

# 从提示 到用途

以智能体AI释放商业价值

如何自动化、优化和扩展代理式AI

Capgemini  ×  snowflake®



# 引言

在企业AI的讨论中，大部分的注意力——以及炒作——都集中在生成式AI和大型语言模型（LLM）：以ChatGPT为代表的AI，以及它的文本写作能力、图像创建能力、代码生成能力、音乐创作能力等等。

智能体AI正在迅速改变对话。根据  
[嘉信的近期研究](#)，

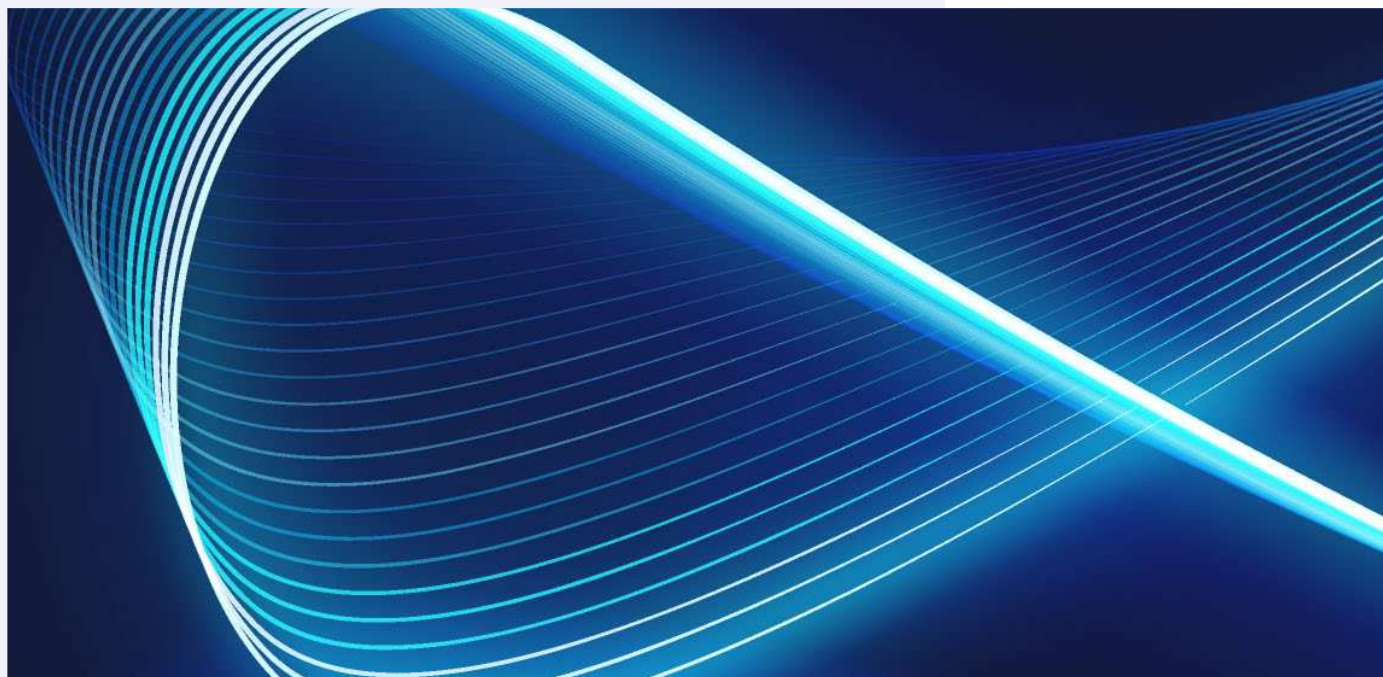
<sup>1</sup>  
93%的领导者认为在未来12个月内成功大规模部署代理式AI将带来竞争优势。的确，采用的速度如此之快，以至于研究人员感到有必要与大多数受访者再次核实。

近期推动这种采用的因素是客户服务、IT、销售、运营以及研发等流程密集型领域。随着组织寻求提升生产力、效率和增长的新水平，预计到2028年，自主智能的价值将达到4500亿美元。

<sup>2</sup>

然而，实现这一价值可能具有挑战性。强大的模型是基础，但为了便捷、连接性和信任，还需要构建一种有效的代理式AI解决方案。这需要系统设计优先考虑无缝集成、清晰的治理框架和可重用、可组合的架构，以及对准备就绪的AI数据的访问和治理。同样重要的是，一个业务用户界面，它允许技术和非技术人员都能高效地进行规模化开发和监控。

在本文中，我们探讨了成功的技术前提条件、最常见的行业障碍，以及正确的平台方法如何解决数据碎片化、不可预测的成本和交付价值的时间问题。我们还提供了一个理想解决方案的实用清单，并分享了现实世界中成功的案例。



<sup>1</sup> <https://www.capgemini.com/wp-content/uploads/2025/07/Final-Web-Version-Report-AI-Agents.pdf>

<sup>2</sup> <https://www.capgemini.com/wp-content/uploads/2025/07/Final-Web-Version-Report-AI-Agents.pdf>

# 什么是自主AI？

自主型人工智能是一种能够自行设定目标、做出决策并采取行动以实现目标，而不需要任何特定指令说明的系统。它由多种不同类型的智能体协同工作，以收集、理解和处理信息。它还能通过自身的行动进行学习，以改进其响应，无需人工干预。

智能代理可以包括机器人流程自动化（RPA）、机器学习模型，以及能够生成、验证和总结内容的生成式AI代理，作为更广泛任务的组成部分。例如，Snowflake Cortex Agents是生成式AI代理，通过分解复杂查询、检索相关数据以及生成精确答案，帮助从复杂的结构化和非结构化数据集中检索数据洞察。

**37%**

在试点或实施 AI 代理的组织中

**25%**

到2028年，实现业务流程的半自动化到全自动化运营

**\$450**

到2028年，由AI代理在14个国家产生的预计经济价值达百亿



## 成功代理式AI的关键特性

在其本质而言，代理式AI可以部署于广泛的应用场景。无论一个组织是实施它来个性化零售体验，加速临床试验，或检测和预防金融欺诈，大多数成功的实施都基于同样的关键特性。

在准备其组织并考虑解决方案和合作伙伴时，考虑代理式AI的企业和技术领导者应该因此寻找包括在内的关键成功因素：

01

获取统一、适用于人工智能的数据：为了获得全面的情况，人工智能代理需要无缝访问来自所有相关来源的完整、一致和连接的数据。没有它，代理式人工智能无法智能地行动或做出相关决策。

02

自优化的AI管道：代理式AI学习、适应和扩展。自优化通过将反馈转化为改进，使得这一切成为可能，这保证了独立代理的安全、有效和按计划进行。

03

实时数据处理：AI代理即时行动。它们需要实时数据来理解当前正在发生的情况，并做出及时决策以确保它们达成既定目标。

04

自动化数据治理：人工智能代理必须遵守规则。强大的治理和安全框架确保运营表现符合道德和法律，并为企业提供对自动化劳动力的监督。

05

可组合ai工作流：在一个智能体ai系统中，没有一个智能体单独工作。可组合工作流让企业混合、匹配和重用ai组件。它是实现真正可扩展性和成本效益的关键。

06

优化计算能力：AI代理处理大量数据并快速做出决策。这是一项计算密集型的工作。高效的计算能力可保持高性能并控制成本。

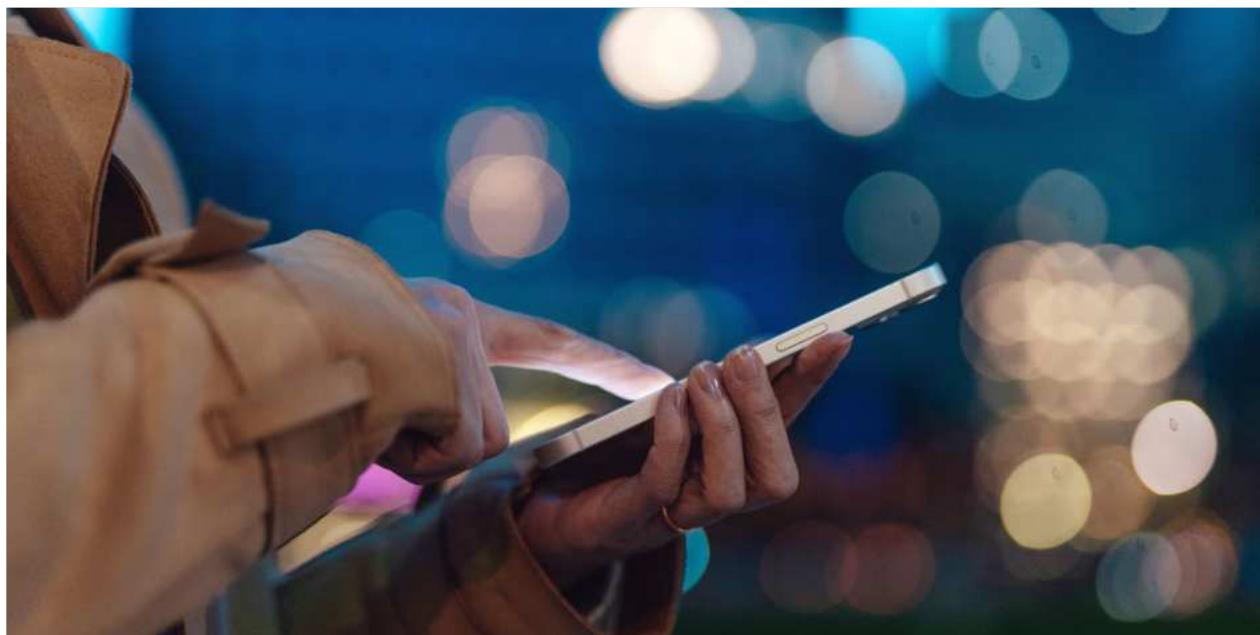
# 三个代理式AI价值的热点行业

## 1. 电信

- ▶ 检测和自愈网络故障：AI代理实时监控、诊断和解决网络问题。
- ▶ 管理客户生命周期：基于实时活动、行为和账户状态的超个性化客户体验，可提升客户旅程。
- ▶ 监控监管合规：AI代理持续扫描广播内容及用户交互，以标记或自动调整针对新兴的内容审核与数据。

## 2. 制造和供应链

- ▶ 优化生产线：AI代理监控需求波动、设备性能和材料限制，然后根据需要调整工作流程、机械和供应路线——实时进行。
- ▶ 追踪环境绩效：AI代理跨供应链收集、分析并报告相关数据，以实现实时合规、与可持续发展目标保持一致以及高效的ESG报告。
- ▶ 预测性维护设备：AI代理不仅标记异常，还规划并安排维修、订购零件，并协调维护团队以防止维护不足和过度维护。



## 3. 零售和电子商务

- ▶ 增强个人购物：AI代理跨多个会话和渠道协助单个客户，策展推荐，管理愿望清单，并协商报价。
- ▶ 管理活动：代理式人工智能根据实时绩效数据设计和启动营销活动，然后根据需要通过调整创意、受众细分和时间来优化它们。
- ▶ 优化退货和退款：智能代理处理购买后的任务，例如处理退款、提供换货、标记滥用或更新库存，近乎实时。

# 最常见的五个障碍和瓶颈

正如存在对代理型AI成功部署至关重要的共同因素一样，也存在常见的障碍。数据碎片化、复杂的治理、成本低效以及可扩展性是实现代理型AI价值的最大障碍。

一个面临以下一个或多个警示信号的项目，其失败的可能性远大于一个没有这些警示信号的项目。

01

**复杂的、低效的扩展：**当实施代理式AI时，许多企业发现他们的时间都花在了“重新发明轮子”或添加很快变得冗余的组件上。由于他们没有标准管道和可重用组件，他们的资源被浪费，数据保持孤岛化，成本上升，最终他们的AI未能达到预期。

02

**缺乏可组合性和治理：**代理式人工智能可以自主行动，但仍需商业领袖进行强有力的监督。如果没有健全的道德框架和明确的限制措施，代理可能会以产生实际业务损害的方式追求目标。除了失控的代理之外，薄弱的监控也打开了数据安全的大门，并使企业面临监管风险。

03

运营壁垒和集成挑战：AI代理旨在相互协作，并与人力共同工作。但断开的AI工具是糟糕的合作伙伴，难以在职能间合作。企业最终重复工作，这减缓了采用进程。转型受阻，企业最终只得到一个更昂贵的“按部就班”版本。

04

不可预测的成本和不一致的表现：人工智能模型会随着时间的推移而退化并且性能下降，就像一个有价值但过度劳累的员工一样。但是，与人工智能代理相关的问题可能要难得多才能发现和修复。优化是必要的，但当项目对性能和支出没有清晰的可见性时，这仅是一种猜测。

05

数据隐私和安全：AI在云端运行效果最佳。但许多企业，特别是那些高度监管行业的企业，持有敏感数据，需要保留在本地。在计算效率与数据隐私和合规控制之间取得平衡是一个持续的挑战——并且可能成为从自主AI中获得实际商业价值的一个重大障碍。

在最好的情况下，这些挑战会减缓价值实现的速度。在最坏的情况下，它们会完全破坏价值实现。计划中的业务效益无法实现，并且AI获得必要信任所需的时间会更长。

## 一种新型人力资源挑战

代理式AI的挑战，对任何领导过高级数据驱动系统和解决方案实施的人来说都耳熟能详。但HR挑战揭示了一种独特的AI风味。

考虑云和移动。这些技术通过协助人类工作者使商业生活更轻松。但代理式AI更接近一个虚拟人类。它不仅仅是提供帮助；它根据被授予的权限独立行动。它可以在更快的速度下完成现有任务，或者承担人类无法完成的全新角色。监控和管理被赋予了新的意义。正如人类劳动力需要一个在职场中明确期望、角色和职责的人力资源框架一样，AI代理也需要一个治理和控制结构来管理其活动，以主动避免安全、合规或伦理风险。

# 集成化、优化和可扩展：理想的智能体AI解决方案

尽管技术复杂，有效的代理式AI解决方案基于六个简单品质：易用性、信任、可重用性、效率、可扩展性和可预测性。这正是为什么在解决方案设计中有必要解决数据碎片化、治理复杂、运营低效和不可见的成本。它们削弱了成功所依赖的品质。

对于商业用户而言，这可以解决这些挑战。与其说是孤立的AI实验，这种集成式解决方案提供了一个可组合、全面受管理的企业资产。它解锁了成本效益高的AI工作流程，确保了更强的安全性，并阻止了组织在将代理式AI应用于更多流程和领域时不断重复活动和组件。

集成工业级AI代理和有效数据平台，并设计了界面的自主AI解决方案

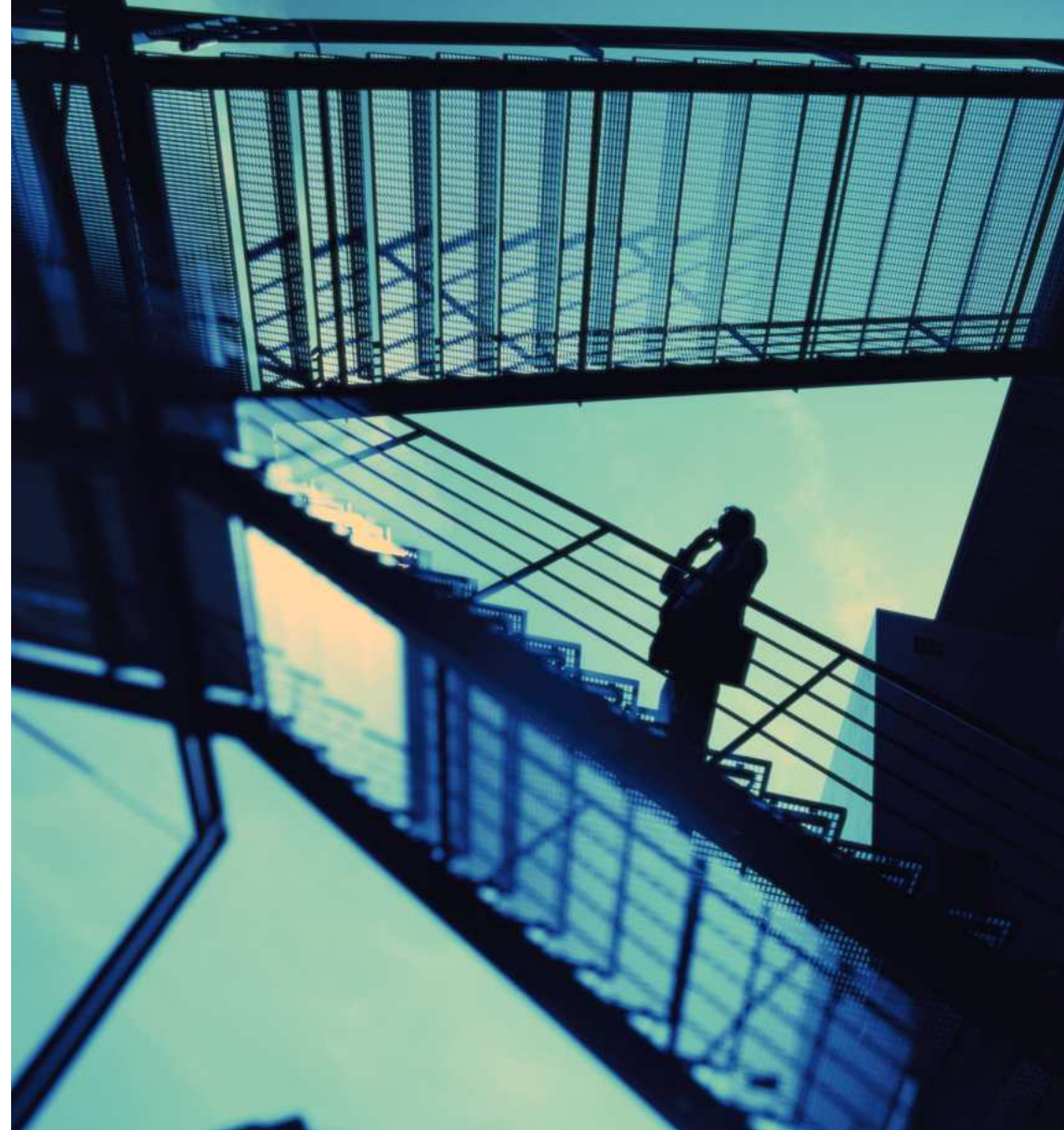
事实上，随着更多业务功能和任务被整合，每个组件的成本就会越低。由Snowflake支持的Capgemini AI可扩展平台(通用人工智能/自主人工智能)是一个已证明这种方法价值的集成解决方案：

- ▶ 至 70%<sup>4</sup> 与单个庞大的LLM相比，每项应用的开发成本更低，成本监控、提示优化和缓存确保财务透明<sup>3</sup> 考虑到使用平台加速器启用的充分生产用例。
- ▶ 60%<sup>4</sup> 与断开连接、各自为政的构建相比，能够更快地部署，这得益于统一的治理、运营仪表板以及用于简化人工智能采用的解决方案加速器。
- ▶ 40%<sup>4</sup> 与手动监控相比，利用模型评估、版本控制和比较工具更快地发现问题，以增强AI的可靠性。

随着智能体式AI的采用加速，像由Snowflake支持、可简化对可重用组件的访问和管理智能体操作的Capgemini AI可扩展性平台这样的解决方案，对于快速扩展和最大化价值至关重要。

<sup>3</sup> 基于使用由Snowflake平台加速器支持的Capgemini AI可扩展性平台实现的充足生产用例。

<sup>4</sup> Capgemini 人工智能可扩展平台：在代理时代开创新生成式人工智能工厂



# 金融服务中的集成智能体AI

## 从有限的扩展性到更快的价值实现

由于数据碎片化和复杂的监管治理要求，银行中传统的人工智能实现可能很慢、成本高昂且难以扩展。结果，在欺诈检测、信用风险建模和客户个性化等领域进行的创新往往受到阻碍。

部署由 Snowflake 支持的 Capgemini AI 可扩展性平台，通过标准化人工智能工作流程来帮助银行克服这一挑战，从而创建一个模块化、可重用且——关键在于——可扩展的人工智能解决方案。包括风险、运营和面向客户的部门在内的各个功能领域都可以访问一个一致的工具来协同创新，同时受益于可预测的人工智能性能，从而更快地实现价值。

## 从不可预测的成本到优化的扩展

许多银行发现，人工智能工作负载增加了计算能力需求，给它们的金融建模和风险分析带来了显著的成本不可预测性和低效性。

通过部署由 Snowflake 支撑的 Capgemini AI 可扩展性平台，与传统金融 AI 模型相比，可最高削减高达 80% 的 AI 模型开发成本。Snowflake 的按需计算模式确保银行仅为其使用部分付费，而 AI 可扩展性平台优化 AI 处理。此外，提升控制力、透明度和可预测性有助于增强组织对 AI 驱动的银行业务运营的信任。

# 医药行业的集成代理式AI

## 从断开的数据到更快的药物发现

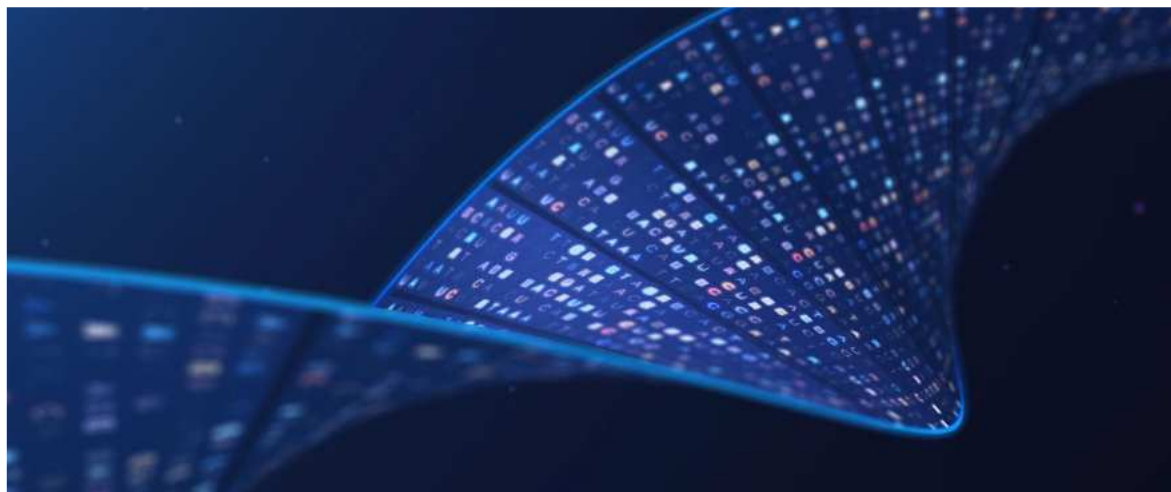
对于希望利用人工智能驱动研发的制药公司而言，数据孤岛仍然是一个重大障碍。临床数据、基因数据和患者数据通常存储在组织内相互隔离的系统中，限制了人工智能可以分析的数据量，并增加了人工智能赋能创新的成本和上市时间。

部署由 Snowflake 赋能的 Capgemini AI 可扩展性平台等集成解决方案，可统一这些数据，从而显著加速 AI 赋能的药物发现。Snowflake 的云原生 AI 数据平台无缝整合了来自传统系统的结构化、半结构化和非结构化数据。AI 可扩展性平台加速了对由 Snowflake Cortex AI 代理访问、分析和处理这些统一数据的重大发现过程的开发与监控。

## 从延迟分析到持续洞察

Precision medicine relies heavily on data, and innovation in this emerging field can be severely limited by delayed data processing. Real-time patient monitoring faces the same critical challenge.

一种集成解决方案，例如由 Snowflake 支持的 Capgemini AI 可扩展性平台，克服了这一障碍，能够显著提高按规模高效进行个性化治疗的能力。Snowflake 通过摄取能力（包括 Snowflake Openflow、Snowpipe Streaming、Kafka 连接器和动态表）使实时数据可用。AI 可扩展性平台促进构建、编排和监控基于 Snowflake Cortex AI 的解决方案，为医疗保健专业人员提供即时按规模访问个性化预测诊断的能力。



# 企业在规模上的 AI 价值发生阶跃式变化

自主型人工智能是企业大规模实施人工智能方式的一次重大转变。集成代理和数据模型同样是一步意义重大的进展。这些模型是为了易于扩展、精准运行和控制成本而开发的，它们是值得信赖的解决方案，以产生商业价值。一次又一次地证明了这一点。

## 人工智能可扩展性平台

思爱普的企业级代理式人工智能基础架构使组织能够超越实验和概念验证，跨业务规模化部署代理式人工智能——同时控制成本，降低IT和业务风险，并通过人机化学反应交付可衡量的成果。借助雪flake的统一人工智能数据平台以及Cortex Agent和文档智能等先进功能，我们加速规模化采用安全、可投入生产的代理式人工智能。

清单：您的代理式AI解决方案应提供：

- ☑ 轻松、高效且值得信赖的 AI 开发，具有端到端集成。
- ☑ 管理数据访问以满足开箱即用的准确性、信任和安全。
- ☑ 针对特定业务环境的可定制代理。
- ☑ 面向数据团队和人工智能团队设计的用户友好界面。
- ☑ 可重用、可组合的AI workflow，用于快速创新。
- ☑ 持续且自动的部署以实现敏捷性和最小化停机时间。
- ☑ 跨数据与AI资产的一体化治理、可观测性与安全。
- ☑ 跨用例的可预测 AI 放大以降低可变成本。
- ☑ 本地直接访问位于数据所在地的LLMs和AI服务。
- ☑ 实时洞察，助决策更智能、更快。

发现由雪花支持、集成、优化和可扩展的代理式AI，以及由雪佛龙支持的可扩展AI平台。

达文西人工智能可扩展平台是达文西的商标。

# 作者



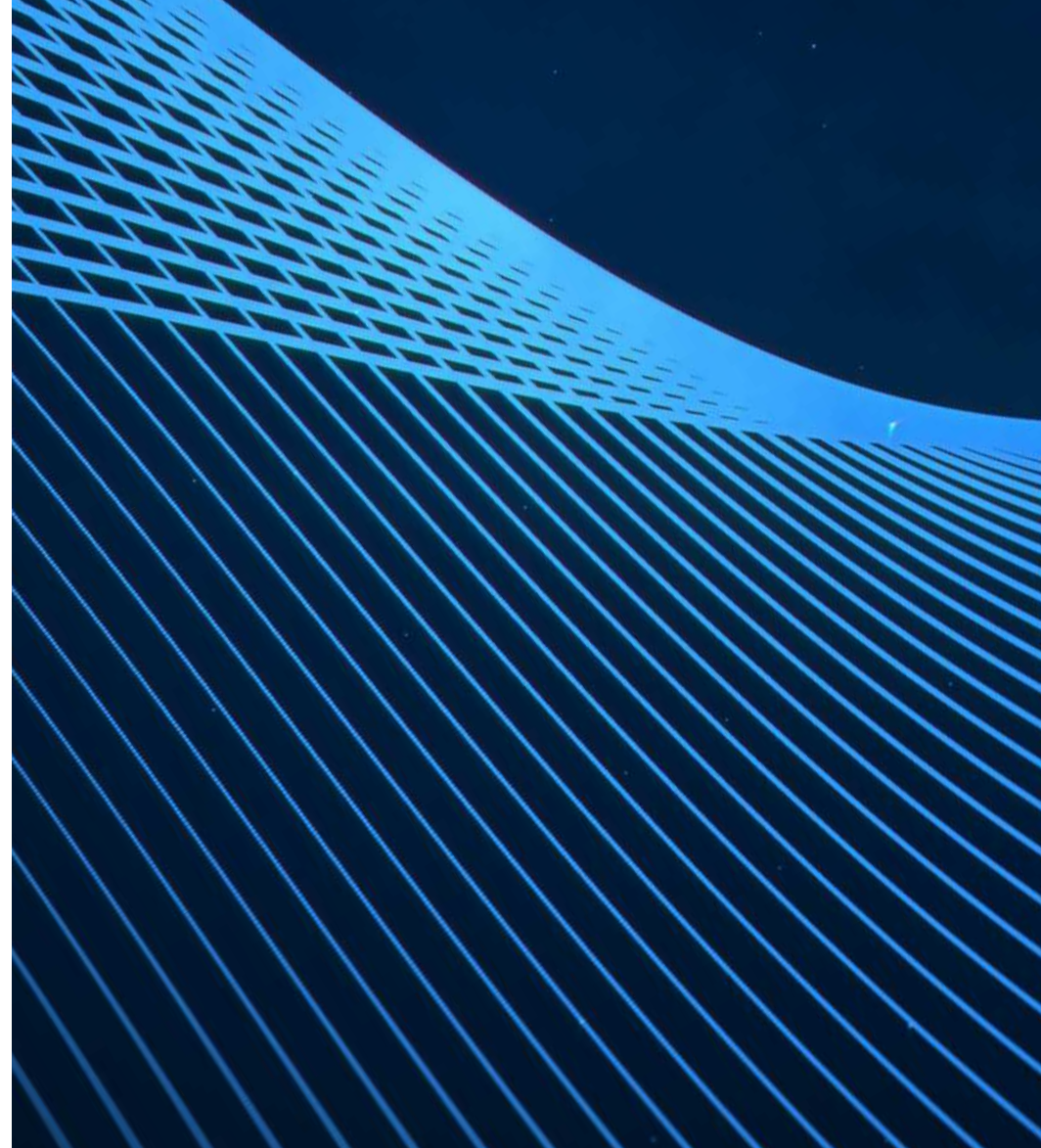
帕拉维·夏尔马

高级合作伙伴解决方案工程师，Snowflake



埃里克·雷希

提供领导者以及全球人工智能与数据工程、洞察与数据负责人，嘉能可



## 关于雪花

雪花是人工智能时代的平台，使企业能够更快地创新并从数据中获得更多价值。全球超过11,000家公司，包括数百家世界最大的公司，都在使用雪花的AI数据云来构建、使用和共享数据、应用程序和人工智能。有了雪花，数据和人工智能对每个人来说都是变革性的。了解更多信息请访问 [snowflake.com](https://snowflake.com) (NYSE: SNOW)

## 关于埃森哲

凯捷是一家以人工智能为驱动的全球商业与技术转型合作伙伴，为客户创造切实的商业价值。我们构想组织的未来，并借助人工智能、技术和人员将其变为现实。凭借近60年的深厚传承，我们是来自50多个国家的420,000名成员组成的负责任且多元化的集团。我们凭借深厚的行业专业知识和强大的合作伙伴生态系统，提供端到端的服务和解决方案，发挥我们在战略、技术、设计、工程和业务运营方面的能力。该集团报告2024年全球收入为221亿欧元。

让它真实 | [www.capgemini.com](https://www.capgemini.com)

