

The background features a dark blue field with intricate, flowing lines in shades of orange, red, and teal. These lines swirl and curve around the central text, creating a sense of dynamic movement and connectivity.

人工智能对广域网的 影响：思科报告2026



执行摘要。
..... 4 主要发现.....
..... 4 关键点。
..... 7 引言.....
.....
..... 8 数据分析与发现.....
..... 9 数据量增长.....
..... 9 数据流长度.....
..... 11 数据流速。
.....
..... 12 流量不对称性。
..... 13 网络延迟影响.....
..... 14 智能体
AI流量模式。
..... 15

展望未来。

代理式AI的企业应用与网络影响：2026-2035..... 18

人工智能的消费使用与自主人工智能及网络影响：2026-2035 21

关于报告

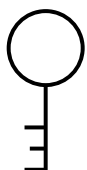
《人工智能对广域网的影响》报告代表了由思科系统公司牵头的研究计划的关键发现和要点。该报告结合了真实世界的流量分析（使用思科Crosswork Assurance用户体验工具）、第三方行业数据，以及思科控制的AI代理实验室测试。它探讨了AI驱动流量与传统非AI网络流量的差异，重点关注代理AI流量的影响以及对网络设计、容量规划和保障的启示。

使该报告与众不同的是，它基于真实的交通数据，其中包括对代理式AI交通的早期洞察（目前规模虽小但增长迅速），这使我们能够开始观察和测量一种新型AI网络流量，并理解其影响。报告包括：

- 跨越服务提供商网络直接测量实时AI推理流量
- 一个可重复的测量框架和基准，用于年度跟踪AI流量演变和预测
- 对AI交通特征进行实证测试，以训练能够精确识别和追踪AI流量的模型

执行摘要

人工智能（AI）的快速采用——尤其是AI推理和代理式AI系统——正以可衡量且结构性的方式开始重塑网络流量模式。到2027年，80%的高管认为其公司的竞争优势和生存将取决于代理式AI，¹而消费者对AI的使用已十分普遍且正在加速。这正推动着流量生成、分发和体验方式的根本性转变，这对管理大规模网络的服务提供商和企业具有影响。



关键发现

连接的未来已不再仅仅是下载速度的提升——它关乎智能体与AI模型之间至关重要的“脊柱”，这是一项关键依赖，正在开辟新的通信路径，并影响我们管理网络的方式。尽管行业数十年来一直致力于优化适合人类节奏、突发式视频流的网络，但自主AI的兴起正在改变网络流量特征和行为。到2035年，预计网络流量的四分之一将是AI推理（来源：思科模型）。这些流量不像互联网那样表现。它们的生命周期更长，需要更多的上游容量，并且以软件速度而非人类速度运行。代理逻辑与AI模型之间的连接实际上成为了代理的“脊柱”——这是一个关键依赖，任何网络退化都会直接损害代理的功能。

1 思科报告 《主动型AI的竞赛：为何基础设施将决定或破坏劳动力整合》 <https://www.cisco.com/c/dam/en/us/solutions/collateral/artificial-intelligence/race-to-agentic-ai-report.pdf>



代币消耗数据同比增长10倍



今天的AI推理流量不大——但增长速度极快。

尽管与视频流等主要类别相比，AI 推理流量仍然可以忽略不计，但观察到的增长率却非常显著。令牌消耗数据显示，同比增长近 10 倍，而在一些服务提供商的测量中，我们仅看到八个月就实现了约 4 倍的增长。以这些速率持续增长意味着，到 2035 年，AI 流量将成为整体网络流量中的一个重要组成部分。

AI推理流程为2倍



更长

网页流量与AI流量相比，流速大10倍



AI推理流量在传输层面与典型的网络交易表现不同。

- AI推理流持续时间大约是后者的两倍，这主要是由于逐个token的内容生成所致。
- 普通网络交易的流量中位数是 AI 推理流量平均值的 10 倍，后者具有更平稳的持续吞吐量，而非突发式传输。

这些特性将要求网络和安全系统（包括防火墙、入侵检测和深度包检测（DPI））随着时间的推移，能够处理更大规模、分布式以及更高的状态效率。

9%

upstream-heavy

AI推理流程

交通对称现状正在改变。

由于大型、内容丰富的提示，AI推理——尤其是具有自主性的AI——引入了显著更多的上游流量。

- 大约 9% 的 AI 推理流量上游流量多于下游流量，而典型网页流量中这一比例约为 0.5%。

随着代理机构维护和传递日益增长的环境状态，这一趋势预计将加剧，这对接入规划及上游容量建模具有直接影响。



#s ~ 5秒用于AI推理延迟，而网络延迟约为20-50毫秒

网络延迟目前还不是主要的瓶颈——但还早着呢。

端到端人工智能推理延迟主要由模型处理驱动，通常在数百毫秒到数秒之间。当前网络延迟已经足够小，尚不足以完全证明纯粹为了降低延迟而进行大规模架构转向边缘推理的合理性。推理硬件的改进将使网络延迟变得更为关键，我们已看到相关证据。衡量和保障人工智能推理体验，对于感知到的服务质量正变得至关重要。

AI代理总流量提升了450%。



70%

该流量由AI推理产生



代理式AI从根本上放大了网络使用。

AI代理充当网络“高级用户”，其运行速度为软件速度而非人类速度。实证测试表明：

- 每个任务由代理执行时，可产生高达450%更多的总流量。
- 大约 70% 的流量是 AI 推理，因此推理路径至关重要。

智能体逻辑与AI模型之间的连接性，实际上已成为智能体的“脊髓”——这是一个关键依赖关系，任何网络退化都会直接损害智能体的功能。

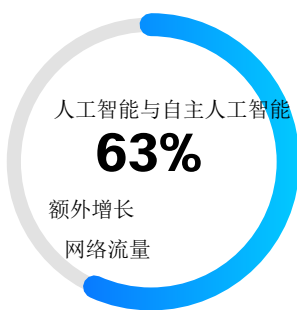
到2035年的AI交通影响* 企业网络：

若没有自主AI，预计未来十年企业流量将增长约2.5倍。而随着自主AI的采用，在自主任务执行和推理密集型工作流的驱动下，流量增长可能达到约9倍。

人工智能与自主人工智能

63%

额外增长
网络流量



消费者网络：

消费者对人工智能的采用正在加速，并将在2030年代中期达到近乎普遍的使用。预计人工智能和代理式人工智能将使消费者驱动的网络流量增加约6.6倍，与非人工智能场景相比，这代表约63%的额外增长——使其成为整体互联网流量扩张的主要驱动因素。预计到2035年，人工智能推理流量将占网络总流量的25%。

思科模型

要点

- 人工智能和具身人工智能不仅会增加流量——它们将改变流量的形态、对称性、持续时间和关键性。
- AI推理路径将成为战略性网络资产，需要具备高水平的韧性、可观测性以及差异化处理，例如服务质量（QoS）和路径安全。
- 传统的规划假设（突发性、下行链路主导、人类主导的交互）将需要适应人工智能和智能体AI流量的新现实，这种流量持续时间更长，对上游容量需求更大，并以软件速度运行。
- 持续监测AI特定的关键绩效指标（KPIs）并自适应网络架构，对于维持性能和用户体验的期望至关重要。

本报告标志着对人工智能对网络流量影响持续研究的开端。随着人工智能在企业界和消费者中的应用日益深化，未来的研究工作将完善预测、验证新兴模式，并指导运营商发展其网络，以适应人工智能驱动的世界。



引言

人工智能（AI）及其应用（如AI代理）的采用正以非常快的速度增长。在一项由思科（Cisco）与奥迪亚（Omdia）联合发布的报告中，80%的高管认为，到2027年，自主AI对于公司生存至关重要²。其他行业报告也证实了企业和消费者AI应用具有类似的增长预期。在这种情况下，服务提供商和运营大规模网络的组织需要评估AI的普及将如何从流量增长、模式或其他特征等方面影响网络。

本报告是年度系列报告的第一份，我们将在此分享关键发现与启示，以帮助网络领导者就其网络基础设施的演进做出决策。思科已启动针对AI流量模式及其与非AI网络交易差异的研究，尤其关注代理型AI流量与应用。本报告呈现的数据基于思科Crosswork Assurance用户体验的实时流量数据分析、思科实证测试以及第三方报告。

2 思科报告 《主动型AI的竞赛：为何基础设施将决定或破坏劳动力转型》 <https://www.cisco.com/c/dam/en/us/solutions/collateral/artificial-intelligence/race-to-agentic-ai-report.pdf>

数据分析与发现

数据量增长

网络运营商正试图回答的首要问题是：“将存在多少AI流量？”以及“这种流量将如何影响我们的网络？”

在此情况下，AI流量指的是在请求模型推理的设备与托管模型并进行推理的数据中心或位置之间跨越广域网（WAN）流动的AI推理流量。在数据中心内部，AI模型训练期间会产生TB级别的流量，但这些流量不会跨越广域网，除非是为了AI训练目的而进行的大规模数据传输。

与视频流媒体等其他类型的流量相比，AI 推理流量的实际量仍然可以忽略不计，在短期内相对而言很可能保持非常低的水平。然而，其增长率却非常显著：

2024年12月至2025年12月，OpenRouter AI的令牌消耗量同比增长10倍。

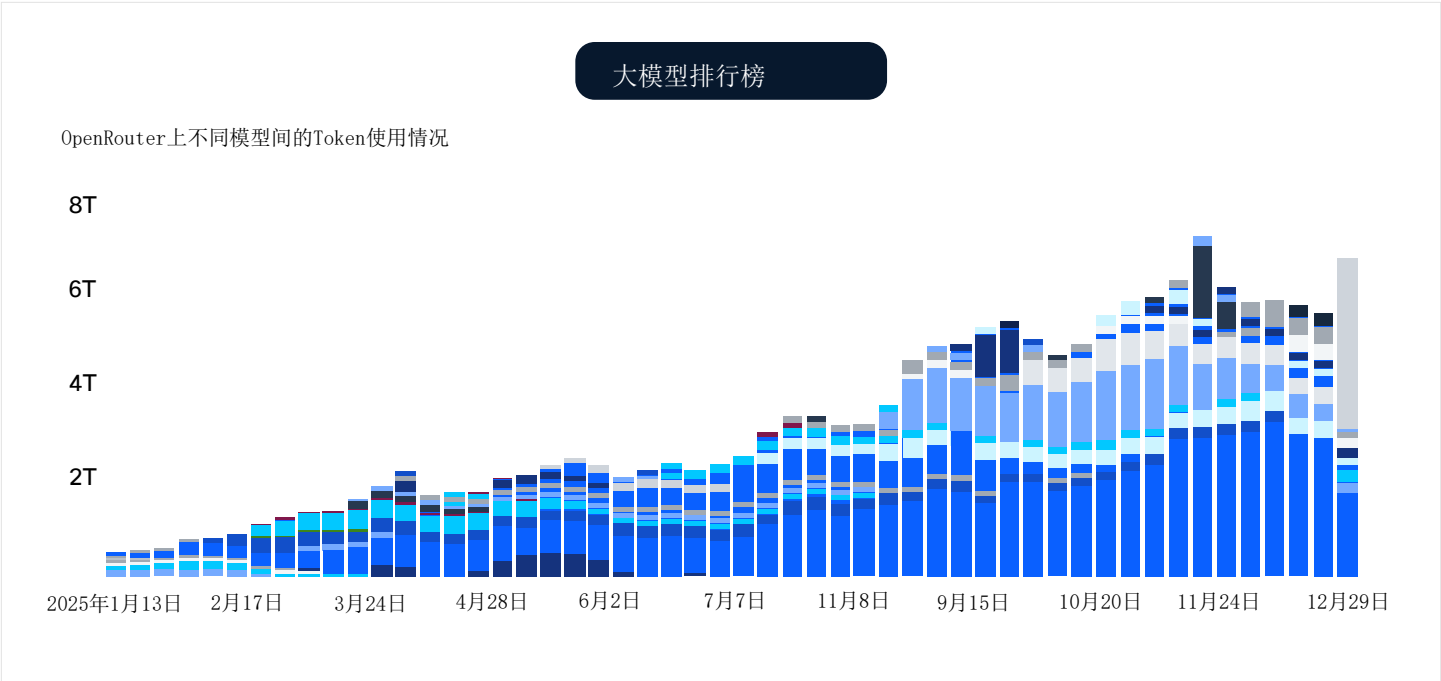


图1. OpenRouter 2026年 token 消耗量 来源：Openrouter.ai

根据两家服务提供商的数据进行流量分析显示，在相似的时间段内，AI推理流量呈现持续增长：在为期八个月的流量监控期间，流量增长了4倍。



图2. AI推理流量演变
数据来源：思科

人工智能推理服务正使用基于传输控制协议（TCP）和快速UDP互联网连接（QUIC）的传输。在流数量方面，两者几乎各占50%，而在数据量方面，QUIC占比57%。

影响：虽然AI推理流量在绝对值上仍然较低，但持续的高增长率将导致其在一段时间内占据相当一部分所有流量。显然，应用程序和用户AI的采用正在推动快速增长。将QUIC用作AI推理的传输方式，给DPI系统应用层检测流量带来了加密挑战。

数据流长度

与非AI网络流量相比，AI推理流量呈现出不同的特征。虽然这些差异并不显著，但它们可能会影响容量规划。数据分析流长度的结果显示，AI推理流量与常规网络交易之间存在差异。

从统计角度看，AI推理的时长是常规网络交易的两倍。

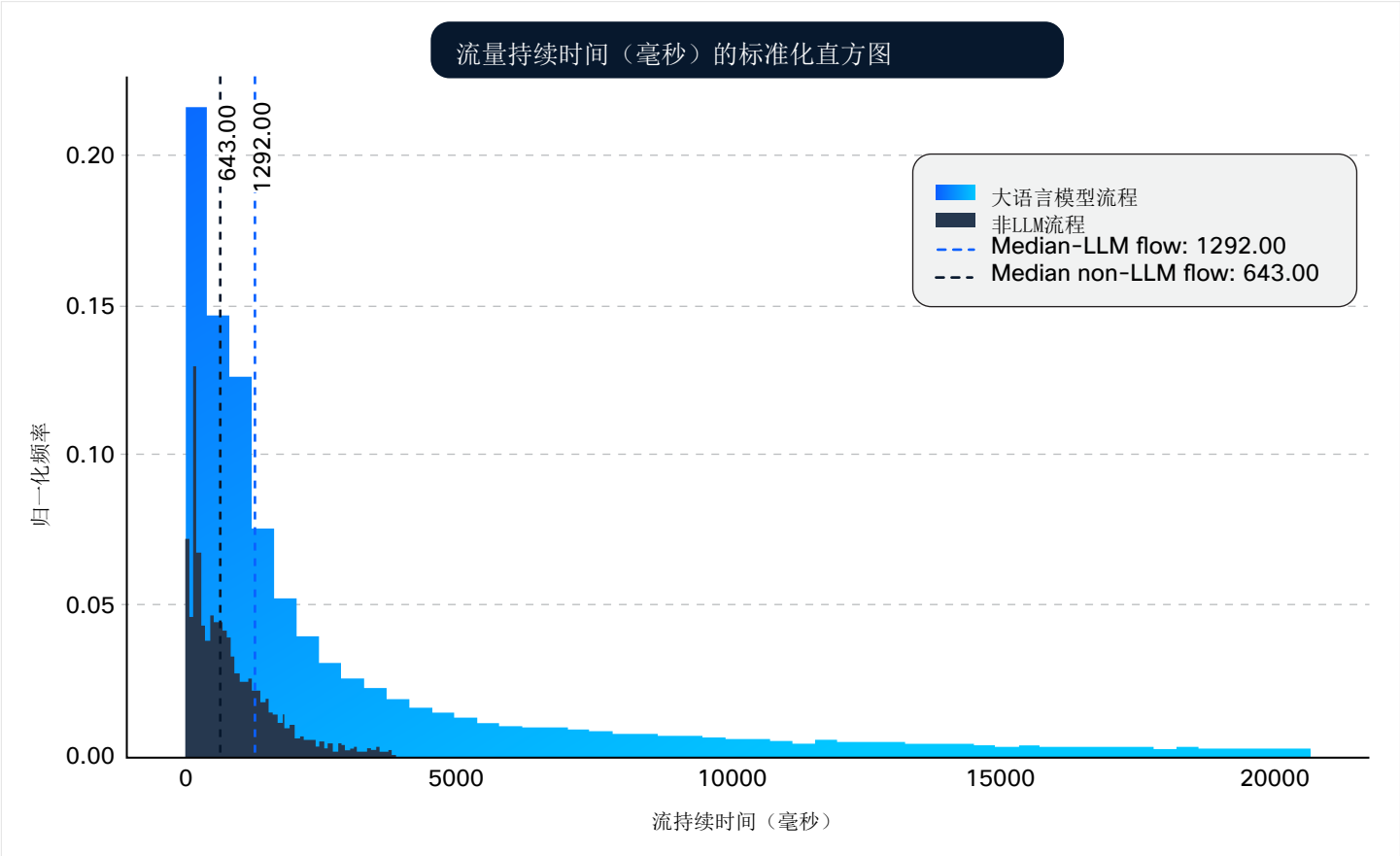


图3. AI推理流与非AI网页交易的时间占比分布 源：思科数据

主要驱动力在于，AI 推理流量是逐个 token 生成内容，与其他流量相比，导致其呈现更长的时长和更低的速率。

影响：对于必须在表中为流保持状态的“流感知”网络系统而言，持续更长时间的AI推理流激增意味着需要更有效地规划不断增长的流表。随着时间的推移，安全和流感知网络系统可能会变得更加分布式，以应对转发状态的增长，类似于Cisco Hypershield™所采用的方法。

数据流速率

与流长度直接相关的是，AI推理流的数据流速与常规网络交易相比呈现出不同的模式。常规网络交易的中位数流速是AI推理流的10倍。

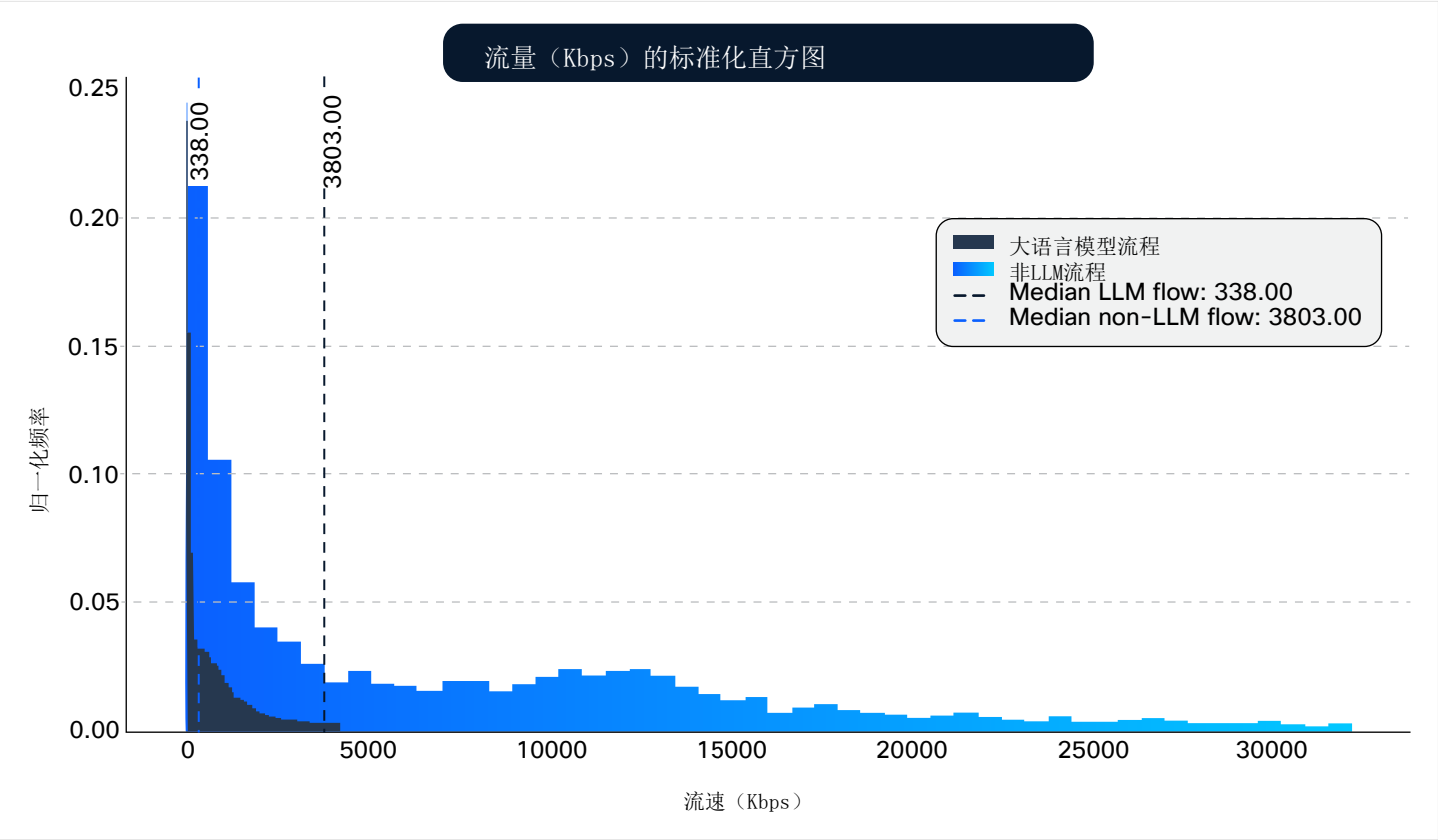


图4. AI推理流量与非AI网络交易流量分布对比
数据来源：思科数据

造成这种差异的主要原因在于人工智能推理中生成数据的过程。一次处理一个数据项的数据流，就像对数据流应用了一个“整形器”。常规网络流量峰值要高得多，因为内容可以从任何存储位置检索并交付给用户。

含义：不同的中位平均流量和峰均比可能需要网络中设置不同的QoS（服务质量）参数，以管理AI推理流量与非AI推理的网页交易。

交通不对称

预计随着人工智能推理流量的出现，流量非对称性可能发生变化。网络流量非对称性因接入类型（移动网络与有线网络）和服务而异。由于社交网络的原因，移动网络流量往往上游流量更重，与有线网络相比，发送到上游的内容更多。

对 AI 推理流量与非 AI 网页交易的流量对称性分析显示出明显差异。如图 5 所示，在 AI 推理流量中，有 9% 的连接（flow）上游流量多于下游流量，而相比之下，在其他 HTTP 交易中，这种情况仅发生在约 0.5% 的连接中。

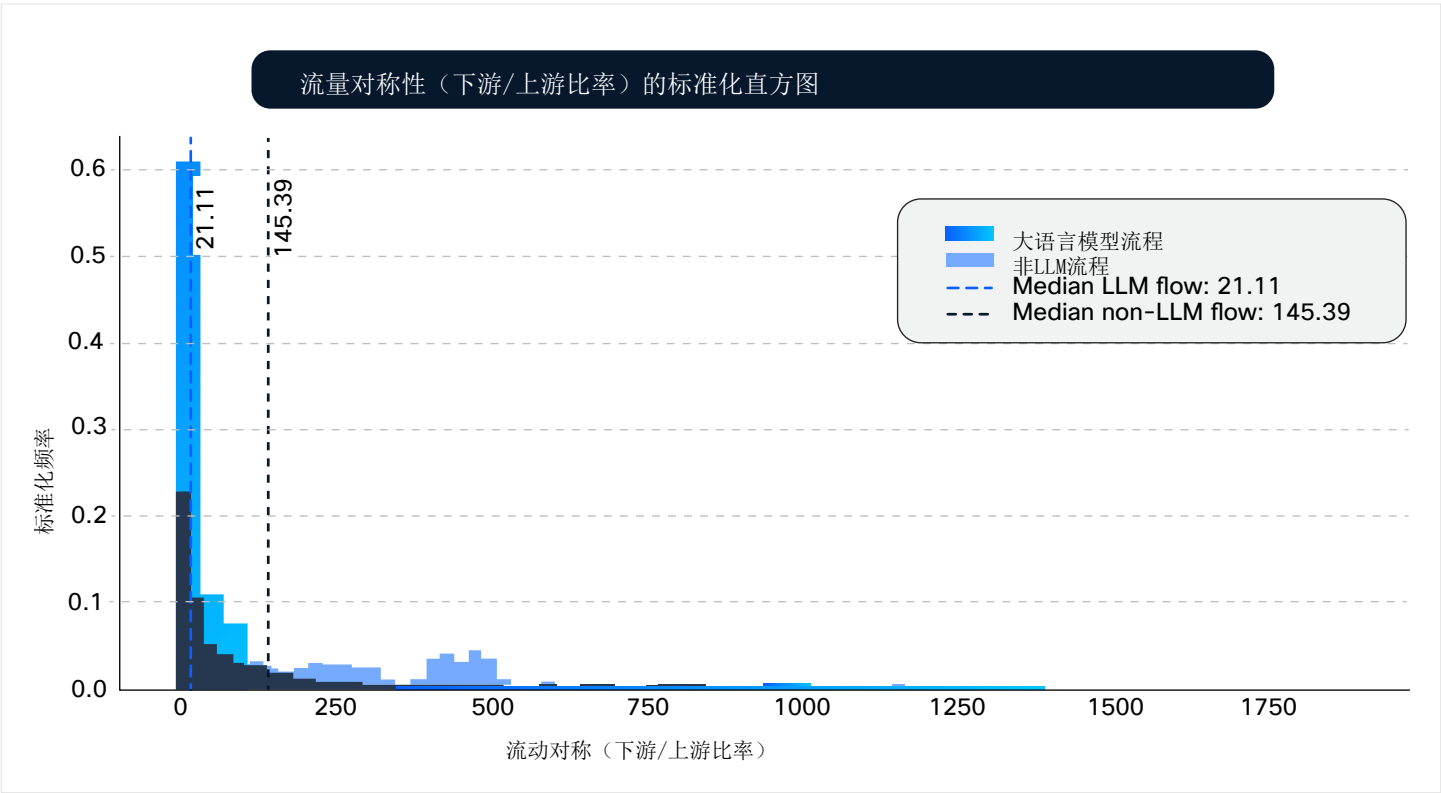


图5. AI推理流量与非AI网页交易流量的流量对称性对比 数据来源：思科

我们预见这一趋势将进一步巩固和增长。造成这种差异的原因是，AI 推理请求通常在提示中捕获越来越多的上下文，以便生成相关的响应。随着自主 AI 的采用增长，可以预期网络流量对称模式将发生更显著的变化。代理将在任务执行过程中捕获持续的上下文，并将这种状态插入到提示中。

影响：特别是在无线电容量规划方面，流量对称性假设是一个相关因素。随着AI推理和代理式AI的采用增长，跟踪流量对称性的演变将很重要，因为它将随着时间的推移而持续减少。

网络延迟影响

大型语言模型（LLM）的推理请求通常比典型的Web API调用具有显著更高的延迟，并且响应时间可能更加多变。传统的Web应用程序REST API通常力求实现亚秒级甚至亚100毫秒（ms）的响应时间。例如，在许多Web场景中，良好的API响应时间通常低于100毫秒。相比之下，即使是短小的LLM查询，也需要数百毫秒才能开始生成输出，而完整响应通常需要数秒时间。LLM的延迟可以按以下方式建模：

$$\text{延迟} = \text{首个令牌时间 (TTFT)} + \text{每个输出令牌时间 (TPOT)} \times \text{输出令牌数量}$$

TTFT是处理提示的初始开销，而TPOT是每个token的生成时间。这意味着延迟随输出长度增长。例如，OpenAI GPT-5.2 4可能需要大约每token 9毫秒，而Anthropic Claude Opus 4.5大约需要每token 20毫秒。一个包含100个token的GPT-5.2答案可能需要大约1秒，而一个最大长度600个token的答案可能接近5.4秒。在各种LLM、模型提供方、实现方式以及用于推理的硬件之间存在很大差异。

影响：AI推理延迟差异很大，从几百毫秒到数秒不等。目前，网络延迟仍然是端到端推理延迟中很小的一部分。然而，推理硬件正在发展，并且变得更快。随着推理加速，网络延迟部分对端到端感知体验的影响更加相关。网络延迟将成为推理分发的一个关键因素，与其他因素（如规模、数据主权和安全）相结合。此外，服务提供商还需要监控客户实际体验的AI推理延迟，因为它是感知用户体验的关键因素。借助Cisco® Crosswork Assurance，可以测量跨不同网络端点的键AI推理KPI，以及网络和服务KPI。

4 LLM排行榜——OpenAI、Google、DeepSeek等100多种AI模型的对比。 <https://artificialanalysis.ai/leaderboards/models>

代理式AI交通模式

除人工智能推理流程外，作为人工智能应用，具有变革性影响且对网络产生启示的智能体式人工智能，也值得特别关注。

AI代理是使用AI模型作为大脑进行决策的软件进程。然后，根据代理的类型，会使用多种工具来完成特定的任务或目标。

一个代理中有三个主要的数据来源或目的地（见图6）：

该代理本身（代理代码运行的位置）

- 人工智能模型（人工智能推理发生的地方）

代理工具和/或数据来源

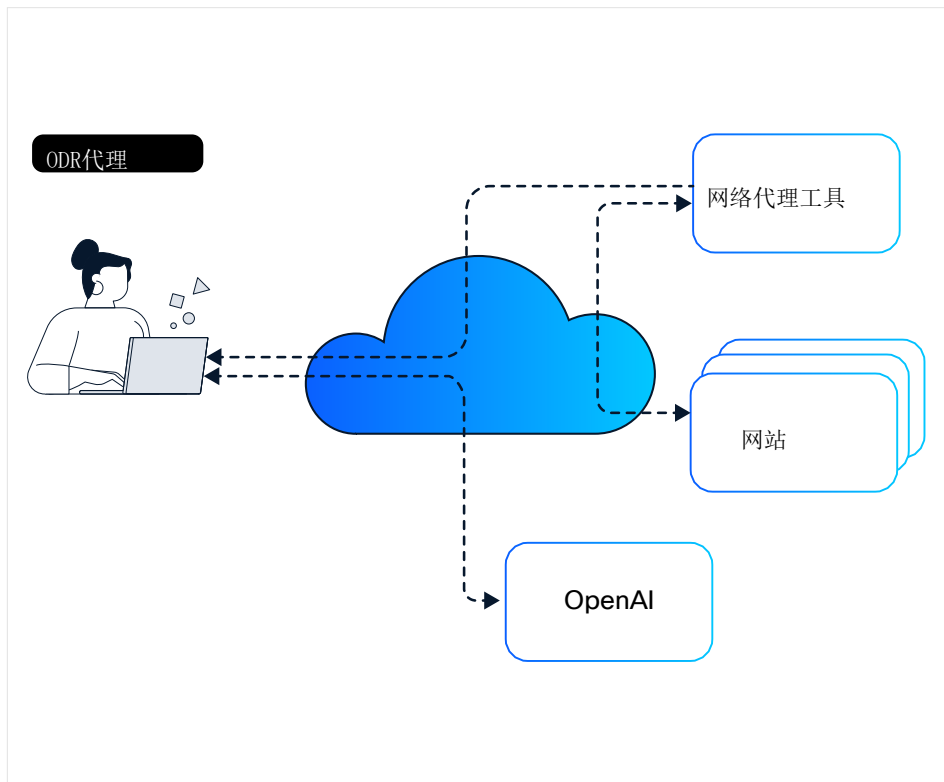
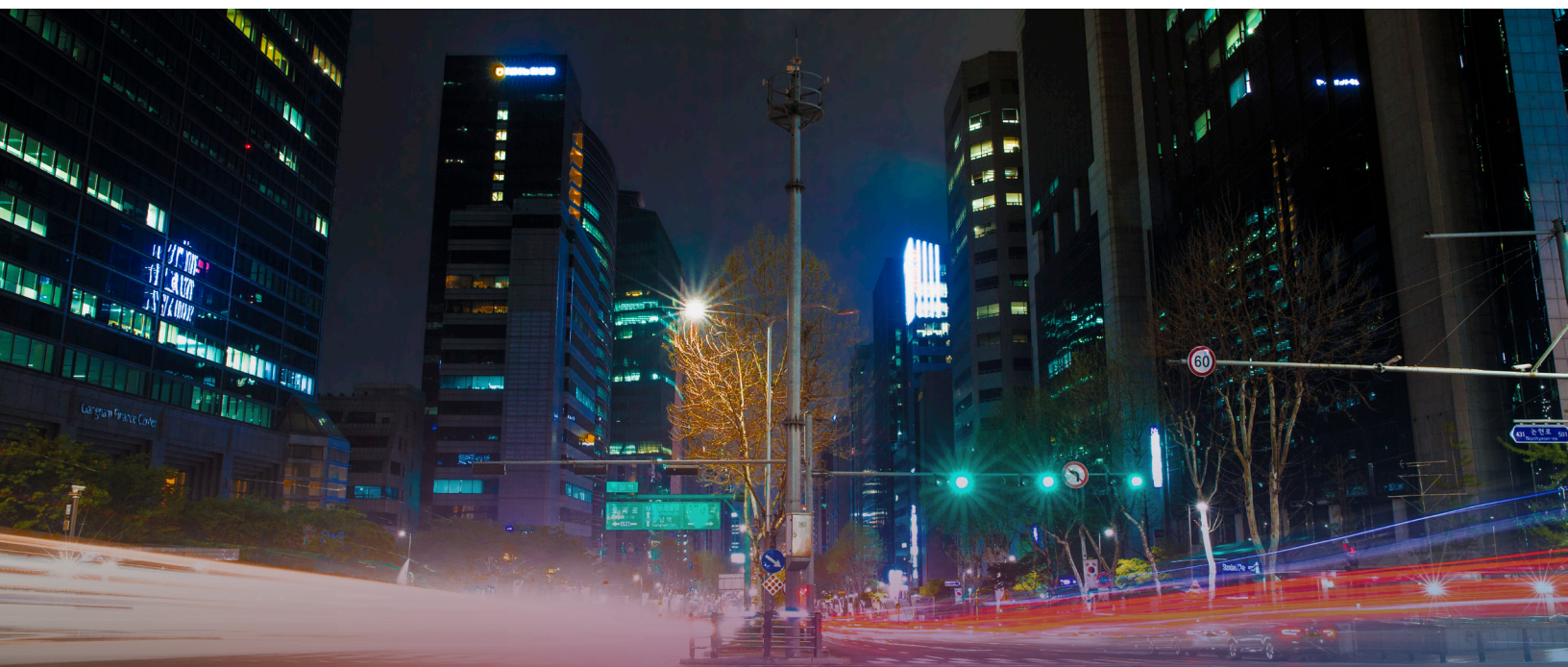


图6. AI代理测试环境 来源：思科



例如，我们正在使用 Langchain5 的一个 Open Deep Research 代理（其他代理的行为也类似）。让我们分析一下当向代理请求特定任务时会发生什么（见图 7）。

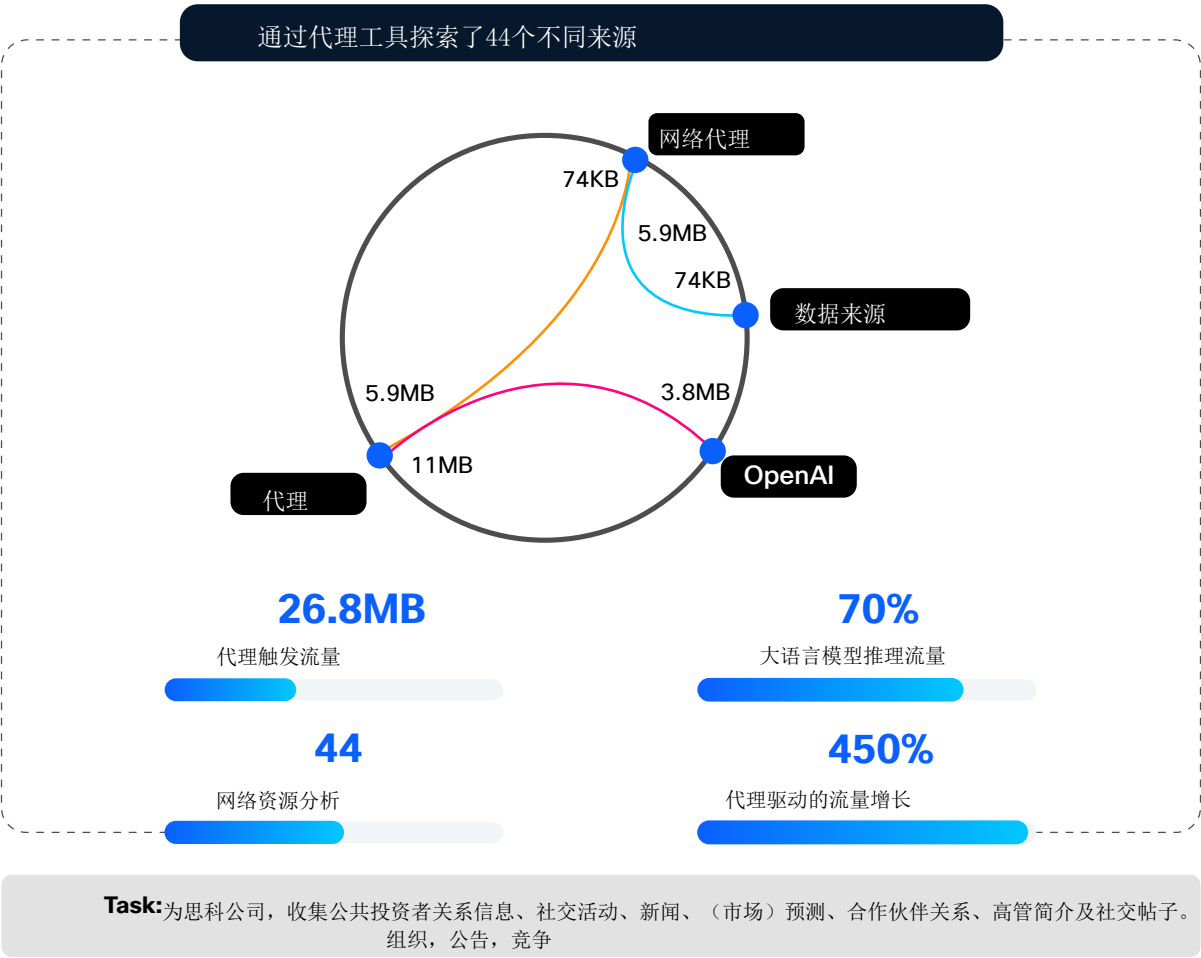


图7. AI代理流量来源：思科

向代理发出请求会在代理的不同组件之间产生大量网络流量。总体而言，由代理产生或触发的网络流量远大于如果由人工手动执行相同任务所产生的网络流量。在我们的示例中，使用代理相关的流量增加了450%，值得注意的是，其中新增流量的70%是AI推理。

以下是与人工智能代理行为及其与网络的交互和依赖相关的关键考量：

- AI代理充当网络超级用户。与手动以人类速度执行这些任务相比，代理以软件速度消耗云和网络资源，这导致网络流量显著增加。
- AI推理流量不仅是流量的重要组成部分，更是代理运行所必需的关键部分。代理代码与AI模型之间的网络路径可以被视为代理的“脊髓”。该通信路径中的任何故障或限制都将显著影响代理的运行能力。

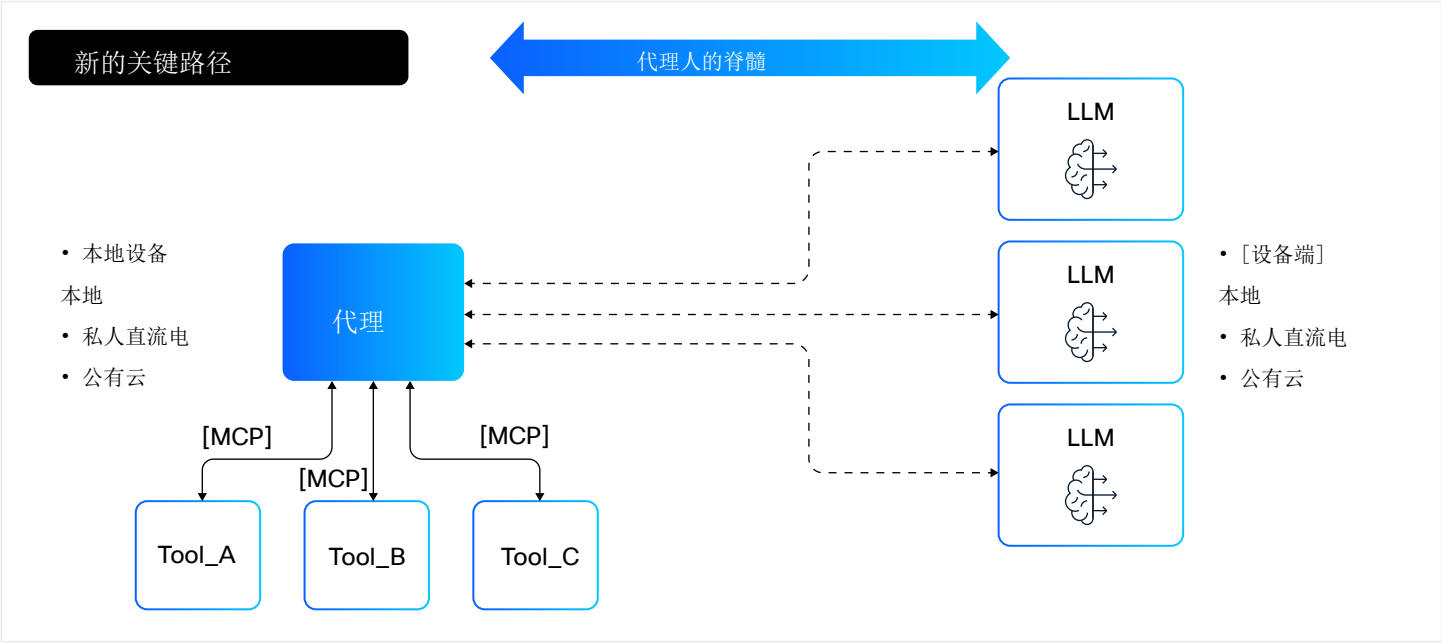


图8. AI代理流量模式和新关键路径
来源：思科

影响：随着自主AI在企业级和消费者应用中的采用日益增长，网络流量也将相应增加，这主要是由用户任务中代理的采用所驱动。在此过程中，由于AI推理流量占AI代理驱动流量的很大一部分，因此其将显著增长。对于通信路径在网络层面的弹性，对于在代理“脊柱”之间、即代理代码和模型之间的AI推理至关重要。



展望未来

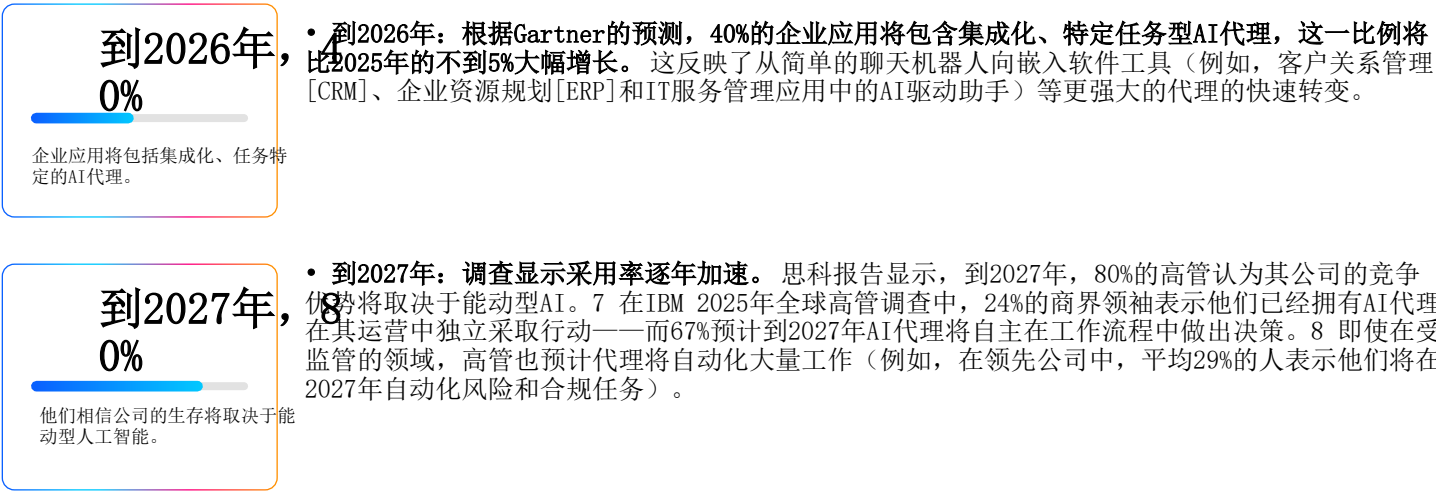
本报告探讨了AI推理流量和智能体AI的行为差异与特殊特征，旨在了解该流量对网络可能产生的影响。

我们的下一步是设想随着这些技术面向消费者和企业的采用率不断提高，网络流量未来可能呈现何种景象。

本报告提供了至2035年的预测。虽然长期建模涉及固有不确定性，但目标是为服务提供商和网络运营商提供长期规划支持。随着更多数据的可用和模型的完善，未来报告中将发布更新的预测。

代理式AI的企业应用与网络影响：2026-2035

自主AI，凭借其自主决策和行动能力的AI代理，在企业应用中尚处早期阶段，但其发展势头正迅速增强。分析师指出，2024-2025年的采用情况尚处萌芽阶段（到2025年，不到5%的企业应用中部署了AI代理[6]），然而各行业的组织已开始规划其广泛实施。2020年代中期的主要预测包括：



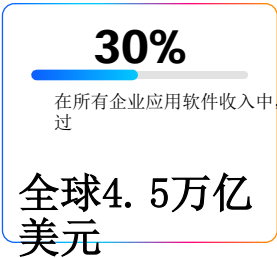
6 Gartner新闻稿：Gartner预测，到2026年，40%的企业应用将具备任务特定AI代理功能，这一比例将从2025年的不到5%大幅提升。⁷ 来源：思科报告《代理式AI竞赛：为何基础设施将决定员工整合成败》

这些早期预测强调了代理式AI的使用正从实验阶段转向战略层面。企业正通过投资生成式AI平台和自动化来奠定基础，这将铺平部署自主代理的道路。⁹ 成本降低和生产率提升是早期采用者（39%将其列为主要驱动力）的关键动机。到2027年，许多企业将完成从规划和试点阶段向在日常运营中积极使用AI代理的过渡。

人工智能代理将在业务流程中变得普遍。高德纳分析人士预测，到2028年，“三分之一的用户与软件的交互将从传统应用程序转向由代理驱动的AI界面”，¹⁰ 这说明了这些代理在日常工作流程中将有多么普遍。在客户运营方面，高德纳预计到2029年，80%的常规客户服务问题将由人工智能代理自主处理。¹¹

展望未来，行业专家预计，到2030年代，自主AI的采用将加速发展，这些AI“同事”将成为各行各业普遍存在。

2030年代中期的主要预测和预期包括：



• **企业软件革命：**根据Gartner的预测，到2035年，自主AI将驱动约30%的企业应用软件收入，全球收入将超过4500亿美元（相较于2025年软件收入的2%）。本质上，到2035年，近三分之一的企业软件市场可能归因于AI代理功能——这是一个十年内的根本性转变。在那个阶段，预计大多数企业软件都将深度嵌入AI代理，而新的软件商业模式将围绕自主功能展开。

• **工作新常态：**到2035年，企业环境很可能在大多数业务流程中呈现人类与AI代理并肩工作的状态。Gartner描述了一种民主化的未来状态，员工可以按需常规地创建或定制AI代理，而自主代理则跨部门协调复杂的业务流程。

⁹ Omdia: 新Omdia分析显示，代理式AI增长速度超越传统生成式AI，数据更新于2025年9月。

¹⁰ 高德纳新闻稿

¹¹ 高德纳新闻稿

¹² 高德纳新闻稿

到2030年代中期，自主AI不再被视为一项新兴技术，而是被视为企业运营的标准组成部分。那些缓慢拥抱自主AI的公司可能会发现自己处于竞争劣势。总的来说，共识是2020年代是自主AI的采用十年，而2030年代将是其成熟和全面融入商业肌理的时代。

根据从行业报告中捕获的数据，以下展示了一个关于2026年至2035年企业任务中自主AI渗透率的预测模型。

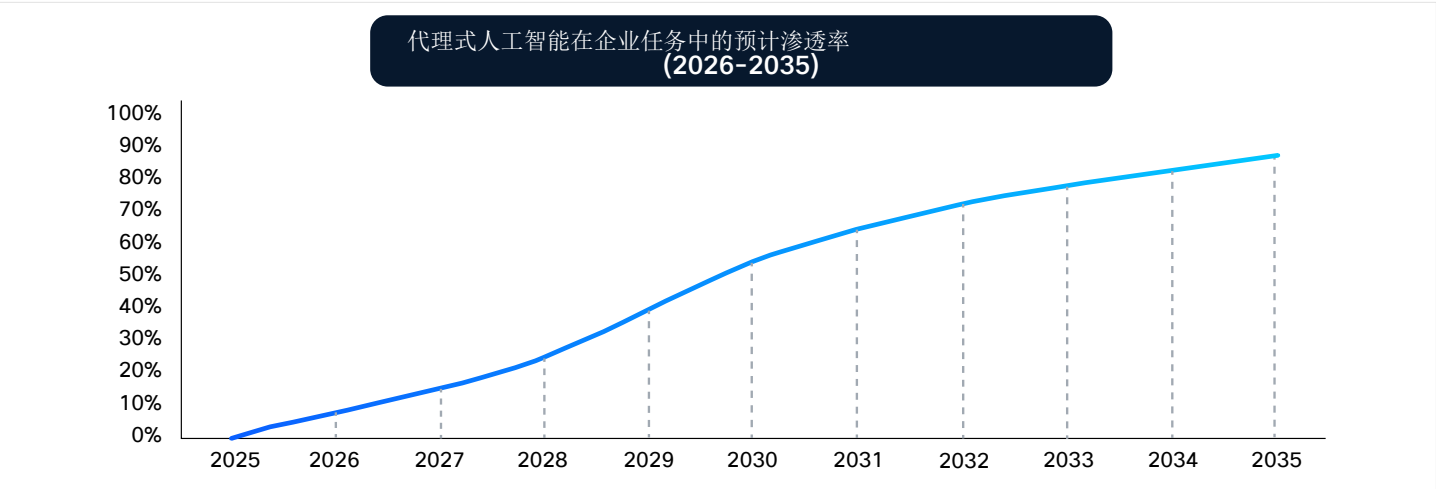


图9. 企业任务中代理式AI的预期采用情况
来源：思科模型

考虑到使用AI代理自动化任务的网络影响以及预测的AI代理渗透率，图10展示了2026年至2035年间，由企业服务和AI代理采用所驱动的网络流量总体影响。

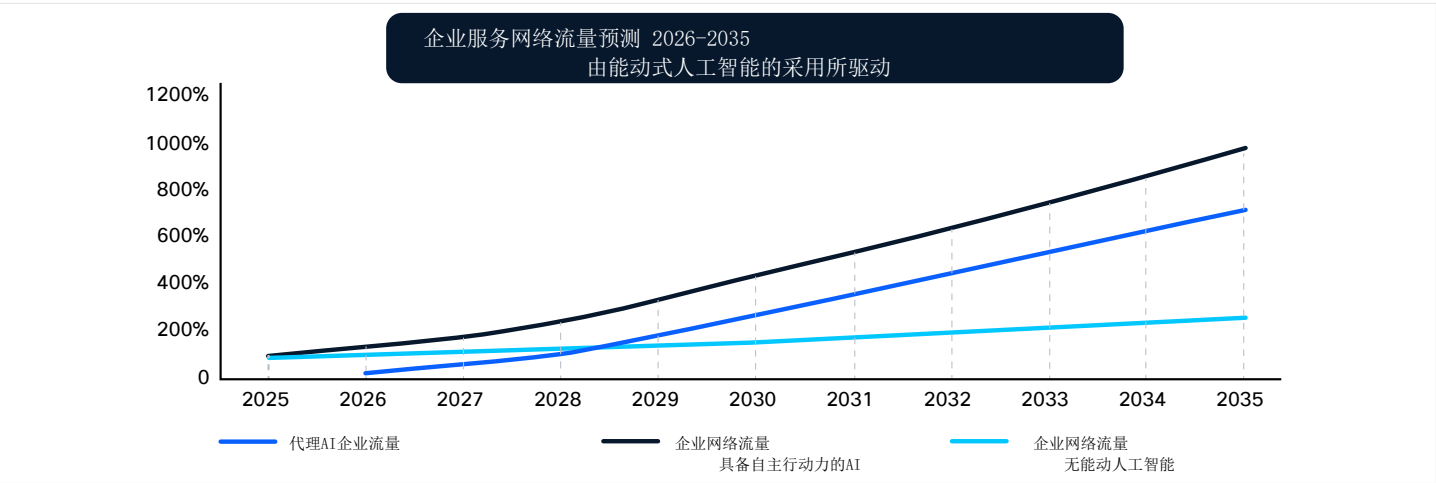


图10. 代理式AI对企业服务网络流量预期的冲击 (%增长)
思科模型

如果没有具有自主行动能力的AI，预计2026年至2035年期间，企业流量将增长约250%。然而，具有自主行动能力的AI将给企业网络使用带来重大颠覆和影响，推动流量增长9倍。

人工智能的消费使用与自主人工智能及网络影响：2026–2035

人工智能正迅速融入日常消费生活，并有望迎来采用量的进一步激增。主要驱动力包括生成式AI聊天机器人和助手（例如ChatGPT）、游戏和娱乐中的AI、AI驱动的图像和视频生成、用于购物和银行的智能应用程序，以及依赖云端AI推理的增强现实（AR）体验的兴起。能够代表用户行动的自主AI代理驱动的智能AI也正在个人应用中涌现。

61%

过去六个月中，美国有多少成年人使用过人工智能？

人工智能的消费应用在2020年代中期达到了一个转折点。2025年中期的调查显示，美国61%的成年人过去六个月里使用过人工智能，其中约19%的人每天都在使用。按全球规模计算，这意味着全球约有17亿至18亿人接触过人工智能工具，其中包括5亿至6亿日活跃用户。这一爆发式增长发生在主流生成式人工智能（ChatGPT于2022年末推出）出现后的短短几年内，表明了消费者技术采用前所未有的速度。

46%

到2030年，在美国所有消费交易中

到2030年，许多分析师预测，人工智能在日常生活中的普及程度将堪比今天的智能手机。高德纳（Gartner）建议，人工智能将深度融入日常活动；例如，到2028年，大部分商业采购可能将由人工智能代理处理，¹⁴ 消费领域也正沿着类似的轨迹发展。科迅（Cognizant）和牛津经济学（Oxford Economics）预测，到2030年，由人工智能驱动购买将占美国所有消费者交易的大约46%。¹⁵ 换句话说，近一半的消费支出可能受到人工智能的影响或由其执行（德国和澳大利亚的类似数据分别为46%和55%）。

到2030年代初，我们可以预期大多数消费者将常规性地依赖人工智能

到2030年代初，我们可以预期大多数消费者将常规性地依赖人工智能——无论是自觉使用（例如与聊天机器人互动或使用人工智能助手），还是潜移默化地（应用程序、家电和汽车中的人工智能驱动功能）。随着人工智能的信任度提升和日益集成，如今偶尔使用人工智能的消费者很可能到2030年将成为日常用户。门洛投资（Menlo Ventures）的分析师预计，鉴于其更广泛地覆盖了普通大众，消费者人工智能的采用最终将在使用率和经济价值两方面赶上企业人工智能。

到2035年，预测倾向于趋同于这样一种假设：对于使用连接技术的任何人来说，人工智能将几乎无处不在。虽然2035年的精确渗透百分比很少，但预计到2030年代中期，全球消费者人工智能的使用率将接近互联网用户的90%至100%，这意味着数十亿人将每天与人工智能互动。人工智能可能会成为设备和服务的默认层——就像宽带或电力一样——很大程度上是不可见的，却是不可或缺的。

13 梅洛风险投资公司，2025年：消费者人工智能现状。¹⁴ 网络世界，AI的阴暗面在IT组织的Gartner顶级预测中显现，2025年10月。

15 柯尼卡美能达，柯尼卡美能达研究报告显示，拥抱人工智能的消费者可能在五年内推动4.4万亿美元消费。2025年1月。

到2030年

23%

过去一个月内，有百分之多少的美国人使用人工智能进行了购买。

本时期的一个主要趋势是代理型AI的兴起，自主AI代理能够主动为用户执行任务。这些任务范围从为您寻找和购买产品的个人购物机器人，到能够管理日程、财务或家庭设备且仅需极少监督的AI助手。行业未来学家预测，到2030年，个人AI助手将充当消费者代理人——筛选优惠、协商交易以及自动化决策。摩根士丹利报告称，截至2025年末，大约有23%的美国人过去一个月曾使用AI进行购买。该报告预计，“代理型购物者”到2030年将占美国所有电子商务支出的10%至20%。

到2030年

3万亿至5万亿美元

全球零售收入的一部分可能由人工智能代理操控。

在全球范围内，自主型消费者AI的影响力可能非常巨大。麦肯锡的研究预测，到2030年，全球零售收入中可能有30万亿到50万亿美元是由能够预见需求并自主执行购买的AI代理来推动的。¹⁷ 这是对消费者如何寻找和购买产品方式的重大转变——其变革程度堪比最初的电子商务革命，但可能以更快的速度发生。自主型AI不仅限于购物；类似的自主辅助服务还包括个人金融（AI机器人优化投资或预算）、旅行（AI代理提供全程旅行规划）甚至医疗健康（AI助手安排看医生或管理药物）。

到2030年

全球大部分消费者可能会像使用互联网一样常规地使用人工智能助手和工具。

如果该代理介导的范式哪怕只有一部分延伸到消费者身上，到2030年，一个典型的人可能会常规地依赖AI代理来处理各种日常任务，从补充杂货到筛选传入信息。简而言之，虽然代理式AI的采用目前尚处萌芽阶段，但预计在2026-2035年的时间范围内将成为消费者生活的主流方面，基本上为人们提供一个“数字个人助理”，该助理在为他们持续工作。

消费级人工智能的设备格局也在不断拓宽。虽然就数量而言，智能手机仍是王者，但人工智能正渗透到我们的家居、可穿戴设备、汽车以及新兴的AR界面中。在2026年至2035年期间，这种多设备普及将确保人工智能几乎无处不在、无时不有。其最终效果是，采用率将不仅因为个人偏好而上升，更因为我们所处的基建和设备本身就具备人工智能功能。

2026年至2035年间，人工智能将有望从一项新技术转变为消费领域的通用工具。预计通用人工智能的使用和代理人工智能的采用都将呈指数级增长，覆盖数十亿人，触及全球所有人口群体。到2030年，全球大部分消费者可能会像使用互联网一样常规地使用人工智能助手和工具——年轻一代将率先采用，年长一代则稳步跟进而来。以个人自主代理形式存在的代理人工智能，预计将处理越来越多的数字和物理任务，从购物到日程安排，预示着一个便捷的新时代（同时也引发了关于安全、信任、控制以及模型偏差等新问题的讨论）。

摩根士丹利：购物机器人来了，2025年12月。<https://www.morganstanley.com/insights/articles/agent-commerce-market-im-pact-outlook>

¹⁷ 量子黑，麦肯锡的AI，代理式商业机遇：
消费者与商家，2025年10月

物理人工智能将在机器人技术进步时对网络基础设施产生重大影响。随着工业和消费领域的更广泛商业应用，对人工智能推理的依赖将增加。许多系统将结合嵌入式实时推理以实现即时行动，以及基于云的推理能力。在此分析中，机器人可以被视为配备实体工具的物理代理，而非纯粹的数字代理。从网络角度来看，我们预计其总体影响将与数字代理相当。这是未来报告中需要密切关注的关键领域。

至关重要，这些采用趋势并非孤立发生；它们与AR可穿戴设备等设备的进步以及人工智能更深入地融入每个应用程序和服务同步发生。因此，消费者人工智能的渗透率可以被理解为并非简单的线性增长，而是一条加速趋近饱和的曲线。报告和预测一致指出，在2020年代，人工智能的使用率和市场规模将保持高双位数增长率。¹⁸ 如果这些趋势持续，到2035年，我们将生活在一个世界，在这个世界里，几乎每个人，无论年龄和地区，都会定期依赖人工智能——无论是通过直接互动，还是通过嵌入其环境中的无形辅助。实际上，人工智能将成为无处不在的副驾驶，成为普通消费者的常伴。

根据行业报告数据，图11展示了一个关于2026至2035年间消费者任务中AI及智能体AI应用采纳的预测模型。

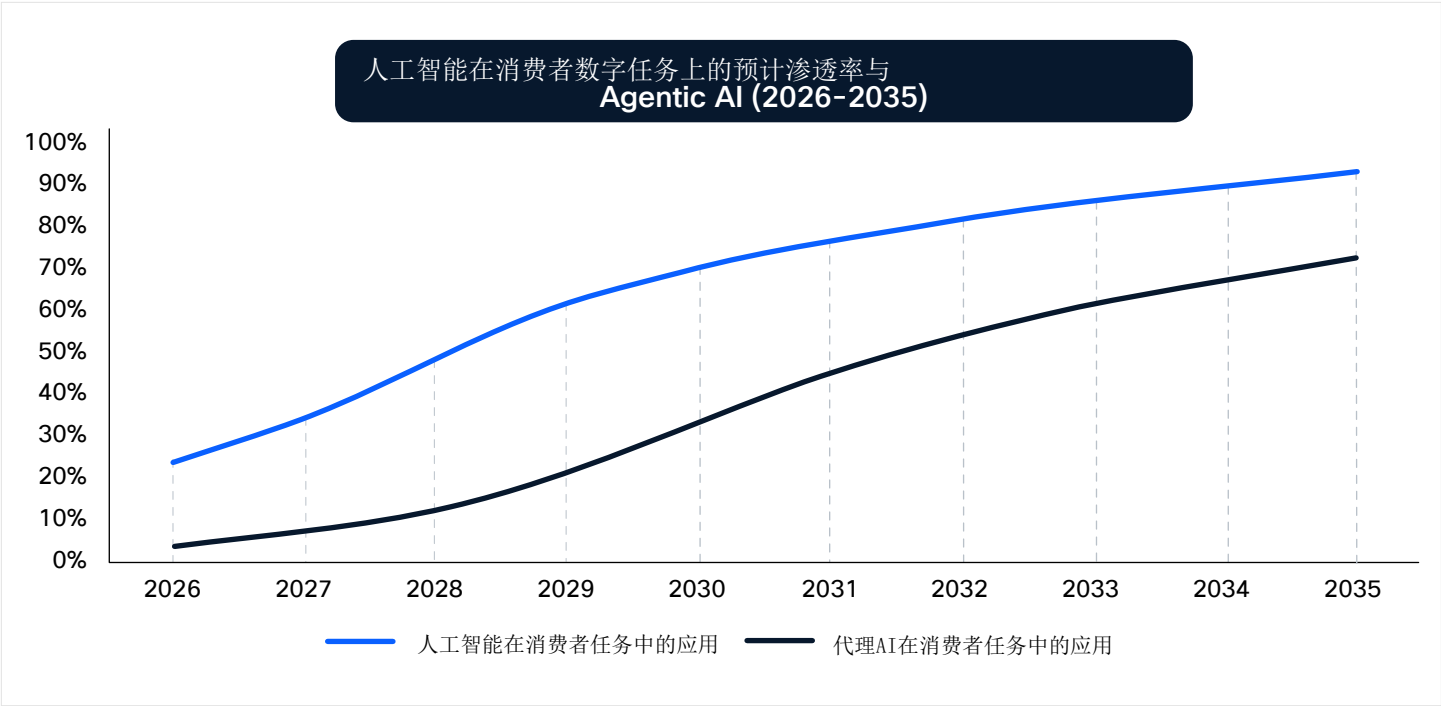


图11. AI及自主AI在消费者数字任务中的预计采用情况
来源：思科模型

18 GlobeNewswire，智能语音助手应用市场报告2025-2035：行业收入将从2025年的81亿美元增长至2035年的1535亿美元，2025年12月。

并非所有消费者任务或数字服务对网络的影响都相同，它们也并非都会受到人工智能（AI）或代理式人工智能（agentic AI）的同等影响。图12中的模型显示了在2026年至2035年期间，人工智能（AI）和代理式人工智能（agentic AI）的采用对整体网络流量产生的综合影响。

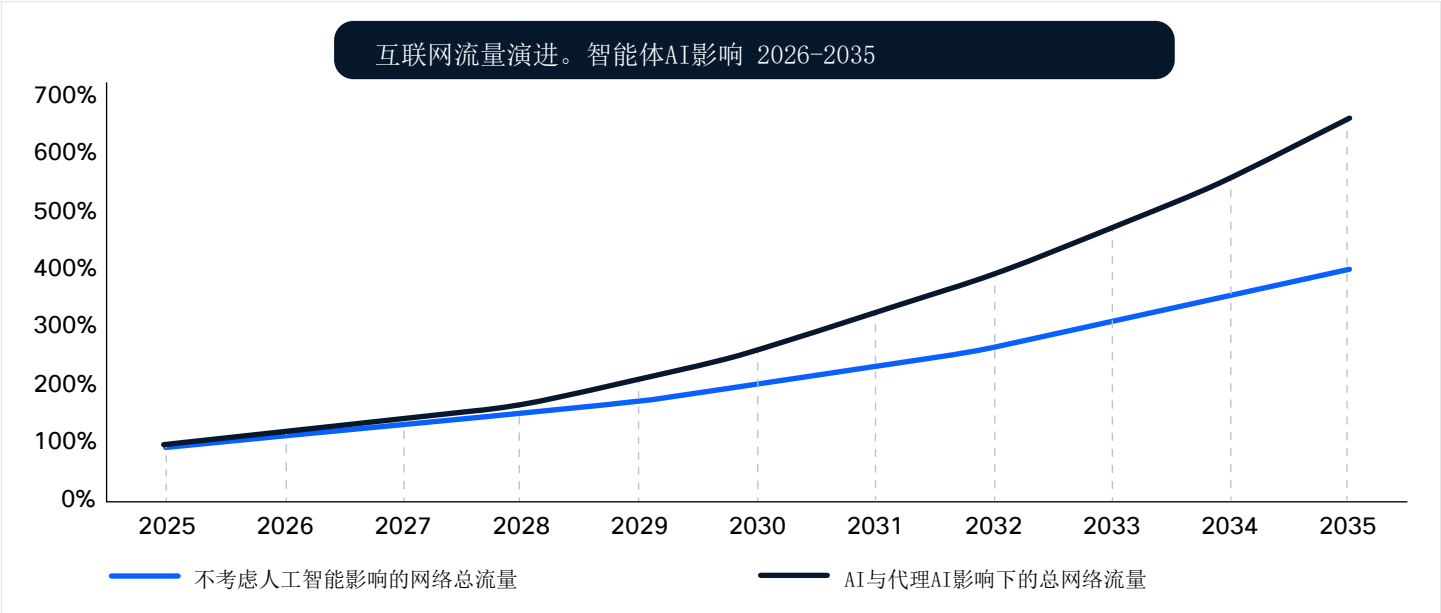


图12. 人工智能与自主人工智能对互联网网络流量预计影响
来源：思科模型

其影响重大，因为消费者流量是整体互联网和服务提供商网络流量的主要贡献者。若不考虑人工智能（AI）或代理式人工智能（agentic AI）的影响，预计在2025年至2035年期间，网络流量将增长4倍，主要由数字服务的采用所驱动。然而，随着人工智能和代理式人工智能在许多数字服务中的采用，预计流量将从当前水平增长6.6倍，或比非人工智能影响的预测高出63%。其中相关部分流量将是人工智能推理，预计到2035年将占总流量的25%。

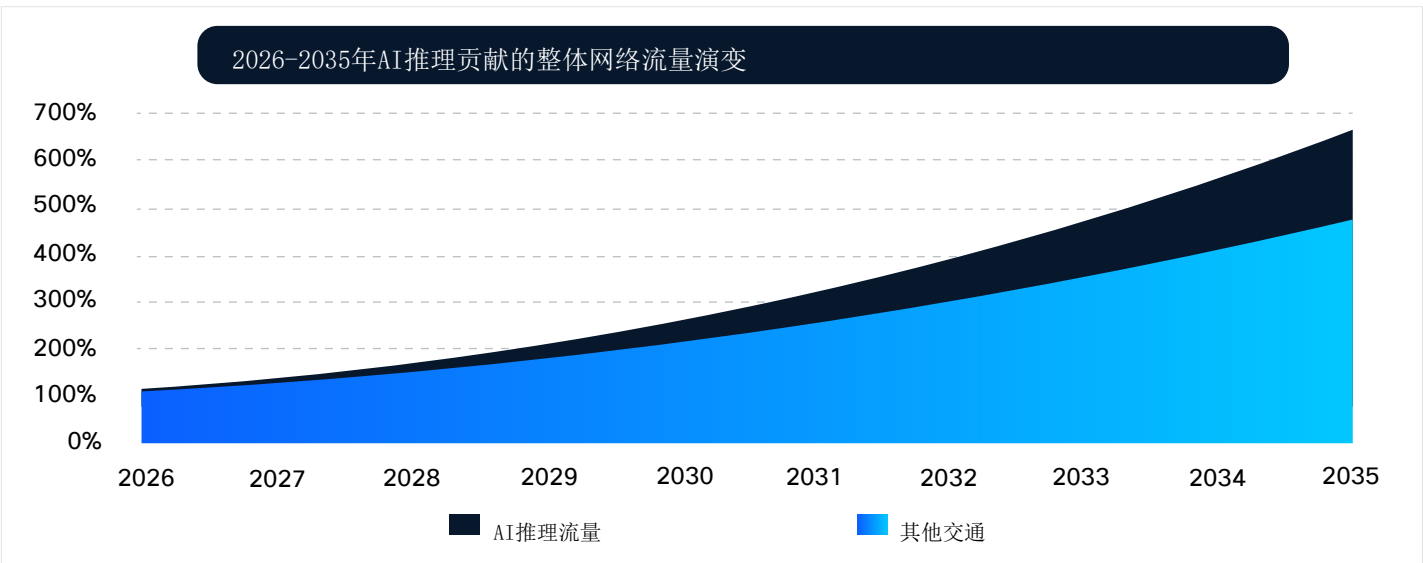


图13. 预测的网络流量增长，AI推理流量与其他流量对比 来源：思科模型

2029年至2032年期间，总网络流量将经历显著增长，此时自主AI的采用将出现更明显的增长，AI推理流量年复合增长率（CAGR）约为25%。

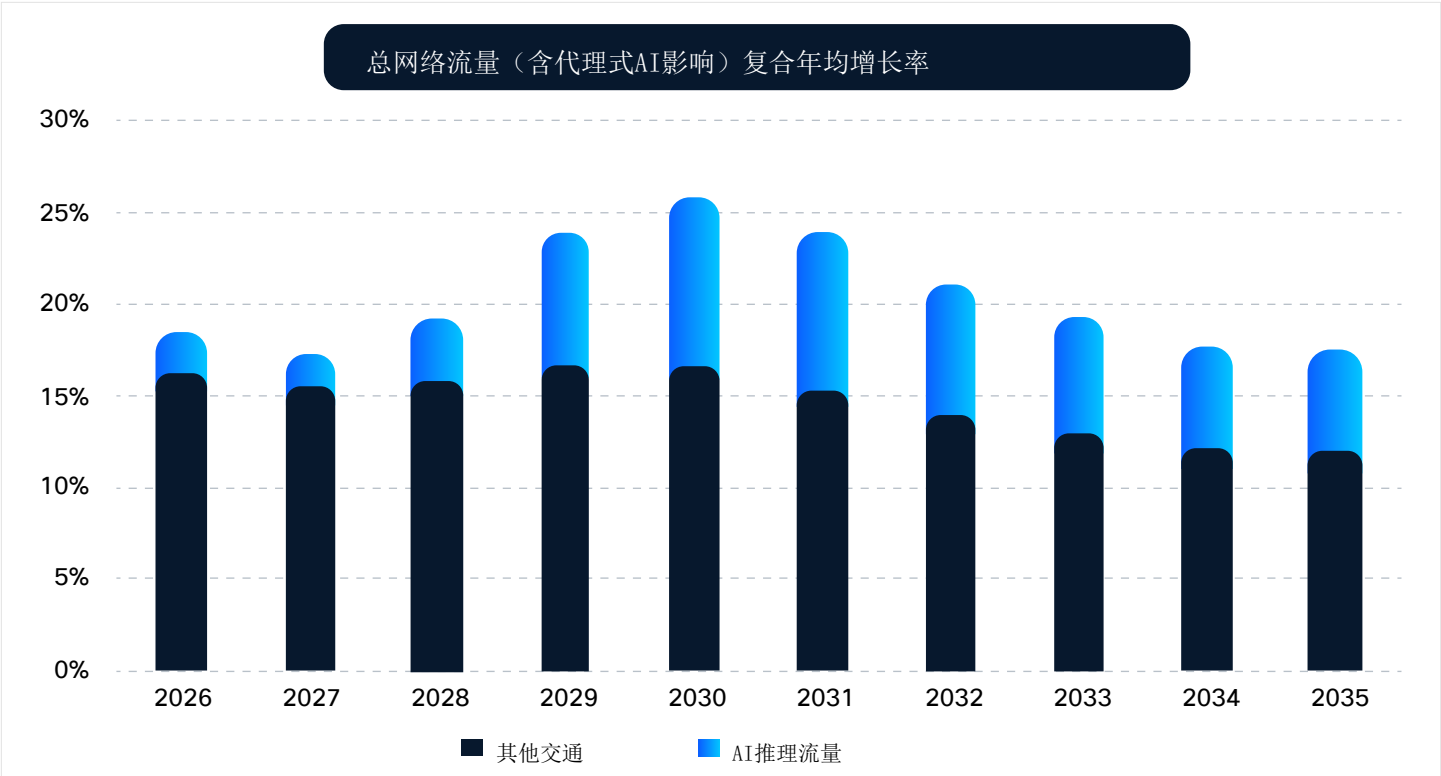


图14. 目标期间AI推理流量和其他流量的复合年均增长率（CAGR）
来源：思科建模

随着人工智能（AI）和自主AI在企业及消费者数字服务中的采用持续进行，未来的转型在网络流量量方面意义重大，并对流量对称性、流持续时间以及关键路径的弹性产生了新的影响。

进一步阅读：思科报告2026《代理式AI的竞赛：为何基础设施将决定或破坏劳动力转型》<https://www.cisco.com/c/dam/en/us/solutions/collateral/artificial-intelligence/race-to-agentic-ai-report.pdf>。

本研究是在与Opanga Networks紧密合作下完成的。

报告作者：约汉·古斯塔夫松（思科）、哈维尔·安蒂奇（思科）、瓦里斯·萨格尔（思科）

贡献者：Ben Hadorn（Opanga Networks）、John Burnette（Opanga Networks）、Ryan Johnston（Opanga Networks）

