KPMG
毕马威

CISCO

人工智能就绪度白皮书

企业数智化转型的AI变革路径与评估指南

序言

驾驭AI浪潮，筑牢数智基石

侯胜利

思科大中华区资深副总裁
暨首席技术官

在人工智能技术的驱动下，特别是大模型的广泛应用，我们正步入一个充满无限可能与深刻变革的时代。AI不再是遥不可及的未来畅想，而是赋能千行百业、重塑核心竞争力的关键引擎。AI浪潮带来了前所未有的机遇，同时也对数字基础设施提出了严峻的挑战。海量数据的处理、极致的低时延交互、复杂的模型训练与推理，都呼唤着一个更加敏捷、高效、可靠且智能的AI基础设施底座。

《人工智能就绪度白皮书：企业数智化转型的AI变革路径与评估指南》的发布，恰逢其时。它凝聚了思科与毕马威在企业AI变革转型的深度思考与实践洞察。我们深刻认识到，企业若想在这场AI变革中占得先机，坚实可靠的基础设施与配套服务是成功的基石。

针对AI应用数据量大、低时延的特性，我们必须重新考虑基础设施的建设理念。首先，我们倡导选择开放的、标准化的信息技术。这不仅能够满足AI应用日新月异的需求，更能通过标准化、通用化来降低企业的初始投入和长期运维成本，显著提高基础设施的利用率，避免技术锁定带来的风险。其次，为了保障AI应用的高性能和优质服务，我们必须积极引入和部署防止网络拥塞、避免数据丢包以及高效处理多媒体数据的先进技术。这些技术的应用，将直接关系到AI模型训练的效率和推理的精度。

伴随AI新技术的广泛应用，安全与合规问题也日益凸显。我们必须构建完善的AI安全防护体系，通过实时检测与监控，确保AI应用的合法合规运行。同时，要着力排除错误数据和恶意干扰，保障AI大模型推理结果的稳定性和可靠性，这对于维系用户信任、保障业务连续性至关重要。

本白皮书深入剖析了“AI Ready”的数据中心在计算、网络、存储及相关治理组件等关键要素，并前瞻性地提出了AI变革就绪度评估体系。我们希望，这不仅能为企业决策者提供清晰的指引，更能为技术管理者提供具体可落地的行动方案。我们相信，通过构建面向未来的AI基础设施，企业定能在这场波澜壮阔的智能化转型中，乘风破浪，行稳致远。

序言

共绘AI蓝图，加速变革征程

张令琪

毕马威中国科技咨询业务
主管合伙人

人工智能，尤其是生成式AI和大模型的崛起，正以惊人的速度重塑商业模式、运营流程乃至整个产业生态。这不仅仅是一次技术迭代，更是一场深刻的认知革命与生产力变革。企业如何在这场浪潮中找准定位、把握机遇、从容应对挑战，成为摆在每一位领导者面前的时代课题。

《人工智能就绪度白皮书：企业数智化转型的AI变革路径与评估指南》白皮书，是我们与思科中国智慧碰撞的结晶。我们深知，AI的成功落地，绝非单一技术或部门的努力，而是一个涉及企业战略、组织文化、数据要素、技术架构乃至人才培养的系统工程。本白皮书旨在为企业提供一个清晰的导航图，助力其全面审视自身的AI变革之路。

在白皮书中，我们系统梳理了企业在拥抱AI过程中的关键维度。从企业决策侧的战略决心与顶层设计，到组织体系侧的人才培养与流程再造；从基础设施侧的算力、网络、存储支撑，到数据语料侧的高质量数据准备与合规利用，每一个环节都至关重要。我们特别强调，坚实的基础设施是AI应用高效运行的“底盘”，而科学的评估体系则是确保“底盘”稳固、方向正确的“罗盘”。

第三章重点聚焦及解析的关键要素，以及第四章提出的AI变革就绪度评估体系，是我们基于大量行业实践和深度研究的核心成果。我们认为，仅仅拥有先进的AI算法和模型是不够的，企业更需要一个能够支撑这些“智慧大脑”高效运转的强大“身体”，以及一套能够客观衡量自身AI成熟度并指导持续改进的科学方法。通过这套评估体系，企业可以清晰地认识到自身在AI变革征程中所处的阶段，识别关键短板，并获得针对性的改进建议，从而更精准地投入资源，加速实现AI价值。

我们非常荣幸能与思科中国联袂发布此白皮书。我们相信，通过将思科中国在基础设施领域的深厚积累与我们在管理咨询、行业洞察方面的专业能力相结合，定能为广大企业提供兼具前瞻性与实操性的指导建议。愿本白皮书能成为您在AI变革道路上的得力助手，共同迎接智能时代的无限可能。

目录

摘要

“致广大”以谋事，“尽精微”以成事。在人工智能革命浪潮这一宏大叙事下，企业及个人等微观主体面临着一道时代必答题，即如何摆脱思维惯性和路径依赖以实现“AI Ready”（人工智能就绪）？

本白皮书从偏宏观层面的发展趋势分析出发，按照产业机遇研判、企业AI变革实践调研、关键变革要素细致剖析、拆解产品服务案例提炼有益方法论的研究思路逐层深入，以求助力企业从精微处快速采取行动。主要洞察如下：

紧跟前沿态势，认识把握AI浪潮下的产业机遇

从技术演进来看，AI发展遵循着“感知AI (Perception AI) -- 生成式AI -- (Generative AI) --代理式AI (Agentic AI) -- 物理AI (Physical AI)”的路线。当前中国AI大模型发展正由“暴力美学”转向“成本效益革命”，由此带来了“基础模型开源化+垂直领域私有化定制”加速AI普惠化、生成式AI加速渗透至千行百业等趋势。

从治理环境来看，中国内地在AI领域已形成以政策战略、专项法规、标准建设和技术赋能为核心的多维度治理体系，港澳AI治理体系则注重国际化和本地特色产业的融合，整体上有助于打造健康安全发展的AI生态。

从企业变革来看，大模型技术迭代升级使得企业对于AI技术从“看得见却用不起”到“用得起也用得好”转变，AI Agent等应用的崛起助力企业的应用场景加速向纵深发展，企业提前布局AI生命周期安全防护等多重机遇下，企业的AI Ready水平关乎其能否在AI浪潮中夺得先机。

具体而言，企业的AI Ready共包含两大一级能力，即企业的“硬实力”和“软实力”，以及七大二级能力，即企业AI Ready的七大核心评估维度，其中技术、数据、业务主要对应“硬实力”；战略、治理、人才和组织结构则对应“软实力”。

聚焦企业AI实践，调研总结企业AI变革之路

毕马威与思科携手开展专项调查，面向全国范围内已进行AI战略部署并且在业务中有明确AI落地场景，并对期望借助大模型技术进一步扩大的泛行业企业进行了问卷调研，以期深入了解企业AI战略认知与布局、AI变革需求与核心挑战、AI体系变革路径等最新实践。

综合调研结果来看，部分受访企业在推进AI变革过程中已逐步形成覆盖技术架构侧、数据语料侧、基础设施侧、组织体系侧的系统化推进思路。主要呈现以下特征：

- **在技术架构侧**，企业通常以落地场景的体系设计为起点，采取混合部署方式推进大模型落地应用，并统筹考虑云端风险应对、私域安全防护、模型幻觉应对等问题，以重构技术底座，适应创新之变；
- **在数据语料侧**，企业着眼于数据治理框架搭建和数据质量提升，推进数据语料的深度治理，并且已初步形成数据标准化体系、智能清洗工具等共性选择；
- **在基础设施侧**，企业兼顾创新与务实，充分权衡AI基础能力建设与业务需求满足，通过混合部署和协作管理措施等实现协同优化；
- **在组织体系侧**，企业立足敏捷性和协同性的组织机制要求，大力推进AI相关的团队能力建设和员工风险应对等，以软实力建设护航组织AI的硬核转型。

立足企业AI Ready硬实力变革视角，精准剖析关键要素

硬实力是企业AI Ready的底层基础，在AI发展阶段性态势下，以“预测下一个token”为核心的技术范式兴起，技术、数据、业务等硬实力要素不断突破固有边界，在token化的演变中迎来价值重构。

在业务应用层目标驱动下，纵向深挖基础设施层、模型服务与编排层的关键模块，横向细究服务治理层的重要构成，可以精准把握各要素的核心价值逻辑和变革新特性，助力相关企业锚定前进方向。

基础设施层包含计算资源、网络架构、存储系统及数据语料等核心要素，旨在为上层AI应用提供稳定、高效、可拓展的运行环境。其中，计算要素的AI变革新特性包括开箱即用、一物多用、安全保障等；网络要素则表现为由“训推一体”的算力架构演进到网络架构、由软件定义网络演进到意图网络；存储要素表现为存算一体、冷热数据自治；数据要素表现为数据价值链重构、数据资产化。

模型服务与编排层包括多模型管理与服务化、智能体与应用编排以及模型通信协议与集成，旨在屏蔽底层基础设施的复杂性，赋能上层应用的敏捷创新，确保企业能够高效、安全、可扩展地利用各类AI模型驱动业务价值。其中，多模型管理与服务化具有动态资源适配与调度、模型市场与发现机制、面向特定场景的模型优化服务等AI变革新特性；智能体与应用编排要求自适应与自学习编排、可解释性与可追溯性等。

服务治理层包括安全可信AI和AI全栈治理两大支柱，旨在推动企业AI安全及治理从原则规范走向工程化落地。安全可信AI要求实现可视化、强检测、广覆盖、可落地等AI变革新特性。AI全栈治理包括从底层基础设施到上层应用的全技术栈联合治理和全栈指标定量建模，需要为上层AI应用构建起无形但又无处不在的体验保障，并实现AI价值显化。

面向泛行业企业发布AI Ready变革评估体系，助力快速行动

AI Ready变革评估体系包括企业架构、数据语料、基础设施、组织体系等4大评估维度，每一维度下拆分出不同的评估指标（共计13项），并进一步细分出二级评估指标（共计41项），评估每项指标的就绪度等级后综合计算得出企业整体AI Ready的对应等级。

企业可结合该体系和相关模板客观评价自身的AI发展水平，明晰在行业竞争中的相对位置，并结合“以评促建--价值为锚--安全为纲--架构先行--筑牢底座--内外兼修--快速迭代”的AI Ready“七步”变革行动指南制定面向更高等级的能力提升计划。

01

新兴AI浪潮下的产业机遇

科技浪潮滚滚向前，人工智能（AI）发展更是日新月异，从ChatGPT问世到DeepSeek发布，AI领域的每一次技术突破，都堪称举世瞩目。一幅“感知AI (Perception AI) -- 生成式AI -- (Generative AI) --代理式AI (Agentic AI) -- 物理AI (Physical AI) ”的技术跃迁图景已悄然铺开，正在引发底层架构变革、模型能力演进、基础设施格局调整、应用场景创新等连锁反应。

1.1 中国AI发展的阶段性态势

模型架构与能力演进

大模型发展由“暴力美学”转向“成本效益革命”。2025年以来，DeepSeek凭借“低成本、高性能、开源共享”的技术路线轰动全球，其在硬件条件有限的情况下，通过混合专家架构（MoE）和强化学习等技术，实现了训练效率的大幅提升和算力成本的有效缩减。这不仅引发了科技界对于Scaling Law是否依然有效的质疑，更促使国内外基础模型厂商加快调整竞合策略，提高各自模型服务的成本效益。例如，国内多家互联网大厂和大模型创企宣布将旗下产品接入DeepSeek-R1，上线“深度搜索”功能、提供“满血版”服务等，以较低成本融合DeepSeek技术优势的同时，强化了自身的差异化服务能力。相应地，大模型市场的竞争焦点由模型参数等底层技术比拼，逐渐转向用户流量争夺、垂直场景的轻量化解决方案等，免费策略成为各家抢占市场份额的一大利器。例如，OpenAI不仅向免费用户开放了o3-mini推理模型的API，还宣布GPT-5将向免费用户开放。技术路线和市场竞争的双重转变下，大模型的“成本效益革命”正在拉开帷幕。

“基础模型开源化+垂直领域私有化定制”加速AI普惠化。科技界的开闭源之争由来已久，而近年来国内各类开源社区、开源基金会的设立，极大促进了开源生态的壮大。根据工信部信息，2025年中国已成为全球开源参与者数量排名第二、增长速度最快的国家[1]。根据国际大模型评测榜单LiveBench，全球排名前十的模型仅两款开源且均来自中国，分别是QwQ-32B和DeepSeek-R1[2]。此背景下，数量庞大的开发者和中小企业依靠“基础模型开源化+垂直领域私有化定制”就能快速部署定制化的专属大模型，将加速推动AI能力惠及更多行业和个人。

模型推理能力不断增强将驱动AI持续跃迁。生成式AI强调AI能够“知道”和“理解”信息，从而生成答案；代理式AI则强调AI具备“主动性（Agency）”，能在理解信息的基础上进行推理，思考如何回答或解决问题，并进一步制定和执行计划。例如，DeepSeek通过显性的思维链展示了推理过程，用户不仅能观察到其认知活动，还能通过提问来深挖底层逻辑，并利用多轮对话来参与优化推理逻辑，在一定程度上反映出了AI推理能力的提升。进一步地，物理AI强调AI能在物理世界中发挥作用，例如赋能自动驾驶和机器人等，这将要求AI能基于多模态信息实现更为复杂的推理能力。例如，DeepMind团队推出的机器人控制模型Gemini Robotics，是基于Gemini 2.0的多模态世界理解能力，接入物理动作输出模态而实现。此外，该团队还推出了聚焦空间推理的Gemini Robotics-ER，显著提升了模型对世界的理解能力。

基础资源发展与结构性变化

算力需求仍将持续攀升，算力生态将逐步转向分布式算力网络。AI普惠化和持续跃迁趋势下，大模型预训练、微调、推理的算力需求仍将持续爆发。分布式计算架构能将对单一节点的算力需求分散到多节点，通过连接各计算节点实现高效协同计算，进而降低计算成本，正在成为新一代计算范式，也是国产算力突围的关键。公开数据显示，中国的算力总规模早已超过230EFLOPS（FP32），位居全球第二[3]。但是，国内智算中心平均算力使用率仍较低（约30%），远低于大型数据中心（50%-60%）[4]。预计随着分布式计算架构与边缘计算、存算一体等技术深度融合，超大规模的中心式算力会有所减少，算力生态将从“超大规模中心垄断”逐步转向“分布式算力网络”。

数据质量决定模型核心竞争力，亟待全方位升级。目前数据已被归入新质生产力的范畴，但高质量数据仍较稀缺，主要问题有中文开源语料稀缺、数据来源受限、垂直领域数据匮乏、数据更新迟缓、数据标准尚未统一等，致使模型训练常遇瓶颈。长期来看，数据质量是决定模型竞争力的核心要素，需在规模、质量、广度、深度、效度、精度等方面不断升级。此外，各国对数据隐私与安全问题的监管趋严，欧盟出台了《通用数据保护条例》、中国则颁布了“三大基本法[5]”，意味着相关各方在促进数据质量提升的同时，必须时刻兼顾数据安全。

[1] 工信部：中国已成为全球开源参与者数量增长速度最快的国家，中国新闻网，2025年3月23日，https://www.chinanews.com/cj/shipin/cns-d/2025/03-23/news1016488.shtml
[2] https://livebench.ai/#/
[3] 《中国算力发展报告（2024年）》，中国算力大会，2024年9月
[4] 突破五大制约因素：智算中心如何提高利用率，通信产业网，2024年11月9日
[5] 《中华人民共和国数据安全法》《个人信息保护法》《网络数据安全管理条例》

基础设施即“AI工厂”，企业级基础设施需积极应变。如果将“智能”比作一种产品，那么推理的本质就是在生产token，AI推理需求的爆发意味着要大规模生产token。此背景下，传统的以硬件设备为核心的基础设施正在演变为“AI工厂”，需要以算力、数据等基础资源为原材料，以极高效率实现token的集约化、规模化生产。纵向扩展（Scale Up）和横向拓展（Scale Out）都很重要，可以先将单机、单系统的规模尽量扩大，然后再做分布式扩展[6]。其中，企业级AI基础设施既是这个庞大且复杂的系统中的关键节点，又会是相对独立的AI工厂，企业需积极变革集中式的IT基础设施以适应新变化。具体来说：

计算方面

越来越多企业倾向于在云端进行大模型预训练，而在本地进行基于本地化数据的模型微调、RAG、推理等，最终形成更小规模、集约化和高知识密度的AI训推一体形态，所占用的资源会低于大型智算中心几个数量级。以本地化智算中心（AI DC）为例，在以GPU为计算核心的服务器架构中，当前主要有四类组织形态（表 1）。其中，算力更为强大的集成式GPU服务器和多GPU互联服务，更适合大规模的模型预训练任务。而对于尚处AI发展初期阶段的企业或中小型企业而言，大规模预训练的投入仍会超出其承受范围，则更适合选用模块化GPU服务器。

表 1　AI DC以 GPU 为计算核心的服务器架构组织形态

组织形态	特点
传统 GPU 服务器	采用 x86 架构的CPU与GPU组合。GPU通过PCIe接口连接到主板，形成高性能计算节点。 该类架构在2024年之前广泛应用，服务器厂商可根据需求选择CPU和GPU组合。
集成式 GPU 服务器	将CPU和GPU集成在同一芯片上，减少数据传输延迟，提高计算效率。
模块化 GPU 服务器	采用模块化设计，支持多种GPU形态和CPU选择。
多 GPU 互联服务器	利用多GPU之间的高速互联，提升数据传输速度和计算性能。

资料来源：Cisco，毕马威分析

[6] 2025 NVIDIA GTC 主题演讲

存储方面

为满足AI任务在实际业务中的高效运行和低成本实现，企业的存储系统需满足能支持多类型数据、多任务并行处理、本地化数据安全管理等需求，分布式存储支持横向扩展以适应企业业务增长，正在成为必选项。

网络方面

企业部署大模型通常需要配备高性能的多核CPU和多个GPU，以获取强大的并行计算能力，因此也需要高带宽、低延迟的网络环境来支持节点间的数据交换。随着大部分企业选择从训推一体的融合架构中逐步演进AI网络，以太网具备通用性、灵活性等优势，渗透率有望持续提高。

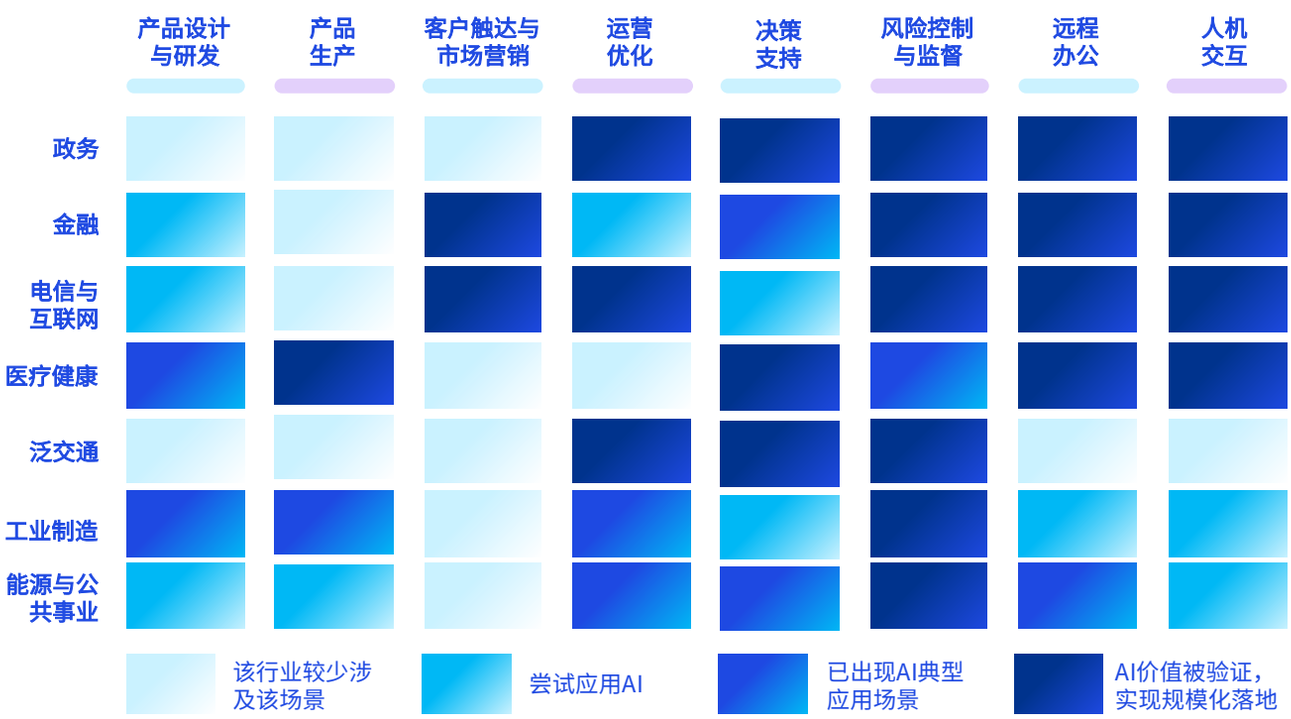
一体化IT基础设施解决方案愈发成为刚需。对于企业自身而言，比建设AI基础设施更为重要的是让AI创造出业务价值。因此，需要将有限的时间、精力投向更为重要的地方，面向企业AI发展需求的一体化IT基础设施解决方案愈发成为刚需。当前，部分厂商会基于模块化GPU服务器提供其验证的最佳实践设计，即按不同AI任务的规格要求，把GPU、AI平台软件、云原生系统软件等全栈生态用最佳实践设计打包形成一体化算力平台解决方案，抑或是提供覆盖深度学习领域推理和训练全流程的“端-边-云”全场景AI基础设施方案，极大降低了企业使用AI的门槛。

微服务架构下AI业务的全栈治理尤为关键。企业应用架构大致经历了从单体架构到SOA架构，再到微服务架构的变迁。当前大部分企业的AI应用都采用了微服务结构，可以把AI功能打散后，将每个功能（即“微服务”）分布到多个计算节点上，由此形成了较长的微服务链条，保障AI应用可以根据需求弹性伸缩。但是，这也带来了潜在风险，这对安全和运维提出了极高要求。一方面，微服务链条上任何节点的性能和安全缺陷都会损害整体体验；另一方面，AI业务的复杂性不仅体现要依赖大规模计算和存储资源，还表现为需协同多层服务和组件，以最终交付清晰的业务价值。因此，实现企业AI业务的全栈治理将是确保AI应用高效、可靠、安全运行的关键。

应用场景趋势

生成式AI加速渗透至千行百业，应用场景持续拓宽加深。截至2024年底，共302款生成式AI服务在国家网信办完成备案，当年新增238款，生成式AI产品的用户规模达2.49亿人，约占整体人口的17.7%[7]。从应用的广度来看，鉴于各行业智能化转型成熟度各不相同，生成式AI应用呈渐次铺开的态势，主要由电信与互联网、政务、金融等行业，逐步拓展至医疗健康、泛交通、工业制造等多个行业。从应用的深度来看，生成式AI应用当前侧重于解决实际操作层面的降本增效需求，包括人机交互较频繁且重复性任务较多的智能客服、风险合规等场景，且有望愈发深入参与到企业核心业务流程中，满足决策支持、战略管理等场景需求。但值得注意的是，部分企业在实践过程中常面临内外部双重挑战，在一定程度上限制了生成式AI应用的拓宽加深。在企业自身禀赋方面，其可能存在智能化转型基础较薄弱、数据资产积累不足、算力投入有限等短板；在外部发展环境方面，大模型技术仍在持续迭代，且算法黑箱、模型幻觉等问题仍存在较大优化空间，十分考验企业紧跟技术前沿、精准研判趋势、科学合理决策的能力。

部分行业智能化转型的场景成熟度



资料来源：毕马威分析

[7]《第55次中国互联网络发展状况统计报告》，中国互联网络信息中心，2025年1月17日

专题

生成式人工智能在企业前中后台的应用案例

生成式人工智能在营销领域的五个新兴应用案例

精心打造的客户体验有助于吸引更多客户并提高留存率，从而提高销售利润

1 提升个性化和精准营销的覆盖度

利用人工智能驱动的数据分析来开发潜在客户并回答客户问题，以提高销售漏斗上层的转化率。

增加潜在客户数量

2 更多的个性化内容

利用人工智能提供更好的个性化体验，例如通过使用满意度高的和优化的定价来加速购买决策过程。

提高转化率

3 更好的用户体验

使用人工智能进行自动化验证和用户引导，以降低成本、提高客户满意度并加快实现获利。

加快用户引导进程

4 增强用户隐私保护

使用人工智能增强客户隐私、安全和优化自助服务支持，并预测潜在问题，例如欺诈活动。

增加信任

5 提升客户留存率降低流失率

使用人工智能，通过提供专注于提升客户终身价值和忠诚度的个性化体验，来主动留住客户。

提高留存率

潜在效益和生产率提高

生成式人工智能在采购领域的五个新兴应用案例

生成式人工智能具有实现采购流程自动化、优化和转型的潜力，进而提升采购效率并推动战略决策

1 更新品类管理

人工智能的使用自动化了类别管理，综合内部/外部数据以制定动态策略，并通过持续评估预先识别潜在风险。

动态的品类策略

2 提高供应商选择效率

利用人工智能实现采购中的 RFx（Request for x，需求申请/报价请求）流程自动化，简化了需求生成、RFx创建、供应商评估和决策的环节。

简化的采购决策

3 动态合同周期管理

人工智能在合同管理中的使用自动化了合同创建、谈判支持、合规性验证、风险检测、绩效跟踪和续签流程。

优化的合同管理

4 透明可视化的供应链洞察

在购买过程中可利用人工智能帮助客户发现所需产品，简化购买流程，并指导用户选择合规供应商。

简化的合规采购

5 革新财务采购

人工智能在采购财务中的使用优化了发票处理、付款计划、对账和现金管理，并增强了预测洞察力。

优化的财务管理

潜在效益和生产率提高

专题

生成式人工智能在企业前中后台的应用案例

生成式人工智能在供应链中的五个新兴应用案例

生成式人工智能具有显著提高供应链效率、辅助供应链决策、提升合规性、优化供应链维护和推动可持续性的潜力

1 人工智能驱动的集成规划和场景建模

利用人工智能分析数据，实现预测性洞见和规划，生成战略性供应链场景，以便做出明智的、适应性强的决策。

先进的预测性规划

2 分布式订单履约

人工智能在订单履约中的使用通过实时分析优化了分拣、包装和交付路线，提高了效率和客户满意度。

简化的订单履约

3 人工智能强化质量管理

人工智能的使用彻底改变了质量控制，实现了文件处理自动化，并促进了实时异常检测，从而提高了效率和产品可靠性。

自动化质量鉴证

4 AI强化的供应链可见性

使用人工智能驱动的工具绘制多层次供应商网络，分配风险评分，整合监管要求，并提高对遗传风险 和集中风险的可见性。

集中风险可见性

5 智能需求与网络优化

使用人工智能进行精确的需求预测、细分市场分析、生成服务成本洞见、战略网络和库存优化，确保实时的适配能力和透明度。

自适应预测优化

潜在效益和生产率提高

生成式人工智能在财务领域的五个新兴应用案例

财务部门应成为整个企业采用生成式人工智能的中心；在财务职能中，生成式人工智能可以提高整个企业的理解能力和叙事能力

1 跨职能绩效评述

使用人工智能为财务报告和预测生成辅助性的叙述内容，从而增强对风险和趋势的叙述和理解。为投资者电话会议创建内容。

提高对财务业绩的理解

2 合同生成和审查

使用人工智能提升合同生命周期 — 生成具有首选条款和措辞的标准合同 — 监控和审查现有非标准合同的风险和利润 情况。

降低财务风险水平

3 报价至收款流程优化

使用人工智能分析大数据集，以确定应收账款账龄及拖欠率的模式和趋势。

改善现金流

4 金融监管合规

利用人工智能将新法规（如与美国一般公认会计师准则、国际财务报告准则或税务相关的法规）简化为业务职能的关键要点。通过生成式人工智能整合数据，创建每位客户的精准画像，从而改善金融犯罪，提高合规性。

识别和管理合规风险

5 行业对标分析

使用人工智能分析季度盈利报告（和其他公开文件），以获取竞争对手的见解并了解趋势，从而为自己的战略方向提供信息。

更好地了解自己如何与他人进行比较

潜在效益和生产率提高

专题

生成式人工智能在企业前中后台的应用案例

生成式人工智能在人力资源领域的五个新兴应用案例

一些人力资源领导者认识到人工智能在塑造未来劳动力方面的作用，能够改善员工体验并提高留存率

1 个性化入职培训

为新员工创建个性化的入职培训计划、核对表、培训课程和入职流程以及欢迎信息。回答新员工所遇到的常见问题，并提供政策、程序和福利方面的指导。

简化入职流程

2 加强员工学习

在有需要时，利用人工智能开发有吸引力的个性化学习课程。个性化的职业发展建议可以提高员工满意度和工作投入度。

使课程内容与需求相匹配

3 提升招聘体验

创建符合企业战略目标的定制职位描述。根据职位角色和要求生成筛查问卷和面试指南。

实现更高效的招聘

4 以人为本的员工支持

及时反馈、有同理心地对待员工（或候选人）的具体情况或问题。人工智能还可以主动将问题转给相应的团队，在需要时立即上报，并用恰当的语言进行回答。

增强员工和候选人的体验

5 全面且规范的劳动力分析

使用人工智能整理来自多个来源的数据，以建立更完整的人力资源报告和更深入的智能数据，为劳动力提供信息支持。

全方位了解人力资源和劳动的表现情况

潜在效益和生产率提高

生成式人工智能在IT管理领域的五个新兴应用案例

生成式人工智能能够提高IT 方案的交付质量和IT投资的回报率，加快创新科技 部署 并支持 人才 发展

1 提高服务可用性和性能

使用智能运维（AIOps）和预测性维护，通过智能警报、根因分析、异常和威胁检测、自动修复、自动补丁和容量优化功能来提高服务性能。

最大限度提高用户可用性

2 推动创新，加快功能开发

使用生成式人工智能生成代码和测试运维（TestOps）功能快速进行原型设计，加快新功能的上市时间，同时使得企业有更多时间用于发现和创新。

以市场速度部署

3 预防安全事故以及事后的快速响应

利用人工智能作为安全运营中心（Security Operations Center，SOC）的增强器， 提升威胁监测和解决的速度，以及利用人工智能的高级监测能力自主采取预防措施，并在必要时切换至SOC验证。

减少安全事故和风险

4 自动化IT文档和日程安排

将机器人流程自动化（Robotic process automation，RPA）和生成式人工智能用于IT知识资产的开发和管理，如故障排除指南、终端用户常见问题解答、流程文档或职位描述，或用于IT劳 动力调度和优化。

提升IT生产率并降低成本

5 员工体验的提升与个性化

使用人工智能驱动的内容搜索、定制和汇总功能，为不同的用户角色提供个性化的结果。这包括使用对话式人工智 能为不同的员工原型定制培训和终端用户支持。

提升用户和IT员工的技能并提高满意度

潜在效益和生产率提高

资料来源：毕马威分析

企业对生成式AI应用的态度回归理性，亟需验证场景价值。技术越是迅猛进步，应用越是蓬勃发展，相关主体越是要谨防盲目跟风。当前，各类企业对待生成式AI的态度已逐渐摆脱早期的技术狂热，愈发聚焦于实际场景价值。从投入侧来看，企业在大模型投入价值计量、算力需求评估、数据成熟度评估、安全对齐成本测算等方面仍较缺乏体系化思维；从产出侧来看，风控、投研等决策类场景存在ROI难以量化的问题，非决策类场景则存在短期收益和长期成本失衡的问题。由此，企业往往会陷入“不会投、不敢投、不能投”的困境。Gartner的调研数据显示，中国企业在部署生成式AI方面尤为谨慎，2024年6月中国企业的生成式AI采用率约为8%，而全球水平约为21%[8]。然而，生成式AI的长期战略价值不容忽视，相关企业亟需建立有效的成本统筹机制、大模型价值指标体系等，筛选出真正契合业务需求且能实现相应商业价值的应用场景。

[8] 目前仅8%的中国企业将生成式人工智能部署在生产环境中，Gartner，2025年1月8日

1.2 区域性AI治理环境及发展趋势

中国AI产业发展概况

从AI产业链来看，可以分为上游的基础层，中游的技术层，以及下游的应用层。目前，中国的AI产业链呈现出“基础层攻坚、技术层突破和应用层分化”的阶梯式发展特征。在基础层，中国的算力规模已经占到全球的30%左右，居全球第二位，但仍然需要突破高端芯片制造和数据标注自动化等技术；在技术层，DeepSeek开启了通用大模型开源和效率优化的先河，但如何产生垂直领域的落地价值仍亟需产业打通“端到端”的闭环；在应用层，B端市场如工业质检、智能制造等已经实现了深度渗透和大规模的效率优化，在C端市场的金融风控、医疗诊断等也不断实现场景创新，但如何从工具赋能升级到生态重构，从而在全球竞争中实现“领跑”仍是一大挑战。

从产业规模来看，根据2025年3月两会数据披露，中国AI核心产业规模已接近6,000亿元，并且仍在快速增长。今年的《政府工作报告》中还提出，要大力发展智能网联新能源汽车、人工智能手机和电脑、智能机器人等新一代智能终端以及智能制造装备，今年有望成为AI端侧应用的爆发年。

区域性AI治理环境

内地AI治理环境逐步成熟，坚持发展与治理并重

当前，中国内地在AI领域已形成以政策战略、专项法规、标准建设和技术赋能为核心的多维度治理体系。

中国将人工智能的发展和治理已上升至国家战略，成为引领新质生产力发展、加快构建现代化产业体系的战略性技术。2024年7月国家提出，要完善人工智能发展政策和建立安全监管制度等治理体系，强调了发展与治理并重的理念；2024年12月的相关会议和2025年3月的政府工作报告均提出要推进“人工智能+”行动，明确了AI发展的目标和路径。

法律法规围绕伦理规范、数据安全、行业特殊监管等方面逐渐优化监管及治理环境，旨在打造健康安全发展的AI生态。比如，2025年3月14日颁发的《人工智能生成合成内容标识办法》，明确了人工智能生成合成内容标识主要包括显式标识和隐式标识两种形式，显式标识是指在生成合成内容或者交互场景界面中添加的，以文字、声音、图形等方式呈现并可以被用户明显感知到的标识；隐式标识是指采取技术措施在生成合成内容文件数据中添加的，不易被用户明显感知到的标识。2023年7月，国家网信办等七部门新公布的《生成式人工智能服务管理暂行办法》，针对AI大模型，对内容的安全性、准确性和可靠性提出明确要求；同年9月，科技部等部门发布《科技伦理审查办法（试行）》，针对科技活动中的科技伦理审查主体、审查程序、监督管理等方面进行了规范；2024年6月，国家药监局发布的《药品监管人工智能典型应用场景清单》，明确了AI技术在药品监管领域的应用边界。

国家标准体系建设和行业规范不断完善，引导人工智能产业规范化、高质量发展。2024年3月，全国网络安全标准化技术委员会出台《生成式人工智能服务安全基本要求》，该标准在语料采集、语料标注、内容安全监测、服务稳定性等方面提供了量化评估标准；同年7月，工信部等四部门联合印发《国家人工智能产业综合标准化体系建设指南（2024版）》（以下简称《指南》），该指南明确了人工智能产业各环节的重点标准方向，并提出到2026年，标准与产业科技创新的联动水平持续提升。

政策驱动AI应用拓展和产业化发展，促使AI治理环境向更加全面、灵活和包容的方向发展。2022年以来，科技部先后出台了《关于支持建设新一代人工智能示范应用场景的通知》《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》等相关政策支持建设新一代人工智能示范应用场景，推动AI与实体经济深度融合。据不完全统计，2024年全国至少有13个省份出台了人工智能专项政策，涉及“AI+金融”“AI+教育”“AI+医疗”等多个领域[9]。这些政策不仅加速了AI的应用落地和场景创新，同时也促进监管机制的创新、加强跨部门跨领域协作、提升市场各类主体与公众的参与度，推动AI治理体系精细化发展。

随着人工智能产业发展进入新阶段，未来中国内地政府可能将在以下方面持续完善政策供给，以促进人工智能创新发展与监管规范相协调，推动AI产业高质量发展。

持续细化技术标准，推动基础研究与关键技术突破。工信部发布的《指南》明确提出，到2026年新制定国家标准和行业标准50项以上，预计将会有更多细分标准推出。此外，依托国家自然科学基金、政府引导基金、国家级创新平台等，为基础研究和关键技术攻关提供资金支持、建设绿色通道，推动人工智能方法在科学领域的创新应用。

[9]《地方AI布局突进：13省份出台“AI+”政策，具身智能、智能驾驶或率先落地》，21世纪经济，2025年1月15日

深入挖掘开放场景应用，放大政府基金效应，加速AI产业化发展。2024我国先后发布了关于药品监管、高等教育、新型工业化等三个方面的场景应用清单和案例名单，深入挖掘开放应用场景，推动应用项目落地。同时，政府产投基金将持续发力，带动社会资本加大投入，推动AI应用落地。2025年1月，出资额高达600.6亿元的国家人工智能产业投资基金成立，将聚焦AI算力构建和应用场景开发进行投资，预计将撬动更多社会资本加大投入。

围绕数据标注、国产算力生态等重点领域，优化支撑体系。数据、算力作为AI产业发展的两大支柱产业，政府将围绕财税金融、数据开放共享、人才培养等方面优化支撑体系，推动构建高质量数据集和自主可控的算力生态。

风险防控将进一步收紧，细分领域的监管将持续深化。随着AI应用的进一步落地，政府可能会通过建立“分类分级、动态响应”的精准监管体系，加强对伦理规范、数据隐私安全、内容安全等领域的监管，以防范技术滥用带来的社会风险。

港澳AI治理体系注重国际化和本地特色产业的融合

与内地相比，香港和澳门作为中国的特别行政区，在AI监管及治理环境上既受国家整体战略影响，又因其法律体系、经济结构和社会需求的差异呈现出独特特点。

港澳对AI监管及治理灵活性较高，国际化导向明显。当前，香港、澳门并无专门针对人工智能发展的专项法规，AI监管主要依赖现有法律（如香港的《个人资料（隐私）条例》）的适应性解释，强调行业自律作为补充监管，灵活性较高。此外，香港在AI金融应用中严格遵循巴塞尔协议等国际标准，同时通过与东南亚地区国家签订协议、合作备忘录等，推动数据跨境流动便利化。

优势产业驱动下，AI监管具有本地化、市场化特色。由于经济结构不同，香港、澳门在AI治理方面的融合了行业发展需求和本地特色。香港作为国际金融中心，侧重于金融科技方面的监管与合规，比如，香港金管局推出“生成式人工智能GenA.I.沙盒”，以促进银行业负责任地创新发展生成式人工智能；澳门则聚焦于旅游业、博彩业方面的AI监管和治理。

由于人工智能的快速发展和突破性进展，港澳特区政府可能将围绕个人数据保护、数据安全共享、确保算法公平透明、AI应用监管等方面建立并完善相关监管法规，强化风险防控；同时，依托粤港澳大湾区优势，进一步融入国家战略，加强与内地政策的衔接，推动港澳与内地协同发展。

1.3 企业数智化转型机遇下的AI Ready

企业数智化转型所带来的机遇

机遇一：大模型技术迭代升级使得企业对于AI技术从“看得见却用不起”到“用得起也用得好”转变

在大模型技术快速演进下，基础设施层的计算、网络、存储架构持续优化，尤其在DeepSeek出现以后，其突破传统模型“高成本换高性能”的模式，重塑了大模型的成本效益曲线，打破越强越贵的价格定律。因此，传统企业对AI技术的认知也正在发生重塑 —— 即从“看得见却用不起”到“用得起也用得好”的转变。

据不完全统计，截至2025年2月，全球已有200多家企业接入DeepSeek10，涵盖互联网、金融、汽车等多个行业，并形成了多层次的部署方案。其中大型企业倾向采用私有化部署方案，深度利用DeepSeek的开源特性进行场景定制，本地化提供更强大和可定制化的功能，并保留了数据主权，对于金融、医疗等受严格监管的行业，数据主权和隐私保护是刚需，本地部署能规避第三方服务器泄密风险，但对运维团队的要求也更高。中小企业则主要借助API快速接入模式，用云服务来管控平面，一来简化了运维复杂度，对团队要求更低；二来管理的地域灵活性和整体运营成本都更优化，但同时需要接受SaaS服务方式的管理模式。

机遇二：AI Agent等应用的崛起助力企业的应用场景加速向纵深发展

企业早期的AI应用主要为智能客服、文档处理等通用型应用，未来随着大模型在垂直领域的进一步深耕，企业可以探索更具行业特色的AI应用新范式。近期备受市场关注的AI Agent，作为新型智能交互方式，具备自主决策、多任务协同和持续的学习能力，就可以在AI应用场景发展的早期帮助企业通过树立样本的模式，打造AI速赢案例，释放新质生产力。Gartner预计到2028年，至少15%的日常工作决策将由AI Agent自主完成[11]。

首先，企业聚焦一个特定业务领域开展AI Agent试点，快速赋能业务提质增效，打造出标杆示范项目，为后续更大范围的项目推广提供重要参考。其次，AI Agent驱动的样板工程一旦落地，能大力提振管理层与员工对数字化转型的信心，形成良性循环，推动更多转型项目的成功。例如，某大型零售企业率先在供应链管理方面引入AI Agent，通过精准预测销售趋势和库存需求，优化采购与销售策略，显著提升了销售预测准确率和库存周转率。项目实施后，库存周转率大幅优化、过剩库存减少、工作效率明显提升，员工得以专注于高价值工作，对数字化转型的满意度与积极性也显著提高。

机遇三：提前布局AI生命周期安全防护，释放AI的创新潜力

伴随AI应用的广泛部署，企业从单一模型走向多模型、多云协作的复杂环境。这样的架构转变不仅提升了技术灵活性，也带来了新的安全挑战。传统的安全体系常以固定规则和静态防护为核心，但AI的动态性使得漏洞可能随时发生，无论是在模型层还是应用层。根据思科发布的《2024年人工智能就绪指数》，中国大陆地区仅有29%的受访企业能够在AI应用中全面检测并防止未经授权的篡改。

全球多数国家在AI应用领域的治理仍以非强制性的软性约束为主，尚未建立统一的、具有强制执行力的监管框架和使用标准，这使得企业在AI技术的合规性应用方面面临较大挑战。尤其是当企业从利用公共数据转向自有专有数据的模型训练时，数据泄露、隐私侵害以及知识产权纠纷等问题将变得更加复杂且难以预防。例如某韩国知名半导体企业在引入ChatGPT后，接连发生三起数据泄露事件，致使半导体设备测量资料和产品良率等关键数据被纳入其训练数据库，可能导致敏感数据泄露或滥用。

为释放AI的创新潜力并推动其更广泛的应用，企业亟需一个能够覆盖整个AI生命周期的安全防护层。这一防护层的意义不仅在于应对已知威胁，更在于预防潜在的安全隐患，为用户和应用构建可靠的信任基础。IDC预测到2027年，65%的组织将正式制定政策和监督措施，以应对AI风险，并确保AI治理与战略业务目标相契合[12]。

[10] 众多企业纷纷接入DeepSeek，释放了什么信号？，新浪财经，2025年2月12日
[11] Gartner认为企业必须探索的十大战略科技趋势，澎湃新闻，2024年11月
[12] IDC FutureScape：2025年中国生成式AI市场十大预测，金融界，2025年1月

何为企业的AI Ready

迈进工业4.0时代，AI已成为企业数智化升级的关键驱动力。然而，AI技术的应用并非简单的工具部署，而是需要企业在多个维度上做好充分准备。这种准备状态被称为“AI Ready”（人工智能就绪），即一个企业在引入和应用AI技术前，需在战略、技术基础设施、数据、治理、员工和文化等维度做好全面准备的状态。它不仅要求企业具备先进的技术能力，更要求企业从组织能力、业务流程和文化认知等层面实现全方位的适配与优化。根据毕马威《2024年中国首席执行官展望》，只有10%的受访企业处于数字化转型领先阶段，36%受访企业仅实现了关键业务数字化与系统集成[13]。企业的AI Ready之路仍任重而道远。

此外，由于AI投资巨大，因此在没有明确的行业AI杀手级应用出现之前大部分企业的CxO一方面对直接进行AI基础架构投资持审慎的态度，另一方面又表现出对AI投入的关注度和积极性。在这样一种复杂情绪的裹挟下，了解企业是否已经AI Ready，哪些方面应加强，对于企业能否在AI浪潮中夺得先机至关重要。

企业的AI Ready共包含两大一级能力，即企业的“硬实力”和“软实力”，七大二级能力，即企业AI Ready的七大核心评估维度，其中技术、数据、业务主要对应“硬实力”；战略、治理、人才和组织结构则对应“软实力”（表2）。

表2 企业AI Ready能力的多维度拆解

硬实力	技术（Technology）	企业在AI技术要素（如平台、工具、框架、方法）的选择、使用与创新效果，同时建立AI速赢示范案例，确保AI的可解释性和可信度。
	数据（Data）	企业在AI方面的数据获取、存储、管理、治理等方面的能力，能确保AI数据安全性、准确性、多样性、可访问性和时效性。
	业务（Business）	以客户为中心，从业务前、中、后台全流程实现智能化提升。以创新的产品与服务吸引客户，以智能技术驱动实现个性化营销触达客户，以交叉式体验留住客户，全流程在线连接，数据闭环，以智能化洞察与决策提高决策精准度，业务在线与闭环确保业务敏捷与可控。
软实力	战略（Strategy）	企业为实现AI愿景而制定的任务、政策及措施，与业务战略相契合；能构建起“政产学研用”协同机制；能评估机遇与风险，并监测其对业务的影响及价值。
	治理（Governance）	企业在AI开发与应用中，为实现风险防范、合规运营及伦理责任而建立的规范与机制，涵盖技术风险全生命周期管理、可信AI准则、数字伦理规范及社会责任履行等要素。
	人才（Talent）	企业在AI人才储备与管理方面的策略，包括人才招聘、培养、激励与留用，以及通过内外部资源提升员工技能、弥补人才缺口。
	组织机构 （Organization）	企业在AI应用与创新方面所形成的系统性支持体系，体现为领导及员工价值观的支持、AI的组织规范与职能定义、跨部门协作与知识共享，以及构建AI驱动的生态系统，以实现企业的自适应性和创造性发展。

资料来源：《思科AI就绪指数报告》，MITRE[14]，毕马威分析

[13] 2024年中国首席执行官展望, 毕马威中国, 2024年
[14] The MITRE AI Maturity Model and Organizational Assessment Tool Guide, https://www.mitre.org/news-insights/publication/mitre-ai-maturity-model-and-organizational-assessment-tool-guide

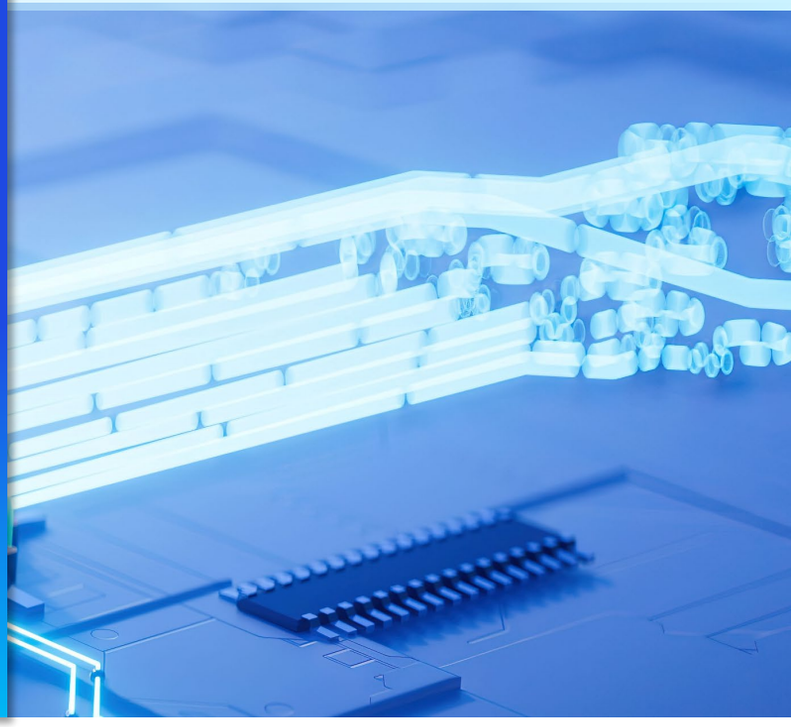
02

企业AI变革之路

在大模型重构产业格局的当下，AI Ready对于企业而言是涵盖以技术和业务为核心的硬实力，以及以组织能力和战略思维为核心的软实力相结合的系统性升级。当前，大企业在AI Ready的多个关键维度上仍存在显著不足，需要把握窗口期完成转型，避免在智能化浪潮中掉队。在此背景下，毕马威与思科携手开展专项调查，向全国范围内已进行AI战略部署并且在业务中有明确AI落地场景，并期望借助大模型技术进一步扩大的泛行业企业发放调研问卷，问题主要涵盖企业AI战略认知与布局、AI变革需求与核心挑战、AI体系变革路径等三大维度。

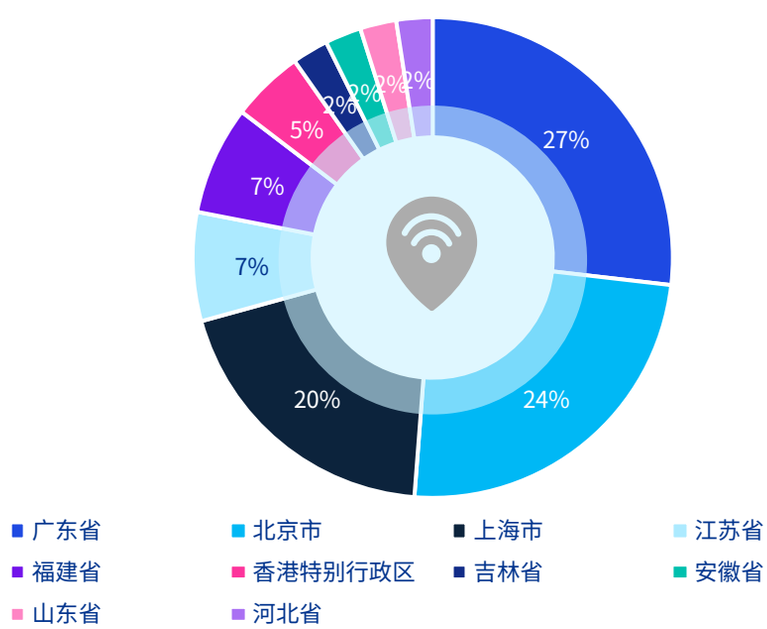
调研历时一个月，共计回收42份有效问卷样本，联合课题组通过对问卷数据的归纳、整理和分析，希望能够为有志在AI浪潮中夺得先机的企业提供有益参考，明晰前进方向和变革路径。

2.1 受访企业基本情况分析



受访企业来自全国范围内十个省市自治区，其中包括香港特别行政区。广东、北京和上海的企业占比最高，分别为27%、24%和20%。从受访企业成立年限来看，10年以上的企业占比88%以上，占据绝对主导。从受访企业所在行业分布来看，排名前三的是制造、金融、电信和互联网，分别占比39%、20%和12%，是AI应用落地场景最丰富的行业代表。

图1 调查企业所在区域分布



*调研数据中的单选题可能由于四舍五入问题加总不等于100%

图2 调查企业成立年限分布

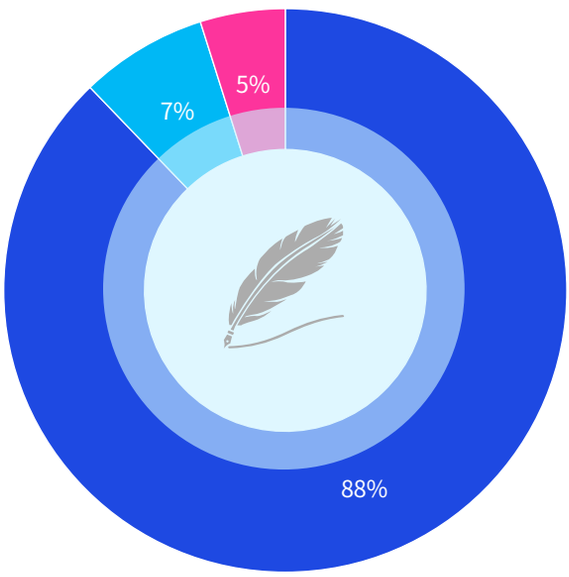
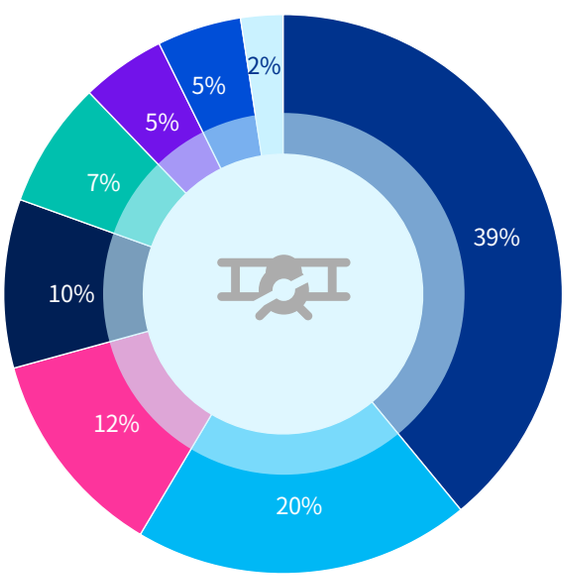


图3 调查企业主要业务领域分布



2.2 企业AI Ready现状及发展需求

战略认知及布局现状

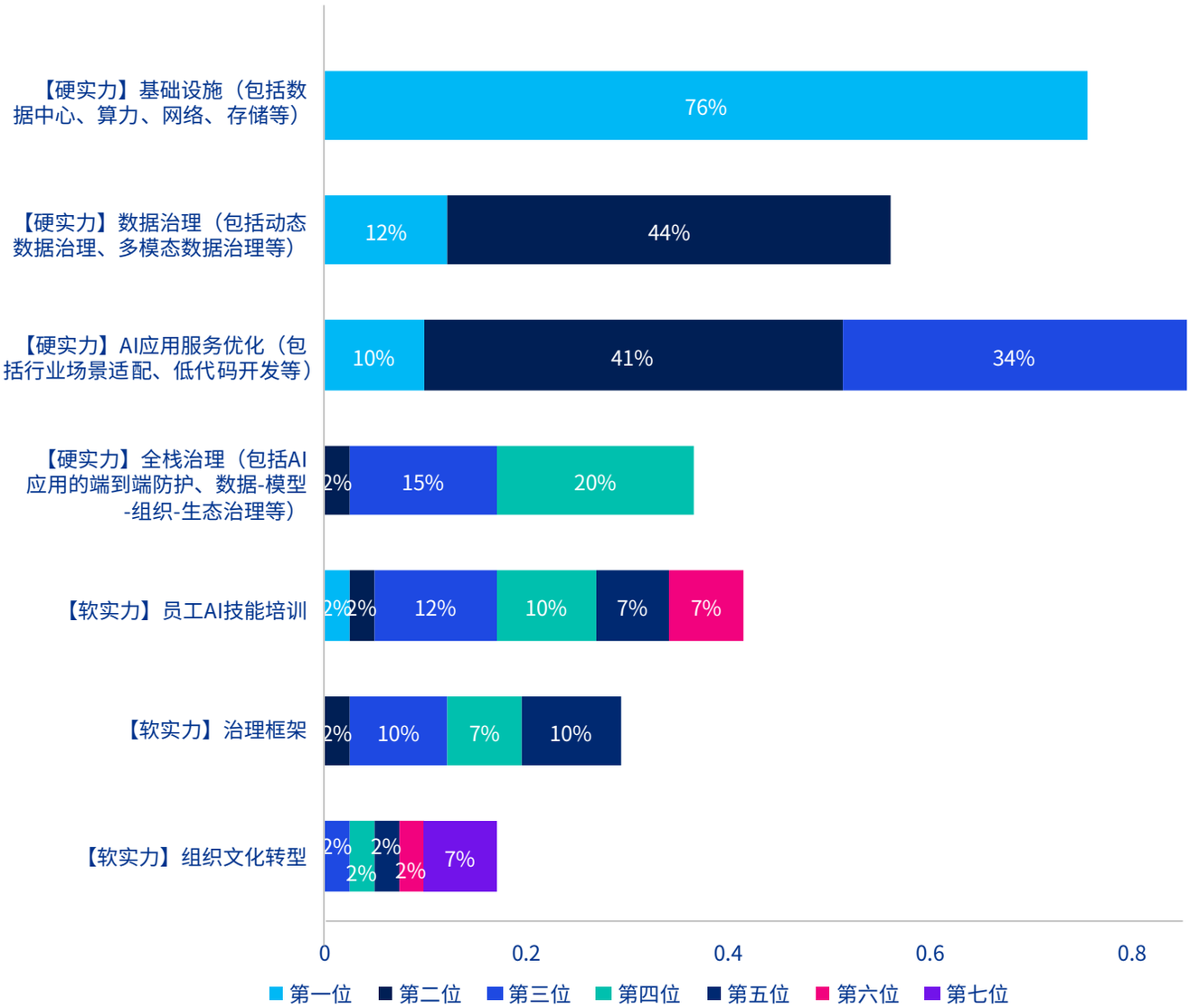
战略认知：

核心为基础设施建设，硬实力优先于软实力

受访企业在AI Ready的战略布局中，将体现硬实力的“基础设施建设”作为第一选项的比例高达76%，由此可见，以算、网、存为代表的基础设施建设是企业AI布局中最为看重的战略核心。硬实力的另外三个选项中，分别有44%和41%的受访企业将“AI数据治理”和“AI应用服务优化”列为第二选项，另外有合计37%的受访企业选择了“AI全栈治理”。

受访企业对于AI软实力布局的重视程度整体弱于硬实力。具体来看，共有40%的受访企业选择了“员工AI技能培训”，其中14%的受访企业将其作为位列前三的首选项。此外，选择“治理框架”和“组织文化转型”的受访企业分别为29%和15%，且排位相对靠后。这表明企业在AI转型中更倾向于先夯实技术基础，再逐步推进企业软实力的适应性调整。

图4 调查企业AI Ready的硬实力与软实力布局优先级排序（多选排序）



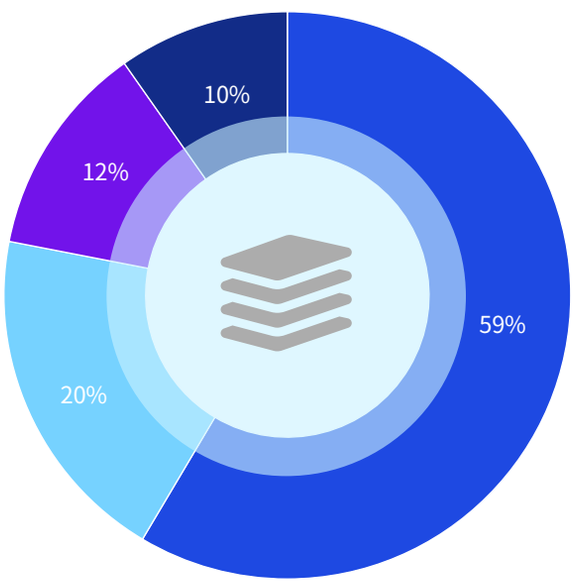
就绪程度：

绝大多数受访企业的AI Ready程度仍需追赶行业平均水平

整体来看，多数企业尚未达到AI Ready的行业领先水平，仍需努力提升自身实力以追赶行业平均水平。具体而言，仅有30%的受访企业认为自身在AI硬实力和软实力的各个维度超出行业平均水平，其中10%的受访企业认为自身处于同行业领先水平。

绝大多数受访企业的AI Ready程度位于行业平均水平以下，占比为70%。其中59%的受访企业认为自身准备基本就绪，但尚未达到行业平均水平，仍需进一步提升；另有12%的受访企业认为准备不充分，存在较多阻碍因素，面临较大挑战。

图 5　对比预期和实际的布局效果来看，调查企业在硬实力和软实力各个维度的准备程度（单选）



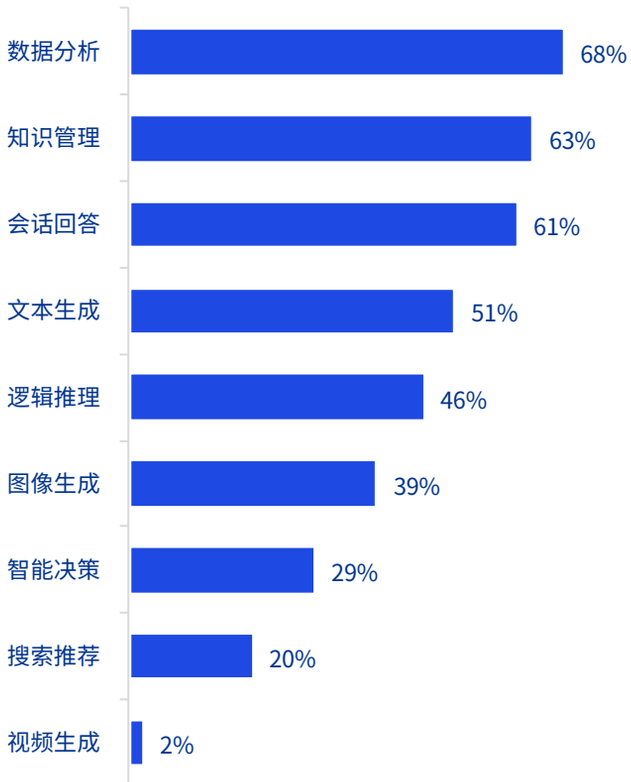
- 准备基本就绪，但仍需追赶同行业领域平均水平（31-60分）
- 准备较充分，且超出同行业领域平均水平（61-85分）
- 准备不充分，阻碍因素过多（30分以下）
- 准备充分，且处于同行业领域领先水平（86分以上）

场景渗透情况：

业务应用层的落地场景多样，数据分析、知识管理、会话回答位列前三

随着技术的不断进步，大模型正在为企业数字化转型和智能化升级提供更加丰富的落地场景，应用的边界被不断拓宽。从各垂直领域来看，大模型在数据分析、知识管理、会话回答等场景中的应用位列前三，分别为68%、63%、61%。其次，文本生成、逻辑推理、图像生成等三类应用的占比分别为51%、46%、39%，这些领域涉及感知、认知和创造能力，能帮助企业探索新的商业模式和应用场景。最后，智能决策、搜索推荐、视频生成等应用的占比均在30%以下，这几类由于对于底层模型推理和多模态处理的要求更高，目前在企业端的应用仍有待进一步的普及。

图6 调查企业已投产的大模型应用主要集中在以下哪些垂直领域（多选）



能力突破情况：

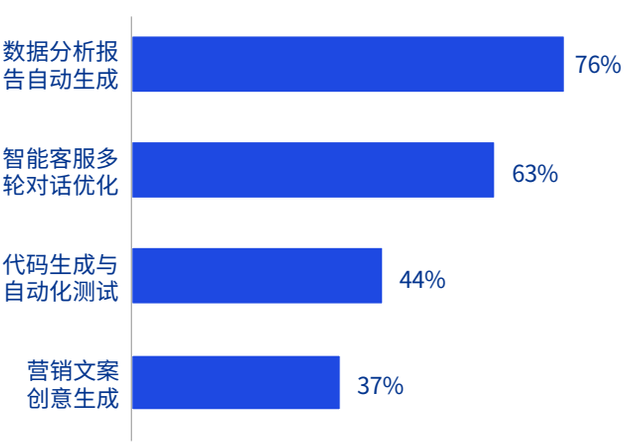
大模型技术在企业场景端的核心突破体现在自动化能力增强和场景适配性的提升

有76%的受访企业将“数据分析报告自动生成”作为最能体现大模型能力的应用，其次是“智能客户多轮对话优化”和“代码生成与自动化测试”，各占比63%和44%， 分列第二和第三位。“营销文案创意生成”占比37%，位列第四。

其中自动生成数据分析报告、生成代码和营销文案生成这三大场景的突破集中体现出大模型技术自动化能力的增强。例如自动生成报告主要源于大模型技术能够自动解析海量数据，识别深层规律，并通过自然语言生成结构化报告。在此基础上，模型还能通过持续训练优化来自动修正报告中的偏差，提升数据可信度。

多轮对话优化则体现出大模型对场景适配性的提升。大模型通过长窗口记忆技术，可以实现多轮对话的上下文追踪。在此基础上，大模型还能在智能客服场景中结合情感分析模型识别用户情绪并调整回复策略，从而提升客户对话的满意度。

图7 调查企业引入AI大模型后，在哪些场景实现能力突破（多选）



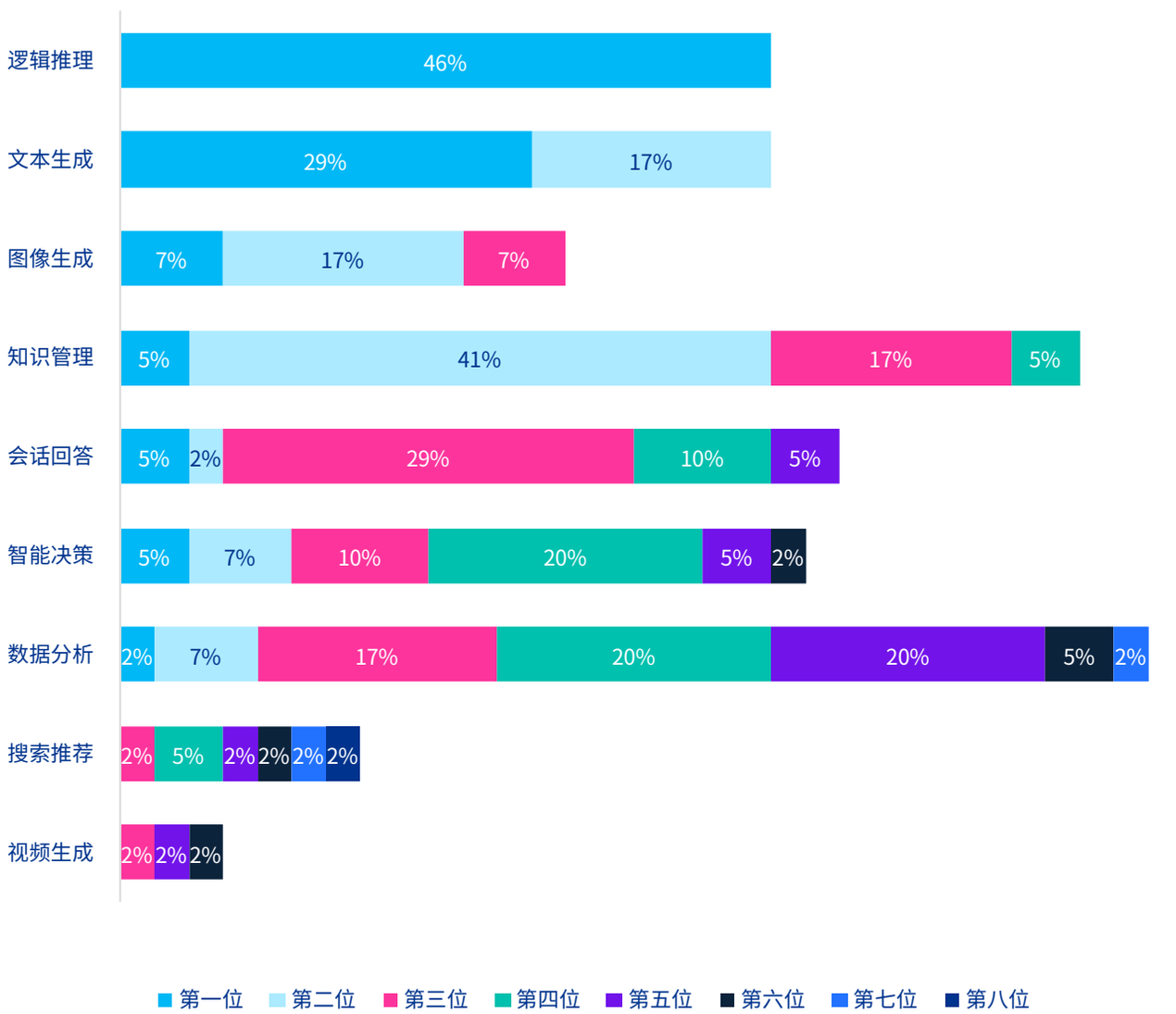
价值收益情况：

逻辑推理、文本生成、图像生成和知识管理等大模型应用对企业的价值贡献度最大

在大模型应用的多个场景中，46%的受访企业认为“逻辑推理”的价值贡献度最大，将其作为了首选项，这也是大模型“智能涌现”出的核心能力之一。另有46%的受访企业将“文本生成”作为了第一和第二选项，其中首选的比例为29%。31%的受访企业在前三位中选择了“图像生成”，其中将其列为首选的比例是7%，显示出企业对于大模型文本和图像等多模态能力所带来价值的重视。此外，分别有5%的受访企业将知识管理、会话问答和智能决策列为了第一选项，另有41%的受访者将知识管理列为了第二选项，在一定程度上反映企业对知识库管理的需求和重视。

除此以外，共有73%的受访企业选择了“数据分析”，但选项较为靠后。数据分析自大数据时代以来在企业端已得到广泛应用，但不属于大模型创新能力的体现。

图 8 **按价值贡献度为以下大模型应用场景排序（多选排序）**



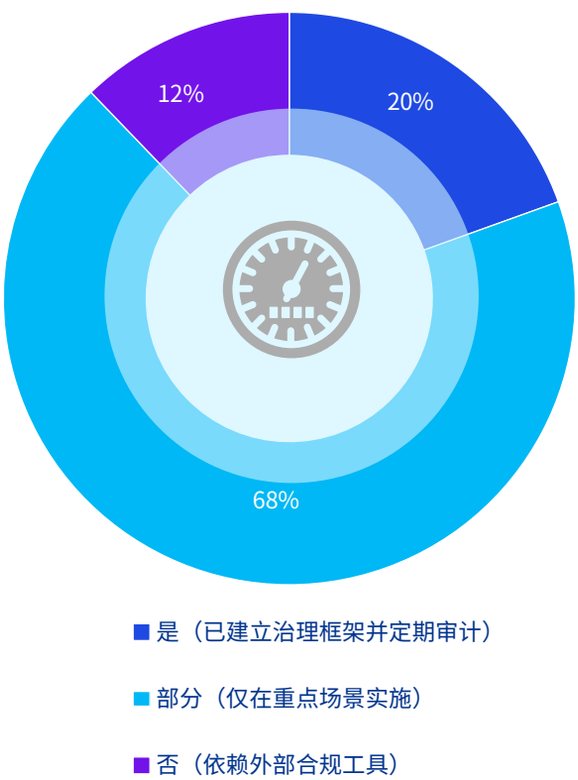
安全治理情况：

AI全栈治理概念普及度较高，目前主要在企业的重点场景中采用

AI全栈治理的本质是“技术-业务-安全”三位一体的动态平衡，有68%的受访企业优先选择在部分重点场景中采用AI全栈治理，而非全面铺开。一方面源于企业在资源投入与业务价值间的平衡，更倾向于选择能快速验证价值的核心场景进行展开，且AI治理依赖高质量数据，而核心场景往往数据成熟度更高。另一方面可能考虑到技术复杂性和实施门槛。AI全栈治理在技术上考验企业在基础设施、智能运维和完全防护等方面的全链条能力，这种复杂性使得企业需要分阶段进行推进。

此外，有20%的受访企业选择是“（已建立治理框架并定期审计）”，而仅有12%受访企业选择对AI全栈治理选择了“否（依赖外部合规工具）”，反映出AI全栈治理概念在企业中的普及度较高，未来随着企业AI部署的进一步深入，将得到更广泛的应用。

图9 调查企业是否将AI全栈治理纳入战略优先级（单选）



需求分析与核心挑战

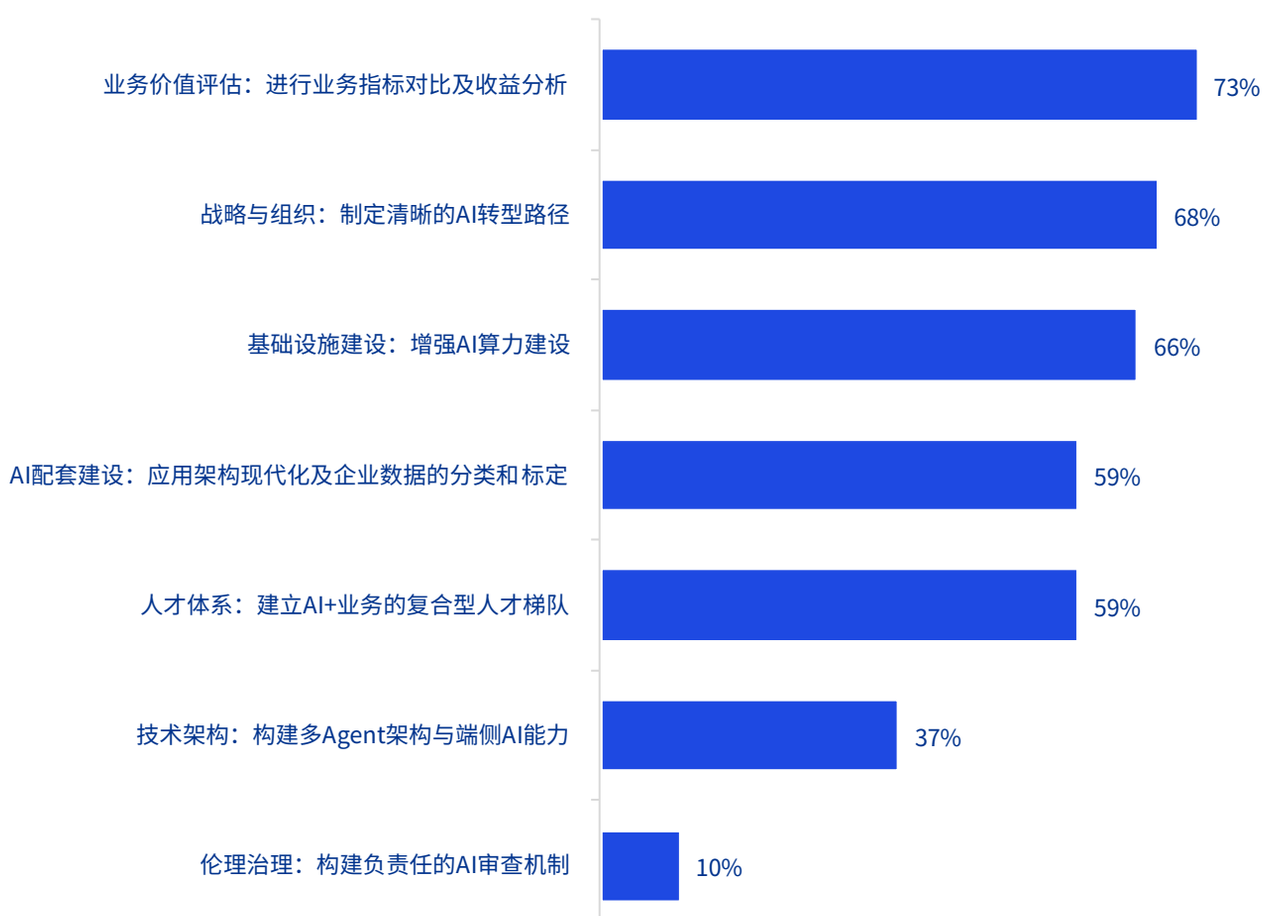
变革需求：

业务价值评估、清晰的AI转型路径及基础设施建设为推动企业AI变革的三大核心要素

受访企业认为，推动企业AI变革的前三大核心需求分别为业务价值评估、清晰的AI转型路径以及AI算力建设。其中，73%的受访企业将“业务价值评估”，即AI投入的ROI 列为核心需求，排名第一。排名第二和第三的需求分别为“战略与组织：制定清晰的AI转型路径”（68%）和“基础设施建设：增强AI算力建设”（66%）。

此外，分别有59%的受访企业选择“AI配套建设：应用架构现代化及企业数据的分类和标定”和“人才体系：建立AI+业务的复合型人才梯队”，并列第四，表明完善的企业AI应用架构、数据治理体系和人才储备是企业实现AI变革的重要支撑。另有37%和10%的受访企业选择 “技术架构：构建多Agent架构与端侧AI能力”和 “伦理治理：构建负责任的AI审查机制”，占比相对较低。

图 10 调查企业认为推动企业AI变革的核心需求为（多选）



价值需求：

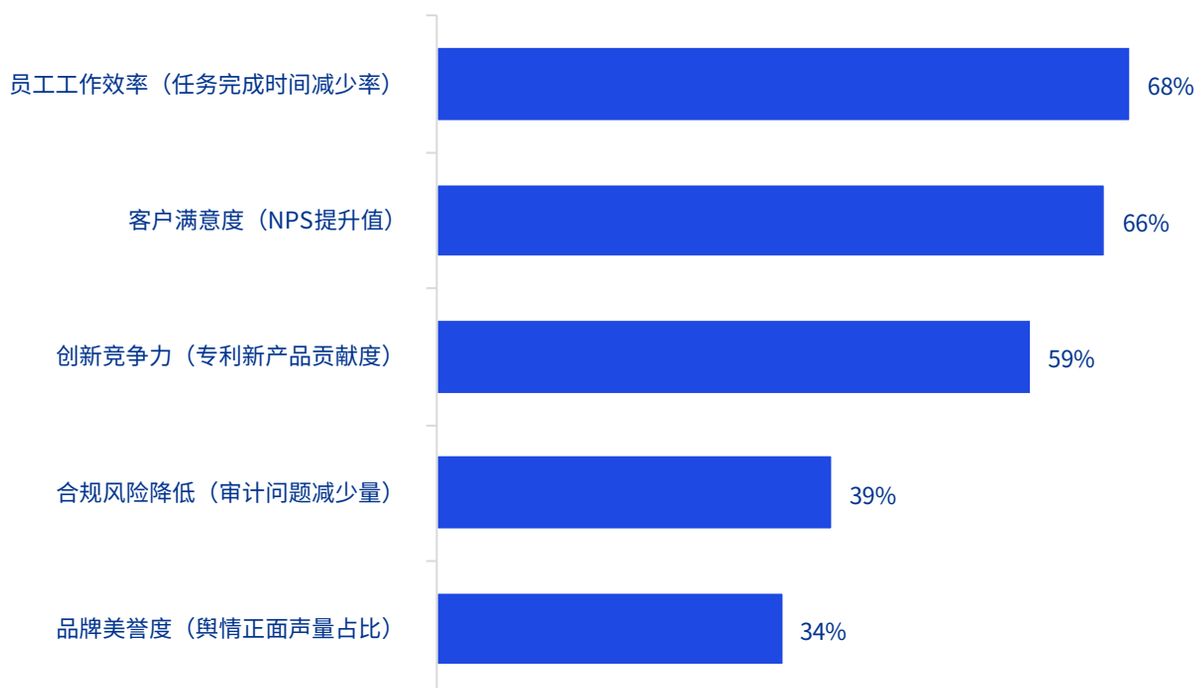
企业最看重员工工作效率、客户满意度和产品创新竞争力等大模型场景价值的隐形收益指标

隐性收益是指科技成果在企业运作中所带来的非直接可见的好处，比如提高客户满意度、员工工作效率等，通过隐性收益的积累，企业可以在市场中获得更强的竞争优势。

在大模型场景所带来的隐形收益中，“客户满意度”“员工工作效率”和“创新竞争力”这三个指标最被受访企业重视，分别占比68%、66%和59%。主要是因为这三个隐形收益指标最贴近业务，其为企业业务端所带来的收益也最易于通过量化手段来进行收益分析，例如创新竞争力这一指标可以通过企业部署大模型后产品专利申请的增加来进行量化。

相比之下，“合规风险降低”与“品牌美誉度”等两个指标由于难以量化和与企业业务关联度相对较低，占比仅为39%和34%。

图11 调查企业量化大模型场景价值时，主要纳入哪些隐性收益指标（多选）



变革挑战：

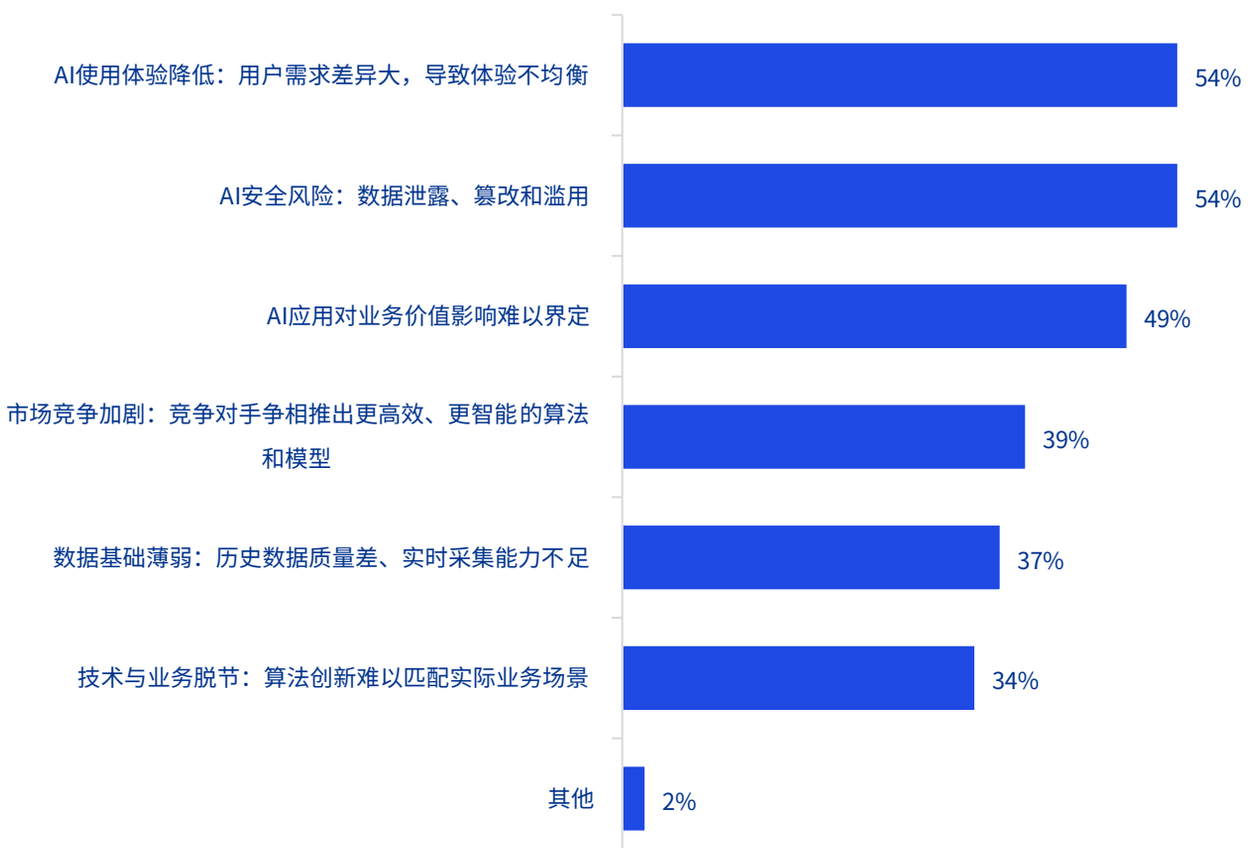
AI体验优化、AI安全风险治理和AI价值显化是企业AI变革亟需解决的三大核心问题

企业在AI变革中面临诸多挑战，其核心障碍主要集中在用户体验、安全性和业务价值的界定上。其中，各有54%的受访企业将由于个体需求差异大所导致的“AI使用体验降低”和由于数据风险所导致的“AI安全风险”列为首要挑战。另有近半数（49%）的受访企业将“AI应用对业务价值影响难以界定”列为第二大挑战，表明AI应用是否能为企业带来业务价值是影响企业AI投入的重要决定因素之一。

此外，有39%的受访企业选择“市场竞争加剧”，表明随着智能算法和模型的不断迭代，AI市场的竞争已经逐渐由蓝海进入红海状态。另有37%和34%的企业将“数据基础薄弱”和“技术与业务脱节”选择为核心挑战。

总体来看，企业AI变革面临多维度挑战，既有技术层面的瓶颈，也有市场和安全方面的压力，需在加强数据治理、提升技术与业务融合能力、优化用户体验、强化安全防护等方面多措并举，从而在AI变革浪潮中赢得市场竞争优势。

图12 调查企业认为目前企业AI变革所面临的挑战（多选）

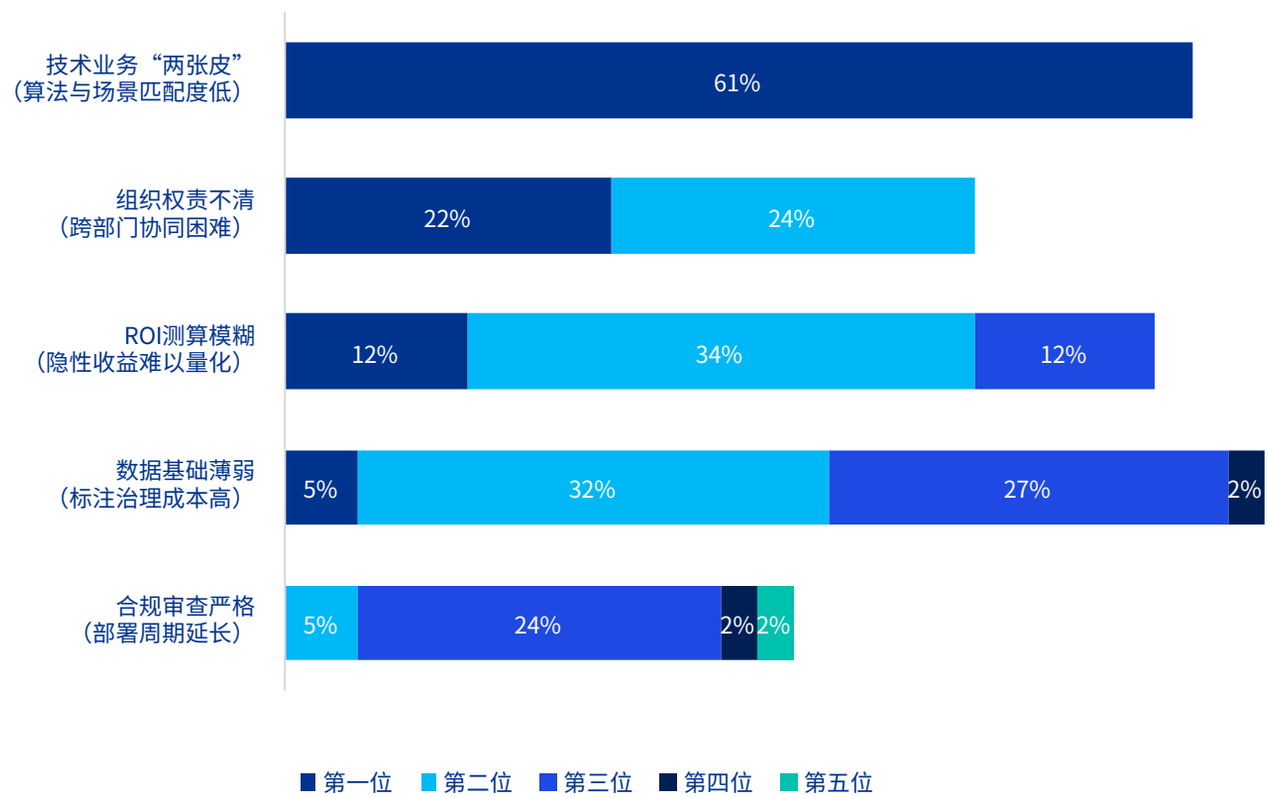


价值挑战：

企业AI落地应用面临着与场景的匹配度低、跨部门协同困难和隐形收益难以量化等多重挑战

从研发到落地，企业也会遇到不少的挑战。受访企业中有61%将技术业务“两张皮”，即技术和业务场景匹配度低作为首要挑战。46%的受访企业将跨部门协同困难选为前两大挑战，其中将其作为首要挑战的比例为22%，表明技术驱动组织机制改革的重要性。58%的受访企业选择“ROI测算模糊”，但将其作为首选挑战的比例仅12%，表明如何将大模型落地价值量化仍是企业的核心困扰之一。此外，66%的受访企业将“数据基础薄弱”作为主要挑战，将其作为第二和第三大挑战的比例为32%和27%, 表明数据质量仍是大模型质量和落地效果的核心制约。

图13 从实验室到规模化推广，调查企业遇到的前三大挑战（多选排序）



2.3 企业的人工智能体系变革路径

技术架构侧的统筹规划

场景体系设计：

实现大模型价值闭环的起点和终点

目前大模型的应用整体处于“初步的思维链可控阶段”，由于应用初期探索的局限性导致应用场景呈现同质化和单点式等特征，例如多数企业优先对标成熟产品，导致应用主要集中于聊天、内容生成等场景浅层化领域。随着未来应用场景的进一步深入，需要在场景侧进行体系化的设计，一是可以将分散的场景需求与企业的战略目标深度融合，二是通过统一数据标准、接口规范和数据底座等，打破单点式场景导致的重复建设和“信息孤岛”,实现算力、算法、数据等的共享复用。企业在进行场景体系设计时不用局限在已知的模型能力、已就绪的数据中，要从企业业务发展战略、AI 技术核心原理、行业发展趋势的角度构思和规划，以始为终来看待场景体系设计，从而实现价值闭环。

体系化的场景设计和企业架构部署，是企业在AI规模化应用中实现技术价值向业务价值转化的核心路径，既能规避碎片化投入的风险，又能为未来的业务创新预留扩展空间。

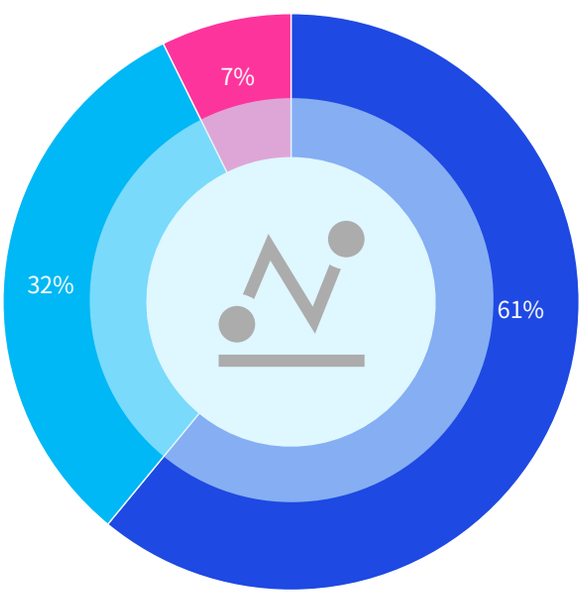
模型部署方式：

混合部署由于兼顾灵活和稳定性成为企业AI落地的主流选择

在AI落地路径的选择中，“混合部署，本地+公有云”以61%的占比成为最受企业青睐的部署策略，远高于“自建，完全本地部署”的32%和“公有云部署”的7%。这表明企业在AI部署过程中更倾向于企业在保留原有本地数据中心资源的同时，又能够借助公有云平台来实现资源的弹性扩展，以实现灵活性与稳定性的平衡。相比之下，完全本地部署由于较高的建设和运维成本，限制了其普及程度；而单一的公有云部署则因难以满足企业对安全防护、数据存储备份等方面能力需求而排名靠后。

在AI落地路径的选择中，“混合部署，本地+公有云”以61%的占比成为最受企业青睐的部署策略，远高于“自建，完全本地部署”的32%和“公有云部署”的7%。这表明企业在AI部署过程中更倾向于企业在保留原有本地数据中心资源的同时，又能够借助公有云平台来实现资源的弹性扩展，以实现灵活性与稳定性的平衡。相比之下，完全本地部署由于较高的建设和运维成本，限制了其普及程度；而单一的公有云部署则因难以满足企业对安全防护、数据存储备份等方面能力需求而排名靠后。

图 14 在AI落地路径选择中，调查企业的部署策略倾向（单选）



云端风险应对：

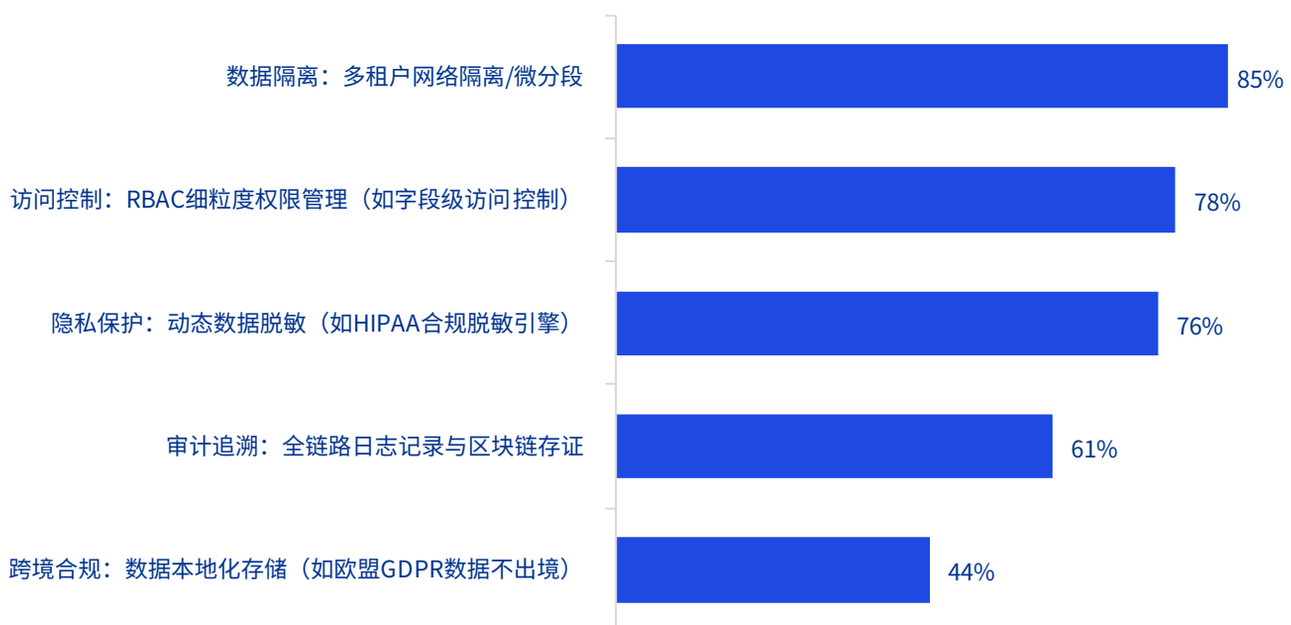
企业需构建多层次、全方位的安全防护体系以应对云部署所带来的数据风险

企业在享受上云所带来的便利性的同时，也需要积极构建多层次、全方位的数据安全防护体系，包括数据隔离、隐私保护、访问控制、审计追溯及跨境合规等多个维度。

调查数据显示， 85%的受访企业在云部署中最为关注数据的安全隔离与独立管理，选择采取“数据隔离：多租户网络隔离/微分段”措施，以实现不同用户数据之间的物理或逻辑隔离。其次，“访问控制：RBAC细粒度权限管理（如字段级访问控制）”和“隐私保护：动态数据脱敏（如HIPAA合规脱敏引擎）”分别以78%和76%的占比紧随其后，显示企业希望通过严格的权限管理和数据隐私保护，提高数据安全性。

此外，“审计追溯：全链路日志记录与区块链存证”以61%的占比位列第四，强化了数据操作的透明性和可追溯性。最后，“跨境合规: 数据本地化存储（如欧盟GDPR数据不出境）”以44%的占比位列第五，反映企业普遍采取数据本地化存储策略以应对数据跨境流动的合规性要求。

图 15　在混合云/公有云部署中，调查企业对于保障数据监管合规的主要措施（多选）

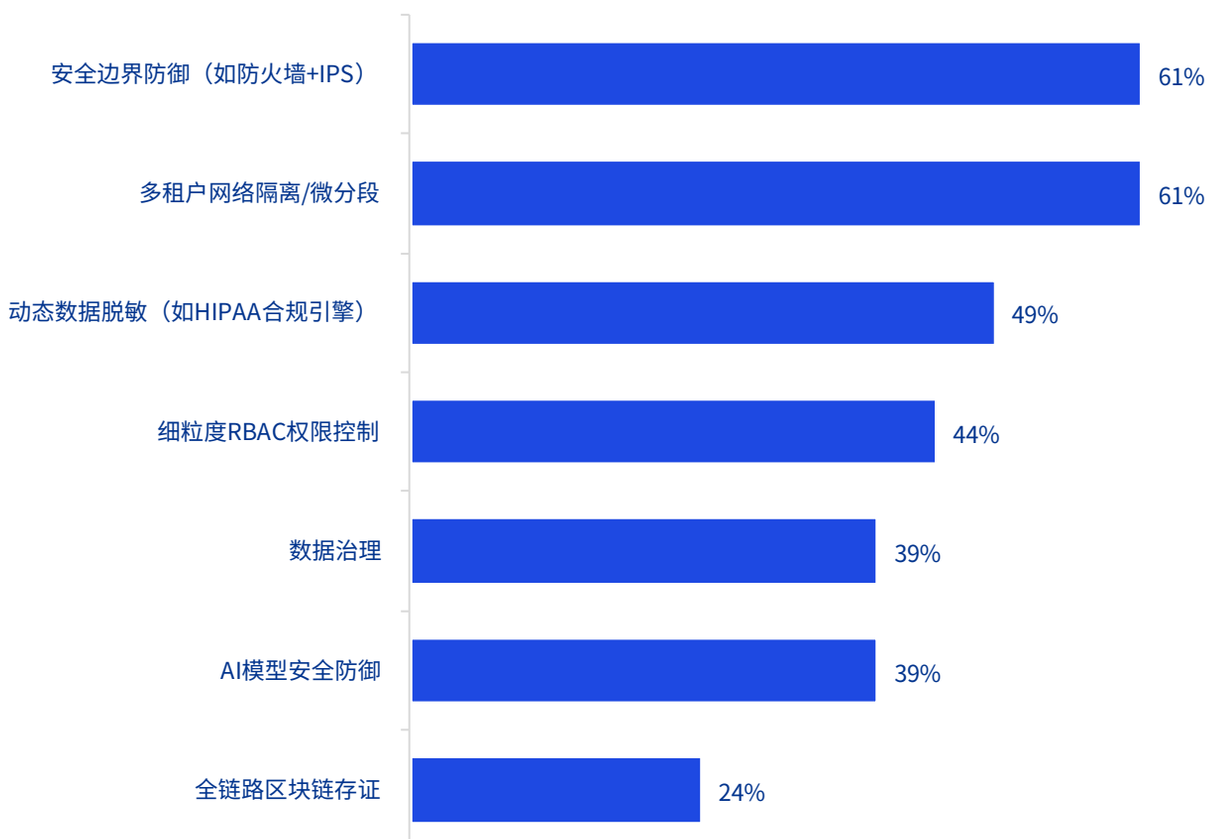


私域安全防护：

安全边界防御和微分段是企业大模型私域部署的关键防护措施

大模型能力的快速演进也随之带来潜在的安全风险，企业在进行大模型私域部署时需要根据具体的业务需求和技术栈灵活组合，持续评估安全策略的有效性。调研结果显示，受访企业首选“安全边界防御（如防火墙+IPS）”与“多租户网络隔离/微分段”两项措施，占比均为61%，并列第一。安全边界防御体系中的防火墙负责过滤已知的恶意流量，而IPS则用于检测和响应更复杂的攻击，两者结合用于边界防护；而微分段将网络划分为细粒度安全域，强化内部隔离，使攻击面最小化。选择“动态数据脱敏（如HIPAA合规引擎）”与“细粒度RBAC权限控制”防护措施的受访企业分别为49%和44%，分列第二和第三。两项措施通过实时的隐私保护和最小权限原则形成纵深防御体系。“数据治理”与“AI模型安全防御”两项措施的占比均为39%，并列第四。“全链路区块链存证”技术的占比最小，为24%。

图16 调查企业在AI大模型私域部署中采取的关键防护措施（多选）

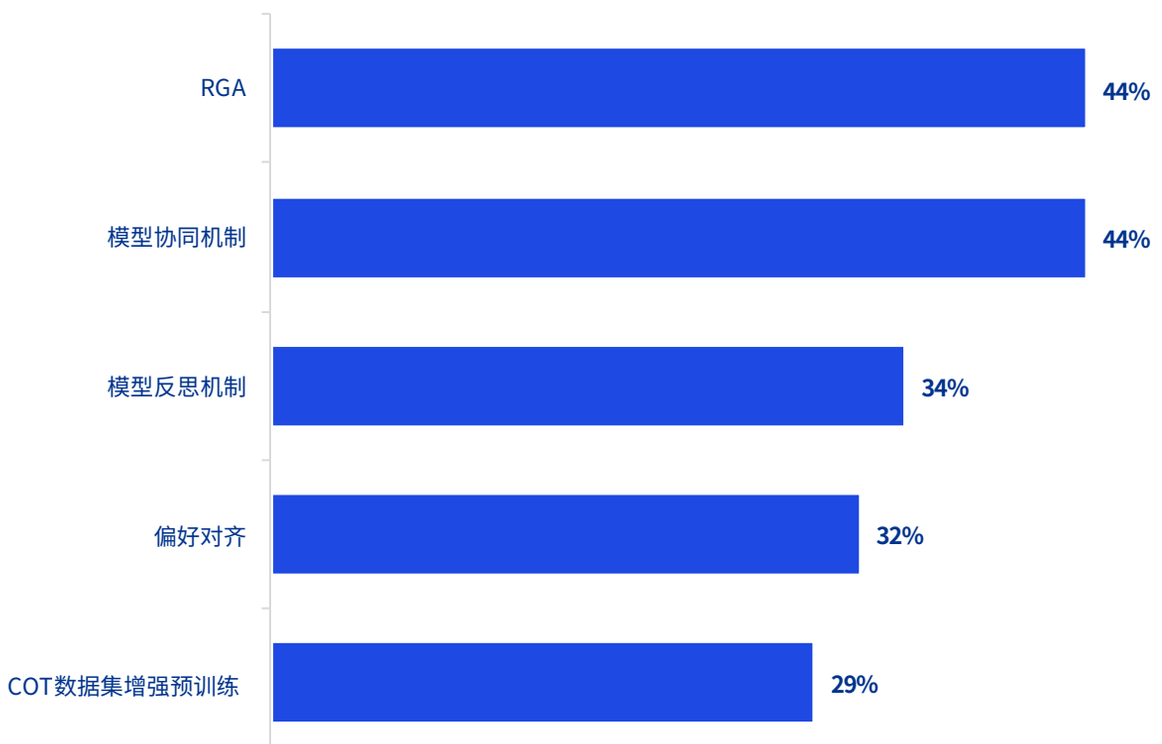


模型幻觉应对：

大模型幻觉的多元化应对策略中RAG和模型协同机制为首选

大模型幻觉主要指的是大模型在生成内容时偏离事实或逻辑的现象，这种现象可能在企业的应用端带来严重的后果。其中“RAG”（检索增强生成）和“模型协同机制”是受访企业优先选择的解决方案，两者均以44%的比例并列首位。此外，“模型反思机制”和“偏好对齐”也受到较高重视，分别占比34%和32%，前者通过自我验证减少错误输出，后者则通过调整模型行为使其更符合用户预期。此外，“COT数据集增强预训练”以29%的占比位列第五，表明部分企业开始探索大规模数据增强，以提升模型的逻辑推理能力。

图 17　在解决大模型幻觉时，调查企业优先采用的策略（多选）



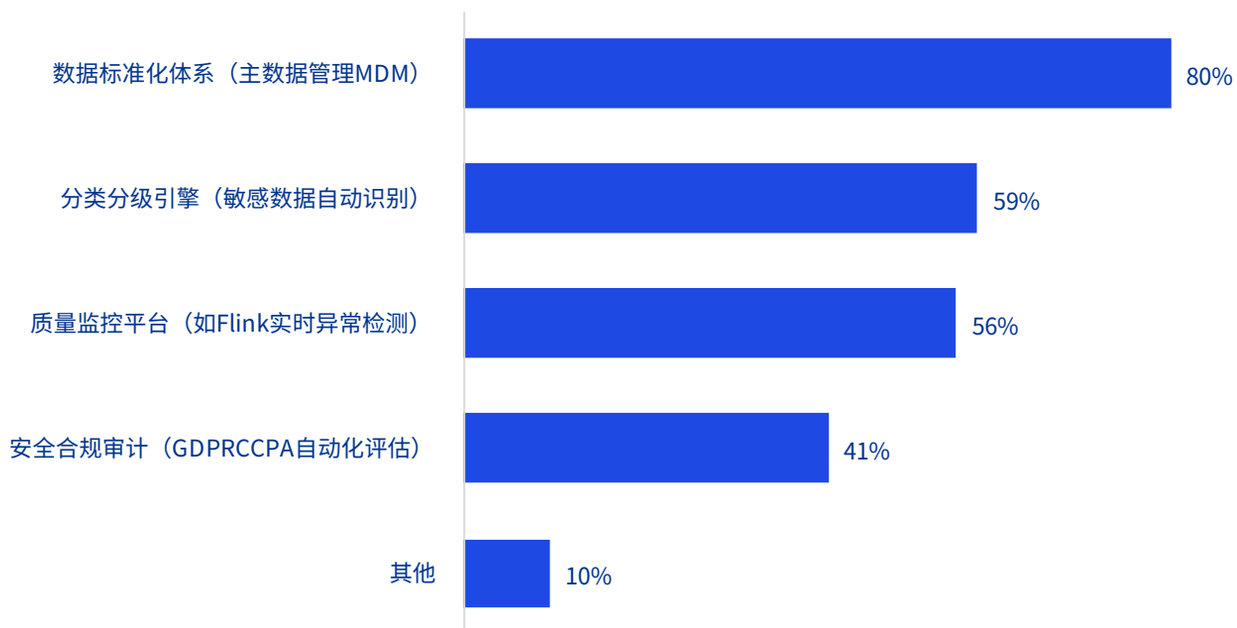
数据语料侧的深度治理

数据治理框架：

数据标准化体系是大模型数据治理框架的核心模块

在企业的大模型数据治理框架中，80%的企业将“数据标准化体系”列为其中最为核心的基础模块。该体系的核心为建立企业的主数据管理（MDM）体系，即使用技术、工具和流程来创建统一的主数据服务，以整合关键的企业数据资产，例如客户信息、产品详细信息和位置数据等。其次，敏感数据自动识别为代表的“分类分级引擎”和以Flink实时异常检测为主的“质量监控平台”分别以59%和56%的占比位列第二和第三，主要是指企业通过实时检测技术和自动识别技术进行动态监控，确保数据在存储和使用过程中的安全性。此外，41%的企业还将以GDPRCCPA自动化评估为代表的“安全合规审计”纳入框架，表明企业正逐步加强对数据使用的合规性管理，以应对全球化的监管要求。

图 18 调查企业的大模型数据治理框架是否包含以下模块（多选）

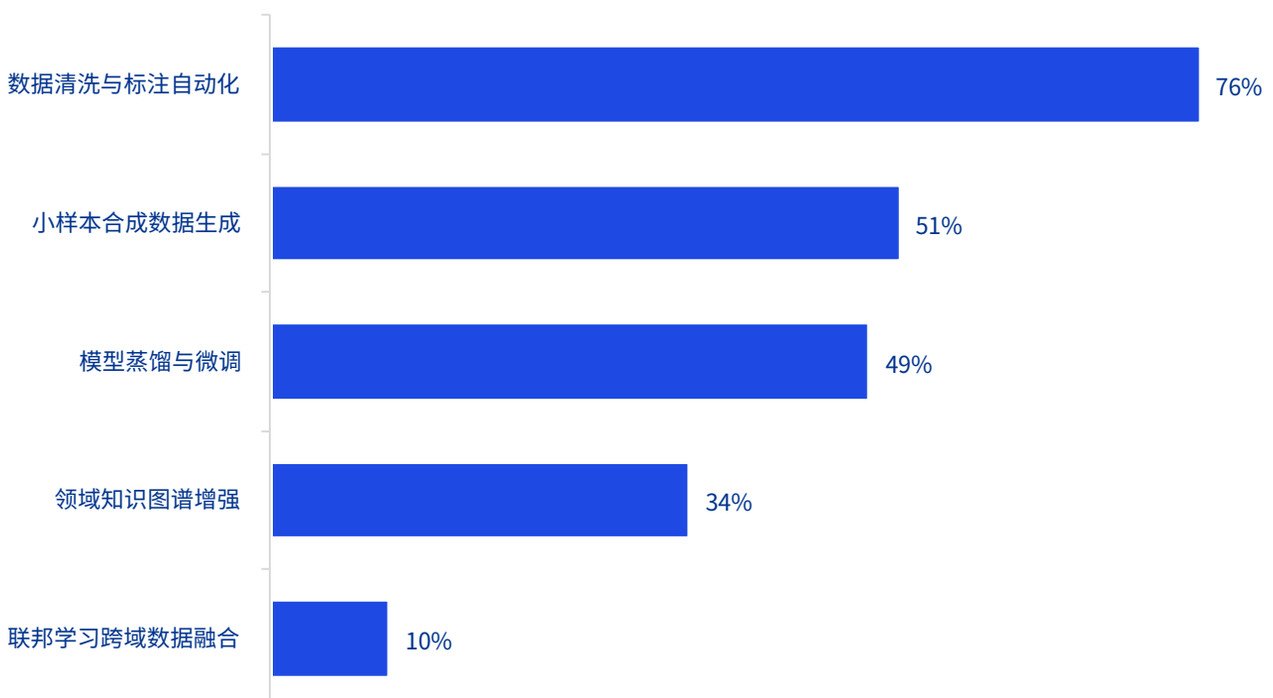


数据质量提升：
智能清洗工具成为企业提升垂直领域数据质量最常用的方法

为系统解决垂直领域数据质量的问题，受访企业主要通过自动化处理、合成数据生成和模型优化等常用手段，支持AI模型的高效训练与应用。调查数据显示，“数据清洗与标注自动化”以76%的占比成为最常用的方法，表明企业普遍采用如DeepSeek-R1等智能清洗工具以提升数据的准确性和可用性。

其次，“小样本合成数据生成”以51%的占比位列第二，反映出企业倾向于利用Diffusion等模型进行数据增强，生成高质量的合成数据，以补充真实数据的不足。同时，49%的企业选择“模型蒸馏与微调”，通过将大型复杂模型的知识迁移到轻量化模型，如BERT-base蒸馏至TinyBERT，进一步提升模型在垂直领域的适应性。此外，34%的企业选择“领域知识图谱增强”，例如医疗领域的术语图谱构建以补全数据在特定领域的知识图谱。相比之下，仅10%的企业采用“联邦学习跨域数据融合”方法。

图19 针对垂直领域数据质量不足问题，调查企业主要采用的方法（多选）



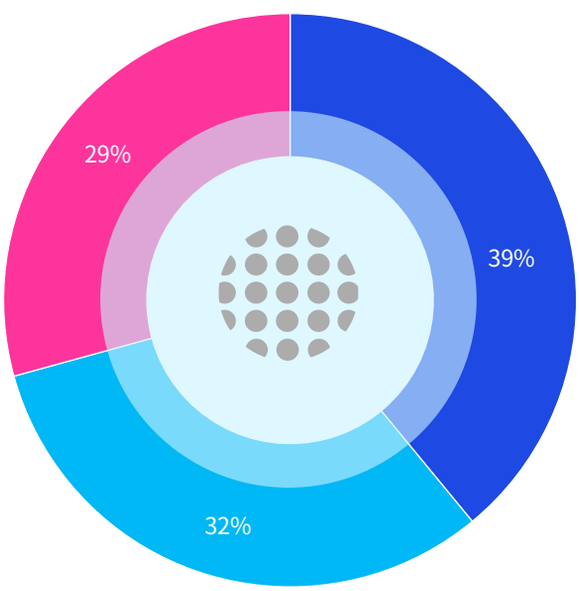
基础设施侧的按需配置

基础设施建设：

企业AI基础设施建设首选以核心业务投入为主的混合模式，兼顾务实与创新

企业的AI基础设施建设通常涉及到较高资金、人力和时间等投入，需权衡自身当前业务需求与未来发展战略后谨慎开展。就投入模式而言，通常可分为前瞻布局、渐进升级和以核心业务投入为主的混合模式等三种。调查数据显示，39%的受访企业为实现资源的灵活配置和高效利用，更倾向于选择“核心业务先导投入+长尾场景按需扩展”的混合模式。其次有32%的受访企业选择“渐进升级：基于现有数据中心扩展AI算力”，位列第二。这种模式的优点为成本可控且易于实施。最后，29%的受访企业更倾向于选择“前瞻布局：3-5年技术预研”，表明部分已有技术储备的企业更加看重AI基础设施建设的长期规划和战略布局。

图20 调查企业AI基础设施建设倾向采用的模式（单选）



- 混合模式：核心业务先导投入+长尾场景按需扩展
- 渐进升级：基于现有数据中心扩展AI算力（如GPU集群分批扩容）
- 前瞻布局：3-5年技术预研（如万兆光网+智算中心一体化）

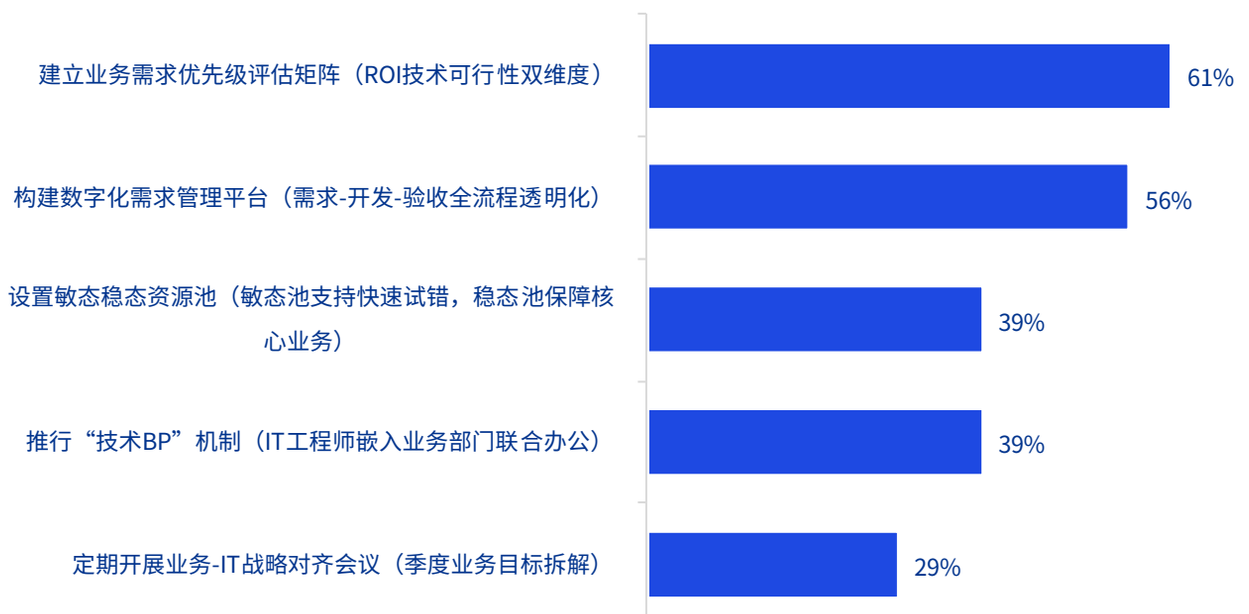
业务需求协同：

优先级评估、数字化平台、跨部门协作、资源池管理等多措并举强化AI基础设施与业务需求的协同

为确保AI基础设施与业务需求的有效衔接，企业普遍采取需求优先级评估、数字化平台、跨部门协作、资源池管理等多种措施以实现业务与技术的协同。

调查数据显示，61%的企业首选采取“建立业务需求优先级评估矩阵”措施，通过ROI和技术可行性等两个维度进行科学评估，明确业务需求的优先级，从而确保技术投入与业务价值产出的匹配。其次，56%的企业倾向于通过“构建数字化需求管理平台”，通过将需求-开发-验收全流程透明化来提升需求与开发的对接效率。此外，“设置敏态稳态资源池”，其中敏态池支持快速试错而稳态池保障核心业务和“推行‘技术BP’机制”，即IT工程师嵌入业务部门联合办公等两种方式的以占比并列第三，均为39%，表明部分企业通过灵活的资源分配和跨部门协作，进一步强化技术与业务的协同。相比之下，“定期开展业务-IT战略对齐会议”占比最低，为29%。

图21 为避免基础设施与业务脱节，调查企业会优先采取的措施（多选）



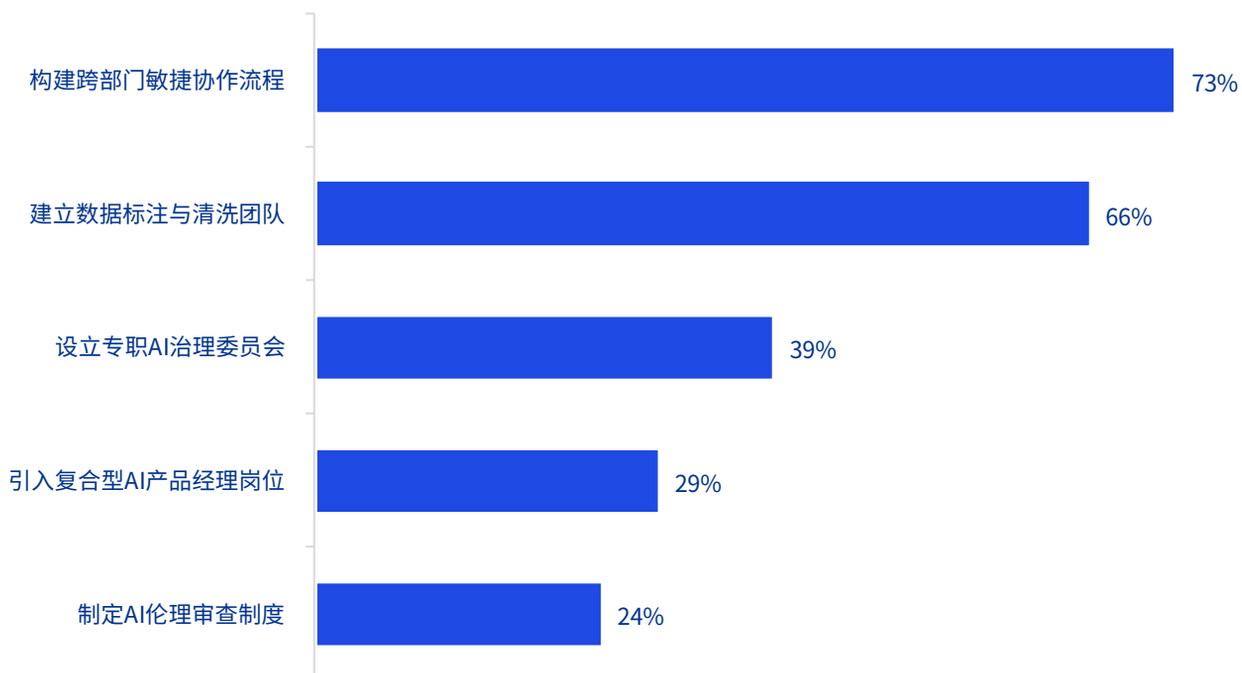
组织体系侧的机制保障

组织机制特点：

敏捷性和协同性是大模型应用对组织机制的核心要求

73%的受访企业认为“构建部门敏捷协作流程”是大模型应用对组织机制最核心的要求，表明大模型的高效应用对组织的敏捷性和协作性有较高要求，这就需要企业打破传统层级结构，转向扁平化或矩阵式组织架构，以提升跨部门协作效率并对需求敏捷响应。受访企业的第二选择是“建立数据标注与清洗团队”，占比66%，表明高质量的数据供给对于解决大模型幻觉和增强应用效能至关重要；“设立专职AI治理委员会”排名第三，占比39%；最后“引入复合型AI产品经理岗位”及“制定AI伦理审查制度”的占比均在30%以下，分别为29%和24%。目前大模型的行业应用仍处于发展早期，成熟的AI产品仍有待时日，可能是企业对于复合型AI产品经理类人才需求较低的原因之一。

图22 大模型应用对调查企业组织机制的关键要求（多选）

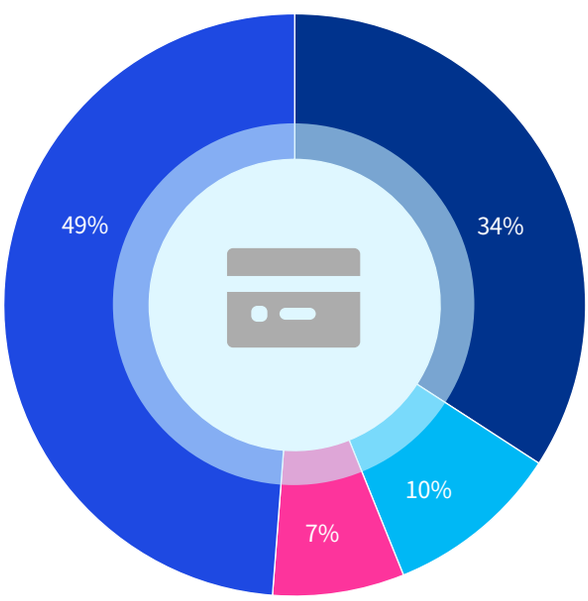


能力建设模式：

以关键能力自研+通用模块外包为核心的混合模式为企业AI能力建设主流

从企业AI能力建设的几种主要模式来看，以关键能力自研+通用模块外包为核心的 “混合模式”为受访企业首选，占比49%，该模式由于具有核心算法自研保障技术壁垒，通用模块外包降低人力成本的优势，可以达到成本与效率平衡，最受企业青睐。受访企业的次优选择为“自建全栈团队”,占比34%，该模式具有技术自主可控且长期竞争力强等优点，但由于成本较高，更适合大型企业采用。另外两种模式为聚焦业务场景落地的“外包核心开发” 和“采购SaaS化服务”，分别占比10%和7% ，这两种模式主要依赖供应商，由于可能具有黑箱风险和数据隐患等，导致企业的采用率较低。

图23 调查企业当前AI能力建设模式（单选）



员工风险应对：

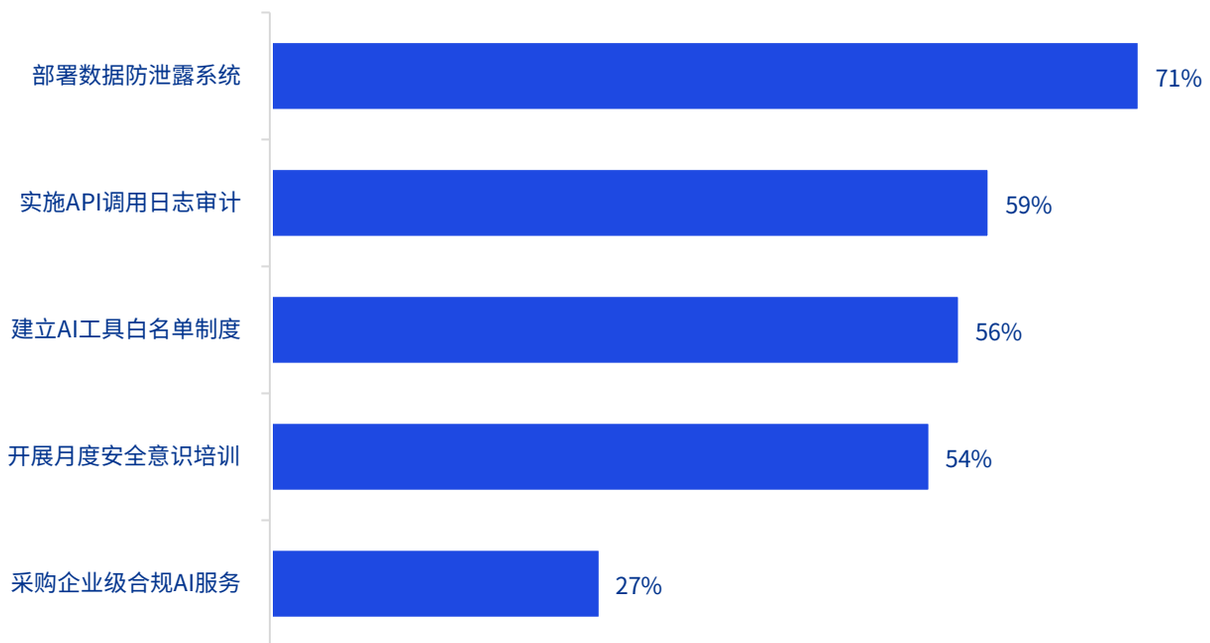
企业多采用部署数据防泄漏系统和实施API调用日志审计等措施来应对员工使用AI工具的风险

随着技术的进步，AI将越来越深度地与企业业务相结合，成为员工日常工作的核心工具。AI工具像一把双刃剑，在提升员工工作质效的同时也会为企业带来安全风险和隐患。

71%的受访企业表示会“部署数据防泄漏系统”来应对员工使用AI工具所带来的风险，排名第一。该系统可以通过数据的全生命周期防护和动态风险拦截来支持高精度防护，减少企业的法律风险。其次有59%的受访企业选择 “实施API调用日志审计”，排名第二。该措施具有全链路可追溯和合规举证等特点，可以快速识别员工的数据泄露或API滥用等风险。

排名第三和第四的措施分别为“建立AI白名单制度”及“开展月度安全意识培训”，占比56%及54%。AI白名单仅允许经安全评估的AI工具在企业环境中运行，在降低供应链风险和合规管理难度方面具有优势，但灵活性较低。安全意识培训成本可控且有利于企业AI安全文化的渗透，但容易流于形式。选择 “采购企业级合规AI服务”的受访企业占比最少，仅有27%。

图 24 调查企业应对员工使用互联网AI工具的主要措施（多选）



综合调研结果来看，部分受访企业在推进AI变革过程中已逐步形成覆盖技术架构侧、数据语料侧、基础设施侧、组织体系侧的系统化推进思路。主要呈现以下特征：

- **在技术架构侧**，企业通常以落地场景的体系设计为起点，采取混合部署方式推进大模型落地应用，并统筹考虑云端风险应对、私域安全防护、模型幻觉应对等问题，以重构技术底座，适应创新之变；
- **在数据语料侧**，企业着眼于数据治理框架搭建和数据质量提升，推进数据语料的深度治理，并且已初步形成数据标准化体系、智能清洗工具等共性选择；
- **在基础设施侧**，企业兼顾创新与务实，充分权衡AI基础能力建设与业务需求满足，通过混合部署和协作管理措施等实现协同优化；
- **在组织体系侧**，企业立足敏捷性和协同性的组织机制要求，大力推进AI相关的团队能力建设和员工风险应对等，以软实力建设护航组织AI的硬核转型。

但是，调研结果也显示出绝大多数受访企业的AI Ready程度仍需追赶行业平均水平。此外，有相当多企业受限于场景渗透不足、技术能力欠缺、AI价值收益不清、安全治理过难等问题，尚未形成对AI变革路径的体系化认知，对AI变革相关的需求和挑战也缺乏合理估计。立足AI变革发展的时代潮头，推进千行万业AI变革已成共识，本白皮书试图在后续章节深入分析如何推进泛行业企业AI Ready进程，以便更好与各类合作伙伴共谋发展。

03

AI Ready硬实力变革关键要素剖析

企业AI Ready需硬实力和软实力双向共驱，硬实力是底层基础，软实力是上层建筑，企业硬实力的突破往往能催生出新型软实力形态，而软实力的前瞻布局又能反哺硬实力发展，形成相互强化的DNA双螺旋式发展。在AI发展阶段性态势下，以“预测下一个token”为核心的技术范式兴起，技术、数据、业务等硬实力要素不断突破固有边界，在token化的演变中迎来价值重构，部分企业已逐渐摸索出下一轮螺旋式成长路径，但还有更多企业仍在迷雾中徘徊。本章立足企业AI Ready硬实力变革视角，以业务应用层目标为驱动，纵向深挖基础设施层、模型服务与编排层的关键模块，横向细究服务治理层的重要构成，以求精准把握各要素的核心价值逻辑和变革新特性，助力相关企业锚定前进方向。

AI Ready硬实力变革分析框架

基础设施层是企业AI技术落地的物理与虚拟资源支撑体系，是承载AI模型训练、推理与部署的高可用性底座。它包含计算资源、网络架构、存储系统及数据语料等核心要素，为上层AI应用提供稳定、高效、可拓展的运行环境。

AI模型服务与编排层（AI Model Service and Orchestration Layer）是企业AI能力体系的中枢，扮演着承上启下的关键角色。它下接坚实的AI基础设施层（算力、网络、存储、数据），负责将底层资源与AI模型能力进行封装、管理与优化；上承多样化的业务应用场景，提供标准、高效、灵活的AI能力调用接口与智能编排服务。该层核心目标是屏蔽底层基础设施的复杂性，赋能上层应用的敏捷创新，确保企业能够高效、安全、可扩展地利用各类AI模型（包括基础大模型、行业模型、自研模型等）驱动业务价值。其关键要素包括多模型管理与服务化、智能体（Agent）与应用编排以及模型通信协议与集成。

服务治理层是保障企业AI系统合规、安全、可持续运行的安全防护和治理体系，涵盖安全可信AI和AI全栈治理两大支柱，旨在推动企业AI安全及治理从原则规范走向工程化落地。

3.1 基础设施层

计算

计算资源是AI模型处理海量数据、执行复杂运算的核心动力。

要素价值剖析

AI算力芯片

AI算力芯片是AI发展的“核心引擎”。企业AI变革必然带来海量数据处理和分析需求，尤其随着DeepSeek、GPT-4 系列等大模型加速普及，亟需依托充沛的底层算力实现模型高效训练和推理。不难看出，AI Ready的必要前提是AI算力芯片Ready。与擅长通用计算任务的CPU不同，GPU等AI算力芯片具备更高效的并行运算能力，针对特定的AI应用场景，还可以实现深度定制（如TPU、NPU等），达到更高效能比。当前，AI各类应用正不断从云端向边缘端和终端下沉，由此对AI算力芯片提出了不同的能力要求。云计算中心和企业私有云中心等云端场景常兼顾大模型训练和推理任务，强调芯片的高性能和高计算密度。智能制造、智能驾驶、智能家居等边缘端场景以推理任务为主，则更强调芯片的低功耗、小尺寸及实时性等特性，且越贴近终端场景，此类需求越发明显。

软件定义算力

面对AI算力短缺和成本高昂的挑战，软件定义算力通过虚拟化技术将物理资源（CPU、GPU等）池化，实现异构算力的统一调度和灵活分配。通过将算力模块化（如标准化刀片）、利用高速互联技术（如PCIe交叉矩阵）和算力规格模板等配置工具，可实现算力的按需定义、动态调整和弹性扩展，提升资源利用率，在“软件定义硬件”的基础上，实现“软件调用硬件”。

算力集群化管理

算力集群化管理是构建企业级AI基础设施的“必经之路”。从全周期视角看，企业AI变革是复杂的系统性工程，粗放型的“堆算力”模式将难以为继，存在成本高昂、资源浪费严重等问题，算力集群化管理强调集约化管理和智能化调度，有助于提高企业对算力资源的掌控力。在集约化管理方面，企业可利用异构算力统一接入、构建标准化接口、建立多层级集群协同机制、强化故障预防与恢复能力等，提高算力集群的集成性和稳定性。在智能化调度方面，可推动算力集群兼容多种操作系统、开发工具和算法框架，实现软硬件高度耦合，利用动态调度策略充分盘活算力资源，并结合液冷技术等优化算力能效。

AI变革新特性

开箱即用

目前，大部分企业面临着AI人才匮乏、前置资源投入压力大的困境，往往需要准确判断AI变革的主要任务，尽量做到“力出一孔”，借助各类“开箱即用”的计算解决方案可减少重复开发等不必要探索，降低试错成本，从而更专注于AI价值创造。

一物多用

伴随数字化转型的深入，越来越多企业对算力的认知已远超硬件范畴，“一物多用”的算力，正在成为企业优化资源配置的关键。“一物多用”的典型表征有单个机架即可支持多种AI任务，既能用于小模型的快速训练，又能满足大模型的精细调优需求，同时还能承担各类AI推理和应用服务。

安全保障

随着生成式AI 、物理AI的快速发展，AI推理需求将远超AI训练需求，且计算节点越发分散在边缘端和终端，安全入侵风险会日趋严重，将倒逼企业不断提升对计算环境的安全保障能力，包括实时监控资源使用状态、强化运维能力、及时预测安全风险等。

企业案例

某制造企业将思科的UCS-X模块化AI全栈解决方案AI Pod部署到非常靠近产线的位置，利用强大的AI能力实现产线的数字孪生预测性维护。基于思科验证设计的一体化最佳实践，不仅解决了工厂当地IT人才匮乏的挑战，还通过本地化处理数据降低了延迟、增强了企业产线数据的安全保护，AI Pod上线后显著减少产线设备停机时间和维护成本。

网络

网络是AI基础设施中数据传输的通道，其效率直接影响算力价值的释放。

要素价值剖析

智能网络架构

智能网络架构是企业AI变革的“加速器”，企业当前的人工智能（AI）和机器学习（ML）工作负载呈爆炸式增长，亟需为此构建可实现动态负载平衡、可观察、可扩展且低能耗的智能网络架构。例如，AI GPU通信的流量以“大象流”为主，而企业各类应用中还存在大量的“老鼠流”，两者的网络资源需求截然不同，在企业高度关注的统一化融合化架构中如何实现资源公平分配越来越成为网络智能调度能力的挑战。因而，企业更需要结合优化芯片等措施。具体来说，企业需结合优化交换芯片等措施持续提高交换转发的集约化程度，强化网络的全栈观察能力以实现对AI工作负载的实时检测，通过提高端口适配的灵活性来增强网络架构平滑演进的能力，并尽可能以少量设备覆盖更多端口需求，进而降低能耗和成本。

软件定义网络

企业AI应用往往与大型AI算力中心不同，从前后端融合或同质架构出发逐渐延展的AI网络发展道路更适合企业启动门槛低、按需增长、逐步迭代的建设思路，这就造成了独特的AI网络资源冲突问题。这要求交换机厂商在软件设计上针对AI模型流量进行深度优化，即在不影响性能的前提下，让流量更加灵活可控。软件定义网络（Software-Defined Networking, SDN）应运而生，即通过分离网络控制功能与硬件设备，基于软件编程实现网络的控制与管理，可极大提高企业网络管理灵活性，进而实现网络资源的优化配置与开放共享。

AI变革新特性

由“训推一体”的算力架构演进到网络架构

“训推一体”在网络架构层面，指的是构建能够同时高效的支持AI模型训练和推理需求的统一网络基础设施。AI训练网络需要承载大规模的GPU间通信，需要高带宽、无损无阻塞的高性能网络架构。而推理网络则更侧重于实时的推理数据交换，需要超低延时的网路架构。不管是训练还是推理网络不仅需要高可扩展性的架构来满足AI应用的敏捷性和可扩展性，还需要具备实时监控，异常检测和故障排查的综合可视化运维工具。随着AI应用加速融合和企业期望不断降本增效，训练和推理网络融合成为趋势。“训推一体”的网络架构旨在通过高性能交换设备、智能调度算法和软件定义能力，打破网络壁垒，实现对不同AI流量模式的统一承载和优化调度，支持分离部署向融合部署的平滑过渡，提升网络资源利用率和整体性能。

由SDN演进到意图网络

随着AI应用渗透，网络管理复杂度提升，需要网络具备自我演进和学习能力。意图网络（Intent-Based Networking, IBN）正是SDN智能化演进的下一站：SDN能实现代码级的软件定义，具备机器级别的自动化能力；IBN则能实现业务级的软件定义，具备解析人类意图和业务价值的自主化能力。IBN由意图翻译、方案执行、网络状态感知、自动优化四个流程构成，意味着该网络架构能基于意图翻译自动生成并部署网络策略，实时分析网络状态，实现从静态架构向业务驱动的智能动态系统转型，快速响应业务需求。

企业案例

某超大型互联网企业与思科合作，采用思科的Silicon One 51.2T 芯片开发并部署了其AI数据中心定制交换机，以增强其数据中心网络的性能和可扩展性。该交换机具备更高的前面板端口密度和性能，尤其在能效方面有显著的提升效果，满足了该企业超级算力中心未来的扩展需求。

某制造业集团在部署AI质检和预测应用时，遇到预训练、微调任务突增时，前端AI应用出现明显卡顿，影响一线操作体验。引入思科交换机后，前端微服务消息流响应延迟降低40%，即使在后端大规模训练高峰期，应用体验依然流畅，整体生产效率提升12%。

存储

存储系统负责高效、可靠地存取AI模型训练和应用所需的海量数据。

要素价值剖析

高性能存储

当前，AI大模型的数据处理模式已经从单一类型转向包含文本、图片、音频、视频等在内的多模态数据，导致原始数据量呈现指数级增长，往往达到PB级别。此外，各类AI计算任务相互交织，实时运算要求也愈发严苛。数据采集、清晰、加工到模型预训练、微调、推理等各阶段，GPU高并发、高吞吐量的I/O需求不断爆发，企业亟需借助高密度固态存储和NVMe架构等实现 “高速、高容量、高可靠”的高性能存储。其中，高速对应的数据传输速率要求一般超过1000Mbps，高容量对应的数据存储规模可达PB级，高可靠则意味着及时存储并快速处理海量数据，并尽可能保证数据的完整性及可靠性。

分布式存储

传统集中式存储系统因性能瓶颈和可靠性问题，愈发难以应对PB级乃至EB级的数据存储和管理需求。企业可借助分布式存储在多个计算节点间分散数据存储，应对不断增长的数据量和访问负载。当部分节点出现故障时，则可通过其他节点恢复数据，增强存储可靠性。随着企业AI业务等不断增长，还可通过增加存储节点实现存储容量的按需拓展和灵活管理。

弹性伸缩存储

企业AI开发与应用中，不同团队、不同业务场景对基础设施资源的需求差异大且调用频率一直在调整变化，要求存储系统高效实现资源的扩展或缩小，以保障应用在负载变化时始终高效运行，即弹性伸缩存储。具体而言，弹性伸缩存储具备自动化、实时响应和高度灵活等特性。自动化意味着无需人工干预，系统就能根据预设规则自动调整资源；实时响应则体现在系统对负载变化的快速感知与处理上，确保资源始终与需求匹配；高度灵活则体现在根据多种指标触发伸缩操作，适用于云计算、Web应用、数据库等多种场景。

AI变革新特性

存算一体

传统冯诺依曼架构中计算与存储分离导致数据交换瓶颈（存储墙）和高能耗（功耗墙），尤其影响关联性强的AI推理任务 。存算一体通过在存储器中嵌入计算能力，直接在存储介质处理数据，消除数据搬运开销，正随AI变革加速渗透。

冷热数据自治

随着大模型等AI技术不断与企业实际业务流程深度融合，所需数据资源越来越庞杂多样，企业开始以全新思维和视角挖掘原有的冷数据，实现冷热数据的智能识别与动态管理成为关键。这要求存储系统具备覆盖数据全生命周期的管理能力，能动态监测数据使用频率和价值的变化，实现冷热数据的智能化分层，及时将冷数据迁移至低成本介质，从而优化存储资源配置，降低存储成本并提升数据价值。

数据

数据是AI模型的“燃料”，其质量和治理水平直接决定AI应用的成败。

要素价值剖析

动态数据治理

AI大模型等技术快速迭代并不断深化应用，企业不仅要应对快速增长的实时数据处理需求，还面临着数据环境复杂化、动态化带来治理挑战，强化动态数据治理成为必要之举。动态数据治理强调实现数据的可得、可见、可管与可用。可得性是企业获取高质量数据资源的首要前提，核心在于实现多源异构数据的动态整合；可见性强调组织内数据使用过程的透明度和可追溯性，需借助元数据管理、各类数据可视化工具等实现数据的动态监测；可管性则依赖于技术防护体系和数据治理制度的双向支撑，确保企业的数据安全可控；可用性则强调数据赋能实际业务，可基于数据共享平台等，促进组织内数据流通与合作，以精准支持决策需求，提高协作效率。

多模态数据治理

AI模型从单模态向多模态演进、实际业务场景不断催生跨模态任务需求等背景下，多模态数据融合是大势所趋，正不断驱动企业推进相应的数据治理实践。但是，由于大多垂直行业或场景中原有数据基础较为薄弱，企业首先可能还需要解决从无到有的问题（包括数据采集整合等），其次才能开始建设多模态数据资产，核心在于通过跨模态数据关联，实现不同数据模态之间的信息互补，并保持数据间的一致性和可比性，进而挖掘出多模态数据的更多潜在价值。

合成数据生成

结合企业当前的AI变革实践来看，数据治理存在“不够用”“不好用”“不能用”三重困境，具体表现为数据消耗速度远超增长速度、质量参差不齐影响使用效果、各类隐私合规风险限制数据有效应用。对此，合成数据可满足企业增量、提质、拓应用的多重需求。合成数据是由算法生成的模拟数据，可将现有数据增强为规模更大、更具代表性的数据集，还能凭借与现实世界非直接对应的特性，一定程度上克服隐私问题。合成数据生成的技术路线多样，主要包括基于大语言模型（LLMs）、生成对抗网络（GANs）或扩散模型，以及统计与模拟的方法。实际应用中，这些方法常结合使用，优势互补，以提升合成数据的质量。

AI变革新特性

数据价值链重构

基于大模型技术，各类生成式AI应用以数据语料为原料，同时又不断产出数据语料作为价值交付物，而企业在场景侧对此类数据的分析、应用又可产生新数据，以及最终用户侧产生的交互和反馈数据等，都会反过来被投入到模型训练参数中，形成数据飞轮。长期来看，数据和应用的边界将逐渐模糊，数据价值与业务价值相互映射，会形成基于主业务流的信息价值链条（融合数据、流程和IT等信息），企业的数据安全治理体系也将随之发生结构性变化，具体体现有原始数据语料、模型训练数据集、应用服务数据、业务数据等的全域打通和综合治理，亟需在域间实现可溯源、可管理的访问控制等。

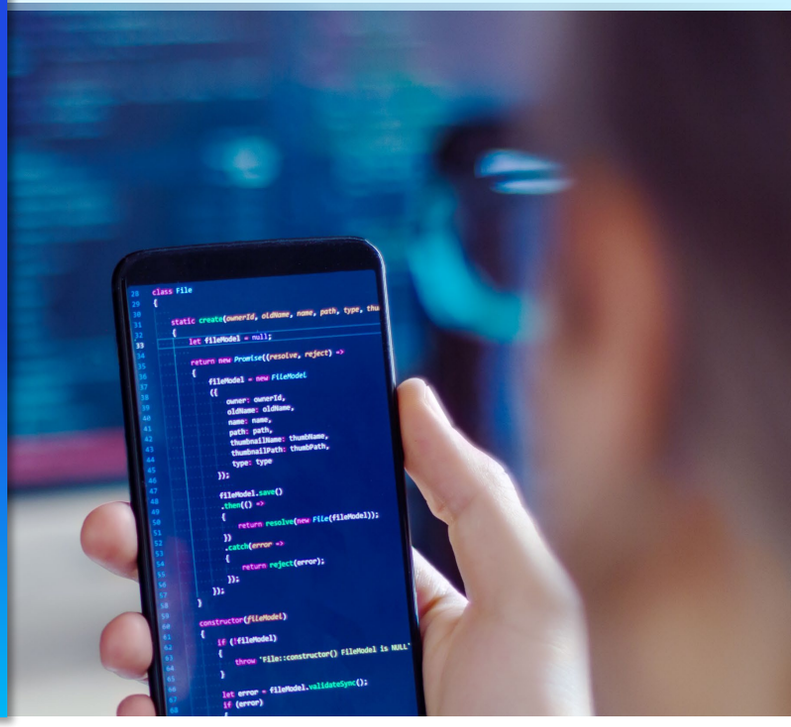
数据资产化

从数据资源到数据资产，意味着数据真正意义上成为企业核心竞争力的关键组成部分，企业的AI变革正在加速这一过程，不仅涉及技术层面的数据治理与产品创新，还关乎企业财务与合规管理的核心。这要求企业建立从战略规划到管理执行的全局认知，推动IT底座、业务流程、人才组织等全域变革，加快制定数据资产化战略、构建数据资产管理平台、加强数据资产评估和治理、培养数据资产相关的财务和管理人才等。

企业案例

某快消品公司在打造智能供应链时，利用Splunk挖掘销售、库存、物流等业务数据，为行业供应链AI大模型训练提供高质量数据源。同时，通过Splunk集成的AI功能，实时预测缺货风险并优化补货策略，库存周转率提升15%，大促期间断货率下降40%。

3.2 模型服务与编排层



模型管理与服务化

随着企业AI应用的深入，往往需要同时管理和使用来自不同来源、具备不同能力的多种AI模型。

要素价值剖析

多模型统一纳管

企业面临管理内部自研模型、外部采购模型（如商业闭源模型API）以及开源模型的复杂挑战。多模型统一纳管平台提供模型注册、版本控制、资源分配、权限管理、性能监控等功能，实现对异构模型资源的集中视图和统一管理，降低管理成本，提升模型资产的可控性。

模型即服务

将训练好的模型封装成标准化的API服务是实现能力复用的关键。模型即服务（Model as a Service, MaaS）平台负责模型的部署、弹性伸缩、负载均衡和健康检查，确保模型服务的稳定性和高可用性。得益于Kubernetes和微服务架构的技术支撑，上层平台具备创造卓越体验、可观测性和安全运行的基础，开发者无需关心底层复杂的模型部署和运维细节，即可通过简单的API调用来使用强大的AI能力。

模型性能与成本优化

对模型服务进行持续的性能监控（如推理延迟、吞吐量、资源利用率）和成本分析至关重要。平台应具备根据实际负载动态调整资源（如GPU实例类型和数量）的能力，并提供优化建议（如模型压缩、量化、蒸馏），以在满足业务需求的前提下，最大化资源利用效率和成本效益。

AI变革新特性

动态资源适配与调度

生成式AI等复杂模型对资源需求波动性大。平台需具备更智能的动态资源适配能力，不仅能根据负载自动伸缩，还能结合模型特点、任务优先级、成本预算等因素进行跨集群、跨地域的智能调度，自动匹配最优的计算和存储资源。

模型市场与发现机制

为了促进模型共享和复用，企业内部或行业生态可能构建模型市场。提供清晰的模型能力描述、性能指标、使用文档和评价体系，便于开发者快速发现、评估和选用合适的模型服务。

面向特定场景的模型优化服务

针对特定行业或业务场景（如金融风控、智能客服、工业质检 ），平台可提供预置的微调（Fine-tuning）流程、数据集和优化工具，帮助企业基于基础模型快速构建高性能的场景化模型服务。

智能体与应用编排

单一模型往往难以完成复杂的业务任务，需要通过智能体（Agent）和工作流编排将多个模型、工具和数据源协同起来。

要素价值剖析

Agent智能编排

AI Agent具备感知环境、进行规划、执行动作和调用工具（包括调用其他AI模型API）的能力。Agent编排平台提供定义Agent行为逻辑、管理Agent技能库（Tools）、协调多个Agent协作完成复杂任务（如自动化报告生成、多轮对话式数据分析）的框架和工具，是实现复杂AI应用的关键。

低代码/零代码应用构建

结合Agent和MaaS能力，低代码/零代码平台允许业务人员通过拖拽、配置等方式快速构建AI驱动的应用，将模型能力便捷地嵌入业务流程，进一步降低AI应用门槛，加速业务创新。

工作流与业务流程集成

AI能力需要无缝融入企业现有的业务流程和IT系统。编排层需要提供强大的工作流引擎和集成能力（如API网关、消息队列、事件驱动机制），将模型服务、Agent能力与CRM、ERP、自动化运维等系统连接起来，形成端到端的智能化解决方案。

模型通信协议与集成

标准化模型通信协议（Model Context Protocol，MCP）类似于网络中的TCP/IP，定义模型服务接口、数据交换格式、调用方式等的标准化协议（如OpenAI API规范、或企业自定义的MCP）对于实现异构模型和应用之间的互操作性至关重要。MCP协议库简化了不同服务间的集成难度，提高了开发效率；API网关作为模型服务的统一入口，提供请求路由、认证授权、流量控制、熔断降级、协议转换等功能，保障模型服务的安全、稳定和可管理性。

AI变革新特性

自适应与自学习编排

编排平台需要支持人类在Agent执行过程中进行干预、监督和反馈的机制（Human-in-the-loop），实现AI辅助决策和人机共同创造。 同时更高级的编排系统将具备一定的自适应和自学习能力。能够根据任务执行效果、用户反馈和环境变化，动态调整编排逻辑、优化工具选择或触发模型再训练，实现持续改进。

可解释性与可追溯性

随着编排流程日益复杂，确保决策过程的可解释性和结果的可追溯性变得尤为重要。平台需要提供清晰的日志记录、可视化执行路径和决策依据解释，满足合规审计和问题排查的需求。

企业案例

某制药企业在推进新药研发智能化过程中，采用了内置OpenShift、Cisco Cilium与Hubble的Cisco AI Pod，采用基于Cisco Validate Design（CVD）的设计快速搭建起符合NVAIE标准的全栈AI平台。相比传统部署方式，AI模型上线周期缩短了40%，总拥有成本下降30%，加速了新药筛选与验证进程，大幅提升了研发效率。

3.3 服务治理层

安全可信AI

AI 带来了新的机遇，但也带来了新的安全风险。与传统感知类、决策类AI应用相比，生成式AI的智能水平更高、适用领域更广，对企业业务乃至社会影响力更大，对应的AI安全可信问题将随着AI在企业的深入应用越发关键，企业亟需实施端到端、全流程、全方位的AI安全防护。

要素价值剖析

基础设施层安全

随着越来越多企业实施大模型本地化部署，大模型技术的基础设施属性逐渐凸显，但其代码开源化、API接口开放化的属性，也使得企业基础设施层的风险暴露更加泛化。例如，攻击者通过逆向工程系统性可分析出基础模型架构的漏洞，进而实施DDoS、HTTP代理攻击和僵尸网络流量等多样化攻击，并且模型开源程度越高，攻击门槛可能越低。对此，企业亟需结合AI大模型特性升级改造传统的基础设施层防护体系，不仅需要确保底层硬件设备安全，还要保证物理环境、网络环境、AI训练环境等的安全，包括针对关键设备设计和部署必要的安全性监控程序、优化数据中心环境控制系统、及时检测和防御针对AI系统的网络攻击等。

模型服务与编排层安全

模型服务与编排层的安全挑战呈现出技术脆弱性与业务危害性相互交织的特点。技术脆弱性方面，大模型相关应用的开发及部署不仅面临第三方库漏洞、数据泄露等传统安全风险，还引发了如提示词注入攻击、模型窃取及篡改等新型风险。业务危害性方面，越来越多的不同来源的模型能力将被封装在应用内部，也会给企业自身运营和业务价值带来各类安全隐患。考虑到大模型的应用广度和深度仍在持续拓展，相关企业迫切需要对AI平台及系统风险实施的全链路扫描与识别，保证模型纳管、服务化、组件编排等过程的“可见可溯”，利用模型加密、模型水印等进行专项防护，还要注重大模型平台与其他平台的网络隔离，以尽量减少被攻击面。

应用层安全

应用层安全与业务场景需求和终端用户体验紧密相关，但当前企业对于生成式AI应用上线后对组织运营和业务价值影响如何、应用风险发生后对商业竞争力的破坏程度如何等问题普遍缺乏较清晰认知和合理估计，也导致对应的预防及防护措施相对不足。面向业务场景需求，结合领域知识库、场景工作流和应用组件等实施定制化的安全解决方案成为必要。面向终端用户体验，大模型生成内容是影响用户体验的关键，可通过恶意攻击意图识别、对话内容安全检测、对话内容安全加固等确保内容输出安全和合规，还可通过优化交互流程、引导式提问等减少用户误操作和不必要的信息输入，提升用户体验的同时降低潜在风险。

可信AI

AI伦理已在世界范围内引起监管机构的广泛重视，也正成为企业商业竞争力的关键要素之一。企业面临的相关风险包括：AI缺乏解释性引发的监管和合规风险；AI辅助决策错误引发的经营和财务风险；算法偏见与社会道德、价值观分歧引发的声誉风险。可信AI强调以负责任和合乎道德的方式设计、构建、部署和使用人工智能解决方案，企业可强化对AI适应性（跨平台灵活性和对抗网络攻击威胁）、公正性（如人机对齐）、完整性（如数据全生命周期管理）、可解释性（如AI系统透明化）等要素的监督，确保其在实际应用中的可靠性和合规性。

AI变革新特性

可视化

AI领域在技术本质上仍存在透明度问题，可视化则强调企业能够发现公有云、私有云、混合云等各类环境中所有人工智能用户访问、工作负载、应用程序、模型和数据等，并清楚掌握其运行状态，对于提高AI技术透明度（如算法架构的可解释程度）、过程透明度（如数据处理流程的可追溯性）、结果透明度（如输出结论的推理过程）至关重要。

强检测

针对愈发复杂多变的AI安全风险和威胁环境，企业需在相关重大损害可能发生前，通过强检测及时发现风险异常并采取响应措施，包括精准识别可能使AI应用程序面临风险的配置错误、安全漏洞和对抗性攻击等。

广覆盖

各类生成式AI工具免费化、简易化趋势下，影子AI问题愈发普遍，企业需构建广覆盖的防护方案，以保护全体员工对第三方人工智能工具的访问，包括通过管理控制台配置拦截器，使员工流量必须经过安全模型检测等。

可落地

各类AI安全风险快速演进趋势下，不仅考验企业安全团队结合最新威胁情报及时创新安全策略的能力，更考验AI安全解决方案动态调整、灵活应变的落地能力。企业可通过搭建安全模型套件实现安全防护能力解耦，进而基于最新安全策略配置定制化安全模型进行流量检测。

某全球领先的芯片架构和知识产权提供商对IT和AI安全高度敏感，利用思科的解决方案构建了端到端的零信任AI应用运行环境。借助Cilium的Identity Based Network Policy能力，该企业对分布在多个海外云平台上的AI应用实施了精细化的零信任网络访问控制。同时，该企业引入了Splunk Enterprise Security，可收集并分析从内核级别到安全设备的全方位数据，并基于MITRE ATLAS的AI安全框架实现了精确的关联分析与告警。在此基础上，该企业正计划全面引入AI Defense，进一步增强AI应用自身的安全能力，以确保在复杂多变的跨境业务环境中构建强大而稳定的安全平台，确保内外AI应用的安全性。

企业案例

某出海企业希望其AI应用能够在本地数据中心、国内云端与海外云端之间实现业务冗余、负载均衡、灰度上线以及统一的安全策略。在考察了众多技术路线和方案后，该企业选择与思科合作，采取了以Cilium为核心的精简部署方案，由此实现了在公有云与私有云上的统一调度与策略制定，从而可以内置解决过往需要多个工具才能实现的功能，大幅提升AI应用的灵活性和可靠性。

AI全栈治理

当前企业AI应用普遍采用云原生架构，终端用户对AI应用的体验感知与IT基础架构走向分化。分化并非割裂，而是走向内生，从底层基础设施到上层应用的全技术栈联合治理和全栈指标定量建模，将为上层AI应用构建起无形但又无处不在的体验保障，并实现AI价值显化。

要素价值剖析

AI体验优化驱动的全技术栈联合治理

相较传统应用的独体式架构，云原生架构下的AI应用功能可以实现微服务化，功能更新能精准到具体节点上，且支持弹性拓展和分布式协作，从而为终端用户创造更加丝滑流畅的应用体验。但是，微服务链条存在节点依赖性，单个节点的风险安全极易威胁到整体应用体验，传统的单节点补丁式的治理模式，不仅难以实现跨节点的大规模检测，更是要求系统本身开放更多授权，会直接导致攻击面指数级扩张。因此，面向AI微服务体验优化的IT治理，应当既能保障上层微服务组件间的灵活调度，又能充分维持底层系统的稳定性，全技术栈联合治理成为必要。

其中，拓宽拓深全栈可观测能力至关重要。广度方面，AI微服务化、算力异构化等趋势下，企业的可观测力要求早已不局限于AI数据中心，而是要向园区、广域网、互联网延展，覆盖用户桌面、移动终端甚至使用AI能力的IoT设备。深度方面，企业需要打破传统的竖井化观测体系，避免云原生应用、云原生平台和底层基础设施“各自为战”，从底层基础架构芯片级的遥测，到微服务和应用层的观测，做到“深入贯通”，建立起一体化的观察力体系。

AI价值显化驱动的全栈指标定量建模

传统AI治理常陷于技术与业务指标割裂的困境，前者聚焦业务负载、模型精度等纯技术参数，后者关注收入增长、转化效率等商业结果，认知鸿沟难以弥合。此外，当前企业AI变革也面临着长期发展潜力可期，但短期量化价值难显的窘境。在全技术栈联合治理基础上，推动全栈指标定量建模，有望打通AI技术释能到价值显化的转换链路，更好实现企业AI应用及治理的定量化衡量和决策。具体来说，企业可在治理域构建全栈治理平台，实现全栈健康状态的定量化与关联，进一步地，可基于治理平台将底层的技术指标与业务关键指标（如用户转化率、收入增长率）相关联，通过数据建模与分析，量化各技术环节对业务目标的实际贡献。这一过程中，跨服务依赖关系的治理、系统异常的实时性检测和预测性分析等水平越高，则越有利于实现全面、稳定、精准的量化评估。

AI变革新特性

大模型赋能AI治理

AI既是治理对象，又是治理工具，尤其是大模型技术兴起以来，凭借其强大的语义理解与生成能力，正在重构企业AI治理的底层逻辑。以底层元数据管理为例，AI大模型能自动生成带语义标签的元数据网络，可大幅减少对人力维护的依赖。而在上层人机交互方面，企业AI治理平台可基于大模型开发自然语言交互界面，让业务人员直接参与治理，进一步打破业务与技术的沟通壁垒。结合企业AI全栈治理来看，大模型技术还可驱动自动化的运维AIOPS，与各类管理软件无缝集成，实现全栈安全监测、根因分析与决策建议。

企业案例

某零售企业CIO在推动AI项目时，曾面临一项巨大挑战，即企业投入了大量资源构建AI能力，却始终无法清晰地了解资源投入对最终AI用户体验、安全性和业务成果的影响程度。更糟糕的是，AI投资与企业的业务目标是否一致也难以量化评估。为了解决这一难题，该企业引入了Splunk ITSI。通过Splunk ITSI平台，其将所有AI全栈指标统一汇集，构建了从业务目标到技术栈的全链条映射，并通过定量化分析得出了每个组成部分对业务目标的影响程度。由此，该企业可以识别出影响最终业务价值的AI体验与安全的薄弱环节，并精准优化。最终，该企业CIO借助Splunk ITSI所体现的系统化治理方式对齐了AI资源投入与业务价值，企业也得以真正实现AI从投入到转化为业务成果的全方位提升。

04

AI Ready 变革评估体系

本章制定了面向泛行业企业的AI Ready变革评估体系，旨在结合企业AI Ready从初始状态发展到最优化的过程，基于各阶段企业的能力特点来定义就绪度等级。企业可结合此评估体系客观评价自身的AI发展水平，明晰在行业竞争中的相对位置，并结合等级评估结果，制定面向更高等级的能力提升计划，认准变革方向，抓住关键问题，找对发展路径。

4.1 评估体系概述

AI Ready变革评估体系包括企业架构、数据语料、基础设施、组织体系等4大评估维度，每一维度下拆分出不同的评估指标（共计13项），并进一步细分出二级评估指标（共计41项），评估每项指标的就绪度等级后综合计算得出企业整体AI Ready的对应等级。

评估维度

评估维度	评估指标	评估标准
企业架构	架构设计	企业能否将AI战略解码为具体的、可执行的目标，并基于此设计出成配套的全栈式企业架构，包括业务架构、应用架构、数据架构、技术架构等。
	架构治理及管控	企业能否通过系统化的流程设计、规范岗位职责、实施全生命周期管控等对业务架构、应用架构、数据架构、技术架构等的AI变革实施路径进行补充完善。
	架构评价	企业能否就架构落地后对组织AI变革和AI相关业务的支撑力度和实际价值进行综合全面的考察。
数据语料	数据需求	企业能否针对AI变革相关的业务运营、经营分析和战略决策等环节，识别所需的数据，确定数据需求优先级并以文档的方式对数据需求进行记录和管理。
	数据设计和开发	企业能否设计并实施数据解决方案，以持续满足AI变革的数据需求，数据解决方案包括数据库结构、数据采集、数据整合、数据交换、数据访问及数据产品(报表、用户视图)等。
	数据运维	企业能否为AI变革的提供持续可用的数据资源，即在数据采集、数据处理、数据存储等过程中保障相关数据平台和数据服务的日常运行和维护。
	数据退役	企业能否根据法律法规、业务、技术等方面要求，执行历史数据的归档、迁移和销毁工作，确保对历史数据的管理符合外部监管机构和内部业务用户的需求,而非仅满足信息技术需求。
基础设施	按需配置	企业能否以AI变革相关业务目标为导向，制定合理的基础设施技术路线和实施方案。
	技术选型	企业能否以投入产出最优化原则，兼顾AI变革需求和资源限制，制定合理的基础设施技术路线和实施方案。
	灵活调整	面向各类业务场景需求，企业的基础设施能否及时与业务挑战的性质和需求相匹配，确保有效的解决方案。
	稳定运行	面向业务连续性发展的需求，企业的基础设施能否充分降低系统性风险，确保企业整体的稳健运营。
组织体系	组织与制度	企业是否具备统筹管理与推进职能转型的企业级敏捷管理组织结构和配套治理机制。
	组织与人才	企业是否就AI变革明确了人才的关键能力要求，并建立了体系化的培养机制，以及配套的绩效考核与激励机制。

评估指标

企业架构评估指标体系		
一级指标	**二级指标**	**评估说明**
架构设计	战略解码的合理性	战略愿景与架构设计的匹配度。
	配套架构的完备性	架构体系的系统完善度及各组成部分间的协同性。
	目标及执行的可落地性	目标和执行计划是否切实可行。
架构治理及管控	架构需求管理能力	能基于AI战略和业务变革方向及时梳理出对各类架构的关键需求。
	架构一致性管理能力	能基于关键需求保障各类架构在目标及原则、标准与规范、组织与流程等层面的逻辑一致性。
	架构资产管理能力	能对架构资产实现元模型级的全周期管理和应用分析。
	架构工具建设能力	能打造标准化的工具链（建模/管控/评估）将抽象的架构原则转化为可执行、可验证的工程实践。
架构评价	组织影响	对企业现有的管理理念、模式和人员思想等的支撑力度和实际价值。
	业务影响	对企业核心业务或AI变革强相关业务的支撑力度和实际价值。

数据语料评估指标体系		
一级指标	**二级指标**	**评估说明**
数据需求	及时性	数据语料采集的及时程度，即能否满足数据使用的时间要求。
	支撑性	已采集数据语料对AI应用的支撑程度。
	可用性	已采集数据语料的可用程度，即数据是否有明确的定义和口径，数据的粒度和精度是否满足使用需求，数据是否可以被理解和妥善管理。
	多元性	外部数据语料补充利用程度，是否有通过第三方合作或购买第三方数据等方式补充数据维度。
数据设计和开发	扩展性	数据解决方案设计时对数据语料扩展的考虑完善程度，即是否估算未来增长情况，增长情况是否符合预期。
	共享性	对跨域的数据语料共享机制及共享管理机制的完善程度。
	安全性	数据语料安全策略的完善程度，包括全生命周期监测、合理的人员权限设置等。
	周期性	系统内数据语料全生命周期管理机制的完善程度，包括重要的数据语料可追溯等。

数据语料评估指标体系		
一级指标	二级指标	评估说明
数据运维	指标化	指标体系及报表体系的完善程度，以及指标体系及报表体系维护机制的完善程度。
	预置化	数据语料预先加工的可使用程度。
	智能化	数据分析、数据挖掘等模型的完善程度及准确程度，以及配套管理维护机制的完善程度。
	融合化	多元异构数据语料的关联融合程度。
数据退役	标准化	企业级数据语料标准的制定与落地情况，包括制定数据标准管理制度、建设数据语料管理工具等。
	精准化	能否在就数据语料质量进行不同管理颗粒度（企业级、项目级、产品级）的评估和管理。
	体系化	数据语料质量评估及管理制度建设的系统化程度。
	信息化	元数据管理制度及工具的完善程度。

基础设施评估指标体系		
一级指标	二级指标	评估说明
按需配置	协同度	企业现有基础设施采用的技术路线与AI变革战略和AI业务发展的协同程度。
	满足度	企业现有基础设施对不同优先级业务需求催生的运算任务的性能满足程度。
技术选型	匹配度	企业现有基础设施采用的技术路线与自身资源禀赋（硬件储备、资金实力等）的匹配程度。
	成熟度	技术路线的系统化程度和规划完善程度，以及相关技术组件的通用性。
	领先性	技术路线在同行业的领先程度。
	开源性	内部源代码和开源代码的有效利用程度。
灵活调整	灵活性	系统功能是否能够灵活应对复杂的业务场景。
	集成性	系统内功能模块间的集成程度，自动化运作程度。
	扩展性	在有新的分析需求时系统功能能否复用。

基础设施评估指标体系		
一级指标	二级指标	评估说明
稳定运行	操作规范化	是否有配套的文档约束和支持操作，从而实现操作上的规范。
	运维常态化	是否配套有稳定、安全的运维管理能力。

组织体系评估指标体系		
一级指标	二级指标	评估说明
组织与制度	组织结构	组织定位及构成的清晰程度、组织机构设立和岗位职责设立的完善程度、组织演进路线的敏捷程度等。
	治理机制	组织内各部门、单元等组成部分的融合协作程度，针对岗位人员的绩效考核和激励方案的完善程度等。
技术选型	选人	AI变革相关人才能力评估标准的完善程度。
	育人	AI变革相关人才培训评价机制的完善程度，包括是否根据不同人才类型设施差异化内容、是否定期优化等。
	留人	AI变革相关人才的绩效考核与激励机制的完善程度。

评估方法

企业可基于当前AI发展现状，参考AI Ready变革评估体系，通过问卷调研、员工访谈、数据分析等方式开展就绪度的自评估。

- **评估准备阶段：**企业可依据整体评估框架梳理出AI变革相关的过程文档、会议记录等，组建评估团队并制定评估计划。
- **正式评估阶段：**首先，对各二级评估指标进行就绪度等级评分，得出一级指标评估均分（一般为5分制）；其次，通过设置指标权重计算出各评估维度的得分，以反映企业在各维度的数字化能力；最后，将各维度得分加权计算后，得出企业整体的AI Ready等级（一般可分为5级）。关于权重设置，企业可结合实际情况灵活调整。
- **评估报告阶段：**企业可将自身AI Ready水平自低向高划分为1-5级，依次为初始级、受管理级、稳健级、量化管理级和优化级，除形成总体性的等级评估报告外，还可形成各维度评估报告，以便针对性地进行AI变革优化。

4.2 评估报告模板

总体等级评估

就绪度等级	对应评分区间	整体描述	能力提升重点
初始级（一级）	N ≤1	企业主要以项目制形式推动AI变革，没有正式的AI发展规划、管理组织及流程等，严重依赖部分人员的个人经验和临时决策。	• 统一内部变革意识 • 提高战略驱动力 • 强化底层技术基础
受管理级（二级）	1<N ≤2	企业开始意识到AI发展的重要性，初步建立了AI变革组织架构和职责分工，但尚未形成系统化路径。	• 深化自身业务需求分析 • 明确规划AI应用场景 • 强化技术底座和数据语料基础
稳健级（三级）	2<N ≤3	企业将AI视为实现发展目标的重要工具，开始将AI技术与业务流程深度融合，形成了较为完善的发展路径。	• 打通端到端的AI工作流 • 加强AI模型、算力、数据等的资产化管理 • 挖掘数据价值并形成闭环，聚焦价值合理规划投入
量化管理级（四级）	3<N ≤4	企业将AI视为获取竞争优势的战略资源，在多场景中实现了AI技术的深度应用，显著提升了业务效率，且建立了较完善的AI治理体系。	• 打造内生AI基础设施 • 推动AI跨场景、跨业务、跨行业的规模化应用 • 向外输出最佳实践和方案
优化级（五级）	N>4	企业将AI视为生存与发展的重要基础，在多场景、多业务领域实现数智化转型，AI技术及治理水平达到行业领先，开始或已经能够主导行业相关标准的制定工作。	• 构建核心产业生态圈 • 创新传统管理模式、业务模式、商业模式等 • 引领解决行业共性问题

四大维度评估

企业架构就绪度评估

评估指标	评估部门	评估均分（5分制）	权重	整体现状描述
架构设计	战略部门	2.42	**	某企业初步在制度、组织、流程方面提出了企业级AI变革战略规划，基于此开展配套的企业架构设计，并设置了动态调整机制。
架构治理及管控	所有部门	2.03	**	某企业目前以部分AI变革重点项目为核心，开始推进业务架构、应用架构、数据架构、技术架构等的过程管控。
架构评价	所有部门	1.85	**	某企业对当前的AI变革可能为自身业务带来的影响有一定研究，从偏定性角度得出了分析结论，但定量分析结论较少。

注：评估均分和整体现状描述仅作为示例，不构成对任何实体企业的评估分析

企业架构就绪度评估

评估指标	评估部门	评估均分（5分制）	权重	整体现状描述
数据需求	所有部门	1.96	**	某企业目前建立了数据需求管理制度，并结合大模型相关训练需求建立了专项管理制度，但在数据的定义和口径统一管理方面上不完善。
数据设计和开发	IT部门	1.28	**	某企业目前主要根据特定AI项目需求设计、实施数据解决方案，并未在部门层面建立数据解决方案设计及实施数据语料共享机制。
数据运维	IT部门	2.05	**	某企业目前在部分AI业务领域建立了数据提供方管理流程（数据溯源、职责分工与协同工作等），并建立了数据运维规范。
数据退役	IT部门	1.38	**	某企业目前只在个别项目和系统中开展数据退役管理，主要是收集数据保留和销毁的内外部需求，设计并执行方案等。

注：评估均分和整体现状描述仅作为示例，不构成对任何实体企业的评估分析

基础设施就绪度评估

评估指标	评估部门	评估均分（5分制）	权重	整体现状描述
按需配置	所有部门	3.16	**	某企业目前已基于部分业务场景、业务功能已搭建可复用的通用化AI服务组件，可根据业务需求灵活配置智能化模版。
技术选型	IT部门	2.78	**	某企业目前采用了部分业界成熟的智能化基础设施解决方案，但关键核心技术对外依赖程度较高。
灵活调整	IT部门	2.53	**	某企业目前已探索运用服务化、组件化系统云原生技术架构，建成统一的、可扩展的、高可用的AI基础设施平台。
稳定运行	IT部门	2.17	**	某企业目前已通过事件告警、日志告警方式监控AI网络、AI计算、AI存储、AI系统状态，对各类资产、配置信息数据实现规范化管理，但尚未完成以场景化方式实现各类运维场景的端到端自动化。

注：评估均分和整体现状描述仅作为示例，不构成对任何实体企业的评估分析

组织体系就绪度评估

评估指标	评估部门	评估均分（5分制）	权重	整体现状描述
组织与制度	所有部门	2.38	**	某企业目前初步建立了AI变革的统筹协调机制，成立了非常设性的AI战略委员会，对应的组织架构、岗位设置、管理流程等较为清晰。
组织与人才	所有部门	3.46	**	某企业目前就AI变革明确了人才的关键能力要求，并建立了体系化的培养机制，但未设置配套的绩效考核与激励机制。

注：评估均分和整体现状描述仅作为示例，不构成对任何实体企业的评估分析

4.3 AI Ready 变革行动指南

以评促建

谈及企业AI Ready，千头万绪当止于一端，企业可着手搭建客观合理的全方位评估体系，通过“评价-诊断-行动-再评价”的闭环机制精准判断自身AI发展水平，方能精准制定AI变革策略。

价值为锚

企业并非为了AI而AI，落地价值才是锚点，企业可结合最小可行产品（MVP）测试、小范围试点、理论预演等方式推进 “低成本试错”，快速锁定价值相对高、复杂度相对可控的高潜力场景。

安全为纲

AI技术迭代带来的安全风险已远超技术范畴，不仅会造成经济损失，还会损害企业最宝贵的资产——信任，基于AI全生命周期视角，统筹多维度安全防护，构建全栈式AI安全，平衡技术创新与风险管控，是企业AI变革的基本纲领。

内外兼修

AI变革涉及企业生产经营方方面面，对内需打破各部门壁垒，让各类型员工都能理解AI发展目标和任务，提升运营管理效率，对外需优化产业链、生态圈资源配置效率，积极向投资者和客户展示相关成果，内外兼修，方能全面进化。

架构先行

AI变革的复杂性决定了单点式、碎片化发展必然无法实现价值最大化，企业可凭借前瞻性的架构设计强化系统性支撑，推动技术创新、数据融合、业务增长、组织变革的协同共振。

快速迭代

AI技术前景与商业价值尚存双重未知，而企业的快速迭代能力可转化为风险缓冲器，助力企业在不确定性中锻造确定性，针对AI技术创新加速、应用需求波动以及全球化竞争加剧等挑战，迅速做出反应并调整策略。

筑牢底座

企业需立足长远发展，实现AI变革由局部优化到全局智能，夯实技术底座至关重要，可推动基础设施层、模型服务与编排层、业务应用层及服务治理层等的分层解耦，通过算力集约、数据贯通、算法进化构建起坚实的AI底座。

总结与展望

“鉴过往”而明道。本白皮书从技术与应用、基础资源与架构调整、创新与治理等多角度深入研究企业当下可以把握的AI发展机遇，并对企业的AI Ready能力按照硬实力、软实力进行了细致拆分，力求在助力企业理解宏大趋势的前提下，找到具体务实的落脚点。更进一步地，我们基于对泛行业企业AI变革实践的调研，总结了有益经验和发展路径，深入剖析计算、网络、存储及相关治理组件等关键要素的价值，并结合过往服务案例与经验提炼了 AI Ready 变革评估体系，以便企业能根据自身情况创新应变。

“知未来”方致远。人工智能作为指数型技术的发展速度极快，我们预见到开源生态带来的普惠力量将加速渗透到各行各业，带来人机协同深度融入企业运营与管理、产品服务从“功能实现”演变到“体验重构”、数据资产作为企业核心竞争力的价值愈发凸显、安全可信AI倒逼企业强化内部治理和风险管理体系等多重趋势。毫无疑问，企业AI Ready变革路径还将动态调整，我们唯恐力有不逮，将持续关注各类企业AI发展动态，及时总结客观规律和科学方法，创新更多开箱即用、务实求效的“一体化解决方案”“AI Ready变革评估”等产品和服务，携手各界同仁行远自迩，笃行不怠。

企业介绍

思科

（NASDAQ：CSCO）

思科（NASDAQ：CSCO）是致力于以AI技术重塑企业连接与防护方式的全球科技领导厂商。40余年来，思科始终以安全为核心连接世界，通过AI驱动解决方案帮助客户、合作伙伴及社区释放创新、提升生产力并增强数字韧性。秉持“包容性未来”愿景，思科持续推动全球互联。

思科在2024《财富》世界500强中排名第246位，并屡次在全球著名职场研究认证机构卓越职场®研究所（Great Place to Work® Institute）发布的“最佳职场”榜单中名列前茅。

毕马威

毕马威中国在三十一个城市设有办事机构，合伙人及员工超过14,000名，分布在北京、长春、长沙、成都、重庆、大连、东莞、佛山、福州、广州、海口、杭州、合肥、济南、南京、南通、宁波、青岛、上海、沈阳、深圳、苏州、太原、天津、武汉、无锡、厦门、西安、郑州、香港特别行政区和澳门特别行政区。在这些办事机构紧密合作下，毕马威中国能够高效和迅速地调动各方面的资源，为客户提供高质量的服务。毕马威是一个由独立的专业成员所组成的全球性组织，提供审计、税务和咨询等专业服务。

kpmg.com/cn/socialmedia

如需获取毕马威中国各办公室信息，请扫描二维码或登陆我们的网站：
https://kpmg.com/cn/zh/home/about/office-locations.html

刊物编号：CN-TMT25-0001

二零二五年五月印刷