

协调悖论:人工智能“智能-安全”的战略选择

李垣¹,李平^{2,3},许雅楠¹

(1. 上海交通大学安泰经济与管理学院,上海 200030; 2. 东北财经大学,辽宁大连 116025;
3. 厦门大学,福建厦门 361005)

摘要:人工智能(AI)技术的快速发展在创造机遇的同时,也加剧了对其安全风险与对齐问题的担忧。现有研究常将AI智能性(执行自动化决策的技术能力)与AI安全性(防范风险的责任属性)割裂讨论,忽视其内在悖论关系。本研究旨在探究如何通过战略设计协调智能性与安全性的张力并获取可持续竞争优势。本研究提出AI“智能-安全”悖论框架,识别出智能与安全交互的四类核心情境;同时,整合悖论理论与动态能力理论,形成AI“智能-安全”悖论核心观点;在此基础上,针对不同情境提出动态匹配的战略路径,阐释了不同战略下AI应用时间、内外能力与企业绩效的非线性关系及平衡机制。本研究主要贡献在于构建了整合智能与安全维度的AI技术分析框架,形成了AI“智能-安全”悖论的动态协同观点,建立了AI情境与协同战略的动态匹配。本研究为企业悖论协同管理和AI情境适配战略提供了理论指导和实践启示。

关键词:人工智能;AI智能性;AI安全性;动态能力;悖论理论

中图分类号:C93

文献标识码:A

人工智能(Artificial Intelligence, AI)技术的迅猛发展正在深刻重塑组织运行和企业发展的范式。AI技术的快速渗透与广泛应用释放出巨大潜力,但同时也伴随着日益凸显且复杂的安全性问题^[1,2]。在此情境下,受到有限资源约束的企业在开发和采用AI技术时面临“智能性-安全性”的悖论困境。如何在促进智能发展的同时确保系统安全,已成为企业AI战略管理中亟需破解的核心问题。

本研究认为,智能性与安全性是AI技术价值实现过程中不可分割的一体两面,企业在拥抱先进AI、追求技术红利的同时,必须将安全性置于同等战略高度予以系统性考量。然而,现有研究普遍将智能与安全单独考虑:一些文献关注AI赋能潜力与价值创造^[3-5];另一些研究则聚焦AI伴生的风险隐患^[6-8]。这种“智能-安全”二元对立构成了AI战略管理中难以回避的悖论。虽然一些研究触及AI技术带来的复杂悖论^[9-11],但均未能从系统整合的视角审视智能性与安全性的悖论关系,因而缺乏在

战略层面动态平衡二者的统一分析框架。

为弥补这一理论缺口,本研究整合动态能力理论与悖论理论,构建AI“智能-安全”动态协同模型。该模型视智能性与安全性为一对既冲突又可协调的因素,主张企业通过培育特定的动态能力,在资源约束和外部环境变迁的双重压力下,实现二者间的动态平衡与协同演进。同时,本文提出了一个基于智能性与安全性双维度的情境分类框架,识别出四种典型的技术状态组合,并针对每种情境设计了与之匹配的协同战略及能力建设路径。

本研究借助悖论理论与动态能力理论观点,系统性地整合了AI智能性与安全性维度,提出的战略框架拓展了AI技术催生的悖论问题的研究。同时构建了一个情境化协同战略框架,一方面拓展了悖论理论和动态能力理论在AI语境下的内涵与应用,另一方面也为AI战略研究提供了一个情境化的整合理论视角。同时,本研究为企业管理AI技术的复杂张力提供了新的理论视角和实践指引。

收稿日期:2025-08-05;修回日期:2025-10-25

基金项目:国家自然科学基金面上项目(72372102)

作者简介:李垣(1961-),男,教授、博士生导师。李平(1957-),男,教授、博士。许雅楠(2000-),女,博士研究生,通讯作者, E-mail: annasjtu@sjtu.edu.cn.

1 AI 战略研究现状

当前战略管理领域的学术研究聚焦于 AI 智能性带来的先进生产力如何促进企业成长与发展,着重围绕技术与产品创新^[12,13]、生产运营效率^[14,15]和绩效增长^[16]等方面。与此同时,也有一些学者注意到了 AI 技术固有的风险与隐患^[1,6,17,18],呼吁对 AI 进行合理监管^[8]。这种对立的现象揭示了围绕 AI 战略的深层次的悖论:对更高智能性的追求常伴随安全风险的放大,而严格的安全约束又可能限制智能快速进化,即“智能-安全”悖论。

围绕 AI 技术带来的复杂组织悖论,部分学者尝试进行探讨,但在阐释“智能-安全”悖论方面仍存局限。例如,Raisch 和 Krakowski 提出的“自动化-增强”悖论主要回应“技术替代人类还是辅助人类”的问题^[9],其焦点在人机关系,未能深入 AI 技术内核属性所带来的安全风险。Vanneste 与 Puranam 发掘的“代理-信任”关系锚定在个体或团队层面的微观心理感知^[10],未能提升至企业战略的宏观架构,且未将技术安全性作为关键变量。同时,鉴于 AI 战略的复杂性质,最新的理论研究提出情境化 AI 的理论视角^[19],呼吁后续战略研究开展对 AI 情境和对应战略的深入分析。

综上,现有成果尚不能全面把握智能性与安全性这一对核心战略悖论。这些理论空白表明 AI 战略研究缺乏一个能够整合技术属性、组织战略与治理框架的协同视角,以系统性地分析和应对 AI 的“智能-安全”悖论。

2 AI“智能-安全”动态协同模型

2.1 基于“智能性-安全性”框架的 AI 情境

本研究基于智能性与安全性这两个核心维度,构建了一个情境分类框架。其中,AI 智能性表征系统在特定任务场景中实现目标的有效性、适应性与创新性^[14];AI 安全性表征系统在数据保护、行为可控、伦理合规及风险抵御等方面的综合保障强度^[7,17]。根据二者的交互组合,可将 AI 技术与产品的状态系统地划分为四类情境:

(1) 低智能-低安全

该状态下,企业在 AI 技术基础与安全机制构建上均处于初级阶段,其核心特征为双重薄弱的“能



图 1 “智能-安全”协同框架战略设计

Figure 1 “Functionality - safety” framework strategy design

力真空”。在功能层面,系统仅能执行规则明确的简单任务,尚未充分利用自主决策与学习能力;在安全层面,缺少系统化的防护机制,风险监控体系缺位。该状态常见于转型初期,通常源于技术不成熟或资源投入不足,使企业面临“低效-高危”的双重困境。

(2) 低智能-高安全

此状态以风险最小化为优先目标,表现为企业虽具备相对完善的风险管控体系,但 AI 技术创新与应用能力明显不足。该状态通常出现在银行、医疗等对风险高度敏感的行业。其核心特征在于,过度的安全导向导致了“资源的配置僵化”,虽然保障了安全,但也抑制了智能性的发挥与进化,使企业陷入安全冗余但创新不足的效能瓶颈,面临因技术滞后而丧失市场竞争力的风险。

(3) 高智能-低安全

此状态体现了智能性与安全性的背离形式。该状态下的组织追求对 AI 技术智能性的最大化利用,从而帮助其快速抢占市场,却忽视了潜在的技术失控与数据安全风险。企业虽能推出高智能产品,但因安全管理机制的缺失而暴露于显著风险之下。这种状态的核心特征是“技术短视”与“治理缺位”,常见于部分急于抢占市场的新兴企业。

(4) 高智能-高安全

此状态代表了企业 AI 应用的理想协同境界,即同时实现高水平的智能性与高强度的安全性。在此状态下,技术先进性与风险可控性得以兼顾,AI 技术通过安全嵌入的架构设计实现了二元统一。企业既规避了“失控型”风险,也突破了“保守型”的创新天花板,从而获得可持续的 AI 赋能优势。能达到此状态的企业通常是资源雄厚且具备强大创新能力的行业领军者。

2.2 协同战略框架的理论基础

(1) 悖论理论

悖论理论关注组织中同时存在且持续存在的矛盾性与相互依存性关系。悖论被定义为在组织系统中并存的对立却相互关联的元素,这些元素在单独考察时各自合理,而当它们并置时则显得非理性或不一致^[20]。

悖论的核心性质包括矛盾性、相互依存性与持久性^[21]。矛盾性表现为伴随组织形成而产生的逻辑对立要素,普遍存在而无法消除^[20]。其次,相互依存性意味着这些对立要素在本质上互为条件、彼此构成,在更大系统中实现协同^[21]。最后,持久性表明悖论并非暂时性冲突,而是无法被彻底消除的持续性动态关系,是组织发展的恒常特征^[20]。

悖论关系还具有动态性,其显性化与外部环境密切相关。首先,悖论中相互依存的对立力量会不断相互影响,形成周期性的“此消彼长”关系。同时,环境动态进一步强化了二者之间的变化^[20]。在多元性、变化性与稀缺性的环境条件下,潜在的矛盾张力更容易被激发并显性化。因此,悖论不仅是内生的组织特征,也受到外部环境动态的影响。

在应对悖论张力时,组织可能采取“非此即彼”或“两者兼顾”的战略导向^[20,21]。前者体现对悖论的防御态度,表现为盲目坚持悖论中的一侧,容易产生路径依赖与组织惯性^[20]。相反,“两者兼顾”战略承认并接受悖论的持续存在,并主动利用这种张力^[22]。Smith 和 Lewis 将该战略描述为动态均衡模型,指出有效的悖论管理是有意识地迭代和平衡:行为者在短期内做出选择,同时敏锐地意识到在长期内接受矛盾^[20]。这种动态管理强调迭代式的平衡而非一次性的解决,使组织能够在张力中持续适应外部环境变化^[23]。

从结果来看,单边策略下,防御性反应和路径依赖将强化单一取向,导致惯性、结构僵化与学习停滞^[24,25]。尽管这种聚焦可能带来短期效率,但长期将削弱组织的灵活性与创新力^[20]。相反,当组织通过动态过程有效管理悖论时,将形成良性循环。对矛盾的包容认识激发了组织学习、创造力与组织弹性,促进资源重组、结构调整与目标再思考^[23,26],从而推动组织实现可持续性 & 超越性发展^[20,27,28]。

(2) 动态能力理论

动态能力理论强调企业仅拥有资源并不足以形成持续竞争优势,还必须具备动态调整资源配置的

能力^[29,30]。动态能力被定义为公司为了匹配或创造市场变化而对使用资源的流程。与普通能力不同,动态能力是“有目的地创建、扩展或修改其资源基础”的高阶能力^[31]。Teece 将动态能力概念化为“感知-捕获-重构”的核心机制框架^[32]。具体而言,企业需要首先识别并评估环境中的新机会和潜在风险,接着动员并配置资源抓住机会、防范风险,最后持续改造组织资源和业务结构,以实现战略性变革^[30,32]。

动态能力的性质主要反映在动态变化、环境依赖以及外部塑造方面。首先,动态能力具有生命周期与演化性,能够随环境变化和组织学习不断重构。Helfat 和 Peteraf 提出能力生命周期的概念^[33],表明动态能力并非固定不变,而是在企业发展中不断更新和演化的。其次,动态能力并非放之四海而皆准的万能能力,其形式和表现取决于行业和市场的动态程度^[25,34],不同行业和环境其具体实现方式会有所差异。最后,动态能力具有外部塑造性,即企业不仅能被动应对环境变化,更能通过创造性地使用能力来主动影响和塑造外部环境,即进化性适应^[35],这体现出动态能力在内外互动中的双向动态作用。

动态能力被认为是企业获取持续竞争优势的关键机制。具有强大动态能力的企业不仅可以响应变化,还能塑造环境,从而建立并周期性更新竞争优势^[36,37]。通过敏捷地重构资源和流程,企业能够在不确定环境中快速响应并推动创新,在快速变化中保持灵活性和创新能力^[22,38]。这种快速感知市场机会、把握机会并重组资源的能力,进一步帮助企业在不确定市场中占据先机。

(3) 理论整合

动态能力理论与悖论理论具有深刻的内在关联。悖论理论认为悖论关系是不断变化的,明确指出有效的悖论管理需要持续动态地平衡悖论关系、适应环境变化^[20,23,27]。而动态能力理论所包含的感知、捕获与重构机制,恰恰为管理此类悖论提供了操作性路径^[27,32,39]。通过动态能力的迭代运用,组织能够在探索与利用、稳定与变革等对立要素间实现有意识的循环调整^[32],从而避免陷入“非此即彼”的刚性陷阱,推动悖论张力向良性循环演化^[20]。

2.3 AI“智能-安全”动态协同模型的核心观点

整合动态能力理论与悖论理论的精髓,本文构建的 AI“智能-安全”动态协同模型提出以下核心

观点,旨在为理解和管理 AI 智能性与安全性之间的复杂关系提供一个系统性的理论框架。

首先,依据悖论理论,资源稀缺性是使潜在矛盾显化的关键条件^[20]。在 AI 战略中,智能性追求开放与探索,而安全性要求保守与可控,技术的先进性会天然地引发对技术安全和技术背叛的担忧,这是 AI 技术固有属性带来的矛盾。同时,智能性与安全性的建设都需要大量的资源投入,智能性的提升需要持续的数据投入、算法优化与算力支撑,而安全性的保障也需要在风险控制、漏洞监测、合规审查等方面大量部署。因此,在资源总量与时间窗口受限的约束下,组织难以对智能与安全进行同步充分的投入,导致其潜在对立性显性化。综上所述,在资源和时间有限的条件下,AI 的智能性和安全性客观上存在一定程度的悖论。

此外,正如悖论理论所指出,组织中的悖论元素间具有深刻而持续的相互依存性,并在不同时间与环境下呈现不同的组合状态^[21]。智能性与安全性之间既相互对立矛盾,又构成一种共生关系。其中,智能性的发展为安全性提出了新的挑战与需求,而安全性的完善又为智能性的进一步突破提供了可信赖的基石与边界。这种相互塑造、共同演进的特性,使得二者在组织中以各种不同的组合状态共存。综上所述,AI 的智能性和安全性同时存在,两者缺一不可。这些看似矛盾的因素在组织中以不同组合方式共存,并相互影响,不断演进,形成不同特点的悖论组合状态。

悖论理论还强调悖论持续存在且不可消除,需要依赖相应的能力实现动态均衡和有效管理,而动态能力理论为此提供了实现路径^[39]。具体而言,企业需要系统培育三类核心动态能力,即感知、捕获与重构,以同步提升 AI 的智能性与安全性;通过“AI 感知”识别智能机遇与安全风险;通过“AI 捕获”在资源约束下配置资源兼顾二者;通过“AI 重构”调整组织与技术架构,实现安全内嵌于智能开发。因此,动态能力实质上构成了将智能与安全的悖论张力从管理挑战转化为可持续竞争优势的桥梁,培养相应的动态能力是同步推动 AI 智能性与安全性进阶的必要途径。综上可得,企业通过 AI 技术建立竞争优势需要同时培育和提高相应的动态能力,从而不断提升 AI 智能性和安全性。

根据悖论理论,有效的悖论管理能促使组织超越“非此即彼”的困境,实现可持续性,而这依赖于

动态能力的支撑。通过灵活配置与所处情境相匹配的动态能力组合,组织可以敏锐地感知智能性与安全性之间张力的变化,并据此调整资源部署与战略重点^[23]。这使得组织能够避免僵化固守,发掘矛盾中的创新机会^[40],实现对悖论的动态管理,最终促使智能性与安全性形成协同,并为可持续的竞争优势创造价值^[23,36]。因此本研究认为,通过有效管理不同的动态能力,可以在不同“智能-安全”悖论状况下影响组织的价值创造和价值维护。

此外,动态能力理论与悖论理论的整合观点认为,组织通过动态能力对悖论进行有效管理可以产生可持续性和超越性,进一步对组织绩效与竞争优势产生影响。这些理论观点同时强调了外部环境对悖论关系以及动态能力实施效果的显著影响,即不存在放之四海皆准的动态能力,组织需要结合外部环境与自身资源能力不断调整^[41]。这意味着,在 AI 技术快速迭代、监管框架逐步完善、社会期望持续提升的复杂环境下,企业不能依赖某种一劳永逸的能力配置。相反,它必须持续评估外部变化对智能-安全悖论关系的具体影响,并有选择性地激活或调整其动态能力的各个维度^[33]。因此可以得出观点,不同悖论条件下的动态能力在影响企业绩效的同时,企业需要根据 AI 技术发展、政策规定和利益相关者多重要求,动态平衡相关动态能力的影响效果。

最后,悖论理论的“悖论持久性”观点意味着,任何在特定时刻建立的“智能-安全”平衡状态都是暂时的,其稳定性将持续受到来自技术革新、主体行为模式转变以及制度环境调整这三方面力量的冲击与重塑。动态能力理论进一步阐释了组织应如何应对这种永恒的动态性。组织必须依靠其持续演化的感知、捕获与重构能力,来追踪、解读并适应这些变化,不仅能够被动响应波动与平衡,更能主动预判并进行前瞻性调整,在动态中寻求并维持一种积极的、适应性的平衡。综上,智能与安全的平衡是动态变化的,这些变化取决于 AI 技术、主体行为和制度规定变化。

3 推进企业 AI 战略协同

根据上述核心观点,企业需要在不同的悖论状态中发展相应的动态能力来管理悖论,因此,针对前文的 AI 情境分类提出以下战略设计:整体突破战略

针对“低智能-低安全”状态下的能力真空,智能强化战略破解“低智能-高安全”状态下的技术局限,安全提升战略遏制“高智能-低安全”状态潜在的风险失控,持续平衡战略维系“高智能-高安全”状态下的二元共生。

3.1 整体突破战略

(1) 情境适用性

该战略聚焦打破低水平均衡,同步提升智能性与安全性的基础水平。在“低智能-低安全”状态下,企业在AI技术与治理体系上均处于初级阶段的能力断层状态,既无法创造价值,又暴露于风险中。因此,企业亟需通过“整体突破战略”在智能性与安全性两个维度上同步构建基本能力,实现从“零基础”到“可控发展”的根本性跃迁。同时,从理论角度来看,悖论理论指出,组织的有效成长始于“矛盾的承认与整合”。整体突破战略要求企业将智能与安全纳入同一能力建设体系,在认知层面实现从对立分离到协同整合的转变。

(2) 能力建设

动态能力理论为这一战略导向提供了操作路径,通过内部认知更新与外部资源获取的双循环,形成突破性能力基础。内部而言,组织亟需构建能力突破,激活动态能力的“感知”与“捕获”机制;通过“感知”力建立面向未来的战略认知,识别智能与安全的双重机会与风险;通过“捕获”机制将有限资源集中于关键领域,实现AI技术利用与安全防护体系的共同起步。从外部游说支持来看,组织应当寻求政策、技术与生态网络的外部链接,打破内部约束,形成能力杠杆。根据悖论理论,资源局限会放大组织的内在张力。在智能与安全双低的初期阶段,企业的能力建设往往受资源基础限制,需要通过获取外部技术支持、政策扶持或资本资源,缓解早期的资源稀缺约束。

(3) 战略效能

在“低智能-低安全”情境下,企业通过整体突破战略实现从“能力真空”到能力突破,但该战略的实施存在代价、依赖性与时滞性,其效果受制于组织的内部突破能力与外部支持能力。

首先,该战略追求智能性和安全性的能力突破,需要大量资源投入,从而影响企业绩效。正如前文观点所述,悖论理论认为资源稀缺性使智能性与安全性的张力显化,同步开发建设需要企业进行资源分流与多目标管理,必然带来短期成本上升。同时,

动态能力理论指出了能力的生命周期与“时滞性”,即动态能力并非瞬间形成。在上述资源投入的短期增大和能力建设的时滞性的叠加影响下,企业绩效在整体突破战略初期呈现负向趋势,且持续投入时间越长,短期压力越显著。因此,本文提出:

假说1a:在整体突破战略下,AI应用(技术开发投入、AI产品与服务的应用等)与企业绩效呈负向关系,且持续时间会加速这一关系。

此外,该战略能否有效施行并带来预期效果取决于相应能力建设情况。悖论理论指出,组织在面对张力时的关键转折在于能否跳出“非此即彼”的单边逻辑、实现认知整合^[20,23]。若企业无法摆脱认知僵化,即便投入资源,也难形成真正的协同突破。因此,整体突破战略的有效性取决于组织能否通过搜寻新出路来打破认知僵化,从而重塑战略方向与资源配置模式。同时,有效的悖论管理离不开动态能力的建设。在此战略中,内部能力提供认知与行动的方向性,外部能力提供资源与制度的可行性。内外部能力共同决定战略的效率与创新性,使组织在“资源稀缺”环境中实现突破。因此,本文提出:

假说1b:整体突破战略有效性(效率、时效、幅度、创新等)取决于企业搜寻新出路或摆脱内部原有惯性的能力和寻求外部支持的能力。

最后,该战略的实施成效高低取决于能力建设的质量与运用水平,而不仅仅是能力的存在。悖论理论指出,组织在感知并接受张力后,需通过动态、持续地调适将协同措施付诸实践^[21,22];动态能力理论进一步表明,战略绩效差异源于能力成熟度与灵活度的差异^[25,35]。具体而言,内外部搜寻能力作为动态能力早期构建的核心,其应用频率、幅度、创新性和时效反映了能力建设的质量与成熟度。当搜寻更频繁、更广泛、更具创新性且响应具时效性时,响应能力建设更加系统和完善,这使得组织对环境变化的敏感度和调适力显著增强,从而提升智能性与安全性双向发展的动态协同性。因此:

假说1c:搜寻能力的应用频率、幅度、创新性和时效决定整体突破战略的效果。

3.2 智能强化战略

(1) 情境适用性

该战略的核心目标是在维持既有安全优势的前提下,优先提升技术智能性,以避免因技术滞后而丧失市场竞争力。在“低智能-高安全”状态下,企业已具备相对完善的风险防控与合规体系,但也暴露

了 AI 技术应用能力的不足,面临的问题更多在于资源的配置僵化而非不足。从理论视角来看,“低智能-高安全”情境体现了悖论理论所描述的“非此即彼”的单边取向战略选择,过度强调安全而抑制了智能发展。然而,正如前文观点所述,智能与安全并非对立,而是共生演化的;组织需根据不同悖论情境匹配相应能力组合以实现动态平衡。智能强化战略通过提升智能性,打破旧有的单边逻辑,使组织在稳健中释放创新潜力。

(2) 能力建设

实现智能强化战略需要组织建设相应的能力,在内部形成功能性智能提升能力,并在外部培育安全门槛协同能力,通过二者的循环互动实现智能性与安全性的动态再平衡。从内部能力建设来看,企业需要依托动态能力理论核心框架的“感知-捕获-重构”机制,激活被安全框架固化的技术潜能。具体而言,企业应强化感知机制,发现算法迭代、数据利用与智能决策的新机会;转变捕获机制,将资源分配和能力建设的中心转向智能创新领域;完善重构机制,进行资源再分配的同时调整组织流程与治理模式,使安全体系内嵌智能模块。从外部能力建设来看,企业应发挥动态能力的“外部塑造性”机制,通过参与标准制定、政策反馈与跨界合作,借助外部信任与政策共识,为智能化创新提供合法性空间。内外能力的动态协同使企业得以在安全约束中释放创新潜能,从而推动 AI“智能-安全”关系由“安全抑制智能”向“安全赋能智能”的转化。

(3) 战略效能

智能强化战略旨在实现从“安全主导”向“智能驱动”的转型,但其实施过程面临路径依赖与制度约束的双重张力,战略成效取决于组织内部功能性能力的重构深度与外部安全标准协同能力的成熟度。

首先,在“低智能-高安全”情境下,企业推进智能强化战略需在稳固的安全体系中重构技术与流程,其初期往往面临显著的学习成本与转换成本。根据动态能力理论,能力生成与应用是一个逐步演化的过程,早期阶段组织需经历感知、捕获与重构的学习曲线,短期投入与绩效提升之间存在时滞。悖论理论指出,当组织从安全主导转向智能驱动时,会打破原有稳定平衡,引发新旧机制间的摩擦与效率波动。因此,智能强化初期可能造成绩效下滑;但随着能力积累与智能体系成熟,组织逐步实现智能与

安全的再平衡,绩效改善趋于显著。由此,AI 应用对绩效的影响呈现出“先抑后扬”的非线性特征,而学习累积与制度适应的效应将进一步强化这种曲线特征。因此:

假说 2a:在“低智能-高安全”状态下,AI 应用与企业绩效呈 U 型关系,且随着应用时间增加,U 型特征越显著。

同时,智能强化战略的实施效果取决于企业能否在内部功能性提升与外部安全协同之间实现动态平衡。根据动态能力理论,不同情境下的能力组合需与组织面临的悖论特征相匹配。如前文所述,对于“低智能-高安全”情境下的企业而言,内部智能提升能力通过技术识别、知识整合与流程重构等机制推动组织打破安全锁定、提升智能水平;而外部安全标准协同能力则通过影响制度规则与行业门槛,缓解创新受限的外部约束。动态能力理论进一步揭示,能力的成熟度、效率与协同性决定了战略产出,只有当内部功能性提升能力与外部安全协同能力均达到较高水平并形成有效互动时,智能强化战略方能释放出预期的协同效能。因此,本文认为:

假说 2b:智能强化战略的有效性取决于内部功能性水平提升能力与外部安全标准协调能力,相关能力发展的深度、效率和程度决定该战略的效果。

从结果来看,当企业通过智能改进实现更高的透明性、响应速度与合规智能化时,外部感知到的风险降低,信任水平上升,从而促进绩效提升。外部利益相关者是影响组织悖论的重要方面,而动态能力理论中的“外部塑造性”指出,企业可以主动影响和塑造外部制度与市场预期,从而构筑有利环境。智能改进实质上是一种外部塑造行为,对处于“低智能-高安全”情境的企业来说,智能改进会向外部利益相关者传递积极信号,表明企业既具备持续创新的实力又能够管理智能、安全多重目标,进一步强化了外部信任。这种信任的提升会转化为实质性的绩效改善。于是,本文提出:

假说 2c:有效的智能改进会增强外部对企业的信任并增加绩效。

最后,“低智能-高安全”情境的企业呈现出明显的单边战略,在智能与安全的悖论框架下,这种防御性战略可能引发组织持续强化惯例的恶性循环。悖论理论指出,在悖论语境下非此即彼的战略选择可能在短期内体现出其价值,但是长期来看会带来意想不到的困难和问题。具体到该情境中,强化安

全短期内或可提升外部合法性与信任感,但长期的防御性应对会加剧组织固化,形成路径依赖。只有在持续动态的平衡中,企业才能维持适应性。由此,过度的安全强化并非真正的稳健,而是以牺牲智能弹性为代价的短期安全。因此:

假说 2d:单纯的安全强化虽会带来更好的外部信任,但会引起与智能水平更大的不平衡而导致绩效下降。

3.3 安全提升战略

(1) 情境适应性

该战略的核心目标是遏制高智能带来的潜在风险失控,迅速提升系统的安全保障强度与治理成熟度。在“高智能-低安全”情境中,企业已具备较高的 AI 技术开发与应用能力,但安全防控相对滞后。从理论视角来看,“高智能-低安全”情境体现了悖论理论所描述的“非此即彼”的单边取向战略选择,过度强调智能性而忽视安全治理的后果。在这种悖论状态下,企业继续单边追求智能优化将陷入恶性循环。只有通过承认并整合矛盾双方,将安全建设纳入智能发展的内在逻辑,才能遏制智能与安全的背离趋势,向动态持续的协同发展转化。安全提升战略正是这一转化的组织机制:它并非否定智能,而是通过强化安全防控、伦理治理与责任机制,使智能发展具有制度约束与社会合法性,使智能与安全动态平衡。

(2) 能力建设

实现安全提升战略,要求组织转变思想,将对智能的追求转移到对安全的夯实上来,并通过动态能力将战略设想投入实践,资源投入和能力培养的中心从技术性能的“先进性”转向风险治理的“可控性”。在该战略下,内部能力建设的重点在于将安全机制嵌入智能体系之中。企业需通过“感知”机制识别风险,通过“捕获”机制将资源配置于安全环节,并通过“重构”机制实现智能开发与安全防控的融合。外部能力方面,组织应当利用动态能力的“外部塑造性”机制,有序优化合作生态,吸取外部资源并学习外部技术与经验,进而弥补自身安全保障和风险防范的不足。通过与外部主体的协同,企业能够在共享信息、共同监测与联合治理中提升系统性安全水平,从而修复因智能快速扩张导致的环境不信任。

(3) 战略效能

安全提升战略旨在弥补治理短板,实现技术先

进性与风险可控性的动态再平衡。其战略成效取决于组织内部能力建设的效果和外部利益相关者的信任与认可。

首先,在“高智能-低安全”情境下,企业的 AI 应用已具备市场竞争力的智能性水平,在初期能够快速抢占市场份额并释放技术红利,从而提升企业绩效。然而本研究提出的悖论观点认为,智能性与安全性之间的悖论关系持续存在。智能性带来的收入虽然能够缓解资源局限,使智能安全的矛盾不那么突出,但二者的张力仍然以隐性悖论的状态存在于组织中。在这种“技术短视”的状态下,潜在的技术失控、数据隐私和伦理风险将随时间推移而不断累积,当风险累积到一定程度时,技术失控和安全漏洞带来的冲击可能会对企业价值造成意想不到的破坏,侵蚀其长期竞争优势,导致绩效下降。因此,AI 应用对绩效的影响呈现出先增加后减小的非线性特征,即倒 U 型关系,即:

假说 3a:在“高智能-低安全”状态下,AI 应用与企业绩效呈倒 U 型关系,且应用时间将强化这一趋势。

同时,根据本研究的理论观点,悖论协同需依靠动态能力,而动态能力的发挥受外部环境影响。因此,安全提升战略的有效性,取决于企业在内部完善安全措施与外部信任环境之间实现协同。在内部能力方面,安全能力建设是动态能力中“重构”维度的具体体现,要求企业对其现有的高智能技术架构、组织流程与治理模式进行根本性调整,实现从“技术短视”向“风险可控”的认知转变。同时,动态能力的有效性依赖于外部环境,而政府和用户是企业最大的风险感知者,他们的信任则构成了企业实施安全提升战略时最关键的外部调节。因此,本文提出:

假说 3b:安全提升战略的有效性取决于企业完善安全措施的能力,而政府和用户对智能技术的信任范围与深度正向调节 AI 应用的效果。

从结果来看,有效的安全改进,如建立系统化的防护机制、增强算法透明度、完善风险监控体系,实质上是向外部利益相关者传递企业具备“风险可控性”与“治理成熟度”的积极信号。本研究的核心观点指出,通过动态能力的灵活配置,组织能够在二者的矛盾中发掘出协同效应,实现可持续性的竞争优势。这种治理能力的提升,能够增强外部对企业能够持续管理高智能复杂系统的“地位认可”,使其被视为技术先进且发展稳健的组织,进而改善客户忠

诚度、融资条件和合作伙伴关系,从而改善绩效。

假说 3c:有效的安全改进会通过提升外部对企业地位的认可而改善绩效。

最后,本研究的核心观点明确了智能性与安全性“同时存在、缺一不可”。因此,在此“高智能-低安全”情境下,若组织坚持过去的组织惯例,采取单纯的智能强化策略,将资源持续投入于技术性能的极大化,而忽视安全合规体系的建设,则会陷入悖论理论所描述的“恶性循环”。尽管这种单边投入在短期内可能表现为更高的效率,但它会扩大智能性与安全性之间的失衡,并导致潜在的技术伦理、数据隐私和系统可靠性风险加速累积和爆发,最终使企业因合规失败或信任危机而遭受重大损失,从而损害其技术优势的长期维护,导致整体绩效下降。因此,本文提出:

假说 3d:单纯的智能强化虽会带来更好的智能发展效率,但会引起与安全合规更大的不平衡而降低绩效。

3.4 持续平衡战略

(1) 情境适用性

该战略的核心目标是持续巩固并发展“高智能-高安全”的高水平动态均衡,在此情境下,组织实现了技术先进性与风险可控性的兼顾。从理论视角来看,该状态是 AI“智能-安全”悖论得到有效管理的结果,悖论并非解决或消失,而是以动态的形式持续存在。即使达到了理想的协同状态,技术发展、主体行为和制度环境的不断演进仍会持续冲击现有的平衡,催生新的矛盾。因此,组织面临的挑战从“如何到达”转变为“如何维持”,这正是持续均衡战略的核心目标。在该战略下,组织需要通过持续的“感知”来预见新风险与新机遇,通过高效的“捕获”来协同配置创新资源与治理资源,并通过灵活的“重构”来不断调整技术架构与治理框架,从而在更高层面上实现智能与安全的动态平衡,维持“良性循环”。

(2) 能力建设

根据本研究的核心观点,组织需要培育高阶动态能力,在持续创造价值的同时维护系统的稳健性。这要求组织在内部实现持续创新与治理协同,并在外部积极构建行业标准以扩展优势。从内部能力来看,企业应当具备敏锐感知,既要预见智能技术的新机遇,更要前瞻性地洞察其伴生的新安全风险;同时优化能力捕获与重构,将“安全内嵌设计”理念融入

技术创新的最前端。从外部能力来看,该状态下的企业同样需要发挥其动态能力的“外部塑造性”机制。根据本文的核心观点,组织需持续调适与平衡动态能力以适应外部环境,利用其在平衡“智能-安全”悖论上的成功实践,主动参与甚至引领行业规范的制定,输出安全标准从而确立优势地位,将短暂的均衡状态转化为可持续的竞争优势。

(3) 战略效能

在“高智能-高安全”情境下,企业已通过有效的悖论管理实现了智能性与安全性的协同。持续平衡战略旨在通过内部创新与外部影响的协同,维持 AI 智能性与安全性的动态协同,从而巩固长期竞争优势。

首先,在该状态下,智能性与安全性之间的矛盾张力得到了有效平衡。根据悖论理论,对矛盾的有效管理能够实现超越性,带来可持续的卓越绩效。在该状态下,AI 的智能性不仅能直接提升运营效率、创新速度与决策质量,同时完善的安全机制还能降低潜在风险,使智能收益得以持续转化为稳定绩效。而随着应用时间的增加,组织基于动态能力的学习效应不断累积,其对智能-安全悖论关系的协同和治理能力日益成熟,使这种正向影响持续增强,形成一个“安全赋能智能、智能提升安全”的良性循环,进一步构成了企业长期竞争力的一环。因此:

假说 4a:在“高智能-高安全”状态下,AI 应用对企业绩效有正向影响,且随着应用时间增加,正向影响持续增强。

同时,持续均衡战略的有效性取决于企业能否将其高阶动态能力同时应用于内部创新与外部塑造。前文核心观点指出,智能与安全悖论关系需要通过动态能力同步提升智能性与安全性,在该战略下具体表现为在内部持续创新能力和在外部影响标准、规则的能力。其中,内部持续创新确保技术与治理的并行进步,而外部标准构建能力则通过影响行业规范与制度结构,实现外部塑造。更重要的是,理论观点表明智能与安全平衡是长期动态,而非静态终点。因此,持续均衡要求企业不仅要同时具备相应的内外部能力,更需要两种能力之间保持协同和共振,持续地应对悖论关系的动态变化。因此:

假说 4b:持续平衡战略的有效性取决于内部持续创新和外部影响标准、规则的能力,这两种能力的协同程度正向影响战略实施效果。

从结果来看,持续均衡战略的最高目标是实现

动态平衡并创造新优势源泉。本研究的核心观点指出,组织需持续调适与平衡动态能力的影响效果,以应对悖论关系的持续动态。当内部持续创新与外部标准影响实现高度协同时,企业便能在自身技术演进路径与外部市场需求、监管要求之间建立起一种动态平衡。根据本研究的核心观点,动态能力最终影响价值的创造与维护。在此,企业创造了新的竞

争优势源泉,即从单纯的技术与合规优势,跃升成为一种生态系统优势,形成更稳固、更难被模仿的长期竞争力。于是,我们认为:

假说 4c:有效的战略实施(即内部持续创新和外部影响标准同步提升)会在企业与外部需求之间建立动态平衡,并催生新的竞争优势源泉。

表 1 AI“智能 - 安全”协同框架
Table 1 AI “functionality - safety” framework

状态组合		低智能 - 低安全	低智能 - 高安全	高智能 - 低安全	高智能 - 高安全
能力建设	对应战略	整体突破	智能强化	安全提升	持续平衡
	内部	寻求突破	提升智能水平	安全能力建设	持续创新协同
	外部	游说外部支持	呼吁安全门槛	优化合作生态	构建标准与扩展优势
绩效影响		负向	U 型	倒 U 型	正向
理论假说		1a, 1b, 1c	2a, 2b, 2c, 2d	3a, 3b, 3c, 3d	4a,4b,4c

4 总结与讨论

本研究以 AI“智能性”与“安全性”悖论为议题,整合动态能力理论与悖论理论,构建了 AI“智能 - 安全”动态协同模型。研究指出,智能性与安全性构成一种组织悖论,企业应通过“两者兼顾”思维与动态能力配置实现双元共生^[21,27,28]。本文进一步提出了一个双维情境分类框架,识别出四类技术状态及对应的协同战略,构建了可操作的 AI 战略协同理论体系。

本研究的理论贡献主要体现在:第一,以悖论管理为核心视角,突破了“智能 - 安全”割裂的传统框架,深化了对 AI 战略悖论的系统理解。第二,整合了悖论理论与动态能力理论,为企业实现 AI 战略目标的动态平衡提供了实践化的理论路径。第三,响应 AI 情境化战略理论号召^[19],构建了可操作的“智能 - 安全”情境分类与协同战略框架,并提出系列假说,推动了 AI 管理研究向动态协同的理论转向。在实践层面,本研究为企业管理 AI 战略提供了指引。企业应认识到智能性与安全性是长期协同而非简单对立;应根据自身所处情境状态,匹配相应的动态能力建设路径;同时,应建立学习与反馈机制,使智能创新与安全治理形成循环共生的演化系统。

总体而言,本文以悖论视角揭示了 AI 智能性与安全性的组织逻辑,融合动态能力理论构建了其协

同管理机制,深化了对 AI 战略悖论的理论理解,也为企业构建动态平衡与可持续优势提供了理论支撑。

参考文献:

[1] 王国燕. 欺骗性人工智能:技术拟真性的风险谱系[J]. 科学学研究,2026,44(5):921 - 930. Wang G Y. Deceptive AI: The risk spectrum of technical realism [J]. Studies in Science of Science,2026,44(5):921 - 930.

[2] 孙丽文,李少帅. 风险情景下人工智能技术信任机制建构与解析[J]. 中国科技论坛, 2022: 150 - 160. Sun L W, Li S S. Construction and analysis of artificial intelligence technology trust mechanism under risk scenarios [J]. Forum on Science and Technology in China. 2022: 150 - 160.

[3] 陈燕萍,邵云飞,黄乐仪. 人工智能能力对企业突破性技术创新的影响[J]. 技术经济, 2025, 44(6): 28 - 39. Chen Y P, Shao Y F, Huang L Y. The impact of artificial intelligence capabilities on breakthrough technological innovation in enterprises[J]. Journal of Technology Economics, 2025, 44(6): 28 - 39.

[4] 文雯,肖瑶,牛煜皓. 人工智能技术水平与企业投资效率[J]. 经济管理,2025,47(6):103 - 122. Wen X, Xiao Y, Niu Y H. Artificial intelligence technology level and corporate investment efficiency [J]. Business and Management Journal,2025,47(6):103 - 122

[5] 陈衍泰,衡予婧. 人工智能与企业竞争优势:研究进

- 展、理论框架与未来展望[J]. 研究与发展管理, 2025, 37(3): 53 - 66. Chen Y T, Heng Y J. Artificial intelligence and enterprises' competitive advantage: Research progress, theoretical framework, and future prospects [J]. R&D Management, 2025, 37(3): 53 - 66.
- [6] 林伟. 人工智能数据安全风险及应对[J]. 情报杂志, 2022, 41(10): 105 - 111, 88. Lin W. Artificial intelligence data security risks and countermeasures [J]. Journal of Intelligence, 2022, 41(10): 105 - 111, 88.
- [7] 孙那, 鲍一鸣. 生成式人工智能的科技安全风险与防范[J]. 陕西师范大学学报(哲学社会科学版), 2024, 53(1): 108 - 121. Sun N, Bao, Y M. Technological security risk and prevention of generative artificial intelligence [J]. Journal of Shaanxi Normal University (Philosophy and Social Sciences Edition), 2024, 53(1): 108 - 121.
- [8] 顾男飞. 善治促善智: 人工智能异化的可信治理[J]. 科学学研究, 2025, 43(12): 2465 - 2472. Gu N F. Good governance for good intelligence: Credible governance of artificial intelligence alienation [J]. Studies in Science of Science, 2025, 43(12): 2465 - 2472.
- [9] Raisch S, Krakowski S. Artificial intelligence and management: The automation - augmentation paradox [J]. Academy of Management Review, 2021, 46(1): 192 - 210.
- [10] Vanneste B S, Puranam P. Artificial intelligence, trust, and perceptions of agency [J]. Academy of Management Review, 2025, 50: 726 - 744.
- [11] Du S, Xie C. Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities [J]. Journal of Business Research, 2021, 129: 961 - 974.
- [12] 束超慧, 王海军, 金姝彤, 等. 人工智能赋能企业颠覆性创新的路径分析[J]. 科学学研究, 2022, 40(10): 1884 - 1894. Shu C H, Wang H J, Jin S T, et al. Analysis on the path of disruptive innovation enabled by artificial intelligence [J]. Studies in Science of Science, 2022, 40(10): 1884 - 1894.
- [13] 杨伟, 刘嫣然, 李垣. 替代或互补: 人工智能应用管理对创新的影响[J]. 科研管理, 2021, 42(4): 46 - 54. Substitution or complementation: The impact of AI application and management on innovation [J]. Science Research Management, 2021, 42(4): 46 - 54.
- [14] 赵剑波, 刘钊. 人工智能渗透率对企业创新效率的影响研究[J]. 经济学动态, 2025(5): 55 - 69.
- [15] 唐露源, 徐源. 中国生成式人工智能产业高效率发展路径研究——基于 TOE 框架的组态分析[J]. 科学学研究, 2026, 44(3): 471 - 481. Tang L Y, Xu Y. Research on the path of efficient development of China's generative artificial intelligence industry - Configuration analysis based on TOE framework [J]. Studies in Science of Science, 2026, 44(3): 471 - 481.
- [16] Babina T, Fedyk A, He A, et al. Artificial intelligence, firm growth, and product innovation [J]. Journal of Financial Economics, 2024, 151: 103745.
- [17] 童云峰. 生成式人工智能技术风险的内部管理型规制[J]. 科学学研究, 2024, 42(10): 2038 - 2046. Tong Y F. Internal management regulation of generative artificial intelligence technology risk [J]. Studies in Science of Science, 2024, 42(10): 2038 - 2046.
- [18] 孙晓宇, 夏保华. 智能化背景下隐私政策的困境及消解路径[J]. 科学学研究, 2026, 44(3): 652 - 660, 672. Sun X Y, Xia B H. Dilemma and resolution approaches of privacy policy in the context of intelligentization [J]. Studies in Science of Science, 2026, 44(3): 652 - 660, 672.
- [19] Kemp A. Competitive advantage through artificial intelligence: Toward a theory of situated AI [J]. Academy of Management Review, 2024, 49(3): 618 - 635.
- [20] Smith W K, Lewis M W. Toward a theory of paradox: A dynamic equilibrium model of organizing [J]. The Academy of Management Review, 2011, 36(2): 381 - 403.
- [21] Schad J, Lewis M W, Raisch S, et al. Paradox research in management science: Looking back to move forward [J]. Academy of Management Annals, 2016, 10(1): 5 - 64.
- [22] 罗瑾琰, 唐慧洁, 李树文, 等. 科创企业创新悖论及其应对效应研究[J]. 管理世界, 2021, 37(3): 105 - 122, 8. Luo J L, Tang H J, Li S W, et al. Research on the paradox of science and technology innovation enterprises and its coping effects [J]. Journal of Management World, 2021, 37(3): 105 - 122, 8.
- [23] 武亚军. “战略框架式思考”、“悖论整合”与企业竞争优势——任正非的认知模式分析及管理启示[J]. 管理世界, 2013(4): 150 - 163, 166 - 167, 164 - 165. Wu Y J. "Strategic framed thinking", "paradox integration" and corporate competitive advantage: An analysis of Ren Zhengfei's cognitive model and its management implications [J]. Journal of Management World, 2013(4): 150 - 163, 166 - 167, 164 - 165.
- [24] Henderson R M, Clark K B. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms [J]. Administrative Sci-

- ence Quarterly, 1990, 35(1): 9-30.
- [25] Eisenhardt K M, Martin J A. Dynamic capabilities: What are they? [J]. Strategic Management Journal, 2000, 21(10-11): 1105-1121.
- [26] Weick K E, Sutcliffe K M, Obstfeld D. Organizing for High Reliability: Processes of Collective Mindfulness [M]. Research in Organizational Behavior, Vol. 21. US: Elsevier Science/JAI Press, 1999: 81-123.
- [27] Smith W K, Binns A, Tushman M L. Complex business models: Managing strategic paradoxes simultaneously [J]. Long Range Planning, 2010, 43(2): 448-461.
- [28] 王象路,罗瑾琰,李树文,等. 组织数字化变革的悖论关系与协同机制研究[J]. 中国人力资源开发, 2023, 40(2): 112-125. Wang X L, Luo J L, Li S W, et al. Research on paradox relationship and synergy mechanism of organizational digital change [J]. Human Resources Development of China, 2023, 40(2): 112-125.
- [29] 弋亚群,李垣,刘益. 企业动态能力的构建及其对战略变化影响的理论框架[J]. 管理评论, 2006(10): 30-34, 64. Yi Y Q, Li Y, Liu Y. The construction of dynamic capabilities and its effects on corporations' strategic change: A theoretical framework [J]. Management Review, 2006(10): 30-34, 64.
- [30] Teece D J, Pisano G P. The dynamic capabilities of firms: An introduction [J]. Industrial and Corporate Change, 1994, 3: 537-556.
- [31] 张亚莉,李辽辽,卢迪. 元知识开发能力对企业颠覆性创新的影响——资源到能力的视角[J]. 科学学研究, 2023, 41(10): 1864-1874. Zhang Y L, Li L L, Lu D. The impact of meta-knowledge development capabilities on disruptive innovation of enterprises —— From resources to capabilities [J]. Studies in Science of Science, 2023, 41(10): 1864-1874.
- [32] Teece D J. Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance [J]. Strategic Management Journal, 2007, 28(13): 1319-1350.
- [33] Helfat C E, Peteraf M A. The dynamic resource-based view: Capability lifecycles [J]. Strategic Management Journal, 2003, 24(10): 997-1010.
- [34] 高洋,葛宝山,蒋大可. 组织学习、惯例更新与竞争优势之间的关系——基于不同环境不确定水平的研究 [J]. 科学学研究, 2017, 35(9): 1386-1395. Gao Y, Ge B S, Jiang D K. A study about the relationships between organizational learning, organizational routines renewing and competitive advantage — under different levels of environmental uncertainty [J]. Studies in Science of Science, 2017, 35(9): 1386-1395.
- [35] Helfat C E, Peteraf M A. Understanding dynamic capabilities: Progress along a developmental path [J]. Strategic Organization, 2009, 7(1): 91-102.
- [36] Fainshmidt S, Wenger L, Pezeshkan A, et al. When do dynamic capabilities lead to competitive advantage? The importance of strategic fit [J]. Journal of Management Studies, 2019, 56(4): 758-787.
- [37] Teece D J, Pisano G, Shuen A. Dynamic capabilities and strategic management [J]. Strategic Management Journal, 1997, 18(7): 509-591.
- [38] Teece D J. Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world [J]. Research Policy, 2018, 47(8): 1367-1387.
- [39] 张志菲,罗瑾琰,李树文,等. 企业数智化变革悖论应对的能力基础:悖论协同能力探索 [J]. 西安交通大学学报(社会科学版), 2024, 44(1): 100-110. Zhang Z F, Luo J L, Li S W, et al. The capability-based view of enterprises to respond to digital intelligence change paradoxes: Developing paradox synergy capability [J]. Journal of Xi'an Jiaotong University (Social Sciences), 2024, 44(1): 100-110.
- [40] 魏龙,党兴华,闫海,等. 悖论中的抉择:如何突破惯例复制困境实现创新催化? [J]. 管理评论, 2025, 37(2): 275-288. Wei L, Dang X H, Yan H, et al. Choice in paradox: How to break through the routines replication dilemma to achieve innovation catalysts? [J]. Management Review, 2025, 37(2): 275-288.
- [41] Ambrosini V, Bowman C, Collier N. Dynamic capabilities: An exploration of how firms renew their resource base [J]. British Journal of Management, 2009, 20: S9-S24.

Secondly, focusing on historical cases where dominant designs were successfully established during similar transitional windows – such as the Wintel alliance, the Android system, and the European GSM standard – this study summarizes general winning principles that can inform the current development of AI in China. These include: technological feasibility as a prerequisite for the formation of a dominant design; the critical importance of dynamically balancing market acceptance and production economics for its widespread adoption; the shift in competitive paradigm from zero – sum game to ecological collaboration as a hallmark of the transitional phase; and the pivotal role of rule – setting power in transforming tri – dimensional coupling outcomes into enduring competitive advantages.

Finally, addressing core challenges such as high computing costs, lagging standardization, and immature ecosystem governance mechanisms in the development of China’s AI industry, this study proposes an adaptive strategy framework encompassing technological breakthroughs, ecosystem development, and rule – making engagement. Specifically, in technological advancement, efforts should focus on overcoming computational bottlenecks and pursuing forward – looking development of key technologies. In ecosystem building, the emphasis lies on creating an open yet controllable industrial ecosystem. Regarding rule – making, it is essential to strengthen standard setting and enhance international governance cooperation. These recommendations aim to overcome critical bottlenecks during the transition phase, foster leapfrog development of China’s AI industry, and build an independent, controllable, and globally influential AI ecosystem.

Key words: artificial intelligence; A – U process innovation model; dominant design; stage assessment; promotion strategies

(上接第 2027 页)

Coordinating paradox: Strategic choices in artificial intelligence “functionality – safety” framework

LI Yuan¹, LI Ping^{2,3}, XU Ya – nan¹

(1. Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai 200030, China;

2. Dongbei University of Finance & Economics, Dalian 116025, China; 3. Xiamen University, Xiamen 361005, China)

Abstract: The rapid development of Artificial Intelligence (AI) technology creates opportunities while simultaneously intensifying concerns regarding its safety risks and alignment problems. Existing research often discusses AI functionality (the technical capability for automated decision – making) and AI safety (the responsible attribute for risk prevention) in isolation, overlooking their inherent paradoxical relationship. This study aims to explore how to coordinate the tension between functionality and safety through strategic design to achieve sustainable competitive advantage. We propose an AI “Functionality – Safety” paradox framework and identify four core scenarios of interaction between functionality and safety. Meanwhile, by integrating perspectives from paradox theory and dynamic capability theory, we formulate the core viewpoints on the AI “Functionality – Safety” paradox. On this basis, the study puts forward dynamically matched strategic paths for different scenarios, explains the non – linear relationships and balance mechanisms among AI application timing, internal and external capabilities, and corporate performance under different strategies. The main contributions of this study lie in constructing an AI technology analysis framework that integrates the functionality and safety dimensions, forming dynamic synergy perspectives on the AI “Functionality – Safety” paradox, and establishing the dynamic matching between AI scenarios and collaborative strategies. This study provides theoretical guidance and practical implications for firms’ paradox collaborative management and AI strategies.

Key words: artificial intelligence; AI functionality; AI safety; dynamic capabilities; paradox theory