



AMIA
应用现代化产业联盟

CSIA 中国软件行业协会
China Software Industry Association

CAICT 中国信通院

应用全生命周期智能化 研究报告



编写单位

应用现代化产业联盟
中国软件行业协会
中国信通院人工智能研究所

顾问组成员

陈 纯 应用现代化产业联盟 会长
陈宝国 中国软件行业协会常务副秘书长
魏 凯 中国信通院人工智能研究所 所长
张宇昕 华为云 CTO
黄 琪 华为云副总裁（战略与产业发展部总裁）
徐 峰 华为云 PaaS 服务产品部部长

编写组成员

文永新、梁广泰、王一男、乔蔚云、高天赐、刘 恒、黄慧娴、
秦思思、闫东伟

贡献者

蒋 涛、黄 颖、王翩宇、杨国龙、朱少民、汪维敏、孙彬彬、
翟传璞、李丹丹、林 扬、罗 斌、曹 峰

出版时间

2025 年 6 月

前言

当前，以大模型和智能体为代表的人工智能技术正驱动新一轮产业变革，推动应用架构从功能自动化向认知智能化跃迁。本研究报告系统阐述应用智能化的框架，关键技术和实现路径，旨在帮助企业应对应用智能化三大核心挑战：

1. 技术链断裂导致的落地周期过长：算力、大模型、智能体的新概念，开发、运维、安全合规等等环节的割裂都导致企业落地智能化的周期依然较长
2. 场景选择困难：单点试点难以穿透企业核心业务链的转型挑战，ROI 的评估难度大
3. 产业协同低效：技术供应商能力分散与企业需求的差距

通过对全球智能化尤其是应用智能化实践的实证研究，本研究报告提出智能应用的六大特征，并以此构建支撑智能化应用孵化和落地的关键技术和技术底座——智能应用平台。

本研究报告旨在为企业智能化转型提供系统性参考。通过解析智能应用的核心特征、构建分级成熟度模型、设计分步实施路径，我们希望为不同数字化基础的企业厘清转型逻辑。在技术快速迭代的背景下，本框架将持续演进，期待与产业界共同探索最佳实践，务实推进应用智能化落地进程。

——《应用全生命周期智能化研究报告》编写组

序言一

迈向 AI-Native 时代：智能体驱动的应用现代化新范式

近两年来，随着 ChatGPT、DeepSeek 等大模型技术的快速发展，人工智能正以前所未有的速度重塑产业格局，而 Agent 智能体的崛起标志着应用现代化迈入全新阶段。未来大于 50% 的人类工作任务场景将被大模型影响，据麦肯锡预测，生成式 AI 每年可以增加 2.6~4.4 万亿美元经济收入，AI 驱动的自动化决策与交互将成为业务常态。与此同时，传统应用与 AI 应用的融合不再局限于简单的功能叠加，而是依托数据与 API 的深度协同，构建起“感知 - 决策 - 执行 - 学习”的闭环智能体系。在这一趋势下，应用现代化的核心命题已从“云原生”升级为“AI-Native”，即应用的全生命周期——从开发、运行到运维、集成——均需围绕智能体的自主性、协同性与进化能力重构。

从技术视角看，AI-Native 架构的关键在于数据与 API 的智能融合，传统企业系统沉淀的海量数据，需通过统一的可信数据资产目录实现跨域流通，为 Agent 提供实时、高质量的训练与推理燃料；而 API 则从传统的服务接口演进为“智能连接器”，通过动态编排与语义理解，支撑 Agent 间的自主协商与任务分解。此外随着云原生 AI 技术（如多模态大模型、智能 Agent 开发与运行、检索增强生成、智能应用管理引擎、智能组装与集成、统一可观测、多模态交互、AI 内生安全等）的成熟，使得 Agent 的开发效率提升数倍以上，支撑智能化应用的多模交互、自主运行、智能进化、环境理解、智能协作、智能研发等特征落地，形成更广泛、更灵活的智能体系统，从而能够在降本增效、体验优化等各方面得到大幅提升。

目前，AI 在医疗、科研、教育、物流等各个领域的应用，为社会带来了巨大价值的同时，又进一步支持了 AI 技术的快速发展。AI 技术的应用离不开社会的宽容和鼓励，我们要继续保持开放和包容的态度；同时关注风险，“汽车上路先装上刹车”，遵循道德和法律规范。

当今，应用现代化的推进目标已从“数字化 IT 提效”提升至“重塑智能商业价值”。未来，企业需以“应用现代化的三驾马车”，即：“敏捷架构、智能数据、生态开放”为基石，建立支持 Agent 自治与进化的全生命周期的数字基座。应用现代化产业联盟的各成员，要加强智能体应用、跨域协作标准等前沿议题的探索研究，以《应用全生命周期智能化研究报告》为起点，共同推动中国在全球智能化竞争中占据制高点，携手迈向应用现代化的智能时代！

陈纯
浙江大学信息学部主任
中国工程院院士
应用现代化专业委员会主任

序言二

从 2022 年底 ChatGPT “横空出世”以来，国内外大模型创新风起云涌，学术界、科技企业和无数开发者探索数十年的 AI 技术，迎来了应用大规模落地的奇点时刻。以应用智能体为代表的 AI 新范式，已成为推动行业变革的重要力量，智能化必将成为驱动数字经济发展的核心引擎。

过去十年，云原生重构了 IT 基础设施；未来十年，AI 原生将重新定义企业运营范式。这场变革不是简单的技术迭代，而是认知维度的跃迁——企业需要从“功能驱动”转向“价值创造”，从“流程优化”升级为“智能决策”。

华为云始终践行“一切皆服务”战略，致力于将最前沿的技术转化为企业触手可及的服务能力。在应用现代化领域，我们聚焦六大核心技术：组装式交付、数智驱动、DevOps、服务化架构、安全可信、韧性。这些能力已在金融、制造、能源等行业得到充分验证，帮助客户构建面向未来的智能化应用体系。

大模型和智能体技术的发展，让应用的智能化从愿景走向现实。这不仅是技术架构的升级，更是产业生态的重构。令人振奋的是，越来越多国内企业在技术变革中展现出强大的创新能力。《应用全生命周期智能化研究报告》的发布，正是产业界集体智慧的结晶。这本书系统梳理了技术趋势、架构设计和实施路径，将为行业的智能化转型提供重要参考。

面对千行万业的智能化浪潮，华为云将持续夯实“云原生 +AI 原生”的融合底座，让智能化应用深度融入企业核心业务流。我们将长期坚持深耕根技术，在 AI、云计算等基础领域持续突破；坚持开放协作，与伙伴共建智能化生态；聚焦客户价值，让技术创新真正服务于业务发展。

感谢应用现代化产业联盟所有企业和专家的共同努力，为业界探索出面向 AI 时代的创新之路。让我们携手加快应用现代化，释放数字经济的无限潜能。

张平安
华为常务董事、华为云 CEO

目录

1 应用智能化发展趋势	01
1.1 大模型技术加速全球产业价值链重构	02
1.2 数字基础设施逐渐成为国际竞争制高点	03
1.3 数据联接正在影响行业转型和应用安全	04
1.4 人机协作成为社会融合发展的主流趋势	04
1.5 客户群体转型加速驱动应用开发新范式	05
2 智能应用企业端和个人端双爆发，市场空间巨大	07
2.1 智能应用双端爆发，孕育万亿市场空间	08
2.2 智能体在 B 端和 C 端加速落地	08
2.3 央国企正成为应用智能化的主力军，领跑多个场景落地	09
2.4 智能研发、运维、知识管理，企业智能化落地三大高价值场景	11
3 智能应用的形态和核心特征	12
3.1 基于大语言模型的智能体（Agent）预计成为智能化应用的主流形态	13
3.1.1 大模型能力的突破性提升，为 Agent 提供智能化基座	13
3.1.2 Agent 从单一工具向自主协作系统演进	13
3.1.3 行业需求从效率优化到体验升级，Agent 成为优先选择	13
3.1.4 Agent 技术生态的成熟支撑	14
3.2 智能应用的核心技术架构概览	14
3.3 智能应用的六大核心特征	15
3.3.1 特征 1：多模交互	16
3.3.2 特征 2：智能协作	16
3.3.3 特征 3：自主运行	17
3.3.4 特征 4：环境认知	17
3.3.5 特征 5：自动进化	18
3.3.6 特征 6：智能研发	18
4 智能应用平台，应用智能化的技术底座	19
4.1 云原生与 AI 原生双擎驱动，加速应用全面智能化	20

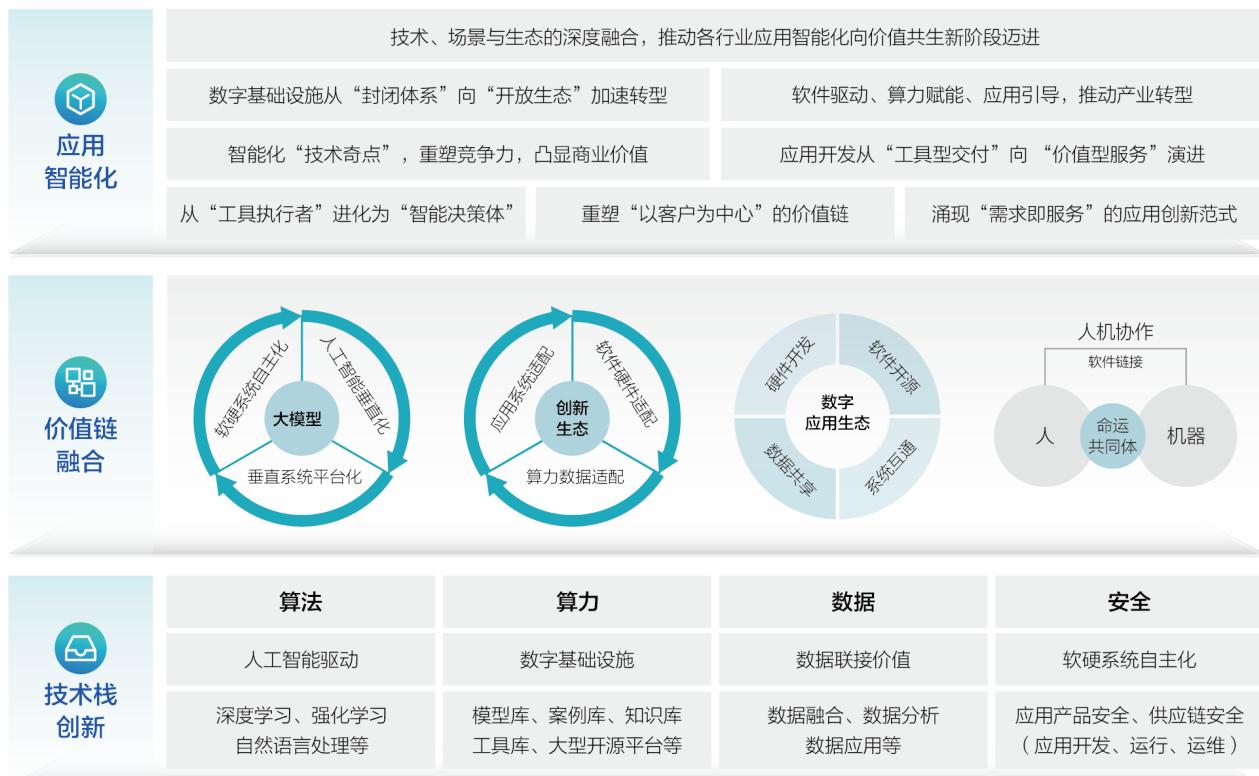
4.2 智能应用平台的核心定位：应用全生命周期的“智能驾驶”	21
4.3 智能应用平台的八大关键技术	24
4.3.1 多模态融合技术，实现多模态智能	24
4.3.2 智能 Agent 开发与运行，实现群体智能	26
4.3.3 检索增强生成，增强智能体个性化记忆	27
4.3.4 智能应用管理引擎，让应用自动运行与自主优化	28
4.3.5 智能组装与集成，扩展应用能力边界	29
4.3.6 统一的智能体可观测，支撑海量智能应用高效运维	31
4.3.7 多模态交互，建立双向人机协同体验	32
4.3.8 AI 内生安全，确保应用全生命周期可信	34
5 应用智能化的实现路径与关键考量点	35
5.1 应用智能化实现路径	36
5.1.1 应用智能化成熟度技术演进路径	36
5.1.2 应用智能化的 6R + 9A 实施路径	38
5.2 应用智能化的关键考量点	41
5.2.1 战略与业务价值对齐	41
5.2.2 用户体验重构	41
5.2.3 组织流程适配	42
5.2.4 渐进式演进策略	42
5.2.5 法律法规与伦理合规	42
6 应用智能化最佳实践案例	43
6.1 流程 IT 智能化应用升级	44
6.2 某智慧家庭解决方案提供商提升运维管理效率，实现资源优化	45
6.3 社区平台打造训推一体化平台	45
6.4 CodeGenie：聚焦鸿蒙应用开发场景的智能化研发助手	47
6.5 智驾 AI 算力平台，保障智驾业务	48
6.6 新疆某油田基于大模型构建统一知识库	49
6.7 云原生技术赋能香港智慧医疗，构建高效服务新生态	49
6.8 软通动力携手同济大学打造云原生研一体化教务系统	50

01

应用智能化发展趋势

PART ONE

应用现代化与智能化的深度融合，正通过开放生态、软硬协同、垂直平台、人机协作与人工智能服务化等特征重构技术、产业与社会生产生活方式。从政务服务到社会治理，从智能制造到智慧生活，从算力基建到数据价值释放等等，应用智能化不仅加速了行业变革，更催生出了“技术 - 场景 - 生态”三位一体的创新范式，为数字经济高质量发展注入持续动能。



1.1 大模型技术加速全球产业链重构

人工智能技术的突破性进展正在推动应用现代化工程理论进入新一轮创新周期。以大模型为核心的技术范式变革，正在衍生出“人工智能垂直化、垂直系统平台化、软硬系统自主化”发展新趋势，加速重构全球产业链。大模型在垂直领域的深度应用，正在重塑金融风控、智能制造、自动驾驶、医疗诊断等各行业领域的工作模式。系统平台化则促进提升资源整合效率，工业互联网平台、城市大脑等载体将分散的数据、算力、算法资源池化，实现按需调用。

软硬系统自主化是安全可控的基石，其核心在于构建从芯片设计到操作系统的全链路自主可控技术栈，涵盖硬件（如芯片设计、制造）和根技术软件（如操作系统、数据库、中间件、IDE、开发工具链等）的自主研发。通过形成完

整的技术生态闭环，确保技术独立性与完整性，为关键领域提供安全、稳定的技术支撑，保障应用智能化发展不受外部因素制约。软硬协同结合自主可控的根技术软件，将促进关键领域的应用智能化发展，提升产业韧性，抵御外部供应链风险，为应用智能化生态的长期安全稳定保驾护航。

推动人工智能蓬勃发展的前提是应用现代化的发展成熟和壮大。未来，随着深度学习、强化学习、自然语言处理等人工智能技术的不断迭代创新发展，AI 将在更多垂直领域实现深度应用，不仅将提升生产效率和服务质量，还将创造新的商业模式和市场机会，使人工智能成为推动社会进步的重要力量。应用现代化走向智能化发展也将迎来“技术奇点”，爆发出更大的潜力，重塑竞争力，持续提升商业价值，推动应用全生命周期智能化从效率革命迈向价值共生新阶段。

1.2 数字基础设施逐渐成为国际竞争制高点

随着数字化转型步入深水期，模型库、案例库、知识库、工具库、大型开源平台等数字基础设施正在成为智能时代的关键战略资源。数字基础设施不仅是技术创新与产业升级的“数字公地¹”，更成为各国争夺未来竞争力的关键抓手。硬件开放与软件开源的协同演进，通过降低技术准入门槛、激发市场创新活力，推动数字基础设施从“封闭体系”向“开放生态”加速转型。

软硬适配、算力数据适配、应用系统适配正在重塑全球数字竞争的底层逻辑。软硬适配通过深度整合软件性能与硬件功能，达到软件与硬件的协同优化，实现以软补硬、以硬促软的竞争优势格局，重构智能时代的系统级创新生态，推动普惠算力时代到来，促使中小企业与开发者以更低成本获取高性能计算资源，加速创新应用的开发。算力数据适配通过动态分配计算资源、优化数据流动路径，实现算力、算法与数据的高效协同。随着异构计算与智能调度技术的突破，算力正成为千行百业数字化转型的核心驱动力，催生更多高精度决策支持系统，赋能产业智能化升级。应用系统适配通过标准化接口与协议，将实现跨平台、跨领域的数字系统互联互通，加速产业互联网向消费互联网的价值渗透，催生个性化、场景化的智能服务新业态，重塑数字时代的产业价值网络。

随着云计算、边缘计算、5G 等技术的快速发展，数字基础设施将更加完善，为智能应用的部署和运行提供强大的支持。未来，数字基础设施将不仅仅是技术创新的基石，更将成为国家竞争力的重要标志。各国将加大在数字基础设施领域的投入，通过建设高速网络、数据中心、智能传感器等，打造更加开放、协同、安全的数字生态。掌握先进适配能力的国家，将在全球数字经济格局中占据领先地位，主导未来十年的创新和发展方向。

1. 数字公地（Digital Commons）指基于互联网与数字技术形成的公共资源形态，其核心是通过共享基础设施、数据要素和知识资源，实现社会效益最大化。



1.3 数据联接正在影响行业转型和应用安全

数据联接将深刻影响各行业的转型路径与应用安全的保障体系。进入数字化时代，数据以其独特属性成为继土地、劳动力、资本、技术之后驱动社会进步与经济发展的第五大核心生产力要素。数据联接不仅加速了应用全生命周期智能化的进程，更实现了跨领域的数据融合、分析及应用。**硬件开放、软件开源、系统互通、数据共享**将构建新的数字化应用生态。硬件开放将促进数据源的多样性、开放性、统一性及互联性，支持更广泛的数据收集和共享，为构建开放协同的数据生态奠定基础。软件开源则推动知识共享与模块化复用，促进开发者协作和加速技术迭代，降低开发成本，两者融合将催生出更丰富的数据应用场景，持续释放数据价值，加速知识向生产力的转化。算力与数据匹配将提高处理效率，系统互通和数据共享则打破数据孤岛，推动数据联接的深化。数据联接不仅促进商业创新，还将深刻影响政府决策、公共服务及社会治理的方方面面。随着新一代信息技术的发展进步，数据联接不同领域、不同行业的能力日益增强，这种跨界融合正在重塑数字经济时代的核心竞争力，推动行业转型与产业升级。

值得注意的是，随着应用智能化在社会各领域的深度发展，应用产品和供应链安全问题也日益成为不容忽视的挑战。各个行业对智能应用产品的依赖性显著增强，从智慧医疗、智能制造以及智慧城市等领域，核心技术栈的自主可控性直接关乎产业安全命脉。针对智能应用开发、运行、运维全生命周期的攻击手段正日益复杂和隐蔽，给安全防护带来了新的挑战。供应链安全威胁尤为突出，软件供应链攻击已能穿透传统防御体系，直接影响国家关键基础设施的安全。开源生态的漏洞、第三方组件的不可控性，更放大了系统性风险。

随着数据量的爆炸性增长和数据联接技术的不断发展，数据将成为推动行业转型和升级的关键要素。未来，数据将在更多领域实现跨界融合，推动智能化升级和创新发展。然而，随着数据联接的深化，智能应用产品安全和供应链安全问题也将日益凸显。政府、企业和个人将更加注重数据安全和隐私保护，推动相关法律法规的完善和执行。

1.4 人机协作成为社会融合发展的主流趋势

随着人工智能技术的快速迭代发展，人与机器的界限正在日益消融，二者在多元场景的深度交互与融合过程中，将逐步构建起人机协同进化的命运共同体，标志着技术层面的革新，更预示着社会结构与生产方式的深刻变革。软件作为人机链接的枢纽，不仅实现物理空间与数字世界的无缝对接，更赋予机器深度感知、自主决策与协同共创的类人化能力。

单机系统智能化、智能系统群体化、人机交互语言化、复杂系统模块化正在人机协作新范式。硬件革新与算法突破推动单机系统从“工具执行者”进化为“智能决策体”，能够自主感知环境，分析数据，做出决策，甚至进行自我优化和升级。物联网与5G技术催生智能系统的群体化演进，形成协同工作的群体，共同完成复杂的任务，实现资源的优化配置和高效利用。自然语言处理与多模态交互技术的突破，则加速重塑人机协作的认知界面，使得人机协作更加便捷和高效。复杂系统向“组装式”模块化架构演进则为人机协作提供了灵活可扩展的载体，不同的功能模块可以独立开发、测试和部署，然后根据需要进行组合和配置，以适应不同的应用场景和需求。

未来，随着智能技术的持续进步和应用场景的裂变拓展，人机协作理念将贯穿应用智能化发展的全生命周期，不仅将提升生产效率和服务质量，还将改变人们的生活方式和工作模式。战略性新兴产业和未来产业新赛道的涌现也将不断拓展人机协作的边界，催生“人机协同设计”“混合增强智能”等创新模式，同时，也将激发商业模式重构与产业生态变革。企业将更加注重人机协作能力建设，通过引入智能技术，提升生产效率和服务质量，创造新的商业价值。政府和社会各界也将加强对人机协作的引导和支持，推动相关法规和标准的制定，保障人机协作的健康发展，推动社会生产力的全面提升。

1.5 客户群体转型加速驱动应用开发新范式

客户需求的变化正在驱动应用现代化向应用智能化转型，促使应用开发从传统的“工具型交付”向更具前瞻性的“价值型服务”演进。企业客户群体日益加速的数字化转型需求和个人消费群体的需求改变，迫使智能化应用从单一功能型产品向深度定制化、场景化解决方案迁移。应用智能化发展不仅将重塑“以客户为中心”的价值链，更将服务体验从单纯的功能满足提升至企业战略赋能的新高度。新一代企业客户正逐渐将智能化应用视为数字化转型的核心载体，对开发范式提出了更高的要求，从流程固化转向组织赋能，从单纯的技术实现转向业务共生，实现技术与业务的深度融合。

技术革新是推动应用开发新范式的重要力量，催生了三大核心驱动力。一是算力资源的水电化供给。随着云计算、边缘计算等技术的普及，算力资源水电化供给成为企业可随时获取的基础设施，解锁场景创新的无限潜能，使开发者能够更加聚焦业务逻辑，而非性能调优，从而加速创新步伐。二是软件工程范式的延伸与重塑。DevOps范式逐渐向DataOps、MLOps等领域延伸，不仅重塑了软件开发模式，更推动了产业向“小步快跑、快速迭代”的敏捷智能模式转型，使得企业能够更快地响应市场变化，满足客户需求。三是智能化应用形成迭代闭环。通过A/B测试、用户旅程地图²、特征工程等工具，智能化应用构建了“需求感知-价值设计-效果验证”的迭代闭环，使应用成为企业业务创新

2. 用户旅程地图是可视化的用户流程，展示用户在使用产品时的各个阶段和接触点，帮助开发者理解用户需求和行为。



的探测器与加速器，不断推动着业务向前发展。

在金融服务、智能制造、智慧城市、医疗健康、教育科技以及零售消费等越来越多的标准化行业领域，创新应用正涌现“需求即服务”的新范式。通过创造数字孪生决策、智能供应链协同、沉浸式教育等场景，智能化应用正在重塑产业认知边界，推动千行百业向更高层次发展。

未来，随着企业级客户和个人消费级客户数字化转型需求的加速提升，应用开发新范式将成为推动产业发展的力量。客户群体应用开发将更加注重业务流、数据流与智能流的深度融合，通过云原生架构、低代码平台、智能运维等技术手段，构建“感知 - 决策 - 行动”的闭环服务体系，使客户能够更精准地感知市场变化，更快速地做出决策，更有效地执行行动，从而在激烈的市场竞争中脱颖而出。在**软件驱动、算力推动、应用引导**的协同作用下，客户群体正在开启智能化应用开发服务的新纪元，推动产业向服务化、平台化、生态化方向深度进化，显著提升行业客户的核竞争力，为客户带来更加广阔的机会和竞争优势，更将推动社会各行业的转型升级。

总之，在人工智能浪潮席卷全球的当下，社会各行各业正迎来前所未有的变革与机遇。要推动人工智能科技创新与应用智能化深度融合，构建企业主导的产学研用协同创新体系，以更加开放的姿态拥抱人工智能技术变革，通过应用全生命周期智能化建设重构核心竞争力。为进一步释放智能化应用变革带来的红利，统筹推进算力基础设施建设，深化数据资源开发利用和开放共享。同时加强关键技术研发投入，推动产业协同创新，构建跨领域人才梯队，完善法律与标准体系以及广泛开展国际合作等共同驱动产业转型升级，形成技术、场景与生态的深度融合，推动各行业应用智能化向价值共生新阶段迈进。

02

智能应用企业端和个人端双爆发， 市场空间巨大

PART TWO



2.1 智能应用双端爆发，孕育万亿市场空间

根据 Market Research Future 的《Applied AI Market Report》报告³，智能应用市场预计将从 2025 年的 2251 亿美元跃升到 2034 年的 1.7 万亿美元，年均复合年增长率（CAGR）高达 25.32%。这场由 AI 驱动的变革已经深入社会肌理。智能应用从工具进化为新生产力引擎，其爆发性增长不仅体现于数字跃迁，更标志着人机协同范式在企业运营与个人生活的全面确立。

在个人端，开启了创造力普惠的时代，文本生成、图像创作、视频制作等 AI 工具已深度融入大众生活，普通人仅需简单指令即可释放专业级创作能力。这种“零门槛智能化”正深刻改变内容生产、知识获取与娱乐消费的底层逻辑。在 2025 年 4 月的“AI 产品榜”⁴，国产 AI 应用在全球前十中占据半壁江山，表现亮眼，从这个榜单的数据增长方面来看，全球 AI 应用 TOP100 中有 68% 的应用实现了增长，而国产 AI 应用在其中占据了显著位置，如夸克、豆包、DeepSeek、腾讯元宝、Talkie 等。

在企业端，全链路的智能变革已经悄然启动，从研发创新、精准营销到智能制造、智慧供应链，智能应用正在重构企业核心价值链。通过实时数据分析与自动化决策，企业运营效率获得突破性提升，传统管理范式迎来颠覆性升级。根据 Menlo Ventures 2024 年《2024: The State of Generative AI in the Enterprise》报告⁵，企业在行业垂直场景的智能应用以及部门级智能应用的支出金额的年增长分别达到了 12 倍和 9 倍，增速已经超过数据 + 基础设施领域和基础模型领域。

2.2 智能体在 B 端和 C 端加速落地

作为智能应用的典型代表，智能体市场迎来爆发式增长。根据 Precedence Research 统计⁶，2024 年全球 AI 智能体市场规模达到了 54 亿美元。预计这一市场将持续高速增长，于 2034 年超过 2300 亿美元。10 年间市场将以 45.82% 的复合年增长率持续扩张。

3.<https://www.marketresearchfuture.com/reports/applied-ai-market-12221>

4.<https://www.aicpb.com/>

5.<https://menlovc.com/2024-the-state-of-generative-ai-in-the-enterprise/>

6.<https://www.precedenceresearch.com/ai-agents-market>

同时根据华经产业研究院统计⁷，中国AI智能体市场规模在2023年达到59.8亿元，在2024年已经超过100亿元，预计市场规模将持续高速增长。

Gartner在《Top 10 Strategic Technology Trends for 2025》报告⁸，将Agentic AI确定为将在未来2年内重塑企业技术格局的最核心技术之一，文中预测到2028年，33%的企业应用将融入智能体，而在2024年这一比例还不到1%。到2028年，20%的电商交互将由智能体提供，15%的日常决策将由智能体自主做出，这将从根本上重塑决策过程。

另外根据Grand View Research报告，预计到2025年底，大约85%的企业将采用智能体，以提高效率和客户满意度。在零售行业，76%的零售商将在未来一年增加对智能体的投资，重点用于客户服务应用，如自动查询和订单跟踪。大型咨询公司正在把智能体能力整合到运营流程中。某咨询公司部署的150个税务智能体将协助合规和数据审查流程。

2.3 央国企正成为应用智能化的主力军，领跑多个场景落地

据中国信通院统计，2024至2025年大模型相关项目持续攀升，从需求方来看，2024年全年近6成数量的采购方为大型央国企，以通信、金融和能源三大行业为主，中标项目金额占总金额超一半⁵。所需的大模型相关项目内容和场景主要分为平台、应用、算力、大模型、数据五个类型，其中平台和应用项目数约占总数的一半以上，但算力项目金额最高。知识问答、智能分析决策、数字人客服、智能运维、智能编程和内容生成六大应用场景的项目数量最多。

从供应方来看，应用交付成核心赛道。据中国信通院2025年Q1统计，超半数供应商聚焦“应用交付”，另外工程化平台是大模型落地关键，细分类型包括大模型应用开发、大模型调优、大模型API服务及模型资产管理等平台，助力企业降低开发成本。行业落地聚焦五大领域，场景应用百花齐放。政务、能源、工业、教育、医疗成为大模型落地前五大行业，占比达57%⁵。政务领域以智能文档与咨询服务为主，能源行业侧重设备维护与研发设计，教育行业则深耕教学辅助与管理优化。

在已落地的实践案例中，排名居前五位的行业领域依次为政务、能源、工业、教育及医疗。从主要行业落地场景进行分析，在政务行业，咨询服务和智能文档是位列前二的AI应用场景。能源行业则更关注设备维护和设计研发场景，教育行业的落地场景主要集中在教育管理和教学辅助两大场景。

7.<https://caifuhao.eastmoney.com/news/20241217103730207581820>

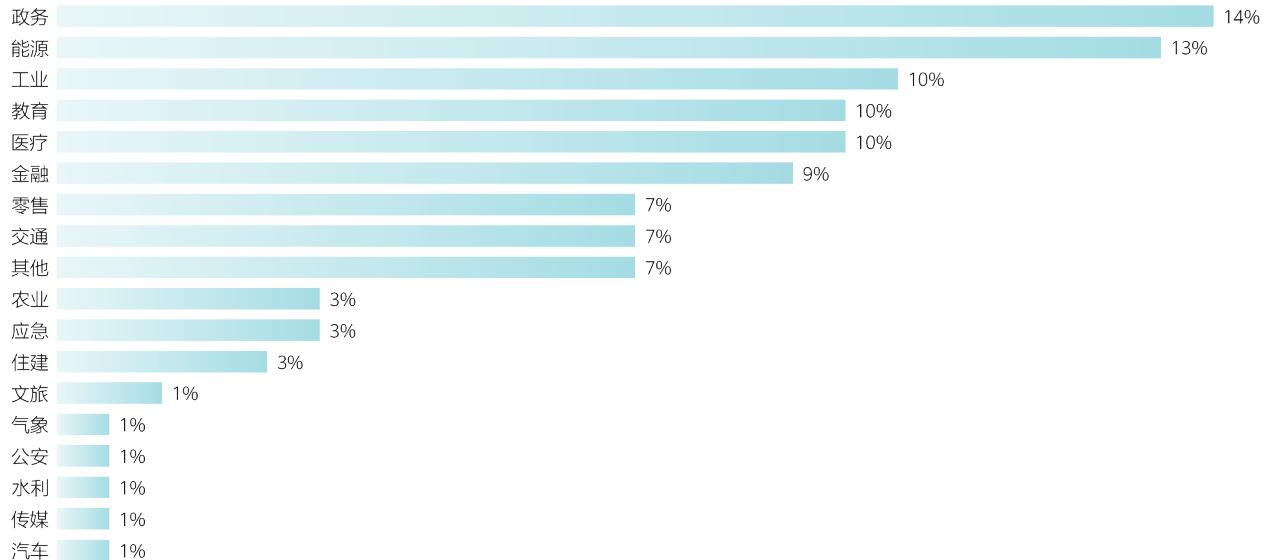
8.<https://www.gartner.com/en/articles/top-technology-trends-2025>



大模型应用交付供应商落地行业及场景分析

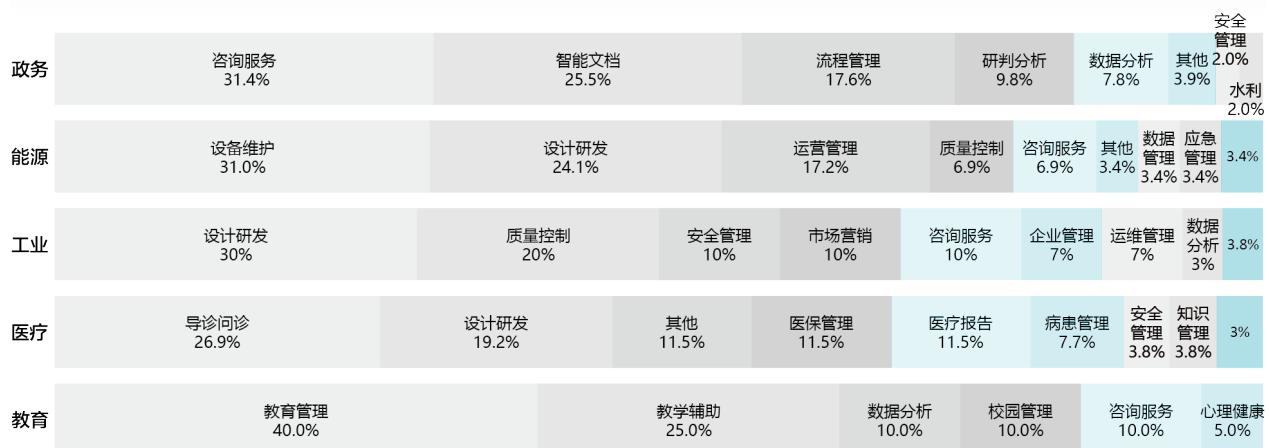
行业分布

从行业来看，落地案例中，排名前五的行业分别是**政务、能源、工业、教育、医疗**，共占整体的**57%**。



Top 5行业内场景分布

不同行业中的落地场景各有侧重，以TOP5行业为例，**政务行业咨询服务和智能文档位列前二占比超过50%**。能源行业则是**设备维护和设计研发**位列前二，而教育行业的落地场景主要集中在**教育管理和教学辅助**两大场景，合计超60%。



2.4 智能研发、运维、知识管理，企业智能化落地三大高价值场景

企业智能化转型正聚焦可量化、快回报的核心场景，头部企业实践表明：智能辅助开发、AIOps 运维与企业级知识管理已成为落地成效最显著的三大领域，其共性在于直击业务痛点、深度嵌入主流程、ROI 周期短（6-12 个月）。

智能辅助开发：根据 Verified Market Research 的《AI Code Tools Market Size And Forecast》报告⁹，智能辅助开发场景是最多受访企业使用的场景，全球 AI 辅助开发工具市场规模在 2024 年达到 49.1 亿美元，并将以 27% 的年复合增长率持续增长，预计在 2032 年市场规模达到 300 亿美元。这类工具通过自然语言描述快速生成代码片段或者完整脚本，显著提高开发效率。某股份制银行基于大模型平台构建了代码生成研发辅助工具体系，代码生成采纳率和代码注释率显著提升，应用迭代周期大幅缩短。

智能运维（AIOps）：根据 Fortune Business Insights 的《AIOps Market Size, Share & Industry Analysis》报告¹⁰，2024 年全球智能运维 AIOps 市场规模约 18.7 亿美元，预计到 2032 年达到 86.4 亿美元，年复合增长率 21.4%。AIOps 借助人工智能技术，能实现故障预测、智能告警、自动化运维等功能，提升运维效率和精准度。某智能电网项目采用 AIOps 系统，实时监控设备状态并自动化修复故障，从而避免大范围停电。

企业知识管理智能化：通过 AI 深度解析技术文档关键参数与应用场景，实现对企业知识库核心信息的精准定位与智能提取，彻底打破信息孤岛。某制造集团实践表明：构建基于 NLP 语义理解的研发知识地图与多级分类知识仓库后，系统可支持「设计参数 + 失效模式 + 应用场景」等多维度组合检索，使关键技术定位效率提升 8 倍，重复设计率下降 25%，推动企业知识资产从被动存储转向主动赋能创新。

当然，各个行业的落地场景和节奏不尽相同，但是初步可以判断企业选择优先落地场景的底层逻辑：大部分企业优先落地见效快（6-12 个月 ROI），技术成熟度相对高的领域，如上三个典型场景的共同特征是：直击痛点、切入主流程、价值可量化。

9.<https://www.verifiedmarketresearch.com/product/ai-code-tool-market/>

10.<https://www.fortunebusinessinsights.com/aiops-market-109984>



03

智能应用的形态和核心特征

PART THREE

3.1 基于大语言模型的智能体（Agent）预计成为智能化应用的主流形态

随着大模型能力的突破性提升、Agent 技术的快速演进、开发技术生态的成熟支撑等因素，预计基于大语言模型（LLM）及智能体（Agent）的 AI 原生智能应用架构将成为未来应用的主流形态。支撑上述洞察的核心要素如下：

» 3.1.1 大模型能力的突破性提升，为 Agent 提供智能化基座

自然语言理解与生成能力的跃迁：LLM 通过海量数据训练和参数规模的扩大，具备了更强的上下文理解、推理和生成能力。GPT-4 在自然语言理解及生成能力上的突破，使得 Agent 能够更自然地进行人机交互。如医疗领域的 Agent 可根据患者症状快速生成诊断建议，教育领域的 Agent 能提供个性化学习路径规划。

函数调用与工具使用能力的增强：大模型的函数调用功能使其能够直接调用外部工具或服务，例如调用 API 接口完成订单处理等。这种能力显著提升了 Agent 的实用性、能力范围及自动化水平。如电商场景中，Agent 可自动调用支付、物流等模块，实现端到端的订单处理。

多模态融合与长上下文处理：近期大模型在图像理解、视觉计算、语义理解等维度持续突破，同时长文本处理能力也大幅提升，这一系列的演进使得 Agent 能够结合图像、语音、视频等多模态信息，处理更复杂的场景需求。

» 3.1.2 Agent 从单一工具向自主协作系统演进

自主性与规划能力的提升：现代 Agent 架构通过引入“规划层”（如 MetaGPT 的任务分解机制）和长期记忆模块（如向量数据库、知识库等），支持自主制定目标、分解任务并动态结合已有经验知识，多轮迭代式自主完成复杂任务。

多 Agent 协作的成熟：复杂业务场景中，多 Agent 协作通过任务分配、信息共享和动态协调，显著提升了处理效率。例如，在智能制造领域，多个 Agent 可分别负责生产调度、质量监控和资源优化，协同完成全流程管理。

» 3.1.3 行业需求从效率优化到体验升级，Agent 成为优先选择

企业端：复杂流程智能响应与降本增效

传统 B 端软件难以应对动态业务需求，而 AI Agent 通过自动化流程（如供应链管理、客户服务）和智能决策（如金融风险评估），显著降低人力成本并提升效率。

用户端：个性化与拟人化体验

Agent 能够基于用户历史行为动态调整服务模式，例如个性化推荐、情感陪伴等。多模态交互（如语音助手、虚拟形象）进一步增强了用户体验的沉浸感。

新兴场景的拓展

从智能制造到智慧城市，Agent 在复杂环境中的适应性使其成为数字化转型的核心组件。例如，自动驾驶领域的大模型赋能车辆实现更智能的感知与决策。

» 3.1.4 Agent 技术生态的成熟支撑

开发工具链：如 OpenAI Agents SDK、Dify 零代码 Agent 开发平台、LangChain/LangGraph 等 Agent 框架，极大降低了 Agent 开发门槛。

架构设计的模块化与可扩展性：结合 Kubernetes 和 Dapr 技术，实现了微服务化部署、弹性扩缩容和服务间异步通信，为 Agent 系统的稳定性和可维护性提供了技术保障。

开源社区与行业标准的形成：AutoGPT、MetaGPT 等开源项目加速了 Agent 技术的普及，同时行业逐渐形成关于数据隐私、模型可解释性的规范，推动技术健康落地。

3.2 智能应用的核心技术架构概览

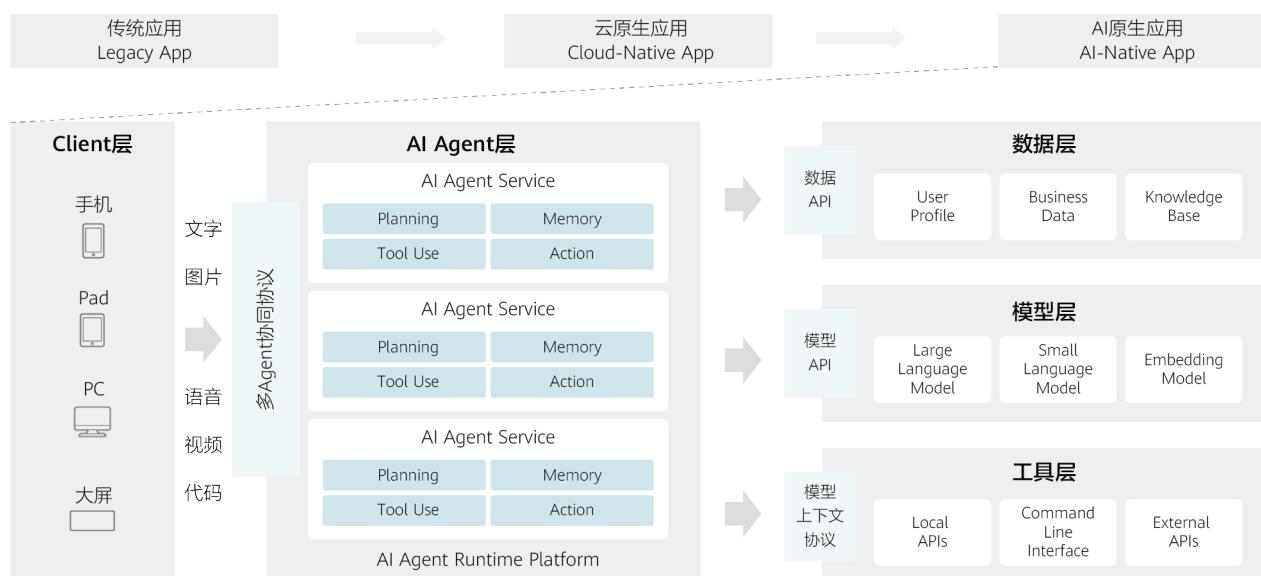


图 1：应用形态变化趋势概览

基于大语言模型及智能体的 AI 原生应用架构，已成为未来技术演进的必然方向。图 1 提供了应用形态变化趋势概览，未来 10 年主流应用的架构将逐步从云原生应用技术栈过渡到基于 LLM 与 AI Agent 的 AI 原生应用技术栈。

数据层：提供多模态数据支持，包括结构化数据（数据库、表格）、非结构化数据（文本、图像）、领域知识（如科研文献、行业标准）和支持动态更新的外部数据（如专利、新闻）。数据层可为模型层提供训练与推理所需的原始数据，并通过向量化技术增强 Agent 的上下文理解能力。

模型层：构建智能化基座，涵盖通用大模型与垂直领域模型，实现模型的高效部署与推理。

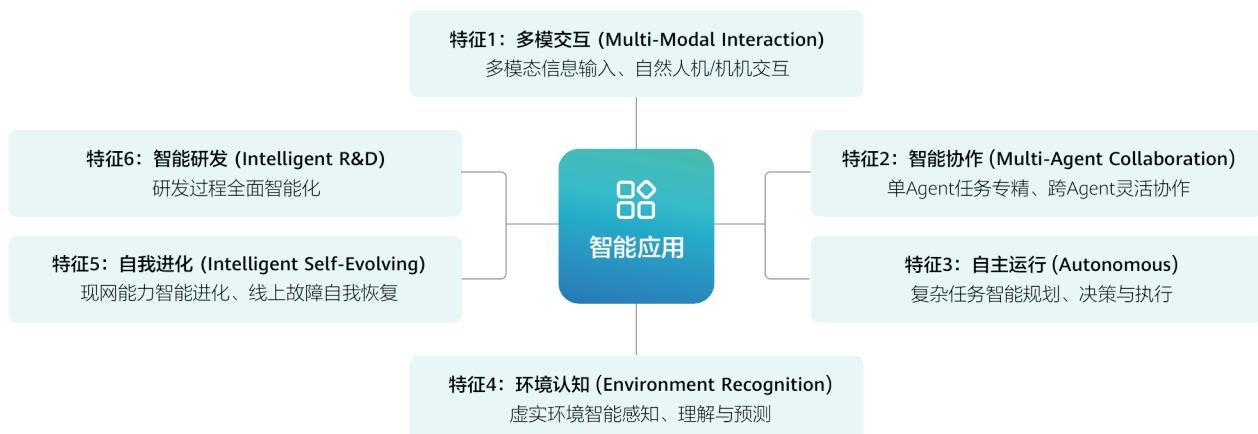
工具层：将本地可调用命令及外部 API 作为 tools 进行集中管理，并通过 MCP 协议面向 LLM 提供工具 API 访问调用，支持为 LLM 提供不同类型的 Context 扩展。

AI Agent 层：驱动任务规划与自动化执行，实现从感知到行动的闭环。其关键能力包括：

- 规划模块：分解复杂任务（如 MetaGPT 的任务拆解机制），生成多步骤执行计划。
- 记忆模块：结合短期记忆（对话上下文）与长期记忆（向量数据库），支持动态信息检索与更新。
- 工具使用：支持调用外部工具或命令灵活获取上下文，支持模型更好地生成表现。
- 行动模块：调用外部工具（如 API、代码执行环境）完成操作，支持多模态输出（如语音合成、图表生成）。

Client 层：为外部用户提供更加自然的交互流程，支持多模态信息的输入输出，充分发挥后台 agent 智能水平，满足业务场景诉求，极大提升用户体验。

3.3 智能应用的六大核心特征





未来基于大模型与 Agent 架构的智能应用软件，将重新定义人机交互与业务运行的方式，展现出六大核心特征。特征 1- 多模交互：这类应用具备多模态环境感知能力，能够同时理解语音、图像、文本、手势甚至生理信号等多种输入形式，并以自然、流畅的方式进行反馈，显著提升用户体验的沉浸感与智能化水平。特征 2- 自主运行：智能应用能基于复杂任务进行智能规划与推理，自动分解目标、调度资源并执行操作，在多变环境中实现高效决策和自适应响应，减少人工干预。特征 3- 智能进化：系统不仅具备业务能力的动态扩展能力，还可通过线上学习持续优化自身性能，甚至在出现故障或异常时实现自我诊断与修复，确保长期稳定运行。特征 4- 环境理解：应用能够实时感知和理解其运行环境，包括用户状态、业务上下文、物理空间乃至社会趋势，并通过预测机制提前做出调整，增强系统的主动性和前瞻性。特征 5- 智能协作：多个 Agent 能够根据角色分工各司其职，彼此之间灵活沟通、协同决策，形成群体智能，从而完成更复杂、更大规模的任务。特征 6- 智能研发：从需求分析、代码生成、测试验证到部署运维，整个研发流程都将被全面智能化重构，大幅降低开发门槛，显著提升研发效率与质量，推动应用快速迭代与持续交付。这六大特征共同构建了面向未来的 AI 原生应用新范式，标志着智能软件进入一个高度自主、协同与进化的新时代。

» 3.3.1 特征 1：多模交互

随着人机交互技术的快速演进和用户需求的日益多样化，多模交互（Multi-Modal Interaction）正在成为新一代智能应用的核心特征之一。

视觉图像分析、语言语义理解、姿态运动感知等领域技术的快速发展与融合，使得系统能够同时接收并理解语音、图像、GUI 界面、文本、手势、生理信号等多模态信息。这些技术的成熟将为应用提供更加丰富的交互形式：从之前的“单模态”信息交互演化为“多模态”信息交互。

多模交互通过灵活运用多种信息感知手段，可以更加智能地进行外部环境感知、用户意图识别。多模交互不仅可以极大提升人机交互过程的自然度与灵活性，还可以极大提升应用与外部环境、应用与应用之间的信息传递、感知与理解能力。这一特征变化，将极大拓宽未来应用在各行各业中的落地场景（如机器人采矿场景：智能感知矿井环境并开展智能安全地采矿活动）及用户范围（如面向残障人士）。

» 3.3.2 特征 2：智能协作

智能协作（Intelligent Collaboration）是智能应用核心特征之一：通过多智能体（Multi-Agent）之间，以及人机之间的互联互通与角色分工，应用可以在复杂、多变的环境中实现动态调度、决策与执行，拓展能力边界，为业务带来更高层次的灵活性与创造力。

当前智能协作依托于多智能体系统架构及其在交互和学习机制，包括：

1. 自然语言驱动的多模态协同：借助语言、视觉等多模态信息进行交流与任务分配，使不同智能体之间，以及智能体与人类之间能够以更自然的方式共享上下文和意图。
2. 人机联合决策：将机器计算能力与人类专家经验相结合，在关键决策环节实现优势互补。
3. 知识建模与共享：通过构建统一的知识表示和交换机制保障协作者的信息一致，避免知识孤岛并支持持续学习改进。

4. 联邦学习：在不共享原始数据的前提下，各节点或机构共同训练模型，实现数据隐私保护与群体智能同步提升等关键技术的融合与应用。

智能协作在众多领域已展现出显著价值。例如，多智能体系统可分工完成代码安全审计、自动化测试与缺陷修复，并与开发团队实时联动，持续在复杂的分布式协作环境中实现高效迭代与交付。尽管目前在数据合规、跨平台标准和算法可解释性等方面仍存在挑战，但随着技术迭代与生态体系的持续完善，智能协作将在未来3-5年内加速规模化落地，成为推动产业升级与社会创新的关键引擎。

» 3.3.3 特征 3：自主运行

自主运行 (Autonomous Operation) 是指智能应用在符合业务目标和安全合规要求的前提下，具备高度的自主决策与自适应独立运行能力。它作为新一代智能应用最显著的特征，也标志着AI原生时代，人与机器在生产与服务方式上发生的颠覆性变革。

传统应用只能按照预先编排的流程或规则被动响应，当引入具备智能体 (Agent) 特性的组件后，应用可以通过实时感知环境、设定目标，并在无需人工干预的情况下制定行动方案并执行任务。同时，为了自主运行更充分地落地，还需在底层引入更加智能的运维和管理机制。例如，采用任务规划引擎、数据驱动决策以及无损变更等技术手段，可以对应用在不同阶段（从部署到运营）的状态进行全局优化。

随着模型规模、算法能力、计算资源和生态体系的不断成熟，智能应用所蕴含的各项能力特征将得到进一步释放，助力人类迈向下一阶段的创新浪潮，对于企业而言，如何安全、可控地落地自主运行能力，将是数字化战略的核心挑战之一，也蕴含着巨大的增长机遇。

» 3.3.4 特征 4：环境认知

环境认知 (Environment Recognition) 是指智能应用系统对其所处物理和数字环境进行实时感知、融合建模、趋势预测与策略动态调整的能力。在新一代智能应用中，环境理解能力至关重要，它让系统能够基于全局上下文自主决策，从而大幅提升运行质量、安全保障和用户体验。这一能力的实现得益于：

1. 多模态感知与语义理解等技术的突破，使系统可以融合来自视觉、语音、传感器等多源数据以深度感知环境。
2. 结合数字孪生建模对物理世界进行虚拟映射与精细仿真，利用边缘智能分担海量数据的实时处理，并通过时空预测模型对未来态势进行精准预判，确保应用对环境变化作出及时优化响应。典型应用包括：智慧园区系统根据人流密度预测能耗需求并结合天气预报提前启动供能装置，以平抑用能峰值；智能工厂通过传感器监测设备运行状态并运用高级预测模型提前识别最佳维护窗口，避免设备突发故障停机。

随着感知硬件迭代、边缘计算与云端服务协同增强以及大模型技术加速演进，环境认知将在未来发挥更关键作用。多模态感知与数字孪生的精细映射，使物理与数字世界无缝融合；边缘侧分布式实时推理与云端超大规模模型协作，可以更低时延、更高可靠性支撑系统决策；时空预测模型吸纳海量环境数据，助力各行业更精准地预判趋势并优化策略。



同时，新一代通信网络、数据治理与隐私计算的协同发展，将进一步拓展环境认知的应用边界，从工业制造、城市治理到出行管理，多元场景不断涌现。未来，智能应用不仅被动适应环境，还将通过深度交互与主动干预，与环境形成共生关系，全面提升产业协同与数字化升级，为各行各业创造可持续的高质量增长。

» 3.3.5 特征 5：自动进化

与传统依赖版本迭代实现局部优化的模式不同，自主进化（Self-Evolving）通过持续学习、动态适配和自我优化，为智能应用构建了更具弹性和自愈能力的技术框架。这一转变不仅可以显著提升系统的自主决策与故障响应效率，还可在复杂多变的业务场景下，帮助企业更敏捷地适应市场变化并保障服务质量。

从技术层面来看，一方面，完善的智能化运维体系能够为自动进化提供完整的数据循环支持，涵盖数据采集、模型训练、部署到持续迭代的全流程管理，让系统能基于环境反馈快速演化。另一方面，联邦学习、强化学习等前沿方法在隐私保护与复杂决策场景中大放异彩，为大规模、多场景下的智能进化提供了可行且安全的路径。在此基础上，结合有效的数据管道与模型治理机制，企业能够在保持合规与安全的前提下，为应用的自动进化奠定坚实基础。

在业务层面，具有自动进化能力的智能应用可通过实时更新用户画像与行为数据，实现更精准的内容推荐和个性化决策支持。在流量高峰或资源紧张时，能根据策略自动调整分配方案，从而最大化资源利用率与服务稳定性。此外，通过对运行数据的动态监测与学习，应用可提前识别潜在风险并进行预测性维护，降低突发停机与安全漏洞带来的损失。随着技术与生态的不断完善，企业若能持续完善自动化运维体系与数据治理规范，便可最大程度释放自动进化的潜力，不断攀升业务效率与创新能力，为应用全生命周期智能化奠定坚实基石。

» 3.3.6 特征 6：智能研发

智能研发（Intelligent R&D）是未来应用的核心特征之一，指通过人工智能技术重构软件研发全生命周期流程，包括需求分析、系统设计、代码开发、测试验证、部署交付与智能运维等环节。其核心价值体现在显著提升研发效率和软件质量，并增强研发敏捷性。

这一范式的转变依托于大模型驱动的代码生成、自然语言需求建模、自动测试生成、软件知识图谱分析以及 XOps（如 MLOps、AIOps 等）体系等关键技术，为各个研发阶段注入了全新的智能能力。例如，企业级应用中已能够从自然语言需求自动生成代码框架，极大降低开发门槛；智能测试系统可自动生成并执行覆盖复杂业务边界的测试用例，大幅提升测试覆盖率和可靠性；在生产环境中，AIOps 通过智能日志分析辅助定位故障根因并动态实施扩容策略，显著增强系统的高可用性和弹性。面向未来，智能研发将进一步迈向“零代码化”和多模态交互，通过 Agent 架构将代码仓库演化为智能化“知识库”，使得函数和模块能够被动态检索、组合与验证，从而真正实现“AI 原生研发”时代的全面到来。

04

智能应用平台， 应用智能化的技术底座

PART FOUR

人工智能与大模型技术正驱动全栈技术体系革新——从 AI 基础设施、大模型数据服务，到应用运行时与开发框架，技术链条全面升级。在此背景下，融合云原生与 AI 原生能力的智能应用平台，凭借 1) 复杂技术的封装和标准 API 化，2) 动态资源的池化和调度，3) 统一的运行时管理三大特征，正成为各行各业实现智能化的关键底座。

4.1 云原生与 AI 原生双擎驱动，加速应用全面智能化

数字经济时代，企业正经历从信息化→数字化→智能化的范式跃迁，在这个进程中，通过容器化、微服务、DevOps 为代表的云原生技术帮助企业构建敏捷数字基座，支撑业务快速迭代。而以智能体（Agents）、实时决策引擎、自适应工作流为核心的 AI 原生，正在重塑应用价值内核、形态和体验。这场技术与应用的跨时代变革，将推动企业数字化与应用现代化进入崭新的阶段。

从技术理念、服务架构、工程方法、应用形态等多个维度看，云时代和人工智能时代的软件应用，有着各自鲜明的特点，对比如下：

对比维度	云时代	人工智能时代
技术理念	Cloud-Native、机器学习、深度学习、小模型	AI-Native、生成式 AI、基础大模型 & 大语言模型 & 多模态大模型 & 世界模型
服务架构	微服务化、容器化、SaaS、Serverless	智能化、Agentic AI、Agent 架构、MaaS（模型即服务）、RAG
软件工程方法	DevSecOps、AIOps	DataOps、MLOps&LLMOps、DevSecOps 融合，数据工程、模型工程、提示工程
应用形态	数据 + 算法逻辑代码 单模态独立	数据 + 模型 + 数字内容 + 代码、 文本 & 图像 & 音视频多模态融合

随着大模型与 AI 技术的发展，从弱人工智能逐步走向强人工智能，AI 的应用领域也从内容生成走向自主决策和复杂问题解决，Agentic AI 成为人工智能新范式，AI Agent 即智能体，已经成为智能化应用的典型代表。未来大量的智能体相互协作，与成百上千的大模型和工具，组成有机的整体，共同构成 Compound AI Systems(复合 AI 系统)。

云与 AI 原生技术相互融合，推动应用全面走向智能化。基于“AI-First”和“AI-Driven”的理念，未来的软件应用

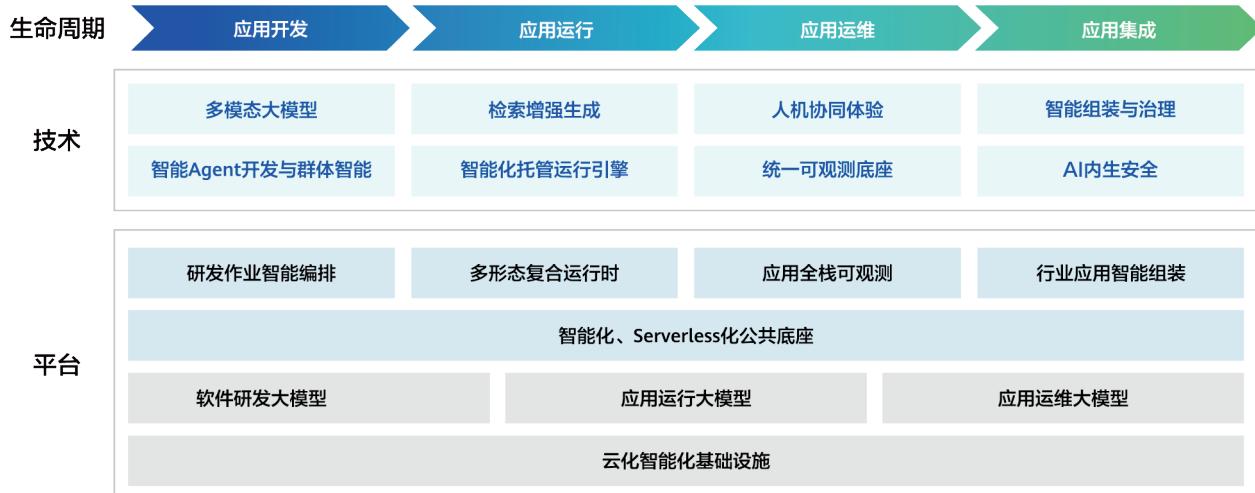
将继承 Cloud-Native 的微服务化、容器化、Serverless 化等能力，增强自然语言 UI、提示工程、RAG 检索增强生成、预训练大模型、XPU 异构混合算力，集成知识库与外部工具等，演进为智能应用。各行各业的智能应用，借助 AI 实现自运行、自学习与群体协作，大幅提升复杂任务处理能力，从“被动服务”进化为“主动智能体”。

智能应用以预训练大模型作为内生智能中枢，赋予系统自主理解意图、规划任务和协同执行的能力。这种变革不仅是技术架构层面的升级，更意味着数字化战略重心从基础架构敏捷转向智能能力内生。



4.2 智能应用平台的核心定位：应用全生命周期的“智能驾驶”

匹配智能时代的技术发展趋势，现代化的智能应用，需要新一代的智能应用平台，为 AI 应用提供 AI 全栈服务，承载云与 AI 原生关键技术，支持智能应用开发、运行、运维、集成等全生命周期与全场景。智能应用平台、参考框架如下图：



新的智能应用平台，内部可分为基础设施服务层、模型服务层、平台核心服务层。

基础设施层，在原有计算、存储、网络等传统云化基础设施服务之外，增强智能化的基础设施和 AI 数据与模型服务，算力架构从单一 CPU 扩展到以 GPU/TPU/NPU 为代表的多处理器混合形态。

模型层，面向应用智能化，从传统的软件研发大模型，扩展到应用运行大模型和应用运维大模型，通过开源、商用、自研大模型，构建多种来源、多种规格、多模态按需组合的大模型服务，支撑应用全生命周期和全场景的智能化。

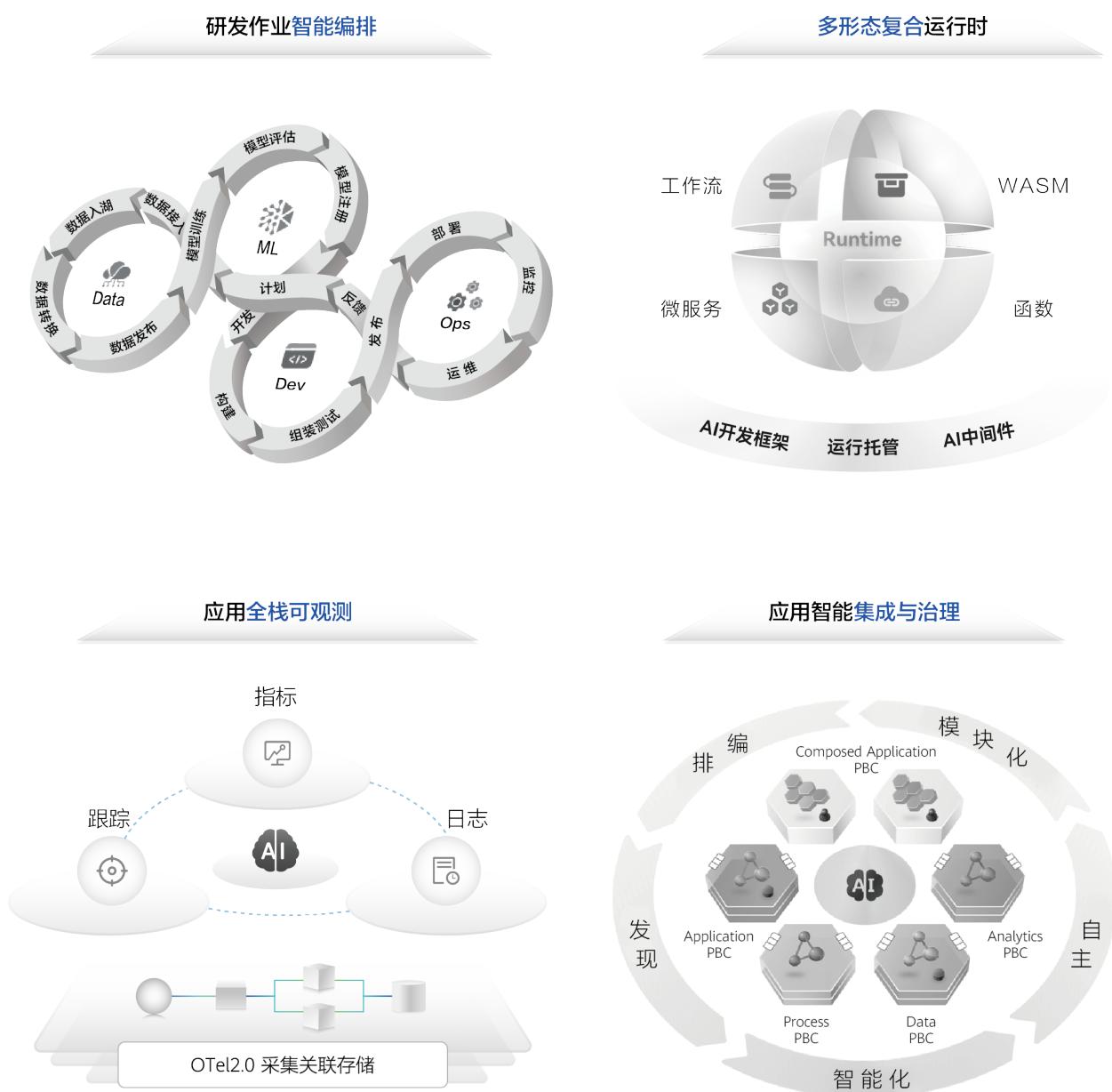
平台核心服务层，为智能应用的各个生命周期阶段提供关键服务。

- 1) 应用开发阶段：提供需求、设计、到开发、测试、发布、部署完整的软件开发工具链能力。面向 AI，数据生产线生产模型训练所需的数据和模型推理所需的知识库，通过模型生产线进行模型的训练与微调，与软件应用生产线共同构成完整的 XOps 融合生产线，支持容易研发作业任务的智能编排。
- 2) 应用运行阶段：平台提供统一的多形态复合运行时，支持工作流、容器、函数、WASM 等多种应用运行形态，支持智能应用的发布、部署与托管，支持应用运行状态的智能感知、智能弹性伸缩、智能流控与安全防护，提供智能管理引擎，提供知识库、应用运行动态信息库，支持多 Agent 协同，实现智能应用管理新范式。
- 3) 应用运维阶段：平台提供 AI 全栈可观测能力，采用统一的 OTel 行业标准，支持 AI 调用与日志的采集、跟踪、指标分析，覆盖 XPU 基础设施层、模型层、平台层和应用层，实现 AI 应用运行环境感知与预测，支持 AI 应用的自学习、自维护、自演进。
- 4) 应用集成治理阶段：平台提供身份认证、API 集成、业务编排能力，支持 AI 应用开发框架和 AI 技术栈，支持云侧、端侧、边缘设备、数据和应用的集成。面向行业场景，提供企业私有数据的全连接能力，提供可复用的 AI

公共能力部件，实现行业应用的智能化组装和高效集成。

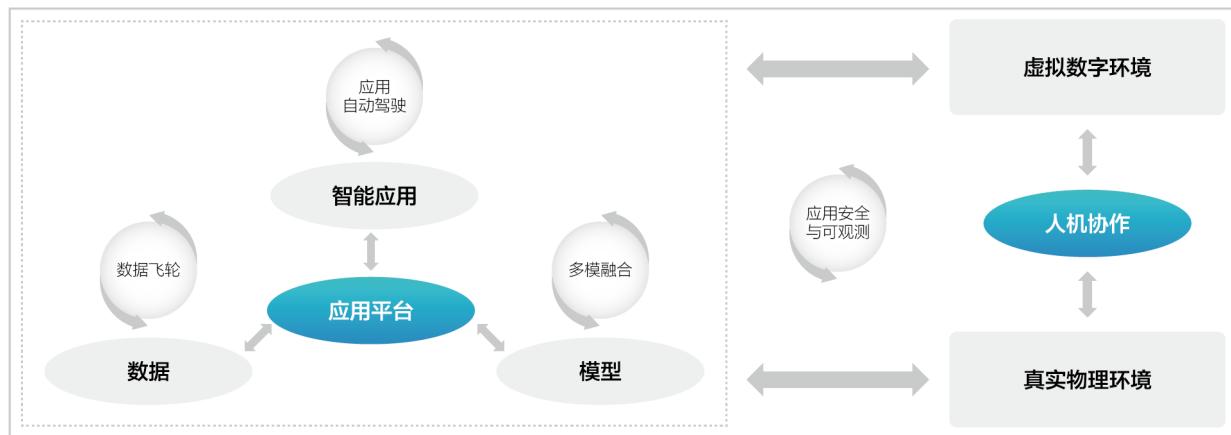
5) 智能化和 Serverless 化的底座：为应用开发、运行、运维、集成提供公共的平台服务，集成消息、事件、缓存等 AI 原生中间件，提供多模型融合、多 Agent 通信、推理加速、实时 RAG、云边一体等能力，统一管理并高效利用 XPU 资源，支持数据、模型与应用的内生安全。

应用开发阶段的研发作业智能编排，应用运行阶段的多形态复合运行时，应用运维阶段的应用全栈可观测，应用集成阶段的智能集成与治理，共同构成智能应用平台全生命周期的四大支柱。



技术层为应用和承载业务提供云和 AI 的关键技术，通过数据服务、模型服务和平台服务，充分释放基础设施的算力价值。在继承云时代应用现代化核心技术的基础上，智能时代的应用现代化包括如下关键技术：多模态大模型、智能 Agent 与群体智能、检索增强生成、智能应用管理引擎、智能组装与集成、统一可观测底座、多模态交互与双向人机协作、AI 内生安全。各关键技术融入到应用平台服务、数据服务和模型服务中，通过 AI 驱动的研发工具链助力智能应用高效开发；通过 AI 运行时、AI 中间件、AI 开放网关、iPaaS 服务等，助力智能应用高性能运行；通过可观测平台助力智能应用高效运维。

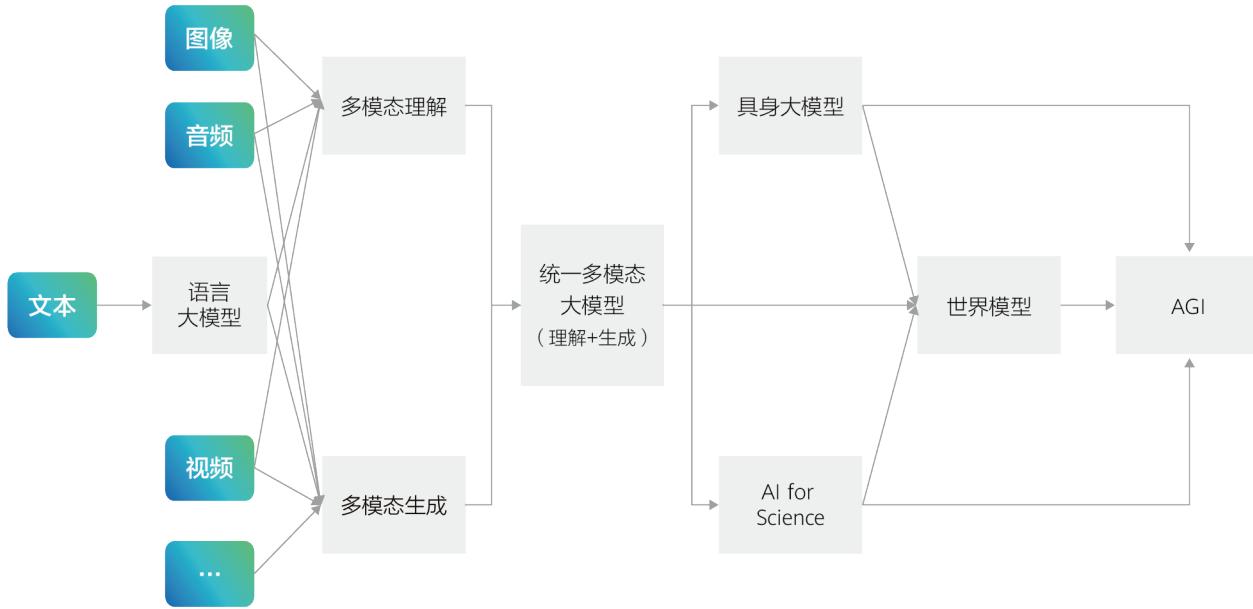
智能应用平台，助力应用从云时代向智能时代演进，端到端覆盖应用全生命周期，无缝融合云原生与 AI 原生技术，实现智能应用全自动驾驶与人机智能协作。



从智能应用场景上看，在 AI 数据与大模型技术服务的驱动下，数据生产、模型训练、软件应用开发，类似汽车的智能驾驶，实现 AI 驱动的多模态大模型、多智能体群体协作式的“应用智能驾驶”，包括：软件开发从“需求分析、系统设计、代码开发、构建打包、测试验证”研发全过程的“应用开发自动驾驶”，真实数据与模拟数据的“数据生成自动驾驶”，大模型开发微调的“模型训练自动驾驶”，智能应用平台的工具与服务，实现软件开发全流程自动化涵盖开发、运行和运维，为开发者提供全场景的应用智能驾驶体验。借助智能应用平台的可观测服务，智能应用从 XPU 基础设施层、模型层、平台服务层到应用层，实现垂直全栈可观测，具备自主运行、环境感知、提前预警、自主决策和主动响应。借助多模态世界大模型，人与智能应用双向协作，与虚拟数字环境和真实物理环境深度互动。

4.3 智能应用平台的八大关键技术

» 4.3.1 多模态融合技术，实现多模态智能



语言大模型是新一代智能应用的“大脑”和语义引擎。LLM 通常指参数规模巨大的深度语言模型，基于海量文本语料训练，具备卓越的自然语言理解和生成能力。随着 LLM 技术的不断发展，LLM 也正在从支持单模态的模型到多模态融合模型进行转变，多模态融合技术是支持应用实现多模态智能的重要一环。多模态大模型是一种能够处理和理解多种类型数据（如文本、图像、音频、视频等）的人工智能模型，与传统的单一模态模型（如仅处理文本的模型）不同，多模态大模型通过整合不同模态的信息，能够实现更复杂的任务。其技术特点在于：（1）多模态数据整合：多模态大模型的核心在于如何有效整合和处理多种数据类型。这需要通过数据预处理技术将不同模态的数据（如文本、图像）转化为统一的表示形式，以确保模型能够理解和协同处理这些信息。（2）模型的架构设计：为了实现多模态的协同，模型架构通常需要引入专门的连接器（Connector），将非文本模态（如图像特征）转化为与文本模态兼容的表示形式。这种设计使得模型能够在不同模态之间进行信息融合和交互。（3）跨模态任务支持：多模态大模型能够支持复杂的跨模态任务，例如图像描述生成、语音识别和视频理解等。这些任务需要模型具备跨模态的感知、理解和推理能力。

多模态大模型当前已经在多个领域得到了广泛应用，例如医疗、教育、研发等。在医疗领域，多模态模型可以整合患者的文本病历和医学影像，辅助医生进行诊断；在教育领域，模型可以结合文本和图像内容，提供更丰富的学习资源。

在研发领域，多模态大模型也显著提升了应用的开发效率。大模型结合应用数据进行调优，能够显著提升模型的能力和应用的智能化效果与体验。通过将应用场景的高质量数据与大模型相结合，可以实现模型的持续优化和适应性增强，对应用数据的预处理和清洗可以确保数据的准确性和一致性，从而为模型提供可靠的基础支持。随着 AI 辅助生成的应用占比越来越多，也促进了应用从云原生向 AI 原生演进，预计到 2027 年，AI 原生应用将占比一半以上。

» 4.3.2 智能 Agent 开发与运行，实现群体智能



1、智能 Agent 开发工具的演进

智能 Agent 开发正经历从专家化到平民化的范式迁移，形成全码 - 低码 - 零码协同进化的技术矩阵。全码开发凭借其对复杂逻辑的深度定制能力，依然在一些对性能和算法有严苛要求的高精度工业场景中占据主导地位，例如复杂物理模型的工业智能 Agent 开发。低码开发通过可视化的编排界面与预制的组件实现业务逻辑的敏捷封装，使更多具备一定业务知识但编程技能有限的人员参与到智能 Agent 的开发中，例如数据分析师使用低码平台快速搭建数据分析智能 Agent。零码开发更是将开发门槛降至最低，非技术人员通过对话即可生成满足特定需求的智能 Agent，常见于办公自动化场景，如对话生成自动回复邮件智能 Agent。

2、多 Agent 智能体协同的崛起

在复杂研发场景中，单智能体的能力边界日益显现，多智能体协同正成为解决系统性难题的核心范式。以 Manus 为例，其通过多 Agent 协作的生态系统，实现了专业能力的模块化分工与动态协作，部分单元专注于数据收集与分析，部分单元聚焦于模型训练与优化，另有单元负责决策路径的智能映射与执行反馈。再比如在药物研发场景中，数据收集 Agent 从医学文献、临床实验数据抓取关键数据，数据分析 Agent 对关键数据进行深度挖掘，找出潜在的药物靶点与作用规律，模型训练 Agent 基于分析结果构建预测模型，最后，决策执行 Agent 依据模型预测结果制定最优的药物研发方案，并自动推进实验流程。这种多 Agent 协同模式极大地提高了研发效率，同时避免了人为操作可能产生的疏漏。

3、群体智能下的多 Agent 协作优势

群体智能驱动的多 Agent 协作系统展现出强大的自适应能力与创新优势。在复杂研发过程中，面对动态需求变更与非线性技术挑战，多 Agent 系统通过实时信息交互构建起动态策略网络，能够基于环境变量进行策略调优。当研发路径出现偏差时，多 Agent 协作系统依托局部感知与信息协同，提出调整建议，通过群体决策机制迅速输出修正方案。同时，多 Agent 协作系统通过差异化功能定位和智能信息协同，有效激发创新潜能，针对复杂研发问题输出多元化解决方案，推动技术路径与研发方法的创新，这种协作创新是单智能 Agent 难以实现的。

4、全码、低码、零码在多 Agent 开发中的融合

在多 Agent 智能体协同开发体系中，全码、低码、零码并非独立的技术路径，而是相互融合、优势互补。全码开发聚焦于构建复杂算法逻辑与高性能计算能力的核心 Agent 模块，为系统提供性能保障和基础能力底座。低码开发通过模块化组件实现标准化功能 Agent 的高效构建，完成系统基础框架搭建与核心功能整合。零码开发则赋能非技术人员基于实际需求对 Agent 进行精细化参数配置与行为逻辑调整，确保多 Agent 系统具备动态响应业务环境变化的能力。通过这种融合，可显著提升多 Agent 智能体协同系统的开发效率与应用部署灵活性，有效推动智能 Agent 在研发场景等重点领域的规模化应用进程。

5、A2A 协议，构建多智能体自主协作的通信标准

随着人工智能技术的快速发展，多智能体系统（Multi-Agent-System）正从单点应用加速向复杂协作生态演进。A2A（Agent-to-Agent）协议作为业界首个标准化的多智能体通信标准，其发展将重构 AI 协作范式，推动智能体从“孤立工具”向“自主协作的数字生命体”的升级。中间件作为 A2A 协议中的核心支撑组件（通信桥梁 / 能力增强器 / 系统稳定器），通过标准化接口、协议适配、安全增强和资源管理四大核心能力，有效应对 A2A 协议在分布式环境中的通信复杂性、可靠性与扩展性等挑战。在云端，中间件（如消息队列，API 网关）基于 Agent Card 的能力，通过智能路由算法动态选择最优 Agent，同步保障高并发场景下的负载均衡。在边缘端，边缘 Agent 依托消息队列 MQTT 协议与云端 Agent 建立低延时通信通道，解决云边 Agent 之间的可靠通信问题；针对 A2A 协议的任务（Task）管理要求，消息队列通过任务状态的全生命周期（如创建、执行、完成、失败）持久化存储与状态追踪机制，从技术底层确保任务（Task）执行的最终一致性，有效避免任务丢失或重复执行风险。未来，随着 Agent 与 AI 原生中间件的融合（如支持 A2A 的服务网格、智能消息队列），多智能体系统将实现从“连通”到“认知”的跨越，全面释放群体智能的潜力。

» 4.3.3 检索增强生成，增强智能体个性化记忆

RAG（Retrieval-Augmented Generation，检索增强生成）是一种结合信息检索与生成模型的人工智能技术，旨在通过动态调用外部知识库，提升大语言模型的准确性和可靠性。它有效解决了传统模型的“幻觉问题”，同时增强了信息的时效性和安全性，成为当前 AI 领域的重要趋势。RAG 的工作流程一般分为四步：（1）索引：将文档分块并转化为向量（高维数字表示），构建可快速检索的数据库。（2）检索：根据用户问题，从知识库中匹配最相关的文本片段（支持关键词、语义或混合检索）。（3）增强：将检索结果整合为上下文信息，输入生成模型。（4）生成：模型结合上下文与自身知识，输出最终回答。RAG 技术的发展从基于“检索 - 生成”链式结构到模块化 RAG 再到 Agentic RAG 持续在发展，Agentic RAG 可以自主智能体动态决策，如根据问题复杂度自动选择检索策略，或调用外部 API 验证答案，实现闭环优化。RAG 技术的优势可以总结为以下四点：

- 1) 减少幻觉：答案基于检索证据，而非仅依赖模型记忆。
- 2) 动态更新：仅需更新知识库，无需频繁训练模型，成本更低。
- 3) 安全可控：企业数据可本地存储，避免敏感信息泄露。
- 4) 可解释性：提供答案来源，增强结果可信度。

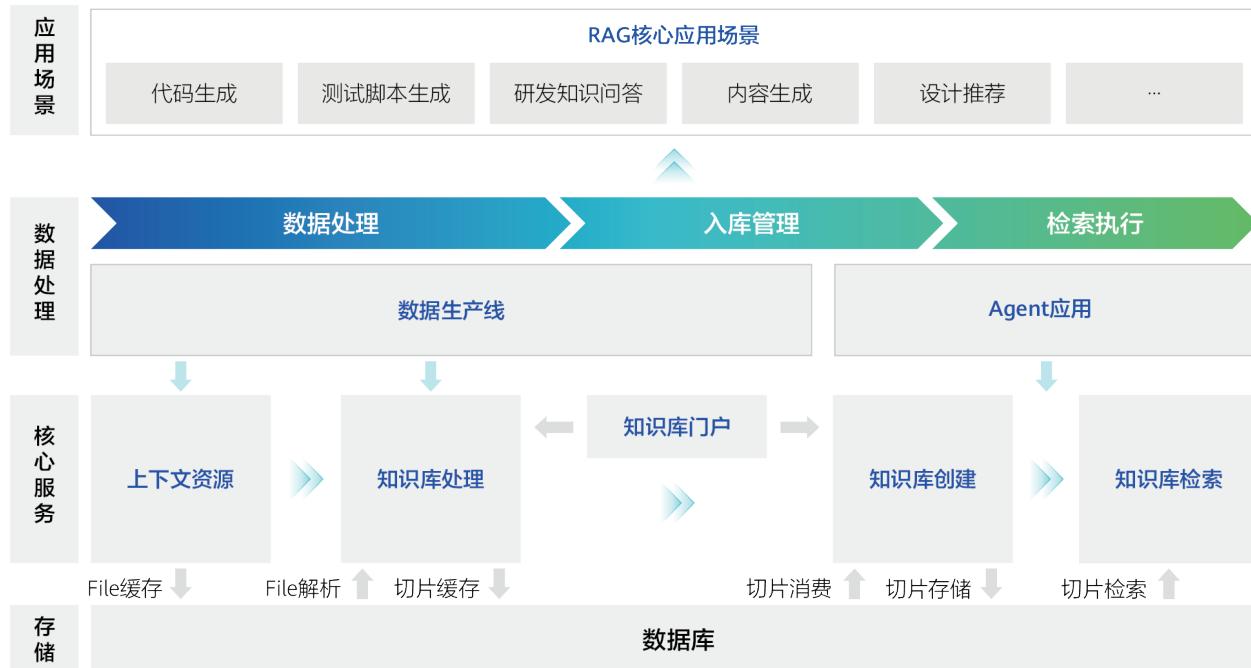


图 检索增强生成 (RAG) 的基本流程

RAG 技术目前在多个领域都发挥了重要作用，如智能客服、医疗、金融和研发领域。在研发领域，如上图所示，通过数据飞轮与 RAG 技术的结合，构建知识数据清洗、转化和评测流水线，可以支持各类研发作业，迭代优化高质量领域知识，淘汰冷门知识，实现知识的 E2E 生命周期闭环，持续提升大模型推理的效果；通过为应用层抽象并统一多种知识数据与存储，可以实现安全高效地访问，简化并提升 AI 应用开发效率。

» 4.3.4 智能应用管理引擎，让应用自动运行与自主优化

在云计算和人工智能技术迅速发展的背景下，传统的应用管理方式已难以满足企业对高效、稳定和灵活性的需求。智能应用管理引擎应运而生，作为一种创新的系统，它通过自动化和智能化的手段，持续优化应用的运行状态，提升系统的整体性能和稳定性。

核心架构与技术实现

智能应用管理引擎结合了大模型 Agent 和 MCP (Model Control Protocol) 技术，构建了一个高效、可靠的管理

系统。Agent 通过调用大模型对应用的运行情况进行实时分析和决策，确保系统能够快速响应各种变化。MCP 协议则作为引擎与应用运行平台之间的桥梁，负责收集运行信息并执行变更操作，实现了平台与引擎的高效协同。

数据驱动的智能决策

引擎通过平台能力收集应用的运行状态信息，确保数据的准确性和可靠性。同时，系统设立了多个管理角色，形成复合 AI 系统架构，从多个角度进行决策分析，提高决策的准确性。引擎还通过 RAG (Retrieval-Augmented Generation) 机制，持续扩展应用运行的最佳实践，并总结历史经验，不断优化管理策略，实现应用的全栈优化。

幂等执行与无损变更

为了避免生成式 AI 可能带来的臆想性决策，智能应用管理引擎采用了幂等执行策略，即多次执行相同操作不会对系统状态产生不必要的影响。此外，系统支持无损变更，无论是架构调整还是资源配置优化，都能在不中断业务的前提下进行，确保系统的稳定性和用户体验。

自主优化与持续迭代

引擎通过持续分析应用的运行数据，结合行业最佳实践，对应用的配置、架构和资源进行调优，实现自我优化。系统还支持历史经验的总结和迭代，使得管理策略能够随着时间的推移不断进化，适应不断变化的业务需求。

灵活的扩展性与平台兼容性

智能应用管理引擎采用分层解耦的设计，通过 MCP 协议实现与 Agent 的独立交互，未来可以接入用户自带的或第三方的 Agent，提供更灵活的管理方案。无论是中小企业还是大型企业，都可以根据自身需求定制管理策略，提升 DevOps 效率。

随着技术的不断进步，智能应用管理引擎将不断发展壮大。未来，每一个在平台上运行的应用都将拥有一个专属的智能管理引擎，实现全自动地运行和调优。通过平台能力的深度整合和持续优化，应用的自动化运行和自我调优将成为常态，极大提升企业的运营效率和竞争力。

» 4.3.5 智能组装与集成，扩展应用能力边界

在数字化转型的深水区，企业正面临两大核心问题：业务敏捷性需求指数级增长和技术体系复杂度几何级攀升。传统开发模式下，新功能上线周期平均周期 6-8 周，难以匹配市场变化的节奏。与此同时，企业应用将需要动态融合 AI 能力、IoT 设备与第三方服务，形成跨域协同的“数字能力体”。基于上述背景下，智能组装与集成技术成为了破局关键，通过模块化封装、智能编排和动态架构，将离散的技术要素转化为可自由组合的业务能力 PBC，使得企业能够像搭建乐高积木一样构建业务能力，同时保持架构的弹性和进化能力。

外部工具集成：打破生态壁垒，可连接一切的超级连接器

在开放协同的数字化生态中，企业需要连接许多不同种类的外部系统，但传统点对点集成模式存在三大痛点：对接成本高、安全风险大和维护难度高等。新一代集成体系通过协议抽象、语义理解和智能治理三层解耦方式设计，解



决上述问题。协议抽象通过标准化中间件统一转换 REST/SOAP/GraphQL/gRPC 等协议；语义理解基于领域本体论构建语义映射库，实现业务概念的跨系统对齐；智能治理实现动态凭据管理支持 OAuth2.0/OIDC/IAM/JWT 等多模式自动续期、数据合规引擎内置合规规则模板、异常熔断机制等。利用通用协议 OAS 的 RPC、内置高效隔离 WASM 自定义运行时和外置通用托管连接器框架三种模式支持任意生态外部工具对接。

智能组装平台：业务创新的加速引擎

当前业务需求变化速度超过 IT 交付能力时，交付容易陷入困境。为此，新一代智能组装平台将在 AI 辅助设计、弹性抽象和多 Agent 协同编排等技术上突破。AI 辅助设计通过 LLM 理解需求信息，自动生成业务流和生成对应的连接器组件代码；弹性抽象将对业务逻辑与底层实现解耦，实现动态运行资源和自动化调整特性资源能力；多 Agent 协同编排支持统一的 Multi-Agent workflow 的 DSL，可实现多模态业务编排与工具调度。

AI 网关：全域连接的神经中枢

AI 网关将成为 Agent 时代的核心基础设施，需兼具传统 API 网关的稳定性与 AI 原生能力（模型编排、多模态优化、幻觉治理等）。其核心演进方向包括：

- 1) 协议扩展：**从 HTTP 流量管理升级为 AI 服务治理，支持 gRPC/GraphQL 等 AI 协议，优化流式响应（SSE）与非结构化数据（文本 / 图像 / 音视频）传输。
- 2) 动态编排：**从单一服务路由转向多模型动态调用链编排（如 LLM+ 语音 + 图像组合），实现意图驱动的模型调度。
- 3) 智能调度：**基于模型性能、成本、SLA（响应时延 / 准确性）自动选择最优大模型供应商，支持混合云边协同推理。
- 4) 深度可观测：**从基础监控升级为 AI 全栈追踪，覆盖 Token 消耗、推理延迟、上下文缓存效率、提示词防护率等特有指标。

MCP 动态服务架构：动态资产的服务智能生成与开放

集成平台拥有丰富的客户业务生态资产，资产包括 API、连接器、组合应用等。当前客户都是单点方式订阅基于提供的文档学习如何使用对应的能力，再结合自身业务进行手工代码方式调用使用，过程漫长，且对每个资产使用时都需要关注凭据权限等问题。

动态 MCP 服务架构，可基于大模型的方式学习集成平台中注册和托管的资产，并基于当前这些资产在系统中的使用和集成关系，生成对应的 MCP 服务能力，供客户和集成平台的 AI 智能体使用。使得集成平台的资产可自动形成可供智能体使用的工具集。

能力边界扩展的飞轮效应

通过上述四层技术体系的协同，企业可构建持续扩展的“数字化 DNA”。实现：横向裂变，每个接入的外部系统，可通过组件市场转化为可复用能力；纵向跃迁，AI 网关的智能化路由是传统系统获取“数字神经”，响应速度大幅提升；立体进化，MCP 架构自动智能工具化，使得业务编排更简单。

» 4.3.6 统一的智能体可观测，支撑海量智能应用高效运维

随着 AI Agent 元年的到来，客户大量的 LLM-based 应用上线，未来应用将逐步从云原生转型成为 AI 原生，形成 Event-Driven、Data-Driven、LLM-Driven 三分天下的应用形态，将产生复杂的应用链路拓扑和组件依赖，同时催生出多元且复杂的运维需求，面向未来海量 AI-Native 应用的智能化运维场景，将形成新的智能化可观测技术框架，该框架包括统一且极致性价比的可观测统一存储分析底座（数据层），运维大 / 小模型（算法层），由运维智能体编排而成的运维助手（应用层）以及各种智能运维场景（场景层），如下图所示：



可观测数据统一存储分析底座是支撑海量运维的基础，指标、日志调用链等可观测数据需要遵循统一的数据模型，并实现统一的存储与分析，以消除数据孤岛，支撑上层智能化运维算法发挥最大价值。

智能化运维算法层将不仅包括特定场景化的小模型算法，也包括由运维知识库以及各类可观测数据微调而来的各类运维垂域大模型，实现价值互补。

在 AI 原生应用时代，智能化运维的主体将由 SRE/ 开发人员逐步向由各类 AI 智能体编排而成的运维助手迁移，同

时运维领域经验将由以 SRE 为代表的“碳基人”逐步转移到以运维助手为代表的“硅基人”，从而实现运维领域的“无人驾驶”。

» 4.3.7 多模态交互，建立双向人机协同体验

多模态交互能力和人机协作体验是推动工业向通用人工智能（AGI）方向发展的基础。AGI 旨在复制类人智能，具有跨不同任务和环境理解、学习和适应能力。使用多模态各种输入和输出通道（如视觉、语言、听觉信号和手势）的集成，使 AGI 系统能够以更接近人类认知的方式感知和解释世界。这一功能显著增强了 AGI 的鲁棒性、灵活性和上下文理解，使其能够自然有效地与人类和复杂环境交互。通过多模式交互建立无缝的人机协作，使各行业能够利用 AGI 的适应性和全面理解，促进医疗保健、机器人、自动化和教育等行业的创新、效率和生产力。新型的 GUI Agent 技术在实现这种多模态交互和人机协作方面起着至关重要的作用。GUI Agent 使用多模态语言模型来像人类用户一样解读图形用户界面 (GUI) 并与之交互。它们能够仔细检查程序元素，并执行人类操作，例如分析屏幕上的元素、点击按钮、输入文本等，操作数字环境交互执行任务。模拟人类方式操作计算机执行各类任务。GUI Agent 关键模块解析如下：

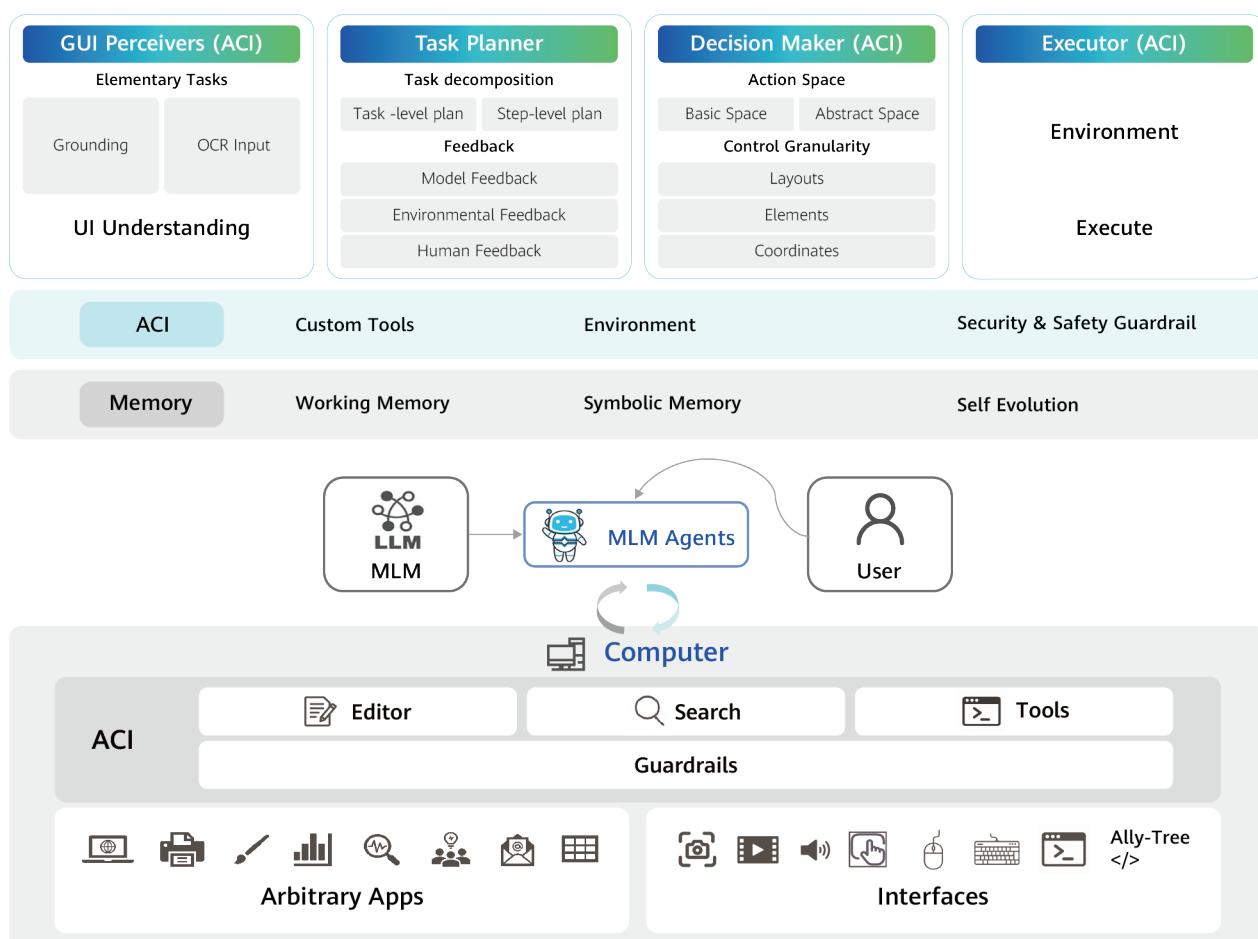


图 1 GUI Agent 模块与架构图

GUI Agent 技术框架由六个组件组成：GUI 感知器、任务规划器、决策制定者、记忆检索器和执行器和 LM (Language Model) Agent- 计算机接口模块 ACI (Agent-Computer Interface)。这种框架也可以有变体，例如，以多代理方式控制框架，包括一个规划代理、一个决策代理和一个反思代理，以解决移动设备操作中的导航挑战。

1. GUI 感知器：设备任务执行的关键组件，需精准解析用户输入并实时响应 UI 变化。受限于单模态 LLM 的能力，该模块通常作为独立增强单元集成至框架中。单模态 LLM 需依赖 XML 可访问树 (Accessibility Tree) 实现操作系统环境导航，而多模态模型 (MLM) 可直接通过 屏幕截图与像素级解析 交互。SOM (Set Of Mark) 方法则通过创建视觉参考点定位 UI 元素。实际应用中，多模态 LLM 代理需 微调 UI 感知模型 以适应具体场景。这种感知机制使代理能 " 看见 " 并操作无法通过程序化接口访问的界面元素。

2. 任务规划器：GUI 代理应有效地分解复杂任务，通常采用链式思维 (CoT) 方法。由于任务的复杂性，通常引入了一个额外的模块来支持更详细的规划。在整个 GUI 代理的过程中，计划可能会根据决策反馈动态调整，通常通过 ReAct 风格实现。例如可使用屏幕上的观察结果来增强 CoT，以改进决策；或开发了一个反思代理，提供反馈以优化计划。

3. 决策制定者：决策者提供控制设备的下一步操作。大多数策略是将与 UI 相关的动作（如点击、文本和滚动）定义为一组基本动作空间。在更复杂的情况下可封装了一系列动作来创建标准操作程序 (SOPs) 以指导进一步的操作。随着 GUI 代理功能的增强，操作的粒度变得更加精细。

4. 记忆检索器：记忆检索器 (Memory Retriever) 被设计为一种额外的信息来源，以帮助代理更有效地完成任务。GUI 代理的记忆通常分为内部和外部两类。内部记忆包括执行过程中的先前操作、屏幕截图和系统状态，而外部记忆包括与 UI 或任务相关的知识和规则，为代理提供额外的输入。

5. 执行器：执行器将输出映射到相关环境。在桌面系统，ACI 为目前最前沿的解决方案但在移动环境大多数使用 Android 调试桥 (ADB) 来控制真实设备，但也有使用模拟器来访问额外的与 UI 相关的信息。

6. ACI：内嵌 ACI 允许人工智能代理在与计算机软件交互时像人类用户一样工作。ACI 使 AI 代理能够通过导航屏幕读取和操作动态图形用户界面 (GUI) 而执行命令和输入数据，而无需固定系统连接。AI 的自适应特性使其能够通过不需要自动化开发的框架与任何数字工具交互。ACI 为 LM Agent 提供定制的工具使能 Agent 和计算机系统之间的无缝交互与确保多模态模型的接地性与响应的准确性同时也提供必要的安全对齐的护栏。

以模型适配物理世界是一个未来应用的发展趋势，多模态模型提供的强大的视觉语音的能力提供了新颖的拟人人工智能智能体。当前由 MLM 模型赋能的 GUI-Agent 的应用案例包括 Computer-Use Agent (CUA) 如 OpenAI Operator, Anthropic Computer Use，和国内的 Manus AI 项目已相继问世，近期宣布推出试用版，应用场景包括 Web 浏览器交互，包括填写表单、在线下单、预约和其他基于浏览器的重复任务，如网站创建、股票分析、旅行计划和日程管理，或软件工程级代码开发，其功能表现均不凡深获业界好评。GUI Agent 使能显然成为未来 AI-Native 应用的趋势。



» 4.3.8 AI 内生安全，确保应用全生命周期可信

应用智能化演进中，安全有了新的定义，安全防护技术上面临全新挑战。在数据开发、模型开发、模型运行、内容生产、AI 应用智能体、模型服务与应用的运行和环境、应用的运维等方面，需要新的安全防护能力，以支持智能软件供应链安全、数据模型可追溯、AI 软件全生命周期的内生安全防护。

AI 应用安全技术框架如下图：



软件工具与软件供应链方面，需要具备数据开发工具链、模型开发工具链、AI 应用开发工具链的安全，保障开源的数据、开源模型、开源的数字媒体内容、开源软件的“来源安全、过程安全和结果安全”，整体端到端可信。

在数据安全技术，需要支持数据来源可信，数据内容合规，数据版权合规，敏感信息识别与关键信息资产检测，数据来源可追溯、数据防篡改与防泄漏等。

模型开发安全方面，需要保障模型来源可信、可溯源，模型安全评测与版本管理，支持防篡改与防泄漏，模型领域隔离强化训练与合规等。

生成内容安全方面，需要控制领域性的安全风险，例如软件应用开发中代码的安全治理、版权合规，支持 AI 攻击检测、内容风控、用户与企业私域数据保护，提示词防攻击与优化等。

05

应用智能化的实现路径与 关键考量点

PART FIVE



在人工智能技术持续演进并加速产业变革的时代背景下，企业正处于数字化与智能化转型升级的关键阶段。应用现代化与智能化转型的协同推进，已成为企业重塑数字竞争力的突破口。通过深度融合云原生与 AI 原生的技术理念、基础设施、服务架构、研发流程、工程方法以及研发工具，企业得以构建智能化应用形态，高效释放生产力，并在快速迭代的市场环境中建立持续创新优势。本章将深入探讨应用现代化为企业带来的价值、实现路径与关键考量点，为企业应用现代化升级提供可落地的实施框架。

5.1 应用智能化实现路径

» 5.1.1 应用智能化成熟度技术演进路径

应用智能化成熟度从 L1 至 L5 逐级演进，以交互方式、自主程度、任务复杂度等维度对五个级别进行定义，以中国信通院人工智能研究所《软件智能化成熟度模型》标准为主要参考依据。

L1 级智能应用以固定内容生成、固定交互方式（如按钮、表单）为特征，主要依赖传统机器学习模型运行，典型场景包括基础客服机器人、人脸识别等应用。此类应用缺乏动态学习能力，需人工全程主导，适用于低复杂度、高重复性任务。

L2 级智能应用引入基于大模型的自然语言交互（如多轮对话方式）方式，能理解用户意图并为人类提供辅助能力，典型场景如智能搜索、代码生成补全等。此类应用需要人类大量参与（如确认和修改生成内容），适用于知识类场景的任务。

L3 级智能应用实现多种模态生成（如文本及图像或语音等），可通过外挂知识库，并基于智能体等方式处理单一领域复杂任务，典型场景如编码智能体、AI 原生应用等，此类应用需要人类参与关键决策（如结果审核）。

L4 级智能应用具备跨领域复杂系统任务处理能力，通过多智能体协同与动态知识库调用，仅需人类设定初始目标。

L5 级代表智能应用的终极形态，能自主完成全领域未知任务（如科学发现），通过全域知识库和自学习实现自我迭代，人类零干预。

○ 应用智能化的实现路径与关键考量点



图：应用智能化成熟度分级

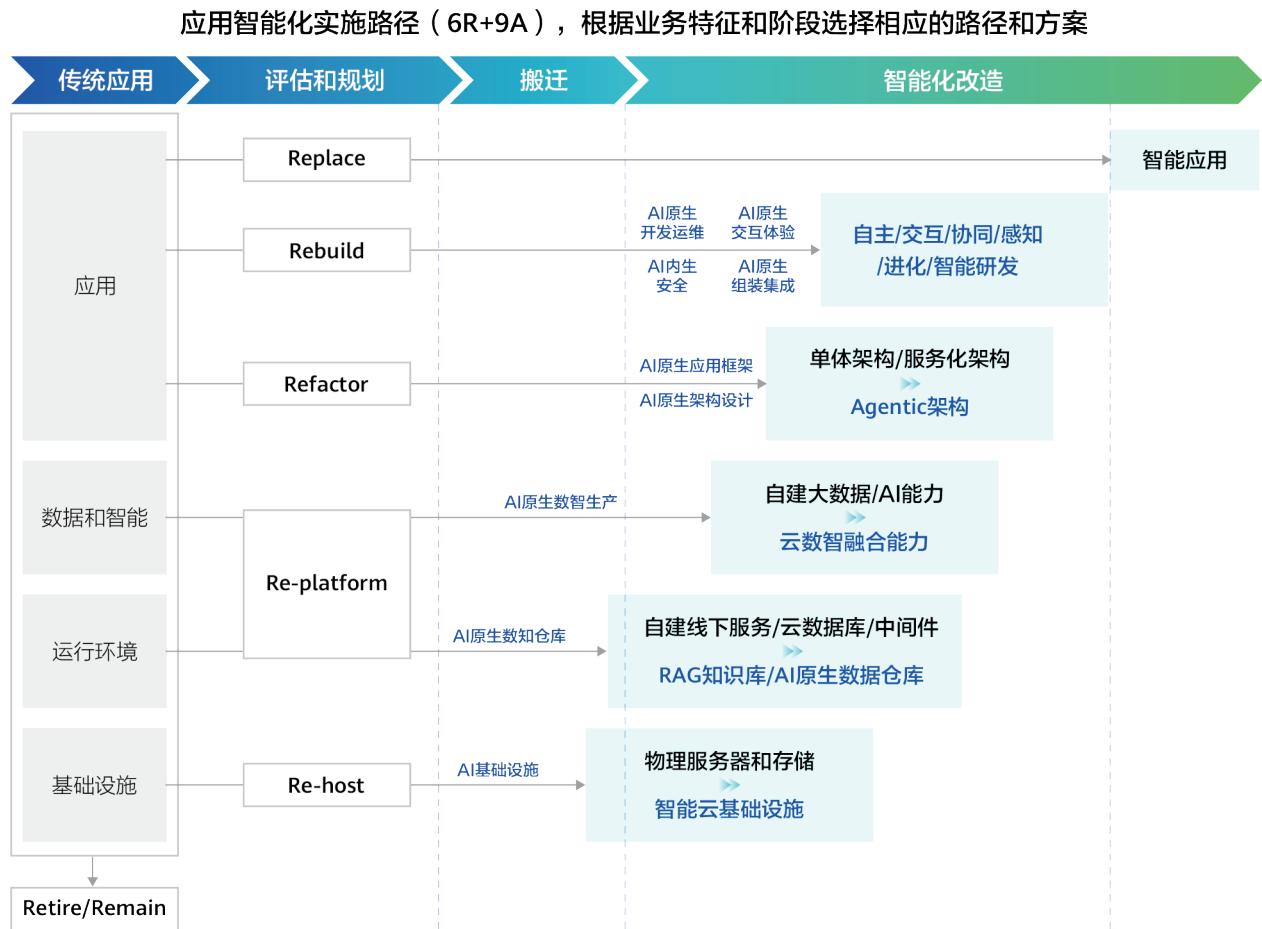
各个级别存在关键差异，从 L1 至 L2，实现从规则驱动转向自然语言交互，从 L3 至 L4，突破单一领域限制，通过多智能体协同解决系统级问题，从 L4 至 L5，实现自我迭代与跨领域自学习能力，从而迈向通用人工智能。

表 1：智能应用成熟度分级特征

成熟度等级	技术特征	人类参与度
L1 基础智能化	传统机器学习模型、按钮 / 表单交互	人工主导
L2 辅助智能化	单模态生成、自然语言交互、检索内部知识库	人工较多参与
L3 部分智能化	至少两种模态生成与交互、大模型与工具调用协同、外挂知识库或联网、处理单领域复杂任务	人工审核输出结果
L4 高度智能化	多模态生成与交互、跨知识库协同、处理跨领域复杂任务	人工设定初始目标
L5 超级智能化	全模态生成与交互、全域知识库调用、自主完成未知领域任务、自主进化、自主迭代	零人工干预

» 5.1.2 应用智能化的 6R+9A 实施路径

在上一节的 L1-L5 成熟度阶梯中，我们勾勒了智能应用能力的纵向演进。为真正推动传统系统踏上这一阶梯，本节聚焦横向“路线图”——6R 策略与 9A 能力矩阵，它们共同构成从传统应用到智能应用的落地方法论。



传统应用的智能化升级经历传统应用、评估规划、搬迁和智能化改造四个阶段；6R 策略贯穿其中，决定具体业务在何时、何地、以何种方式融入智能能力。Replace 意味着直接替换，当市场存在成熟智能解决方案时，此策略可最快见效；Rebuild 在保留业务价值的同时重塑智能化应用，以 AI 原生开发运维、AI 原生交互体验、AI 内生安全和 AI 原生组装集成为基础，快速构建兼具六大特征的智能应用；Refactor 更强调架构层面调整，将单体或传统应用通过 AI 原生架构设计和 AI 应用框架重构过渡到 Agentic 架构，使应用对智能体友好并确保可观察性；Re-platform 将数据与模型生产力搬上云智算平台，通过 AI 原生数智生产与 AI 原生数智仓库构建贯通数据、算法、算力的统一生产线；Re-host 解决基础设施迁移，将物理服务器与专用存储平滑切换到可弹性的 AI 基础设施，为后续智能改造奠定基础；Retire/Remain 则提供现实主义出口——对价值低、风险高或无法改造的遗留系统果断退役，对仍具商业价值却暂不适宜改造的部分予以保留，以资源调度方式优化整体投入产出。



» 9A 能力栈介绍

AI 原生开发运维：以软件智能化生产线贯穿需求、开发、测试与发布，低码 / 零码组装让业务人员参与迭代，XOps 将监控与回滚纳入同一流水线，交付节奏在安全边界内实现小时级循环。

AI 原生交互体验：将文本、语音、图像与视频模型封装为一致的对话接口，前端仅需调用统一 API 便可实现自然语言、视觉理解与内容生成，用户交互从点击转向语义协作。

AI 内生安全：在数据、模型、流程、环境与内容五个层面同步加固：静态脱敏与动态水印保护数据，模型审计与红队评测遏制漏洞，可信执行与内容校验贯穿推理全过程，使安全成为系统默认属性。

AI 原生组装集成：通过智能组装平台与 AI 网关，将传统 API、外部 SaaS 与多智能体编排在同一策略面板；MCP Server 动态选择最优调用链，减少手工胶合，提升吞吐与可观测性。



AI 原生应用框架：提供 LLM SDK、LangChain 及智能体编排模板，开发者用少量代码即可组合检索、推理与行动；内置缓存与探针确保链路可控、状态可追踪。

AI 原生架构设计：在微服务与 Serverless 之上引入 Agentic 架构与 A2A 协议，使智能体既能独立运行又能灵活协同，并实现全链路弹性扩缩与事件驱动。

AI 原生数智生产：将数据治理、特征工程与模型训练形成自动流水线；数据湖、特征仓与模型仓统一管理，指标反馈实时驱动模型快速重训与上线。

AI 原生数智仓库：结合向量数据库、模块化 RAG 与 Agentic RAG，将非结构化知识与结构化数据统一检索；知识更新与数据生产线同步，保障推理的新鲜度与一致性。

AI 基础设施：由 GPU / NPU 池化与高速互连构成云智算底座，统一调度弹性算力与推理加速；全链路可观测平台同时呈现资源利用率、延迟与能耗，为上层能力提供稳固支撑。

以上九项能力与 6R 路线相互配合，为传统应用智能化改造提供可组合、渐进式升级路径，并为应用迈向更高成熟度奠定可持续技术基座。6R 确定“怎么走”，9A 决定“带什么”。它们相互配合，使企业在评估、迁移与重构过程中灵活切换策略，并在技术落地阶段针对痛点逐层加固。依托该实施路径，组织不仅可在短期内取得局部智能化成效，更可为迈向更高成熟度打下可持续的技术与管理基础。

过去十年，云原生技术的演进已显著加快软件研发节奏。相比之下，生成式 AI 驱动的智能应用浪潮正以前所未有的斜率攀升。模型推理成本持续下探、多模态能力迅速普及以及 Agentic 架构的场景化落地，使应用从“可用”到“智能”的跃迁周期由年缩短至月甚至周。企业能否在最短时间内让核心系统跨过 L3 门槛、迈向多 Agent 协同的 L4 乃至闭环自治的 L5，将决定其在未来竞争格局中的生存空间与议价能力——智能应用成熟度已成为与营收增长、运营效率并列的第三条生命线；

回望本章，我们自上而下描绘了 L1-L5 的纵向阶梯，又以 6R+9A 构建了横向路线图与能力栈：6R 回答“先动谁、何时动、如何动”，9A 回答“用什么技术把它落地”。唯有将两者耦合，企业才能在评估、迁移、重构与扩张各阶段形成闭环：Replace 与 Rebuild 为高价值场景快速注入智能火种；Refactor 与 Re-platform 夯实架构与数据底座；Re-host 释放算力弹性；Retire/Remain 确保资源投入与业务价值始终匹配。与此同时，9A 九大能力从开发运维、交互体验、安全治理到数智生产和基础设施，为每一次策略选择提供可组合、可量化、可演进的技术支柱，确保智能化成果既能在短期见效，也能在长期保持可持续演进；

由此可见，依循 6R+9A 这一兼顾战略节奏与工程细节的实施路径，组织才能在智能化浪潮冲击中稳住节奏、加速转型——既抢占新赛道，又守住核心业务韧性，最终在瞬息万变的市场中保持持续竞争力。

5.2 应用智能化的关键考量点

» 5.2.1 战略与业务价值对齐

企业在推进应用现代化时，应避免“为现代化而现代化”，必须以明确的业务目标为导向。通过顶层设计明确 ROI 指标，优先改造高业务价值系统，确保现代化改造切实为企业带来实际效益。在成本效益分析方面，需要评估云 - 边 - 端协同架构的算力分配，平衡性能与成本，并选择高性价比的模型训练与运维方案。例如，构建混合云架构以平衡本地 IDC 与公有云资源，实施训练 - 推理分离，降低 GPU 消耗；在满足业务需求的同时，优化资源投入，提升投资回报率。

» 5.2.2 用户体验重构

在交互方式上，应用应支持语音、图像、文本等多模态输入，构建自然交互界面，提升用户友好性。同时，要确保多模态融合的深度，使应用能够处理复杂的现实场景——如医疗影像分析中的多模态融合、自动驾驶等复杂场景的实时响应能力，复杂场景任务完成率应提升。此外，还需建立用户行为数字孪生，实现界面自适应优化；通过 AI 辅助界面简化设计，减少操作步骤，为用户提供更加流畅、便捷的使用体验。

» 5.2.3 组织流程适配

企业需推动研发团队从职能型向产品型转型，建立跨职能敏捷小组，形成“业务 + 数据 + 算法”铁三角，并明确端到端责任机制，以促进团队协作和沟通。通过开展 Prompt 工程、AI 运维、MLOps/DataOps 等专项培训，帮助相关技术团队掌握相关技能，提升团队在 AI 开发、模型运维与数据治理等领域的技术储备。同时，建立 AI 伦理审查机制，将安全合规纳入开发流程，如在金融行业对风控模型进行审计，确保技术应用符合社会价值观及相关法规要求。

» 5.2.4 渐进式演进策略

应采用“分步导入”原则，优先改造高价值、低风险业务模块，避免全盘颠覆式升级带来的业务中断风险。例如，可先行改造部分核心业务系统，并建立灰度发布与回滚机制；在小范围验证技术路线的可行性与稳定性后再逐步推广。通过微服务化与 API 治理，设计可扩展架构并构建架构感知中枢，实时可视化技术债，确保系统具备自我进化能力，能够适应快速变化的业务需求，并支撑 3-5 年的技术迭代周期，从而降低长期运维成本。

» 5.2.5 法律法规与伦理合规

企业应确保应用符合《数据安全法》《生成式 AI 服务管理暂行办法》等相关法律法规，并考虑设立 AI 伦理审查委员会，制定详细合规检查清单。通过部署模型监控工具（如幻觉检测、偏见消除），规避数据滥用与算法歧视等法律风险。应设计透明可控的 AI 决策机制，例如采用可解释性模型 XAI，并建立允许用户追溯 AI 决策逻辑的机制，使关



键场景的人工审核覆盖率达到 100%，确保 AI 行为透明、可控，提升技术的社会接受度，从而在合法合规前提下赢得用户与社会的信任。

应用现代化是企业在智能化时代实现转型发展的必然选择。通过“3A+6R”的实现路径，结合 AI 原生关键技术能力，企业可在提升用户体验、加速价值实现与优化成本等方面取得显著成效。在推进应用现代化过程中，企业需重点关注战略与业务价值对齐、用户体验重构、组织流程适配、渐进式演进策略以及法律法规与伦理合规等关键考量点，采取科学且合理的实施策略，平衡创新与风险，逐步构建智能、弹性且可持续的现代化应用体系，在数字化竞争中占据先机，实现业务价值的可持续增长。

06

应用智能化最佳实践案例

PART SIX



各行业积极探索应用现代化转型之路，并结合智能化的前沿技术落地实践，本章将对应用智能化面向各行各业的落地案例进行阐述和分析。

6.1 流程 IT 智能化应用升级

作为全球领先的 ICT 基础设施和智能终端提供商，公司关注数字化建设与变更。公司从 2005 年开始投入到 AI 数据与 AI 建设的相关工作中，构建了公司的统一数据仓库，并在供应链、制造、采购、财经、营销服务等领域试点用传统的决策式 AI 改变业务运作。流程 IT 部门不仅通过 600 多种智能应用高效支持公司产业，同时也紧跟 AI 大模型技术变革的机遇，利用 AI 技术辅助内部业务，在内部多个业务场景使能建设 AIGC 应用。在推进应用智能化升级的过程中，也遇到了如下挑战：

1. 大量小模型离散，共享模型占比只有 2% 左右。场景化定制导致 AI 应用的维护成本高。
2. 小模型主要应用于判断和感知，企业场景适用性优先，数字员工进入瓶颈。

» 解决方案

流程 IT 抓住 AI 大模型技术变革的机遇，通过 AI 辅助内部业务，提升各领域的作业效率和质量。聚焦价值场景，识别 AI 主战略，共同制定 AI 辅助内部应用的顶层框架；统一 AI 服务内部业务框架、AI 应用框架；统一平台基础设施，包含算力、模型能力服务化、货架化，明确 SLA 等。将经验总结为“三层五阶八步”转型方法论，其中：

“三层”是指重新定义智能业务、AI 开发与交付、持续运营智能应用，定义了转型所触及的层次和深度。

“五阶”是指场景、流程、组织、数据、IT 等五大阶段，能够帮助企业在转型过程中掌握顺序、抓住重点。

“八步”是指明确目标、场景识别、重塑流程、组织变革、数据和知识工程、AI 建模与发布、AI 融入业务应用、AI 持续运营等八项具体工作，是转型流程的深度细化。



图 6-1 AI 数字生产线业务框架图

6.2 某智慧家庭解决方案提供商提升运维管理效率，实现资源优化

某智慧家庭解决方案提供商致力于为亿万家庭构建“智慧生活”，该企业的大运维部承接了整个集团基础资源运维，以及集团办公软件的管理。随着整个集团数字化转型的加速，海量应用系统上线，如何保障平台稳定性和可靠性，助力数字化转型，成为客户的关键诉求。当前客户在平台的使用中存在以下挑战：

1. 日志分析系统性能不足，运维成本高：基于自建系统开发的日志查询和分析 API 速度较慢，导致定位问题等待时间较长，影响故障处理。
2. 日志分散，存储管理困难：客户公司内部业务比较多，各业务部门日志分散且总量较大，使用自建的日志存储系统对日志管理和汇聚难。
3. 开发运维成本高：自建日志管理系统，开发大量的 API 用于日志的查询、分析等，开发和维护成本会大幅增加。

» 解决方案

该企业可观测运维系统实现海量的日志管理，系统支持百亿级日志秒级搜索和千亿级日志迭代搜索，可实现计算资源的弹性获取。未来，客户也将持续基于智能全栈可观测平台，借助盘古智能运维助手，实现业务异常识别，告警降噪，以及根因分析诊断的功能，达成 1-5-10 运维目标。

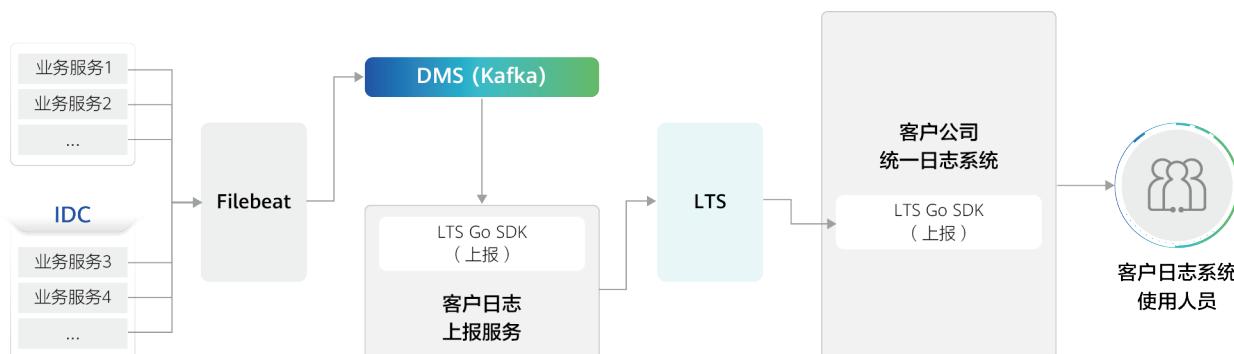


图 6-2 某智慧家庭解决方案提供商一站式可观测运维平台解决方案

6.3 社区平台打造训推一体化平台

某国内领先生活方式社区平台的核心商业模式是社区 UGC (User Generated Content) 和电商结合。目前，该平台 MAU (Monthly Active User) 超过两亿，汇聚了全球数十万品牌，千万级的用户，每日新增百万条笔记，月活高

达亿级别。庞大的用户群体每天要产生上亿级的训练样本，TB 级的数据量。如何提升开发运维效率，快速迭代，实现海量内容高性能查询分析，秒级扩容，应对流量洪峰，提升大数据的实时分析能力，是该平台需要解决的问题。

» 解决方案

大数据存算分离：针对客户业务场景，构建高弹性能力，提升系统整体业务容量。通过大数据存算分离、容器化改造等实现业务和数据架构升级，通过技术降本增效。实现架构优化，综合成本下降 30%，资源使用率提升 40%。大数据提效，日志分析场景性价比提升 90%。

容器批量计算：搭建统一的 Volcano 调度架构，用户只需调用对应 API 即可便捷地在同一个集群内下发大数据或者机器学习任务，提升 AI 训练和内容审核效果。AI 训练性能提升 30%，算法模型更新时效从天级优化至分钟级。运维团队压力释放，运维节约成本 30%。



图 6-3 某生活方式社区平台搭建大模型训练推理一体化平台

6.4 CodeGenie: 聚焦鸿蒙应用开发场景的智能化研发助手

随着鸿蒙星河版正式发布，原生应用全面启动。鸿蒙生态进入快车道，各应用正在加速适配开发，迁移工作量巨大，研发人员对支持鸿蒙开发语言的智能化开发助手工具诉求强烈。

» 解决方案

DevEco Studio 联合小艺，推出 DevEco CodeGenie AI 辅助编程工具，为开发者提供高效的应用 / 元服务 AI 辅助编程能力，支持 HarmonyOS NEXT 领域的智能知识问答、ArkTS 代码补全 / 生成和万能卡片生成能力，提高开发者编码效率。

DevEco CodeGenie 致力于使开发者编码效率更高（持续提升采纳率），编写的代码性能更优（提升鸿蒙最佳实践检测覆盖率），让开发者只用聚焦于核心业务代码逻辑实现（多模输入生成 ArkUI 界面）。



图 6-4 CodeGenie: 聚焦鸿蒙应用开发场景的智能化研发助手

6.5 智驾 AI 算力平台，保障智驾业务

智能驾驶一直是汽车行业的热门话题，ADS 智驾系统也一直是问界、享界等系列电车的亮点。在云助智驾场景，从商用生产到研发端到端的数据闭环过程中，可能会面临以下问题：例如智驾过程对体验要求高、端云处理耦合重、故障爆炸半径不容易控制；智驾数据飞轮从车端驱动，迭代周期较长；智驾数据规模大，类型多，难以管理和资产转化；泊车过程通常为弱网环境，视频体验不流畅。

» 解决方案

开展云上架构优化改进、确定性运维专项保障，并通过汽车专区 ADS 云可靠性提升、最佳智驾 AI 算力平台、最佳智驾数据湖、自动化标注、智驾确定性运维 5 大核心方案，助力车 BU 实现 ADS 云车辆规模接入、提升智驾能力。

目前智驾能力已经贯通全场景，当智驾一键开启时，ADS 云提供实时的交互处理，确保智驾功能端到端高可用。当高阶城区智驾穿行于闹市街区时，云端大规模智算资源池支撑极致训练效率和迭代下的模型使得智驾更类人。当大规模数据上传到云端时，湖仓一体解决方案使能百 PB 级数据高效治理。当泊车代驾只需车主隔空远观时，实时流视频加速让车主可实时查看车辆行驶轨迹。



图 6-5 实现智驾算力上云

6.6 新疆某油田基于大模型构建统一知识库

新疆某油田作为我国重要的能源基地，其信息化建设一直走在行业前列。为提升当地各委办局的数字化办公水平，该基地搭建了统一的政务云平台，为政府部门提供基础云服务与 IT 技术支持。但在实际管理过程中，各委办局普遍面临知识管理难题：政策文件、技术文档等分散在不同系统中，纸质档案与电子文档并存，工作人员需要耗费大量时间进行资料检索，严重影响办公效率。特别是涉及油田安全生产、能源政策等专业领域时，传统检索方式难以快速获取精准信息。

» 解决方案

基地引入 AI 大模型技术，重点构建智能政务知识库系统。该系统将分散在 OA 系统、档案库、PDF 文档等各处的政务知识进行结构化处理，通过大模型的自然语言处理能力，实现“政策一键查”“规程智能问”等功能。例如安全监管等部门可通过自然语言提问，即时获取最新的油田安全操作规范。利用大模型构建知识库，不仅解决了知识碎片化问题，更通过 AI 技术将传统文档转化为可交互的智能知识资产，使各委办局的公文处理效率提升 40% 以上，成为政务应用现代化的标杆案例。



图 6-6 新疆某油田统一知识库架构

6.7 云原生技术赋能香港智慧医疗，构建高效服务新生态

香港医院管理局（HA）是负责管理全港多家公立医院及相关医疗服务的香港特区政府机构。医管局自主研发的临床管理系统历经 25 年以上的发展，已实现全院级部署。面对医疗需求增长，系统承载产品量预计从 2015 年 900 项增至 2027 年 1800 项，支撑智慧医疗发展。目前，香港医院管理局面临数据治理及跨机构协作等挑战，需要突破信息系统扩展性，以应对数字化转型压力。

» 解决方案

医管局通过云原生技术驱动医疗数字化转型，构建弹性可扩展的智慧医疗生态，实现成本优化与服务质量跃升。AI 与数据能力赋能临床创新，最终构建以患者为中心的全场景数字化医疗生态，达成如下效果：

- 降低研发成本。通过技术架构升级实现标准化、模块化，通过优化技术栈复杂度消除遗留系统负担，实现开发成本降低约 25%，显著减少资源投入。
- 提高软件质量。引入测试驱动设计（TDD）与自动化测试，系统性提升代码健壮性，同时优化高可用架构实现分钟级容灾能力，整体软件质量跃升约 40%。
- 提升交付速率。依托 CI/CD 全链路集成与云端自动化流水线，结合敏捷发布模式，从月度升级至每周高频迭代，交付效率提升约 3 倍，加速业务需求响应。



图 6-7 香港智慧医疗云原生解决方案架构图

6.8 软通动力携手同济大学打造云原生本研一体化教务系统

同济大学作为国内顶尖高校，以培养拔尖创新人才作为使命和责任。学校的教学管理系统需支撑超 5 万名在读学生、近 5 千名教师的全生命周期管理。原系统存在本研业务分离、架构僵化、扩展性不足等问题，选课等高峰场景曾

因并发压力导致服务卡顿，运维效率低下，难以满足教育数字化转型需求。目前教务系统在使用过程中存在以下问题：

1. 本科生与研究生教务系统独立，数据互通困难，管理成本高。
2. 性能瓶颈：传统单体架构无法应对选课等瞬时高并发，扩容效率低。
3. 运维低效：故障排查依赖人工，升级需停机，影响师生体验。

» 解决方案

软通动力基于全栈技术，构建云原生本研一体化教务系统中台：

1. 微服务重构：拆解学籍、选课等 50+ 核心模块为独立微服务，支持本研业务灵活组装，系统耦合度降低 80%。
2. 弹性扩容：通过 CCE 容器服务实现秒级资源扩展，选课高峰期承载能力提升 2 倍。
3. 智能运维：集成 APM 监控和自动化治理工具，故障诊断效率提升 50%，支持无感升级。



图 6-8 软通动力携手同济大学打造云原生本研一体化教务系统

系统上线后支撑 865 万次选课、7.8 万次成绩录入，资源利用率大幅提升，应用部署从小时级缩短至分钟级。项目成为高校教务云原生转型标杆，助力同济大学实现教学管理效能的数字化跃迁。软通动力以金融级高可用架构经验赋能教育行业，未来将持续输出全栈应用现代化能力，推动智慧校园建设。

目前围绕新生入学等场景，软通动力正在联合同济大学构建 AI 智能迎新解决方案。该方案基于 AI 智能体、主线支线任务动态分配，为新生打造智能化、个性化的数字迎新服务；基于大语言模型的问答系统、语义理解精准解析，支持复杂场景的新生问题查询；基于知识图谱整合校内资源，通过协同过滤算法个性化推荐生活服务。智能化新生迎新系统的推进将会减少人工咨询量，提升系统响应速度，语义搜索可以解决大部分新生常见问题，高效便捷。后续系统将接入 AI 选课助手、智能缴费等更多功能。



如何加入联盟

请通过以下方式联系我们

联盟对接人：李丹丹

邮箱：Linda.lidandan@ami-alliance.org.cn

官网地址：<https://www.ami-alliance.org.cn>



应用现代化产业联盟官网



扫码申请加入联盟