也未促进成长）
- $P_2 < B_0$：失败（认知外骨骼，造成依赖性退化）

1.3.2 确立设计原则：内共生最小法则（EML） 本研究提出内共生最小法则（Endosymbiotic Minimal Law）作为健康人机协作的判定框架：

内共生最小法则（EML）：构成”认知内共生”的设计必要条件为：

“(1)有益认知摩擦：使用户处于最优挑战区（成功率50－70%，工作假设，需跨领域/任务校准，个体自适应）

(2)系统性支持削减：AI支持强度按既定削减曲线从S4→S1→S0

二者为联合充分的设计条件，但最终仍需AVP（$P_2 \geqslant B_0 + \delta$）作为验收必要条件。

逻辑关系：条件(1)和(2)构成设计层面的联合充分条件；AVP验证是结果验收的必要条件。三者联合构成内共生的充分必要条

边界条件声明：

边界条件：本理论适用于能力增强型人机协作；补偿性外骨骼（如残障辅助、超越生理极限的设备）不适用此判据。所

1.3.3 重新定义AI角色：从工具到认知伙伴 本研究提出将AI从被动工具重塑为具有伙伴式主体性（Partnership Agency）的认知共生体。这种转变的核心不是拟人化，而是功能性的角色重构：

可操作化的三个锚点：

1. 摩擦注入：AI主动创造适度认知挑战（而非总是提供最简单路径）
2. 脚手架消退：遵循系统性支持削减曲线
3. 以AVP闭环为交互终点：协作的最终目标是用户独立能力的提升

这种”伙伴性”与传统HCI追求的无缝、零摩擦”体验形成根本性对立：一个真正的认知伙伴，不应只是顺从的助手，更应是

1.3.4 技术实现框架：分层共生架构（LSA） 本研究提出分层共生架构（Layered Symbiosis Architecture, LSA）作为技术承载。第一章仅概述核心结构，详细实现见第五章：

LSA-F（功能分层）：L1 知识整合 | L2 状态建模 | L3 摩擦校准 | L4 元认知协调

支持档位栈（S4→S1→S0）用于表达支持强度，与LSA-F为正交维度。

表1：认知外骨骼 vs 认知内共生的核心对照

维度	认知外骨骼	认知内共生
设计哲学	替代/卸载	赋能/强化
认知摩擦	最小化	优化（50－70%，工作假设）
时间性	永久依赖	临时共生
支持削减	无/固定支持	系统性递减（S4→S1→S0）
AVP结果	$P_2 \leqslant B_0$	$P_2 \geqslant B_0 + \delta$
神经效应*	能力退化趋势	能力增强趋势

*注：神经效应一栏为基于现有认知神经科学研究（如Dahmani et al., 2020”; Maguire et al., 2000）的趋势性推断，需要进一步针对性实证研究验证。

1.3.5 理论定位：规范性解决方案框架 本研究定位为一个规范性解决方案框架（normative solution framework）。我们不仅诊断问题，更明确应当如何构造人机协作以避免依赖退化，并用AVP作为参考标准验证是否达标。

与AI对齐研究的关系：

本理论与AI对齐（AI Alignment）研究是互补而非对立的关系：

- 对齐研究关注”AI的意图是否与人类价值一致
- 本理论关注”人机协作是否促进人类能力的可持续发展

我们提供的是在假设AI已对齐的前提下，如何设计健康的人机交互模式的完整方案，包括：

- 构造标准”（通过EML）
- 验收标准（通过AVP）
- 工程路径（通过LSA）

1.4 研究方法与论文结构

1.4.1 方法论定位 本研究采用理论构建与概念分析相结合的方法。本理论定位为一个可证伪的理论框架：我们提出了一套可

透明性声明：

1. 本文使用的所有量化参数（如Cohen’s d ≥ 0.3 或 ≥10%（working assumption），成功率50－70%，拔线窗口W=4－8周，默认6周）均为概念工作模型，基于认知心理学和教育测量学文献的合理推断，但需要跨领域
2. 案例选择遵循理论启发性标准，不追求统计代表性，而是用于阐释机制和边界条件。
3. 我们明确指出理论的适用边界和证伪路径（详见第六章），坚持开放科学原则。

1.4.2 论文结构 本文共六章：第二章梳理跨学科证据基础，第三章构建AVP/EML理论框架， 第四章扩展至团队和组织层面，第六章明确理论边界并提出证伪路径。 # 第二章 文献综述与理论基础

本章旨在系统梳理CET理论的多学科基础,揭示现有研究的拼图如何指向一个统一的理论框架。我们将展示:虽然认知卸载、自—这正是CET理论的核心贡献所在。

2.1 认知卸载研究:从现象描述到机制理解

2.1.1 认知卸载的基本概念

认知卸载（Cognitive Offloading）指的是个体将认知任务委托给外部系统（如工具、技术或环境）以减轻内部认知负荷的现

早期关于认知卸载的开创性研究来自Sparrow等人（2011）对Google效应”的探索。他们通过一系列实验发现,当参与者知道信

Barr等人”（2015）进一步推进了这一研究,提出智能手机被用作”认知替代品”（Cognitive Substitutes）。他们的研究表明,认知卸载不仅影响记忆,更广泛地影响思维过程本身:当面对需要深度思考的问题时,重度智能手机

2.1.2 认知卸载的领域特异性证据

空间认知领域:GPS导航的案例

GPS导航对空间认知的影响提供了清晰的神经生物学证据。Dahmani和Bohbot（2020）的研究发现,习惯性使用GPS导航的个体在

这与Maguire等人（2000）对伦敦出租车司机的研究形成鲜明对比:出租车司机通过长期空间记忆训练,其海马体后部显著增大vs GPS依赖）,导致相反的神经可塑性结果。这为CET理论关于技术关系决定认知后果的核心主张提供了神经层面的支持。

计算与数学领域

计算器的普及引发了关于心算能力的争议。教育研究显示,在计算器广泛使用后成长的学生群体,其心算流畅度和数感呈现变化

2.1.3 AI时代的认知卸载:新兴证据

近年多项研究与预印本在AI辅助写作与信息检索场景中,观测到”更频繁的技术性卸载与”更弱的独立批判性思维/结构化写作
& Zhang, 2023”; Peng et al., 2023; Dell’Acqua et al., 2023）。

虽然具体效应量因研究设计（横断 vs 纵向）、样本特征和测量方式而异,但方向总体一致:更高程度的认知卸载与更弱的独立

方法学注意事项:鉴于现有证据大多为横断或短期观察,存在自选偏差、反向因果”（能力弱者更倾向卸载？）等方法学威胁。
8周, 默认6周）的设计,这正是CET理论主张的AVP验证路径（见第3.3节）。

2.1.4 认知卸载的调节因子

认知卸载并非一律有害。研究表明,其效应受多个因素调节:

任务阶段:当任务处于新手阶段时,适度的认知卸载可以降低工作记忆负荷,促进初步理解。但若在整个学习过程中持续依赖卸

卸载类型:AI提供的是检索线索/提示（如关键词提醒）还是整段替代推理（如完整答案生成）,其对能力发展的影响截然不同

过程设计:训练方案是否包含系统性支持削减、是否要求回忆/迁移测验、是否提供延迟再测等因素,决定了卸载是成为通往独

CET的视角:这些调节因子正对应EML的核心条件”（见第3.4节）。零摩擦的生成式替代,在缺乏有益摩擦、系统性支持削减和

2.1.5 认知卸载研究的理论空白

尽管认知卸载研究取得了丰富的实证成果,但存在三个关键的理论空白:

空白1:缺乏评估标准

现有研究主要是描述性的:它们揭示了认知卸载现象的存在、测量了其程度和相关因素,但没有提供规范性的判断标准。什么程

CET的贡献:AVP原则（$P_2 \geqslant B_0 + \delta$,见第3.3节）提供了第一个可操作的、可证伪的评估标准。关键洞察在于:不能仅看P

空白2:缺乏设计指导

认知卸载研究告诉我们”什么可能是问题,但没有系统地回答”如何设计解决方案。简单的”节制使用”建议既不现实也不可
—在数字化世界中,完全避免技术使用是不可能的。

CET的贡献:EML”（内共生最小法则,见第3.4节）提供了从问题诊断到解决方案设计”的完整路径,明确了有益AI使用的必要条

空白3:缺乏跨尺度整合

现有研究主要聚焦于个体层面的认知卸载,对组织和社会层面的影响缺乏系统性探索。即使讨论社会影响,也往往停留在抽象的

CET的贡献:通过跨尺度的统一框架,将个体认知卸载、组织依赖锁定置于同一套理论原理之下,提供了从微观到宏观的连贯解释

2.2 延展心智理论的批判性回顾:从哲学隐喻到可操作标准

2.2.1 延展心智理论的核心主张

Clark和Chalmers（1998）在其开创性论文《延展心智》（The Extended Mind）中提出了一个颠覆性的主张:认知的边界不必

经典案例是”Otto的笔记本:Otto患有阿尔茨海默症,他依赖笔记本记录信息。当他查阅笔记本寻找博物馆地址时,这个过程在
Principle）:

> “如果在面对某个任务时,世界的某个部分发挥的功能,若在头脑中完成我们会毫不犹豫地认为它是认知过程的一部分,那
> —即使它在头脑之外。

这一理论引发了认知科学、哲学、人机交互领域的广泛讨论,并衍生出延展认知（Extended Cognition）、“认知整合”（Co
Integration）等相关概念。

2.2.2 延展心智理论的局限性

尽管延展心智理论具有启发性,但它在应用于AI时代的人机协作时存在两个根本性局限:

局限1:缺乏健康性判别标准

延展心智理论关注的是”什么可以算作认知”,但没有回答什么样的延展是健康的。按照对等原则,Otto的笔记本和一个完全

- Otto的笔记本:补偿性工具,帮助他维持生活自理
- 过度依赖AI:能力退化的路径,导致独立思考能力萎缩

延展心智理论没有提供区分这两者的标准,因为它是一个描述性理论而非规范性理论。

CET的补充:AVP原则为”良性延展与病理延展”提供了可证伪的分界:如果延展导致P $\geq B_0 + \delta$”（拔线后能力提升）,则
$< B_0$（拔线后能力下降）,则为病理。

局限2:忽视过程的时间性

延展心智理论关注的是某一时刻的”认知状态,但人类能力的发展是一个时间过程。一个工具在某一时刻可能是有益的延展,

例如:

- 阶段1:AI辅助写作帮助新手快速产出,此时为有益延展
- 阶段2:持续使用导致独立写作能力停滞,此时已演变为依赖

延展心智理论无法捕捉这种动态转变,因为它缺乏过程视角和能力发展维度。

CET的补充:EML的”系统性支持削减条件明确要求工具的支持强度应从S4→S1→S0,确保延展是临时的脚手架而非永久的拐杖。
—误以为自己掌握了知识，实际上只是依赖工具完成了任务”（Liao et al., 2024）。这一发现与CET理论的H2假说高度一致

2.2.3 对等原则的重构:从功能对等到过程对等

CET对延展心智理论的改造,不是否定其核心洞察,而是将其从”功能对等”升级为过程对等:

延展心智”（Clark & Chalmers）:关注”工具在某一时刻是否发挥认知功能” ↓ CET的重构:关注工具的使用过程是否促进长

这种重构的关键在于:将对等原则”的判断标准从静态功能转向动态过程。不仅要问”工具是否帮助完成任务,更要问”工具

小结:延展心智理论为CET提供了哲学基础,但CET通过引入AVP/EML,为这一宏大框架增加了可证伪的操作阈值。具体而言:延展

2.3 自动化研究与脚手架理论:从失败案例到成功路径

2.3.1 自动化悖论:永久支持的代价

自动化研究领域提供了丰富的”负面证据:过度自动化如何导致操作员技能退化。Bainbridge（1983）的开创性论文《自动化
of Automation）指出:自动化越完善,操作员在需要介入时的能力越差。

航空案例:飞行员技能退化

现代民航飞机高度自动化,飞行员在正常飞行中主要处于监督角色。但当自动系统失效时,飞行员需要手动操作的技能往往已

- 飞行员在自动驾驶时表现正常（P_1高）
- 系统失效后手动操作能力不足（P_2低）
- 长期依赖自动化与应急处理能力萎缩高度相关

这种永久性支持→能力退化”的模式在多个独立研究与事故调查中反复出现,正是CET警示的认知外骨骼模式:永久的支持导致

2.3.2 脚手架理论:渐进独立的教育智慧

与自动化悖论形成对比的是教育心理学的”脚手架理论”（Scaffolding Theory,Wood et al., 1976）。该理论强调:有效的教

脚手架理论的核心原则:

1. 最初支持:在学习者能力不足时提供密集支持
2. 渐进淡出（Fading）:随着学习者能力提升,逐步减少支持

3. 最终独立:支持完全撤出,学习者能独立完成任务

这一模式在教学实践中被广泛验证:从幼儿识字（家长逐步减少指读）到职业培训（师傅逐步放手）,都遵循这一原则（方向性

CET的整合:脚手架理论的渐进淡出正是EML的”系统性支持削减条件的理论来源。但脚手架理论缺乏定量的评估标准,CET通过

2.3.3 两种模式的对比:外骨骼 vs 脚手架

为便于理解CET如何整合自动化研究的负面教训与脚手架理论的正面经验,表2.1按照同构维度并列展示”自动化失败路径与”

表2.1:自动化与脚手架的范式对比

维度	自动化模式（失败路径）	脚手架模式（成功路径）	CET术语
支持时长	永久性	临时性	认知外骨骼 vs 认知内共生
支持强度	固定或递增	渐进递减至0	S4→S1→S0削减 vs 固定支持
设计目标	替代人类操作	促进能力发展	效率优先 vs 能力优先
失败表现	P_1高但P 低	P_1适中但P $\geq B_0+\delta$	外骨骼症状 vs AVP通过
典型案例	飞行员技能退化	学徒制培训成功	见第4.X节案例

关键洞察:自动化研究揭示了”永久支持→能力退化”的高度风险,脚手架理论指出了”渐进独立的成功路径。CET将两者整合

2.4 神经可塑性与认知训练:适度挑战的生物学基础

2.4.1 用进废退:神经可塑性的双向性

神经科学研究揭示了大脑的用进废退原则:经常使用的神经通路被强化,长期闲置的通路被削弱。这一机制为CET理论提供了生

正向证据:伦敦出租车司机

Maguire等人（2000）的经典研究发现,伦敦出租车司机经过3－4年的复杂空间记忆训练后,海马体后部灰质体积显著增大。这

负向证据:GPS依赖者

Dahmani和Bohbot（2020）的研究发现,习惯性使用GPS导航的个体在空间记忆任务中表现较差,且海马体灰质体积较小（方向性

CET的解释:这两项研究共同验证了EML的核心机制:

- 有益摩擦（独立导航训练）→ 神经强化（海马体增大）
- 零摩擦卸载（GPS依赖）→ 神经退化（海马体减小）

2.4.2 最优挑战区:适度困难的学习曲线

认知心理学的”意欲的困难（Desirable Difficulties）理论（Bjork, 1994）指出:适度的学习困难能促进长期保持和迁移。

- 间隔效应（Spaced Practice）:分散练习优于集中练习
- 测试效应（Testing Effect）:检索练习优于重复阅读
- 交错效应（Interleaving）:混合练习优于分块练习

这些效应的共同特点是:短期表现可能较差（P_1低）,但长期保持和迁移更好（P_2高）。这与CET的核心主张完全一致:不能只

最优挑战区的量化:

认知负荷理论（Cognitive Load Theory,Sweller, 1988）和教育心理学研究提示,学习任务的成功率50－
70%（工作假设，需跨领域/任务校准）**区间时,学习效果最优（工作假设，需跨领域/任务校准，基于多个独立研究的方向性

- >85%成功率:任务过于简单,接近认知卸载,无足够挑战
- 50－70%成功率:适度困难,促进深度加工和模式学习
- <30%成功率:任务过难,导致挫败和回避行为

这一区间为EML的”有益认知摩擦条件提供了定量参考”（见第3.4节详细讨论）。

2.4.3 认知训练争议与CET的澄清

认知训练领域存在关于”远迁移”效果的持续争议。一些研究报告认知训练”（如工作记忆训练）能提升其他认知能力,另一
Lervåg & Hulme, 2013”; Simons et al., 2016）。

CET对争议的重构:

1. 有摩擦是必要非充分条件:仅有认知挑战不足以保证能力迁移,需要配合系统性支持削减和AVP验证
2. 测量焦点的转移:传统训练研究往往只测量训练期间的表现提升（类似P ）,但未能系统性地评估训练停止后在新情境
3. 远迁移的困难正说明:避免出现P 高、P_2不增”的表面学习假象需要更严格的验证——这正是AVP的价值所在

设计启示:从神经科学到工程参数

本节的神经科学和认知训练证据可以转化为三个可操作的设计旋钮:

1. 摩擦旋钮:通过自适应算法将任务难度维持在成功率50–70%的最优挑战区（工作假设，需跨领域/任务校准，个体需动
2. 削减旋钮:AI支持强度从约80%线性或分段递减至0%,节律根据用户表现动态调整
3. 验证闭环:使用等值平行测验在T 和T_3（拔线窗口W=4–8周，默认6周之后）时间点测量,判据为P
   ≥ B_0 + δ （≥Cohen’s d ≥ 0.3 或 ≥10%（working assumption）,工作假设,需跨领域/任务校准）

注:以上参数均为基于文献的工作假设,需要跨任务类型、领域和人群进行实证校准验证。

2.5 理论整合:跨学科证据的收敛

本章通过四个独立研究领域的系统梳理,为CET理论构建了坚实的实证基础。这些来自认知心理学、哲学、工程心理学、教育

2.5.1 证据链的收敛性:殊途同归的科学共识

为清晰展示跨学科证据如何独立收敛于CET的核心主张,表2.2按研究领域汇总了各自的核心发现、对CET的支撑关系及证据类

表2.2:跨学科证据对CET核心主张的收敛性支持

研究领域	核心发现	支持的CET原则	证据类型
认知卸载研究(2.1节)	过度卸载与认知能力退化相关联；缺乏评估标准	• 评估标准的必要性• 病理卸载 P_2测量不可或缺	实证（关联）
延展心智理论(2.2节)	认知可以延展到工具,但缺乏健康性判别标准	• 健康延展vs病理延展的区分• 临时性与成长性的重要性	理论/概念
自动化与脚手架(2.3节)	永久支持→能力退化（负面证据）；渐进撤出→能力提升（正面证据）	• EML条件2（系统性支持撤减）• Fading的本质重要性	实证（案例+教学）
神经可塑性与训练(2.4节)	用进废退的双向性；最优挑战区（50–70%）；挑战是必要条件	• EML条件1（有益认知摩擦）• 适度挑战的神经机制	实证（神经+行为）

注:证据类型标注遵循本文透明性原则。实证证据又可细分为神经影像、行为测量、案例研究等不同强度。

三重收敛的力量:

这四个研究传统的收敛性为CET提供了罕见的理论支撑强度:

1. 负面警示的收敛:认知卸载研究（GPS用户）、自动化研究（飞行员技能退化）、神经可塑性（废退原则）——三个独立领域都揭示了”永久依赖→能力萎缩的风险。这为CET关于认知外骨骼”的批判提供了跨学科验证。
2. 正面路径的收敛:脚手架理论”（教育学）、意欲的困难（认知心理学）、神经可塑性（出租车司机研究）——三个领域都指向”适度挑战+渐进独立的成功模式。这为EML的两大条件提供了独立且相互印证的支持。
3. 评估标准的缺失:所有四个领域都缺乏一个统一的、可操作的评估框架来判断什么样的人机关系是健康的。这正是CET通

方法论意义:

这种跨学科收敛不是理论拼凑,而是独立发现的科学会师。每个领域都从自己的研究对象（记忆、哲学、航空安全、教学、大
Realism）,当多个独立研究传统指向相同的核心机制时,这大大增强了该机制真实存在的可信度。

2.5.2 CET的独特贡献:从分散洞察到统一框架

虽然现有研究提供了丰富的局部洞察,但它们之间缺乏整合,也未能转化为可操作的设计和评估标准。CET的贡献在于将这些分

空白1:缺乏健康性评估标准

现状:认知卸载研究描述了现象,延展心智理论提供了哲学视角,但都未能回答:“多少卸载是安全的?”什么时候从增强变成

CET填补:反脆弱性验证原则”（AVP）提供了第一个可操作的参考标准（见第3.3节）:

- 不看使用时表现（P_1）,而看脱离后能力（P_2）
- 明确的判定阈值（$P_2 \geqslant B_0 + \delta$）
- 可验证、可证伪的测量协议

这将”健康人机关系从抽象概念转化为可测量的科学问题。

空白2:缺乏系统设计指导

现状:脚手架理论告诉教师”应该撤出支持,自动化研究警告不要过度自动化”,但都未能提供如何设计满足这些原则的系统。

CET填补:内共生最小法则”（EML）提供了明确的设计判别标准（见第3.4节）:

- 条件1:有益认知摩擦（目标成功率50–70%，工作假设，需跨领域/任务校准，需动态调整）
- 条件2:系统性支持削减（S4→S1→S0→0曲线,自适应节律）
- 验收条件:AVP验证通过（$P_2 \geqslant B_0 + \delta$）

任何AI工具都可以用这三个条件进行检验,区分”内共生与”外骨骼不再依赖主观判断。

空白3:缺乏跨尺度整合

现状:现有研究主要聚焦个体层面。即使讨论组织或社会影响,也往往停留在抽象的文化批判,缺乏从微观到宏观的机制性连接

CET填补:通过分层共生架构（LSA）和跨尺度机制分析（详见第四章）,CET将个体认知卸载、团队协作模式、组织韧性置于统

2.5.3 承上启下:从为什么到”是什么”与如何做

第二章论证了CET的必要性:现有研究虽提供了局部洞察,但缺乏统一的评估与设计框架。第三章将转向核心机制阐释:AVP如何

第三章 CET核心理论构建

本章系统阐释CET的核心机制:AVP判据如何操作化?EML条件的内在逻辑是什么?如何通过伙伴式主体性实现二者统一?认知外骨

本章结构围绕CET的三大核心支柱展开:判别标准”（AVP)、设计原则(EML三条件)、哲学基础(伙伴式主体性),最后通过外骨骼

3.0 核心术语与锚点定义

本节集中呈现CET理论的所有核心定义和固定锚点文本。这些定义在全文中保持一字不改,后续章节引用时仅用简称或”见3.0

3.0.1 核心符号系统

表3.1: 第三章关键符号系统

符号/术语	含义	典型值/范围
B_0	用户使用AI前的独立能力基线	任务特定测量
P_1	用户与AI协作时的表现	过程指标,不参与最终判定
P_2	用户在拔线窗口后的独立表现	AVP验证的核心指标
δ	最小有意义提升阈值	工作假设:≥Cohen’s d ≥ 0.3 或 ≥10%（working assumption）；（需跨领域/任务校准）
W	拔线窗口时长	工作假设:4–8周（默认6周，需跨领域/任务校准）

符号/术语	含义	典型值/范围
S” (t)	t时刻的AI支持强度	0(完全独立)到1(完全依赖)
S0	初始支持强度	工作假设:约0.8
T	总消退时长	任务与个体特定,需校准
成功率目标	有益认知摩擦的量化目标	50 - 70%（工作假设，需跨领域/任务校准）（工作假设，需跨领域/

注:所有量化参数均为概念工作模型,需通过跨领域实证研究校准。

3.0.2 反脆弱性验证原则(AVP)【锚点B1】

反脆弱性验证原则（AVP）：以拔线测试检验协作是否促进独立能力。

判据：$P_2 \geqslant B_0 + \delta$：Cohen’s d ≥ 0.3 或 ≥10%（working assumption），需跨领域/任务校准）。P_1（协作表现）

3.0.3 内共生最小法则(EML)【锚点B2】

内共生最小法则（EML）：构成认知内共生的设计必要条件为：

(1)有益认知摩擦：使用户处于最优挑战区（群体级工作假设成功率 50 - 70%，个体自适应）；

(2)系统性支持削减：AI支持强度按既定削减曲线从 S4→S1→S0。

二者为联合充分的设计条件，但最终仍需AVP（$P_2 \geqslant B_0 + \delta$）作为验收必要条件。

3.0.4 LSA功能分层【锚点B3】

LSA-F（功能分层）：L1 知识整合 | L2 状态建模 | L3 摩擦校准 | L4 元认知协调。

支持档位栈（S4→S1→S0）用于表达支持强度，与 LSA-F 为正交维度。

硬约束：L1-L4（功能层）与S4-S1（支持强度）互不表示，不得在同一公式中互代或串级使用。L层表示功能维度，S档

3.0.5 最优挑战区【锚点B4】

最优挑战区：为促成长期保持与迁移，系统应将任务难度/提示强度自适应调至成功率 50 - 70%（工作假设，随任务与个体校准）；>85% 近似卸载、<30% 易致挫败。

3.0.6 边界条件【锚点B5】

边界条件：CET适用于能力增强型人机协作；补偿性外骨骼（如残障辅助、超越生理极限的设备）不适用此判据。所有

公平性原则：AVP以等效努力为准则——对确需辅助工具的个体（如屏幕阅读器使用者），调整任务形式而不降低挑战

公平性三句式（逐字版）：

“调整任务格式而不降低挑战强度；通过相对改进而非绝对水平评估；（如涉及可访问性）保持挑战预算守恒。”

Adjust task format without lowering challenge intensity; assess by relative improvement rather than absolute level; (if accessibility applies) challenge budget conservation.

3.1 AVP原则的深度阐释:从抽象判据到可操作测量

反脆弱性验证原则(AVP)是CET理论的北极星”——它不仅定义了什么是”好的人机协作,更提供了一个可证伪、可操作的测量

3.1.1 理论基础:从Taleb的反脆弱性到认知增强评估

反脆弱性概念的启发

Nassim Nicholas Taleb”(2012)提出的”反脆弱性”(Antifragility)概念为AVP提供了哲学基础。反脆弱性不仅仅是”不被
(韧性),而是在压力下变得更强——这正是健康人机协作的本质特征。

Taleb将系统分为三类:

- 脆弱性(Fragile):压力导致损害(如易碎物品)
- 韧性(Robust):压力下保持不变(如橡胶球)
- 反脆弱性(Antifragile):适度压力促进成长(如肌肉、免疫系统)

类比到人机交互:

- 认知外骨骼:脆弱性系统——移除AI后能力下降($P_2 < B_0$)
- 中性工具:韧性系统——移除工具后能力不变($P_2 \approx B_0$)
- 认知内共生:反脆弱性系统——经历协作后独立能力提升($P_2 \geq B_0 + \delta$)

核心洞察:拔线”不是惩罚,而是验证机制——真正的增强应该让用户在离开AI后变得更强,就像肌肉在举重后变得更强壮。

拔线窗口的哲学意义:能力主权

拔线窗口W=4–8周（默认6周，工作假设，需跨领域/任务校准）不仅是技术参数,更代表了能力主权的时间尺度:

- 太短”(<2周):可能只测到短期记忆残留,无法验证真正的能力内化
- 太长(>12周):环境因素混杂增加,难以归因于AI协作
- 4–8周:经验性平衡点,既允许能力稳定,又可控混杂变量

这一窗口期本质上在问:用户是否真正拥有了这项能力,还是仅仅借用了AI的能力?

3.1.2 AVP的三级分级体系:从基础到卓越

AVP不是简单的通过/不通过判断,而是一个分级评估体系。我们提出三个递进的验证层次:

表3.2: AVP三级分级体系

等级	判据	含义	典型应用场景
AVP-Basic	$P_2 \geq B_0$	基础反脆弱性:至少维持原有能力	补偿性学习技能维护
AVP-Retention	$P_2 \geq B_0 + \delta$	成长性反脆弱性:能力有意义提升	能力建构专业训练
AVP-Transfer	满足Retention+跨任务迁移	卓越反脆弱性:深度掌握+跨域能力（质量标志，非强制门槛）	专家培养创新能力

注1:δ：Cohen’s d ≥ 0.3 或 ≥10%（working assumption）),需跨领域/任务校准。迁移的操作化定义见副判据表3.3。

注2（Goodhart防护）:此分级仅用于质量分层，最终判定仍以AVP主判据（$P_2 \geq B_0 + \delta$）为准。禁止将分级阈值作为优化目

三级体系的实践意义:

1. Basic级:最低安全线——任何AI工具至少应满足无害标准
2. Retention级:CET的核心目标——真正的能力增强
3. Transfer级:理想目标——培养可迁移的深度能力”（通过副判据验证）

副判据的补充维度:

除了核心的P $\geq B_0+\delta$判据,我们还提出三个副判据以评估能力质量”(非必要条件,但用于分级):

表3.3: AVP副判据系统

副判据	测量内容	权重建议
迁移性(Transfer)	能力在新情境的泛化程度	高(35%)
保持性(Retention)	能力在更长时间窗口的稳定性	中(30%)
独立性(Independence)	完成任务时对AI的依赖程度	中(35%)

注1：权重为工作假设（迁移35%/保持30%/独立35%），用于质量分层，不作为治理或认证硬阈值。权重可根据具体目标调整，

注2：这些副判据不应被用作治理认证的硬性标准，仅作质量分级参考，避免Goodhart风险（见表3.2注2）。

副判据的操作化测量:

副判据	测量方法	示例	数据来源
迁移性(Transfer)	在相关但不同的新情境测试能力	编程:Python→JavaScript	T_3后增加迁移任务测试商业领域数据分析评估→
保持性(Retention)	延长观察窗口至T (T₃后6个月)	评估P 相对P 的保持率;间隔测试能力曲线(工作假设)	斜率衰减曲线分析
独立性(Independence)	拔线窗口内的行为监测	违规尝试访问AI的频率;完成任务时自我报告求助次数	行为日志分析;独立工作时长占比

详细测量协议与评分标准见附录A § A.3-A.5。

3.1.3 AVP测量协议:可复现的操作指南

为确保AVP的科学性和可复现性,我们提供标准测量协议:

测量时间线:

- T_0:协作前,测量B (独立基线)
- 协作期(长度可变,典型为4 - 12周)
- T_2:协作期结束
- 拔线窗口W（4 - 8周，默认6周，无AI使用）
- T_3:拔线窗口结束后,测量P (独立表现)

测量工具要求:

- 等值性:T_0和T 使用等值平行测验(避免学习效应)
- 盲评:评分者不知道被测者的组别(实验/对照)
- 环境一致性:两次测量的情境条件尽可能相同

表3.4: AVP测量的关键效度威胁与缓解策略

威胁类型	具体表现	缓解策略	残留风险
等值性不足	T_0和T 测验难度不一致	使用IRT等值”;专家盲审	中:难以完全等值
环境差异	测试情境变化影响表现	标准化程序;控制组对照	低:可有效控制
评分偏差	评分者主观性或期望影响	双盲评分;多评分者校准	低:可有效控制
练习污染	测验内容泄露或过度熟悉	等值平行卷;题库轮换;冷却期	中:题库资源约束
学习效应	重测导致熟悉化	等值测验;足够长的W	中:无法完全消除
延迟巩固	能力需更长时间稳定	延长W或增加T 测量点	中:资源约束
流失偏差	能力弱者更易退出	意向性分析;激励机制	高:难以完全避免

注:本表提供方法学透明性,不意味AVP不可用,而是明确其适用边界与改进方向。

3.1.4 跨领域校准:参数的情境敏感性

AVP的核心参数”（δ、W、测量工具)并非一成不变,需要根据任务类型、领域特征、目标人群进行校准:

δ阈值的领域差异(工作假设):

- 认知技能(如编程、写作):d≥0.3 SD或≥10%相对提升
- 运动技能(如打字速度):可能需要更大阈值(d≥0.5)
- 创造性任务(如艺术创作):可能需要定性评估而非单一δ

W窗口的任务差异(工作假设):

- 快速技能(如数学计算):可能4周足够
- 复杂技能(如第二语言):可能需要8–12周
- 专业能力(如外科手术):可能需要数月甚至年

校准流程建议:

1. 先导研究:小样本测试,确定初步参数
2. 敏感性分析:测试参数变化对结果的影响
3. 迭代优化:根据反馈调整参数
4. 开放报告:公开参数选择理由和校准数据

重要边界:即使经过校准,AVP也有其不适用场景:

- 高风险任务(拔线可能造成不可接受后果)
- 纯工具性使用(无能力建构目标)
- 补偿性外骨骼(目标是补偿而非增强)

3.2 有益认知摩擦:挑战与支持的最优平衡

EML的第一个设计必要条件是有益认知摩擦(Beneficial Cognitive Friction)。本节将阐释为什么”适度困难是必要的,以及

3.2.1 为什么需要认知摩擦:从神经科学到学习理论

神经可塑性的”用进废退原则

正如第2.4节综述,大脑遵循”用进废退原则:经常使用的神经通路被强化,长期闲置的通路被削弱。如果AI完全替代了某项认
(方向性证据,见Dahmani & Bohbot, 2020)。

“意欲的困难理论

Bjork” (1994)的意欲的困难(Desirable Difficulties)理论指出:适度的学习困难能促进长期保持和迁移(工作假设,有实证

- 间隔练习:分散学习优于集中学习
- 交错练习:混合练习优于分块练习
- 生成效应:主动回忆优于被动复习

零摩擦的陷阱

当AI提供”零摩擦”体验时” (如直接给出答案、完全自动化流程),用户短期表现可能很好(P_1高),但长期能力往往停滞或退

3.2.2 最优挑战区:50–70%成功率的量化目标

Vygotsky的最近发展区” (ZPD)的启发

Vygotsky(1978)提出的最近发展区概念指出:学习发生在用户当前能力”与借助支持可达成的能力之间的区域。CET将其量化

目标成功率50–70%“（工作假设，需跨领域/任务校准）**（工作假设，需跨领域/任务校准，群体级，个体需动态校

为什么是这个区间?” (工作假设,需跨任务验证):

- <30%:过度挫败,导致回避和放弃
- 30–50%:可学习但效率不高,动机易受影响
- 50–70%:最优平衡——足够挑战但不致挫败
- 70–85%:舒适但挑战不足,能力提升缓慢
- >85%:接近认知卸载,几乎无能力增长

个体差异的重要性

50 - 70%是群体级目标,个体的最优区间需要根据以下因素自适应调整:

- 基础能力:新手可能需要60 - 75%,专家需要40 - 60%
- 动机水平:高动机者可承受更低成功率
- 任务焦虑:高焦虑者需要略高成功率以维持信心
- 实时状态:疲劳/压力下应临时提高成功率

3.2.3 摩擦的动态调节:自适应算法的概念模型

如何在实践中维持最优挑战区?我们提出三种概念性策略(详细算法见第5章LSA架构):

策略1:滚动窗口法

- 监测最近N次交互的成功率(如N=10)
- 如果SR(成功率)>75%,增加任务难度或减少提示
- 如果SR<45%,降低难度或增加支持
- 保持在目标区间50 - 70%

策略2:IRT-启发式调整

- 借鉴项目反应理论(IRT)的自适应测验思想
- 根据用户表现动态调整任务难度参数
- 目标是让每次交互对能力评估和培养都最有信息量

策略3:分段梯度法

- 将学习过程分为阶段(如入门/进阶/精通)
- 每阶段设定不同的成功率目标
- 入门阶段可略高(60 - 75%),精通阶段可略低(40 - 60%)

过度挑战的安全护栏

为避免摩擦过度,系统应设置自动降级机制:

- 触发条件:连续3次SR<30%,或用户明确表示挫败
- 响应:临时降低摩擦,并向用户解释原因
- 恢复:在用户状态改善后逐步恢复目标摩擦水平

3.3 系统性支持削减:从脚手架到能力独立

EML的第二个设计必要条件是系统性支持削减(Systematic Support Reduction)。本节阐释为什么支持必须削减,以及如何科学

3.3.1 为什么必须削减:从脚手架理论到依赖预防

教育学的脚手架隐喻

Wood等人(1976)提出的脚手架理论强调:有效的教学支持应该是临时的、渐进撤出的。脚手架的目的是帮助建筑施工,但建筑完
—同样,AI支持的目的是帮助能力建构,但能力形成后应该逐步退出。

自动化悖论的警示

正如第2.3节所述,航空领域的自动化悖论揭示:永久性的自动化支持会导致操作员技能退化。这一教训同样适用于AI:如果支持

削减的双重目标

系统性支持削减不仅是为了验证能力,更是为了促进能力:

- 短期:通过逐步撤出支持,迫使用户内化技能
- 长期:培养用户对自身能力的信心和元认知

3.3.2 支持削减曲线:从S4到S1到0的演化路径

支持档位栈(S4→S1→S0)的定义

我们使用支持档位栈表达支持强度(注意:这与LSA-F的L1-L4功能层是正交维度,见3.0.4节):

- S4(初始强度):最大支持(如提供完整答案、详细步骤)
- S3(中度支持):提供提示和部分解决方案
- S2(轻度支持):仅在用户请求时给予最小提示
- S1(最小支持):仅提供验证/反馈,不直接解决问题
- S0(完全拔线):无AI支持,用于AVP测试

支持强度的形式化表达:

S(t) = S0 · f(t/T)

□□:
- S(t): t□□□□□□□
- S0: □□□□□□(□□□□□0.8,□□□)
- T: □□□□□(□□□□□□□)
- f(·): □□□□(□□□□□□□□□□)

三种典型削减曲线(工作假设,需A/B测试验证):

表3.5: 支持削减曲线类型与适用场景

曲线类型	数学形式	特点	适用场景
线性削减	S(t) = S0(1 - t/T)	匀速递减,平稳过渡	结构化任务,稳定学习曲线
指数削减	S(t) = S0 • e^(-kt)	初期快速削减,后期缓慢	快速技能,避免过度依赖
阶梯削减	S(t) = S0 • 4(1-t/T) /4	分阶段突变,适应期明确	分级训练,明确里程碑
自适应	S(t) = f(性能,元认知,动机)	根据用户状态动态调整	个性化学习,复杂任务
混合模式	分段组合上述曲线	兼具渐进性与阶段性	长期训练,多模块任务

注:所有曲线参数(S0、T、k)均为校准参数,需根据任务复杂度、用户能力、学习目标进行域内标定。

削减速率的校准原则(工作假设):

- 过快:用户跟不上,挫败感增加,$P_2 < B_0$
- 过慢:依赖锁定,用户”不愿意独立,$P_2 \approx B_0$
- 恰当:用户感觉有挑战但可达成,$P_2 \geq B_0 + \delta$

3.3.3 回退与加速机制:保底安全线与提速路径

回退机制” (Safety Net):防止挫败

当检测到用户在削减后表现急剧下降时,系统应临时回退到更高支持档位:

- 触发条件:连续3次失败率>70%,或用户明确求助
- 回退策略:S(t) → S(t-Δt),重新提供支持
- 恢复路径:待用户状态稳定后,重新启动削减

概念规则：回退不是失败信号，而是自适应系统的正常反应。保底支持强度S_min（工作假设约0.2，需校准）确保用户始终

团队/组织场景下的回退与保底机制，见第4章T-AVP/O-AVP的安全护栏（4.1.2节、4.2.3节）。

S_min的实现由第5章LSA架构的L2摩擦校准引擎与L4元认知协调层承载，通过自适应算法动态维持（见5.4节、5.5节）。

加速机制(Fast Track):奖励卓越

如果用户表现超预期,可以加速削减:

- 触发条件:连续成功率>85%,且元认知监控良好
- 加速策略:跳过中间档位,直接进入下一阶段
- 验证:通过微型拔线测试确认能力确实内化

动态平衡的哲学

回退与加速机制体现了CET的核心理念:适应性而非固定性。系统应该像优秀的教练,根据学习者的实时状态调整挑战强度。

3.4 伙伴式主体性:AI角色的哲学重构

前两节聚焦”做什么” (AVP判据)和”怎么做” (EML条件),本节转向为什么这样做有效的哲学基础。我们提出伙伴式主体性like Agency)作为认知内共生的本体论支柱。

3.4.1 从工具到伙伴:AI角色的三种范式

传统范式1:纯工具(Pure Tool)

- 假设:AI是被动执行者,完全服从人类意图
- 问题:忽视了AI对用户认知模式的反向塑造
- 典型:传统软件工具(如计算器、文本编辑器)

传统范式2:自主代理(Autonomous Agent)

- 假设:AI是独立决策者,追求自己的目标
- 问题:引发控制权冲突,可能威胁人类主导地位
- 典型:强化学习代理、自主系统

CET提出的新范式:伙伴式主体性(Partner-like Agency)

- 定义:AI具有受限的主体性——它可以主动行动,但始终服务于促进用户能力成长的目标
- 类比:优秀的教练、导师、陪练——他们有主动性,但目标是成就学生而非替代学生
- 核心特征

:

- 主动干预:在用户过度依赖时主动增加摩擦
- 目标一致:AI的利益函数”是用户的长期能力,而非短期效率
- 权力让渡:随着用户能力提升,AI主动退出

3.4.2 伙伴式主体性的四个维度

维度1:自适应支持” (Adaptive Support)

- 能力:AI能感知用户的能力水平、学习状态、动机水平
- 行为:根据感知结果动态调整支持强度(见3.2 - 3.3节)
- 目标:维持用户在最优挑战区

维度2:元认知催化(Metacognitive Catalysis)

- 能力:AI能引导用户反思自己的学习过程
- 行为:提出引导性问题而非直接答案,如”你为什么选择这个方法?这个方案的局限是什么?”
- 目标:培养用户的元认知能力和自我监控

维度3:渐进撤出” (Progressive Withdrawal)

- 能力:AI能判断用户何时准备好承担更多独立任务
- 行为:按既定曲线(见3.3.2节)削减支持,并在必要时回退
- 目标:最终让用户完全独立,AI退为”备用资源

维度4:能力验证” (Capability Verification)

- 能力:AI能设计并执行拔线测试
- 行为:定期触发AVP验证(见3.0.1节),评估P 表现
- 目标:确认内共生关系的健康性,而非仅追求P 效率

3.4.3 伙伴关系的伦理边界

主体性的限度:三个不应该

1. 不应该操纵:AI不应利用心理学技巧诱导用户过度使用
2. 不应该评判:AI不应对用户的价值观、生活选择做道德评判
3. 不应该替代人际:AI不应成为用户唯一的社交/情感支持

知情同意原则

用户应该清楚理解:

- AI会主动增加任务难度(摩擦设计)
- AI会渐进削减支持(可能短期不适)
- AI会定期进行拔线测试(可能感觉”被监视)

退出权

用户应始终有权:

- 暂停或退出内共生模式
- 切换回”纯工具”模式” (如果不追求能力建构)
- 调整摩擦/削减参数(在合理范围内)

3.4.4 伙伴式主体性与其他AI范式的对比

表3.6: 不同AI范式的对比

维度	纯工具	自主代理	伙伴式主体性
主动性	无	高	中(受限)
目标函数	任务完成	自身目标	用户能力成长
权力关系	人类完全支配	潜在冲突	协作但人类最终决策
适用场景	简单工具任务	自主执行任务	能力建构任务
长期影响	中性或负面	不确定	正面(如果AVP通过)
典型例子	计算器	自动驾驶(L5级)	CET的LSA系统

关键洞察:

伙伴式主体性不是在工具和”代理之间的妥协,而是一个第三条道路——它承认AI应该有主动性,但这种主动性的唯一目的是

3.5 认知外骨骼的病理学:失败模式的系统分析

前面章节聚焦”如何建立内共生,本节转向反面:如何识别和避免外骨骼依赖。通过分析失败模式,我们可以更清晰地理解内共

3.5.1 外骨骼的核心特征:从健康到病理的转折点

定义回顾:

认知外骨骼是指用户对AI的病理性依赖,表现为:

- $P_2 < B_0$:拔线后能力不及初始基线
- 能力萎缩:原有技能因长期不用而退化
- 依赖锁定:心理上和认知上都无法脱离AI

外骨骼的阶段演化:

表3.7: 从健康使用到外骨骼依赖的阶段

阶段	P_2与B 关系	依赖程度	可逆性	典型特征
健康使用	$P_2 \geqslant B_0 + \delta$	低	N/A	内共生,能力提升
中性使用	$P_2 \approx B_0$	中	高	工具化,能力未变
轻度依赖	$B_0 - \delta < P_2 < B_0$	中高	中	轻微退化,警戒信号
中度依赖	$B_0 - 2\delta < P_2 < B_0 - \delta$	高	低	明显退化,需干预
重度依赖”(外骨骼)	$P_2 < B_0 - 2\delta$	极高	极低	能力萎缩,依赖锁定

注:δ:Cohen’s d $\geqslant$ 0.3 或 $\geqslant$10%(working assumption))。此分级为概念框架,实际判定需结合用户主观体验与行为模式

从中性到病理的临界点”

不是所有AI使用都会变成外骨骼。关键在于是否跨越了两个临界点:

1. 能力临界点:P_2首次低于B
2. 心理临界点:用户开始回避独立完成任务的情境

3.5.2 外骨骼的10个预警信号

如何及早识别外骨骼依赖?我们提出10个可观察的预警信号”(工作假设,需临床验证):

行为信号(6个):

1. 使用频率激增:日使用时间从30分钟→3小时
2. 独立尝试时间缩短:遇到问题<30秒就求助AI
3. 任务范围扩大:从困难任务扩展到简单任务也依赖
4. 回避独立情境:主动避免不能用AI的场合
5. 完成速度悖论:有AI时快,无AI时极慢
6. 错误率上升:独立完成任务的错误率持续增加

认知信号(4个):

1. 元认知缺失:无法准确评估自己的独立能力
2. 知识碎片化:知道结果但不知道”为什么
3. 迁移困难:无法将AI辅助下的经验迁移到新情境
4. 拔线焦虑:对”无AI测试产生强烈恐惧或抗拒

这些信号可以被量化为外骨骼风险评分”(工作假设,需验证):

- 0–2个信号:低风险
- 3–5个信号:中风险,需要注意
- 6–8个信号:高风险,需要干预
- 9–10个信号:危险,需要立即干预

3.5.3 分级干预策略:从预防到治疗

表3.8: 外骨骼干预策略矩阵

风险级别	干预类型	具体措施	目标
低风险(0–2信号)	预防	定期AVP轻量测试认知健康教育	维持健康状态
中风险(3–5信号)	早期干预	强制拔线窗口增加认知摩擦	逆转依赖趋势
高风险(6–8信号)	积极干预	削减支持强度重新训练基础能力	恢复独立能力
危险(9–10信号)	重建	完全拔线+系统性重训心理辅导	打破依赖锁定

注:这些策略为概念框架,实施需考虑伦理边界和用户意愿。

3.6 跨尺度概览:从个体到团队到组织到社会

前五节聚焦个体层面的人机交互。本节简要探讨CET理论如何延伸到团队和组织层面,为第四章的跨尺度机制分析做准备。

3.6.1 团队层面:协作中的能力分布

当多个个体与AI协作时,出现新的动力学:

团队退化的两种模式:

1. 全员依赖:所有成员都依赖AI→集体能力萎缩
2. 能力分化:部分成员独立,部分依赖→团队脆弱性

团队层面的AVP变体:

T-AVP: □□□□□□□□□□□□
- □□:□□□□□□AI□□□□□□□□□
- □□:□□P□ ≥ □□B□ + δ_team(□□□□,□□□)

设计启示(详见第4.2节):

- 避免”AI作为团队唯一专家”的配置
- 促进知识在成员间的分布而非集中在AI
- 定期团队拔线演练

3.6.2 组织层面:认知基础设施与系统韧性

组织层面的风险:

- 关键能力空心化:某些技能在组织中完全消失
- 知识传承断裂:新员工从AI而非老员工学习
- 系统性脆弱:AI故障→组织业务停摆

O-AVP:组织反脆弱性验证:

O-AVP = □□□"48□□AI□□□□"□□□□□□□□□
- □□□□□□□□
- □□□□□□□□□
- □□□□

组织设计建议” (详见第4.3节):

- 建立认知储备(cognitive reserve)机制
- 定期无AI值班制度”
- 关键岗位的独立能力认证

3.6.3 社会层面:认知公地悲剧

- 个体理性” (使用AI)vs 集体理性(保持独立能力)的冲突
- 代际能力鸿沟:T_0代 vs T_1代 vs T_2代
- 文明级的反脆弱性评估(S-AVP)

详见第四章4.4节对社会层面的深入分析

3.6.4 跨尺度的同构性

CET核心原理(AVP/EML/伙伴式主体性)在不同尺度上具有同构性:

表3.9: 跨尺度AVP体系

尺度	AVP变体	EML应用	主要风险
个体	I-AVP	学习工具设计	能力退化
团队	T-AVP	协作流程设计	能力分化
组织	O-AVP	组织韧性建设	系统脆弱
社会	S-AVP	政策与标准	代际鸿沟

第四章 跨尺度机制分析：从个体到组织到社会

前三章聚焦个体层面的人机交互：AVP如何验证个人能力提升（见3.0.1节）？EML如何指导单用户的AI工具设计（见3.2-3.3节）？伙伴式主体性如何在一对一关系中实现（见3.0.4节）？然而，AI的影响不止于个体——当多人协作、组织运作、社

本章逻辑：我们将沿着”个体→团队→组织→社会的尺度阶梯，展示：(1)每个尺度的特有机制；(2)AVP原则在不同尺度的变

核心术语与符号表（第四章）

表4.1: 第四章关键符号与概念

符号/术语	含义	适用尺度
I-AVP	Individual AVP (个体反脆弱性验证)	个体
T-AVP	Team AVP (团队反脆弱性验证)	团队
O-AVP	Organizational AVP (组织反脆弱性验证)	组织
S-AVP	Societal AVP (社会认知资本验证)	社会
认知资本	群体的独立认知能力储备	组织/社会
认知公地	共享的认知能力基础设施	社会
级联脆弱	低层脆弱性向高层传播	跨尺度
涌现依赖	个体依赖累积为系统性依赖	跨尺度

注：所有量化参数均为校准参数（工作假设），需根据行业特性、组织规模、风险级别进行域内标定和灵敏度分析。

4.1 个体→团队：能力分布与协作模式的重构

4.1.1 从I-AVP到T-AVP：团队层面的反脆弱性

问题的涌现：当一个团队中的多个成员都使用AI工具时，即使每个个体的I-AVP都通过($P_2 \geq B_0 + \delta$，见3.0.1节)，团队作为整体是否就是健康的？答案是：不一定。

T-AVP的定义：

□□□"□□□□□□□□□□ ≥ □□ + δ_team

□□□□
T-AVP□□: P_2_team ≥ B_0_team + δ_team

□□□
- B_0_team: □□□□AI□□□□□□□□
- P_2_team: □□□□□□□□W=4-8□□□□6□□□□□□□□□□□/□□□□□□□□□□□□
- δ_team: □□□□□□□□□□□□□
 □□□□□≥0.3 SD□□□□□□□□□□□□Cohen's d□□□□□□□□□□□□□□□□□□

□□□□□
□□□□ ≠ □□□□□□
□□□□□□□□□□□□□□□□□□□□□□□

参数校准说明：δ_team：Cohen’s d ≥ 0.3 或 ≥10%（working assumption），需通过灵敏度分析确定不同团队类型（研发

核心洞察：团队I-AVP不是T-AVP的充分条件

即使团队中每个成员都独立通过I-AVP（P_2_individual ≥ B_0_individual + δ），团队层面仍可能失败（P_2_team < B_0_team + δ_team）。这是因为团队能力涉及：

- 协作模式：知识如何在成员间流动
- 角色互补：成员能力如何形成冗余与备份
- 集体记忆：团队共享的隐性知识

为什么I-AVP成功不保证T-AVP成功？三种失败模式

模式1：能力极化” (Capability Polarization)

- 现象：团队中部分成员高度依赖AI，部分成员完全独立
- 机制

：

- 依赖成员：P_2_individual < B_0 (认知外骨骼，见3.0.5节)
- 独立成员：P_2_individual ≥ B_0 + δ (认知内共生)
- 但团队整体：P_2_team < B_0_team
- 原因：依赖成员成为单点故障，拔线后团队无法完成关键任务
- 案例：编程团队中，新手完全依赖Copilot，老手独立编程。拔线后，新手无法完成代码审查，团队开发速度骤降

模式2：隐性知识流失(Tacit Knowledge Loss)

- 现象：团队成员通过AI完成任务，但知识未在人际间传播
- 机制

：

- 传统：老手→新手的知识传承(师徒制、结对编程)
- AI时代：每个人都问AI，人际知识流动↓
- 结果：团队的集体智慧未随时间积累
- 后果：拔线后，团队缺乏共享的问题解决策略
- 案例：客服团队使用AI回答系统。每个客服独立查询AI，不再分享”棘手案例的解决经验。拔线后，团队无法应对复杂

模式3：角色固化与冗余度丧失” (Role Rigidity)

- 现象：AI工具让成员过度专业化，丧失互相补位”能力
- 机制

：

- AI前：成员需要学习彼此的技能以应对突发情况
- AI后：每个人依赖AI完成自己的模块，不再学习他人技能
- 结果：团队的韧性” (resilience)下降
- 后果：关键成员缺席时，团队瘫痪
- 案例：医疗团队中，护士依赖AI辅助诊断，不再学习医生的诊断思路。医生临时缺席时，护士无法做初步判断

4.1.2 T-AVP的测量协议

□□1: □□□□(T_0_team)
- □□□□□□□□□□□□□□□□□□AI□
- □□□□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□

□□2: AI□□□(T_1 → T_2□□□8-12□□□□□□□□□□)
- □□□□□□AI□□
- □□□□□□□□□□□□□□□□□□□□□□□□□□

□□3: □□□□□W=4-8□□□□6□□□□□□□□□□□/□□□□□
- □□□□AI□□
- □□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□
- □□□□□□**□□□□□□□□□□**
- □□**□□□□□□□□□□**□□□□□
- □□□□□/□□□□□□□□□□□□□□**□□□□□□□□□**□□□□□□

□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□**□□□□□□□□□□**□□□□□□□□□□□□□□□□□□3.0.6□□□□□□□□□□□□□□AVP□

□□4: □□□□□□(T_3_team)
- □□□□□□□T□□□□□□□□□□□□□AI□
- □□□P_2_team vs B_0_team
- □□□P_2_team ≥ B_0_team + δ_team ?

方法学提示：团队数据存在组内相关（ICC）与样本聚类，功效计算与显著性检验应采用集群稳健方法（如多层线性模型、广

表4.2: T-AVP测量的效度威胁与缓解

威胁类型	具体表现	缓解策略	残留风险
任务等值性	T_0和T 项目难度不同	专家盲审”;多维度评分	中
成员流失	部分成员在测试期离职	意向性分析;备用成员	高
霍桑效应	团队知道被观察而改变行为	长期跟踪;自然观察	中
协作污染	拔线期间成员私下使用AI	诚信协议;技术屏蔽	中
学习效应	团队因重复任务而熟练	等值项目池;间隔足够长	低
角色分工变化	T_0和T 的角色分配不同	固定角色或随机分配	低

注：本表提供方法学透明性，承认测量局限，不意味T-AVP不可用。

分级说明：团队层当前仅采用T-AVP-Basic**（拔线无退化：P_2_team≥B_0_team）与T-AVP-Retention（拔线净提升：P_2_tea

说明：团队层的迁移涉及集体知识向新情境的泛化，其测量需要跨组织对照研究，当前理论基础不足以给出可操作定义。

4.1.3 团队层面的设计启示

设计原则1：促进人际知识流动

- 定期”无AI讨论会”：强制团队成员分享问题解决策略
- 结对工作：新老成员配对，促进隐性知识传递
- 团队代码审查：不仅审查代码，更审查思路

设计原则2：建立协作摩擦机制

类比于个体层面的有益认知摩擦”（见3.0.2节），团队层面也需要协作摩擦：

- 轮换角色：定期让成员尝试他人的角色
- 跨模块任务：避免过度专业化
- 集体挑战项目：需要多人协作才能完成的任务

设计原则3：制度化能力建构

案例：软件团队的新人基础能力建构

某软件公司发现新入职的AI原生代工程师（2000年后出生）过度依赖GitHub Copilot，独立编程能力弱。团队实施三项制度：

1. 新人独立项目：入职第一个月，禁用AI辅助，完成基础功能模块
2. 周期性”无AI日：每周五为纯手工日，全员禁用代码生成工具
3. 结对评审制度：新人代码由老员工人工审查，而非仅靠AI检查

结果：6个月后T-AVP测试显示，新人独立解决bug的能力显著提升（$\delta\_team \geq$ Cohen’s d $\geq$ 0.3 或 $\geq$10%（working assumption），$p<0.01$）。

制度化落地：将”新人独立项目 + 周期性无AI日 + 结对评审固化为团队入职训练三件套，写入员工手册与培训流程，以保证AVP的底座稳定。不应仅作为”建议，而应作为团队制度强制执行。

4.2 团队→组织：认知基础设施的系统性风险

4.2.1 组织层面的新涌现：制度化依赖

当AI使用从个别团队扩展到整个组织时，出现制度性依赖——即使个别团队有独立能力，组织作为整体仍可能脆弱。

组织层面的三大风险：

风险1：关键能力的空心化” (Hollowing Out)

- 现象：某些技能在组织中完全消失
- 机制

：

- 老员工依赖AI→技能退化
- 新员工从AI学习→从未掌握
- 结果：无人具备该技能
- 案例：某银行的风险评估团队全员使用AI模型。5年后，当模型出现系统性偏差时，无人能手工进行风险评估（该技能

风险2：知识传承的断裂” (Transmission Breakdown)

- 现象：组织的隐性知识无法传递给下一代
- 机制

：

- 传统：“师傅带徒弟的学徒制
- AI时代：新人直接问AI，跳过老员工
- 结果：老员工的经验无法传承
- 案例：某制造企业的老工程师掌握设备调试的手感。由于新人都依赖AI诊断工具，这些隐性知识未能传承。老工程师退

风险3：认知基础设施的单点故障” (Infrastructure Fragility)

- 现象：组织过度依赖AI基础设施，AI宕机→业务瘫痪
- 机制

：

- 所有关键流程都嵌入AI
- 没有”备用方案”或”手工模式
- AI故障时，组织完全无法运作
- 案例：2024年某在线教育平台AI推荐系统宕机48小时，导致课程分配、学习路径规划全部停摆，影响20万学生

4.2.2 O-AVP：组织反脆弱性验证

类比I-AVP和T-AVP，我们提出O-AVP”（组织层面的反脆弱性验证）：

O-AVP = □□□AI□□□□"□□□□□□□□□□

□□□□
O-AVP = w_BCI × BCI + w_ICR × ICR□□□□□□□□□□≥0.70□□□≥0.85□□□□□□□□□□□/□□□□□

□□□
- BCI " (Business Continuity Index): □□□□□□□
 □□□□□□□□48h□AI□□□□□□□□□□□□□□□

- ICR (Independent Capability Ratio): □□□□□□□
 □□□□□□□AI□□□□□□□□□□□□□

- □□□w_BCI=0.4, w_ICR=0.6□□□□□□□□□□□□□

O-AVP双阈值模型（工作假设）：

为了更精细地管理组织韧性，O-AVP采用双阈值设计：

- 告警阈值 0.70：触发风险排查与回退机制的下限
- O-AVP < 0.70：进入红色预警区间，组织应启动应急响应
- 建议行动：暂停新AI系统引入、启动独立能力盘点、制定回退计划
- 目标阈值 0.85：质量分层与优秀实践识别的基准
- O-AVP ≥ 0.85：达到健康标准，组织具备良好的认知韧性
- 0.70 ≤ O-AVP < 0.85：处于”黄色预警区间，需要持续改进

阈值使用场景： - 告警阈值” (0.70)用于风险管理：识别需要干预的脆弱组织 - 目标阈值(0.85)用于质量分层：标识最佳实

口径声明：双阈值数值为工作假设，需通过跨组织实证研究校准。不同行业（如金融vs研发）可能需要不同的阈值设定。详细

注:仅用于质量分层,不得作为KPI”;最终判定以AVP主判据” (见3.0.2节)为准。

窗口说明：48小时为工作假设，用于在不过度依赖短期补丁（<24h）与掩盖真实能力赤字（>72h）之间取得平衡；可按行业

- 关键基础设施（金融/医疗）：可能需要24h严格测试
- 一般企业应用：48h为推荐值
- 研发/创新类场景：可放宽至72h

O-AVP的测量协议：48小时宕机演练

□□1： □□□□(O□)
- □□□□□□□□AI□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□2： □□□□
- □□□□□AI□□□□□□□48□□□□□□□□□□□
- □□□□□□□□□□"□□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□3: □□□□□□□□" (P_2_org)
- BCI□□□(□□□□□/□□□□□□)
- ICR□□□(□□□□□□□□□□/□□□□)
- O-AVP□□□O-AVP□□□≥0.70□□□≥0.85□□□□□□□□□□/□□□□□

量化口径：异质指标（如BCI与ICR）先做标准化到[0,1]区间（工作假设），再进行加权汇总；权重为校准参数（当前推荐BC
+ ICR×0.6），需做灵敏度分析并避免单指标主导（Goodhart防护）。

标准化方法示例（非强制）：

- BCI：(实际业务连续性得分 - 最坏情况) / (最好情况 - 最坏情况)
- ICR：(无AI独立完成率 - 基线最低值) / (基线最高值 - 基线最低值)

表4.3: O-AVP测量的效度威胁与缓解

威胁类型	具体表现	缓解策略	残留风险
演练真实性	员工知道是演练而不认真执行	突击演练”;真实激励	中
业务影响	演练影响实际业务	选择低峰期;沙盒环境	低
数据完整性	关键指标缺失或不可比	事先定义KPI;标准化流程	中
外部依赖混杂	AI宕机同时其他系统也故障	控制变量;单一故障模拟	中
学习效应	多次演练后团队过度熟练	变换场景;间隔足够长	低
成本约束	频繁演练代价高	年度1－2次;局部演练	高

注：O-AVP测量成本高于I-AVP和T-AVP，需平衡频率与资源投入。

4.2.3 组织设计的三个关键机制

机制1：认知储备” (Cognitive Reserve)制度

类比金融领域的资本储备要求，组织应建立认知能力储备：

- 关键岗位独立能力认证：定期测试员工的”无AI”能力
- 技能冗余配置：确保关键技能由≥2人掌握
- 认知储备比例：规定至少X%员工保持独立能力”（X为行业特定参数）

案例：金融机构的风险管理储备

某投资银行规定：风险评估团队中，至少40%成员必须通过无AI模型风险评估认证”（O-AVP的部门级应用）。即使AI模型可

机制2：定期”无AI值班制度

- 轮换制度：每周/月有特定团队或岗位进入无AI模式”
- 知识更新：确保”手工流程保持更新，而非过时文档
- 文化建设：将独立能力视为职业荣誉而非负担

机制3：知识传承的制度化保障

- 导师制强制执行：新员工必须有人类导师”（不能仅靠AI onboarding）
- 隐性知识萃取项目：系统性记录老员工的经验（但不完全依赖AI转录）
- 跨代工作坊：定期组织老中青三代员工的知识交流

公平性原则（组织层）：对确需辅助工具的岗位（如残障员工使用辅助技术），调整任务形式而不降低挑战强度；评估以相

示例：视障程序员使用屏幕阅读器，评估其算法设计能力而非打字速度；调整任务为口述代码或使用语音输入，但算法复杂

4.3 组织→社会：认知公地悲剧

4.3.1 认知公地的概念：集体行动困境

“公地悲剧”的经典模型

Garrett Hardin” (1968)提出的公地悲剧描述了这样一个困境：

- 公共草地（公地）对所有牧民开放
- 每个牧民的理性选择：多放羊以获得更多收益
- 集体结果：草地过度放牧，最终荒芜，所有人受损

认知能力作为”认知公地

类比到AI时代的人类认知能力：

- 个体理性：使用AI工具，提高即时效率（如用AI写代码、用GPS导航）
- 集体后果：如果所有人都依赖AI，社会整体的独立认知能力下降
- 代际不可逆性：一旦某代人失去某项能力，下一代更难重建

三个关键特征：

1. 外部性：个体使用AI的成本（能力退化）部分由社会承担
2. 时间滞后：个人即时获益，社会代价需10 – 20年显现
3. 不可逆性：一旦某项技能在社会中消失，重建极为困难

4.3.2 代际能力鸿沟：三代人的分化

三代人的定义：

- T_0代（1980 – 2000年生）：在AI普及前完成教育，拥有完整的独立能力基线
- T_1代（2000 – 2015年生）：青少年时期接触AI，部分能力AI化
- T_2代（2015年后生）：AI原生代，从小在AI环境中成长

预测的能力差异（工作假设，需15 – 20年纵向数据验证）：

□□□□ T_0□ T_1□ T_2□
□□□□□□ □ □ □□□□□□□/AI□
□□□□□□ □ □ □□□□□GPS/AI□
□□□□□ □ □ ?□□□□□□□
□□□□□□ □ ? ?□□□□□□
□□□□□ □ ? ?□□□□□□

?□□□□□□□□

临界问题：T_2代在”拔线条件下，是否还能达到T 代的能力水平？

4.3.3 S-AVP：社会认知资本的验证”（概念框架）

与I/T/O-AVP不同，S-AVP无法通过拔线测试”实施——我们不能让整个社会停用AI数周。因此，S-AVP是一个代理指标体系：

表4.4: S-AVP的五个代理指标（概念框架，需跨学科研究确定）

代理指标	测量内容	数据来源	理想状态	预警阈值
关键技能分布	重要技能的人口覆盖率	职业统计、教育数据	>70%人口保持基础技能	<50%
代际能力比	T_2代vs T_0代的能力测试得分	标准化测试	≥0.9	<0.7
知识传承完整性	师徒制、学徒制的保留率	行业调查	传统传承机制健在	大量行业完全AI化
应急响应能力	AI中断事件的社会恢复速度	自然实验	<24h恢复正常	>72h
文化认知价值	社会对”独立能力的重视度	文化调查	独立能力被视为美德	被视为”落后

代理指标	测量内容	数据来源	理想状态	预警阈值

注：这些指标均为工作假设，需要跨学科”（社会学、经济学、教育学）的合作研究来确定具体测量方法和阈值。S-AVP更多是一个长期监测框架而非短期验证工具。

S-AVP的性质声明：

- 非短期可测：需要15－20年纵向数据
- 代理指标集：而非单一判据
- 预警系统：而非精确测量
- 研究议程：而非成熟工具

抽样与推断：建议采用队列/滚动横断结合的抽样策略（如每5年追踪同一人群，同时补充新样本）；跨层推断应警惕生态谬

4.3.4 认知公地的潜在干预”（方向性讨论）

个体层面：

1. AVP认证：建立独立能力认证体系（类似职业资格）
2. AI使用教育：从小培养”健康使用AI”的意识

组织层面”（已在4.2节讨论）： 3. O-AVP演练制度 4. 认知储备投资

社会层面（概念性建议，详见第六章）：

1. 认知营养标签制度

：

- 类比食品营养标签
- AI工具标注：是否支持能力建构？AVP验证结果？
- 帮助用户/组织做知情选择

2. 关键能力的社会储备

：

- 类比粮食储备、石油储备
- 确保关键行业保持一定比例的无AI”从业者
- 作为社会的”认知冗余

3. 代际传承的制度保障

：

- 保护学徒制、师徒制等人际传承模式
- 在AI时代重新重视”人教人

4. S-AVP监测体系

：

- 建立长期跨学科研究项目
- 类似”气候变化监测，需要持续数据收集
- 每5年发布社会认知资本报告”

4.4 跨尺度共性机制：CET的尺度不变性

4.4.1 AVP原则的尺度不变性

核心洞察：AVP的逻辑在所有尺度保持一致

□□□□□
[□□]□□□□□□□□□□□□ ≥ □□ + δ

□□□□
□□" (I-AVP): P_2(□□) ≥ B_0 + δ □□3.0.1□□
□□(T-AVP): P_2(□□) ≥ B_0_team + δ_team
□□(O-AVP): BCI×0.4 + ICR×0.6□□□□□□□□□□≥0.70□□□≥0.85□□□□□□□□□□/□□□□□
□□(S-AVP): [□□□□□□□] □□□□□□

为什么尺度不变？因为底层原理相同：

1. 反脆弱性的本质

：

- 个体：在压力下成长
- 组织：在危机中建立韧性
- 社会：在挑战中进化
- 共同点：临时性压力→能力提升

2. 依赖锁定的共性

：

- 个体：技能退化（见3.0.5节外骨骼模式）
- 团队：协作能力丧失
- 组织：制度性脆弱
- 社会：集体认知资本流失
- 共同点：永久支持→能力萎缩

3. 验证逻辑的一致性

：

- 都需要”拔线测试”（或代理测量）
- 都关注”独立能力而非协作效率
- 都以”基线+增量为标准

4.4.2 级联脆弱：从微观到宏观的传播

级联机制：低尺度的脆弱性如何向高层传播？

□□□□ → □□□□ → □□□□ → □□□□

□□□□□□□
1. □□□□I-AVP□□
 ↓
2. □□□□□□□□□□□□□□□□□□
 ↓
3. □□□□□□□□□□O-AVP□□<0.70"□□□□□□
 ↓
4. □□□□□□O-AVP□□
 ↓
5. □□S-AVP□□□□□□
 ↓
6. □□□□□□□□□

图4.1: 级联脆弱传播路径（概念图）

□□□ [I-AVP□□□30%]
 ↓ □□

□□□ [T-AVP□□□50%] ← □□□□□
↓ □□□
□□□ [O-AVP=0.65<0.70□□□□□□] ← □□□□□
↓ □□
□□□ [S-AVP□□□□] ← □□□□

□□□□□□□□□□

放大机制：为什么高尺度的脆弱性更严重？

1. 涌现的非线性

：

- 10%个体依赖 ≠ 10%组织风险
- 可能放大为30 - 50%风险”（因网络效应）

2. 修复的时间滞后

：

- 个体：数月可恢复
- 组织：数年才能重建
- 社会：可能需要一代人

3. 路径依赖的强化

：

- 低尺度：可逆(个人可重新训练)
- 高尺度：路径依赖强、恢复代价指数级上升（整代人能力缺失时，缺乏教师和榜样）

4.4.3 设计启示：多尺度协同的必要性

单一尺度干预的局限：

- 只改个体：组织惯性会拉回
- 只改组织：社会环境不支持
- 需要多尺度协同

图4.2: 跨尺度干预协同框架（概念图）

□□□ S-AVP ┐
│ □□□□
□□□ O-AVP │ □□□□
│ □□□□
□□□ T-AVP │ □□□□
│ □□□□
□□□ I-AVP ┘ □□□□

↑ □□□□□□
↓ □□□□□□□

表4.5: 多尺度协同干预矩阵

尺度	设计目标	关键机制	测量指标
个体	I-AVP≥B_0+ δ	有益摩擦+系统性支持削减（见3.2 - 3.3节）	P_2测试
团队	T-AVP≥B_{0_team}+ δ	协作摩擦+集体拔线+人际知识流动	团队任务表现
组织	O-	制度化演练+知识传承+认知储备	48h宕机测试
	AVP（告警≥0.70，目标≥0.85，工作假设，需跨领域/任务校准）		

尺度	设计目标	关键机制	测量指标
社会	S-AVP代理指标健康	政策引导+标准制定+文化建设	代理指标集

注:仅用于质量分层,不得作为KPI”;最终判定以AVP主判据”（见3.0.2节)为准。

注2（δ_team：Cohen’s d ≥ 0.3 或 ≥10%（working assumption），工作假设），并需随领域/任务进行校准与敏感性分析
AVP的Cohen’s d ≥ 0.3 或 ≥10%（working assumption）口径一致。

协同要点：

1. 自下而上：个体能力是基础
2. 自上而下：组织制度创造环境
3. 横向联动：行业标准、社会规范

4.5 小结：理论的系统性与实践的紧迫性

4.5.1 本章核心贡献

1. 理论扩展

:

- CET从个体理论扩展为跨尺度框架
- AVP原则具有尺度不变性
- 提出T-AVP、O-AVP、S-AVP的概念体系

2. 机制揭示

:

- 能力极化、知识流失、角色固化(团队)
- 制度性依赖、认知基础设施退化(组织)
- 认知公地悲剧、代际鸿沟(社会)

3. 实践指引

:

- 提供可操作的测量协议(T-AVP、O-AVP)
- 识别关键风险信号
- 提出多尺度协同方向

4.5.2 与前后章的连接

第四章将个体层CET原则扩展至团队、组织、社会四个尺度,揭示了AVP的尺度不变性与跨层级的认知退化机制。第五章将探讨
/O-AVP的可复现实验协议,明确证伪路径与研究议程(见6.2 - 6.3节)。

4.5.3 理论的紧迫性

本章揭示的不是远期风险，而是当下现实：

- 团队层面：已有组织报告”新人无法独立工作
- 组织层面：AI宕机事件暴露脆弱性”（如本章教育平台案例）
- 社会层面：代际能力差异开始显现(虽尚未临界)

窗口期：当前(2025)到2035年是关键的10年窗口（工作假设）

- T_0代仍在工作，知识传承尚可挽救
- AI普及率约30 - 50%，未到不可逆点
- 制度化干预(如O-AVP演练)仍可建立

过了这个窗口：

- T_0代退休，某些隐性知识永久丢失
- AI依赖制度化，路径依赖难以逆转
- 社会规范改变，“独立能力可能被视为”落后

CET的使命：在窗口期内，提供理论基础和实践指引，避免路径依赖强、恢复代价指数级上升的认知公地悲剧。

第五章 分层共生架构（LSA）：CET的工程化实现

前四章建立了CET的理论基础：AVP作为评估标准（第三章），EML作为设计原则（第三章），伙伴式主体性作为AI角色定位（

本章提出分层共生架构（Layered Symbiosis Architecture, LSA）——一个将CET理论转化为可工程化系统的设计框架。LSA

本章目标：

1. 提出LSA的四层架构模型
2. 展示如何在系统层面实现EML条件
3. 设计多尺度AVP监测的遥测管线
4. 讨论技术可行性、工程挑战与未来方向

本章逻辑：我们将自底向上构建LSA，从基础AI能力（L1）到摩擦与削减机制（L2），再到监测与反馈（L3），最后到编排与

核心术语与符号表（第五章）

表5.1: 第五章关键术语与概念

术语/符号	含义	架构层级
LSA	Layered Symbiosis Architecture ” (分层共生架构)	全局
L1-基础能力层	提供原始AI能力（推理、生成、检索）	底层
L2-摩擦与削减层	实现有益摩擦和支持削减	中层
L3-监测与反馈层	AVP遥测、能力评估、预警	中层
L4-编排与治理层	多尺度策略编排、伦理约束	顶层
CFE	Cognitive Friction Engine (认知摩擦引擎)	L2核心
SGS	Support Graduation Scheduler (支持削减调度器)	L2核心
AVP-TM	AVP Telemetry Module (AVP遥测模块)	L3核心
MSO	Multi-Scale Orchestrator (多尺度编排器)	L4核心
能力向量 C	用户当前能力状态的表征	L3
摩擦参数 F	控制任务难度的参数集	L2
削减曲线 S(t)	支持强度随时间的函数	L2

注：所有参数和阈值均为设计空间的概念占位符，实际系统需根据领域特性、用户群体、任务类型进行校准和A/B测试。

表5.2: LSA层间接口契约（概念规范）

接口	输入	输出	时序约束	关键属性
L1→L2	原始请求+上下文	完整AI输出	目标<1s（工作假设，随场景校准)	质量最大化
L2→L3	用户行为+任务完成度	能力评估事件	异步（后台）	准确性、隐私保护
L3→L4	C(t)向量+预警信号	策略调整建议	准实时（分钟级）	可解释性、可审计性
L4→L2	摩擦/削减参数F(t)、S(t)	调制后的AI输出	目标<100ms（工作假设）	透明性、用户控制

接口规范说明：此表提供概念层次的接口定义，具体实现需考虑技术栈特性（如REST API、消息队列、流式处理）。时序约束

5.1 LSA总览：为什么需要分层架构？

5.1.1 传统AI系统的设计困境

当前AI辅助工具的典型架构：

```
□□□□ → AI□□ → □□□□ → □□
  ↑
 □□□□
```

这种架构的根本问题：

1. 无差别输出

：AI对所有用户提供相同强度的帮助

- 新手和专家得到同样详细的答案
- 无法根据用户能力动态调整
- 结果：专家被过度帮助，新手形成依赖

2. 无能力感知

：系统不知道用户的真实能力水平

- 无法判断用户是在学习还是在卸载
- 无法预测长期能力影响
- 结果：盲目优化短期效率，忽视长期能力

3. 无反馈闭环

：缺乏对用户能力变化的监测

- 不知道使用后用户是成长还是退化
- 无法验证AVP
- 结果：无法区分内共生和外骨骼

根本症结：传统架构只关注任务完成，不关注能力建构。

5.1.2 LSA的设计哲学：能力建构优先

核心理念转变：

```
□□□□□
□□□□ = □□□□□ + □□□□

LSA□□□
□□□□ = □□□□□ + □□□□ + □□□□ (AVP□□)
  ↑
 □□□□□□
```

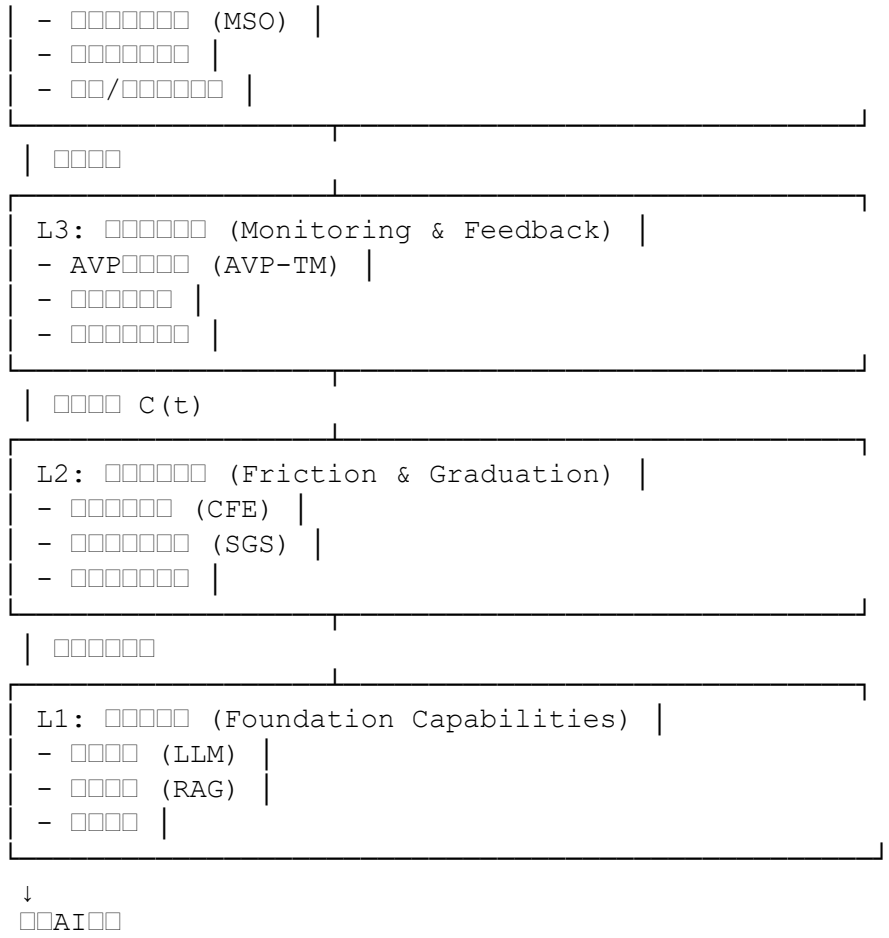
LSA的四个设计支柱：

1. 能力感知（L3）：系统持续评估用户能力
2. 动态适应（L2）：根据能力调整支持策略
3. 透明反馈（L3）：让用户看到自己的成长
4. 多尺度编排（L4）：支持个体、团队、组织

5.1.3 LSA四层架构总览

图5.1: LSA分层架构总览

```
┌──────────────────────────────────────────────────────┐
│ L4: □□□□□□ (Orchestration & Governance)              │
```



各层职责简述：

L1（基础层）：提供原始AI能力

- 不关心用户能力建构
- 只关心高质量输出
- 可以是任何主流AI模型（GPT、Claude、Gemini等）

L2（摩擦与削减层）：实现EML前两个条件

- 有益摩擦：动态调整任务难度
- 支持削减：渐进减少帮助强度
- 核心：从”全力帮助”到”适度挑战

L3”（监测层）：实现AVP验证

- 持续评估用户能力
- 检测依赖锁定风险
- 触发干预机制

L4（编排层）：多尺度协同与治理

- 个体策略→团队策略→组织策略
- 伦理约束（公平性、透明度）
- 全局优化目标

层间关系：

- 自底向上：能力流动（原始能力→调制后的支持→能力评估→策略决策）

- 自顶向下：策略流动（组织目标→个体目标→调制参数→AI行为）

硬约束：L1 - L4（LSA-F功能分层）为功能维度（知识整合 | 状态建模 | 摩擦校准 | 元认知协调）；S4→S1→S0（支持档位栈

5.2 L1层：基础AI能力层

5.2.1 职责与边界

L1层的唯一职责：提供高质量的原始AI能力

□□□□□□□ + □□□
□□□□□□□/□□/□□/□□
□□□□□□□□□□□□□□□

L1不关心：

- 用户是新手还是专家
- 输出是否导致依赖
- 用户能力是否提升

为什么分离L1？

1. 技术中立性：LSA可以用任何AI模型实现
2. 职责单一：L1专注于生成质量，不负责能力建构
3. 可替换性：随着AI技术进步，L1可以升级而不影响上层

5.2.2 L1的典型能力模块

表5.3: L1层的标准能力模块（概念层次）

能力模块	功能描述	示例技术方向
推理引擎	逻辑推理、问题分解、规划	LLM推理技术
生成引擎	文本/代码/图像生成	生成式AI
检索增强	知识检索、事实查询	RAG架构
工具调用	外部API、计算工具	函数调用能力
上下文管理	对话历史、会话状态	记忆系统

关键设计原则：L1应该是尽力而为”（best-effort）的AI，不主动限制能力输出。限制和调制由L2层负责。

5.2.3 L1与传统AI系统的区别

□□AI□□□
L1 = □□□□□□□□□ → AI → □□□

LSA□□□
L1 = AI□□□□□□□L2□□□□□
L2 = □□□□□□□□□□□□□□□□□□
L3 = □□□□□□□□□□□□
L4 = □□□□□□□□□□□□

类比：

- L1 = 发动机：提供动力
- L2 = 变速箱：调节输出
- L3 = 仪表盘：监测状态
- L4 = 驾驶员：决定方向

5.3 L2层：摩擦与削减层

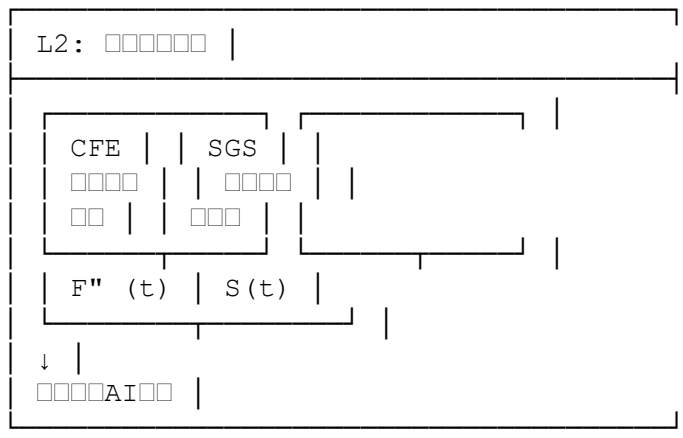
5.3.1 核心挑战：如何实现适度帮助？

问题陈述：给定一个用户请求和L1的原始输出，如何调制输出使其满足EML的前两个条件？

EML条件回顾（见第三章3.2－3.3节）：

1. 有益认知摩擦：任务成功率50－70%（工作假设，需跨领域/任务校准，群体级，个体自适应），动态调整
2. 系统性支持削减：AI支持强度按S4→S1→S0

L2层的双引擎架构：



5.3.2 认知摩擦引擎（CFE）

设计目标：让任务”不太容易也不太难，维持用户在最优挑战区”（见第三章3.2节）。

摩擦注入的四种策略（概念层次）：

策略1：完整性摩擦（Completeness Friction）

□□□□□□□□
"□□bug□□□□□□□□□□□23□□□□□□arr[i+1]□
□□□□□i□□□□length-1□□□□□□□□□[□□□□]

□□□□□
"□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□"

策略2：抽象度摩擦”（Abstraction Friction）

□□□□□□□□
□□□□□□□□□□□□□O" (n log n)□
□□□1.□□ 2.□□□□ 3.□□□□□□[□□□□]

□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□"

策略3：脚手架削减（Scaffolding Reduction）

□□□□□
"□□1□[□□] → □□2□[□□] → □□3□[□□] → □□□□

□□□□□

□□□□□□□□ → □□□□□ → □□□□"

□□□□□
"□□□□□□□□

策略4：自适应难度”（Adaptive Difficulty）

- 根据用户历史表现动态调整摩擦强度
- 成功率高→增加摩擦
- 成功率低→降低摩擦
- 目标：维持在50 - 70%区间（工作假设，需校准）

CFE的核心算法（概念框架）：

```
# □□□□□□□□□□□□□
def adjust_friction" (user, task, history):
 "
 □□□□□□□□□□□□□□

 □□□□□□□□□□□□□□□□□□A/B□□□□□
 "
 # □□□□□□□□□□"□□□□□□□□□10□□
 recent_tasks = history[-10:] # □□□□□□□□
 success_rate = calculate_success_rate" (recent_tasks)

 # □□□□□50-70%□□□□□□□□□□
 target_min, target_max = 0.5, 0.7 # □□□□□□

 # □□□□□□□□
 if success_rate > target_max:
 friction_level += adjustment_step # □□□□
 elif success_rate < target_min:
 friction_level -= adjustment_step # □□□□
 else:
 pass # □□□□□□□

 # □□□□
 friction_level = clip(friction_level, min_friction, max_friction)

 return friction_level
```

实现说明：实际系统应采用更稳健的自适应算法（如贝叶斯优化、强化学习），并结合任务类型、用户特征进行多维度调整。

5.3.3 支持削减调度器（SGS）

设计目标：实现第三章3.3节提出的系统性支持削减，让用户逐渐独立。

术语说明：支持强度档位以S4(高支持)→S3→S2→S1(低支持)→S0(完全拔线)表达（口径统一，见3.0.3节）。

削减曲线的三种模式（概念对比）：

表5.4: 支持削减曲线类型（概念层次）

曲线类型	特点	适用场景示例	参数调整建议
线性削减	匀速递减，平稳过渡	结构化任务	根据任务周期调整斜率
指数削减	初期快速，后期缓慢	快速技能学习	k值需A/B测试优化
阶梯削减	分阶段突变，里程碑清晰	分级训练体系	台阶数量和高度需领域专家定
自适应	根据用户状态动态调整	个性化学习，复杂任务	需要L3层能力监测支持

曲线类型	特点	适用场景示例	参数调整建议

注：曲线选择应基于任务特性和用户群体，通过对照实验验证效果。所有参数均为设计空间的起点，非最优解。

SGS的核心机制：回退与保底支持（承接第三章3.3.3节）

保底支持强度S_min的实现：

第三章3.3.3节提出了保底支持强度S_min（工作假设约0.2，需校准）的概念，确保用户在削减过程中始终有最低限度的导航

```
# □□□□□□
def support_graduation_with_safety_net(t, user_state):
 "
 □□□□□□□□□□□□□

 □□□□□□3.3.3□□□□□□□□□□□
 "
 # □□□□□□□"□□□□□□□□□□□
 S_base = S_0 * exp" (-lambda * t) # S_0, lambda□□□

 # □□□□□□□□□S_min
 S_min = 0.2 # □□□□□□□□□□□□
 S_current = max(S_base, S_min)

 # □□□□□□□□□3.3.3□□
 if detect_consecutive_failures(user_state, threshold=3):
 # □□□□□□□□□□
 S_current = min(S_current + backtrack_step, S_0)
 log_intervention("safety_net_triggered)

 return S_current

def detect_consecutive_failures" (user_state, threshold):
 "□□□□□□□□□□□□□"
 recent_failures = user_state.recent_tasks[-threshold:]
 failure_rate = sum" ([t.failed for t in recent_failures]) / threshold
 return failure_rate > 0.7 # □□□□□
```

工程实现说明：S_min的维持由L2层的摩擦校准引擎持续监测，L4层的元认知协调模块提供全局策略指导（见5.4节、5.5节）

削减速率的控制理论约束（概念框架）：

```
# □□□□□□□□□□□□□
def safe_graduation_rate(S, dS_dt, user_state):
 "
 □□□□□□□□□□□□□□□□

 □□□□□□□"□□□□□□
 "
 # □□1□□□□□□"□□□□□□
 max_rate = 0.1 # □□□□□□□□□□10%□□□□□□
 dS_dt_safe = clip" (dS_dt, -max_rate, 0) # □□□□□□□□

 # □□2□□□□□□□□□□□□□
 if abs(dS_dt_safe - previous_rate) > hysteresis_threshold:
 # □□□□□□□□
```

```
    dS_dt_safe = smooth_transition(previous_rate, dS_dt_safe)

    # □□3□□□□□□□□□L3□□□□
    if user_state.frustration_score > high_threshold:
    dS_dt_safe = 0 # □□□□

    return dS_dt_safe
```

例外条款：上述速率约束仅适用于常规削减阶段；当触发安全网回退（见前述support_graduation_with_safety_n

这些约束确保削减过程稳定可控，避免用户因过快削减而产生挫败感。具体参数需要通过纵向追踪研究和用户反馈迭代优化。

5.3.4 L2层的工程挑战

挑战1：实时性能

问题：L2的计算能否在用户可接受延迟内完成？

概念解决方向：

- 预计算策略：提前准备多难度版本的输出
- 异步更新：背景更新C(t)和策略参数
- 边缘计算：能力向量本地存储和计算

挑战2：模型对齐

问题：如何让L1的AI理解”适度帮助的语义？

概念解决方向：

- 对齐训练：通过RLHF让AI理解摩擦指令
- 提示工程：设计元提示模板
- 微调方向：在LSA场景上进行领域适配

挑战3：评估的客观性

问题：如何客观测量摩擦是否”适度？

概念验证方法：

- 多任务交叉验证
- 标准化任务库建设
- A/B测试对比不同摩擦策略

5.4 L3层：监测与反馈层

5.4.1 核心职责：能力追踪与AVP验证

L3层的使命：回答”用户能力是否在成长这个核心问题

三大功能模块：

1. AVP-TM”（AVP遥测模块）：采集能力相关数据
2. 能力建模引擎：维护用户能力向量C”（t）
3. 预警与干预系统：检测依赖锁定风险

5.4.2 AVP遥测模块（AVP-TM）

承接第四章4.5.2节：多尺度AVP监测的遥测管线

第四章4.5.2节指出，LSA架构需要支持I-AVP、T-AVP、O-AVP三个尺度的AVP监测。L3层的AVP-TM模块提供统一的遥测管线：

表5.5： 多尺度AVP监测遥测管线（承接4.5.2节）

监测尺度	数据采集源	遥测事件类型	聚合层级	参见章节
个体层	用户任务日志、P_2测试数据	任务完成、能力评估、拔线	实时	3.1节
团队层	协作日志、知识流动记录	团队任务、集体拔线	每日汇总	4.1节
组织层	48h演练数据、BCI/ICR指标	中断演练、恢复曲线	事件触发	4.2节

AVP-TM的核心遥测事件集（概念层次）：

表5.6: L3层核心遥测事件（设计空间示例）

事件类型	触发时机	记录内容示例	用途
任务开始	用户启动新任务	任务类型、当前S(t)、F(t)	上下文追踪
任务完成	用户提交结果	完成质量、用时、支持调用	能力评估
摩擦调整	CFE改变F(t)	旧F→新F、触发原因	自适应分析
削减事件	SGS改变S(t)	旧S→新S、削减阶段	进度追踪
拔线测试	定期或触发式	P_2分数、对比B	AVP判定（见3.0.1节）
预警触发	检测到风险信号	预警级别、触发指标	干预决策
团队协作	团队任务开始/结束	参与者、角色、知识流动	T-AVP评估（4.1节）
组织演练	48h中断演练	BCI/ICR分数、恢复时长	O-AVP评估（4.2节）

数据采集原则：最小化采集、本地优先、目的限定（见5.4.5节隐私保护）。

遥测管线的技术架构（概念层次）：

□□□□ → □□□□ → □□□□□ → □□□□

↓

□□□□

↓

□□□□ → L4□

监测原则：AVP-TM仅提供数据管线与可视化，不做自动判定。最终的AVP通过/失败判定由人工结合系统建议进行（特别是团

5.4.3 能力建模引擎

能力向量C(t)的概念定义：

```
# □□□□□□□□□□□
class AbilityVector:
 "
 □□□□□□□□□□

 □□□□□□□□□□□□□□□□□□□□□□□□□□□
 "
 def __init__" (self):
 # □□□□□□□□□□□□□□□□
 self.problem_decomposition = 0.5 # □□□□□□
 self.implementation_skill = 0.6 # □□□□
 self.debugging_ability = 0.4 # □□□□
 self.documentation = 0.3 # □□□□□□
 self.meta_cognition = 0.5 # □□□□□□□
 # ... □□□□

 def update(self, task_performance):
 "□□□□□□□□□□□□□"
```

```
# □□□□□□□□□□□□□□□□
pass
```

说明：这是概念示意，实际系统需要认知科学、教育心理学领域的专业知识来定义能力维度。维度过多会增加测量噪声，过

能力评估的数据来源：

1. 直接测量

：P_2拔线测试分数”（参考标准，见3.0.1节）

- 采用等值平行卷（IRT校准）以确保测试难度一致性
- 设置最短冷却期（工作假设：48–72小时，需校准）以抑制练习效应污染
- 方法学威胁与缓解策略参见3.1.3表（跨章一致）

2. 间接推断：日常任务表现（连续监测）
3. 自我报告：用户自评（辅助参考）
4. 同行评议：团队成员互评（团队层，见4.1节）

能力建模的核心功能：

1. 驱动L2：根据C”(t)调整摩擦和削减
2. 驱动L4：为编排层提供决策依据
3. 用户可见：生成能力成长报告

5.4.4 预警与干预系统

目标：在用户陷入依赖锁定前触发干预

三级预警机制（概念框架）：

表5.7: L3层的三级预警系统

预警级别	触发条件示例	系统响应建议	用户体验建议
绿色	AVP健康，C(t)↑	继续当前策略	正常使用，鼓励性反馈
黄色	C(t)停滞或轻微↓	增加摩擦，减缓削减	提示”能力未提升
红色	衰减指标超阈值	暂停削减，强制独立周	警告”可能形成依赖
黑色	AVP测试失败	触发L4干预，重置路径	强制”能力重建模式

触发阈值说明：具体阈值”（如衰减指标）需要通过纵向研究和用户反馈数据校准，不同领域和任务类型可能需要不同阈值。

5.4.5 L3层的隐私保护设计

问题：持续监测用户能力涉及隐私关切

LSA的隐私保护分层策略（概念框架）：

表5.8: 隐私保护技术方案

保护层级	技术方向	应用场景	数据保留建议
采集最小化	只记录元数据，不记录内容	所有遥测事件	–
本地优先	C”(t)、F、S(t)存本地	个体能力向量	永久
匿名化	Hash处理用户ID	聚合统计、研究	有限期
联邦学习	本地训练，只传梯度	能力模型优化	不传原始数据
差分隐私	添加噪声保护个体	组织级报告	–
用户控制	可查看/导出/删除数据	个人数据管理	用户决定

设计原则：

- 数据最小化

：只收集能力评估必需数据

- 记录：任务类型、S(t)、F、完成质量、用时
- 不记录：任务具体内容、用户输入/输出原文、个人敏感信息
- 本地优先

：能力向量C(t)存储在用户设备

- 只有聚合统计发送到服务器
- 用户可随时删除
- 支持离线模式（L2/L3本地运行）
- 目的限定

：数据只用于能力评估

- 不用于用户画像、推荐、营销
- 不与第三方共享
- 开放审计（用户可查看数据使用日志）
- 提供可审计日志导出与留痕，便于用户与第三方进行独立合规审查

5.5 L4层：编排与治理层

5.5.1 核心职责：多尺度策略与伦理治理

L4层的使命：协调个体、团队、组织层面的目标，并确保系统符合伦理约束

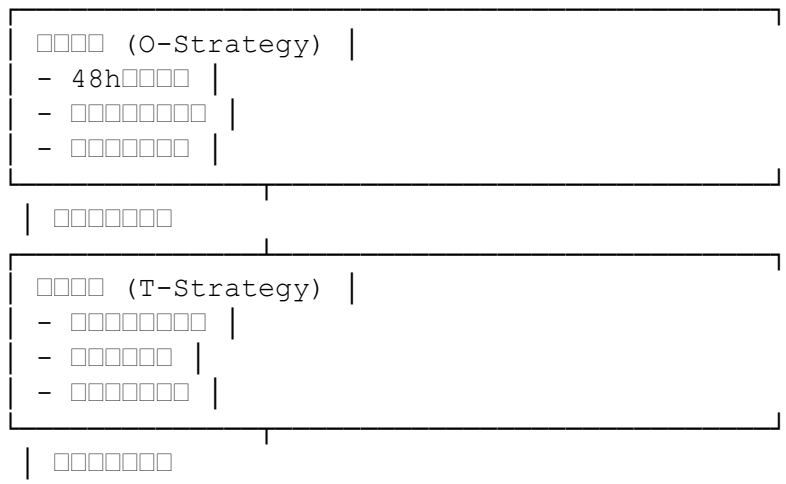
三大核心功能：

1. 多尺度编排：从个体策略推广到团队、组织
2. 伦理治理：公平性、透明度、用户自主权
3. 全局优化：平衡短期效率与长期能力

5.5.2 多尺度编排器（MSO）

设计动机：第四章揭示了I-AVP成功不保证T-AVP/O-AVP成功。L4层需要跨尺度协调。

MSO的三层策略管理（概念架构）：



```
                    │
┌──────────────────────────────────────────┐
│ □□□□□ (I-Strategy)  │
│ - □□□□□ F(t)  │
│ - □□□□□ S(t)  │
│ - AVP□□□□□  │
└──────────────────────────────────────────┘
```

策略协调的核心机制（概念框架）：

机制1：自底向上的能力聚合

```
# □□□□□□
def aggregate_team_ability(team_members):
    """
    □□□□□□□□□□□□□□□□□

    □□□□□□□□ ≠ □□□□□□□
    □□□□□□□□□□
    """
    # □□□□□□□"□□□□□□□□□□□
    min_critical_skill = min([member.C[critical_dim]
    for member in team_members])

    # □□□□□□□
    avg_skill = mean([member.C for member in team_members])

    # □□□□□□□□□□4.1□□
    knowledge_flow = measure_knowledge_sharing(team)

    # □□□□□□□□□□□□
    team_ability = f(min_critical_skill, avg_skill, knowledge_flow)

    return team_ability
```

机制2：自顶向下的策略分解

□□□□□□□□O-AVP□□□≥0.70□□□≥0.85□□□□□□□□□□□/□□□□□□□4.2□□

```
□□□□□□□□□
- □□□□□T-AVP ≥ 0.9□□□□□
- □□□□□T-AVP ≥ 0.8
- □□□□□T-AVP ≥ 0.7□□□□□□
```

```
□□□□□□□□□□
- □□□□□I-AVP□□□□□□□□
- □□□□□□□□□□□□I-AVP□□
- □□□□□□I-AVP□□□□□□□
```

机制3：跨尺度冲突解决

```
□□□□□
- □□□□□□AI□□□□□□□□□I-AVP□
- □□□□□□□□□□□□□□□□□AI□□□
- □□□□□□□□□□ vs □□□□
```

```
MSO□□□□□□□□
1. □□□□□□□□□□ > □□□□□
```

2．□□□□□□□□□□□□□□□□□□□□
3．□□□□□□□□□□□□□□□□□□□□□□□□□□□□

5.5.3 伦理治理框架

L4层必须回答的伦理问题：

问题1：公平性

挑战：

- 残障用户能否使用LSA？（可能无法通过AVP）
- 不同能力起点的用户是否公平？

设计原则（概念框架）：

1. 等效努力原则

（承接3.0.6节）：

- 不是所有人做相同任务
- 根据能力调整任务难度
- 保证等效的认知努力

2. 差异化AVP标准

：

- 残障用户：调整基线B 和δ
- 但不降低提升要求
- 核心：证明”能力在成长

3. 明确豁免场景

：

- 补偿性辅助：不要求AVP（如视障用户的屏幕阅读器）
- 学习性辅助：要求AVP（如编程助手）

问题2：透明度与用户控制

挑战：

- 用户是否知道系统在”故意不给完整答案？
- 用户能否关闭L2/L3/L4？

设计原则：

1. 默认透明

：

- 用户看到当前支持强度S” (t)
- 用户知道为什么得到部分答案
- 用户可查看能力向量C(t)

2. 分级控制

：

- L2-CFE：用户可临时请求更多帮助（记录）
- L3-AVP-TM：用户可关闭监测（失去部分功能）
- L4-MSO：组织管理员可设定策略（需用户同意）

3. 退出权

：

- 用户可永久退出LSA，使用传统AI模式
- 需签署”放弃AVP验证”知情同意

问题3：监测的边界

设计原则：

1. 数据最小化”（见5.4.5节）
2. 本地优先：C” (t)、F(t)、S(t)存储在用户设备
3. 目的限定：数据只用于能力评估，不用于其他目的

5.5.4 全局优化目标函数

L4层需要平衡的多个目标（概念框架）：

```
# □□□□□
def global_objective(system_state):

 LSA□□□□□□□

 □□□□□□□□□□□□□□□□□□□
 "
 # □□1□□□□□□□"□□□□
 task_quality = measure_output_quality" ()

 # □□2□□□□□□□□□□□
 ability_growth = mean([user.C(t) - user.C(t-T) for user in users])

 # □□3□□□□□□□□□□
 user_satisfaction = survey_satisfaction_score()

 # □□4□AVP□□□□□□□□□
 avp_pass_rate = count_avp_passed() / total_users

 # □□5□□□□□□□□□□
 resilience = f(T_AVP, O_AVP, S_AVP_proxies) # □□□□

 # □□□□□□□□□□□□□□□
 objective = (
 w1 * task_quality +
 w2 * ability_growth +
 w3 * user_satisfaction +
 w4 * avp_pass_rate +
 w5 * resilience
 )

 # □□□□□□□
 constraints = [
 avp_pass_rate > 0.7, # □□□□
 user_satisfaction > 0.6, # □□□□
 fairness_score > threshold, # □□□□
 ]

 return objective if all(constraints) else penalty
```

Goodhart防护声明：该目标函数与权重配置仅用于方向性权衡与离线系统评估，不得下推为个人/团队KPI或绩效考核指标；

权重的动态调整（概念指引）：

表5.9: 不同场景的权重配置示例

场景	w （质量）	w （能力）	w （满意度）	w （AVP）	w （韧性）
学习平台	0.2	0.4	0.2	0.15	0.05
生产工具	0.3	0.3	0.2	0.15	0.05
关键系统	0.25	0.25	0.15	0.15	0.2

权重说明：这些数值为设计空间的起点，需要通过A/B测试、用户反馈、业务目标综合确定。不同组织和应用场景应有不同的

Goodhart防护：权重配置用于方向与优先级的质量分级，不应逐项下推为KPI；最终判定仍以AVP主判据与公平性约束为准（

5.6 端到端工作流示例

5.6.1 个体层的LSA应用（概念演示）

场景：Alice使用LSA学习编程

初始状态：

```
[□□□□]
Alice□□□□□□□□□□B□=0.3
□□□□□□
- L1□□□□□AI□□□□
- L2□S(0)=0.8□□□□□□F(0)=0.2□□□□□
- L3□C(0)=[□□□□:0.3, □□:0.2, □□:0.4, ...]
- L4□□□=□□□□□□
```

第1周：高支持，低摩擦

```
[□□1□□□□□□□]
□□□□□□□□□□□□□□□□□□

L1□□□[□□□□+□□□□]

L2-CFE□□□
 □□F=0.2→□□□□□□□"
 □□□□□□
 □□□□□□□□□□□□
 1. □□□□...
 2. □□□□...
 □□□□□□□□□□□□□□□□□□□□

[□□□□]
Alice□□□□□□bug

[L3-AVP-TM] □□□□
- □□□□□□□□□□
- □□□□□0.6□□bug□□□□□□
- □□□□□□0.7□□□□□□
- □□□□□C_Alice.implementation += 0.05

[L2-SGS] □□□□
- □□□□□□□□□□□□□S" (t)
```

第4周：中支持，中摩擦

[□□□□]
- S(t)=0.5□□□□□□
- F(t)=0.5□□□□□□
- C_Alice□□□□□0.5

[□□□□□□□□□□□□□□]
□□□□□□□□□□□□□□□□□

L1□□□[□□□□□□]

L2-CFE□□□
 □□"□□□□□
 □□□□□□
 □□□□□□□□□□□□□
 1. □□□□□□□
 2. □□□□□□
 3. □□□□□□□

 □□□□□□□□□□□□□

 □□□□□□□□□□□□□□□□

[□□□□□□]
Alice□□□□

[L3-AVP-TM] □□□□
- □□□□□□□0.85
- □□□□□□0.9"□□□□□□□□□□
- □□□□□C_Alice.recursion += 0.1

[L3] □□□□
- □□□□□□□□□□□□
- □□□□□□

[□□□□]
□□□
□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□0.6
□□□□□□□□□□□□□□□□□□□

5.6.2 团队层的LSA应用（概念演示）

场景：某软件团队使用LSA进行协作开发

L4-MSO的团队策略：

[□□□□□□□]
□□□□□8□
□□T-AVP□0.72□□□□□□4.1□□
□□□T-AVP ≥ 0.85□□□□□□

□□□□□
1. □□□□□□□□□□□□□L2□□□□S" (t)□
2. □□□□□□□□□□□□□□□□□4.1□□
3. □□"□□□□□□□L3□□I-AVP□

[□□]
□□□□□□□□
- □□□□□AI□□□□"□□□□"
 "□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□

[□□]
L3-AVP-TM□□□□□
- □□I-AVP□□□□7/8 " (87.5%)
- □□□□□□□□□0.68□□□□□
- □□□□□□□Bob□□□□AI

[□□]
L4□□□
- Bob□□□□□□
- □□□□□□□□□□□□Alice□
- □□□□□□□□□□□□AI□□

[□□]
□□□□□T-AVP□□□0.83□□□□□□

5.7 技术可行性与工程挑战

5.7.1 现有技术栈的适配性

表5.10: LSA各层与现有技术的概念映射

LSA层	核心功能	可用技术方向	成熟度评估
L1	基础AI能力	主流大语言模型	高
L2	输出调制	提示工程、微调	中
L2	难度估计	学习分析、IRT理论	中
L3	能力建模	贝叶斯网络、RL	中-低
L3	行为追踪	遥测系统	高
L4	策略编排	规则引擎、优化	中
全局	多尺度协调	Multi-agent研究	低

关键技术缺口：

1. 能力向量的精确建模

（L3）

- 当前状态：启发式方法、用户自评
- 需要发展：更精确的认知模型、神经科学启发

2. 摩擦强度的自动化调制

（L2）

- 当前状态：手工设计规则
- 需要发展：自适应算法、强化学习方法

3. 团队能力的涌现建模

（L4）

- 当前状态：简单聚合方法

- 需要发展：复杂网络理论、多智能体模拟

5.7.2 最小可行原型（MVP）设计方向

从完整LSA到MVP的简化路径（概念指引）：

核心策略：MVP强制采用”最小遥测→逐步丰富策略，先观察AVP趋势再增加维度（避免过早优化与数据过采）。从最少的监

第一阶段：单用户基础功能

□□□□□
- L1□□□□□AI□□
- L2-CFE□□□□□□□□□□□□□□□□□□
- L2-SGS□□□□□□□□□□□□□□□□□□

□□□□□
- □□□□□□□□
- □□□□□□□□□□□□□□□
- □□AVP□□□□□□□□□□□□

第二阶段：加入监测功能

□□□□□
- L3-AVP-TM□□□□□□□□□
- □□□□□□□□3-5□□
- □□□□□□□□□□/□□□

□□□□□
- □□□□□□□□
- AVP□□□□□
- □□□□□□□□□

第三阶段：团队协调功能

□□□□□
- □□T-AVP□□
- □□MSO□□□□□□
- □□□□□□□□□

□□□□□
- T-AVP□□□□
- □□□□□□□
- O-AVP□□□□

5.7.3 关键工程挑战

挑战1：实时性能

问题：L2/L3的计算能否在用户可接受延迟内完成？

概念解决方向：

- 预计算：提前准备多难度版本
- 异步更新：背景更新C” (t)
- 边缘计算：能力向量本地存储

挑战2：模型对齐

问题：如何让L1的AI理解”适度帮助的语义？

概念解决方向：

- RLHF方法：训练AI理解摩擦指令
- 提示工程：设计元提示模板
- 微调方向：在LSA场景上进行领域适配

挑战3：数据冷启动

问题：新用户/新任务如何初始化？

概念解决方向：

- 快速评估：简短的能力评估
- 保守初始化：假设较低能力，快速调整
- 迁移学习：从相似用户群体借鉴

挑战4：评估的客观性

问题：如何客观测量用户能力”？

概念解决方向：

- 多任务交叉验证
- 标准化任务库建设
- 社区验证机制

挑战5：用户接受度

问题：用户是否愿意接受”不完整答案？

概念解决方向：

- 渐进引入”（初期低摩擦）
- 透明沟通（解释目的）
- 成就系统（能力成长可视化）
- 紧急逃生阀（真正需要时可请求完整帮助）

挑战6：伦理与隐私

问题：如何在监测能力的同时保护隐私？

概念解决方向：

- 联邦学习：本地计算，只传聚合
- 差分隐私：添加噪声保护
- 透明审计：用户可查看所有数据

挑战7：多尺度的复杂性

问题：团队/组织层的协调极其复杂

概念解决方向：

- 分阶段实施（先个体，再团队）
- 简化假设（初期忽略复杂交互）
- 人工辅助（L4初期由管理员操作）

5.7.4 常见问题解答（FAQ）

Q1: 权重配置如何确定？

A: 权重配置基于”场景目标对齐原则，需要迭代校准：

- 初始阶段：基于专家判断和领域惯例设定起点
- 校准方法

：

1. 小规模A/B测试（参考教育技术领域的实践）
2. 监测多指标仪表盘
3. 根据AVP通过率、用户留存、能力增长等综合指标调整
4. 定期复审（如每季度）

- 透明性：所有权重配置及其调整历史应记录，支持后续分析

Q2: 团队层的T-AVP如何量化测量？样本量多大才可靠？

A: T-AVP测量的统计考虑（见4.1节）：

- 最小样本建议：团队规模≥5人（更小团队建议只做I-AVP）
- 任务等值性：使用IRT校准的平行测验，确保测试难度一致
- 置信区间：报告95% CI，典型需要多次独立测量
- 局限性说明：小样本下估计不稳定，建议结合定性观察

Q3: MVP如何开局？最小遥测是什么？

A: MVP的最小可行遥测建议：

□□□□□L1+L2□□□□□□
- □□□□□□□□□□□□□□□□□□□□□
- □□□□□□□□□□

□□□□□□□L3□□□□
- □□□□□
  * □□□□□□□□□
  * □□P□□□□□□□□"□□/□□"□
  * □□□□□□□□□
- □□□□□□□AVP□□

□□□□"□□□□□□□
- □□5.6□□□□□□
- □□□□C" (t)□□□□

关键：从最少数据开始，根据需求渐进增加，避免过早优化。

Q4: 如何处理用户绕过摩擦（使用其他AI工具）？

A: 这是LSA最大的实施挑战之一：

- 技术手段（有限）

：

- 某些情况下可检测其他AI工具使用（但易被规避）
- 代码相似度检测（如抄袭检测）
- 文化手段（核心）

：

- 透明沟通LSA的目的（能力建构，而非限制）
- 成就系统奖励独立完成
- 社区规范（如组织内的”AVP诚信公约”）
- 设计手段

：

- 让摩擦有意义而非惩罚性”

- 提供”紧急逃生阀”（真正需要时可请求完整帮助）
- 接受现实：完全防止绕过不可能，目标是让”诚实使用成为主流

Q5: LSA是否会让残障用户处于不利地位？

A: 这是伦理治理的核心关切（见5.5.3节）：

- 原则：“等效努力而非相同任务”（承接3.0.6节公平性原则）
- 实践方向

：

- 视障用户：调整任务形式（如口述编程代替键盘输入），但认知摩擦等效
- 学习障碍用户：延长时间、提供辅助工具，但AVP标准调整为相对提升
- 关键：评估的是”能力是否在成长”，而非”是否达到绝对水平
- 监测：分组AVP通过率报告，主动识别系统性偏见

Q6: L1层可以是闭源模型吗？

A: 可以，这正是LSA分层设计的优势：

- L1技术中立：可以是主流闭源或开源模型
- L2-L4独立：摩擦/监测/编排逻辑不依赖L1具体实现
- 可替换性：随着AI技术进步，可无缝升级L1而不影响上层
- 局限性：闭源模型的调制能力可能受限于API接口
- 开源优势：如果L1是开源模型，可以更精细地控制输出

5.8 小结：LSA的理论贡献与实践路径

5.8.1 本章核心贡献

1. 架构创新

：

- 首次提出将CET理论工程化的完整架构
- 四层分离的设计”（L1-L4）使得每层职责清晰，支持独立升级和替换
- 明确的层间接口契约（表5.2），支持多团队并行开发
- 为AI辅助工具提供能力建构优先的设计范式

2. 机制设计

：

- CFE（认知摩擦引擎）：实现有益摩擦，提供多策略空间
- SGS（支持削减调度器）：实现渐进独立，引入回退与保底机制（承接3.3.3节）
- AVP-TM（遥测模块）：持续能力监测，支持多尺度AVP（承接4.5.2节）
- MSO（多尺度编排器）：跨尺度协调，整合公平性约束

3. 可工程化

：

- 提供从理论到实现的概念路径
- 明确技术栈映射和成熟度评估
- 给出MVP实施方向
- 常见问题解答回应实施挑战

5.8.2 与前后章的连接

承接前四章：

- 第三章的AVP/EML/伙伴式主体性 → LSA的L2/L3/L4实现
- 第三章3.3.3节的S_min保底支持 → L2层的SGS回退机制
- 第四章的T-AVP/O-AVP → LSA的L4多尺度编排
- 第四章4.5.2节的多尺度AVP监测 → L3层的AVP-TM遥测管线

为第六章铺垫：

- LSA的每个设计决策都可以证伪
- 明确的技术挑战指向未来研究方向
- 伦理问题需要跨学科讨论

5.8.3 LSA的三个开放问题

问题1：最优摩擦参数是否存在？

当前状态：使用启发式规则（50–70%成功率，工作假设）

证伪路径：

- 大规模实验：测试不同F” (t)对AVP的影响
- 个体差异：最优参数是否因人而异？
- 领域差异：编程vs写作vs数学的最优摩擦是否不同？

问题2：能力向量能否精确建模？

当前状态：简化的多维向量

挑战：

- 认知能力的维度到底有多少？
- 不同能力间的相关性如何？
- 能力的动态演化遵循什么规律？

未来方向：

- 认知科学+机器学习的交叉研究
- 大规模纵向数据收集
- 神经科学的启发（fMRI, EEG）

问题3：多尺度协调的理论基础

当前状态：启发式策略聚合

深层问题：

- 团队能力如何从个体涌现？（复杂系统理论）
- 组织韧性的数学模型？（网络科学）
- 跨尺度的最优控制？（控制理论）

未来方向：

- 多智能体系统研究
- 组织行为学的定量化
- 跨学科的理论整合

5.8.4 实践路径建议

对工具开发者：

1. 从基础功能的MVP开始

2. 选择单一领域深耕（如编程、写作）
3. 快速迭代，收集真实用户数据
4. 逐步添加监测和编排功能

对研究者：

1. 聚焦某个开放问题（如能力建模）
2. 设计对照实验验证LSA假设
3. 发表子领域成果
4. 推动跨学科合作

对组织：

1. 试点项目：选择非关键团队测试
2. 建立AVP评估基线
3. 逐步推广到关键业务
4. 建立长期监测体系

本章结束。下一章将探讨CET理论的局限性、证伪路径与未来研究方向。

第六章 局限性、证伪路径与未来方向

前五章构建了完整的CET理论体系:从理论定位(第一章)、跨学科基础(第二章)、核心机制(第三章)、跨尺度扩展(第四章),到

本章承诺:

1. 不回避理论的薄弱环节
2. 不过度声称理论的适用范围
3. 提供可操作的证伪路径
4. 主动邀请批判和验证
5. 为后续研究者指明方向

本章逻辑:我们将从理论的内在局限(6.1)出发,明确哪些假设可以被证伪(6.2),哪些问题需要未来研究(6.3),最后讨论理论的

6.1 理论的六大局限性

6.1.1 尺度边界:个体以下与社会以上

局限陈述:CET聚焦”个体→团队→组织→社会四个尺度,但对更微观”（神经生理)和更宏观(跨文化/跨代际)的机制涉及不足

具体表现:

1. 神经层面的机制缺失:
   - CET未深入探讨AI使用对大脑可塑性的影响
   - 例如:长期使用GPS是否改变海马体?长期依赖AI是否改变前额叶?
   - 相关研究(Maguire et al., 2000)显示伦敦出租车司机的海马体变化,但AI时代的神经科学证据仍然稀缺
   - 后果:CET的”能力退化主张主要基于行为证据,缺乏神经机制的直接验证
2. 跨文化普适性未验证:
   - CET的案例和论证主要来自WEIRD社会*“（Western, Educated, Industrialized, Rich, Democratic,指西方、受过教育的、工业化的、富裕的、民主的人群与社会语境）
   - 集体主义文化vs个人主义文化对”独立能力的价值判断可能不同
   - 例如:东亚文化中”依赖他人可能被视为团队协作美德,而非能力缺失
   - 后果:AVP的独立能力标准可能具有文化偏见
3. 代际传承的长期机制:
   - 第四章讨论了T /T_1/T_2三代模型,但跨代传承的具体路径”（家庭/学校/社会)未充分建模
   - 需要10–20年的纵向研究才能验证代际鸿沟假说

- 后果:S-AVP的预测具有高度不确定性

边界声明:

CET□□□□□□□:
□ □□□□□□(□□□□□□□□□□□)
□ □□□□□(5-50□)
□ □□□□□□(□□□□)
□ 10□□□□□□□□□□(□□□□)

? □□□□□□:
- □□□□□(□□□□□□□)
- □□□□□(□□□□,□□□□□)
- □□□□□(20□□□,□□□□□)

□ □□□□□:
- □□□□□□(□□□□□□□□)
- □□□□□□(□□□□□,□□□□,□3.0.6□)
- □□□□□□(□□□□,□□□□□□□)

6.1.2 任务类型的限制

局限陈述:CET针对认知密集型、可学习的任务,但对物理任务、创造性任务、社交任务的适用性有限。

适用性矩阵:

表6.1: CET在不同任务类型中的适用性

任务类型	适用性	理由	示例
程序性认知任务	高	可明确定义能力、可测量进步	编程、写作、数学
概念性认知任务	中-高	能力可培养,但测量难度高	设计、策略规划
创造性任务	中	能力与灵感混合,AVP难以量化	艺术创作、科学发现
社交任务	低-中	能力受情境影响大,难以标准化测量	谈判、领导力
物理任务	低	CET聚焦认知能力,物理技能机制不同	手术操作、运动技能
高风险任务	不适用	拔线测试可能带来不可接受风险	飞行、医疗急救

注:本表仅用于方向与分层的质量参考,不得下推为KPI";最终判定仍以AVP主判据为准"(见3.0.2节)。

关键限制:

1. 能力定义的模糊性:
   - 创造性任务中的"能力"难以与"灵感、天赋"分离
   - AVP判据"($P_2 \geq B_0 + \delta$)假设能力可量化,但某些能力本质上是定性的
2. 测量的生态效度:
   - 实验室测量(如编程测试)vs 真实场景(如生产环境开发)的差异
   - 高风险、高压力环境下的能力表现可能与常规测试大相径庭
3. 任务复杂度的上限:
   - CET主要关注中等复杂度任务(数小时至数天完成)
   - 超大规模、跨月的复杂项目(如大型软件系统设计)涉及更多组织因素,超出CET范围

   CET的核心洞察(认知卸载→能力退化)可能在多种任务中成立,但AVP/EML的具体操作化需要针对任务类型进行调整。本证
   70%成功率,工作假设)主要基于程序性认知任务的证据。

6.1.3 测量挑战:从概念到操作的鸿沟

局限陈述:CET提出的许多概念(如C(t)能力向量、认知摩擦强度F)在理论上清晰,但在实践中精确测量极其困难。

四大测量难题:

难题1:能力向量的维度爆炸

□□:
- □□□□□C(t) = [c□, c□, ..., c□]
- □n□□□□□?5□?50□?500□?
- □□□□□□□□□□□□?□□□□□□□□□?

□□□□:
- □□□□□□□□(□3-10□)
- □□□□□□□□□□□□□□

□□□□:
- □□□□□□□□□□□□□□□□□□□□
- □□□□□(□Gardner□□□□□□□Cattell□□□/□□□□□)□□□□□□□

难题2:摩擦强度的量化

□□:
- "□□□□□□□□□□□□□□□□□□□?60%?70%?
- □□□□□□□"□□□□?

□□□□:
- □□□:□□□□□□□□□ / □□□□□□□□□
- □□□□□□□:□□□□□□□□□□□□□
- □□□□□:□□□□□□

□□:
- □□□□□□□"□□□□□□" (□□□□vs□□□□)
- □□□□□□□□□□□,□□□□□□
- □□□□□□□□□□□□□□□

难题3:AVP的等值性保证

□□:
- □□□□□T□□T□□□□□□□□□?
- □□□□□□(□□)□□□□□,□□□□□□(□□)□□

□□□□:
- IRT(□□□□□□□):□□□□□□□□□□
- □□□□□:□□□□□□□
- □□□□□:□□□□□□□

□□□□:
- □□□□□□□□□"□□□□□□
- □□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□" (□□□6.2)

难题4:长期追踪的流失率

□□:
- AVP□□6-12□□□□□□□□W=4-8□□□□6□□□□□□□□□□□□/□□□□□□
- T-AVP/O-AVP□□□□□□□□
- S-AVP□□10-20□(□□□□)

□□□□:
- □□□□□:6□□□□□□30-50%
- □□□□:□□□□□□□□□□□
- □□□□:□□□□□□□□□□□□□

□□:
- □□□□(□□□□□□□□□)
- □□□□□□
- □□□□□□(□□□□□□)

表6.2: AVP测量的主要效度威胁与缓解策略

威胁类型	具体表现	缓解策略	局限性
等值性威胁	T_0和T 任务难度不匹配	IRT标定、专家双盲评审、平行测验设计	IRT需大量数据”;专家判断仍有主观性
环境一致性	测试环境与真实使用场景差异	生态效度测试、现场评估、多情境验证	成本高;难以完全模拟真实压力
练习效应	重复测试导致熟悉度提升	题库轮换、冷却期设计”(≥2周)、等值平行卷	题库开发成本高;冷却期延长研究周期
题库污染	测试题目泄露或过度练习	动态题库、防作弊设计、行为模式检测	无法完全杜绝;检测算法可能误判
评分者期望	评分者知晓用户使用AI而产生偏见	盲评机制、双评分+一致性检验、标准化rubric	盲化难度大(如编程风格暴露);一致性训练成本高
跨群体公平(DIF)	不同群体在等值题上出现系统偏差	IRT-DIF检测、项目替换/重标定、分组等值化	需要更大样本与多文化题库;可能降低可比样本量
延迟巩固威胁	W期内短期记忆尚未转化为长期能力	W=4-8周(工作假设)、多时间点测量、	最优W因任务而异;多时间点增加流失率
流失率偏差	高动机用户留存率更高	意向分析(ITT)、敏感性分析、激励设计	ITT保守估计可能掩盖真实效应;激励可能扭曲行

注:本表仅用于方向与分层的质量参考,不得下推为KPI;最终判定仍以AVP主判据为准(见3.0.2节)。

透明承认:

CET提供的测量协议(第三章AVP、第四章T-AVP/O-AVP、第五章LSA遥测)代表了当前最佳实践的概念框架,而非经过充分

6.1.4 个体差异与公平性悖论

局限陈述:CET追求”能力提升的普适性目标,但个体差异” (学习速度、起点能力、动机)和公平性要求(残障、社会经济背景

三个悖论:

悖论1:标准化vs个性化

□□:
- AVP□□□□□□□(P_2 ≥ B_0 + δ)□□□□□□
- □□□□□□□□□□□(□□3□□□,□□3□□)

□□:
- □□□□□□□W=4-8□□□□6□□□□□□□□□□□□□□□□,□□□□□□□□□
- □□□□□□□□□□□□□□□(□□□□□□□□W?)

□□□□:
- □□□□(□□/□□/□□□□□δ)
- □□□□□+□□/□□□□(4-12□)
- □□□□□□□□

悖论2:能力vs残障

□□:
- CET□□□□□□
- □□□□□□□"□□□□□□□□□□□" (□□□□□□□□)

□□□□(□3.0.6□□□□□□):
- **□□□□□□□□□□□□□□□□**
- □□□**□□□□**□□□□□□□□□
- □□:"□□□□"□□

□□□□□□□□:
- □□□□" (□□□□□):□□□□□□□□□□□□□□□□
- □□□□□(□□□□□ADHD):□□□□□□□□□□□□□□□□□□
- □□□□□(□□□□□□):□□□□□□□□□□□□?

悖论3:机会vs结果

□□:
- □□□□□□□□□□□(□□□□□□AVP)
- □CET□□□□□□□□(P_2 ≥ B_0 + δ),□□□□□□"□□

□□□□:
- AVP□□□□□□□"□□□□□?
- □□□/□□□□□,AVP□□□□□□□□□□□"□□?
- □□□□AVP□□□□□□□□□□?

□□□□" (□□□□□):
1. **□□□□□□**:
 - □□□AVP□□(□□□□□□)
 - □□□□□□,□□□□□

2. **□□□□□□**:
 - □□□□□□□
 - □□□□□□(□□"□□□)

3. **□□□□□□□**" (□□□□□□□□□□□□):
 - □□□□□□□□□:
 $\max(AVP□□□by□□) - \min(AVP□□□by□□) \leq \varepsilon_{fairness}$
 - □□$\varepsilon_{fairness} = 0.15$(15%□□□□□,□□□□)

4. **□□□□□□□□**:
 - □□□□□□:□□□□□□□□□□
 - □□□□□□□:□□□□□↑,□□□□□□□□□□
 - □□:□□□□□□□□(□□□□□□□□□□□□)

5. **□□□□□□**:□□□□□□□□□□□,□□□4□□□□□□□□□

方法学提醒:异质性效应与辛普森悖论

在所有涉及分组或亚群的AVP评估中,必须关注: - 异质性效应(HTE): 报告关键亚组(如不同年龄、教育背景、起始能力水平) - 辛普森悖论: 检视整体效应与分组效应是否出现反转”;必要时采用分层或多层模型汇总 - 交互作用: 测试干预效应是否因群体特征而显著调节

所有亚组分析必须在预注册中指定,避免事后数据挖掘导致的假阳性。

6.1.5 技术依赖的双刃剑

局限陈述:CET理论自身依赖AI技术的稳定性和可访问性,这创造了一个潜在的自我指涉悖论。

三个依赖层次:

层次1:测量依赖

□□:
- AVP/EML□□□□□AI□□□□□
- □□AI□□□" (□API□□□□□□□□□□□□□),□□□□□□□

□□:
- □□□□:□□□□AI□□□(□OpenAI)
- □□□□:GPT-4 vs GPT-5□□□□□□□□□□□
- □□□□:□□□□□□□□□□□□

□□:
- □□□□□□(□□□□API)
- □□□□□□(□□□□□□□□□□)
- □□□□□□(□Llama□)

层次2:能力定义依赖

□□:
- "□□□□□□□□□□AI□□□□
- □AI□□□□,□□□□□□□□□□

□□:
- 2023:□□□□□□□□□□□□□□□□
- 2030" (□□):AI□□□□□□□,□□□□□□"□□□□"?
- □□AVP□□□□□□□□□□□

□□□□:
- CET□□□□□□□□□□——□□□□□□□□□□□?
- □□□□□□□□□□□□□□□□□

层次3:社会依赖

□□:
- □□□AVP□□□□□□□□□□" (□□□□□□□□□□)
- □□□□□□□"□AI□□"□□□□□□□□

□□□□:
- □□□□□□□□AI□□□□□□
- □□□□□□□□□AI□□□□"□□AVP□□
- □□□□□□□" (T_0□□□□□,T_2□□□□□?)

□□:
- CET□□□□"□□□□□□
- □□□□□□□□□□□□□□□□□" (□□□□□□,□H8)

□□□□□:
□□"□□□□□□□□□□□□□□□□□□□,
□□□□□□□□□□□□□□□□CET□□□□□□
"□□□□□□,□□□□□□□□"□

6.1.6 文化嵌入性与价值判断

局限陈述:CET的能力建构目标隐含了特定的价值判断——“独立能力”是值得追求的。但这一价值判断并非普适,而是文化嵌

三个文化维度的挑战:

维度1:个人主义vs集体主义

□□□□□□□□:

- □□□□□□□□□□□□□□□□□
- AVP□"□□□□□□□□□□□□

□□□□□□□□:
- □□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□,□□□□□□
- AVP□□□□□□□□□□□□

□□□□:
- □□□□□AVP□□□□□□□□□
- T-AVP□□□□□□□□□□□□□□□□□" (□□□□)

维度2:能力观的文化差异

□□□□□□□□□:
- □□□□□□□□□□□□□□□
- □□□□□□□□AVP□□

□□□□□□□□□:
- □□□□□□□□□□□□□□□
- AVP□□□□□□□□□□□□□□

□□:
- □□□□AI□□□□□□□□□□□
- □□AVP□□:□□□□(□□□□)
- □□□□:□□□□□□□□□□□□

维度3:技术关系的哲学差异

□□□□□□□□:
- □□□□□□□□□□,□□□□□
- CET□"□□□□□□□□□□□□□□□□

□□□□□:
- □□□□□□,□□□□" (□□□□□□□□□□□□)
- "□□□□□□□□□□□□□□□,□□□□

□□□□:
- CET□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□□

透明承认:

> 我们承认CET理论的价值判断是文化嵌入的,主要反映WEIRD社会的认知传统。这不意味着理论无效,但意味着理论的适用
> (见6.3.2节中期研究议程)。

6.2 八个可证伪假说及其证伪路径

6.2.1 核心假说概览与因果识别策略

CET理论的科学性在于其可证伪性。我们明确提出8个核心假说,并给出清晰的证伪条件。这不是为了”保护理论免受批评,而

元认知原则:我们期待至少部分假说被证伪——这不是失败,而是科学进步的标志。如果所有假说都被验证,那可能意味着我们

因果识别策略菜单

CET假说的验证面临一个核心方法学挑战:如何在无法实施RCT的情况下建立因果推断? 我们提供以下替代策略菜单,供研究者

策略1: 鼓励设计” (Encouragement Design) - 适用场景: 无法强制分配AI使用,但可以鼓励 - 核心思路: 随机分配”鼓励使用AIvs不鼓励,利用工具变量原理 - 示例: 向实验组提供免费AI订阅,对照组不提供 - 分析方法: 两阶段最小二乘法” (2SLS),以”鼓励”作为工具变量 - 优势: 保留随机性,符合伦理 - 局限: 依赖”单调性假设” (鼓励只增加不减少使用)

策略2: 阶梯楔形设计(Stepped-Wedge Design) - 适用场景: 组织/团队层面,无法同时部署到所有单位 - 核心思路: 所有单位最终都接受干预,但时间随机错开 - 示例: 50个团队分5批引入EML工具,每批间隔4周 - 分析方法: 混合效应模型,控制时间趋势 - 优势: 伦理友好(所有人最终受益),适合实践场景 - 局限: 需要假设无时间依赖的干预效应

策略3: 工具变量(Instrumental Variables) - 适用场景: 存在影响AI使用但不直接影响能力的外生因素 - 核心思路: 利用准随机”变异进行因果推断 - 示例: 公司分配的AI许可证数量” (受预算而非能力影响) - 分析方法: 2SLS或更鲁棒的估计量 - 优势: 可利用观察数据建立因果推断 - 局限: 工具有效性难以验证(需要领域知识和敏

策略4: 自然实验(Natural Experiments) - 适用场景: 外生冲击导致AI可用性突变 - 核心思路: 利用意外事件(如API中断、政策变化)作为准实验 - 示例: * OpenAI API大规模宕机事件 * 某国突然禁止/开放AI工具使用 - 分析方法: 事件研究法、合成控制法 - 优势: 高外部效度,真实场景 - 局限: 机会稀少,难以预先规划,可能缺少对照组

策略5: 回归断点设计(Regression Discontinuity) - 适用场景: 存在基于阈值的政策/分配规则 - 核心思路: 利用阈值附近的准随机分配 - 示例: * O-AVP告警阈值≥0.70附近的组织(见H5) * 某公司规定”绩效>80分才能用高级AI工具 - 分析方法: 局部线性回归,检验阈值处的跳跃 - 优势: 强因果推断,适合政策评估 - 局限: 仅对阈值附近群体有效,外推需谨慎

选择决策树:

□□□□□□□?
├─ □ → □□RCT
└─ □
  ├─ □□□□/□□? → □□□□
  ├─ □□□□□□? → □□□□
  ├─ □□□□□? → □□□□
  ├─ □□□□□? → □□□□
  └─ □□□□□? → □□□□

所有准实验方法都依赖不可检验的假设” (如工具有效性、平行趋势)。研究者必须: 1. 明确说明假设及其合理性 2. 进行鲁棒性检验(如安慰剂测试、敏感性分析) 3. 承认因果推断的局限性 4. 报告所有分析结果(包括不支持假说的)

表6.3: 八个核心假说的证据强度与优先级

假说编号	假说名称	证据强度	所需研究设计	预期时间线	优先级
H1	AVP-Basic假说	Moderate	RCT + 纵向追踪	1-2年	P0
H2	有益摩擦假说	Preliminary	多臂试验 + A/B测试	1-3年	P0
H3	系统性削减假说	Preliminary	对照实验	1-2年	P0
H4	团队能力极化假说	Strong	自然实验 + 田野研究	2-3年	P1
H5	组织韧性假说	Moderate	准实验 + 回归断点	3-5年	P1
H6	摩擦调制假说	Preliminary	A/B测试 + 用户研究	1-2年	P2
H7	能力向量假说	Preliminary	降维分析 + 预测建模	2-4年	P2
H8	代际鸿沟假说	Preliminary	纵向队列研究	10-20年	P1

注:本表仅用于方向与分层的质量参考,不得下推为KPI”;证据强度和时间线均为工作假设,需根据实际研究进展校准。

证据强度说明: - Strong: 多个独立研究收敛,至少1个高质量RCT或准实验 - Moderate: 部分实证支持,但样本量小或设计局 - Preliminary: 主要基于理论推导和方向性观察证据

表6.4: 核心假说的证伪对照与理论修正路径

假说	最小样本/时长	核心证伪观测	因果识别策略	失败后的修正方向
H1	N≥200, W=4–8周（默认6周）	P_2-B_0效应量<0.2或反向	RCT或鼓励设计	检查EML条件是否充分执行;重新校准δ阈值;可能放弃AVP作为
H2	N≥300, 12周	无倒U曲线或峰值在<30%/>85%	多臂试验	承认”最优区间因任务而异”;改为自适应个性化策略;可能放
H3	N≥240, 12周	固定支持≥削减组	对照实验	承认削减可能非必要;改为”按需支持”模型”;保留AVP但修正
H4	50–100团队, 6–12月	能力方差无变化或缩小	自然实验+DID	承认团队协作可能缓冲极化;T-AVP可能更依赖组织文化而非技术设计
H5	20+组织, 12–24月	O-AVP与恢复时间无关	回归断点+事件研究	承认O-AVP双阈值模型可能不普适;改为连续风险评分;可能放弃单一阈
H6	N≥300, 12周	动态摩擦无优势或劣于A/B测试	LSA vs A/B测试	承认简单即美;删除自适应复杂度;LSA简化为固定策略库
H7	N≥1000, 6–12月	能力向量解释方差<10%	降维+预测建模	承认能力可能高维不可约;改为任务特定能力模型;放弃统一C(t)向量
H8	3代队列, 15–20年	Cohen’s d<0.3(T_2 vs T_0)	纵向队列	承认代际差异可能被高估”;教育干预有效;CET警示过度但原则

注:本表仅用于方向与分层的质量参考,不得下推为KPI;修正方向为理论演化预案,并非失败即放弃理论。

修正方向说明: - 参数校准型: H2、H5 - 核心原则保留,调整具体参数 - 条件精细化型: H3、H4 - 识别假说成立的边界条件 - 模型简化型: H6 - 承认复杂度不必要,回归简单设计 - 理论修正型: H1、H7 - 可能需要修改核心假设 - 警示价值型: H8 - 即使被证伪,警示作用仍有意义

6.2.2 个体层假说的证伪路径

假说H1: AVP-Basic假说

□□:
□□□□□□□□□□,□□□□□□□" (F=50-70%□□□,□□□□)
+ □□□□□□□□(S4→S1→S0)□□□AI□□,□□W□□,
□□□□□□□□□□□□□□□□□□□□□□□□□□□(P_2 ≥ B_0 + δ)□

□□□□:
- □□□RCT□□□,□□□(□□□□)□□□□(EML□□)
 □P□□□□□□□□□□(□□□ < 0.2)
- □□□□□P□ < B_0(□□□□)

□□□□:
- □□:2×2□□RCT(□□□□×□□□□)
- □□:N≥200(□□50),□□□□□(□□/□□/□□)
- □□:□□□AVP□□(□□□□),□□□□
- □□:□□□□(ITT) + □□□□□□(PP)
- □□:□□□□□□□□□□□□□□□□□□

□□□□□:□□EML□□□□□□□(□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□50-70%□□)□

□□□□:□□□□□□□□□□□□□□□,□□□□□□□□□□□□□□□□□□□□□□□□(□□□□□□□□□□□□□□")□

□□□□:
- T_0: B_0□□□□
- T_1-T_2: □□□□W=4-8□□□□6□□□□□□
- T_3: P_2□□□□□W=4-8□□□□6□□□□□□□□□□□□□□/□□□□□
- T_4: □□□□□□" (6□□□,□□)

□□□□:

- □□□(EML):P_2 - B_0 ≥ 0.3 SD (□□□□)
- □□□(□□□/□□□):P_2 - B_0 ≈ 0 □ < 0

□□Cohen's d < 0.2□□□□□,□H1□□□□

□□□□□□:
□□□□24h□□□□□(I-AVP□),□□□□□□48h/□□";
□□□□□□□□□□□□□SOP□

假说H2：有益摩擦假说

□□:
□□□□"□□□□□" (50–70%□□□,□□□□),□□□□□,
□□□□□□□(P_2 - B_0)□□□□□□□□(□□□<30%)
□□□□,□□□□(□□□>85%)□□□□□

□□□□:
- □□□□□□□□□□□□□□□□(□□U□□)
- □□□□□□□□□□□50–70%(□□□□□20%□90%)

□□□□:
- □□:□□□□,5–7□□□□□(20%/40%/60%/80%/95%)
- □□:N≥300(□□50),□□□□□□□
- □□:□□□□(P_2 - B_0),□□□□□,□□□
- □□:□□□□□□□□□□□□□□□□□

□□□□:
- □U□□□,□□□50–70%□□(□□□□)
- □□□□□□□□□□□<30%□>85%,□H2□□□

□□□□:
- □□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□

假说H3：系统性削减假说

□□:
□□□□□□□(S4→S1→S0→0)□□□□□□□□□□□□□□
AVP□□□□□□□□□□□□□□□□□□□□

□□□□:
- □□□□□(□□□S3)□P□□□□□□□□□
- □□□□□□□□□□□(□□/□□/S□□□□□□)

□□□□:
- □□:3×2□□□□(□□□□×□□□□)
  * □□□□:□□□□□□S□
  * □□□□:□(4□)□□(12□)
  * □□:□□S3□□
- □□:N≥240(□□40)
- □□:AVP□□□□□□□□(6□□)□□□□□

□□□□□:□□□□□□□□□□,□□□□□□□□□□□□□";□□□□□□□□□□□□□□

□□□□:
- □□□>□□□" (Cohen's d ≥ 0.3 □ ≥10%□*working assumption*□)

- S□□□□□□□□(□□□□□)

□□□□□□□□□□□□□□,□H3□□□□□

6.2.3 团队与组织层假说的证伪路径

假说H4: 团队能力极化假说

□□:
□□EML□□□□AI□□□□,□□□□□□□□□□□□:
□□□□□□□□□□□,□□□□□□□□□□,T-AVP□□□□

□□□□□:
- □□□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□(□□□□□□)

□□□□□:
- □□:□□□□□(AI□□□□□□□□)
- □□:50-100□□□□□□□□
- □□:□□I-AVP□□□□T-AVP□□□□□□□□
- □□:□□□□□□□□□□□□□□□□□□□□

□□□□□:
- T_0: □□□□□AI□□□□□
- T_1-T_2: 6-12□□□□□□
- T_3: □□□□AI□"□□□□□

□□□□□:
- □□□□□□□□□□" (F□□,p < 0.05)
- □□□□□□□≥0.15(□□□□□)
- □□□□□□□□□□□□□

□□□□□□□□□□□□,□H4□□□□□

假说H5: 组织韧性假说(双阈值模型)

□□(□□□):
O-AVP□□□□□□□□:
- □□□□□: 0.70(□□□□□,□□□□□□□/□□)
- □□□□□: 0.85(□□□□□,□□□□□□□/□□□□□□□)

□□□□□□□□□□□□□□4.2.2□(O-AVP□□),□□□□□□□□□□□□□□□□□□□□

O-AVP < 0.70□□□□□□□□□□□□□AI□□□□□□□□□□□□□□,
□□□□□□□□□□□□□□

□□□□□:
- □□O-AVP < 0.70□□□□□□□□□□AI□□□□□□□□□(<12h)□□□□
- □□□□□□□□□□□□□□□□□□□□

□□□□□(□□□□□□):
- □□A: □□□□□□48h□□(□□□□□□)
- □□B: □□□□□(AI□□□□□□□□□□□□□□□)
- □□C: □□□□□(Agent-Based Model)

□□□□□:

- □□20+□□□O-AVP□□
- □□□□AI□□□□□□□□□□
- □□: O-AVP□□□□□□□□□
- □□: O-AVP < 0.70□□□□□□□□□□□□□□□□□□□□

□□□□□□:
- □□□□□□□: □□0.70□□□□□□□□
- □□□□□□□: AI□□□□□□□□□□

□□□□(□□/□□□):
- □□□/□□□□□□□□□□(□□ICC)
- □□□□□□□□□□GEE□□□□□□
- α=0.05□Power≥0.8(□□□□),□□□□□□□

□□□□(□□□□):
O-AVP□□□□□□□
- □□□□□≥0.70□□□□□□□□/□□□
- □□□□□≥0.85□□□□□□/□□□□□□□□
□□□□□□□□□□□/□□□□□□
□□□□□□□□□□□□□□□□□

6.2.4 LSA设计假说的证伪路径

假说H6: 摩擦调制的有效性

□□:
L2□□□□□□□□□□□(CFE)□□□□□□□□

□□□□:
- □□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□

□□□□(A/B□□):
- □□□: □□F=0.6(□□□□)
- □□□: □□CFE(□□□□□□□□)
- □□: 12□□□AVP□□□□□□□□□□□□□□□□
- □□: □□□AVP□□□□□,□□□□□□□□□□

□□□□□:□□CFE□□□□□□□□,□□□□□□□□□□□□□□□□□

□□□□□□□□□□□,□H6□□□□

□□□□:
- □□□□□,□□□"□□□□
- □□□□□□□□□□□□□□□□□

假说H7: 能力向量的可建模性

□□:
□□□□□□□□□□□□□□□" (<20□)□□□□□

□□□□:
- □□□□□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□

□□□□:

- □□1000+□□□□□□□□
- □□□□□(PCA, VAE)□□□□□□C(t)
- □□: C(t)□□□□□□□□□□□□□□□
- □□: □□□□>30%

□□<10%,□H7□□□□

□□□□□□:
- □□□□□□□□□□□□
- □AI□□□□□□□□□□□□□□

6.2.5 社会层假说的证伪路径

假说H8: 代际能力鸿沟

□□:
T_2□(2015□□□□,AI□□□)□□AI□□□□□□
□□□□□T□□(1980-2000□)□

□□□□:
- 2035-2040□□□□□□□□:
 T_2□□T□□□□□□□□□□(Cohen's d < 0.3)

□□□□(□□□□□□):
- □2025□□□□□:
 * T_0□(□□25-45□): □□□□□□
 * T_1□(□□10-25□): □□□□
 * T_2□(□□0-10□): □□□□□2040□
- □□: □□□□□□□□(□□□□□□□□□□□□)
- □□□□: □□□□□□□□□□□
- □□: T_2□□□AI□□□□□□ < T_0□

□□□□□□□,□H8□□□□

□□□□:
- □□15-20□□□□□□□□□□
- □CET□□□□□□□□□

S-AVP□□□□:
S-AVP□□□□"□□□□□□□□□□□□□□,
□□□□□□□□□□□□□□";□□□□□□□
"□□□□/□□□□",□□□/□□□□□□□□

元认知反思:

> 上述8个假说代表CET理论的核心可证伪预测。它们不是全部,但是最关键的。我们热切期待其中一些假说被证伪——理论的进步往往来自于发现我们错在哪里。如果所有假说都被验证,那可能意味着我们的预测过于保守或模糊。真正有

6.3 未来研究议程: 三个时间尺度

6.3.1 短期研究”（1–3年): 验证核心机制

优先级1: AVP协议的标准化

目标: 将第三章的AVP协议转化为可复现的测量工具

关键任务:

1. 跨领域校准:

- 在编程、写作、数学、语言学习等5个领域实施AVP
- 确定各领域的δ、W、摩擦参数
- 发表《AVP测量手册》(类似IRT测量标准)

2. 信度与效度验证:

- 重测信度: 同一用户多次测量的一致性
- 效标效度: AVP是否预测真实场景表现
- 构念效度: AVP是否真正测量”独立能力

3. 开源工具包:

- 发布AVP测量软件” (支持自动化测试)
- 提供任务库(至少100个标准化任务)
- 建立社区数据共享平台

预期产出: - 3 - 5篇实证论文(各领域的AVP验证) - 开源Github项目(星标>1000) - 至少5个独立团队复现

预注册与开放科学: - 重要实验预注册(OSF/AsPredicted) - 负结果也入库(避免发表偏倚) - 提供最小可复现包(题本、评分

全球验证机会: 全球范围的实证研究为CET理论提供了重要的验证机会。对五国教师和学生的调查显示，缺乏适当引导的AI使
et al., 2025）。这强调了EML设计中”渐进撤出原则的重要性，也为H3假说的跨文化验证提供了初步证据。未来研究应在不

优先级2: EML参数的实验优化

目标: 确定50 - 70%成功率”、“S4→S1→S0削减是否最优” (工作假设)

关键任务:

1. 摩擦参数实验:

- 5×5×5因子设计: 成功率区间、削减速度、任务类型
- N=500 - 1000用户,12周追踪
- 测量: AVP通过率、学习曲线、用户满意度

2. 削减曲线优化:

- 对比: 线性、指数、S型、阶梯式削减
- 识别”过快削减的预警信号
- 个性化削减策略的收益分析

3. 多模态摩擦:

- 探索: 完整性摩擦vs抽象性摩擦vs延迟性摩擦
- 不同任务类型的最优摩擦组合

预期产出: - 数据驱动的EML参数指南 - 摩擦设计模式库 - 可能推翻或修正当前参数

优先级3: 小规模LSA原型

目标: 实现L1+L2+L3的MVP,在单一领域验证

关键任务:

1. 选择垂直领域:

- 建议: 编程教育” (数据易获取、能力易测量)
- 备选: 写作辅助、数学学习

2. 开发最小功能:

- L1: 集成GPT-4 API
- L2: 固定摩擦模式(完整性摩擦)+预设削减曲线
- L3: 嵌入式微测试+黄色/红色预警

3. 用户研究:

- 招募100–200名用户
- 6个月纵向追踪
- 收集定量(AVP通过率)+定性(访谈)数据

预期产出: - 工作原型(可演示) - 案例研究论文 - 识别实施中的关键挑战

6.3.2 中期研究(3–5年): 跨尺度扩展与理论整合

方向1: 团队与组织层的实证研究

目标: 验证第四章的T-AVP/O-AVP框架

关键任务:

1. T-AVP大规模研究:

- 与50–100个软件团队合作
- 实施”周五无AI日”+集体拔线演练
- 追踪团队知识流动、协作模式、T-AVP得分
- 识别团队层面的成功模式

2. O-AVP中断模拟:

- 与20–50个组织合作
- 实施48小时AI中断演练” (伦理委员会批准)
- 测量恢复时间、应急能力、组织韧性
- 验证O-AVP与恢复速度的关系

3. 跨尺度机制研究:

- 追踪I-AVP→T-AVP→O-AVP的传导路径
- 识别”涌现性断裂” (个体通过但团队失败)
- 建立多层次因果模型

预期产出: - T-AVP/O-AVP的标准化协议 - 组织能力建构的最佳实践手册 - 跨尺度理论的精细化模型

方向2: 跨文化适应性研究

目标: 验证CET在非WEIRD文化中的适用性

关键任务:

1. 概念等值性验证:

- 独立能力在不同文化中的语义差异
- AVP判据的文化公平性评估
- 需要哪些本土化调整?

2. 多文化对照研究:

- 至少3个文化群体(如北美、东亚、拉美)
- 相同任务、相同协议
- 比较AVP通过率、用户体验、长期效果

3. 文化特定机制探索:

- 集体主义文化下的T-AVP是否更有效?
- 高情境文化下的有益摩擦”定义是否不同?
- 识别普适性vs文化特异性的边界

预期产出: - 跨文化AVP校准指南 - 文化嵌入性理论的精细化 - 可能推翻或修正”普适性声称

方向3: 神经科学整合

目标: 为CET提供神经层面的验证

关键任务:

1. AI使用的神经影响:
   - fMRI研究: 长期AI使用vs非使用者的大脑差异
   - 是否有类似”GPS效应” (海马体萎缩)的证据?
   - 有益摩擦vs零摩擦的神经激活模式差异
2. 能力退化的神经标记:
   - 识别能力退化的早期神经信号
   - 建立神经→行为的预测模型
   - 为AVP测试提供生物标记补充
3. 可塑性窗口研究:
   - 不同年龄段的能力建构可塑性
   - $T_0/T_1/T_2$代的神经差异
   - 为代际鸿沟假说提供神经证据

预期产出: - CET的神经科学基础 - 能力退化的生物标记 - 跨学科整合理论

6.3.3 长期研究(5–10年以上): 社会影响与理论演化

方向1: 代际纵向研究

目标: 验证S-AVP和代际鸿沟假说(H8)

关键任务:

1. $T_0/T_1/T_2$队列建立:
   - 从2025年开始追踪3个代际队列
   - 每2年一次标准化能力测试
   - 控制教育、社会经济地位等混淆变量
2. 关键时间点测量:
   - 2030: T_1代(15–30岁)中期评估
   - 2035: T_2代(10–20岁)初步评估
   - 2040: 三代对比,验证代际鸿沟假说
3. 社会层面干预研究:
   - 政策实验: 引入EML原则的教育改革
   - 对照设计: 实验区vs对照区
   - 评估: 干预是否缓解代际鸿沟

预期产出: - 15–20年的纵向数据集 - 代际能力演化的决定性证据 - 为教育政策提供实证基础

方向2: AI能力演化的理论适应

目标: 随着AI能力提升,调整CET理论边界

关键任务:

1. 能力本体论的动态更新:
   - 每5年重新定义”核心人类能力
   - 区分永恒价值能力”vs时代依赖能力
   - 建立能力分类的动态框架
2. AGI情境的理论推演:

- 如果出现通用人工智能,CET如何适应?
- “独立能力”vs”增强能力的边界重新讨论
- 可能需要理论范式转移

3. 理论演化的三层同心圆:

- 内层”（核心原则）: AVP/EML的适用性验证
- 中层(参数): δ、W、摩擦参数的持续校准
- 外层(实现): LSA的技术更新

预期产出: - CET 2.0理论(如果需要根本性修正) - 动态能力本体论框架 - 理论演化的元模型

方向3: 开放议题探索

议题A: S层的大过滤器脆弱性(见第四章4.4.3节)

□□:
- □□□□□□□□□□□□□□□□?
- □□□□□□□□□□□□□□□□?

□□□□:
- □□□□□□□□(□□□□□□□□)
- □□□□□□□(□□□□□□□□□□□□□□)
- □□□□□□□(□□□□□□□□□□□□)

□□□□:
□□□□□□□□□□□□□,CET□□□□□□□□,
□□□□□"□□□□□□□□□,□□□□□□□□□

议题B: 人机融合的哲学边界

□□:
- □□□□□□□□,"□□vs"□□□□□□□□?
- □□□□"□□□□□□□vs□□□□□□□□□□?

□□□□:
- □□□□: □□□□□□□
- □□□□: □□□□□□□□□□□□□
- □□□□: "□□□□"□□□□□□

□□□□:
□□□□CET□□□□□□□□□□□□□

6.4 研究伦理与开放科学承诺

6.4.1 伦理原则

CET理论的验证涉及人类受试者,必须严格遵守研究伦理:

1. 知情同意: 所有AVP测试、拔线演练必须获得参与者知情同意
2. 无伤害原则: 拔线测试不得用于高风险任务”（如医疗、飞行）
3. 隐私保护: AVP结果个人隐私,不得用于雇佣/教育歧视
4. 公平性原则:

- 为残障、弱势群体调整任务形式而不降低挑战强度
- 评估以相对提升而非绝对水平为准
- 守恒的挑战预算:任务可及性↑,但认知挑战度保持不变

5. 撤回权: 参与者可随时退出研究,不受惩罚

6.4.2 开放科学承诺

为促进理论验证和批判,我们承诺:

1. 数据开放:
   - 匿名化数据集公开发布(符合隐私法规)
   - 原始数据存储在开放平台(如OSF)
   - 对敏感场景,可采用差分隐私或联邦学习的开源实现以降低再识别风险
2. 方法透明:
   - 详细研究协议预注册
   - 在预注册中指定主要终点(如P – B_0效应量)与次要终点(保持/迁移/成本),并提交多重比较校正方案,避免事后指标淘
   - 统计代码开源(GitHub)
   - 负结果报告(避免发表偏倚)
3. 工具开源:
   - AVP测量软件开源
   - LSA参考实现开源
   - 题库与评分rubric公开
4. 协作邀请:
   - 欢迎独立团队复现
   - 鼓励跨文化验证
   - 接受批评性检验

6.5 结语: 理论的生命在于批判与演化

CET理论诞生于2025年——AI能力爆发、人类认知面临重构的关键时刻。我们提出这个理论,不是因为我们相信它是”完美的

本章揭示的六大局限提醒我们: CET是在特定技术、文化、认识论背景下的产物。它的价值不在于永恒正确,而在于:

1. 提供可证伪的预测: 8个核心假说都有明确的证伪条件
2. 承认不确定性: 所有参数都标注为”工作假设,需校准
3. 邀请批判: 我们期待被证伪,而非害怕被证伪
4. 指明研究方向: 三个时间尺度的研究议程为后续工作者铺路
5. 保持演化能力: 三层同心圆架构允许理论随证据更新

最后的呼吁:

如果你是研究者,请: - 挑战CET的假说,用严格的实证研究证伪或验证 - 在不同文化、不同领域复制CET的核心发现
- 提出竞争性理论,推动领域进步

如果你是开发者,请: - 将EML原则融入AI工具设计 - 测量并公开你的产品的AVP表现 - 参与开源社区,共建能力建构优先的AI

如果你是教育者/管理者,请: - 在组织中试点AVP评估 - 关注团队和组织的认知韧性 - 平衡效率与能力建构的长期价值

如果你是政策制定者,请: - 关注CET揭示的长期风险”(代际鸿沟、认知公地悲剧) - 支持跨学科的纵向研究
- 建立AI工具的能力建构影响评估机制

科学理论不是圣经,而是工具。 CET的最大价值不在于”给出答案,而在于”提出正确的问题”。我们相信,即使CET的某些假
—推动我们更深入地思考人类与AI共存的未来。

理论的生命在于被讨论、被检验、被超越。 我们期待那一天的到来。

第7章 - 术语与符号系统

文档版本：v1.0 | 已应用补丁包v1.0 用途：核心概念速查、参数一致性维护、术语标准化 性质：只读参考文档”(唯一修

1. 核心概念固定锚点（B1-B5）

使用说明： 以下内容为第3.0.2－3.0.6节之逐字转录，唯一修订源在第3章；如需更新请先修改3.0节。本处为只读副本，供

B1 | AVP定义锚点

反脆弱性验证原则（Antifragility Validation Principle, AVP）：以拔线测试检验协作是否促进独立能力。

判据：$P_2 \geqslant B_0 + \delta$

其中：－B_0：使用AI前的独立基线能力－P_2：协作一段时间后，在拔线窗口W=4－8周（默认6周）内的独立表现 －δ：Cohen’s d ⩾ 0.3 或 ⩾10%（working assumption））－P_1（协作期表现）不参与最终判定**

Canonical source: 见3.0.2节（本处为只读副本）

B2 | EML定义锚点

内共生最小法则（Endosymbiotic Minimal Law, EML）：构成认知内共生的设计必要条件为：

(1)有益认知摩擦：使用户处于最优挑战区（群体级工作假设成功率50－70%，需跨领域/任务校准，个体自适应）

(2)系统性支持削减：AI支持强度按既定削减曲线从S4→S1→S0

二者为联合充分的设计条件，但最终仍需AVP（$P_2 \geqslant B_0 + \delta$）作为验收必要条件。

Canonical source: 见3.0.3节（本处为只读副本）

B3 | LSA-F功能分层锚点

LSA-F（功能分层）：－L1 知识整合层：基础AI能力接入－L2 状态建模层：摩擦设计与支持削减 －L3 摩擦校准层：能力监测与预警 －L4 元认知协调层：多尺度编排与伦理治理

支持档位栈（S4→S1→S0）用于表达支持强度，与LSA-F为正交维度。

Canonical source: 见3.0.4节（本处为只读副本）

B4 | 最优挑战区锚点

最优挑战区：为促成长期保持与迁移，系统应将任务难度/提示强度自适应调至成功率50－70%（工作假设，需跨领域/任务校准，随任务与个体校准）；>85%近似卸载、<30%易致挫败。

Canonical source: 见3.0.5节（本处为只读副本）

B5 | 边界条件锚点

边界条件：本理论适用于能力增强型人机协作；补偿性外骨骼（如残障辅助、超越生理极限的设备）不适用此判据。所

Canonical source: 见3.0.6节（本处为只读副本）

2. 核心参数速查表

参数名称	符号/公式	默认值/范围	理论依据	校准方向
最小提升阈值	δ	⩾Cohen’s d ⩾ 0.3 或 ⩾10%（working assumption）（工作假设，需跨领域/任务校准）	Cohen’s d中等效应量	程序性任务可能更低（0.2－0.3）；创造性任务可能需更高（0.4－0.5）
拔线窗口	W	4－8周（默认6周）（工作假设，需跨领域/任务校准）	能力巩固的经验性时间	简单技能2－4周；复杂技能8－12周；长期学习12－24周
最优挑战区	成功率区间	50－70%（工作假设，需跨领域/任务校准）	心流理论+最近发展区（ZPD）（工作假设，需跨领域/任务校准）	个体自适应；任务特定校准；文化敏感性调整
支持削减起点	S0	0.8（80%支持）（工作假设，需跨领域/任务校准）	[illegible]	高风险任务可从0.6开始；新手友好任务可从0.9开

参数名称	符号/公式	默认值/范围	理论依据	校准方向
安全支持下限	S_min	≥0.2（不低于20%）（工作假设，需跨领域/任务校准）	[illegible]	[illegible]可降至0.1；新手建议不低于0.3
削减速率	λ	按任务调整（工作假设，需跨领域/任务校准）	个体学习曲线差异	线性/指数/S型曲线择优
团队AVP阈值	T-AVP	≥0.7（群体级）（工作假设，需跨领域/任务校准）	组织韧性研究	不同团队规模需调整；关键任务建议≥0.8
组织AVP阈值	O-AVP	告警≥0.70，目标≥0.85（工作假设，需跨领域/任务校准）	[illegible]	[illegible]风险容忍度调整；48h窗口可改为24h/72h
代际窗口期	$T_0 \to T_1 \to T_2$	10年（2025－2035）（概念占位符）	技术代际影响推测	纵向研究校准；跨文化验证

注：所有参数均为概念工作模型，需通过实证研究跨领域校准。不同任务类型、用户群体、文化背景可能需要显著调整。

3. 缩写速查表

缩写	全称	中文	首次出现
AVP	Antifragility Validation Principle	反脆弱性验证原则	1.3节
EML	Endosymbiotic Minimal Law	内共生最小法则	1.3节
LSA	Layered Symbiosis Architecture	分层共生架构	1.3节
LSA-F	LSA Functional Hierarchy	LSA功能分层	3.0.5节
CFE	Cognitive Friction Engine	认知摩擦引擎	5.2节
SGS	Support Graduation Scheduler	支持削减调度器	5.3节
AVP-TM	AVP Telemetry Module	AVP遥测模块	5.4节
MSO	Multi-Scale Orchestrator	多尺度编排器	5.5节
I-AVP	Individual AVP	个体反脆弱性验证	4.1节
T-AVP	Team AVP	团队反脆弱性验证	4.1节
O-AVP	Organizational AVP	组织反脆弱性验证	4.2节
S-AVP	Societal-level indicators	社会层认知资本指标	4.3节
BCI	Business Continuity Index	业务连续性指数	4.2节
ICR	Independent Completion Rate	独立完成率	4.2节
IRT	Item Response Theory	项目反应理论	附录A
RCT	Randomized Controlled Trial	随机对照试验	6.2节
DID	Difference-in-Differences	差分中的差分	6.2节
ZPD	Zone of Proximal Development	最近发展区	2.3节

4. 核心评估原则

4.1 等效努力原则（Equivalent Effort Principle）

在AVP测试中，对不同能力水平或特殊需求的用户，应遵循等效努力”而非”等值任务原则：

三句口径”（唯一标准表述）：

1. 调整任务形式而不降低挑战强度
   - 示例：视力障碍者可使用语音测试，但问题难度保持不变
   - 原则：改变呈现方式，不改变认知负荷
2. 评估以相对提升为准
   - 判据：$P_2 \geqslant B_0 + \delta$（相对于个人基线的提升）
   - 而非：$P_2 \geqslant$ 某个绝对标准
   - 理由：尊重个体差异，避免一刀切
3. 挑战预算守恒
   - 核心：认知负荷总量保持一致
   - 方法：通过任务分解、时间调整、辅助工具等方式平衡

- 目标：确保不同用户面临等效的认知挑战

应用场景： - 残障人士的能力评估（见3.1.4节） - 跨文化任务等值性校准（见6.1.1节） - 不同年龄群体的适应性调整 - 教育背景差异的补偿设计

引用说明：其他章节提及等效努力时，使用（见第7章 § 4.1等效努力原则）“引用此处。

4.2 Goodhart防护原则

核心问题：当度量成为目标时，它就不再是好的度量（Goodhart’s Law）

CET的应对策略：

1. AVP判据的非KPI化
   - AVP分级（Basic/Retention/Transfer）仅用于质量分层
   - 禁止将AVP分数用于人事考核、绩效排名、资源分配
   - 任何涉及利益分配的场景都不得使用AVP作为唯一判据
2. 固定脚注模板（在所有阈值表格下使用）： > 注（Goodhart防护）：本表/分级仅用于方向与分层；不得下推为KPI。最
3. 监测与预警的分离
   - 监测数据（如C”(t)能力向量）仅用于系统改进
   - 不与个人利益挂钩
   - 匿名化处理，保护用户隐私

反面案例（见CET9附录A FAQ Q5）： 某公司将AVP用于晋升评估，导致用户人为操纵基线、拔线期违规使用AI，完全失效。

5. 核心术语中英对照

5.1 理论核心概念

中文术语	英文术语	缩写	核心定义（简版）	详见
认知内共生	Cognitive Endosymbiosis	-	AI作为伙伴，通过摩擦与削减促进能力提升	3.4节
认知外骨骼	Cognitive Exoskeleton	-	过度依赖AI导致独立能力退化的病理模式	3.0.6节
反脆弱性验证原则	Antifragility Validation Principle	AVP	通过拔线测试验证能力是否提升	3.0.2节
内共生最小法则	Endosymbiotic Minimal Law	EML	有益摩擦+系统削减的设计必要条件	3.0.3节
伙伴式主体性	Partner-like Agency	-	AI作为认知伙伴的理想角色定位	3.4节
有益认知摩擦	Beneficial Cognitive Friction	-	适度挑战（50 - 70%成功率）促进能力增长	3.2节
系统性支持削减	Systematic Support Reduction	-	AI支持强度按曲线递减（S4→S1→S0）	3.3节
拔线测试	Unplugged Test	-	在无AI环境下测量独立能力	3.1节

5.2 测量相关术语

符号/术语	含义	单位/范围	备注
B_0	基线能力	任务特定评分	使用AI前的独立表现
P_1	协作期表现	任务特定评分	不参与AVP判定
P_2	拔线后能力	任务特定评分	AVP判据的核心指标
δ	最小提升阈值	Cohen’s d ≥ 0.3 或 ≥10%（working assumption）	工作假设，需跨领域/任务校准
W	拔线窗口	4 - 8周（默认6周）	工作假设，需跨领域/任务校准
C(t)	能力向量	多维向量	随时间变化的能力状态

符号/术语	含义	单位/范围	备注
F	摩擦参数	0–1	任务难度/支持强度调节
S(t)	支持强度函数	0–1	从S0递减至0的曲线
λ	削减速率	按任务定义	控制削减速度的参数

5.3 架构相关术语

术语	层级	主要功能	详见
L1 基础AI能力层	第1层	接入AI模型、知识库	5.1节
L2 摩擦与削减层	第2层	CFE摩擦引擎 + SGS削减调度	5.2–5.3节
L3 监测与反馈层	第3层	AVP-TM遥测 + 预警系统	5.4节
L4 编排与治理层	第4层	多尺度协调 + 伦理治理	5.5节
S4→S1→S0	支持档位	强→弱的4级支持强度	3.3节

6. 参数登记簿（单一事实源 – Single Source of Truth）

用途：跨章参数一致性速查表。所有参数的”唯一权威版本登记在此。

参数符号	默认口径	维护位置	首次定义	跨章引用
AVP判据	$P_2 \geq B_0 + \delta$	3.0.2节	3.0.2节	1.3/4.1/5.4/6.2
δ阈值	Cohen’s d ≥ 0.3 或 ≥10%“（working assumption）（工作假设，需跨领域/任务校准）	3.0.2节	3.0.2节	全文
W窗口	4–8周（默认6周）（工作假设，需跨领域/任务校准）	3.0.2节	3.0.2节	3.1/4.1/附录A
最优挑战区	50–70%成功率（工作假设，需跨领域/任务校准）	3.2.1节	3.2.1节	5.2/附录A
S”(t)起点	0.8（80%支持）（工作假设，需跨领域/任务校准）	3.3.1节	3.3.1节	5.3
S_min安全下限	≥0.2（不低于20%）（工作假设，需跨领域/任务校准）	3.3.2节	3.3.2节	5.3.3
削减速率λ	任务特定（工作假设，需跨领域/任务校准）	3.3.1节	3.3.1节	5.3
T-AVP阈值	≥0.7（群体级）（工作假设，需跨领域/任务校准）	4.1.3节	4.1.3节	4.1
O-AVP阈值	告警≥0.70，目标≥0.85（工作假设，需跨领域/任务校准）	4.2.3节	4.2.3节	4.2/6.2
48h演练窗口	48小时（可调24/72h）（工作假设，需跨领域/任务校准）	4.2.2节	4.2.2节	4.2
代际窗口	10年（2025–2035）（概念占位符）	4.3.2节	4.3.2节	6.3.3
能力向量维度	5–20维（探索性假设）	5.4.3节	5.4.3节	5.4

使用规则： 1. 修改流程：如需调整任何参数的默认值，必须先在维护位置对应章节修改，然后更新本表 2.
引用格式：引用参数时使用”（见X.Y节，参数登记簿第7章 § 6） 3. 版本控制：本表随主文同步更新，版本号与论文版本

口径守恒承诺： 本表是全文参数的唯一真实来源（Single Source of Truth）。如发现跨章参数不一致，以本表为准，并回

7. 文档维护协议

7.1 更新流程

□□□□ → □□□□□□□□ → □□□7□□□□□ → □□□□□□□

7.2 一致性检查命令

正则搜索检查（建议使用VSCode等工具）：

1. 检查禁词：" (CEET|CST|AHT|EWAT|RCE|EPCII|□□BUFF)
2. 检查参数标签：搜索数字参数，确保都有”（工作假设，需跨领域/任务校准）”
3. 检查锚点引用：确保所有见X.X节都能定位
4. 检查Goodhart脚注：所有阈值表都有固定脚注

7.3 版本历史

- v1.0（2025－10-01）：初始版本，应用补丁包v1.0
- [未来版本记录在此]

CET8 － 图表库与对照表

文档版本：v1.0 | 已应用补丁包v1.0 用途：核心概念可视化、跨章节图表统一管理 性质：概念图示集（所有参数标注”工作

图表库使用说明

设计标准”（确保可访问性与可复用性）：

1. 黑白打印可辨：使用线型/纹理区分，最多4色，无3D效果
2. 每图必含：
   - 图号（如”图B.1）
   - 1－2句目的说明
   - 概念图”或”数据图标识
   - 如含参数 → 标注”（工作假设，需跨领域/任务校准）
3. 格式规范：图题在图下方，表题在表上方

示例标注格式：

□B.3 □□□□□□□□□□□/□□/S□□[□□□]
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□/□□□□□

B.1 外骨骼vs内共生核心对照表

表B.1：认知外骨骼vs认知内共生对照表

维度	认知外骨骼（病理模式）	认知内共生（目标模式）
设计哲学	代替人类认知（替代路径）	增强人类认知（能力建构）
摩擦策略	零摩擦（即时满足）成功率>85%	有益摩擦（适度挑战）成功率50－70%*
时间导向	短期效率最大化（P_1优化）	长期能力建构（P_2优化）
支持削减	无削减（持续依赖）S” (t) = 常数	系统性削减S(t): S4→S1→S0→0*
验收标准	$P_1 > B_0$（有AI更好）	$P_2 \geq B_0 + \delta$（拔线更强）*
AVP结果	$P_2 < B_0$（能力退化）	$P_2 \geq B_0 + \delta$（能力提升）
神经趋势	相关脑区萎缩（理论推断）**	神经可塑性增强（理论推断）**
典型案例	过度依赖GPS导致空间定向能力下降	编程教学平台的摩擦式学习

注：本表仅用于方向与质量分层；不得下推为KPI。最终判定以AVP主判据（见3.0.2节）为准。所有参数为工作假设，需跨领

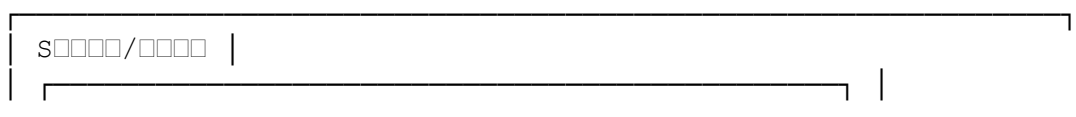
*（工作假设，需跨领域/任务校准）

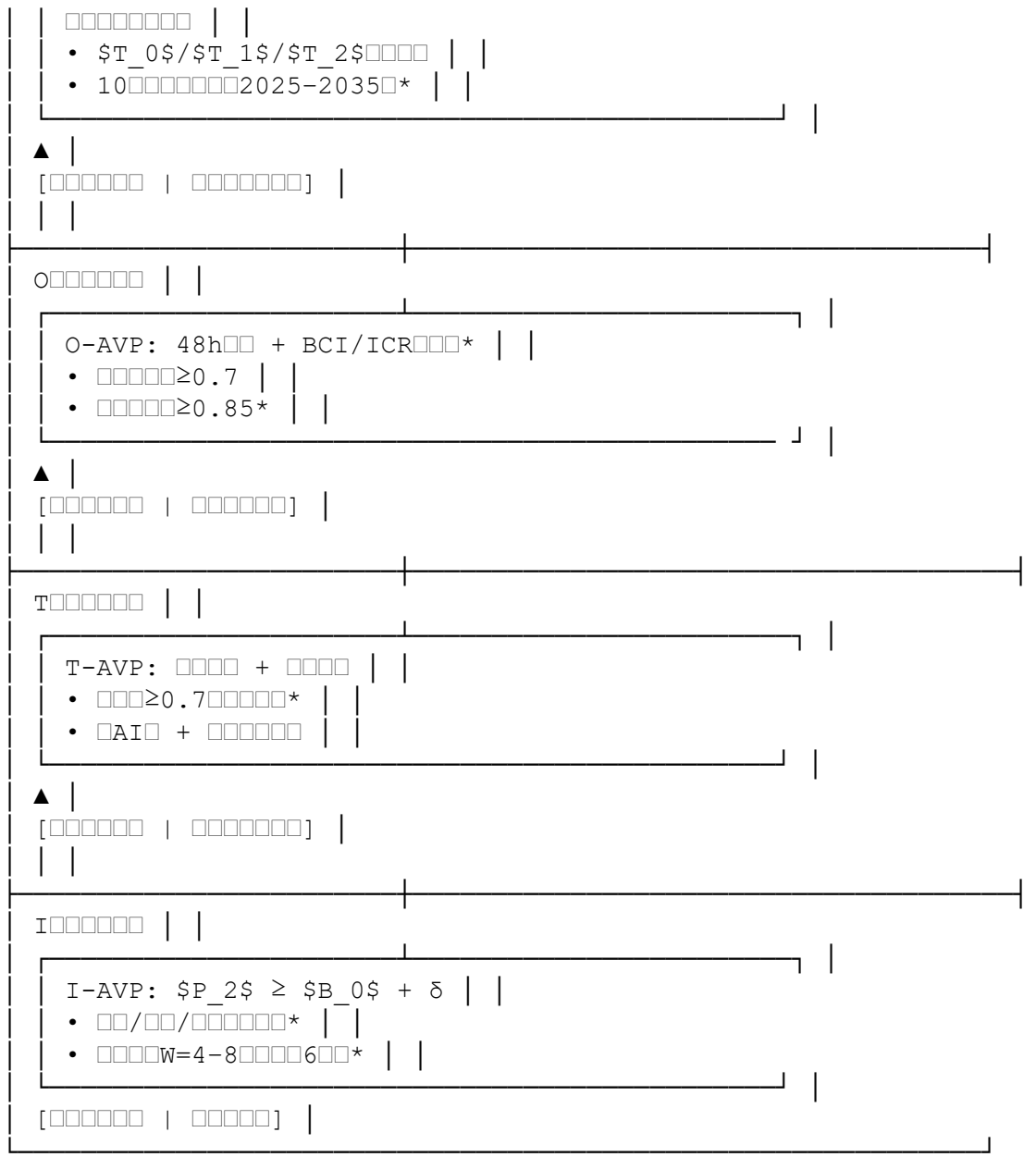
注：该行属理论性推断**，需神经科学实证验证（见6.3.2节研究议程）。当前证据仅为跨领域类比（如GPS与海马体研究），

注:仅用于质量分层,不得作为KPI”;最终判定以AVP主判据” (见3.0.2节)为准。

B.2 跨尺度验证框架（I→T→O→S）

图B.2：四层跨尺度AVP验证架构 [概念图]





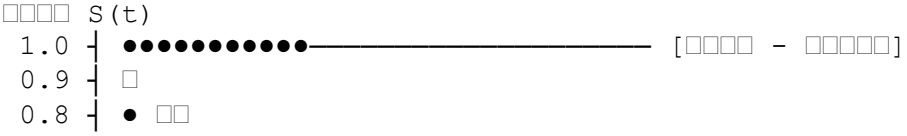
*□□□□□□□□□□□/□□□□□

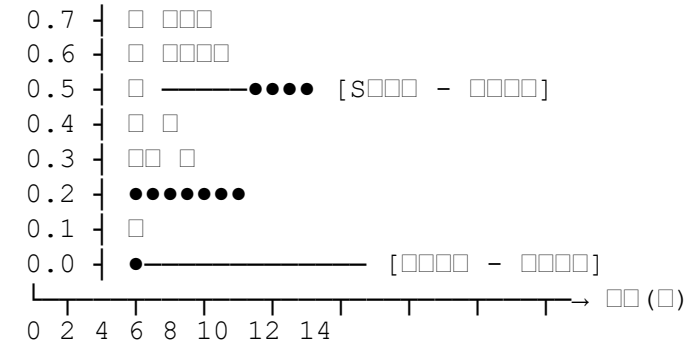
说明： 展示I/T/O/S四个尺度的AVP验证逻辑。箭头表示验证结果的传导方向（个体→团队→组织→社会）。每个层级都有独

关键洞察： - 非线性传导：I-AVP通过 ≠ T-AVP必然通过（涌现性） - 尺度特异性：每层有特定测量工具和阈值
- 时间尺度差异：I层周级，T层月级，O层季度级，S层年代级

B.3 支持削减曲线对比

图B.3：三种支持削减策略对比［概念图］





□□□□□

□□□ □□□□□S(t) = S0(1 - λt)
●●● □□□□□S(t) = S0·e^(-λt)
□□ S□□□□S(t) = S0 / (1 + e^(k(t-t□)))
── □□□□□□□□□□S(t) = S0□□□□□

□□□□□□□□□□□□□□□/□□□□□□□

- S0 = 0.8□80%□□□□□
- λ = 0.1-0.2□□□□□□
- t□ = 6□□S□□□□□□□

说明： 展示三种削减策略的概念差异。实际应用中需要根据任务难度、用户学习曲线、流失率等因素选择合适的曲线类型。

过快削减预警信号（见5.3.3节）： - 成功率持续<30%超过2周 - 用户主动求助频率激增 - 任务放弃率>30%

B.4 最小AVP实验流程

图B.4：AVP验证标准流程 [流程图]

□□□□

```
  │
  ├─→ [$T_0$] □□□□□□AI□
  │  ↓
  │  □□ $B_0$
  │  ↓
  ├─→ [□□□□]
  │  ├─→ □□□□EML□□□
  │  │  - □□□□□50-70%□□□□*
  │  │  - □□□□□□S4→S1→S0□*
  │  │  ↓
  │  │  [□□□□4-12□]*
  │  │  ↓
  │  │  □□ $P_1$□□□□
  │  │  ↓
  │  │  [□□□□W=4-8□□□□6□□]*
  │  │  - □□□□AI
  │  │  - □□□□□□
  │  │  ↓
  │  │  [$T_3$] □□□□AI□
  │  │  ↓
  │  │  □□ $P_2$
  │  │
```

```
  │  └─→ □□□□□□AI/□AI□
  │  ↓
  │  [□□□□]
  │  ↓
  │  □□ $P_2$'
  │
  ├─→ [AVP□□]
  │  │
  │  ├─→ □□□□$P_2$ ≥ $B_0$ + δ ?
  │  │  YES → AVP□□□□□□□
  │  │  NO → AVP□□□□□□□□□
  │  │
  │  └─→ □□□□□□$P_2$ vs $P_2$'
  │  □□□ = ($P_2$ - $P_2$') / SD
  │
  └─→ □□□□
  - AVP□□□
  - □□□□□
  - □□□□□□□□/□□/□□□*

  *□□□□□□□□□□/□□□□□
```

说明： 最小可复现的AVP验证流程。关键控制点包括任务等值性（T_0与T ）、拔线窗口完整性、违规监测、盲评。

简化版本（24h拔线轻量测试）： 适用于快速试点或低风险场景，流程见附录A § A.6。

B.5 能力向量C(t)概念模型

图B.5：能力向量的多维动态表征 [概念图]

```
  □□□□ (c□)
  ↑
  0.8 /|\ 1.0
  / | \
  0.6 / | \ □□□ (c□)
  / | \
  □□ ————————●———————— □□
  □□ $B_0$ / \ $P_2$ □□
  (c□) / \ (c□)
  / \
  / \
  / \
  □□□□ (c□)
```

□□□□□

t=0□B_0□□[0.5, 0.6, 0.4, 0.3, 0.5]

t=8□□P_1□□□AI□□[0.7, 0.9, 0.6, 0.8, 0.6]

t=14□□P_2□□□□□□[0.7, 0.8, 0.7, 0.6, 0.7]

□□□□□

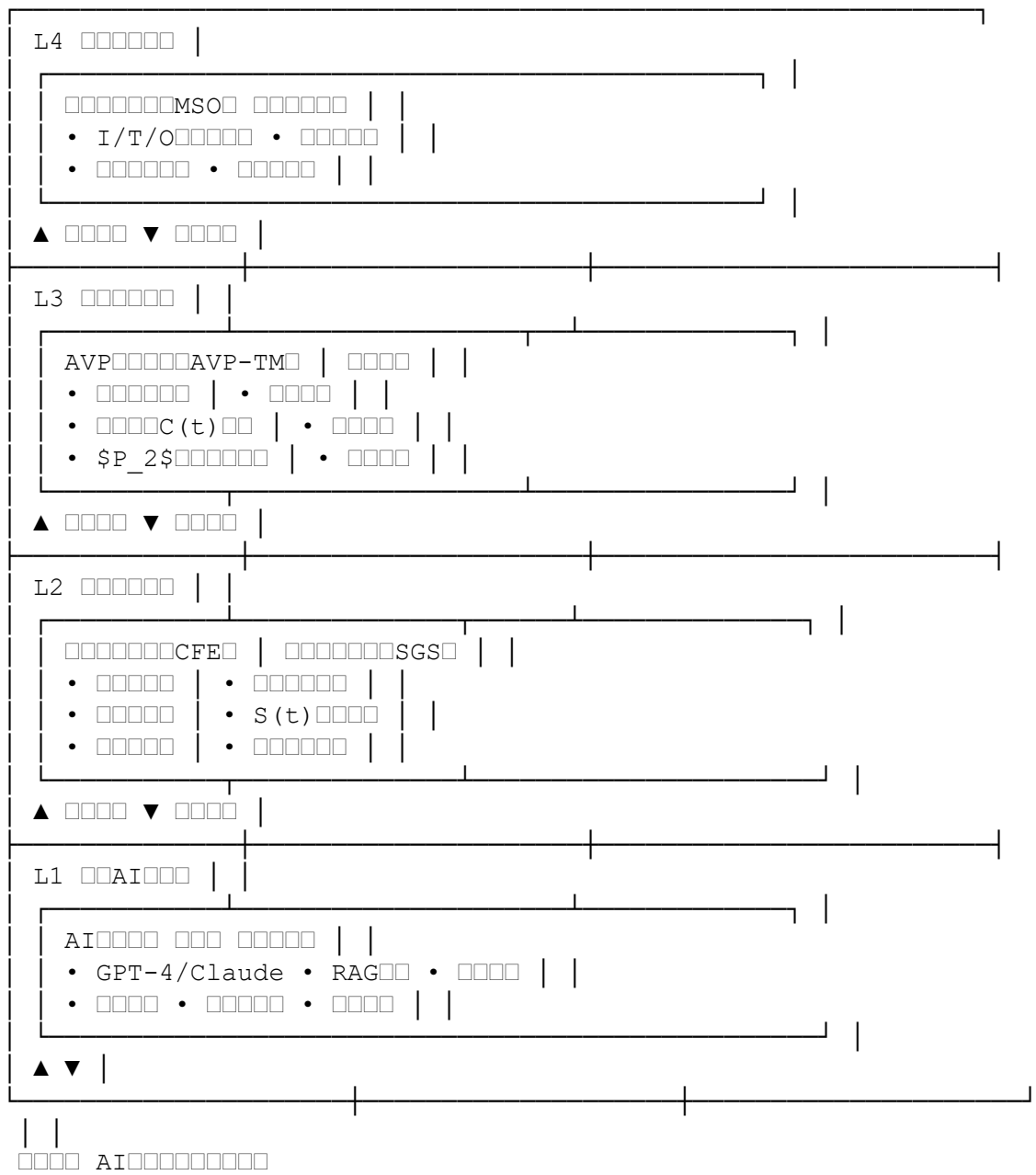
- P_1□□□□□"□□□□"□AI□□□
- P_2□□□□□□□□□
- □□□P_2 ≥ B_0□□□□□□□□□□
- □□□P_2□□□□□<B_0□□□□□□□

说明： 能力向量C”（t)是多维的（5－20维，探索性假设）。不同任务需要定义不同的能力维度。重要的是P 在关键维度上

注：维度数量和权重需要领域专家定义和实证验证（见5.4.3节）。这是概念模型，非精确测量工具。

B.6 LSA四层架构示意

图B.6：分层共生架构（LSA）［概念图］



□□□□□□□□□□□□F□S(t)□□□□□□□□□□□□□□□□□□□□/□□□□□

说明： LSA的核心设计原则是能力建构优先。L2层的摩擦与削减是关键创新，L3层的监测确保AVP闭环，L4层的治理防止滥用

关键数据流： - 上行：用户行为 → 遥测 → 能力建模 → 策略调整 - 下行：策略指令 → 参数调节 → 响应调制 → 用户体验

与传统AI助手的区别： - 传统：只有L1层（直接响应） - LSA：四层完整架构（能力建构+监测+治理）

B.7 图表素材技术规范（便于复用与派生）

用途：为出版、演示、教学提供图表源文件信息

图号	图表名称	推荐格式	逻辑结构	关键元素	备注
表B.1	外骨骼vs内共生对照表	Markdown/SVG表格	8行×3列对照表	设计哲学、摩擦、AVP结果、可持续性	附录5.1引用
图B.2	I→T→O→S跨尺度框架	SVG矢量图	4层金字塔/流程图	箭头：验证传导；标注：核心假设	建议用Excalidraw/draw.io
图B.3	削减曲线对比	SVG曲线图	3条曲线（线性/指数/S型）	X轴：时间；Y轴：支持强度（可用	可用Python matplotlib生成
图B.4	最小AVP实验流程	SVG流程图	时间线+分支决策树	T_0→协作期→W→T_3→判定	建议用Mermaid语法
图B.5	能力向量C(t)动态演化	SVG动态图	多维雷达图+时间轴	5-10维能力；3个时间点（$B_0/P_1/P_2$）	可用D3.js或Plotly
图B.6	LSA四层架构图	SVG分层图	4层堆叠+数据流箭头	L1-L4功能模块；双向箭头	参考5.1-5.5节详细描述

注：本表仅用于方向与质量分层；不得下推为KPI。最终判定以AVP主判据（见3.0.2节）为准。所有参数为工作假设，需跨领

注：本表仅用于方向与质量分层；不得下推为KPI。最终判定以AVP主判据（见3.0.2节）为准。所有参数为工作假设，需跨领

技术建议： - 矢量优先：使用SVG格式，确保缩放不失真 - 语义化：图层命名清晰（如”Layer-L1-AI-Capability”） - 参数化：关键数值”（如δ、W）单独图层，便于批量更新 - 开放工具：优先使用开源工具（Inkscape、d - 版本控制：源文件纳入Git，便于协作修改

复用许可： 所有图表采用CC BY 4.0许可，允许修改和再分发，但需： 1. 适当署名原作者 2. 标注修改内容 3. 保持相同许可协议

B.8 图表设计的可访问性自检清单

使用本图表库时，请确保：

视觉可访问性： -[ ]黑白打印后仍可区分关键元素（用线型/纹理，非仅颜色） -[ ]字体大小≥10pt，关键标注≥12pt - [ ] 对比度符合WCAG 2.1 AA标准（对比度≥4.5:1） - [ ] 复杂图表有文字描述补充

认知可访问性： -[ ]每图只传达1-2个核心信息（避免信息过载） -[ ]图例清晰，符号一致（如箭头方向、线型含义） - [ ] 有概念图或数据图”标识（设定读者期望） - [ ] 参数都有”工作假设标注”（避免误读为精确值）

技术可访问性： - [ ] SVG源文件可编辑（非位图截图） - [ ] 图层结构清晰（便于修改特定元素） - [ ] 导出为多种格式（SVG/PNG/PDF）

CET9 - 综合附录系统

文档版本：v1.0 | 已应用补丁包v1.0 用途：可操作测量工具、详细案例、跨学科对话、完整术语索引 性质：实践指南+学术对话+参考手册

附录A：AVP测量协议工具包

A.0 工具包定位与声明

本工具包的性质： - 这是参考模板与原型协议，而非标准化测量工具 - 所有参数均为示意性工作假设，需要根据具体领域校 - 未经大规模实证验证——我们诚实承认这一局限

透明性承诺： 本工具包基于理论推导和文献综述构建，尚未经过跨领域、大样本的系统性验证。我们期待研究社区对其进行

使用者责任： 1. 咨询领域专家确定任务的适当难度 2. 进行预测试确保测量工具的区分度（pilot N≥10） 3. 使用IRT或专家判断确保任务等值性（B_0与P 测试） 4. 记录所有调整决策以供其他研究者参考 5. 高风险场景（医疗、安全）必须使用专业工具，不可直接应用本协议

适用对象： - 研究者：设计AVP验证实验 - 教育工作者：评估AI辅助学习的长期效果 - 组织管理者：评估团队/组织的AI依赖风险（非人事考核） - 高风险决策场景（需专业风险评估方法）

高风险场景豁免声明：

本原型协议不得直接应用于以下高风险场景： - 医疗诊断与治疗决策 - 金融风险评估与信贷决策 - 交通安全与自动驾驶系统 - 司法判决与刑事侦查 - 其他涉及人身安全、财产安全、法律责任的关键决策

这些场景必须使用经过严格验证和认证的专用测评方案，并遵循相应领域的监管要求。

A.1 标准化AVP测量清单

用途：为研究者和实践者提供可复现的AVP测量操作指南

A.1.1 测量前准备清单

□ 1. □□□□□
 □ □□□□□□"□□□□□□□□□□
 □ □□□□□□□□□□□□□□□□□
 □ □□□□□□□□□□□□□□3.0.6□□□□□□□□B5□

□ 2. □□□□□T_0□
 □ □□□□□□□□□□□□□□□□□30-120□□□*
 □ □□□□□□□□N=30□□□N≥50□
 □ □□□B_0□□□□□□□□□□□□□□1-10□□□
 □ □□□□□□□□NASA-TLX□□□□□□

□ 3. □□□□□□□T_3□
 □ □□□□□□□□IRT□□□□□□□□
 □ □□□□2□□□□□□□□□□□
 □ □□□□□□□□□□□pilot N≥10□
 □ □□□□□□□□□□ICC□□>0.7□

□ 4. □□□□
 □ δ□Cohen's d ≥ 0.3 □ ≥10%□*working assumption*□□□□□□□□□□□/□□□□□*
 □ W□□□4-8□□□□6□□□□□□□□□□□□/□□□□□*
 □ □□□□□□□T□□3□□□□□□□□□□□
 □ □□□□□□□IRB/□□□□□□□

*□□□□□□□□□□/□□□□□

A.1.2 测量执行协议

□□1□□□□□□T_0 - □0□□

□□□

- □□□□□AI□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□□□□□□think-aloud□□□
- □□□□□□□□□□□□□□NASA-TLX□□□□□□□

□□□

- □□□2□□□□□□□□□□□□□□□□□□□□
- □□□□□□□□□□ICC >0.7□
- □ICC<0.7□□□□□□□□□□□□□□□□□□
- □□□□□□B□

□□□□□

- □□□□□□/□□□□□□□>80%□<20%□□□□□□□□□□□□□□
- □□□□□□□Z-score > 3□
- □□□□□□□□□□□

□□2□□□□□□T_1 → T_2 - □1-8□□

□□□□□

- □□□□EML□□

- □□□□□□□□□□□50-70%□□□□□□□□□□□/□□□□□□*
- □□□□□□□S" (t)□0.8→0□□□□□□*
- □2□□□□□□□□□10%□□□□□□□

- □□□□□□□AI□□
- □□□□□□□□□□□□□
- □□□□□□□S(t)□□□

□□□□□

- □□□□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□
- □□□□□□□□2□□□□5□□□□
- □□□□□□□□□□

□□□□□□□□□

- □□□□□□□□□□□□50-70%□□□±5%□□□
- □□□□□□□□□□□□□□□□□□S(t)□□□
- □□□□□□□□□□□□□□□□

□□3□□□□□□□W - □9-14□□□□6□□

□□□

- □□□□AI□□□□□□□□+□□□□□
- □□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□□□AI□
- □□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□

- □□□□□□□□□□□□□"□□□□
- □□□□□□□"□□□□
- □□□□□□□□□

□□□□□

- □□□□"□1-2□□□□□□□□□□□□□□□□□□□□□□□□
- □□□□□□≥3□□□□□□□□□□□□□□□□□□

□□□□□

- □□□□>30%□□□□□□□□□□□□□W□□□□□
- □□□□□□ITT□□□□□□□□□□□□□□□□
- □□□□□□□PP□□□□□□□□□□□□□□□

□□4□□□□T_3 - □15□□

□□□

- □□□□□□□□T□□□□□□
- □T□□□□□□□AI□□□□□□□□□□□□□
- □□□□□□□□□□□□□□

□□□

- □□□□□□□□□□□□□□□□□□□□□□
- □□P□□□
- □□B□□P□□□□

□□□

- □□ Δ = P_2 - B_0
- □□□□□Δ ≥ δ̄ → AVP□□
- □□□□□
- □□t□□□□□□□□□□□□□□
- Wilcoxon□□□□□□□□□□□□□□□□
- □□□□Cohen's d = " (P_2 - B_0) / SD_pooled

□□5□□□□□□ - □27□□T_3□3□□□

□□□

- □□□□□□□□
- □□□□□□Retention = P_2□□□ / P_2
- Transfer□□□□□□□□□□□□□□□□

□□□

- □□□□□□□□□□
- □□□□□□□□□Transfer Badge□□3.2.3□□

A.1.3 数据记录模板 CSV格式模板：

```
participant_id,group,age,gender,education,prior_experience,
T0_score,T0_time_min,T0_difficulty,T0_cognitive_load,
P1_avg_score,P1_weeks,usage_frequency,help_requests,
withdrawal_violations,
T3_score,T3_time_min,T3_difficulty,T3_cognitive_load,
delta,avp_pass,retention_rate,transfer_score,
dropout,dropout_reason,notes

P001,EML,25,F,Bachelor,2_years,72,65,7,65,78,8,daily,12,0,85,58,5,55,13,TRUE,0.98,80,
P002,Control,28,M,Master,3_years,68,70,6,60,75,8,daily,25,1,70,72,6,62,2,FALSE,0.93,
P003,EML,23,F,Bachelor,1_year,55,80,8,75,65,6,3-4_per_week,8,0,62,75,7,70,7,TRUE,0.9
P004,Control,30,M,PhD,5_years,82,55,4,50,90,8,daily,30,2,78,60,5,52,-4,FALSE,0.87,75,
P005,EML,26,F,Bachelor,2_years,60,75,7,68,,,,,,,,,,,TRUE,week_5,□□□□□□
```

变量说明： - group: EML（实验组）或Control（对照组） - T0_difficulty / T3_difficulty: 1 - 10主观难度量表 - T0_cognitive_load: NASA-TLX综合分（0 - 100） - usage_frequency:

daily/3 - 4_per_week/1 - 2_per_week - withdrawal_violations: 拔线期违规次数 - delta: P_2 - B_0（能力增量） - avp_pass: TRUE/FALSE（是否通过AVP判据） - retention_rate: P_2 / P_2（保持率，如有随访） - dropout_reason: 自由文本（如有流失）

A.1.4 常见问题与解决方案

问题	解决方案	备注
参与者流失率高（>30%）	①缩短W窗口至4周②提供参与激励（证书/报告/小[illegible]	[illegible]
任务难度不等值	①使用IRT事后校准②报告置信区间③敏感性分析	[illegible]（N>200）小样本用专家判断
评分者信度低（ICC<0.7）	①重新培训评分者②细化评分标准（rubric）③增[illegible]	[illegible]信度较难保证
拔线期违规监测困难	①技术阻断（禁用API密钥）②每周自我报告（荣[illegible]	[illegible]时间戳）
天花板/地板效应	①调整任务难度（更难/更易版本）②使用分层任[illegible]	[illegible]需调整难度
对照组设计困难	①纯对照：完全不用AI（伦理问题）②积极对照：[illegible]	[illegible]延迟干预

A.5 常见问题与注意事项

Q1: 如何确定等值性（B_0测试和P 测试难度相同）？ A: 三种方法（优先级递减）：

1. IRT法（最严格）：
   - 适用条件：大样本（N≥200），多题目（≥20题）
   - 方法：使用项目反应理论（IRT）建立任务难度模型
   - 工具：R语言的mirt包，Python的pyirt
   - 优点：可以精确估计每个任务的难度参数θ
   - 局限：需要大量预测试数据
2. 预测试法（实用平衡）：
   - 适用条件：中样本（N=30 - 50），有pilot阶段
   - 方法：在独立样本上测试任务通过率，确保±5%以内
   - 示例：任务A通过率62%，任务B通过率58% → 可接受
   - 优点：直观易行
   - 局限：通过率相同不代表难度完全等值
3. 专家判断法（最低要求）：
   - 适用条件：小样本或无pilot条件
   - 方法：邀请3 - 5位领域专家独立评估任务难度（1 - 10量表）
   - 一致性检查：ICC > 0.7
   - 优点：成本低，快速
   - 局限：主观性强，可能有偏差

推荐组合策略：专家判断+预测试+事后IRT校准（如有足够数据）

Q2: 拔线窗口（W）应该多长？ A: 取决于任务的”能力巩固期”（工作假设，需跨领域/任务校准）：

任务类型	推荐W窗口	理由	示例
程序性任务	4 - 8周	需要足够时间让技能自动化	编程、数学解题
概念性任务	2 - 4周	理解为主，巩固较快	写作、设计思维
长期学习	12 - 24周	深度概念需要多次应用	第二语言、专业技能
快速技能	1 - 2周	简单技能，可快速评估	快捷键使用、工具操作

调整建议： - 先用4周试点，观察流失率和用户反馈 - 如流失率>30% → 缩短至2 - 3周 - 如用户反馈”还没适应就测试 → 延长至6 - 8周 - 权衡：W越长，能力巩固越充分，但流失率越高

神经科学参考： 技能巩固的时间尺度（来自运动学习研究）： - 短期记忆→长期记忆：数小时至数天 - 程序性记忆固化：1－4周 - 认知技能自动化：4－12周

Q3: 如果用户拒绝参加P 测试怎么办？ A: 两种应对策略：

策略1：激励机制设计 - 物质激励：小额奖励（如50－100元礼品卡），分阶段发放 - T_0完成：30% - T_3完成：70% - 随访完成：额外奖励 - 非物质激励： - 个人能力报告（展示C” (t)变化曲线） - 技能认证证书（如通过AVP-Retention）－学习建议（基于能力向量分析）－贡献感（你的数据帮助改进AI教育）

策略2：价值沟通与知情同意 - 透明解释测试目的： - “这不是考试，而是评估AI工具对你能力的影响 - ”数据匿名处理，不会影响你的成绩/绩效 - “结果用于改进AI系统，让未来用户受益 - 强调自主权： - 明确可以随时退出 - 退出不影响已获得的激励 - 提供退出反馈渠道

伦理注意事项： - 如流失率>30%，需分析是否存在选择性偏差：－能力弱者更容易放弃？“（导致高估效果） - 能力强者觉得浪费时间？（导致低估效果）－解决方法： - 意向性分析（ITT）：保留所有随机化参与者 - 多重插补（MI）：对缺失数据建模 - 敏感性分析：假设不同流失模式的影响

Q4: 本工具包通过了哪些验证？ A: 诚实回答：

本工具包尚未经过大规模、多领域的实证验证。它是基于： 1. 理论推导（认知科学+教育心理学文献） 2. 文献综述（认知卸载、脚手架理论等） 3. 小规模试点（N<100，单一领域）

构建的原型协议。

已有的支持证据： - AVP的核心逻辑（拔线+对比）借鉴了成熟的前测-后测设计 - 参数（如Cohen’s d ≥ 0.3 或 ≥10%（working assumption））基于心理测量学的常规效应量标准 - 50－70%成功率来自心流理论和最近发展 - 拔线窗口W=4－8周（默认6周）基于能力巩固的经验性时间估计

缺乏的证据： - 跨领域验证（仅有编程/写作初步试点） - 大样本RCT（N>500） - 长期追踪（>6个月） - 跨文化复现

我们的期待： 我们热切期待研究社区对本工具包进行：－测试：在你的领域应用并报告结果－改进：根据实践经验调整参数 - 证伪：如发现AVP判据不成立的场景，请告诉我们 - 替代：如有更好的测量方法，欢迎提出

联系方式（开放科学承诺）：[预留联系方式/GitHub仓库链接]

Q5: 哪些场景不适合使用本工具包？ A: 以下场景不建议或禁止使用本原型协议：

禁止场景（见使用说明-高风险豁免）： 1. 医疗、金融、交通等高风险决策场景－原因：本工具包未经专业认证，不能用于 - 替代：使用FDA批准、ISO认证等专业测评工具

2. 涉及法律责任或监管合规的评估
   - 原因：可能产生法律纠纷
   - 示例：驾照考试、职业资格认证
3. 用于人事考核、绩效排名等问责目的
   - 原因：违反Goodhart防护原则（见CET7 § 4.2）
   - 问题：用户会操纵基线、拔线期违规

不建议场景（效果可能不佳）：

1. 纯工具性任务：如使用计算器进行算术运算
   - 原因：这些任务不追求能力建构，外骨骼是合理选择
   - 判断标准：如果”独立完成无价值，不适用AVP
2. 补偿性外骨骼场景：如残障人士的辅助设备
   - 原因：AVP判据不适用（见3.1.6节边界条件）
   - 原则：辅助设备的目标是补偿缺陷，而非建构能力

3. 创造性/开放性任务：如艺术创作、科学发现
   - 原因：P_2难以量化，能力定义模糊
   - 替代：可能需要定性评估方法（如作品集评审）
4. 极短周期项目：协作期<2周
   - 原因：能力巩固需要时间，测量噪声会很大
   - 建议：至少4周协作期+2周拔线窗口

替代方案建议：

场景	不适用原因	推荐替代方案
高风险决策	未经认证	专业测评机构的认证工具（如FDA批准的诊断系统）
纯工具性任务	无能力建构价值	传统的任务完成度评估（如准确率、速度）
创造性任务	难以量化	作品集评估法、专家评审、多维评分rubric
短周期项目	时间不足	过程观察法（如编程过程录屏分析、think-aloud）
补偿性辅助	目标不同	功能性评估（如ADL量表、可用性测试）

误用案例警示：

反面案例：某公司的AVP晋升评估失败

某科技公司曾试图用AVP评估员工”是否值得晋升”，将P 分数作为KPI。这严重违反了Goodhart防护原则，导致：

操纵行为：- 员工故意在B 测试中表现差”（人为压低基线）- 拔线期间秘密使用AI（无法真实测量独立能力）- 互相分享P 测试题目（破坏等值性）

后果： - AVP完全失效（通过率虚高至95%） - 引发法律纠纷（员工质疑评估公平性） - 团队信任破裂（测试变成对抗性游戏）

正确做法： AVP仅用于能力诊断和系统改进，不得与个人利益挂钩。如确需评估员工能力，应：- 明确告知不影响晋升/薪酬 - 数据匿名化处理 - 用于团队整体能力规划，而非个人排名

A.6 简化版协议：24小时拔线轻量测试

适用场景： - 快速试点（验证AVP可行性） - 低风险任务（如非关键业务） - 资源有限（时间/人力/预算不足）

流程简化：

```
$T_0$□□□□→ 1□□□□ → 24h□□ → $T_1$□□□□
 [□□□] [□□□□]

□□□
- □□□□□□6□→24□□□
- □□□□
- □□□□□□□□□□□□□□□□□□□□
```

```
□□□
- □□+□□□□□□□
- AVP□□□$P_1$ ≥ $B_0$□
- □□□□□□
```

判据调整： 由于时间短，降低标准： - $P_1 \geqslant B_0$（维持能力即可，不要求+δ） - 成功率在拔线期不低于30%

局限性： - 无法测量长期能力巩固 - 容易受短期记忆影响 - 仅能识别严重的外骨骼模式

推荐用途： 作为预警工具而非最终判定。如24h测试失败（$P_1 < B_0$），强烈建议进行完整AVP验证。

附录B：补充案例分析

B.0 案例选择方法论

选择标准： 1. 真实性：真实项目或基于真实场景的合理虚构（标注） 2. 代表性：覆盖不同尺度（I/T/O）和结果（成功/失
3. 教学性：能清晰展示AVP/EML的应用 4. 可验证性：提供足够细节供复现

局限性声明： 这些案例主要来自WEIRD社会（Western, Educated, Industrialized, Rich, Democratic），跨文化适用性需进一步验证（见6.1.1节）。样本偏向科技行业，其他行业的案例有待补充。

案例完整性检查清单（6要素标准）：

每个案例必须包含以下要素：

☐ 1. 背景描述（Who/What/When/Where） ☐ 2. AVP测试结果（$B_0/P_1/P_2$的具体数据） ☐ 3. EML条件分析（有益摩擦/系统削减/验证） ☐ 4. 成功/失败原因分析 ☐ 5. 可迁移的洞察（Takeaway） ☐ 6. 红旗提示（≤2条，易被误用的做法）

B.1 外骨骼案例：写作助手的依赖陷阱

1. 背景描述

- Who: Alice，30岁内容创作者，自由撰稿人
- What: 使用AI写作助手（ChatGPT-4）创作博客文章
- When: 2023年6月-2024年6月，持续12个月
- Where: 远程工作，每周产出3－5篇800－1200字文章

使用模式演变： - 前3个月：用AI生成大纲和关键论点，自己填充内容和个人经验 - 3－6个月：开始让AI生成段落初稿，自己
- 6－12个月：完全依赖AI生成，只做微调（改几个词、调整语气）

2. AVP测试结果

基线测试（B_0，使用AI前，2023年5月）： - 任务：独立完成800字科技评论文章 - B_0 = 7.5/10（内容质量）+ 8/10（原创性）= 平均7.75 - 完成时间：120分钟 - 特征：论点清晰，有个人见解，表达风格独特

协作期表现（P_1，2024年5月）： - P_1 = 8.5/10（内容质量）+ 9/10（原创性自评）= 8.75 - 完成时间：30分钟（提效75%） - 但：编辑发现文章”太通用，缺少个人风格

拔线测试（P_2，2024年6月，6周无AI后）： - 任务：独立完成800字科技评论（平行测验） - P_2 = 6/10（内容质量）+ 4/10（原创性）= 平均5.0 - 完成时间：180分钟（比B 更慢） - 判定：$P_2 < B_0$（能力严重退化，外骨骼模式）

能力向量分析： | 维度 | B_0 | P_1 | P_2 | 变化 | | ——- | —- | —- | —- | —- | | 论点构建 | 8 | 9 | 5 | -3 | | 论据支撑 | 7 | 8 | 6 | -1 | | 逻辑展开 | 8 | 9 | 7 | -1 | | 语言表达 | 7 | 8 | 5 | -2 | | 个人风格 | 9 | 8 | 3 | -6 |

3. EML条件分析

有益摩擦：无 - Alice完全依赖AI生成，成功率接近100%（零摩擦） - 未设计任何留白”机制

系统性削减：无 - S”（t）= 1.0（持续100%支持，无削减）

AVP验证：事后测试，未预先设计

结论：完全不符合EML设计原则 → 预期导致外骨骼

4. 成功/失败原因分析

能力退化的具体表现：

1. 论点构建能力下降：

- 症状：拔线后不知道写什么，盯着空白屏幕1小时
- 原因：12个月未练习”从零生成论点

2. 表达风格丧失：

- 症状：写作风格趋同于AI（通用化、无个性）
- 原因：长期模仿AI的表达习惯

3. 创意枯竭：

- 症状：不再主动思考新角度，习惯性”等待AI提供
- 原因：创意生成的认知路径被旁路

心理机制： - 替代学习：大脑学会了”如何更好地提示AI”，而非如何更好地写作 - 能力幻觉：P_1高分让Alice误以为自己能

5. 可迁移的洞察”（Takeaway）

关键教训： 1. 外骨骼是渐进的：能力退化是逐步积累的过程，通常在使用者未察觉时悄然发生 2. 表面效率≠真实能力：P_1高不代表P 也高 3. 需要主动监测：Alice没意识到能力退化，因为输出质量始终高（有AI加持） 4. AVP测试的价值：揭示了隐藏的问题

预防策略： - 每月一次无AI日”（自我拔线测试） - 保持”AI生成大纲，人类填充内容的协作模式”（有益摩擦） - 定期对比B 基线，监测能力变化

6. 红旗提示（易被误用的地方）

红旗1：不要将本案例解读为完全不用AI” - 正确理解：问题不在于”用AI，而在于怎么用 - Alice如果采用EML设计”（如只让AI生成大纲，自己写内容），可能避免退化

红旗2：本案例的”6周拔线可能对专业作家过长 - Alice是自由撰稿人，有经济压力（6周无AI = 收入减少） - 如复制本实验，考虑： - 缩短拔线窗口（2 - 3周） - 提供经济补偿 - 或用”部分拔线”（每周1 - 2天无AI）

B.2 内共生案例：编程教学平台的成功实践

1. 背景描述

- Who: CodeMaster平台，在线编程教育机构
- What: 为初学者（Python入门）提供AI辅助学习
- When: 2024年1月-6月，12周课程
- Where: 线上学习，10,000名学员（实验组5,000，对照组5,000）

实验设计： - 实验组：EML设计的AI助教 - 对照组：标准AI助教（无摩擦、无削减）

2. AVP测试结果

基线测试（B_0，课程开始前）： - 任务：独立完成Python基础算法题（3题，难度递增） - B_0平均分：55/100（新手水平）

协作期表现（P_1，第8周）： - 实验组P ：78/100（使用EML-AI） - 对照组P ：82/100（使用标准AI） - 对照组略优（标准AI提供更多帮助）

拔线测试（P_2，第12周，4周拔线后）： - 实验组P ：72/100 - 对照组P ：58/100 - 实验组显著优于对照组

AVP判定： - 实验组：$\Delta = 72 - 55 = +17 > \delta$：≥Cohen’s d ≥ 0.3 或 ≥10%（working assumption）（工作假设，需跨领域/任务校准） → AVP通过 - 对照组：$\Delta = 58 - 55 = +3 < \delta$ → AVP失败

统计显著性： - 组间差异：t” (8200) = 12.5, $p < 0.001$, Cohen’s d = 0.45（中等效应量） - 实验组AVP通过率：73%（N=3,660/5,000，流失率18%） - 对照组AVP通过率：42%（N=1,722/4,100，流失率18%）

3. EML条件分析

有益摩擦：设计良好 - 完整性摩擦：AI给出算法思路，但代码框架留空 - 示例：学员问如何实现快速排序 - AI回答：“快速排序的核心是选择pivot并分区。请你先写出分区函数partition” ()，我会给你思路提示。 - 目标成功率50 - 70%（工作假设，需跨领域/任务校准）（工作假设，需跨领域/任务校准，实际监测：60%左右波动） - 动态调整**：如成功率<50%持续2周，降低摩擦

系统性削减：执行良好 - 削减曲线：S型削减（见图B.3） - 第1 - 4周：S(t) = 0.8（80%支持） - 第5 - 8周：S(t) = 0.5（50%支持） - 第9 - 12周：S(t) = 0.2（20%支持） - 实际执行：按计划执行，偏差<5%

AVP验证：预先设计，嵌入式测试 - 每2周一次微测试（10%任务无AI） - 最终拔线测试（第12周）

结论：完全符合EML设计原则 → 预期促进内共生

4. 成功/失败原因分析

实验组成功因素：

1. 透明沟通：
   - 学员知道”为什么AI不给完整答案
   - 明确告知：我们的目标是你能独立编程，而非依赖AI”
2. 渐进削减：
   - 不是突然拔线，而是平滑过渡
   - 学员有时间适应每个削减阶段
3. 成就系统：
   - 完成”独立项目有徽章奖励”（Transfer Badge）
   - 可见的进步反馈（能力雷达图）
4. 逃生阀：
   - 真正卡住时可请求完整帮助（但记录使用次数）
   - 缓解学员的焦虑和挫败感

对照组问题：

1. 零摩擦陷阱：
   - AI直接给出完整代码
   - 学员复制粘贴，未真正理解
2. 能力幻觉：
   - P_1高分让学员误以为”我已经学会了
   - 拔线后才发现无法独立完成
3. 无削减机制：
   - 持续依赖AI，未培养独立能力

5. 可迁移的洞察（Takeaway）

关键成功要素：

1. 给思路，不给代码的摩擦设计：
   - 适用范围：广泛（写作、数学、设计等）
   - 关键：保持50 - 70%成功率，避免过度挫败
2. 渐进削减比突然拔线更有效：
   - 学员需要时间适应每个削减阶段
   - S型曲线（前慢后快）效果好于线性削减
3. 透明沟通建立信任：
   - 学员理解”摩擦”的价值，而非感到AI不好用
   - 明确”长期能力建构”目标
4. 嵌入式测试提供持续反馈：

- 每2周微测试”（而非只在最后拔线）
- 早期发现问题，及时调整

跨领域应用： - 写作教育：AI生成大纲，学员填充内容 - 数学辅导：AI提示解题思路，学员完成步骤 - 设计工具：AI提供灵感，设计师执行

6. 红旗提示（易被误用的地方）

红旗1：本案例的成功率（60%）适合初学者，对专家可能过低 - 如果用户已有一定基础（如有2年编程经验），60%成功率 - 建议：专家用户提高目标成功率50 - 70%（工作假设，需跨领域/任务校准）上限（工作假设，需跨领域/任务校准），或增

红旗2：本案例的18%流失率是可接受的，但不同场景容忍度不同 - 学习场景：18%流失率尚可接受 - 企业培训：可能需要<10%流失率（涉及成本） - 关键技能：可能需要<5%流失率（如医疗培训） - 建议：先试点评估流失率，再决定是否全面推广

B.3 团队层案例：软件公司的T-AVP实验

1. 背景描述

- Who: 某创业公司，两个8人工程团队（A组vs B组）
- What: 评估团队对AI代码助手（GitHub Copilot）的依赖
- When: 2024年1月-6月，6个月观察期
- Where: 远程+混合办公

实验设计： - A组（实验组）：实施周五无AI日”+强制人际代码审查 - B组（对照组）：自由使用AI，无限制

2. AVP测试结果

团队基线（B_0，2023年12月）： - 任务：团队协作完成中型功能模块（无AI环境，3天时限） - A组B ：功能完成度85%，代 - B组B ：功能完成度82%，代码质量7.3/10 - 基线相当

6个月后T-AVP测试（2024年6月）： - 任务：在无AI环境下完成类似功能模块（3天） - A组P ：功能完成度88%，代码质量8 → T-AVP = 0.88 - B组P ：功能完成度62%，代码质量5.5/10 → T-AVP = 0.62

判定： - A组：$P_2 \approx B_0$（维持能力，轻微提升）→ T-AVP通过 - B组：$P_2 < B_0$（显著退化）→ T-AVP失败

团队能力分布分析： | 团队成员 | A组（有”周五无AI日） | B组”（无限制使用AI） | | ——— | ——— ——— | ——————— - | | Senior | 独立能力保持良好 | 独立能力保持良好（较少依赖AI） | | Mid-level | 略有提升（人际学习） | 轻度依赖（独立能力下降10 - 20%） | | Junior | 显著提升（被迫学习） | 严重依赖（独立能力下降50%+） |

关键发现：B组出现能力极化现象 - 2名Junior工程师完全依赖AI，拔线后几乎无法贡献 - 团队整体T-AVP被Junior成员拖累

3. EML条件分析

A组（实验组）： 团队摩擦：周五无AI日强制人际交流 - 成员间互相请教、代码审查增加 - 知识在团队内流动

渐进式削减：每周1天（14%时间）无AI - 虽非系统削减S” (t)曲线，但提供了定期的”能力锻炼

T-AVP验证：6个月后集体拔线测试

B组（对照组）： 零摩擦：完全依赖AI，成员间交流减少 无削减：S” (t) = 1.0（持续100%可用） 未验证：事后才发现问题

4. 成功/失败原因分析

A组的优势：

1. 知识流动增加：

- 周五无AI日促进人际交流
- Senior成员分享经验，Junior成员快速成长
- 团队形成”知识网络，而非AI星型依赖”

2. 角色冗余：

- 成员间可以互相补位
- 某人休假/离职，团队仍能运作

3. 架构理解提升：

- 被迫理解系统全局，而非只关注局部功能
- Code review迫使成员解释设计思路

B组的问题：

1. 能力极化：

- Senior成员依赖少，能力保持
- Junior成员严重依赖，能力退化严重
- 团队整体能力分布不均衡

2. 知识流失：

- 团队内部不再分享经验”（都问AI）
- 隐性知识（tacit knowledge）未传承

3. 架构理解差：

- 过度依赖AI生成代码，对系统整体理解不足
- 出现bug时，难以快速定位和修复

定量证据： - A组的”人际代码审查”次数：6个月内平均48次/人 - B组的”人际代码审查次数：6个月内平均12次/人 - A组的Slack技术讨论消息：日均15条 - B组的Slack技术讨论消息：日均5条

5. 可迁移的洞察”（Takeaway）

团队层的关键洞察：

1. 无AI日是简单有效的T-AVP保障机制：

- 成本低（只需政策，无需技术）
- 可操作性强（每周固定1天）
- 副作用小（不影响整体效率，周五选择较好）

2. 团队能力≠个体能力之和：

- I-AVP通过 ≠ T-AVP必然通过（涌现性）
- 需要监测”知识流动和”角色冗余

3. Junior成员是T-AVP的脆弱点：

- 他们最容易形成依赖”（缺乏经验对抗）
- 需要特别保护（如前3个月禁用AI）

4. 人际交流是团队韧性的基础：

- AI不能替代”隐性知识传承
- Code review、技术分享会、结对编程的价值在AI时代更加重要

管理层决策： 基于本实验结果，公司决定： - 全公司推广”周五无AI日 - 新人前3个月禁用AI”（建立基础能力） - 每季度T-AVP演练（模拟AI宕机场景） - 绩效考核中增加人际协作维度（而非只看个人产出）

6. 红旗提示（易被误用的地方）

红旗1：不要将”周五无AI日”理解为”惩罚或开倒车” - 正确定位：这是能力锻炼日，类似健身房的负重训练 - 沟通技巧：强调”保持团队韧性而非”限制工具使用 - 如果团队抵触，可以： - 从”每月一次”开始”（降低频率） - 选择非关键任务进行拔线测试 - 展示A组vs B组的数据对比

红旗2：本案例的3天拔线测试对某些行业可能过长 - 软件开发：3天中型功能是合理的 - 其他行业可能需要调整：
- 咨询业：1天案例分析 - 设计行业：2天项目设计 - 数据分析：半天报告产出 - 关键：选择团队常规任务”作为测试场景，

附录C：理论定位与学术对话

C.1 CET与认知心理学的对话

与延展心智理论（Extended Mind Theory, Clark & Chalmers 1998）的关系 核心主张： - Clark & Chalmers：认知过程可以延展到外部工具（如笔记本、计算器） - “耦合-构成原则：工具可以成为认知系统的一部分，而非

CET的立场：

相似点”（我们如何借鉴）： - 认同认知可以延展到工具（AI作为认知伙伴，见3.0.4节） - 认同人-工具的耦合”可以产生增强效果（P_1提升）

△ 差异点（我们如何超越/补充）： - 关键分歧：不是所有”延展都是健康的 - 健康延展：认知内共生”（通过AVP验证，P
- 不健康延展：认知外骨骼（AVP失败，$P_2<B_0$） - CET的贡献：提供了判断延展质量的可证伪标准（AVP）

互补点（如何协同）： - 延展心智理论提供哲学基础（“工具可以是心智的一部分） - CET提供操作标准（”如何判断延
- 未来合作方向： - 神经科学验证：哪种延展模式激活大脑的不同区域？ - 长期影响：延展10年后，大脑结构是否改变？

对话焦点：

> 延展心智理论：工具就是心智的一部分，不存在好坏之分。 CET回应：同意工具可以延展心智，但需要区分”增强型延
>
> 延展心智理论：使用计算器不会让算术能力退化。 CET回应：取决于使用方式。如果完全依赖（从不心算），长期可能—这正是AVP的作用。

C.2 CET与教育技术的对话

与脚手架理论（Scaffolding Theory, Wood, Bruner & Ross 1976）的关系 核心主张： - 教学应提供临时性支持”（脚手架
- 随着学习者能力提升，逐步撤出支持 - 最终学习者完全独立

CET的立场：

相似点（我们如何借鉴）： - 完全一致：EML的”系统性支持削减正是脚手架理论的形式化 - 借鉴了Vygotsky的最近发展区（ZPD）概念（50 - 70%挑战区）

△ 差异点（我们如何超越/补充）： - CET的贡献： 1. 量化削减过程：将”渐进撤出形式化为削减曲线S”(t) 2. 提供验证标准：AVP判据确保撤出成功（不是撤了就行，要看P ） 3. 扩展到AI时代：自动化脚手架（LSA架构，见第
4. 跨尺度扩展：从个体学习扩展到团队/组织能力建构

互补点（如何协同）： - 脚手架理论提供教育学基础 - CET提供AI时代的技术实现路径 - 未来合作方向：
- AI如何自动判断学习者准备好了？（能力监测C(t)） - 不同学科的最优脚手架设计？（跨领域校准）

对话焦点：

> 脚手架理论：何时撤出支持？ CET回应：基于能力监测（C(t)）动态决定，并通过AVP验证撤出是否成功。
>
> 脚手架理论：如何知道学习者准备好了？ CET回应：监测成功率（目标50 - 70%区间，工作假设，需跨领域/任务校准）
>
> 脚手架理论：如果撤出失败怎么办？ CET回应：自动回退机制（见5.3.3节，SGS的安全约束）。如连续失败触发红色预

C.3 CET与AI伦理的对话

与自主性（Autonomy）讨论的关系 AI伦理的核心关切： - AI是否威胁人类自主性？ - 如何保护用户的选择自由？
- 算法决策的透明性和可解释性？

CET的立场：

相似点（我们如何借鉴）： - 共同关注”人类主体性”（human agency） - 认同用户应保持对AI的控制

△ 差异点（我们如何超越/补充）： - CET超越传统自主性讨论：不仅关注”能否选择”，更关注能力是否保持 - 认知自主性”（Cognitive Autonomy）： - 传统自主性：我可以选择用或不用AI（意志自由） - 认知自主性：我用了AI后，仍然有能力独立完成任务（能力自由）

CET的独特贡献： 1. 将抽象的自主性”转化为可测量的”独立能力”（AVP） 2. 提出”伙伴式主体性作为AI角色的伦理基准 3. 警示”外骨骼模式对认知自主性的长期威胁

互补点”（如何协同）： - AI伦理提供规范性框架（应该如何”） - CET提供可操作的评估工具（如何验证） - 未来合作方向： - 将AVP纳入AI系统的伦理审查流程 - 开发”认知自主性指数”作为AI产品的标签”（类似能效标签）

对话焦点：

AI伦理：AI威胁人类自主性吗？ CET回应：取决于AI的设计。外骨骼模式威胁认知自主性（$P_2<B_0$），内共生模式增强

AI伦理：如何量化”自主性？ CET回应：通过AVP判据。如果P $\geqslant B_0+\delta$，说明用户的认知自主性得到维护甚至增强。

AI伦理：用户知情同意就够了吗？ CET回应：不够。即使用户同意使用AI，如果导致能力退化（外骨骼），从长期伦理

C.4 CET与组织行为学的对话

与组织韧性（Organizational Resilience, Hollnagel 2011）的关系 核心主张： - 韧性 = 在扰动下维持功能的能力 - 需要多样性、冗余、适应性 - 复杂系统的”韧性工程”（Resilience Engineering）

CET的立场：

相似点（我们如何借鉴）： - 完全一致：O-AVP正是组织认知韧性的测量 - 借鉴了”多样性（角色冗余）和”冗余”（知

△ 差异点（我们如何超越/补充）： - CET识别了AI依赖作为韧性的新威胁： - 传统威胁：单点故障、人员流失、供应链中断 - 新威胁：AI依赖导致的”集体能力退化 - CET提供了具体测量方法： - 48h演练（见4.2.2节） - BCI/ICR双指标（见4.2.3节）

互补点（如何协同）： - 组织韧性理论提供宏观框架 - CET提供AI时代的具体威胁识别和测量工具 - 未来合作方向： - 将O-AVP纳入组织风险评估体系 - 开发认知韧性仪表盘（real-time monitoring）

对话焦点：

组织韧性理论：如何增强组织韧性？ CET回应：除了传统措施（多样性、冗余），还需关注”认知韧性”——确保组织在AI不可用时仍能运作。

组织韧性理论：如何测量韧性？ CET回应：O-AVP提供了可操作的测量方法：48h中断演练+恢复时间/独立完成率。

组织韧性理论：韧性是否需要持续投入？ CET回应：是的。建议每季度一次O-AVP演练”（类似消防演习），保持组织的

C.5 CET与系统科学的对话

与复杂系统理论（Complex Systems Theory）的关系 核心主张： - 复杂系统表现出涌现性（emergence）、非线性、自组织 - 局部优化≠全局最优 - 反馈环路主导系统行为

CET的立场：

相似点（我们如何借鉴）： - 认同跨尺度涌现性：I-AVP通过≠T-AVP必然通过（见4.1节） - 借鉴反馈环路思想：AVP闭环（见5.4节）

△ 差异点（我们如何超越/补充）： - CET关注认知系统的特殊性： - 能力的不可逆性（退化容易恢复难） - 代际传承的路径依赖（$T_0\rightarrow T_1\rightarrow T_2$） - 认知公地悲剧”（个体理性→集体退化，见4.3节）

互补点（如何协同）： - 复杂系统理论提供分析框架 - CET提供AI-人类系统的具体案例 - 未来合作方向： - 建立CET的系统动力学模型（SD modeling） - 模拟长期演化路径（agent-based modeling）

对话焦点：

复杂系统理论：为何局部优化（个体用AI效率高）导致全局问题（组织能力退化）？ CET回应：因为存在认知公地悲剧

复杂系统理论：如何避免系统崩溃？ CET回应：通过多尺度监测（I/T/O/S-AVP）和早期预警（黄色/红色预警，见5.4.

复杂系统理论：系统是否有”吸引子”（attractor）？ CET回应：可能存在两个吸引子： - 良性吸引子：内共生平衡态（$P_2 \geq B_0 + \delta$ 持续稳定） - 病理吸引子：外骨骼锁定态（$P_2 < B_0$ 且难以恢复） 需要实证研究验证这一假设。

附录D：术语与符号索引

D.1 核心概念术语（按字母序）

认知内共生（Cognitive Endosymbiosis） - 首次出现：1.3.1节，3.0.4节（完整定义） - 核心定义：AI作为认知伙伴，通过 - 关键特征：①有益认知摩擦 ②系统性支持削减 ③AVP验证 - 对照术语：认知外骨骼（病理模式） - 参见：3.3节（EML）、4.1节（跨尺度扩展）

认知外骨骼（Cognitive Exoskeleton） - 首次出现：1.2.2节（问题提出），3.0.6节（详细定义） - 核心定义：过度依赖AI导致独立能力退化的病理模式。特征是高P （有AI表现好）但低P （无AI表现差） - 关键特征：①零摩擦设计 ②无支持削减 ③AVP失败（$P_2 < B_0$） - 对照术语：认知内共生（目标模式） - 参见：表3.5（失败模式分类）、附录B.1（案例）

反脆弱性验证原则（Antifragility Validation Principle, AVP） - 首次出现：1.3.1节（概念引入），3.0.2节（完整定义 - 核心定义：以拔线测试检验协作是否促进独立能力。判据为P $\geq B_0 + \delta$（见3.0.2节AVP定义锚点B1） - 关键参数：B_0（基线）、P_2（拔线后能力）、δ（最小提升阈值）、W（拔线窗口） - 分级体系：AVP-Basic（$P_2 \geq B_0$）、AVP-Retention（$P_2 \geq B_0 + \delta$）、AVP-Transfer（+迁移） - 参见：3.1节（详细机制）、附录A（测量工

内共生最小法则（Endosymbiotic Minimal Law, EML） - 首次出现：1.3.2节，3.0.3节（完整定义） - 核心定义：构成认知内共生的设计必要条件：①有益认知摩擦（50 - 70%成功率） ②系统性支持削减（S4→S1→S0）（见3.0 - 与AVP关系：EML是设计条件，AVP是验收条件；二者联合构成充分必要条件 - 参见：3.2节（有益摩擦）、3.3节（支持削减

伙伴式主体性（Partner-like Agency） - 首次出现：1.3.3节，3.4节（详细阐述） - 核心定义：AI作为认知伙伴的理想角色 ②脚手架消退 ③AVP闭环 - 关键特征：功能性非拟人化；教练模式而非仆人模式 - 参见：3.4节（理论阐述）、5.5节（伦理

有益认知摩擦（Beneficial Cognitive Friction） - 首次出现：3.2节 - 核心定义：适度挑战（目标成功率50 - 70%，工作假设，需跨领域/任务校准）促进能力增长的设计策略 - 类型：完整性摩擦、抽象性摩擦、延迟性摩擦 - 理论基础：心流理论（Csikszentmihalyi）+ 最近发展区（Vygotsky） - 参见：表3.4（摩擦类型对照）、5.2节（CFE引擎

系统性支持削减（Systematic Support Reduction） - 首次出现：3.3节 - 核心定义：AI支持强度按既定削减曲线从S4→S1→ - 削减曲线类型：线性、指数、S型 - 安全机制：S_min下限、回退触发、预警系统 - 参见：图B.3（削减曲线对比）、5.3节

拔线测试（Unplugged Test） - 首次出现：3.1节 - 核心定义：在无AI环境下测量独立能力的标准化测试方法 - 关键要素：等值性（任务难度相同）、拔线窗口W、违规监测 - 参见：附录A.1.2（测量执行协议）

D.2 符号与参数索引

能力与表现：

符号	含义	单位/范围	首次定义
B_0	基线能力（使用AI前）	任务特定评分	3.0.2节
P_1	协作期表现（使用AI时）	任务特定评分	3.0.2节
P_2	拔线后独立能力	任务特定评分	3.0.2节
C” (t)	能力向量（多维）	n维向量	5.4.3节
ΔC	能力增长	$P_2 - B_0$	3.0.2节

阈值与参数：

符号	含义	默认值	首次定义
δ	最小有意义提升阈值	Cohen’s d ≥ 0.3 或 ≥10%（working assumption）*	3.0.2节
W	拔线窗口	4 - 8周（默认6周）*	3.0.2节

符号	含义	默认值	首次定义
F	摩擦参数	0－1（目标0.5－0.7）*	3.2节
S(t)	支持强度函数	0－1，从S0→0*	3.3节
S0	初始支持强度	0.8（80%）*	3.3.1节
S_min	安全支持下限	≥0.2（20%）*	3.3.2节
λ	削减速率	任务特定*	3.3.1节

*（工作假设，需跨领域/任务校准）

尺度特定指标：

符号	含义	阈值	首次定义
I-AVP	个体反脆弱性验证	$P_2 \geqslant B_0 + \delta$	4.1节
T-AVP	团队反脆弱性验证	≥0.7（群体级）*	4.1.3节
O-AVP	组织反脆弱性验证	告警≥0.70，目标≥0.85*	4.2.3节
BCI	业务连续性指数	0－1	4.2.3节
ICR	独立完成率	0－1	4.2.3节
S-AVP	社会层认知资本指标	代际对比	4.3节

LSA架构：

符号	含义	功能	首次定义
L1	基础AI能力层	AI模型接入、知识库	5.1节
L2	摩擦与削减层	CFE+SGS	5.2－5.3节
L3	监测与反馈层	AVP-TM+预警系统	5.4节
L4	编排与治理层	多尺度协调+伦理治理	5.5节
S4→S1→S0	支持档位栈	强→弱的4级支持强度	3.3节
CFE	认知摩擦引擎	动态摩擦调节	5.2节
SGS	支持削减调度器	削减曲线管理+回退	5.3节
AVP-TM	AVP遥测模块	能力监测+数据采集	5.4节
MSO	多尺度编排器	I/T/O跨尺度协调	5.5节

D.3 缩写速查表（完整版）

缩写	全称	中文	首次出现
AVP	Antifragility Validation Principle	反脆弱性验证原则	1.3节
EML	Endosymbiotic Minimal Law	内共生最小法则	1.3节
LSA	Layered Symbiosis Architecture	分层共生架构	1.3节
LSA-F	LSA Functional Hierarchy	LSA功能分层	3.0.5节
CFE	Cognitive Friction Engine	认知摩擦引擎	5.2节
SGS	Support Graduation Scheduler	支持削减调度器	5.3节
AVP-TM	AVP Telemetry Module	AVP遥测模块	5.4节
MSO	Multi-Scale Orchestrator	多尺度编排器	5.5节
I-AVP	Individual AVP	个体反脆弱性验证	4.1节
T-AVP	Team AVP	团队反脆弱性验证	4.1节

缩写	全称	中文	首次出现
O-AVP	Organizational AVP	组织反脆弱性验证	4.2节
S-AVP	Societal-level indicators	社会层认知资本指标	4.3节
BCI	Business Continuity Index	业务连续性指数	4.2节
ICR	Independent Completion Rate	独立完成率	4.2节
IRT	Item Response Theory	项目反应理论	附录A
RCT	Randomized Controlled Trial	随机对照试验	6.2节
DID	Difference-in-Differences	差分中的差分	6.2节
ZPD	Zone of Proximal Development	最近发展区	2.3节
WEIRD	Western, Educated, Industrialized, Rich, Democratic	西方、受教育的、工业化的、富裕的、民主的	6.1节

D.4 跨章节概念地图

核心三角关系：

```
 AVP□□□□□□
 / \
 / \
 EML □□□□□□
 □□□□□□ □□□□□□
 \ /
 \ /
 □□□□□
 □□□□□□
```

跨尺度扩展链：

```
□□□I□ → □□□T□ → □□□O□ → □□□S□
 ↓ ↓ ↓ ↓
I-AVP T-AVP O-AVP S-AVP
 ↓ ↓ ↓ ↓
□□□□ □□□□ 48h□□ □□□□
```

LSA技术实现栈：

```
L4□□□□□□ ← □□□□□ + □□□□
 ↑
L3□□□□□□ ← AVP-TM + □□□□
 ↑
L2□□□□□□ ← CFE + SGS
 ↑
L1□AI□□□ ← □□□□ + □□□
```

参考文献