



2026科技、传媒和电信行业预测

2026科技、传媒和电信行业预测： 差距缩小，但仍将持续存在

德勤预测，到2026年，随着AI规模化应用的持续推进，其理想与现实之间的差距将缩小

德勤预测，到2026年，随着AI大规模应用的工作持续开展，围绕AI的喧嚣将逐渐趋于理性与务实。理想与现实之间的差距将会缩小，但不会消失：进展将更多地来自于基础能力的夯实，而非引人注目的新模型。这种更为务实的关注点之所以重要，源于科技、传媒和电信（TMT）行业日益增长的重要性不仅关乎芯片和代码本身，更在于其他行业如何利用这些TMT能力来实现自身的增长、效率和创新能力。

AI助力跨行业转型

展望2026年及未来，我们似乎正从“软件颠覆世界”迈向由AI——尤其是代理式AI（Agentic AI）引领的“TMT颠覆世界”阶段。在美国，AI数据中心的支出目前几乎占上半年GDP增长的全部。¹ 科技股约占标普500指数市值的19%，而如今TMT行业市值占比已接近53%。² 情况可能会发生变化，但照此趋势，TMT行业不仅将超越其他任何单一行业的规模，更有望在价值和经济增长贡献方面超过其他所有行业的总和。原因之一是其他行业正利用TMT（特别是科技和电信）来驱动自身的AI创新，而TMT恰好是AI热潮中的硬件、软件和服务提供商。

尽管如此，其他行业也发挥着关键作用。在2025年和2026年的TMT行业预测中，我们汇聚了来自德勤其他研究中心及行业的专家见解，涵盖能源、采矿与化工，制造与建筑，国防与航空航天，政府与公共服务，以及生命科学与医疗保健。要准确预测生成式AI（GenAI）和代理式AI的趋势及影响，需要真正深入的跨行业协作。

在我们2026年的13个预测主题中，超过半数围绕AI展开。总体而言，我们看到了一个关于推动AI规模化的核心叙事。新的基础模型，乃至光鲜的企业级智能体应用，持续令人惊叹，但要将这些应用推广到试点和试验之外，则需要完成通常被认为不那么激动人心的工作，例如数据治理、与现有工作流的集成、合规管理、新定价模式以及应对监管要求。这些工作可能不如“AI在科学测试中击败人类”的新闻稿那样引人注目，但在短期内可能更为实用。

生成式AI和代理式AI正在驱动许多当下正在发生的深刻变革，同时我们也关注更远的未来。尽管德勤预测机器人和无人机在未来一两年内将呈现缓慢而稳定的增长，但“物理AI”模型的出现有望以巨大的增速和实用性变革这两个行业。与此同时，短剧等新型媒体形式正从亚洲走向全球。此外，尽管生成式AI创作的图像在社交媒体上的传播可能引发关注，但它也可能刺激相关监管措施的出台。

2026年的13个主题速览

现有应用植入生成式AI, 打破独立AI的“独大”局面

生成式AI (*Gen AI*) 或许是这个时代最具影响力的技术之一，通过将其集成到现有主流数字应用（如搜索引擎）中，其用户群体的扩展速度可能会超过作为独立应用使用时的速度。

德勤预测，2026年及以后，使用内嵌于现有应用（如搜索引擎）中的生成式AI的人数将超过使用独立生成式AI工具的用户。在日常使用层面，通过搜索引擎调用生成式AI（例如搜索时获得综合结果摘要）的普及度将比使用任何独立生成式AI工具高出300%。独立生成式AI工具通常需要用户掌握提示词工程技巧并保持耐心尝试，而被动式生成式AI则更为隐性、体验更贴近传统操作，因此其可及性更高、需求也更为广泛。未来，独立生成式AI应用的运营方或将面临选择：要么将其功能嵌入其他应用程序中，要么继续维持独立的操作界面。

为何生成式AI下一阶段将要求更高算力？

世界正从单纯训练生成式AI模型转向大规模应用。许多人认为这意味着消费者边缘计算增加，而数据中心计算将减少。但在2026年，这两者都不太可能发生。

德勤预测，到2026年，“推理”（即运行AI模型）将占据全部AI计算能力的三分之二。尽管存在不同预测，但大多数推理仍将在价值近5000亿美元的新型数据中心，以及使用价值超过2000亿美元的高成本、高功耗AI芯片的企业本地服务器中进行，而非通过廉价、低功耗的边缘芯片完成。未来将出现价值数十亿美元的推理专用优化芯片，但它们同样会部署在数据中心和企业服务器中，且部分芯片的功耗将与通用AI芯片相当甚至更高。

借力智能体协同调度，驱动价值指数组攀升

自主智能体 (*Autonomous AI agents*) 或许具有变革性力量，但协调机制可能是实现智能自动化的关键。开源通信协议与专有通信协议将展开竞争，争夺主导地位。

到2026年，估计自主智能体市场规模可能会达到85亿美元，到2030年达到350亿美元。德勤预测，如果企业能够更好地协调代理并审慎应对其相关挑战和风险，到2030年这一市场预期可能会增长15%至30%，高达450亿美元。在2026年，企业可能会努力提升自身在协调具有一定自主性代理方面的能力。与此同时，对于那些注重代理互操作性与管理，并能有效实现在工作流程和人才方面所需变革的企业而言，多代理系统很可能将发挥积极作用。

AI革命正引领机器人和无人机领域的创新浪潮

更强大的AI模型和芯片能否为这个相对停滞的行业带来变革契机？

德勤预测，到2026年，全球工业机器人的累计安装量可能达到550万台，但自2021年以来，全球工业机器人年销量始终徘徊在50万台左右。发展拐点可能早在2030年就会出现——届时年度新型机器人出货量将翻倍至100万台，这主要由以下增长催化剂推动：(i) 发达国家专业工业应用中的劳动力短缺，以及 (ii) 计算能力的指数级进步以及专用基础AI模型的出现。机器人可渗透至包括自主无人机在内的多个行业与应用场景，但除非更广泛的技术、人工智能及机器人生态系统能解决数据质量、系统集成和网络安全等瓶颈问题，否则工业机器人市场可能仍将维持在当前相对平缓的年增长率水平。

当SaaS邂逅智能体：预算体系、客户体验与劳动力生态的深度变革

随着智能体 (AI agent) 在软件即服务 (SaaS) 市场中的普及，企业使用和利用软件的方式可能会发生变化，从而改变商业模式、能力和用户期望。

AI继续颠覆SaaS市场。随着代理式AI (Agentic AI) 能力的不断成熟，以及企业级SaaS供应商持续扩展其平台以创建、集成和编排智能体，企业采购和使用软件的方式可能发生根本性转变。到2026年，SaaS应用有望变得更加智能、个性化、自适应且具备自主性，逐步演变为能够从自身经验中学习的实时工作流服务联盟。这一演变将可能颠覆传统的定价模式，基于订阅和按席位授权的模式或将被融合了按使用量和按成果计费的混合模式所取代。从长远来看，一些人甚至认为，足够先进的代理式AI可能会取代现有的企业SaaS。所有这些变化都将增加财务规划、运营、生态系统管理和价值衡量方面的复杂性。

半导体供应链：新技术涌现，挑战依旧

随着针对下一代关键AI芯片技术的贸易限制不断升级，企业领导者应迅速调整，以增强供应链的韧性。

长期以来，制造最先进的芯片一直意味着要应对脆弱的供应链体系，而如今这一挑战的紧迫性已显著加剧。极紫外光刻 (Extreme Ultraviolet Lithography, EUV) 技术多年来一直受到严格限制，预计到2026年，某些支持先进人工智能模型的其他尖端技术及软件工具将成为新的供应链瓶颈。这些高科技工艺与材料大多依赖少数供应商主导，其在关键地区的垄断地位已促使多国政府为保护战略利益、降低对外依赖而实施贸易壁垒，这凸显了它们在全球半导体供应链中扮演的关键角色。

简短而精彩：微剧凸显短视频内容的演变

从独立创作者到大型平台，微短剧正在帮助重新定义观众与内容的连接和消费方式。

微短剧——这种以片段化、移动优先的剧集形式呈现的剧本化视频系列——正在重塑全球观看习惯。当前微短剧应用已创造数十亿美元营收，其中美国市场增长最为显著。德勤预测，到2026年，应用内微短剧的收入增长将翻倍，达到78亿美元。德勤还预测，美国将在2025年占据全球收入的一半，但随着其他市场将更多观看量和下载量转化为收入，其份额将下降至40%。微短剧融合了短内容的便捷性与连

续剧的叙事魅力，精准契合当代观众碎片化时间、移动端为主的观看特性。在新技术的赋能下，独立创作者正在组建精简灵活的制作团队，未来可能对规模更大、模式更传统的内容制作机构形成挑战。

视频播客称霸：品牌迎机遇，传统视频陷竞争

播客正逐渐转变为以视频为主、多语言的媒介，其覆盖范围迅速扩大，有助于品牌触达全球受众，同时占据观众更多的屏幕使用时间。

视频播客（vodcasts）正通过融合音频叙事与视觉吸引力改变受众的参与方式，并可能与电视及流媒体平台形成直接竞争。德勤预测，2026年全球播客及视频播客年度广告收入将达到约50亿美元——同比增长近20%。印度、尼日利亚和巴西等新兴市场正通过本地化、多语言内容推动这一增长。克服内容可发现性、盈利模式及规模扩展等方面的挑战，很可能成为行业持续增长的关键。

自主新纪元：引领技术主权征程

各国和地区集团正竞相建设自身的主权技术和AI基础设施。这将带来哪些影响，全球企业又该如何应对？

随着全球地缘政治环境日益复杂和不确定，企业和政策制定者正敦促各自国家和地区加强对数字基础设施的直接控制，尤其是与AI相关的部分。对主权的追求并非新现象，但向技术主权的转变可能会在2026年进一步加速。在未来的十年中，大量投资将涌入云计算、半导体、数据中心、AI模型、连接技术以及卫星通信等领域。在相互关联的世界中，任何国家或地区都难以实现完全的主权，但许多国家和地区正致力于至少实现更大的自主性。

生成式AI视频完美适配社交媒体，或成社交媒体公司威胁

最新的生成式AI视频模型正接近好莱坞水准，似乎正在推动独立视频的快速发展，但也可能引发针对社交媒体平台更严厉的监管反应。

生成式视频技术可能赋能独立创作者并提升平台广告收入，但同时也存在淹没受众、削弱内容真实性及助长虚假信息的风险，这可能引发更严格的监管审查。德勤预测，2026年生成式视频或将在美国触发监管应对措施，具体可能包括：推动更多州实施更广泛年龄验证机制，重启对1996年《通信规范法》第230条保护条款的联邦层级挑战，以及要求社交平台对发布的AI生成内容进行标识。成功的关键可能在于平衡创新与监管，若生成式视频技术无节制发展，可能冲击商业模式、加速虚假信息传播，并进一步瓦解社会对现实的共同认知基础。

当公共媒体遇上流媒体巨头：传统电视可持续发展的破局之道

公共服务广播机构（PBSs）正积极在社交媒体平台发布内容，与流媒体服务商联合制作节目，并与大型视频发行商建立合作关系。它们可为美国营利性媒体公司提供可借鉴的经验。

公共服务广播机构正通过三种核心策略应对传统电视网络面临的普遍压力：与流媒体平台联合制作节目、在社交平台推广内容、尝试分阶段发布模式。这些策略有助于扩大传播范围、吸引年轻受众，并将本土内容注入全球平台。2025年合作进程明显加速，短短几个月内就出现了三项具有标志性的广播机构与流媒体合作协议。德勤预测，2026年将再出现几项广播机构与流媒体的重要合作。此外，我们也预期会看到更多以公共服务广播机构为主导的联合制作及其他创新举措。其灵活应变的能力，可为面临流媒体和社交视频冲击的美国广播机构及小众制片方提供借鉴。然而，公共服务广播机构在探索可能影响其公共服务使命的营利性合作关系时，仍需保持审慎态度。

下一代卫星互联网正在全球范围内改变定价、容量及监管格局

卫星连接技术在直连设备（D2D）领域持续增长，但常面临变现难题；与此同时，低地球轨道（LEO）数据扩展和技术进步正在重塑部署方式与系统韧性，并带来复杂的监管挑战。

德勤此前曾预测，2024年直连设备（D2D）网络基础设施（主要为卫星）支出为30亿美元，但实际已达约40亿美元，预计到2026年将增至60亿至80亿美元。德勤同时预测，约1000颗D2D卫星将为缺乏地面蜂窝网络覆盖的区域提供低带宽连接服务（SOS紧急呼叫、短信及语音），部分D2D网络更致力于提供高速服务。目前D2D服务的用户接受度与付费意愿仍存在不确定性，这意味着其盈利模式与商业模式尚未明晰。我们进一步预测，到2026年底，低地球轨道通信卫星数量将达到1.5万至1.8万颗，为全球超过1,500万用户提供连接服务。2026年低轨卫星领域的另一趋势是，新兴运营商可能以低成本月度宽带服务冲击新兴市场电信商，而非像部分其他卫星供应商那样与地面电信商合作。

部分移动用户消费新取向：赠礼狂欢vs 千兆孤寂

发达国家市场的部分消费者难以察觉到网络性能的提升。电信公司应考虑推出更具创意的服务以扩大市场份额。

德勤预测，到2026年，在发达市场，移动运营商的奖励计划对主流消费者的重要性可能会与网络性能相当，甚至更胜一筹。从中期（截至2030年的未来五年）来看，出现连接移动网络的革命性新设备或变革性网络应用的概率较低。在这十年的剩余时间里，随着网络持续升级，非网络福利对吸引用户、减少流失可能日益重要——对消费者而言，这类福利或许比网络基础设施的升级更具实际感知价值。

Gillian Crossan

Global

Girija Krishnamurthy

Global

Jody McDermott

Canada

Tim Bottke

Germany

Deb Bhattacharjee

United States

1. Nick Lichtenberg, "Without data centers, GDP growth was 0.1% in the first half of 2025, Harvard economist says," *Fortune*, Oct. 7, 2025.
 2. Deloitte analysis of historical S&P500 data. As of December 31, 2008, technology weighting was 15.27% and communications services was 3.83%, for a combined TMT total of 19.1%. As of October 31, 2025, information technology weighting was 35.02%, communications services was 10.94%, and two consumer discretionary stocks that are generally considered tech stocks have a combined weighting of 6.68%, for a total of 52.64%.
-

ACKNOWLEDGMENTS

The authors would like to thank **George Johnston, Pedro Barros, Stephen Hipkiss, Lorraine Barnes, Nick Seeber, Chris Arkenberg, Duncan Stewart, Susanne Hupfer, Debapratim De, Cornelia Calugar Pop, Eytan Hallside, Jonas Malmlund, Steve McMullen, Ian Stewart, Rupert Darbyshire, Andy Cowen, and Ralf Esser** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

现有应用植入生成式AI, 打破独立AI的“独大”局面

生成式人工智能 (*Generative AI*) 或许是这个时代最具影响力的技术之一, 通过将其集成到现有主流数字应用中, 其用户群体的扩展速度可能会超过作为独立应用使用时的速度。

德勤预测, 到2026年, 生成式AI的用户基数将大幅增长, 这一扩张主要归因于集成生成式AI功能的现有应用程序。德勤预计, 通过现有应用程序使用生成式AI的人数将超过使用独立生成式AI工具的用户。简而言之, 2026年及以后被动使用将超过主动、显式的使用。

德勤预测, 在搜索中整合生成式AI (即提供结果综述的搜索) 的日常使用普及度, 将比任何单一功能的独立生成式AI工具 (包括文本、音频、图像、视频、代码或多模态工具) 高出三倍。¹我们预计到2026年, 在发达市场每天将有约29%的成年人发起一次或多次包含生成式AI摘要的搜索, 而使用独立生成式AI应用的比例仅为10%。我们进一步预测, 到2027年这两种搜索形式的日使用率均将提升, 但3:1的比例将保持不变: 每日使用搜索摘要功能的人群将达40%, 而使用任何独立生成式AI应用的比例仅为13%。为便于比较, 本预测仅聚焦于单一被动应用场景 (图1)。

德勤进一步预测, 在目前采用率相对较低的群体中, 尤其是年龄较大的群体, 通过其他应用程序被动使用生成式AI的增速将最为显著。

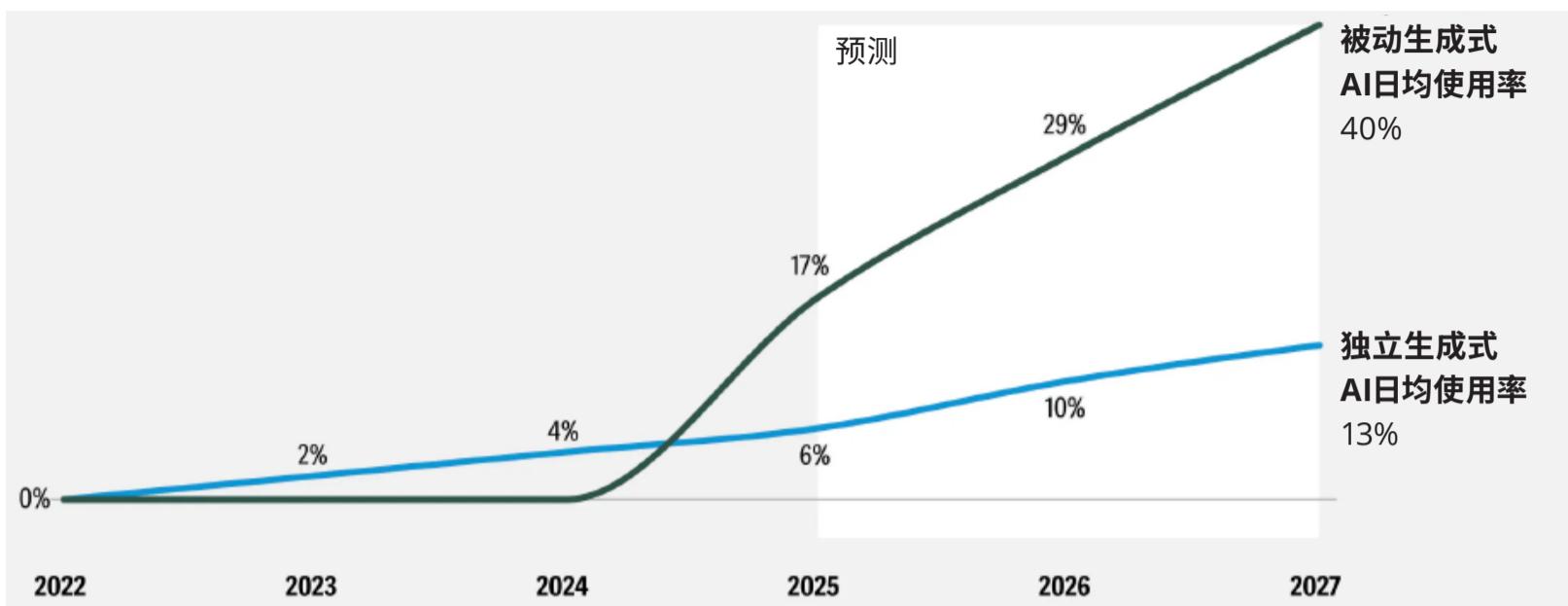
被动式与独立式生成式AI

生成式AI技术被被动使用的常见场景包括搜索、电子商务、社交媒体和在线新闻。这种使用现有应用程序中的生成式AI与我们称之为“传统”的独立生成式AI应用程序 (如Chat GPT或Gemini) 的使用方式形成对比, 用户通常会在设备上打开这些应用, 专门用于生成某种输出, 无论是文本、图像、代码还是其他类型的内容。

在被动使用生成式AI的情况下, 该技术作为嵌入式、不可或缺但不显眼的功能存在于其他应用程序之中。用户并未明确直接使用生成式AI, 但该技术却是用户体验的核心组成部分。例如, 生成式AI可用于整合搜索结果中的多个回答; 也可用于汇总针对某一产品的成千上万条独立评论; 还可用于创建通过社交媒体或在线新闻平台传播的内容。

图1：被动搜索摘要的日均使用量高于独立工具

2023-2027年每日使用被动搜索与独立生成式AI工具的用户比例



2023-2025年数据：加权基数，英国：所有16至75岁受访者，2023年（4,150人），2024年（4,150人），2025年（4,150人）
来源：德勤基于《2025数字消费者趋势，英国》预测

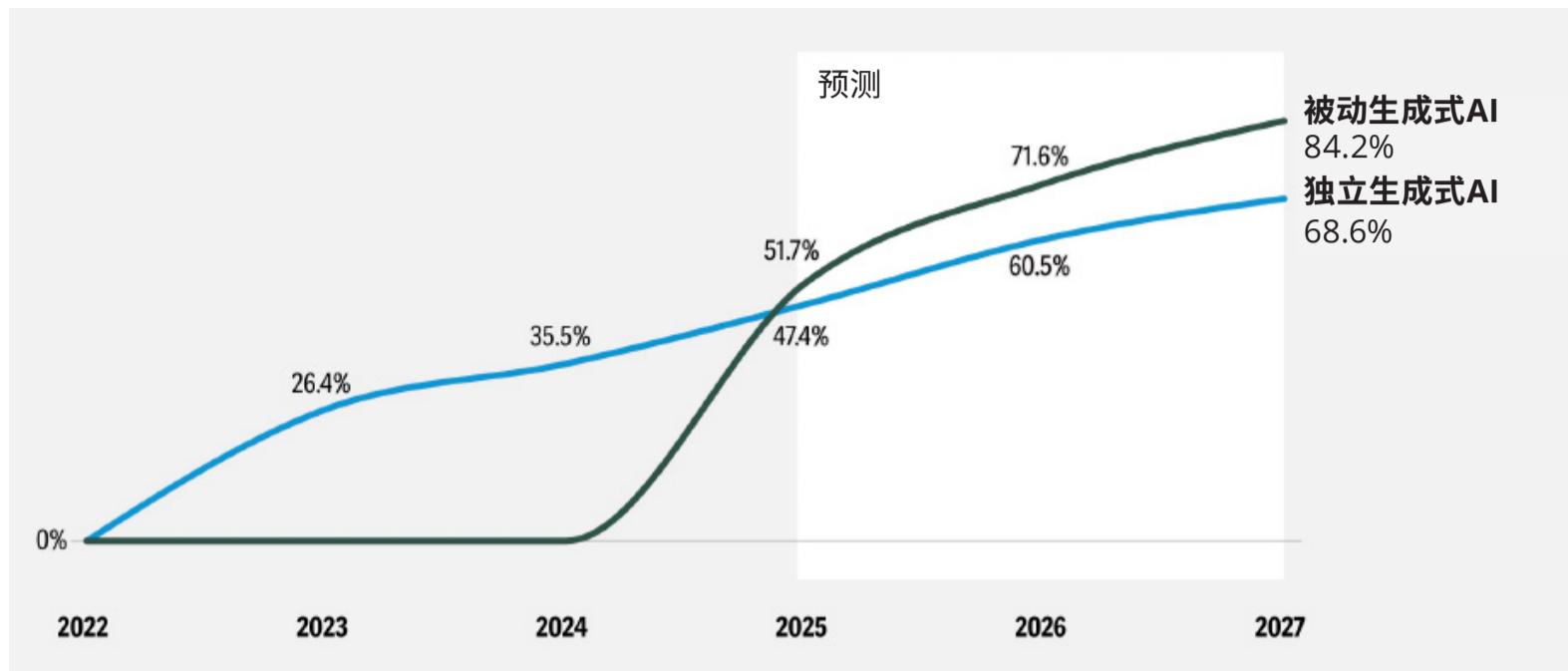
Deloitte. deloitte.com/insights

德勤估算，若将全部被动生成式AI应用与全部独立主动使用的应用进行比较，可发现前者的使用普及度在2025年已显著领先。以英国市场为例——该市场往往是新兴产品的首发试验田——德勤英国的研究发现，截至2025年中，约四分之三的受访者已使用过四种被动生成式AI应用中的至少一种，²这一比例显著领先于曾使用过任何专用独立生成式AI应用的受访者比例（47%）。

衡量新兴应用的另一个指标是将其在相对较短的发展历史中任意时间段的使用情况进行比较。被动式生成式AI应用于2024年5月首次在美国推出。包含首批搜索摘要，³并于当年11月宣布扩展至其他市场，⁴在2022年底首款流行的独立生成式AI应用发布近两年后。尽管独立应用具有先发优势，但我们预测到2026年年中，使用搜索概览功能生成内容的成年人比例（72%）将超过曾使用过独立生成式AI工具的成年人比例（61%）。

图2：使用过被动生成式AI搜索摘要的用户数量多于使用过独立生成式AI工具的用户

2023-2027年使用搜索摘要与独立生成式AI工具的用户比例



加权基数：所有16至75岁受访者，2023年（4,150人），2024年（4,150人），2025年（4,150人）

来源：德勤基于《2025数字消费者趋势，英国》预测

Deloitte. deloitte.com/insights

这一预测表明，生成式AI作为现有主流应用中的一个基本功能，其普及程度和广泛性将远超作为独立应用的形式。如果我们的预测正确，这并不代表独立的生成式AI本身缺乏价值，而是表明当该技术融入已具备主流地位的应用时，其使用频率可能大幅提升，从而创造更大的整体效用。当然，这亦隐含着对专用生成式AI工具中期普及率的思考：它最终会像社交媒体或搜索引擎等在线服务一样普及吗？抑或其日活跃用户量将稳定在全球网民总数的五分之一左右？

从用户对被动搜索的偏好中，我们能学到什么？

搜索、社交媒体和电子商务已经是所有数字应用中最常使用的：平均每天有超过150亿次的搜索量。用户每天在社交媒体上花费超过两个小时。⁵仅2025年第一季度，美国的电子商务销售额就达到了3,000亿美元。⁶相比在不熟悉且新颖的生成式AI聊天机器人中进行搜索，用户可能更倾向于在熟悉的搜索工具中使用生成式AI功能。

到2026年，市场不仅会继续探讨生成式AI对搜索模式的持续影响，将更加关注集成AI的搜索功能对Chat GPT、Synthesia等独立生成式AI工具的冲击。⁷德勤研究显示，当前工作场景中最普及的生成式AI应用正是搜索功能。部分原先的生成式AI独立应用的用户，可能会转而回归主流搜索平台。

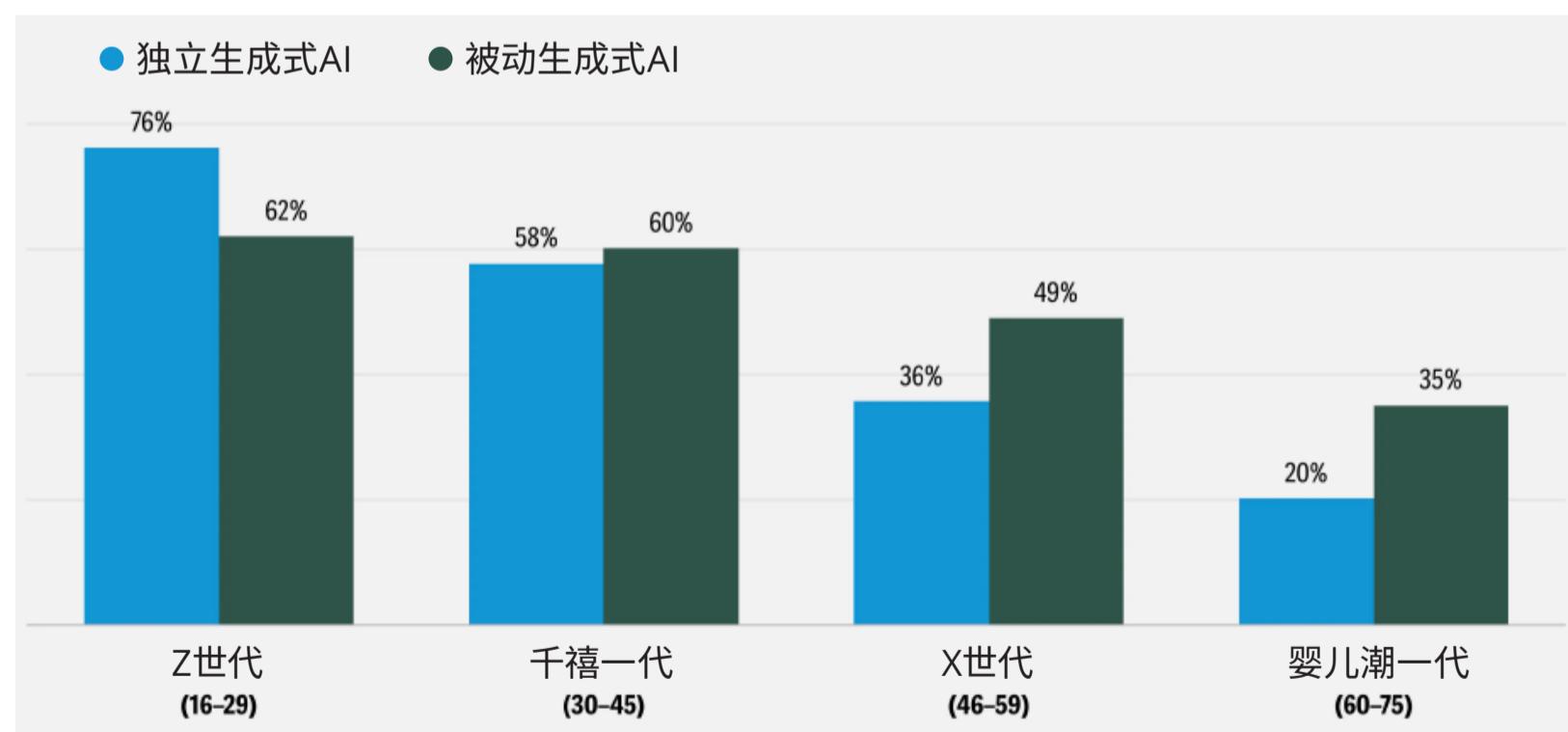
被动生成式AI超越独立生成式AI的速度令人印象深刻，或许也在意料之中。独立生成式AI呈现出新颖性，且被视为相对实验性的技术。这需要技巧和耐心：不尽如人意的输出结果可能源于提示词不佳，而非模型本身存在缺陷，解决方法是重新调整提示词。⁸问题可能出在用户的提示工程技能上，而非产品本身。如果被动式生成式AI作为一项渐进式功能，能够无缝集成到现有的主流数字应用中，例如搜索引擎、电子商务网站、社交媒体应用或办公生产力工具，那么其使用阻力将更低。用户几乎无需重复尝试。该技术更加隐蔽，使用体验也更为熟悉，因此可及性更高，需求也相应更大。将生成式AI应用于搜索结果的摘要生成，可以自动完成一项原本需要用户手动操作的任务，即点击并阅读多个链接以形成个人总结，而对部分用户而言，这一过程繁琐耗时，摘要功能则消除了这一负担。生成式AI在现有应用中的集成，类似于电子商务网站中集成一键结账（包含支付功能），或消费者银行应用程序中整合面部识别认证功能。

各世代的使用趋势

年长一代对被动AI搜索摘要功能的高接纳度尤为明显——这类用户往往不太愿意费心掌握新的独立工具。截至2025年中，婴儿潮一代对独立生成式AI仍持谨慎态度。德勤研究发现，尽管该群体对生成式AI的认知度高达58%，但其实际使用率仅为20%。相比之下，Z世代在2025年使用过生成式AI工具的比例高达76%，近乎前者的四倍（图3）。然而在婴儿潮一代中，搜索摘要功能的使用率（35%）比任何独立工具高出75%。

德勤预测，婴儿潮一代对被动生成式AI的使用率增速将高于独立生成式AI，到2026年搜索摘要的使用率将达到49%，2027年达到59%，后者明显高于独立生成式AI的使用率（32%）。

图3：2025年，年长一代使用被动生成式AI搜索的频率，超过为任何目的使用独立生成式AI工具的频率



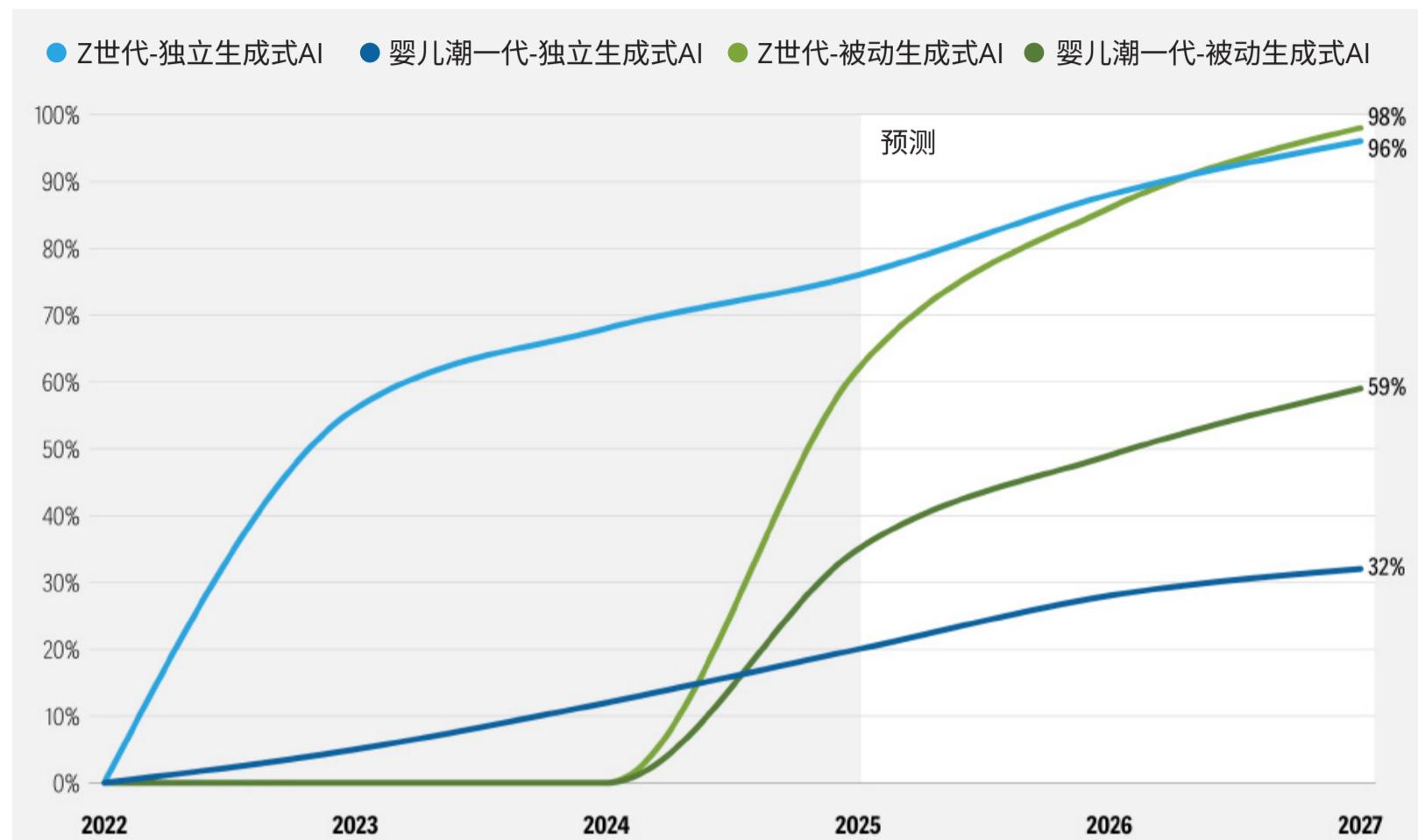
加权基数：所有16至75岁的受访者，2025 (4,150)

来源：德勤基于《2025数字消费者趋势，英国》预测

Deloitte. Insights deloitte.com/insights

图4：在婴儿潮一代与Z世代群体中，首年被动生成式AI搜索摘要的使用频率超过了首年独立AI的使用频率

在婴儿潮一代和Z世代群体中，使用过独立生成式AI或被动生成式AI搜索摘要的受访者比例



来源：德勤基于《2025数字消费者趋势，英国》预测

加权基数：所有16至75岁受访者，2023年（4,150人），2024年（4,150人），2025年（4,150人）

Deloitte. deloitte.com/insights

小结：被动式AI使用对生成式AI市场具有影响

生成式AI是当前最重要的技术之一，但其全部潜力可能只有在以下情况下才能实现，它还作为一项隐蔽但不可或缺的功能，集成于现有的主流应用程序之中。

如今许多重要技术最初都以独立功能的形式出现，往往搭载于专用设备中。不久之前，GPS或卫星导航还意味着需要携带实体设备——它们如此实用，以至于用户出差度假都会随身携带。随后这项功能被整合进智能手机及其应用之中。如今卫星导航已融入路线规划之外的无数应用场景，其使用变得至关重要、无所不在，且隐于幕后。

生成式AI虽未必完美，但能持续优化现有应用。它可以对搜索结果进行摘要处理，尽管可能在此过程中产生错误，但在多数场景下或许无伤大雅。更重要的是，用户或许愿意以生成式AI简化搜索流程的优势，来抵消其固有概率算法可能引发的错误。

对于市场中众多独立生成式AI应用的开发者而言，2026年需要思考的核心战略问题是：究竟将工具功能嵌入其他应用平台，还是继续保持独立界面？后者虽能带来更高的单用户收入，却可能限制用户规模。少数企业或可兼顾两条路径，但对大多数业者而言，必须做出明确抉择。

中国生成式AI前瞻

目前，中国的生成式AI市场正处在爆发式增长的高质量增长期内，生成式AI发展迅速，用户普及度不断升高，但同时相应的高技能人才缺口较大，中国整体生成式AI行业处在蓬勃发展的初期。

用户规模不断扩大，多项数据位列全球第一

截至 2025 年 6 月，中国生成式AI用户规模达5.15亿，在这些用户中，有超90% 用户首选国产大模型，已有 538 款生成式 AI 服务完成合规备案，应用场景持续拓宽，覆盖办公、创作等日常场景及农业、工业等多项不同领域，同时具身智能成为发展热点，在知识产权领域，我国人工智能专利申请量累计达 157.6 万件，以 38.58% 的全球占比位居世界第一。⁹

AI国产算力自主创新提速

中国在AI模型层面已形成显著优势；硬件领域虽尚未完全突破算力瓶颈，但市场格局已出现关键转变——国内科技企业已从此前普遍囤积进口硬件，转向采用国产算力解决方案。中国拥有庞大的市场规模，头部企业盈利能力强劲，为AI算力发展提供充足需求支撑。

中国生成式AI市场占比不断提升

2024年中国生成式AI占AI市场投资总规模18.9%。随着生成式AI技术的快速发展，预计在2028年中国生成式AI投资占比将达到30.6%，投资规模超300亿美元，CAGR51.5%，发展速度迅速。¹⁰

中国生成式AI政策力度加大

中国政府已构建起“发展-安全-治理”三位一体的AI产业发展保障框架，出台《关于深入实施“人工智能+”行动的意见》《生成式人工智能服务管理暂行办法》等支持政策。未来，要在2027年实现人工智能与6大重点领域深度融合，2030年推动智能经济成为经济增长核心极，2035年全面迈入智能社会的三阶段性目标；同时配套出台算力协同共享、公共数据开放、算法备案等落地措施，推动AI技术创新与产业的合规发展。

ROI压力不断加大

AI应用已广泛普及，未来的生成式AI市场，生成式AI正不断从技术探索转向价值兑现，企业将更加注重生成式AI举措的投资回报 (ROI)，更愿意为生成式AI驱动的业务增长付费。

AI人才需求缺口加大

未来中国AI市场内增长不断加快，AI人才需求业不断增加中国目前面临AI人才的较大缺口，当前，中国人工智能人才缺口超过500万，供求比例达1:10。未来通过校企合作、内部‘数字工匠’培养等方式加大培养人才力度将愈发普遍。¹¹

Paul Lee
United Kingdom

Girija Krishnamurthy
Global

Steve Fineberg
United States

Ben Stanton
United Kingdom

Tim Bottke
Germany

Roger Chung
China

Gillian Crossan
Global

ENDNOTES

1. Deloitte's forecast is based on multiple sources, including its proprietary research undertaken as part of Deloitte's Digital Consumer Trends survey, fielded in April and May 2025, and also in 2023 and 2024. This longitudinal data set provides a trajectory for the adoption of standalone gen AI apps. Our proprietary data set includes surveys conducted in multiple developed markets globally. Additionally, we have considered multiple other data points, including Alphabet's reporting on the volume of AI Overviews, which had a monthly usage base of over two billion as of July 2025. See: [Alphabet, "Alphabet announces second quarter 2025 results,"](#) July 23, 2025.
2. Paul Lee and Ben Stanton, "[Digital Consumer Trends 2025, UK edition](#)," Deloitte, June 2025.
3. Elizabeth Reid, "[Generative AI in search: Let Google do the searching for you](#)," Google, May 14, 2024.
4. Hema Budaraju, "[New ways to connect to the web with AI Overviews](#)," Google, Aug. 15, 2024.
5. Josh Howarth, "[Worldwide Daily Social Media Usage \(New 2025 Data\)](#)," Exploding Topics, June 23, 2025.
6. United States Census Bureau, "[Quarterly retail e-commerce sales](#)," press release, Aug. 19, 2025.
7. Danny Goodwin, "[Google search is 373x bigger than ChatGPT search](#)," Search Engine Land, March 11, 2025.
8. Kara Kennedy, "[Poor prompts lead to misleading research](#)," AI Literacy Institute, Aug. 19, 2024; Ulster University, "[Generative artificial intelligence \(Gen AI\): Prompt engineering](#)," Oct. 23, 2025; Haringun Nur Adha, "[You made a specific prompt but the results are disappointing? Maybe you're using ChatGPT wrong](#)," Medium, Sept. 16, 2025.
9. [从 5.15 亿用户到具身智能: 2025 生成式 AI 报告拆解中国产业新图景—第七十二期—中国远大集团](#)
10. [IDC: 2028年中国人工智能总投资规模将突破1,000亿美元, 五年复合增长率达35.2%](#)
11. [人工智能人才需求大 供求比例1:10_行业资讯_数字中国建设峰会](#)

ACKNOWLEDGMENTS

The authors would like to thank **George Johnston, Pedro Barros, Stephen Hipkiss, Lorraine Barnes, Nick Seeber, Chris Arkenberg, Duncan Stewart, Susanne Hupfer, Debapratim De, Cornelia Calugar Pop, Eytan Hallside, Jonas Malmlund, Steve McMullen, Ian Stewart, Rupert Darbyshire, Andy Cowen, and Ralf Esser** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

为何生成式AI下一阶段将要求更高算力？

世界正从单纯训练生成式AI（Gen AI）模型转向大规模应用。许多人认为这意味着消费者边缘计算增加，而数据中心计算将减少。但在2026年，这两者都不太可能发生。

人们普遍预计，到2026年，生成式AI计算将从目前主要依赖海量数据训练模型，转向利用这些模型来思考并回答企业和消费者的问题、提示及任务，这一过程被称为“推理”（inference）。许多人推测，这种计算工作负载（即“计算”）的转变意味着AI生态系统将需要专门针对推理任务优化的芯片，而这些（或许成本更低的）芯片可能会被部署在大型数据中心之外的边缘设备上。这甚至意味着我们未来可能需要更少、更小，或至少是不同类型的计算中心，并减少相关支出。

德勤的看法略有不同：推理工作负载确实将成为2026年的热点，占所有计算量的约三分之二（2023年为三分之一，2025年为一半）。¹ 并且推理优化芯片的市场规模将在2026年增长至超过500亿美元。但德勤同时预测，大部分计算仍将依赖于价值2,000亿美元或更多的尖端、昂贵且高功耗的AI芯片，这些芯片主要部署在价值4,000亿美元或更多的大型数据中心，或价值500亿美元或更多的企业本地化解决方案中，而这些解决方案使用的芯片和机架与数据中心相同，而非边缘设备上的芯片。这意味着我们可能仍需要当前正在规划的所有数据中心和企业本地AI工厂，以及这些数据中心所需的全部电力资源。

不断增长的AI计算需求

尽管对新模型训练算力的需求增长可能已经放缓（很可能仍在增长，但增速低于2023年和2024年），² AI模型正通过训练后改进的先进技术持续演进。这些方法，加上海量的推理查询，很可能意味着计算需求将持续增长而非减少。换句话说，尽管得益于摩尔定律，用于计算的芯片每年都在变得更加高效，但计算需求的增长速度更快，预计到2030年每年将增长4至5倍。³

用于初始训练的计算需求增长正在放缓

2020年的一篇论文表明，模型规模越大、训练数据越多、采用越先进的AI加速芯片，产出的结果就越好：这构成了生成式AI的首个扩展定律。⁴ 到2022年和2023年，训练模型的参数量已从10亿级增长至千亿级，并进一步达到万亿规模。⁵

2024年，两个问题开始浮现：可用的训练数据并非无限，而不断增大的模型正呈现出收益递减的趋势：十倍的训练数据可能仅能产生一个比前一版本略好一点的“最先进的”AI模型，甚至可能毫无提升。⁶与此同时，更小且更高效的AI模型似乎有可能利用更少的数据、更短的时间、更少的资金和更少的芯片，开发出真正达到顶尖水平的AI模型。⁷

如果训练规模的增长放缓，AI计算将越来越侧重于推理。让一个大语言模型对文档进行摘要（推理的一个例子）所消耗的计算资源，仅相当于训练该模型所需资源的一小部分。然而，逻辑在于，随着数十亿消费者和企业员工更频繁地发起此类请求，所有这些推理任务累积起来，将使整体计算工作负载从训练转向推理。其中部分请求可以在消费者和企业设备（如智能手机和个人电脑）上处理，而正如德勤在2024年预测的那样，2025年已有数亿台搭载了支持本地AI加速芯片的个人电脑和智能手机售出。⁸此外，由于推理的计算密集程度低于训练，数据中心内部或许可以采用专门针对推理优化的芯片。这类芯片成本更低，每次推理所消耗的能量也少于用于大规模训练的高性能AI芯片，并且可能不需要配备大量昂贵的共封装内存。⁹

上述所有情况均发生在2025年，并可能在2026年持续。德勤2025年的调查显示，无论是在美国还是全球范围内，使用生成式AI的消费者数量都在增加，且每日使用的人数也在上升。¹⁰诸如个人电脑和智能手机等边缘设备越来越多地搭载了AI加速器。许多针对推理任务优化的芯片（专用集成电路，ASIC）已被设计、制造，并正在数据中心和部分边缘设备中部署。这些芯片包括但不限于来自Meta、Google、Amazon、Intel、AMD、Qualcomm、Groq、SambaNova、Cerebras和Graphcore的产品，其中一些基于Broadcom的封装解决方案，由设计方提供处理核心。¹¹尽管这些不同芯片的销售数据并未公开，但德勤认为，这些芯片在2025年的总收入合计将超过200亿美元，并在2026年达到500亿美元或以上。¹²

既然如此，为何我们仍需在数据中心部署耗电量巨大、单价高达3万美元甚至更高的芯片（到2028年，此类芯片总成本或将达到约4,000亿美元），¹³而仅2026年数据中心的建设成本就预计高达4,000亿美元，且到同年，其年度潜在成本可能攀升至1万亿美元？¹⁴

AI模型的训练比过去更加复杂

第一个扩展定律的目的是生成“更好”的AI模型，而且在最初几年效果非常显著。这种最初的扩展形式过去被称为训练，现在则称为“预训练”，用于生成基础模型。

事实证明，还有两种方法可以构建更优秀的模型：一种是后训练扩展，包括各种技术，如微调、剪枝、量化、蒸馏、基于人类反馈以及越来越多地基于AI反馈的强化学习，以及合成增强数据。¹⁵另一种是测试时扩展，也称为长时思考，即模型在被提问后，通过多种技术在推理过程中逐步进行思考，例如思维链提示、基于多数投票的采样、搜索，甚至一些后训练技术。¹⁶变得更加准确，拥有更多选择、更优质的来源以及更少的错误信息。¹⁷

新的高能耗AI技术可能会超过能效提升的速度

首先，训练后扩展和测试时扩展似乎已成为新的常态：目前大多数AI公司都采用这些方法以多种方式提升AI模型的性能。¹⁸

其次，这两者都非常消耗AI算力。据估计，训练后阶段总体使用的计算量是训练原始基础模型的30倍，而长链推理所消耗的算力更是一次简单推理（例如让AI总结一封邮件）的100倍以上。¹⁹

第三，由于这两种扩展技术被广泛使用且计算资源消耗大，这对AI数据中心及其地理位置和电力需求、用于AI数据中心的芯片（以及需要执行AI的其他场所）、往年的AI芯片、边缘设备等均产生了影响。

关于德勤预测2025年及以后AI数据中心所需芯片的简要回顾

数据中心已经存在了几十年。事实上，全球有数千万平方英尺的数据中心，多年来每年都有价值数百亿美元的半导体组件被售出，用于数据中心。²⁰但新的AI数据中心以及实现这些数据中心的新半导体，往往与昨日的数据中心和半导体截然不同，简直是天壤之别。

下一代AI数据中心的建设成本可能每年高达数千亿美元，并消耗数百吉瓦的电力。在大多数此类数据中心中，冷却方式可能与以往几代数据中心不同，电源和电压可能不同，内部通信技术也可能不同，甚至地板可能需要更厚，以支撑密度更高、重量更大的服务器机架。或许最重要的是，新型AI服务器机架不再采用以CPU为中心、内存就近配置的服务器架构，而是主要由一种名为图形处理单元（GPU）的专用芯片构成。²¹这些组件通常配备专用的高带宽内存（HBM），并与GPU紧密集成，同时还配有专门的CPU来协调庞大的AI计算工作负载。许多组件都是针对新一代AI数据中心的独特需求和规模而设计的。²²

直到2006年，高端GPU仍被认为仅适用于游戏电脑和游戏主机，而非数据中心。²³大多数数据中心的任务由CPU即可很好地完成，因为CPU主要是串行处理器，任务按顺序执行。一些高性能计算机，即“超级计算机”，则配备了称为“大规模并行处理器”的特殊芯片，能够同时执行数百个任务。然而，这些芯片的成本通常是游戏用GPU或数据中心CPU的数十倍甚至数百倍。

2009年，科学家们注意到游戏用的GPU同样是并行处理器，于是尝试在高端GPU上运行机器学习模型，这些GPU与游戏电脑中使用的完全相同。²⁴它们运行良好，几年之内，经过特别优化的GPU（与游戏版本略有不同）开始被用于一些数据中心和部分本地设备上，以执行机器学习AI任务。²⁵但直到2018年，该市场的规模每年仍仅为数十亿美元级别。²⁶

2022年，生成式AI的大语言模型（LLMs）的发展需要更加专业化的GPU，而且通常要求将这些GPU与一种相对较新且专用的内存（HBM）集成在同一个先进的封装中。²⁷这些GPU+HBM组件还需要一个设备来协调和管理数据流。经过优化的专用CPU（虽然核心架构相似，但不同于计算机、智能手机或数据中心中的CPU）也是生成式AI数据中心的重要组成部分，此外还有多个其他可能同样关键的组件。到2025年，全球几乎所有的Top 500超级计算机都采用了类似的GPU、专用内存和CPU组合。²⁸从某种意义上说，正在建设的超大规模AI数据中心可以被描述为一种专用超级计算机的版本。

小结：更多的算力需求可能对AI生态系统产生影响

企业和高管应当为未来计算需求持续增长做好准备——特别是在大型数据中心和企业AI工厂领域，这种增长部分源于训练后阶段及测试阶段的扩展需求。尽管面向推理优化的芯片和边缘处理将实现增长，但对超大规模数据中心和企业级AI专用设备的投资仍不可或缺。值得注意的是，“针对推理优化”并不等同于低功耗：近期某款专为推理预填充计算优化的产品通过采用GDDR7显存替代HBM显存，但其单机柜功率需求高达370千瓦，达到同厂商训练版本产品功率密度的近三倍！²⁹

AI数据中心：预计到2026年，全球AI数据中心的资本支出将达到4,000亿至4,500亿美元。³⁰其中超过一半的支出用于设备内部的芯片（2,500亿至3,000亿美元），³¹其余部分则包括土地、建设、电力和许可等所有其他费用。此外，预计到2028年，AI数据中心的资本支出将上升至1万亿美元。³²当年AI芯片市场规模超过4,000亿美元。³³尽管预训练阶段的增长正在放缓，且计算资源正转向推理阶段，但训练后扩展、测试时扩展以及使用量增加所带来的计算需求表明，全球可能仍需要大量数据中心，从2025年约3,000亿至4,000亿美元增长到2028年约1万亿美元的规模，这一趋势是基本合理的。

AI数据中心的位置：预训练一个拥有100万亿参数的大语言模型可能需要数周时间，且对微小的中断极为敏感：关键组件的故障或处理器之间过高的延迟交接，都可能导致所有工作成果的丢失，而不得不重新开始。大多数基础模型的预训练通常集中在同一栋建筑或园区内进行，所有服务器和机架都位于同一点。然而，越来越多的AI计算任务现在可以在美国不同地区的数据中心，甚至在全球范围内的多个数据中心之间协同完成。³⁴此外，数据中心的规模可能从吉瓦级到较小规模的推理数据中心不等，这些数据中心可用于部署完全训练好的模型，且通常更靠近城市区域，以帮助降低延迟。这有助于形成不断增长的需求。**主权AI解决方案**（每个国家或地区都拥有位于国内甚至本地运营的AI计算能力），以及作为混合云一部分的企业本地边缘解决方案。³⁵

AI数据中心的电力需求：总体而言，进行三种扩展方式的更多AI数据中心仍将需要大量电力。但与需要一次性完成训练过程的预训练相比，后训练和推理时的扩展在一定程度上相对“可中断”。这使得AI公司能够参与需求响应计划，通过将任务转移到其他数据中心位置或降低处理器时钟频率，在用电高峰期减少电力需求。³⁶据估计，提高此类负荷的灵活性，可使大型新建数据中心有助于维持电网的可靠性和经济性。³⁷

AI的训练和推理可以分布式进行，这意味着数据中心无需集中在一个州或一个国家，而是可以更均匀地分布在全球各地，从而分散电力需求。

AI数据中心中的芯片：有些人可能将AI芯片市场视为零和博弈。这种观点通常表现为“虽然我在预训练基础模型时不得不花费数万美元购置搭载HBM的先进GPU，但随着计算重心向推理转移，或许可以采用更廉价、专为推理优化且减少HBM配置的芯片？

芯片市场似乎将不再是“非此即彼”，而是走向“两者兼有”。仅用于推理或针对推理优化的芯片很可能将迎来显著增长；但与此同时，通常最适合基础模型预训练、后训练以及测试时扩展（计算需求包含训练与推理的混合）的芯片，仍将是那些配备高带宽内存（HBM）、每块售价数万美元、性能强大且耗电量高的GPU。对于采购这些芯片的用户而言，到2026年它们的价格可能更加昂贵，预计先进制程晶圆的成本将上涨50%。³⁸

面向智能手机和PC等消费类或企业级设备的边缘AI：如前所述，数以亿计的智能手机和个人电脑正在出货并被购买，这些设备均配备了NPU：³⁹专用于处理AI推理任务的专用芯片或CPU芯片的部分模块（NPU部分价值几美元至几十美元），在功耗合理的情况下进行了优化。

然而，目前NPU的性能仅足以支持前文讨论的单次推理任务（如“总结这封邮件”等）。因此，德勤预测，2026年几乎所有的AI计算仍将主要在大型AI数据中心，或由企业拥有的相对昂贵的高端AI服务器上完成，而非在个人电脑和智能手机上进行。至少在当前所处的高速增长、抢占市场的阶段，成本优化的混合架构似乎并非供应商或企业的优先考虑事项。此外，诸如测试时扩展（test-time scaling）等技术对于绝大多数消费者应用场景，甚至大多数企业级设备端应用场景而言，都可能显得过于冗余。未来某一天，计算机和智能手机或许会发挥更重要的作用，但在2026年之前这不太可能发生。

最近，一家AI公司推出了一款具备推理能力的生成式AI模型，可在个人电脑上本地运行。目前尚不清楚该模型的实际效果如何，对电池续航会产生何种影响，或有多少个人电脑用户更倾向于在本地而非通过云端使用AI功能。⁴⁰

边缘AI与企业本地化解决方案的应用：那些通常被部署于全球大型AI数据中心的高性能、高功耗GPU，搭配HBM和协同工作的CPU模组，如今也可供希望采用本地化、混合式、更具韧性的生成式AI计算架构的企业使用，尤其适用于训练后的应用场景。出于对成本、知识产权归属、数据主权、系统韧性和定制化需求的考虑，企业可以花费约30至50万美元购置一台配备约八块GPU（以及HBM和CPU）的设备，以实现特定规模的AI训练与推理任务。⁴¹或者，他们可以花费300万至500万美元购买一个配备多达72个最先进GPU（以及HBM和CPU）的机架，其性能更强大。⁴²或者，他们甚至可以花费数千万美元购买多个机架，以实现更强大的功能。⁴³德勤预测，到2026年，这一本地混合企业市场规模将超过500亿美元。

用于机器人、无人机和自动驾驶汽车的边缘AI：到2026年规模仍然相对较小，但存在多种需要实时、本地设备推理的应用场景。这些场景涵盖从无人机、机器人到自动驾驶汽车等多个领域。目前所使用的芯片种类繁多：大多数无人机配备的是相对基础且低功耗的AI推理芯片。⁴⁴而大多数自动驾驶汽车使用的GPU芯片解决方案，其性能仅略低于数据中心所采用的芯片。⁴⁵这个非AI工厂市场的规模可能仍然相对较小（2026年不足50亿美元）⁴⁶但可能成为要大得多，尤其是如果机器人市场腾飞，这种情况可能会发生，但很可能是在2030年之后。⁴⁷

值得注意的是，我们仍处于AI发展的初期阶段。截至2025年夏季，对AI算力的需求（从而对更多数据中心、企业本地解决方案以及更高性能AI芯片的需求，无论是用于预训练、后训练、测试时扩展还是推理）增长十分迅猛，即便人们持续努力提升算法效率，这一趋势仍未减缓。⁴⁸在某个阶段，新的技术可能会取得突破，更先进的AI模型将能够在成本更低的芯片上良好运行，从而减少对数据中心的需求并降低功耗。但这不会发生在2026年。

中国AI算力市场前瞻

中国AI算力不断扩容

中国十五五规划，统筹布局全国一体化算力网络，建设国家级AI算力中心，推动算力资源普惠化，降低社会算力总成本。推动中国AI算力的不断扩容。同时中国工业和信息化部于2025年5月印发《算力互联互通行动计划》，旨在加快构建算力互联互通体系，实现不同主体、不同架构的公共算力资源标准化互联，提高公共算力资源使用效率和服务水平，促进算力高质量发展。

AI算力互通互联不断增强提效

到2026年，中国将建立较为完备的算力互联互通标准、标识和规则体系。在设施互联层面，将部署新型高性能传输协议，以提升算力节点间的网络性能；在资源互用层面，建成国家、区域及行业级算力互联互通平台，统一公共算力资源标识，实现全国主要算力服务商资源的接入与汇聚；在业务互通层面，推动计算、存储、网络等业务的协同调度，实现跨主体、跨架构、跨地域的算力智能供给；在应用创新层面，开展算力互联网试验网试点，探索普惠化的算力服务模式。这一系列举措旨在打通算力资源流动的关键环节，最终构建形成全国一体化调度、高效集约利用的算力互联网，为产业数字化升级提供坚实、普惠的算力底座。

国产芯片追赶，与国际巨头差距不断缩小

在外部环境制约和国家战略推动下，国产AI芯片正在加速崛起。以华为昇腾、寒武纪、阿里巴巴含光等为代表的本土力量正不断缩小与国际巨头的差距。未来随着国产芯片在推理赛道的进一步渗透、超节点/集群技术的规模化部署以及产业链闭环生态的完善，有望在更多细分市场实现追赶甚至领先。

国产化替代不断加快

全国多个智算中心项目中，国产化率大幅提升，中国移动呼和浩特智算中心AI芯片国产化率超85%，成渝（成都）天府数据中心国产芯片占比不低于60%，浙江东阳人工智能算力中心更是明确了70%的国产化目标。在AI算力中，国产化替代正不断加速。

Duncan Stewart
Canada

Arpan Tiwari
United States

Karthik
India

Jeroen Kusters
United States

Girija Krishnamurthy
Global

Roger Chung
China

Deb Bhattacharjee
United States

ENDNOTES

1. Rodrigo Liang, "Scaling AI without breaking the grid: The path to sustainable innovation," World Economic Forum, Jan. 3, 2025.
2. Michelle Weaver, "Big debates: The AI evolution," Morgan Stanley, Jan. 10, 2025.
3. Josh You and David Owen, "How much power will frontier AI training demand in 2030?" Epoch.AI, Aug. 11, 2025.
4. Jared Kaplan et al., "Scaling laws for neural language models," OpenAI, Jan. 23, 2020.
5. Amazon Web Services, "What are foundation models?" accessed Sept. 19, 2025.
6. Ashu Garg, "Has AI scaling hit a limit?" Foundation Capital, Nov. 27, 2024.
7. Aixin Liu et al., "Deepseek-v3 technical report," arXiv preprint arXiv:2412.19437 (2024).
8. Chris Arkenberg, Duncan Stewart, Gillian Crossan & Kevin Westcott, "On-device generative AI could make smartphones more exciting—if they can deliver on the promise," *Deloitte Insights*, Nov. 19, 2024; IDC Media Center, "Worldwide smartphone market forecast to grow 1% in 2025, driven by accelerated 3.9% iOS growth, according to IDC," Aug. 27, 2025; Gartner, Inc., "Gartner says artificial intelligence ("AI") PCs will represent 31 percent of worldwide PC market by the end of 2025," press release, Aug. 28, 2025.
9. Amazon Web Services, "AWS Inferentia," accessed Sept. 19, 2025.
10. Paul Lee and Clare Mortimer, "How citizens use devices and AI: what government needs to know," Deloitte UK, Aug. 29, 2025; Steve Feinberg, et al., "In the gen AI economy, consumers want innovation they can trust: Deloitte's 2025 Connected Consumer Survey," Deloitte, Sept. 25, 2025.
11. Wylie Wong, "Data center chips in 2024: Top trends and releases," Data Center Knowledge, April 11, 2024; Reen Singh, "AI inference chips latest rankings: Who leads the race?" Uvation, July 11, 2025; Broadcom Inc., "3.5D XDSiP AI Accelerator Platform," accessed Oct. 23, 2025.
12. Deloitte Consulting LLP performed an analysis of the data center market, including a rough bill of materials for the various components, and market sizes. This analysis is due to be published in December 2025.
13. Skye Jacobs, "NVIDIA Blackwell server cabinets could cost somewhere around \$2 to \$3 million each," TechSpot, July 28, 2024.
14. Beth McKenna, "2 key things from AMD's earnings call that investors should know," The Motley Fool, Feb. 1, 2024; Dell'Oro Group, "AI infrastructure spending sustains strong growth momentum," press release, Feb. 5, 2025.
15. Kari Briski, "How scaling laws drive smarter, more powerful AI," NVIDIA, Feb. 12, 2025.
16. Ibid.
17. Jonathan Farrington, "What is chain of thought prompting – AI prompt engineering," Silicon Dales, July 24, 2025.
18. Briski, "How scaling laws drive smarter, more powerful AI."
19. Ibid.
20. "Data centers: Computing risks and opportunities for U.S. real estate," S&P Global, Oct. 22, 2024; Equinix, Inc., "Form 10-K: Annual report for fiscal year ended Dec. 31, 2023," Feb. 16, 2024; Digital Realty Trust, Inc. and Digital Realty Trust, L.P., "Form 10-K: Annual report for fiscal year ended Dec. 31, 2023," Feb. 23, 2024.

21. Shubham Sharma, "Going beyond GPUs: The evolving landscape of AI chips and accelerators," VentureBeat, Sept. 26, 2024.
22. Deloitte Consulting LLP performed an analysis of the data center market, including a rough bill of materials for the various components, and market sizes. This analysis is due to be published in December 2025.
23. Eric Reed, "History of NVIDIA: Company and stock," SmartAsset, May 22, 2024.
24. Rajat Raina, Anand Madhavan, and Andrew Y. Ng, "Large-scale deep unsupervised learning using graphics processors," In Proceedings of the 26th Annual International Conference on Machine Learning, 2009.
25. NVIDIA, "NVIDIA delivers massive performance leap for deep learning, HPC applications with NVIDIA Tesla P100 accelerators," press release, April 5, 2016.
26. Hannah Wilson, "NVIDIA facts and statistics (2025)," Investing.com, Aug. 28, 2025.
27. Hannah Wilson, "NVIDIA facts and statistics (2025)," Investing.com, Aug. 28, 2025.
28. Top 500, "June 2025," June 2025.
29. Ray Wang, "NVIDIA's new Rubin CPX targets future of large-scale inference," Futurum, Sept. 18, 2025.
30. In 2025, Deloitte Consulting LLP performed an analysis of the data center market, including a rough bill of materials for the various components, and market sizes. This analysis is due to be published in December 2025.
31. Omdia, "New Omdia forecast: AI data center chip market to hit \$286bn, growth likely peaking as custom ASICs gain ground," Aug. 28, 2025.
32. Anthony Di Pizio, "Jensen Huang predicts annual data center spending will hit \$1 trillion by 2028. Here's the ultimate semiconductor ETF to buy right now." The Motley Fool, May 1, 2025.
33. Dave Lawler, "Exclusive: 'Massive ten-year' AI boom is just starting, AMD CEO says," Axios, Sept. 17, 2025.
34. Paul Mah, "AI training is going to multiple data centers," CDO Trends, Sept. 11, 2024.
35. Chris Thomas, Akash Tayal, Duncan Stewart, Diana Kearns-Manolatos, and Iram Parveen, "Is your organization's infrastructure ready for the new hybrid cloud?" Deloitte Insights, June 30, 2025.
36. Mike Robuck, "Google strikes deals for flexible AI data centre power use," Mobile World Live, Aug. 5, 2025.
37. Tyler H. Norris, Tim Profeta, Dalia Patino-Echeverri, and Adam Cowie-Haskell, "Rethinking load growth: Assessing the potential for integration of large flexible loads in US power systems," Nicholas Institute for Energy, Environment & Sustainability, Duke University, February 2025.
38. Anton Shilov, "TSMC could charge up to \$45,000 for 1.6nm wafers — rumors allege a 50% increase in pricing over prior-gen wafers," Tom's Hardware, June 4, 2025.
39. Francisco Jeronimo, "The rise of gen AI smartphones," IDC, July 5, 2024.
40. Dan Shipper, "Vibe check: OpenAI drops two new open-weight models," Every Media, Aug. 5, 2025.
41. Cyfuture Cloud, "NVIDIA DGX H100 price 2025: Cost, specs, and market insights," Cyfuture Cloud Knowledgebase, accessed October 2025.
42. Tae Kim, "NVIDIA's multi-million dollar AI servers are getting more expensive," Barron's, Aug. 28, 2025.
43. Skye Jacobs, "NVIDIA Blackwell server cabinets could cost somewhere around \$2 to \$3 million each," TechSpot, July 28, 2024.

44. Qualcomm, "[Flight RB5 5G platform](#)," accessed Sept. 19, 2025.
 45. Ali Kani, "[NVIDIA DRIVE Thor strikes AI performance balance, uniting AV and cockpit on a single computer](#)," NVIDIA, Sept. 20, 2022.
 46. There are a variety of suppliers for chips for driving assistance, but as one example, NVIDIA's auto segment is at a US\$2 billion run rate as of August 2025: Pras Subramanian, "[NVIDIA's auto business surges 69% from self-driving tech](#)," Yahoo Finance, Aug. 25, 2025.
 47. Karthik Ramachandran, et al, "[AI for industrial robotics, humanoid robots, and drones](#)," Deloitte Insights.
 48. Jameel Rogers, "[AI chips for data center and cloud to exceed US\\$400 billion by 2030](#)," IDTechEx, May 8, 2025.
-

ACKNOWLEDGMENTS

The authors would like to thank **Brandon Kulik, Amy Scimeca, Karan Aggarwal, Kate Hardin, Diana Kearns-Manolatos, Mike Luk, Baris Sarer, Jason Chmiel, Dan Hamling, Jan Nicholas, Jordan Bish, Nitin Mittal, Rohit Tandon, Nicholas Merizzi, and Dan Littmann** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

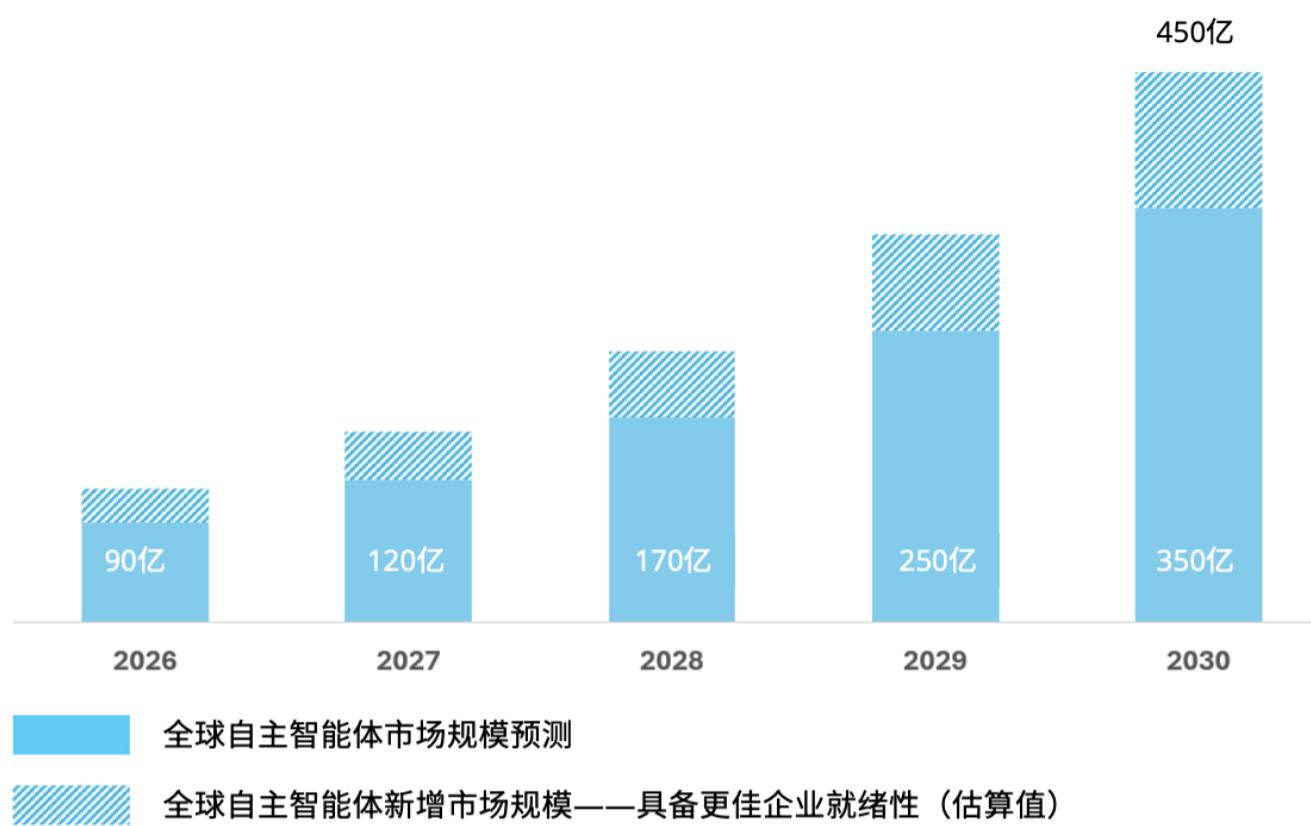
借力智能体协同调度，驱动价值指数级攀升

自主智能体 (*Autonomous AI agents*) 或许具有变革性力量，但协调机制可能是实现智能自动化的关键。开源通信协议与专有通信协议将展开竞争，争夺主导地位。

随着企业整合多代理系统——即不同的AI推理引擎在各个领域之间无缝交互——代理协调（对特定角色的代理进行有效协调）将成为充分发挥其潜力的关键。周密的协调能够释放智能化工作流的潜力，使多代理系统能够理解请求、设计工作流程、分配与协调任务，并持续验证和优化结果。¹相反，若代理协调不当，则会显著限制其为企业创造的价值。

到2026年，估计自主智能体市场规模可能会达到85亿美元，到2030年达到350亿美元（图1）。²德勤预测，如果企业能够更好地协调代理并审慎应对其相关挑战和风险，到2030年这一市场预期可能会增长15%至30%，高达450亿美元。据估算，由于成本超出预期、扩展复杂性或意外风险的出现，到2027年，将有超过40%的代理式AI (*Agentic AI*) 项目可能会被取消。³如果企业能够提前解决潜在隐患，这些项目可能会带来显著的收入增长。

图1. 全球自主智能体市场规模预测——提升准备度后（估算，十亿美元）



来源：德勤分析

Deloitte. deloitte.com/insights

为充分发挥多代理系统的潜力，企业可能需要提升自身协调能力，使代理在特定自主程度下运作，并预先应对早期潜在风险。与此同时，多代理系统更可能适用于那些注重代理互操作性及管理、并能有效推动工作流程与人才结构必要变革的企业。

让企业为多代理系统服务

企业在制定与其代理协调准备相关的决策时，以下三个指导原则可能至关重要：

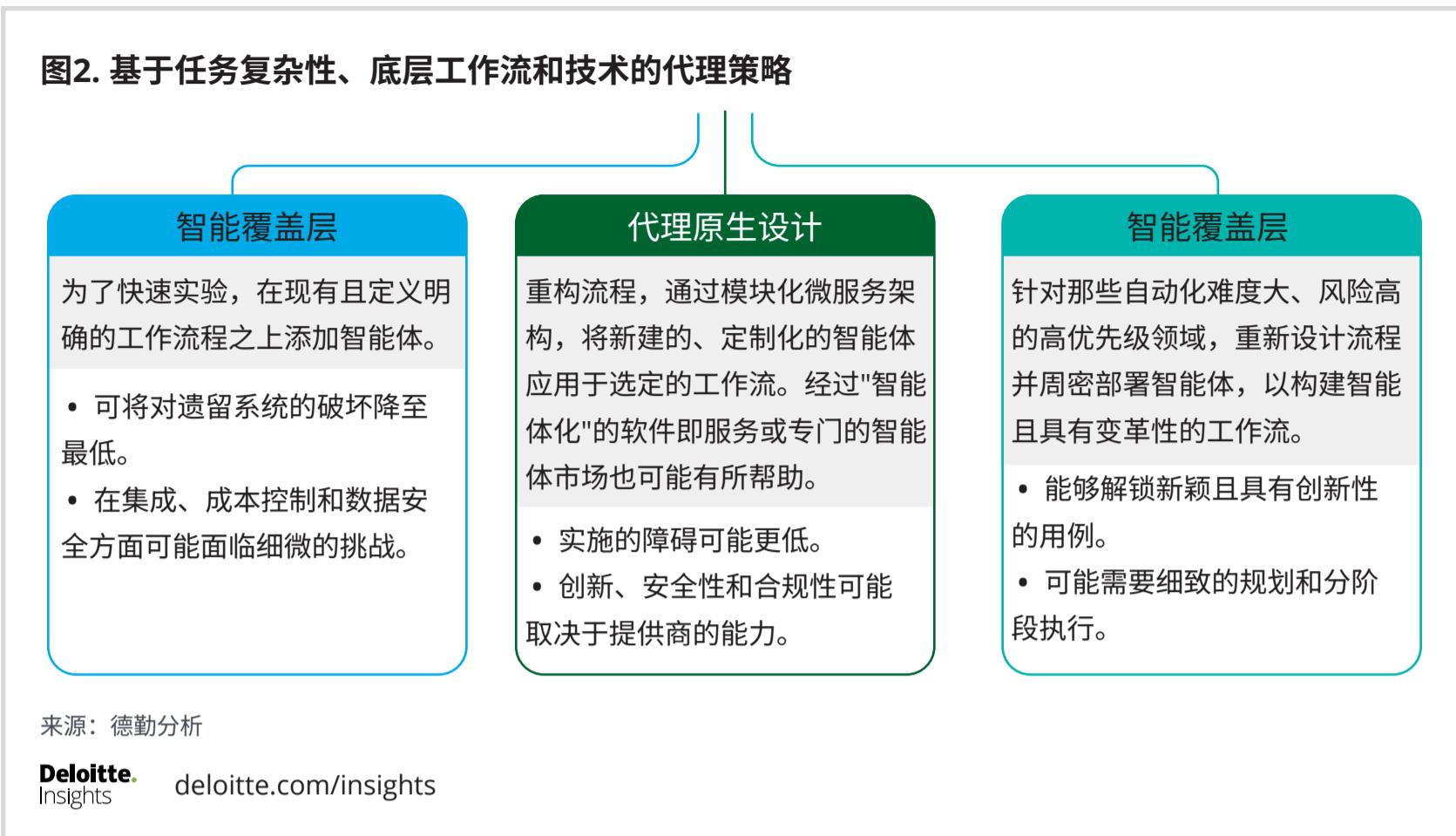
从单一用途代理到多代理系统：企业准备好了吗？

如今，企业可利用单一功能智能体自主执行多步骤任务。⁴ 他们逐渐发现，代理式AI的优势同样适用于多代理系统，能够释放更广泛且呈指数级增长的企业价值。⁵ 然而，对于许多企业而言，技术实施可能还远未成熟。

德勤针对近550名美国跨行业领导者开展的《2025技术价值》调查显示，80%的受访者认为其组织在基础自动化方面已具备成熟能力，而仅有28%的受访者认为其在基础自动化与智能体相关工作方面具备同等能力。此外，在推进各项战略的受访者中，45%的人预计其基础自动化工作可在三年内实现预期的投资回报，而在类似时间内，仅有12%的人预计基础自动化与代理结合的工作能实现相同回报。⁶

他们如何能更快地到达那里？第一步是考虑三种可能的多代理方法（图2）。⁷

图2. 基于任务复杂性、底层工作流和技术的代理策略



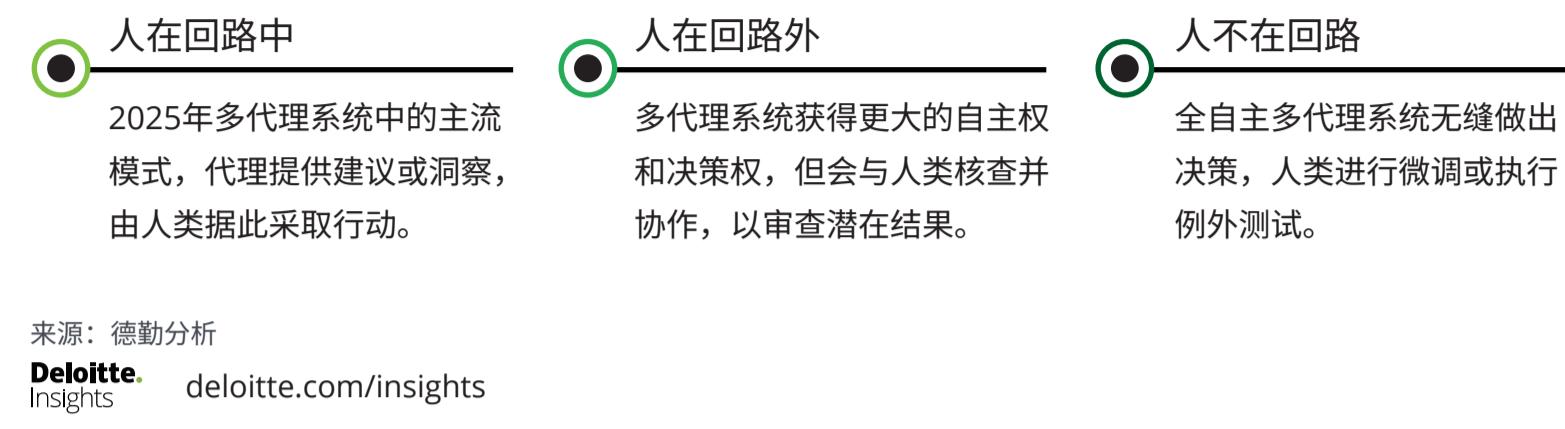
代理协同中的人为因素

2025年，企业在特定领域（如金融投资研究和重大疾病医疗）已开始实施相对简单但前景广阔智能代理协同方案。⁸在这些应用中，代理通常在人类监督或专门的“监督代理（Supervisor agent）”指导下协同工作，为专业人员提供决策依据。而跨越多个业务领域的更复杂、更自主的代理协同，目前主要局限于少数行业领军企业。⁹随着相关实践深入，企业将日益面临平衡代理自主性与人工监督的挑战——需在创新与风险、责任归属及信任建立之间审慎权衡。

研究表明，当前新兴的多代理系统在有人类参与的情况下表现更佳——因为它们能够受益于人类的经验，并与企业复杂的期望保持一致。¹⁰我们预测，在未来12到18个月内，将有更多企业加快对复杂代理协同的试验和规模化应用，同时保持人类的参与。这些企业可能会采用各类框架和解决方案，将人类判断融入代理工作流程中，以提升决策的可信度、质量和可问责制。¹¹

此外，根据任务复杂度、业务领域、工作流程设计及结果关键性等因素，将逐步形成一个“自主性层级谱系”（autonomy spectrum）——包括人在回路中（实时参与决策）、人在回路上（监控并适时介入）和人在回路外（完全自主运行）三种模式（图3）。尽管“人在回路外”模式仍需持续监控，但“人在回路中”和“人在回路上”模式将更多依赖平台与智能体遥测仪表盘，通过结果追溯、协同可视化及其他细节信息，为人类干预提供指导。我们预测，到2026年，最先进的企业将开始向“人在回路上”的协同模式转型奠定基础。

图3. 智能体自主性的渐进式演进



驯服碎片化蔓延的智能体

到2026年，智能体可能会在不同编程语言、框架、基础设施和通信协议中呈现激增态势。更复杂的是，部分代理可能需要具备多模态能力（即解读文本、音频、图像等不同类型和格式信息的能力），才能实现最高水平的智能。此外，针对代理的网络协议开发（例如麻省理工学院的NANDA项目）可以定义代理如何在企业外部的数字界面上进行协调。¹²从长远来看，它能够实现企业内部及外部网络间的战略性代理协同，释放出全新的能力。

这些变量将使多代理的互操作性变得至关重要，同时也更具挑战性。此外，企业将越来越倾向于寻求通过统一平台来指挥、监控和管理不同智能体的方法。缺乏数字劳动力的运营标准可能导致智能体的构建、配置和部署呈现去中心化和缺乏协调的状态，进而可能增加性能下降以及伦理、网络安全和合规性问题带来的潜在风险与成本。

企业可以从影响当今IT和业务架构的过往技术（如云计算和微服务）中汲取灵感。标准化协议（例如https、JSON等）、清晰的API蓝图以及面向特定领域的微服务，实现系统间的互操作性、稳定性和权属划分。服务注册表、分布式追踪和集中式日志则提升了能力发现、错误排查和服务管理的效率。治理机制、服务目录以及“零信任”安全模型确保了系统的健壮性，并避免了版本混乱的问题。这些措施均可为构建弹性、可扩展的多代理系统提供有益借鉴。然而，企业也应采取创新方法，专注于在其企业架构中构建独特的层次结构。

图4：具有弹性且可扩展的多代理系统的企业架构



来源：德勤分析

Deloitte. Insights deloitte.com/insights

- 1. 情境层**—这一强大的知识工程基础对于可扩展的智能体架构至关重要。它将原始且多样化的数据转化为结构化且管理规范的知识表示形式（例如知识图谱、本体论、领域分类体系等），从而为代理提供问题领域的“小世界”模型。经过优化的上下文检索技术可使代理精准、及时地获取相关信息，而上下文塑造则可优化输入，减少噪声和冲突，从而提升代理的准确性与效率。
- 2. 代理层**—该组件利用底层情境层来支持代理操作，重点关注安全性、自主性和互操作性。该层的核心是一种模块化且可组合的架构，能够集成并适应新技术。通过强调工具相关性和抽象化的策略，有助于防止代理过载。此外，精心设计的记忆策略可优化对事实性、经验性和程序性记忆的综合访问，从而增强情境感知能力。该层还选择适当的AI模型（从紧凑型专用模型到大规模高性能模型）以优化代理在各类协同任务中的表现。通过先进的遥测技术实现强大的安全措施和全面的可观测性，确保代理活动的安全性、透明性和可靠性。
- 3. 体验层**—作为企业用户与代理之间的主要交互界面，有助于控制并及时纠正代理的行为。它向用户提供代理状态和上下文数据等相关信息，支持生成提示建议，并以易于审阅的格式呈现清晰易懂的结果。直观的人工监督控制、高级反馈功能以及可解释性特性（如展示代理的推理过程），有助于提升结果的透明度与可信度。此外，当出现错误或模糊情况时，该层还能提供清晰的解释和恢复选项。

让多代理系统为企业发挥作用

随着企业掌握技术基础，这三个指导原则有助于更好地与业务关键需求保持一致：

灵活、可扩展且安全的通信协议

多代理协同需要代理之间以及代理与其他工具或平台之间具备标准化的通信方式。这对于准确传递代理的能力、洞察和行为信息至关重要。在过去一年中，出现了多种代理间通信协议，每种协议都旨在实现基于不同框架或模型构建的代理之间的协调。这些协议包括谷歌的A2A、思科主导的AGNTCY、Anthropic的MCP等。¹³ 科技供应商正在集结其合作伙伴、联盟和客户，以在该领域取得主导地位。此外，其中一些协议正在被扩展，用于在金融交易等特定领域实现可信代理间的互操作性。¹⁴

各协议之间的过度竞争可能导致形成“封闭花园”，使企业被锁定在单一的通信协议和代理生态系统中。¹⁵ 然而，到明年这些协议很可能会开始趋于统一，形成两到三个主导标准，其他技术供应商将需要与之保持一致，以维持竞争力。

哪些选定的协议能够脱颖而出，可能取决于多种参数，以及企业如何根据其多代理应用、行业特点和协同成熟度对这些参数进行优先排序。例如，具备标准API及用于测试和仿真的开发工具的轻量级协议，有助于降低实验门槛。支持点对点和中心辐射式代理交互、具备共享上下文与记忆能力，并内置协商、委托和冲突解决机制的协议，可实现多样化的协同方式。代理注册中心可用于可信发现和负载均衡，异步消息传递、高吞吐量、低延迟以及对链式和嵌套工作流的支持，有助于扩展代理协同规模。此外，身份认证、安全消息传输和访问控制有助于降低安全风险，而代理间的通信记录和解释机制则可确保审计能力和错误可追溯性。

管理平台和可观测性工具

随着多代理系统的扩展，企业将越来越需要管理代理并理解其做出的决策。企业可以利用现有统一且可扩展的平台，借助监督功能或“监督代理”来解释请求、分配任务、授权和管理访问权限，以及执行并行或多步骤流程。¹⁶ 在未来一年里，科技公司很可能会在此领域推出新的功能，企业将需要决定如何搭建此类协同平台。例如，集中式的内部平台可以降低对供应商的依赖，并增强对数据和代理的控制能力。然而，现成的平台有助于加快测试进程，并有效控制创新成本。

无论企业如何选择，代理协同平台都将在跟踪运营指标、提升性能和管理成本方面发挥重要作用。目前，一些平台正在开发整合代理遥测监控的方法，例如延迟、错误率、令牌使用情况以及其他工具洞察信息。¹⁷ 护栏评估及检测异常行为的能力有助于降低风险。随着时间的推移，此类平台可能会引入创新功能，例如分层的商业洞察和额外的控制机制。例如，一种新兴的“守护代理（Guardian agent）”类别，既能自主执行任务，又能监管其他代理，以感知和管理潜在风险行为。¹⁸

代理协调平台还需要纳入监管合规性，该领域目前国际上的相关工作正在推进。欧盟《人工智能法案》对风险评估、透明度措施、技术保障措施以及人工监督等方面提出了具体要求。¹⁹ 此外，欧盟的标准机构正在根据《欧盟人工智能法案》制定统一的法律标准。²⁰

业务流程和员工变动

Gartner® 预测，到2028年，“33%的企业软件应用将包含代理式AI，而2024年这一比例尚不足1%，其中至少15%的日常工作任务决策将由智能体自主完成。”²¹为了实现这一目标，到2026年，更多的企业可能会开始重新构想其工作流程，定义具体且独特的模块。这将有助于根据关键性、依赖关系、任务可预测性以及目标韧性，确定所需代理协调的类型。例如，某些模块可能受益于代理按顺序工作——前一个代理的输出作为下一个代理的输入，而其他模块则可能利用代理并行或协作运行。

另一个主要考虑因素是人类将如何与多代理系统协作。一项针对200名人力资源主管的全球调查显示，86%的首席人力资源官（CHRO）认为，整合数字劳动力（即执行智能工作的技术）是其职责的核心部分。²²早期模型显示，人类充当“代理主管”（Agent boss）或与代理并肩工作。²³到2026年，企业可能会在更多岗位、职能和任务中进一步探索这些协作模式，以明确代理协调在哪些方面能够提升效率，以及在哪些方面人类的优势和协作能够带来更具意义的价值。²⁴

到明年，企业可能还将开始重新构想现有岗位如何借助多代理系统释放更高价值的成果。²⁵例如，人类的贡献可能包括在解决问题和高效制定战略决策时，进行更具创造性的提示，并引导多代理系统。与此同时，企业很可能还将专注于定义人类在代理训练、协调、监督和治理方面的的新技能与职责。²⁶量身定制的培训计划以及培养能够管理人类员工和数字员工的领导者将至关重要，这有助于在利用人类独特技能的同时，将更高质量、问责制和韧性融入多代理决策之中。²⁷

小结：2026年可能是代理协同的关键转折点

代理协同或将引领智能企业迈入下一个发展阶段。明年，我们预计企业将开始规模化部署多代理系统，为其信息技术和商业环境增添更多复杂性。代理通信协议预计将逐步集中于那些能够提供易于实验、灵活性、可扩展性和安全性的方案。企业工作流程将可能变得更加模块化，由内部开发或通过SaaS及其他第三方供应商获取的代理驱动。面向人类员工的新职位和调整后的岗位将开始涌现，以促进与多代理系统的高效协作。

然而，企业和技术供应商应果断采取行动，以主导这一发展进程。

企业采用多代理系统时的考虑因素

- **明确所有权与责任归属。**企业应明确由哪位高管负责主导公司智能体的愿景、战略及实施，并配套相应的激励机制与问责制度。该职责最自然地可归属于负责战略性技术举措和推动创新的领导者，但设立一个整合型职能团队，将有助于展现更全面的影响和风险管理能力。
- **着眼于演进而设计，而不仅仅是部署。**代理和协同能力正在迅速发展。模块化的“即插即用”协同框架可帮助企业提升灵活性、成本效益和创新能力，同时最大限度地减少对系统架构的干扰。
- **严格进行压力测试协同运作。**在规模化部署前，企业应模拟真实商业场景中的复杂状况（如数据不全、目标冲突或对抗情境）对代理协调系统进行测试。通过模拟环境，可在企业全面部署前发现隐藏的故障点，并强化安全保障机制。

- **认真对待治理与衡量。**智能体治理对于确保大规模的安全性、合规性和可靠性协调至关重要。制定明确的智能体角色规则，界定其责任归属，设计应对错误的回退机制，并实施监督，有助于防止滥用、确保可审计性并建立信任。除了技术准备之外，企业还应识别并跟踪将代理协调与价值创造相连接的衡量指标，例如更快的决策、更优的客户体验或更迅速的创新。

技术供应商的注意事项

- **以互操作性为基础进行构建。**技术供应商除了应遵守代理间通信标准外，还应设计出模块化的解决方案，使各代理能够理解彼此的意图和行为背景，从而实现无缝协作。
- **重新思考信任。**仅仅提供洞察是不够的；必须具备理解或验证智能体输出的能力，这对于建立信任和推动应用至关重要。诸如为代理提供数字身份等新型安全措施，对于构建和运行可信的多代理系统也同样重要。
- **使治理成为固有属性。**了解业务长期所需，并与人类价值观和组织政策保持一致，可能是提供相关治理框架的关键。未来的解决方案应具备创新的代理监控、高级治理机制以及伦理保障措施，以确保合规性和有效性。
- **拓展生态系统。**技术供应商应持续建立并加强全行业的联盟，以在通信协议、信任机制和治理方面实现必要的标准。创新的跨平台协同工具正获得越来越多的关注，这为新兴和成熟的技术企业通过并购、合作与协作来巩固其市场地位带来了机遇。²⁸

中国AI Agent发展前瞻

中国市场规模不断增长

中国AI Agent市场增长迅速。预计到2028年，市场规模将达到8520亿元，2023-2028年间CAGR预计为72.7%，成长空间极为广阔。AI技术的飞速进步是根本驱动力。大语言模型（LLM）能力的突破使得AI Agent能够理解和执行更为复杂的任务，从简单的问答交互升级为能够处理多步骤、跨应用的业务流程，从而释放出巨大的自动化潜力。

中国AI 关键技术取得进展

中国人工智能技术正加速与产业融合，形成从技术突破到生态繁荣的完整链条。在工业领域，面向装备、电子、汽车、家电等行业的垂域大模型，破解实际业务痛点，推动解决方案快速落地；同时，支持70个语种、超过4000个方向互译的多语言语音翻译引擎，显著提升了小语种处理能力，突破国际交流的技术瓶颈。在开发工具层面，工程化、标准化、低代码的智能体开发平台不断涌现，极大降低了构建和部署单智能体乃至多智能体协同应用的门槛。

AI Agent需破除商业落地障碍

目前中国的AI Agent在商业落地方面存在一定障碍，一方面，AI Agent的收益量化困难，企业难以准确量化AI Agent所带来的收益，另一方面AI存在变革阻力，商业化落地存在跨部门协作障碍。针对此问题，企业需进行AI基础场景测试与全链路价值评估。同时针对跨部门的协作，要建立跨职能Agent团队，同时制定AI协作的完整SOP。

智能体安全风险不断引起重视

随AI Agent的应用不断深入，信任危机与合规风险不断引起人们的重视，未来AI Agent行业需强化用户数据安全意识，技术端加强安全保护，降低数据泄露风险，同时加大对行业数据使用的规范性治理。

Sayantani Mazumder

India

Girija Krishnamurthy

Global

Diana Kearns-Manolatos

United States

China Widener

United States

Baris Sarer

United States

Roger Chung

China

Gillian Crossan

Global

ENDNOTES

1. Deloitte, "[The cognitive leap: How to reimagine work with AI agents](#)," December 2024.
2. The baseline projection is derived from a Deloitte analysis of global autonomous AI agent market projections as per seven publicly available and third-party research reports. The estimated increase of 15% to 30% in the projected market is modeled on future scenarios where fewer agentic AI projects are cancelled owing to improved enterprise readiness.
3. Gartner, "[Gartner predicts over 40% of agentic AI projects will be canceled by end of 2027](#)," press release, June 25, 2025.
4. Bojan Ceric and Prakul Sharma, "[Generative AI meets the virtual world: A model for human-AI collaboration](#)," Deloitte Insights, Feb. 10, 2025.
5. Abdi Goodarzi and Nitin Mittal, "[A new digitally-enabled workforce era: How AI agents can help deliver functional efficiency and value across the enterprise](#)," Forbes, Aug. 18, 2025.
6. Tim Smith, Gregory Dost, Garima Dhasmana, Parth Patwari, Diana Kearns-Manolatos, and Iram Parveen, "[Digital budgets are rising, but investment strategies may need a recalibration](#)," Deloitte Insights, Oct. 16, 2025. The survey asked respondents about four types of AI automation and their incremental actions across each: mature or very mature respondents for basic automation (n = 443) and basic automation and AI agents (n = 153); and those with up to three-year expectations for basic AI automation (n = 245) and basic automation and AI agents (n= 68).
7. Prakul Sharma, Val Srinivas, and Abhinav Chauhan, "[How banks can supercharge intelligent automation with agentic AI](#)," Deloitte Insights, Aug. 14, 2025; Kausik Chaudhuri, "[Applying agentic AI to legacy systems? Prepare for these 4 challenges](#)," CIO, July 16, 2025; [SaaS meets AI agents: Transforming budgets, customer experience, and workforce dynamics](#); Bojan Ceric and Prakul Sharma, "[Scaling AI agents may be risky without an enterprise marketplace](#)," Deloitte Insights, Sept. 15, 2025.
8. Julian Horsey, "[AI investment research agent "Ask David" built by JP Morgan](#)," Geeky Gadgets, May 30, 2025; Irene Iglesias Álvarez, "[The agentic AI assist Stanford University cancer care staff needed](#)," CIO, May 30, 2025.
9. Isabelle Bousquette, "[Why Walmart is overhauling its approach to AI agents](#)," The Wall Street Journal, July 24, 2025.
10. Henry Peng Zou et. al, "[A call for collaborative intelligence: Why human-agent systems should precede AI autonomy](#)," arxiv, June 11, 2025.
11. Jesus Olivera, "[Ensuring accuracy in AI with human-in-the-loop](#)," Medium, Sept. 27, 2024.
12. John Werner, "[They're making TCP/IP for AI, and it's called NANDA](#)," Forbes, May 01, 2025
13. Emilia David, "[Google's Agent2Agent interoperability protocol aims to standardize agentic communication](#)," VentureBeat, April 9, 2025.
14. Emilia David, "[Google's new agent Payments Protocol \(AP2\) allows AI agents to complete purchases — is your enterprise ready?](#)," VentureBeat, Sept. 16, 2025.
15. Leslie Joseph and Rowan Curran, "[Interoperability is key to unlocking agentic AI's future](#)," Forrester, March 25, 2025.
16. Alfred Shen and Anya Derbakova, "[Design multi-agent orchestration with reasoning using Amazon Bedrock and open source frameworks](#)," Amazon Web Services, Dec. 19, 2024; IBM, "[Multiagent orchestration](#)," accessed Oct. 7, 2025.
17. Amazon Web Services, "[Observe your agent applications on Amazon Bedrock AgentCore Observability](#)," accessed Oct. 13, 2025.

18. Gartner, "Gartner predicts that guardian agents will capture 10-15% of the agentic AI market by 2030," press release, June 11, 2025.
 19. The Future Society, "How AI agents are governed under the EU AI Act," June 4, 2025.
 20. CEN-CENELEC, "Artificial intelligence," accessed Oct. 7, 2025.
 21. Daniel Sun, "Capitalize on the AI agent opportunity," Gartner, Feb. 27, 2025. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.
 22. Salesforce, "HR leaders to redeploy a quarter of their workforce as agentic AI adoption expected to grow 327% by 2027," May 5, 2025.
 23. Ibid; Atikah Amalia, "The marketer's new job title: AI boss," Content Grip, April 29, 2025.
 24. Kyle Forrest, Brad Kreit, Abha Kulkarni, Roxana Corduneanu, and Sue Cantrell, "AI, demographic shifts, and agility: Preparing for the next workforce evolution," Deloitte Insights, Aug. 25, 2025.
 25. Michael Caplan et al., "The technology operating model of the future: Rise of the agentic enterprise," The Wall Street Journal, Aug. 23, 2025.
 26. Ritu Jyoti, "The rise of the agentic economy: How autonomous AI is reshaping the future of work," CIO, Sept. 8, 2025.
 27. Isabelle Bousquette, "Digital workers have arrived in banking," The Wall Street Journal, June 30, 2025.
 28. Marina Temkin, "Why AI agent startup /dev/agents commanded a massive \$56M seed round at a \$500M valuation," TechCrunch, Nov. 28, 2024; Hui Wong, "Questflow secures \$6.5M seed round to build AI agent economy for every workflow," Marketers Media, July 24, 2025.
-

ACKNOWLEDGMENTS

The authors would like to thank **Prakul Sharma, Rajib Deb, Mark Szarka, David Jarvis, Abhinav Chauhan, Michael Steinhart, Ankit Dhameja, and Iram Parveen** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

AI革命正引领机器人和无人机领域的创新浪潮

更强大的AI模型和芯片能否为这个相对停滞的行业带来变革契机？

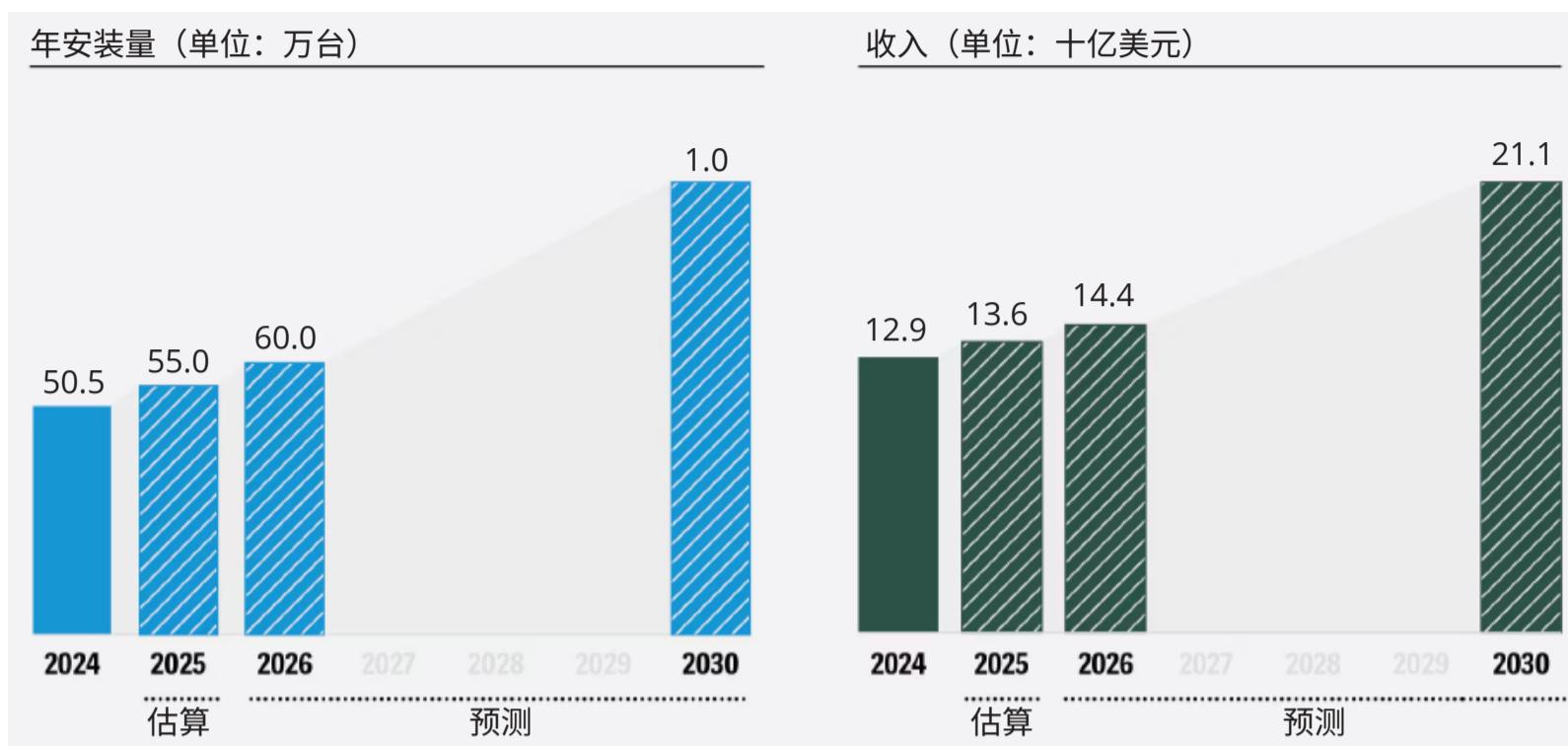
到2030年或2040年，工厂车间遍布能够像人类智能一样感知和行动的人形机器人，这一愿景令人向往，甚至可能实现，但2026年的现实情况却有所不同。德勤预测，到2025年，全球工业机器人的累计安装量将超过500万台，到2026年可能达到550万台。¹

随着AI能力在机器人系统中的进一步融合以及专用基础模型的出现，机器人可渗透到从智能工厂、公共服务到自主无人机等多个行业和应用领域。然而，如果整个技术、AI和机器人生态系统不能解决与数据质量、系统集成、网络安全以及监管相关的问题，工业机器人市场可能仍将维持目前相对温和的年增长率水平。

先进及专用AI模型：工业机器人的催化剂

自2021年以来，全球工业机器人年销量始终徘徊在50万台左右，这与德勤在《2020科技、传媒和电信行业预测》中关于工业机器人增长放缓的判断相符。²长期预测表明，该领域将在近期迎来爆发式增长，有预估指出人形机器人产业规模到2050年有望达5万亿美元。³但更值得关注的是，发展拐点可能早在2030年就会出现——届时年度新型机器人出货量将翻倍至100万台，预计实现210亿美元营收，较2024年水平增长近一倍（图1）。⁴

图1：到2030年，工业机器人的年出货量（其中许多由AI驱动）可能达到100万台，每年产生超过200亿美元的收入



注：E = 估计值；P = 预测值。

来源：德勤根据包括以下来源公开信息进行的分析国际机器人联合会、Interact Analysis 和 IPF Online。

Deloitte. deloitte.com/insights

本文涵盖哪些“机器人”？

值得注意的是，“机器人”是一个广泛的术语，涵盖的范围从洗碗机，到价格几百美元、更智能且具有自主能力的家用吸尘器，再到每台价值数百万美元的生产线工业机器人。此外，这一定义有时还包括飞行机器人（即无人机）、驾驶机器人（即完全自动驾驶汽车），以及能够完成人类几乎一切任务甚至更多功能的人形机器人。

在本预测章节中，重点主要是工业机器人、用于工业用途的人形机器人以及无人机。目前，物理AI、机器人和无人机领域呈现出上升趋势，而关于自动驾驶汽车的文章和分析师报道已经非常丰富。因此，本章将聚焦于工业机器人和无人机领域。

两个增长催化剂可能在2026年至2030年间推动工业机器人普及率出现拐点。首先，发达国家由于人口老龄化而面临持续的劳动力短缺问题。⁵随着这些地区日益加强国内制造业并且构建更具韧性的供应链，对能够执行日益复杂任务的机器人需求可能只增不减。其次，或许更为重要的是，计算能力的指数级进步以及专用基础AI模型（不同于典型的大语言模型（LLMs））的出现，正在加速AI机器人和具身智能系统的发展。⁶专用模型可能正在为高度复杂的AI引擎铺平道路，使机器人能够超越简单的命令与控制，像人类一样理解自然语言、感知物理环境，并以通用的方式学习和完成复杂任务。⁷

然而，尽管人们充满热情且先进技术不断涌现，机器人技术的发展仍面临某些障碍。例如，将机器人系统集成到现有的工业工作流程中十分复杂，尤其是在数据质量、互操作性以及与遗留系统的兼容性方面。许多企业在获取干净、统一的数据集方面存在困难，而这些数据集对于有效的AI训练和推理至关重要。⁸此外，联网机器人网络面临的安全与隐私泄露风险，或遭受恶意网络攻击的可能性，仍然是一个关键问题。⁹此外，保障人类工作人员的安全是工业机器人和人形机器人需要解决的重要方面。¹⁰

德勤认为，将生成式AI（Gen AI）和代理式AI（Gentic AI）更紧密地与机器人及自动化工具相结合，将有助于推动AI驱动的机器人设备走出科幻领域，进入现代化的工作场所。¹¹作为一个例证，德勤位于威奇托的智能工厂正在帮助各类组织探索先进的制造与运营技术，以获得切实可观的商业效益。智能工厂配备了多种技术能力，包括生成式AI、代理式AI、无限现实，以及无人机、自主移动机器人（AMR）、四足机器人和人形机器人等机器人技术，提供一条完全可操作的生产线，助力企业解决实际业务问题。¹²

工业机器人似乎已经在制造业、医疗保健、仓储甚至国防等多个行业创造了价值（图2）。¹³但可能正在为工业机器人创造新机遇的，是一些科技公司所展示的技术创新，尤其是多模态AI模型以及先进芯片和硬件的出现。

图2：由AI驱动的工业机器人正在为多个产业环境创造价值

行业	AI驱动机器人的优势与应用场景
制造业	包括配备仿生手的认知型人形机器人，可通过合成数据进行自主学习，并在高端工厂中以协调的方式协同工作，同时具备3D物体识别能力，可协助完成机器上下料、注塑成型及设备维护等任务。
医疗保健	机器人可以协助护士跑腿（如取样本和递送药品），帮助外科医生执行需要高精度的精细手术，支持工作人员处理危险材料（如病毒样本），并执行消毒房间等高风险工作。
仓库与物流	由AI驱动的机器人可以利用深度学习视觉技术识别并处理各种形状和尺寸的物品。自主移动机器人则通过AI算法以及摄像头和3D视觉生成的先进传感器数据，实时绘制并导航仓库环境，能够精确定位自身位置，而无需依赖任何固定基础设施或标记物。
国防与军事	机器狗可通过先进的传感器和基于AI的分析技术，发现并处理炸弹，对威胁性物体进行分类和识别。AI机器人可执行监视和侦察任务，运输物资，并协助伤员撤离。

来源：根据与行业领域专家的交流所得的见解，以及多个公开可获取的资料，包括：American Machinist、Admedica、世界经济论坛、PHS Innovate 和 ASDNews。

视觉-语言-动作模型有望使人形机器人变得更智能、更自主

一些AI初创公司和大型科技企业正在开发视觉-语言-行动（VLA）模型，有望使机器人从执行预编程任务向理解上下文并自主决策迈进。VLA技术使机器人获得更高的自主性，能够实现更高级别的规划和空间推理，并具备灵活应对复杂地形的能力。¹⁴通过在模拟环境中进行大规模强化学习和多模态学习，机器人可以在海量数据集上进行预训练。

VLA 将视觉感知（观察环境和物理规律）、自然语言理解（口头指令和理解）以及实际操作能力（响应视觉和文本指令）相结合，以执行任务。¹⁵一般，截至2025年年中，视觉语言模型（VLA）的参数规模在5亿到70亿之间，使人形机器人能够学习、感知和行动。¹⁶目前在美国，已有一些选择性案例将VLA模型用于增强机器人技术的开发，预计在2026年至2030年间有望实现更广泛的商业应用：

- 英伟达（NVIDIA）的开放人形机器人的基础模型结合了推理与行动能力，有助于推动机器人技术的发展。¹⁷像波士顿动力公司这样的机器人公司正在利用英伟达的模型库及支持技术来开发人形机器人。¹⁸
- Figure AI 的 Helix 是一种VLA模型，通过视觉和自然语言提示来训练机器人，使人形机器人能够深入理解真实世界的场景和物体，并发展出精细的运动控制能力。¹⁹
- Hugging Face 开发了专门用于机器人的开源数据和模型，同时仍在持续研发和测试其自身的开源人形机器人，²⁰且允许开发者自定义机器人。²¹

在美国以外，亚洲和欧洲也在积极开发人形机器人，重点在于定制基础模型以及基于物理世界数据的训练。例如，总部位于韩国的初创公司RLWRLD正在开发基础AI模型，使机器人能够通过自动化学习和模仿人类专业知识，自主执行传统的劳动密集型流程。²²日本FANUC公司致力于开发一系列适用于工业环境、具备AI功能且尺寸各异的机器人。²³欧洲的Neural Foundry（伦敦）和NEURA Robotics（德国）正通过整合认知能力并开发定制模型，正着力打造适用于工业环境的AI机器人。

在中国，AgiBot和MagicLab等初创企业正在设计能够执行制造环境中复杂任务的人形机器人。²⁴像宇树科技和优必选科技这样的企业正在推进大规模生产，使人形机器人变得更加普及和实惠，从而推动更广泛的应用。²⁵制造人形机器人（图3）需要多种芯片组件和硬件，这表明该市场为半导体（包括芯片硬件及相关软件和服务）带来了巨大的收入潜力。²⁶

尽管用于大规模工业应用的AI人形机器人仍处于早期阶段，但德勤预计，2025年年度出货量将在5,000至7,000台之间，到2026年可能增至15,000台。²⁷每台平均价格为14,000美元至18,000美元，²⁸用于工业领域的AI人形机器人市场在2026年可能价值约2.1亿至2.7亿美元。²⁹随着机器人行业在2026年至2030年间逐步克服技术、价格和操作方面的障碍，人形机器人市场到2032年可能达到6亿至7亿美元（约为2026年基准情景的3倍），甚至在乐观情景下达到10亿美元（为2026年水平的4倍）。³⁰

图3：能够实现人形机器人和嵌入式人工智能的芯片与硬件



注：视觉语言动作（VLA）；图形处理单元（GPU）；中央处理单元（CPU）。

来源：德勤分析

Deloitte. Insights deloitte.com/insights

先进AI技术可能使无人机更加自主和功能多样化

目前大多数无人机（又称 UAV 或无人驾驶航空器）仍由人工操作，但其自主能力似乎正在迅速发展。如今，许多无人机已采用AI实现实时导航，彼此之间进行通信，规避障碍物和碰撞，不久后甚至可能在无需人工干预的情况下执行任务。例如，匈牙利的科学家研究了包括鸽子和野马在内的多种动物的行动模式和移动规律，并利用这些研究成果开发出一种算法，用于引导一群能够自主做出机载决策的无人机。这些无人机不仅能在空中安全地导航、避免碰撞并悬停，还能在不同环境中执行任务，如土地测绘、气象观测以及野火管理等。³¹

无人机：内部有什么？

新一代先进的AI无人机通常配备多种技术和芯片，包括作为飞行控制器的单核或双核微控制器；机载电源系统，可能包含锂基电池和电源分配板，为各个组件供电；射频模块，用于实现无人机与地面控制单元之间的通信；用于导航和定位的GPS模块；传感器（如加速度计、陀螺仪、磁力计、光流传感器以及激光雷达和超声波传感器）；以及用于管理空中操作的机载飞行控制软件和平台。³²

正如德勤在2024年关于农业科技的预测中指出的那样，安装在无人机或飞行器上的光谱传感器、芯片和摄像头可收集大量数据（如土壤湿度、植物健康状况等），AI模型可对这些数据进行分析，从而为精准喷洒作业提供决策支持。³³除了农业，无人机还可用于检查风力涡轮机和电力线路，从而减少人工巡检的需求。³⁴中国、南澳大利亚和英国正在试验能够完全自主执行远距离远程高压输电线路巡检的无人机。这些无人机不仅能承担危险且关键的任务，协助人类工作人员和工程师，还可自动拍摄并传输数十张图像，帮助工程师利用AI和高级分析技术检测和分析腐蚀情况。³⁵

多国正致力于部署自主无人机进行空中监测，以协助灾害救援（例如2022年9月佛罗里达西南部遭遇伊恩飓风后，自主无人机绘制受灾区域图像，为应急人员提供支援），同时用于侦测及反制潜在边境威胁。³⁶在这些应用场景中，无人机多数任务阶段仅需人工远程操控：AI系统如同商用客机的自动驾驶仪，在将无人机引导至目标区域后，便将控制权转交人类操作员。然而多国近期在无人机蜂群军事应用领域的投入表明，³⁷该技术同样可能推动非军事（工业及民用）领域发展：例如，自主无人机群可对偏远艰险地区的高压输电线进行巡检，甚至能在恶劣气象条件下监测海上风力发电机组。

小结

尽管近年来工业增长相对温和，工业机器人已成为半导体公司的一个重要终端市场。例如，一台价值约20万美元的工业机器人，可能包含价值约2.5万至5万美元的芯片及相关电子元件。³⁸此外，提升工业机器人的性能可能将依赖于日益先进的芯片技术，涵盖处理器、网络和传感器等多个方面，每台机器人所含的半导体内容量也可能会增加。同时，截至2025年，半导体产业本身也是工业机器人的主要消费者和终端用户，广泛将其应用于晶圆厂制造流程、晶圆搬运、测试与分拣、先进封装以及洁净室等各个环节。³⁹在迈向“熄灯制造”的进程中，芯片行业可能会在其运营中使用更多的工业机器人。

随着包括人形机器人和无人机在内的工业AI机器人市场前景看好，许多半导体和技术公司正在积极长期投资这一领域。机器人初创企业已在仓库、物流和空中自主等实际场景中进入试点阶段。机器人领域的风险投资正在增长，预计这可能成为2025年唯一一个融资额将增加的非AI市场类别。⁴⁰云计算和IT基础设施也正在逐步完善，同时合成数据生成和物理仿真技术可能正在加速开发进程，并降低对高成本现实世界测试的依赖。

为应对工业机器人商业化应用过程中的潜在挑战，并解决数据整合、隐私安全、防护保障及人才储备等相关问题，AI、机器人及科技行业领袖可考虑采取以下五项行动措施：

1. 以开放创新证明商业价值：科技与AI企业应通过推广开放式的全栈机器人生态系统，实现机器人的大规模部署与协同运作，依托更广泛的商业化应用证明投资回报率；同时构建协作型通用生态系统，推动通用具身AI发展。⁴¹
2. 提升数据质量并解决整合难题：行业参与者应优先推进数据标准化，协作开发通用平台与中间件，使各类机器人能更无缝地融入工业环境。
3. 修复网络安全漏洞：企业应采用通用互操作性协议，践行隐私保护与安全设计理念，并主动联合网络安全专家构建灵活清晰的安全框架。由AI工具生成的机器人代码需经过人工严格审查，排查潜在网络安全漏洞。
4. 将安全作为核心基础功能：从研发、测试到原型设计阶段，应通过编程确保机器人的本质安全——无论是实现人机协作时的零伤害保障，还是避免机器人间意外碰撞。基于模拟的训练、计算机辅助安全规划工具及主动防碰撞运动规划等创新技术与方法，可有效提升机器人安全性能。
5. 前瞻性赋能现有人才队伍：针对新兴AI技术开展员工技能再培训与提升，对每家企业都至关重要。在新一轮工业AI自动化浪潮中，随着机器人与人类协同作业日益深化，企业需定期评估并提升员工的AI技能水平，以保持在工业机器人应用与企业整体融合方面的领先优势。

发展路径已十分明晰：AI、机器人及科技产业应当率先迈出步伐——当前既具备必要的先进AI技术条件，也存在商业层面的迫切需求与浓厚兴趣。唯有通过360度系统思维和基于生态圈的整体推进策略，才能在上述五大关键领域（开放创新、数据、网络安全、安全保障及人才培育）取得实质性进展，从而加速工业机器人在2026年及以后的商业化应用进程。

中国机器人、无人机发展前瞻

中国已成为世界最大机器人市场

中国在工业机器人领域具有全球领先地位。根据国际机器人联合会（IFR），2024年中国再次成为全球工业机器人最大市场，全球超54%的新增安装量来自中国。预计到2030年，中国具身智能产业市场规模有望达到4000亿元。

人形机器人与AI结合成为未来主力方向

中国人形机器人市场发展迅猛，政策支持力度不断加大，中国工信部人形机器人标准化技术委员会的成立，是中国系统性布局未来产业、将先发技术优势转化为产业和市场优势的关键举措。头部企业AI大模型提升机器人多线程任务处理能力。AI算法与机器人硬件深度结合。

AI大幅提升智能无人机的应用空间

随AI技术的不断发展与应用，智能无人飞行器的功能与应用空间不断拓展，如在地理测绘领域在未来AI技术加持下，智能无人飞行器可以实现测绘数据的智能化、自动化处理；同时AI可以通过算法，提升无人飞行器的路径规划、精准避障的能力。在安防领域，无人机搭载AI算法实现对火点、烟雾、人员、车辆等的自动识别和预警。

AI机器人与中国产业链基础相结合，带动全产业链的协同发展

在产业供应链与技术基础层面，具身智能的本体制造与原有制造业产业存在高度协同效应，两者在供应链上的重合度预计可达70%-80%。如新能源汽车行业，其开发的智能驾驶所依赖的环境感知、多模态融合与决策规划等核心技术与具身智能具有高度同源性。因此，新能源汽车企业在布局具身智能领域时，具备显著的供应链整合与核心技术迁移的先天优势，AI机器人的发展也将带动东全产业链条企业发展。⁴²

全栈整合趋势愈发明显

未来的具身智能，真正的领先者必须是能够将先进“大脑”与高性能、低成本的“身体”进行正向设计的软硬一体全栈型企业。将“最强大脑”与“最强身体”相结合成为主流发展方向。

Karthik Ramachandran

India

Tim Gaus

United States

Girija Krishnamurthy

Global

Duncan Stewart

Canada

Gillian Crossan

Global

Roger Chung

China

Jeroen Kusters

United States

ENDNOTES

1. Deloitte analysis and estimates based on data from multiple publicly available information sourced from International Federation of Robotics IFR, Interact Analysis, and Automation.com.
2. Duncan Stewart et al, Robots on the move: [Professional service robots set for double-digit growth](#), Deloitte TMT Predictions 2020, November 2019. To read further, see Professional services robots on the move, The Wall Street Journal-CIO Journal, April 8, 2020.
3. Morgan Stanley research, [Humanoids: A \\$5 trillion market](#), May 14, 2025.
4. Methodology and assumptions: From 500,000 annual installations each year, in 2025 and 2026, we anticipate annual industrial robot installations could grow by 100,000 units every year between 2027 and 2030, reaching 1 million installed units in 2030. These calculations are based on insights gathered from IFR press release dated September 24, 2024 [Record of 4 million robots in factories worldwide](#). From our conversations with industry experts, we believe growth and availability of computing power, especially new types of AI models LLMs, but also VLAs and world models, plus the active role that some major tech and robotics companies are playing to invest and bring forth robotics chips and solutions to market, will help drive robotics adoption during 2026 to 2030 and beyond. Additionally, average unit price per industrial robot has declined by approximately 3.2% between 2018 and 2024. We expect average price to continue to decline in that range through 2030, given the broader availability of chips, sensors, and other components, including open model-based robots. Between 2025 to 2030, we have assumed average annual price per industrial robot could decline approximately 3.1 to 3.2 percent based on information gathered from IPF Online's article dated June 27, 2025 [Global industrial robot shipments down in 2024, recovery likely in 2025](#).
5. OECD, [OECD Employment Outlook 2025: Can we get through the demographic crunch?](#), July 9, 2025.
6. Deloitte analysis of the various foundational models released by technology companies and niche LLM players during 2024 and 2025.
7. Standard bots, [The most advanced robots in 2025](#), August 7, 2025.
8. Cem Dilmegani, [Data quality in AI: Challenges, importance, & best practices](#), AIMultiple research, July 9, 2025.
9. Ainsley Lawrence, [AI's impact on robots in manufacturing](#), September 11, 2024.
10. Brian Heater, [Figure AI details plan to improve humanoid robot safety in the workplace](#), TechCrunch, January 28, 2025.
11. Tammy Whitehouse, [AI robots in the workplace: Preparing for humanoid colleagues](#), Deloitte-WSJ CIO Journal, July 26, 2025.
12. [The Smart Factory by Deloitte website](#).
13. Deloitte analysis based on insights gathered from interviews and conversations with industry subject matter experts, and supplemented with information gathered from multiple publicly-available sources including: [American Machinist](#), [Admedica](#), [World Economic Forum](#), [PHS Innovate](#), and [ASDNews](#).
14. Reyk Knuhtsen, Dylan Patel, et al, [Robotics levels of autonomy](#), SemiAnalysis, July 30, 2025.
15. Sudhir Pratap Yadav, [Vision-Language-Action VLA models: LLMs for robots](#), Black Coffee Robotics, April 17, 2025; Raman Thakur, [How Vision-Language-Action models powering humanoid robots](#), Labellerr, March 5, 2025.
16. Deloitte analysis based on information gathered about multiple VLA models that are commercially available in the market.

17. Andrew Liszewsk, [NVIDIA says 'the age of generalist robotics is here'](#), The Verge, March 19, 2025.
18. Automation World, [Boston Dynamics working with NVIDIA on next-gen humanoid robots](#), May 21, 2025.
19. Brian Heater, [Figure's humanoid robot takes voice orders to help around the house](#), TechCrunch, February 20, 2025; Wei Sun, [Figure AI Unveils its 2nd-Gen Robot, Extending Focus from Factory to Home After OpenAI Split](#), Counterpoint Research, August 14, 2025.
20. Rebecca Szkutak, [Hugging Face unveils two new humanoid robots](#), TechCrunch, May 29, 2025.
21. Michael Nunez, [Hugging Face just launched a \\$299 robot that could disrupt the entire robotics industry](#), VentureBeat, July 9, 2025. The company launched a sub US\$ 300 robot, which can integrate with the Hugging Face Hub, enabling its developer community to access pre-built AI models, hardware designs, and software and assembly instructions.
22. Kate Park, [RLWRLD raises \\$14.8M to build a foundational model for robotics](#), TechCrunch, April 14, 2025.
23. The Robot Report, [RBR50 Spotlight: FANUC produces one-millionth industrial robot](#), August 12, 2024.
24. domainB, [China's AI-powered humanoid robots set sights on transforming global manufacturing](#), May 13, 2025.
25. Based on publicly available secondary sources that reference Unitree and UBTECH.
26. Based on multiple publicly available data and research reports that highlight the various chip components and hardware that are used to build humanoid robots.
27. Deloitte analysis based on data and information gathered from select major AI humanoid robot makers in the US and China.
28. Deloitte analysis based on data and information gathered from select major AI humanoid robot makers in the US and China.
29. Note to calculations: Using the 2026 estimated price range of US\$14,000 to US\$18,000 per unit, and 15,000 unit shipments, we multiplied the two variables to arrive at US\$210 to US\$270 million as overall revenue opportunity.
30. Using the variables and methodology noted in end note No. 26, we took the baseline scenario range of US\$210-270 million for 2026, and multiplied it by 3X and 4X to arrive at the other two probable 2032 market revenue potential presented in this paragraph. Our underlying assumptions for these relatively optimistic scenarios are mainly based on how fast the broader AI, robotics and tech industry might be able to address and workaround data, integration, safety, and cyber related challenges, and as price points become relatively attractive over time.
31. Justin Spike, [Data on animal movements help Hungarian researchers create a swarm of autonomous drones](#), AP News, December 19, 2024.
32. Deloitte analysis based on information gathered from publicly available sources about AI-enabled drones.
33. Karthik Ramachandran, Gillian Crossan, Duncan Stewart, Ariane Bucaille, [On solid ground: AgTech is driving sustainable farming and is expected to harvest US\\$18 billion in 2024 revenues](#), Deloitte TMT Predictions 2024, November 29, 2023.
34. Damon Johnson, [From Sci-Fi to reality: The latest in drone technology for 2024](#), Raising Drones, July 12, 2025.
35. Yahoo! Finance, [Britain to allow drones to inspect power lines, wind turbines](#), October 15, 2024; Joe Macy, [Autonomous UAS inspection system for power lines introduced](#), Unmanned Systems Technology, March 14, 2025.

36. Damon Johnson, [From Sci-Fi to reality: The latest in drone technology for 2024](#), Raising Drones, July 12, 2025.
 37. Aja Melville, "[Drone Wars: Developments in Drone Swarm Technology](#)," Forecast International, January 21, 2025.
 38. Deloitte analysis based on publicly available price information of select major industrial robots in the market.
 39. Gregory Haley, "[Increasing roles for robotics in fabs](#)," Semiconductor Engineering, Aug. 19, 2024.
 40. Rebecca Szkutak, "[We are entering a golden age of robotics startups — and not just because of AI](#)," TechCrunch, September 12, 2025.
 41. Deloitte China, "[Open Full-stack Intelligent Service Robot Ecosystem white paper](#)," April 24, 2025.
 42. [具身智能从新概念到新赛道: 人形机器人如何迈向规模化应用_新闻频道_中国青年网](#)
-

ACKNOWLEDGMENTS

The authors would like to thank **Dan Hamling, Rohini Prasad, Viswanath Anakkara, Joe Mariani, and Adam Routh** for their contributions to this article.

Cover art by: **Jaime Austin**; Adobe Stock

当SaaS邂逅智能体：预算体系、客户体验与劳动力生态的深度变革

随着智能体 (AI agent) 在软件即服务 (SaaS) 市场中的普及，企业使用和利用软件的方式可能会发生变化，从而改变商业模式、能力和用户期望。

随着代理式AI (Agentic AI) 能力的不断成熟，以及企业级SaaS供应商持续扩展其平台以创建、集成和编排智能体，企业采购和使用软件的方式可能发生根本性转变。到2026年，SaaS应用有望变得更加智能、个性化、自适应且具备自主性，逐步演变为能够从自身经验中学习的实时工作流服务联盟。这一演变将可能颠覆传统的定价模式，基于订阅和按席位授权的模式或将被融合了按使用量和按成果计费的混合模式所取代。这些进步将很可能在软件实施和商业化方面引入新的复杂性，甚至可能重新定义整个SaaS商业模式。

什么是智能体？

在AI领域，智能体是一种能够感知其环境、自主采取行动以实现目标，并可能通过机器学习或获取知识来提升自身性能的实体。¹

智能体可能会从2026年开始逐步推动SaaS市场的变革

为了更全面地理解现状，让我们先审视一下AI在市场中的整体应用趋势。德勤《2025年技术价值》调查显示，57%的受访者将年度数字化转型预算的21%至50%投入AI自动化，20%的受访者投入比例达到50%或更高（对于年收入130亿美元的公司而言，平均投入约为6亿美元）。²近四分之三的受访领导者表示，其组织在过去十二个月中为AI和生成式AI (Generative AI) 技术提供了资金支持（位居首位），另有39%的组织为代理式AI提供了资金支持。

基于此，德勤预测，到2026年，近半数的企业将把其数字化转型预算的50%以上投入AI自动化领域，而代理式AI将吸引更高比例的企业投资，可能达到75%。尽管《技术价值》调查仅聚焦于美国受访者，但我们认为全球企业将遵循类似的发展路径，可能仅延迟一到两年。SaaS通常是数字化转型工作的基础，若将这些更广泛的支出变化视为参考指标，我们预计在未来一年内，作为SaaS组成部分的自主智能体相关支出也将相应增加。

这些投资和技术进步最终会走向何方？目前，一些关于未来的乐观愿景正受到关注。有人指出，部分甚至整个企业级应用最终都可能被智能体所取代。³德勤预测，就部分企业应用而言，这一未来图景或许终将实现，但不会在2026年到来。即便在智能体人工智能技术快速发展和投资的推动下，这一愿景至少也需要五年或更长时间才能成为现实。该愿景仍面临诸多挑战，因为传统SaaS提供商在复杂工作流程中占据着庞大份额，这些份额可能难以被取代。⁴

到2026年，我们可能会看到大量实验性尝试，整体能力得到提升，同时SaaS市场将缓慢重构，届时以人工智能为核心的公司将展开竞争。这种相对温和的进展速度，可能是因为SaaS的“智能代理化（agentification）”不仅涉及技术变革，还牵涉到供应商和用户双方的商业模式与运营模式的转变。

随着代理式AI的发展，SaaS将变得更加复杂

许多首席信息官（CIO）和首席技术官（CTO）仍面临降低成本并精简所使用供应商数量的压力。⁵在代理式AI时代，人们常常会问，组织应何时以及如何开始将投资转向采用智能体的解决方案，以实现更高的效率？

一部分大型SaaS提供商在向客户提供这些功能时采取了不同的路径。许多公司正在将智能体添加到现有产品中，并推出全新的智能体驱动的产品（近期的示例包括Salesforce Agentforce、SAP Joule Agents、ServiceNow Now Assist 智能体以及Workday Illuminate Agents）。⁶许多企业还在现有服务基础上构建代理开发框架，并引入新的数据管理和编排功能，以简化智能体的创建与管理（近期的示例包括Google Cloud Agent Development Kit、Oracle AI Agent Studio、SAP Business AI、Workday Build以及Adobe Experience Platform Agent Orchestrator）。⁷

在代理式AI时代，人们经常会提出这样一个问题：组织何时以及如何开始将投资转向配备智能体的解决方案，以提高效率？

此外，一些新兴的原生AI公司似乎正在开发具有潜在颠覆性的智能体解决方案，可能对现有企业造成冲击。短期内，客户服务等“较简单”的业务流程更有可能被颠覆，但这种颠覆未来可能蔓延至ERP（企业资源规划）和CRM（客户关系管理）等更为复杂的市场。大量投资正推动着这些初创企业的发展。⁸随着现有企业寻求扩展其代理产品组合并实现差异化，未来几年，许多这些新兴公司可能会被收购。事实上，Gartner®指出：“到2030年，35%的独立SaaS工具将被智能体取代，或被整合进主要SaaS供应商的大型代理生态系统中。”⁹

如今，企业可通过现有的软件即服务供应商获取人工智能智能体，借助其内置功能更便捷地测试并学习如何构建智能体解决方案。尽管企业初期可能采取这种“默认启用智能体”的策略，但随着经验积累，它们很可能会转向更为审慎的路径。德勤预测，企业将以自身数据为核心，从庞大复杂的智能体生态系统中筛选所需能力，开发自有智能体，并将所有组件整合为一个集成化、自主化的多智能体系统。

向代理式AI的过渡

为了更顺利地实现这一未来，需要解决若干挑战：

定价变得更加复杂

一个可能对SaaS用户和供应商都产生重大影响的领域，是智能体的定价与付费方式。在软件主要为本地部署的时代，通常采用永久性软件许可，并为升级和维护另行付费。由云计算推动的SaaS革命，使商业模式转向了订阅制。如今，SaaS的定价方式主要有几种常见模式。通常情况下，企业根据其拥有的用户数或席位数来计费。这些席位可能包含分级定价选项，不同级别根据用户类型提供不同的功能组合。此类定价模式相对简单且可预测。而基于使用量或消耗量的定价模式也正变得越来越普遍，但其费用则较难预测。这种模式通常依据所使用的API调用次数或令牌数（AI模型处理的文本或数据单位）来计费。

随着智能体的广泛应用，这些传统的定价模式将不足以反映供应商与消费者之间真实的价值交换。¹⁰理论上，可使单个用户具备多个用户的能力，并减少企业所需席位数量，从而影响SaaS供应商的收入。此外，智能体能够自主运行，其行为未必可预测——在完成任务时，它们可能会采取新颖或低效的路径。

向这些新模式转型很可能需要投入大量精力，我们预计在2026年及以后，定价方式将呈现多样化并经历诸多实践探索。即便最终会形成标准做法，也可能需要数年时间。其中，基于使用量和基于成果或价值的定价模式预计将越来越受欢迎。Gartner表示：“到2030年，至少40%的企业SaaS支出将转向基于使用量、代理或成果的定价模式。”¹¹

基于使用量定价

在基于使用量的定价模式中，每当智能体执行操作或完成任务时，客户都可能被收取费用。定价也可以基于计算时间、API调用次数、生成任务所使用的令牌数量，或智能体处于活动状态的时长（或以上因素的组合）。此外，也可以按固定周期收取单一智能体使用的固定费用，类似于为数字员工支付薪水。Maxio在最近对SaaS公司的一项调查中发现，目前83%的原生AI SaaS公司提供基于使用量的定价模式。¹²基于使用量的定价通常具有吸引力，因为它是可量化的，因此也是可审计的。

基于结果或价值定价

定价模式的变更将影响企业内的多个职能，并可能改变SaaS供应商的运营方式。

基于结果或价值的定价模式依赖于智能体驱动的SaaS应用所产生的实际业务成果——而这一点往往更难衡量。这可能简单到客户支持工单的解决数量，或由于HR代理而最终聘用的员工人数，也可能复杂到智能体所贡献的整体收入增长。尽管目前已有部分企业正在探索这一模式，但在该模式得到广泛应用之前，可能仍有很长的路要走。¹³代理系统仍需证明其能够产生持续且可靠的价值。

这些定价模式的变更将影响企业内的多个职能，并可能改变SaaS供应商的运营方式。首先，各方需就“代理”、“任务”、“流程”、“交互”和“结果”等基本定义达成一致。应明确定义“价值”的含义及其归属方式，并在合同中予以清晰沟通和约定。这可能需要工程师、销售人员、法务团队及其他相关人员投入大量努力并进行充分协调。证明某个智能体创造了价值或实现了业务成果可能具有挑战性，尤其是在使用由不同供应商提供的多个代理组成的多代理系统时。供应商的收入和客户的成本可能变得难以预测且波动较大。系统监控与计量功能可能需要更加先进，数据可观测性、计费以及财务合规性也可能需要实现实时化和自主化。

许多供应商的销售模式可能需要改变。销售人员必须向客户讲解这些新模式，并说服他们相信智能体能够创造价值，且这种转变不会比基于订阅的服务花费更多。销售人员的考核和薪酬方式也可能需要做出相应调整，并且可能需要与客户建立更深入的关系。

客户体验和用户界面可能会成为更重要的差异化因素

智能体本质上是自主的，那么它们为何还需要用户界面？与API类似，代理是“无头的”，即它们不直接连接用户界面。然而，仍需要有某种方式实现交互和可视化。那么，这种交互方式会是什么样？会存在一个主要的智能体界面，还是多个界面？是由SaaS供应商还是第三方来“控制”通往代理的网关？

未来几年，德勤预测SaaS智能体的用户体验和界面将变得更加：

- **个性化且主动：**界面将根据个人用户自适应地提供基于特定职责和先前操作所需的工具和任务。它将提供量身定制的洞察，并建议用户采取具体行动。
- **对话式交互：**人机交互将从菜单选择和点击操作转向自然语言和语音指令。智能体将把自然语言转化为一系列结构化的API调用，从而无需依赖预定义的工作流程。未来的交互将不再局限于告诉软件如何操作，而是更侧重于要求软件达成某一特定结果。
- **诊断：**由于智能体具有自主性，一旦发生错误或意外情况，用户必须能够还原代理的决策过程，并理解其发生原因。透明性、可解释性、可逆性和可审计性对于建立信任至关重要。

另一个尚未明确的问题是交互层将位于何处。德勤预测，许多交互将在独立的SaaS应用程序中完成。许多SaaS供应商希望尽可能让用户停留在其应用程序内，以维持工作效率并增强用户对其产品的使用黏性。他们将越来越多地不仅提供自身系列的代理，还会接入其他供应商的代理。交互也可能通过一个独立的管理平台进行，该平台可由SaaS供应商提供，或来自第三方公司（类似于当前的SaaS管理平台）。这些“控制中心”能够整合来自多个供应商以及企业内部开发的代理，实现对使用情况、支出、访问权限、性能、状态、安全性和合规性的统一管理。¹⁴未来还可能出现代理市场，内部和外部的代理将在其中发布，企业可以动态地发现并集成新的功能。¹⁵这一交互或注意力层具有提供重要价值的潜力，因此在其周围可能会产生激烈的竞争。

小结

到2026年，通过SaaS应用使用智能体的规模将迅速增长，许多主要的SaaS供应商正致力于与客户合作，部署更强大的代理式AI解决方案。我们预计，所有由AI驱动的自动化领域投资都将增加，并延伸至SaaS应用中。企业将寻求提升流程效率、节约成本、增强灵活性以及为员工提供个性化功能。届时将出现大量实验性尝试和多样化的定价模式。总体而言，德勤预测，未来将逐步迈向由集成化、自主式多代理系统驱动的新阶段。

SaaS 客户需考虑的准备工作：

- 投资数据管理：对于智能体而言，数据的访问、集成、可观测性以及数据治理可能变得更加重要。数据不一定需要集中存储在一个单一的存储库中，但应在整个组织内保持一致且易于访问。
- 拥抱日益增长的复杂性：未来将出现更多模型、智能体和供应商、新的生态系统及数据关系。企业将不得不让来自不同供应商的智能体协同工作，这可能导致定价和运营上的复杂性。
- 预计将出现多元化的定价模式：随着包含多种许可、基于使用量的定价以及基于价值或成果的定价等混合模式成为标准，智能体的定价模式可能带来不确定性。企业应加强其实时财务能力以应对这一趋势。
- 帮助员工成为AI协调者：可以投入更多时间像管理同事一样管理智能体——设定目标、监督工作，并验证和纠正其行为。在重构工作流程时，应明确界定人类负责的任务、代理负责的任务以及双方协作完成的任务。这是一场文化变革，而不仅仅是一次软件升级。

SaaS 供应商需考虑的准备工作：

- 为更激烈的竞争做好准备：随着生成式AI工具让任何人都能越来越容易地编写代码，编写代码的成本趋近于零。这将导致与原生AI公司甚至客户自身之间的竞争加剧，因此需要更强的产品差异化能力。
- 关注互操作性：代理将需要在多个系统之间运行，协调任务，并共享数据和目标，同时保持安全性和合规性。企业应为更加开放和互操作的环境做好准备，以便在无法满足客户期望时，客户能够轻松更换服务提供商。
- 转变销售模式：由于定价模式多样化，收入预测等事项变得更加复杂，因此对销售团队的考核与激励方式也必须随之调整。应帮助客户预测成本，并提供简单灵活的混合定价模式。通过围绕定价模式的沟通，发现客户尚未满足的需求，进一步深化客户关系。

中国SaaS行业发展前瞻

随AI应用的不断深入，当前中国正处在SAAS行业的价值修正期，SaaS行业正面临变革的拐点AI（特别是AI Agent）不再仅仅是SaaS产品的亮点功能，而正在重塑SaaS行业的商业模式和竞争格局。

AI不断重塑生产力，提升效率

当下，中国的AI行业发展迅速，AI与SaaS的融合正不断打造SaaS发展的“黄金十年”。以Deepseek、通义千问为代表的国产大模型，可以有效将SaaS工具从“流程自动化”推向“决策智能化”，使用大模型等AI软件，最高可降低推理成本95%，有效提升工作效率。企业正利用开放式的AI生态，构建可复用的技术底座、高效的数据流转机制、跨部门的协作文化，推动AI在SaaS领域的规模化部署。

AI加速SaaS国产替代，向按成果付费模式的转变

AI技术的不断深度渗透与国产大模型替代的加速推进，形成中国SaaS行业发展新机遇，推动中国SaaS由早前的单一浅层功能向业务流重构一体化功能的迈进。同时中国的SaaS行业也将不断从早期的功能型收费的单一收费模式转向按成果付费模式。实现技术服务与客户业务成效的更深层次绑定。

AI赋能国产SaaS平台的出海，提升国际化竞争力

AI可以通过多模态交互，提升用户体验，大规模代替人力，国产大模型带来的低成本趋势，有效减轻SaaS企业出海成本压力。AI可以更好适应不同海外地区的本地化需求，包括界面语言的转换，产品逻辑、交互方式乃至运营策略与当地文化的深度融合，推动国产SaaS平台的出海，中国SaaS平台的国际化竞争力不断提升。未来，日本、东南亚、中东等地区将是中国SaaS出海的重要海外市场，中国SaaS企业要不断提升在全球网络、合规支持与技术工具方面的核心能力。

David Jarvis

United States

Gopal Srinivasan

United States

Gillian Crossan

Global

Sayantani Mazumder

India

China Widener

United States

Roger Chung

China

Girija Krishnamurthy

United States

ENDNOTES

1. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. (New York, NY: Pearson, 2021).
2. Tim Smith et al., "AI is capturing the digital dollar. What's left for the rest of the tech estate?", Deloitte, October 16, 2025; AI automation includes basic automation, process automation with agents, process reimagination, and organizational reimagination.
3. Eric Newmark, "The agentic evolution of enterprise applications," IDC, April 4, 2025.
4. Dan Gallagher, "Software's death by AI has been greatly exaggerated," Wall Street Journal, August 27, 2025.
5. Zylo, "111 unmissable SaaS statistics for 2025," accessed October 2025; Matt Ashare, "AI drives up compute costs as cloud inflation slows," CIO Dive, February 18, 2025.
6. Salesforce, "Agentforce," accessed October 2025; SAP, "Joule Agents," accessed October 2025; ServiceNow, "AI Agents," accessed October 2025; Workday, "Workday unveils next generation of Illuminate Agents to transform HR and finance operations," press release, May 19, 2025.
7. Oracle, "Oracle introduces AI Agent Studio," press release, March 20, 2025; SAP, "Business AI," accessed October 2025; Adobe, "Adobe launches Adobe Experience Platform Agent Orchestrator for businesses to activate AI agents in customer experiences and marketing workflows," news release, March 18, 2025; Workday, "Workday unveils Workday Build, giving developers the tools to build the future of work," press release, September 16, 2025; Erwin Huijzena and Bo Yang, "Agent Development Kit: Making it easy to build multi-agent applications," Google for Developers, April 9, 2025.
8. Joanna Glasner, "AI autonomous agents are top 2025 trend for seed investment," Crunchbase News, June 17, 2025; Jacob Robbins and Kia Kokalitcheva, "Y Combinator is going all-in on AI agents, making up nearly 50% of latest batch," PitchBook, June 11, 2025.
9. Gartner, AI agents are disrupting SaaS pricing: What must CIOs do?, July 16, 2025 (ID G00834627). GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.
10. Adrian Radu, "Billing infrastructure in the age of co-pilots and AI agents," Lightspeed, March 6, 2025.
11. Gartner, AI agents are disrupting SaaS pricing: What must CIOs do?, July 16, 2025 (ID G00834627). GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.
12. Maxio, 2025 pricing trends: Usage-based models and the path to SaaS growth, 2025.
13. Zendesk, "Zendesk first in CX industry to offer outcome-based pricing for AI agents," August 28, 2024.
14. Salesforce, "Agentforce Observability," accessed October 2025; Google Cloud, Gemini Enterprise, accessed October 2025.
15. Bojan Ceric and Prakul Sharma, "Scaling AI agents may be risky without an enterprise marketplace," Deloitte Insights, September 15, 2025.

ACKNOWLEDGMENTS

The authors would like to thank **Diana Kearns-Manolatos, Faruk Muratovic, Rohan Gupta, Khusro Khalid, Laura Shact, Girish Srinivasan, Pavan Srivastava, and Prakul Sharma** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

半导体供应链：新技术涌现，挑战依旧

随着针对下一代关键AI芯片技术的贸易限制不断升级，企业领导者应迅速调整，以增强供应链的韧性。

地缘政治紧张局势和不断升级的贸易限制正在重塑半导体供应链，对AI芯片创新、全球经济、国家安全以及科技进步产生深远影响。这些高科技工艺和材料中的许多环节依赖于少数供应商，而这些供应商在关键地区的主导地位促使各国政府设置贸易壁垒，以保护战略利益并降低依赖性。长期以来，为下一代AI系统和高性能计算（HPC）数据中心制造世界上最先进的芯片意味着要在脆弱的供应链中艰难前行，但如今风险已大大增加。

德勤预计，包括前端及后端芯片制造中的蚀刻工艺、全环绕栅极（GAA）晶体管、电子设计自动化（EDA）以及支撑先进AI模型的软件工具。此外，德勤预测，受贸易壁垒影响的关键技术投资总额将至少达到300亿美元，涵盖极紫外光刻设备（EUV）和高带宽内存（HBM）协同封装工具等领域。¹然而这些投入与相关技术催生的约3000亿美元AI芯片市场规模相比仍相形见绌，这凸显了这些技术在全球半导体供应链中的核心地位。²

AI重塑全球半导体供应链

德勤与半导体行业协会联合开展的一项关于人工智能数据中心半导体含量的研究指出，全球半导体供应链高度相互依存，各国正着力保障自身对人工智能芯片及硬件组件的获取，这些组件对生成式AI（Gen AI）、高性能计算和自主系统而言至关重要。³因此，与两三年前相比，出口管制及其他贸易限制措施在2025年和2026年已开始对更大范围的半导体设备、材料、软件、设计工具、各类芯片以及封装和组装工具产生影响（图1），这并不令人意外。

图1：美国和欧洲的贸易管制在2025年、2026年及以后已扩大至涵盖多种半导体技术

半导体与人工智能芯片领域在不同时间点（2019 - 2025*）受到出口管制的特定方面



来源：德勤分析。2019年至2025年的数据基于从公开可获取的来源收集的信息，包括联邦公报网站（美国商务部工业与安全局）发布的文件和公告。2026年的信息基于与中小企业交流所获得的对话内容和前瞻性洞察。

*注：2025年的信息截至2025年10月8日。

Deloitte. Insights deloitte.com/insights

AI系统的性能依赖于一个由全球分布的多项技术构成的狭窄堆栈，包括先进的AI逻辑设计、前沿的前端节点制造以及先进封装技术。实现这些能力需要多方利益相关者之间的协作，例如集成器件制造商 (IDMs)、晶圆代工厂、设备制造商、设计供应商、外包半导体封装测试 (OSAT) 供应商、系统集成商、外包渠道分销合作伙伴以及来自不同国家的政府机构。⁴

出口管制正在重新定义先进AI逻辑设计的未来

2024年和2025年，美国对多项关键半导体技术（尤其是EDA工具）的限制先收紧后放松。⁵EDA 流程包括设计逻辑、芯片布局与布线、仿真、AI 增强设计、验证以及集成工作流程，这些对于开发先进的 AI 加速器至关重要。

例如，对于基于环绕栅极场效应晶体管（GAAFET）开发的芯片，此前已存在限制。⁶GAAFET 是一种面向 5 纳米及以下、3 纳米及以下逻辑设计的新兴晶体管架构，可为计算密集型的生成式AI工作负载提供性能和能效优势。2024 年 12 月，美国进一步扩大了出口管制范围，将支持先进计算节点开发与设计的软件和工具纳入管制。⁷随着这些新的出口管制措施的出台，它们可能在2026年对更广泛的EDA生态系统及代工厂合作伙伴产生影响。

2026年及以后的预测与展望

随着对基于GAAFET的芯片限制增加，使用GAAFET工艺设计套件（PDK）开发先进节点的非美国盟国晶圆厂将需要EDA工具支持以进行验证。但如果某一地区无法获得这些工具，则可能不得不依赖更老旧、效率更低的制程节点，或被迫发展本土的EDA能力，这两种情况都可能导致产品周期延长并削弱竞争力。此外，特别是在中国，对先进计算芯片的额外管制以及对AI模型权重的新管制，提高了企业与客户及商业伙伴合作时的合规要求，⁸AI模型及其模型权重的规模和质量正日益影响用于芯片设计的AI驱动EDA工具的能力。⁹

德勤预测，到2026年，电子设计自动化和逻辑设计领域的相关企业可能会受到这些管控措施的影响：针对与AI加速器相关的晶圆厂知识产权（IP）库、PDK以及性能测试输出，在实体、地点和最终用途方面，企业可能面临更严格的审查和更细致的披露要求。用于产品验证和模型微调的评估硬件（包括用于测试的参考模型权重及输出结果）也可能受到更密切的审查。¹⁰涉及AI硬件协同设计的公司可能需要建立可信国家通道，或不得不调整工作流程：例如，可将模型权重保留在美国或盟友的安全IT基础设施内，同时允许代工厂伙伴远程运行测试。¹¹

AI系统前端节点制造中面临的瓶颈问题

美国和荷兰继续限制对EUV设备的获取，该设备被广泛认为是生产最先进工艺节点所必需的。¹²尽管美国不具备本土EUV生产能力，但通过与盟友（如荷兰）协调实施出口管制，美国仍能影响哪些国家可以购买这些设备，主要目的是保障技术和国家安全。与此同时，中国则通过国内芯片设备企业，采用多重图形化技术定制深紫外光刻技术，积极推进光刻设备的研发进程。¹³尽管这些方法看起来有效，但其运行速度较慢且成本较高。¹⁴为保护国家安全利益，美国对用于精密蚀刻的工具加征了额外的出口限制，这些工具对于构建复杂的AI架构至关重要。¹⁵

2026年及以后的预测与展望

先进的刻蚀技术对于在5纳米以下节点制造前沿AI芯片至关重要。芯片行业采用双重、四重以及基于间隔层的图案化技术，在最先进的AI芯片上制造精细结构。¹⁶因此，源自美国的刻蚀工艺设备，以及使用美国刻蚀技术知识产权在海外设计或制造的刻蚀设备和工具，可能在2026年成为新的瓶颈。此外，光学元件（如透镜和反射镜）和光罩（photomask）等作为晶圆制造设备关键组成部分、承载着晶圆上图案蓝图的部件，也可能面临限制。

此外，特种气体（如硅烷和氟化衍生物）¹⁷以及关键矿产（包括镓、锗和锑）¹⁸作为先进节点制造工艺的一部分，为全球芯片供应链引入了额外的摩擦点。

由于前端工艺设备、组件和原材料面临广泛的出口管制，德勤预测，美国、中国台湾和韩国的5纳米及以下、3纳米及以下制程的产能扩张将持续加速，这一趋势将延续至2026年及以后。与此同时，中国预计将继续专注于采用多重成像技术作为替代方案的成熟DUV技术。

因此，跨国芯片设备企业应调整其在区域层面的前道晶圆制造相关资本支出规划。晶圆制造设备供应商、零部件供应商以及代工厂可能面临比2024年和2025年更长的认证、升级和安装周期。随着芯片设计公司适应新要求——开发功能简化或降级的AI XPU（即高端AI芯片的性能降低版本）以及面向特定区域的工艺库，以满足中国及其他非美国盟友国家日益增长的生成式AI芯片需求——前端制造设备供应商需提供更强支持的可能性也将上升。

贸易管制扰乱先进封装和测试环节

先进封装技术已迅速成为出口管制的战略目标。来自荷兰的测量与检测设备正面临出口限制。¹⁹由于其在高密度芯片堆叠中的关键作用，²⁰当前和未来生成式AI芯片的关键基础组件。²¹用于测试和验证先进AI芯片所必需的特定类型的芯片设备（刻蚀、沉积、光刻、离子注入、退火、计量与检测以及清洗设备）受到出口管制。²²这是因为这些技术被视为敏感且具有潜在的两用性，未来可能继续受到额外的贸易管制。

2026年及以后的预测与展望

正如2024年TMT预测中所强调的，芯粒（chiplets）和异构架构正迅速成为面向高性能计算AI工作负载的生成式AI芯片的首选封装模式。²³然而，由于需要从不同地区的多家供应商采购并封装多个芯片裸片和组件，其复杂性可能导致芯粒（chiplets）在2026年成为重大的地缘政治瓶颈。值得注意的是，基于芯粒的解决方案预计在2026年将带来约1000亿至1100亿美元的年收入。²⁴

HBM也已成为生成式AI训练和推理工作负载的关键组成部分。截至2025年年中，业界对HBM的共封装技术进行了更密切的关注，包括识别HBM与逻辑芯片共封装的具体位置。²⁵因此，参与组装、测试和封装的半导体企业可能需要提供额外的信息披露，包括列出参与封装的OSAT供应商或后端制造厂商，明确系统共封装的地点，指明中间产品或最终产品所运往的目的地国家，以及详细说明相关的性能阈值。

到2026年及以后，日益凸显的趋势可能是对后端工艺效率的依赖不断加深，以确保新产品能够按时上市。随着异地封装站点（特别是涉及HBM、逻辑芯片和高速输入/输出的站点）在布线和文件记录方面的要求日趋严格，供应链的每个环节——从前端晶圆厂的生产排程和EDA供应商的设计签核，到终端客户ODM/OEM（原始设计制造商和原始设备制造商）的产品发布——都将更加依赖于先进封装相关工艺审批和流程的完成速度。封装供应商或OSAT（外包半导体封装测试）厂商任何一方的延误，都可能影响良率爬坡和调优进程，进而引发将生产设施迁回本土或转移至盟友国家的“近岸外包”或“友岸外包”举措。

总体而言，这些因素可能会影响2026年及以后多个地区规划中的AI数据中心的部署。预计到2026年，超大规模企业、云服务提供商以及各行业公司将合计在AI数据中心上投入约5,000亿美元，到2028年将达到约1万亿美元。²⁶其中芯片解决方案约占该支出的50%至60%。鉴于预期的增长，供应链中断可能在三年期间影响价值数百亿甚至数千亿美元的半导体产品。

小结

中国加强其国内半导体生态系统

严格的出口管制以及对一系列半导体技术的限制，阻碍了中国获取最先进的AI芯片。这促使中国加快国内半导体创新步伐，尤其是当中国意识到这些举措可能影响其在7纳米以下和5纳米以下制程上的进展，而中国大陆以外的芯片制造商正从2025年的3纳米和2纳米推进至2026年和2027年的1.8纳米。²⁷

随着中国开发应对出口管制的变通方法，它可能会探索全球半导体供应链的多个方面，不仅限于前端制造，还包括芯片设计和先进封装。²⁸尽管使用较旧制造工艺的复杂芯片可用于先进封装，但美国可能会实施额外的管控和审查，以限制面向前沿AI芯片的此类封装系统的性能。

主权技术体系竞争加速，催生区域格局新变局

技术主权是一种理想目标，各国致力于独立开发、掌控和监管数字技术。²⁹由于AI被普遍视为推动经济发展和国家竞争力的下一个主要动力，各国政府正寻求对其数字基础设施实施更直接的控制，因而AI生态系统备受关注。各国和地区都不愿进一步落后，或被迫丧失自身主权。这种紧迫感日益增强，因为目前先进的AI能力主要集中于少数国家和企业手中。此外，随着美国和欧洲纷纷将高端芯片制造回迁本土，预计到2026年及以后，它们将在国内以及印度、越南和马来西亚等国家投资建设替代性的先进封装与测试中心。³⁰

半导体行业亟需增强供应链韧性

整个生态系统中的芯片企业可能需要通过内部压力测试等方式主动提升韧性，主要用以自我评估端到端的供应链，并加强网络安全准备能力。³¹

其次，跨区域的供应链多元化，以及对替代采购策略和渠道合作的投资至关重要。确保关键材料和零部件供应链独立的战略重要性，要求加快本地化进程并提升监管适应能力。此外，地缘政治问题可能导致全球AI生态系统分裂，带来诸如通过灰色市场出口芯片等风险，并加剧企业加强产品及供应链监控与追踪能力的压力。

尽管面向AI推理优化的芯片市场预计到2026年将增长至数百亿美元，但大部分高性能计算仍将依赖最先进的AI芯片，这些芯片主要部署在超大规模数据中心，或使用与数据中心相同芯片和机架的本地服务器中。³²因此，新的和额外的出口管制及要求可能会针对AI推理芯片及相关基础设施，整个半导体行业应为此在从采购到分销的各个环节开发替代性的供应链方案。

随着从训练向推理的转变，软件作为半导体更核心组成部分的重要性也将不断提升，例如，利用软件编程技术将一个主要用于训练的大型单体AI GPU重新配置为多个较小的GPU切片或虚拟GPU实例，以用于推理任务。³³

此外，总部位于美国和欧洲的设备原始设备制造商（OEM）可能需要将生产和组装从中国转移至东南亚和印度等新兴制造中心。这一转变可能在短期内推高成本，进而导致消费类科技设备价格上涨。半导体企业应保持灵活性并实现规模化运营，预判并适应2026年后的贸易格局变化，同时探索替代性的战略性国家联盟，以保障关键物流通道和基础设施的安全。

随着贸易紧张局势重塑全球联盟和渠道合作关系，芯片产业的韧性将在进入2026年前面临前所未有的考验。全球芯片供应链具有高度互联和战略性的特征，这凸显出各方产业利益相关者亟需积极开展合作，以增强半导体供应链的韧性。

Karthik Ramachandran
India

Jeroen Kusters
United States

Girija Krishnamurthy
Global

Duncan Stewart
Canada

Deb Bhattacharjee
United States

Jan Thomas Nicholas
Malaysia

ENDNOTES

1. A note to methodology. Estimates include projected aggregate spending for 2026 on extreme ultraviolet equipment, AI-based etch equipment, select advanced packaging equipment including high-bandwidth memory co-packaging tools, and AI chip design software and tools.
2. In 2025, Deloitte Consulting LLP performed an analysis of the data center market, including a rough bill of materials for the various components and market sizes. This analysis is due to be published in December 2025.
3. Ibid.
4. Ibid. Importantly, an AI server rack is not just a monolithic unit but a far more complex, integrated system that comprises tens of thousands of components ranging from advanced chips, memory dies, analog integrated circuits, controllers, power devices, and passives like substrates and capacitors.
5. Karen Freifeld and Surbhi Misra, "[As trade war truce with China holds, US lifts curbs for chip design software and ethane](#)," Reuters, July 3, 2025; Joe Cash, "[China says successful US trade talks make return to tariff war unnecessary](#)," Reuters, July 18, 2025.
6. Bureau of Industry and Security and US Department of Commerce, "[Federal Register, vol. 89, no. 173](#)," Sept. 6, 2024.
7. New software and technology controls included restrictions on electronic computer-aided design and technology computer-aided design software and technology, especially when these are used for designing advanced node-integrated circuits. To read further, see: Bureau of Industry and Security and US Department of Commerce, "[Commerce strengthens export controls to restrict China's capability to produce advanced semiconductors for military applications](#)," Dec. 2, 2024.
8. Bureau of Industry and Security and US Department of Commerce, "[Framework for artificial intelligence diffusion](#)," Federal Register, Jan. 15, 2025.
9. Wenji Fang, Jing Wang, Yao Lu, Shang Liu, Yuchao Wu, Yuzhe Ma, and Zhiyao Xie, "[A survey of circuit foundation model: Foundation AI models for VLSI circuit design and EDA](#)," arXiv, March 28, 2025.
10. For further information on AI model weights related technology controls, see: US Department of Commerce and Bureau of Industry and Security, "[Federal Register, vol. 90, no. 9](#)," Jan. 15, 2025.
11. Insights based on conversations and interviews with Deloitte experts in the areas of the semiconductor industry, supply chains, and export control impact.
12. Chris Miller, "[How US export controls have \(and haven't\) curbed Chinese AI](#)," AI Frontiers, July 8, 2025.
13. Stefano Lovati, "[China invests €37 billion to develop domestic EUV lithography systems](#)," Power Electronics News, Feb. 11, 2025.
14. Pablo Valerio, "[China semiconductor ambition and adversity](#)," EE Times, May 19, 2025. Additionally, US regulations included restricting and capping the production of advanced AI chips far below the domestic demand in China.
15. See Bureau of Industry and Security and US Department of Commerce, "[Federal Register, vol. 89, no. 173](#)," p. 7. As noted in this document, atomic layer etching helps produce vertical edges required in high-quality, leading-edge advanced devices and structures, including gate-all-around field-effect transistor and similar 3D structures. Anisotropic dry etching is critical for gate-all-around field-effect transistor and similar 3D structure fabrication. It is also an important tool for fin-shaped field effect transistor (FinFET) fabrication.
16. Ibid.
17. US Department of Commerce and Bureau of Industry and Security, "[Foreign-produced direct product rule additions, and refinements to controls for advanced computing and semiconductor manufacturing items](#)," Dec. 5, 2024.

18. Sara Bulter, "[How China's rare earth metals export ban will impact supply chains in 2025](#)," Optilogic, Feb. 17, 2025.
19. Deloitte analysis based on conversations and insights gathered from industry experts and cross-validated with multiple secondary sources, including: Abbie Windsdale, "[Netherlands takes bold step to tighten semiconductor export control](#)," Tech Announcer, Jan. 16, 2025.
20. For example, hybrid bonding is fundamental to developing advanced 2.5D and 3D chip designs and heterogeneous architectures (or chiplets), as it enables ultra-fast data transfers (up to 17 TB/s) that are critical for AI and high-performance computing. To read further, see: Sam Naffziger, "[Future of AI hardware enabled by advanced packaging](#)," IEEE Electronics Packaging Society, May 28, 2024.
21. Duncan Stewart, Karthik Ramachandran, Prashant Raman, and Ariane Bucaille, "[Silicon building blocks: Chiplets could move Moore's Law forward](#)," Deloitte Insights, Nov. 19, 2024.
22. Bureau of Industry and Security, "[Commerce strengthens export controls to restrict China's capability to produce advanced semiconductors for military applications](#)," press release, Dec. 2, 2024.
23. Stewart, Ramachandran, Raman, and Bucaille, "[Silicon building blocks](#)."
24. Xiaoxi He and Yu-Han Chang, "[Chiplet technology 2025-2035: Technology, opportunities, applications](#)," IDTechEx, accessed Oct. 1, 2025.
25. US Department of Commerce and Bureau of Industry and Security, "[Foreign-produced direct product rule additions, and refinements to controls for advanced computing and semiconductor manufacturing items](#)."
26. Duncan Stewart, et al, "[Why AI's next phase will likely demand more computational power, not less](#)," Deloitte Insights.
27. For context, state-of-the-art chip fabs in the United States and Taiwan were already pushing the boundaries toward sub 7 and sub 5 nm as of 2020 to 2021, indicating China is probably at least four to five years behind (see [Deloitte 2024 semiconductor outlook](#)). Therefore, initiatives such as Beijing's Big Fund III actively support the expansion of local semiconductor capabilities, notably electronic design automation (EDA) and lithography tech development. To read further, see: Anton Shilov, "[China to pivot \\$50 billion chip fund to fighting U.S. squeeze as trade war escalates — country to back local companies and projects to overcome export controls](#)," Tom's Hardware, June 27, 2025.
28. The Chinese Academy of Sciences worked with domestic chip design players on an open-source project to develop an AI system that used large language models to accelerate chip design and build fully functional central processing units. To read further, see: Mark Tyson, "[China claims to have developed the world's first AI-designed processor — LLM turned performance requests into CPU architecture](#)," Tom's Hardware, June 12, 2025. Additionally, Huawei's breakthroughs in developing EDA tools capable of supporting 14 nm processes and above mark significant milestones. To read further, see: Omar Sohail, "[Huawei has reportedly developed 14nm EDA tools, which the company will employ to mass manufacture its Kirin 9020, but the company is still limited to the 7nm architecture](#)," WCCF TECH, June 11, 2025.
29. David Jarvis, et al, "[A new era of self-reliance: Navigating technology sovereignty](#)," Deloitte Insights.
30. Analysis based on multiple publicly available secondary sources that discuss the chip industry's plans to commence new AT hubs in countries including India, Malaysia, and Vietnam.
31. Aside from trade-related issues, as we already mentioned in our [2024 Global Semiconductor Outlook](#) report, cyber threats are surging, requiring chip fabs and AI systems to intensify security measures against malware targeting critical infrastructure.
32. Duncan Stewart, et al, "[Why AI's next phase will likely demand more computational power, not less](#)," Deloitte Insights. Deloitte analysis based on conversations and insights gathered from industry experts.
33. Gwangoo Yeo, Jiin Kim, Yujeong Choi, and Minsoo Rhu, "[PREBA: A hardware/software co- design for multi-instance GPU based AI inference servers](#)," arXiv, Nov. 28, 2024.

ACKNOWLEDGMENTS

The authors would like to thank **Nina Zhang, Amy Scimeca, Karan Aggarwal, Jesse Singh, Michael Greco, and Pablo LeCour** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

简短而精彩：微剧凸显短视频内容的演变

从独立创作者到大型平台，微短剧正在帮助重新定义观众与内容的连接和消费方式。

觅食者、零嘴客与漫游者：面对内容的泛滥，社交媒体受众不断寻找值得消费的优质内容。他们从无穷无尽的信息流中筛选，这些内容或许符合他们的元数据偏好，却未必能满足其情感或智力上的需求。像微剧、微短剧这类连载式短篇叙事，能否在高度碎片化的注意力经济中提供更丰富的营养与更强的连续性？

微短剧，有时也称为微戏剧或短剧系列，是一种剧本化视频系列，每集仅有几分钟，以“移动优先”消费和快速吸引为目标。DramaBox、ReelShorts、ShortMax 和 DramaWave 等移动应用程序在亚洲和美国已创造数十亿美元收入，拥有数亿用户。¹这种爆炸性增长正在重新定义观众对数字娱乐的期待，同时也为内容创作者、平台和品牌带来了新的机遇与挑战。²

预计到2025年，微短剧内容的应用内收入将达到38亿美元。³德勤预测，到2026年，应用内微短剧的收入增长将翻倍，达到78亿美元。德勤预计，美国将在2025年占据全球收入的一半，但随着其他市场将更多观看量和下载量转化为收入，其份额将下降至40%。随着越来越多的观众接触到微短剧，我们认为，简短与连续性娱乐的结合将更具吸引力，并受益于社交媒体上病毒式传播的增强。此外，我们预计会有更多微短剧在社交平台上脱颖而出，占据更多用户注意力时长，并登上美国社交媒体参与度排行榜前列。最后，我们预计一些精明的视频流媒体服务提供商将更积极地在其平台上尝试推出短剧式连续性内容。

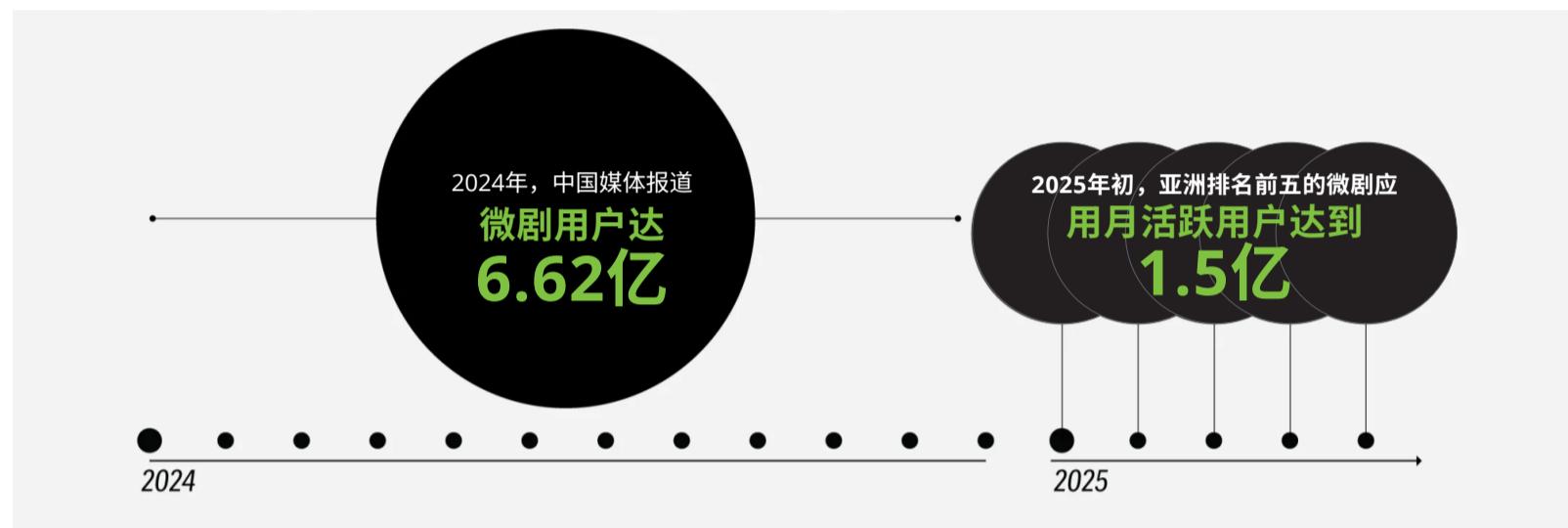
尽管短剧形式看似是专为社交媒体而生的创新，但它可能挑战主流社交平台的主导地位。社交平台算法推送内容反复无常，可能导致用户难以“追剧”并及时跟进后续剧集。这或将促使更多观众和创作者转向竞争性的微剧应用程序。与此同时，已有迹象表明，年轻一代正因社交媒体感到不堪重负，既难以跟上无限滚动的信息流，又无法摆脱它的影响。⁴短剧式的怀旧内容回归线性电视会是解决方案吗？

微短剧正在吸引观众

系列化短剧内容的日益流行似乎正在主流社交视频平台上获得关注，同时也推动了新兴竞争者的发展：这些成功的微短剧服务正在争夺人们用于数字娱乐的有限时间。

微短剧类移动应用正日益流行，提供数百集时长60至90秒的系列剧集，剧情充满转折与悬念，旨在吸引观众持续关注并激发追剧欲望。这些新型微短剧制作快速、成本低廉，根据观众互动不断优化调整，并常常借助主流短视频社交平台推动内容曝光与话题传播。⁵

图1：亚洲微剧的流行助力全球成功



资料来源：Omdia，《亚洲新兴微剧趋势》，2025年3月25日。

Deloitte Insights deloitte.com/insights

中国的爱奇艺提供了超过15,000部免费和付费的微短剧，过去一年中微短剧的观看时长实现了显著增长，并在微短剧生态系统周边拓展了电子商务功能。⁶截至2024年，中国媒体报道全国微短剧用户规模约为6.62亿人。⁷另一份2025年初的报告发现，中国排名前五的微短剧应用的月活跃用户（MAUs）总数为1.5亿。⁸中国领先的视频流媒体平台正与短视频平台合作，共同制作优质短剧，或许预示着长视频与短视频内容融合的未来趋势。

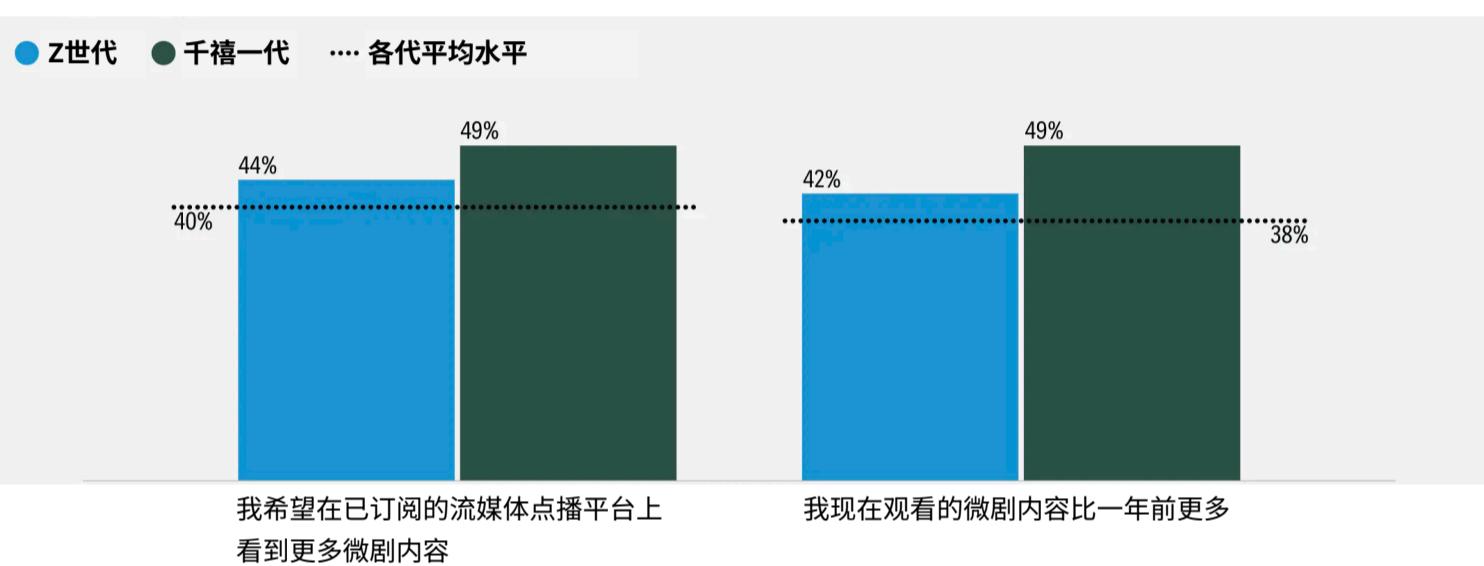
鉴于全球趋势的不断增长，印度正抓住机遇，大量创新平台和成熟的媒体公司纷纷进入市场。⁹Zee Entertainment和Kuku FM等OTT平台已推出专门的微短剧频道，平台报告显示，引入短剧形式的试点内容后，日均观看时长翻倍。¹⁰在印度这个价格敏感度极高的市场，由于每用户平均收入（ARPU）相对较低，各平台正在尝试为单集内容推出小额支付选项，同时也有平台推出灵活的订阅方案，包括将订阅费用与广告收入相结合的混合模式。¹¹观众可以免费观看前几集，但之后需要付费才能继续观看后续剧情的发展。

如今，领先的微短剧应用程序经常出现在美国应用商店下载量前25名中。

对短剧和微剧应用的需求也正在蔓延至亚洲以外的地区。¹²一份报告发现，2025年第一季度，微短剧应用的全球收入从2024年第一季度的1.78亿美元激增至近7亿美元。¹³美国已成为DramaBox、ReelShort和GoodShort等短剧应用收入最高的市场。¹⁴领先的微短剧应用程序目前经常出现在美国应用商店下载量前25名中。¹⁵跨界进入社交视频平台使它们在美国获得了增长动力。¹⁶社交平台本身也观察到用户对微短剧内容的参与度越来越高。¹⁷即使是头部流媒体也开始涉足短视频和竖版内容领域。¹⁸

德勤开展的《数字媒体趋势》对美国消费者的调查显示，截至2025年5月，约有30%的Z世代和千禧一代了解微剧或微短剧。其中，近一半的人目前观看的微剧内容较一年前有所增加，且近一半的人希望在他们已订阅的视频点播流媒体平台上看到更多微剧内容，这表明流媒体平台在此领域具备一定的竞争潜力。（图2）

图2：年轻一代更热衷于系列化短形式视频内容：微剧和微系列
熟悉微剧或微系列的消费者中选择“非常认同”或“有些认同”的百分比



注：在1620名熟悉微剧或微系列的受访者中，Z世代有483人，千禧一代497人，X世代399人，婴儿潮一代230人，成熟世代11人。

资料来源：《2025数字媒体趋势》，19.5版

Deloitte Insights deloitte.com/insights

越来越多的观众被独立创作、形式短小、以叙事为主的内容所吸引。为此，越来越多的创作者正在建立自己的独立工作室，利用数据、AI和社会化平台来扩大影响力。

社交化、高性价比且数据驱动的工作室兴起

媒体和娱乐行业正受到短视频内容的行为模式与经济效应、社交视频平台的功能与覆盖范围，以及优质内容所带来的声望与高昂成本双重影响的重塑。随着观众将更多的娱乐时间投入到短视频内容，并赋予其更高的价值，独立创作者正在演变为现代型工作室，以远低于传统制作的成本大幅提升独立视频的质量。

新的创作者工作室不仅能带来经济优势或利用受众行为的变化，还能快速响应，迅速根据受众反馈进行调整。它们通过利用互动数据来判断哪些内容有效、哪些无效，从而降低内容决策的风险。¹⁹他们通过与观众互动来加强社区联系并扩大粉丝群体。他们在能够缩短上市时间、降低生产成本以及拓展地理覆盖范围的地方均会采用AI技术。²⁰他们可以自由尝试各种编辑策略，以最大程度地提高用户参与度和留存率。

微短剧参与度的提升，可能会催生更多由观众互动驱动并借助技术放大的创作者工作室，使其更快速地发展并触达全球受众。同时，这也将加剧微短剧应用、流媒体视频服务和社交视频平台对顶尖创作者资源的竞争。其中，社交平台可能处于相对劣势，除非它们能更有效地帮助用户发现并持续追更系列剧集。

助力现代短视频内容创作者的工具与策略

创作者和媒体高管应考虑新兴的独立工作室如何从零开始，利用各种工具和平台以最低成本触达并吸引全球受众。

AI赋能的生产流程：影视公司可利用生成式AI工具缩短制作周期，降低实现高制作水准的门槛，例如自动生成辅助镜头或简单动画，以及通过自动化字幕生成、语音克隆或AI配音技术，弥合语言、方言和口音之间的差异。²¹一些能够将长篇故事转化为短篇连载的工具也正在涌现。

编辑策略：这些剧集通过增加“钩子”的密度（例如情节反转和真相揭露），并在每集结尾设置悬念，激发观众“想知道后续”的冲动。成功的微短剧通常借鉴同人小说和网络小说中的桥段，例如贫富恋、穿越复仇等，这些元素已被证实具有广泛吸引力。²²

社群发展：创作者应通过评论与观众互动，甚至可根据粉丝反馈调整后续剧集内容。参与幕后直播及社交媒体讨论，有助于将系列作品发展为活跃的粉丝社群。以稳定频率在固定时间每日更新剧集，有助于培养观众定时观看的习惯。微短剧可发展为具有持续性的迷你连续剧品牌，通过重复出现的角色或主题延伸至多个季度或衍生作品。那些善于跨平台推广、IP衍生品开发以及多渠道分发的创作者，例如推出小说化作品或由AI生成音乐制作的原声带，将能以较低成本持续运营其“迷你漫威”宇宙，并获得竞争优势。

新的成功关键绩效指标：数据驱动的反馈循环可帮助短剧工作室测试多种剧情线，并加大对用户留存率高的剧情的投入。需跟踪完播率、每位用户平均观看的剧集数、系列剧集的订阅转化情况，甚至特定剧情的投资回报率，例如某部剧集是否带动了周边商品销售或提升了平台观看时长。

变现模式：短剧正在探索多种变现途径以支持其发展，包括单集小额付费、月度订阅、原声带销售、周边商品、广告以及产品植入等。²³

对抗大脑退化和沉迷负面信息的解药？

牛津大学评选的2024年度词汇是“brain rot”（大脑退化），该词用于“反映人们对于过度消费低质量网络内容，尤其是在社交媒体上，所带来的影响的担忧”。²⁴一个相关的术语“doomscrolling”（末日刷屏）反映了部分用户在社交媒体上表现出成瘾行为的倾向。²⁵

一些证据表明，越来越多的人正从社交媒体转向规模更小、更私密且受保护的信息、娱乐和社交空间。²⁶也许这是对社交媒体不再“社交”、而仅仅是“媒体”，变得过于碎片化和商品化这一现象的回应。²⁷这可能是由于过度使用对心理健康影响的已知担忧所致。²⁸随着一些合成内容开始侵蚀真相与证据，人们很快可能会因无法信任网络内容和信息而选择离开。又或者，仅仅是因为不断寻觅间歇性奖励的过程本质上就令人疲惫且缺乏满足感。²⁹

微短剧可能并未预示新媒体领域的巨大变革，但人们对更具系列化的短篇独立内容日益增长的兴趣，或将挑战社交视频平台以及传统影视公司和流媒体服务商，改变权力格局，并有望催生一批高质量、低成本的新型独立制作公司。

一些证据表明，越来越多的人正从社交媒体转向规模更小、更为私密且受保护的信息、娱乐和社交来源。

一些证据表明，越来越多的人正从社交媒体转向规模更小、更为私密且受保护的信息、娱乐和社交来源。

小结：创作者和独立工作室在响应用户需求方面正变得越来越得心应手

独立创作者和工作室正日益扩大其影响力，与品牌建立更紧密的联系，并探索触达受众的新渠道。由于在社交平台上面临用户参与度和变现方面的挑战，或出于在自有平台聚合发展的机遇，越来越多的创作者可能转向其他渠道，例如视频流媒体平台和微短剧应用，甚至可能催生由创作者主导的新型娱乐服务。

创作者正在多个平台上聚集受众，并与品牌建立更紧密的关系。³⁰这导致了热门内容创作者与其发布平台之间产生了一些紧张关系，尤其是在利润分成和内容审核方面。³¹

社交媒体平台已从社交关系图谱演变为兴趣图谱，其内容推送不再基于用户主动关注的对象，而是根据用户互动行为提供源源不断的信息流。这种变化可能导致创作者更难触达自己的受众，因为算法可能判定其他内容更能促进用户与平台的互动。创作者可能投入大量时间与资金来培养受众群体和品牌关系，却眼睁睁看着自己的内容被算法降低优先级，甚至被关闭，而这些决策看似反复无常。

在大多数情况下，算法推荐信息流和兴趣图谱并不支持连续性的叙事内容。如果微剧的受欢迎程度持续上升，社交平台可能会调整其趋势算法，以支持持续更新的叙事内容，从而帮助创作者有效触达并留住忠实观众，这可能预示着媒体消费模式的潜在转变。平台或许会引入“系列感知”算法，或设置“继续观看”推荐栏。

否则，可能会有更多创作者转向专注于创作的工作室类应用程序，例如目前在移动应用下载榜上占据主导地位的微短剧应用。他们也可能被流媒体视频平台挖角，因为这些平台正寻求更多短视频内容来充实节目单，并吸引年轻观众。事实上，流媒体服务或许已处于有利地位，因其服务建立在系列化和定时更新内容的基础之上。如果越来越多的创作工作室采用系列化的短视频制作模式，流媒体平台可能将成为主要受益者。又或者，现在正是由创作者主导的新平台涌现的恰当时机，这些平台可依托流媒体和社交媒体的技术积淀与经验而建立。

Chris Arkenberg

United States

Tim Bottke

Germany

Ankit Dhameja

India

Gillian Crossan

Global

ENDNOTES

1. Rui Ma, “[State of short drama apps 2025](#),” Mobile App Insights, July 2025.
2. Stephanie Yang, “[Two-minute TV shows have taken over China. Can they take over the world?](#)” Los Angeles Times, March 16, 2025.
3. Ma, “[State of short drama apps 2025](#).”
4. Gaby Hinsliff, “[It’s the age of regret: Gen Z grew up glued to their screens, and missed the joy of being human](#),” The Guardian, March 7, 2025.
5. Xinhuanet, “[Love, twist and one-minute cliffhangers: China’s micro dramas go global](#),” July 29, 2025.
6. CMB Global Markets’ equity research, May 23, 2025 (private report accessed via AlphaSense).
7. Mandy Zuo, “[China’s addictive micro-dramas show how commercial demand is fuelling a netcasting boom](#),” South China Morning Post, March 27, 2025.
8. Jeff Huang, “[How China’s \\$7 billion micro drama industry is taking on the US entertainment industry](#),” CNBC, July 22, 2025.
9. Kunal Purandare, “[The VC-backed rise of micro dramas in India](#),” Forbes India, Aug. 8, 2025.
10. Systematix Institutional Research and Morning Brew, July 10, 2025 (private report sourced via AlphaSense).
11. The Economic Times, “[Stage set for micro-dramas; WhatsApp’s monetisation bid](#),” June 17, 2025.
12. Robert Steiner, “[Microdrama plot twist: A threat to the apps’ stratospheric US growth](#),” Variety, April 30, 2025.
13. Ma, “[State of short drama apps 2025](#).”
14. Ibid.
15. Appfigures, “[Top ranked iOS app store apps](#),” accessed Oct. 23, 2025.
16. Steiner, “[Microdrama plot twist: A threat to the apps’ stratospheric US growth](#).”
17. Paige Gawley, “[People on TikTok are obsessed with a fake group chat](#),” Vice, April 9, 2025.
18. Lauren Forristal, “[Netflix is getting into short videos with a new vertical feed for mobile](#),” TechCrunch, May 7, 2025.
19. Global economic outlook and investment strategy, 2H 2025; ICBC International Research (private research brief via AlphaSense).
20. Carson Taylor, “[Microdramas: China’s new craze goes global](#),” Naavik, Sept. 8, 2024.
21. Focus on structurally high-growth segments, Huatai Securities, Aug. 23, 2025 (private documents sourced via AlphaSense).
22. Kristian Monroe, “[Told one minute at a time, micro dramas are soap operas designed to fit in your hand](#),” NPR, March 19, 2025.
23. Taylor, “[Microdramas: China’s new craze goes global](#).”
24. Oxford University Press, “[‘Brain rot’ named Oxford word of the year 2024](#),” Dec. 2, 2024.
25. Sian Boyle, “[Is doom scrolling really rotting our brains? The evidence is getting harder to ignore](#),” The Guardian, Dec. 9, 2024.
26. Annalee Newitz, “[Social media is dead – here’s what comes next](#),” NewScientist, July 23, 2025.
27. Rodney Mason, “[Social isn’t social anymore—now what?](#)” Forbes, April 28, 2025.
28. Jessica A. Kent, “[Need a break from social media? Here’s why you should—and how to do it](#),” Harvard Summer School, Aug. 28, 2023.

29. Sanzana Karim Lora, Sadia Afrin Purba, Bushra Hossain, Tanjina Oriana, Ashek Seum, and Sadia Sharmin, "Infinite scrolling, finite satisfaction: Exploring user behavior and satisfaction on social media in Bangladesh," Arxiv, April 15, 2025.
 30. Deloitte Digital, "2025 state of social research: How efficiency can meet impact with the right investments," May 15, 2025.
 31. Gillian Follett, "How creators are shaping Cannes Lions—from business discussions to the campaigns winning awards," AdAge, June 12, 2025.
-

ACKNOWLEDGMENTS

The authors would like to thank **Jeff Loucks**, **Brooke Auxier**, and **Daniela Gonzales** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

视频播客称霸：品牌迎机遇，传统视频陷竞争

播客正逐渐转变为以视频为主、多语言的媒介，其覆盖范围迅速扩大，有助于品牌触达全球受众，同时占据观众更多的屏幕使用时间。

播客不再只是用来听的了：现在你还可以观看它们。视频播客（vodcasts）通过将音频叙事与视觉吸引力相结合，正在重新定义受众消费长篇内容的方式，使内容更具沉浸感和可分享性。随着视频播客逐渐模糊了传统播客、社交媒体和流媒体视频之间的界限，内容创作者正利用跨平台分发来提升用户参与度、构建社群、扩大广告收入，并开拓新的赞助机会。如今，视频播客或许正开始抢占原本由传统电视和流媒体平台垄断的屏幕使用时间。

与此同时，印度、尼日利亚和巴西等新兴市场的听众正因播客移动优先、低带宽需求的特点而逐渐接受这一媒介。本地化、多语言内容的兴起也在推动这一增长，使播客成为真正全球化且具有文化多样性的媒介，尽管在变现、语言可及性和基础设施差距方面仍面临实际挑战。德勤预测，到2026年，播客和视频播客的全球年度广告收入将达到约50亿美元，实现年均近20%的增长。¹

将视频播客的日益流行与播客的全球扩张相结合——我们看到的是一个在受众规模、覆盖范围和广告收入方面均有望实现显著增长的市场。不过，该行业未来的发展路径可能取决于创作者与平台能否有效应对内容可发现性、商业变现及规模化发展等挑战。

视频播客通过创新形式吸引受众

这些以视觉为核心的播客将在2026年及以后持续获得消费者和广告商的青睐。推动这一增长的因素可能有三个方面：它们能够无缝融入现有且受欢迎的媒体平台，利用社交媒体短视频制造话题并实现广泛传播，以及增强视频播客创作者与其忠实受众之间的深度连接。

我们预测，带有视频的热门播客占比将上升，消费者将愈发倾向于选择那些支持视频的平台。

近年来，包括Spotify、Wondery、Podbean和YouTube在内的一些流媒体音乐与音频服务平台，²已将视频播客（vodcasts）直接整合至用户界面，让消费者得以便捷访问这类内容。与此同时，部分头部服务商还为播客创作者配备了创作及变现视频内容所需的工具与经验，进一步丰富了视频播客的供给。³尽管其他支持视频播客的平台中不乏行业长期玩家，但YouTube作为2022年的新入局者，⁴已迅速产生影响：2025年初，其月度视频播客观众规模达10亿；推出了针对美国市场的顶级播客榜单；⁵2025年8月，旗下《New Heights》的一期视频播客直播更创下吉尼斯世界纪录——吸引了130万并发观众。⁶将这些视频播客置于用户本就投入时间与订阅费用的知名平台中，既降低了普通消费者的尝试门槛，也提升了他们对内容及广告的接受度与参与度。不同播客平台对视频与纯音频格式采取了差异化策略：有的仅保留纯音频（或近乎纯音频），有的则为所有播客内容配套视频版本。Spotify虽于2022年才开始推出视频版，但截至2025年9月中旬，其最受欢迎的节目中已有超60%包含视频组件。⁷我们预计，含视频的热门播客占比将持续攀升，消费者也会愈发倾向于选择拥抱视频的平台。另一方面，2025年秋季《数字媒体趋势》报告显示，27%的美国消费者每周都会观看视频播客，这一趋势主要由Z世代与千禧一代引领。⁸

视频播客的视频组件让观众以单纯音频无法实现的方式参与到对话中。观众可以看到主持人的面部表情、肢体语言以及环境的视觉背景，从而产生一种亲近感。当观众看到他们熟悉并喜爱的播客主持人时，这种准社会关系会得到加强，同时提升信任感和真实感，进而推动社区的形成和用户参与度的提升。⁹所有这些互动都会产生累积效应：观看视频播客的用户比仅收听音频播客的用户多消费1.5倍的内容。¹⁰视频还增加了引人入胜的视觉叙事层次，能够吸引年轻的数字原生受众，并帮助创作者跨平台触达新的观众。视觉形式也带来了更多的广告和赞助机会，因为它支持品牌标识和产品植入，并可制作出适合社交媒体分享的短视频片段。¹¹

因此，社交媒体平台对于视频播客的成功也至关重要。¹²短视频播客片段可被重新利用并分享至各大社交媒体平台，以触达不同受众，并突出单集中最吸引人、最具话题性的内容。通过社交媒体平台及具有传播效应的短视频片段，能够扩大播客的覆盖范围，提升内容的可发现性，拓展广告投放空间，同时为播客创作者提供与听众直接互动的机会。

视频播客的兴起可能会持续对其他视频娱乐内容和平台造成压力，因为这些视频资源正在争夺宝贵且有限的屏幕和电视观看时间。早在2024年，近一半的播客观众表示他们会在联网电视上观看节目。¹³这种竞争或许正促使一些传统流媒体视频提供商考虑进军播客领域。¹⁴随着视频播客的普及并逐渐进入客厅，这种形式正在以值得关注的方式改变着消费者的行为：音频播客可以在进行其他活动（如通勤或锻炼）时收听，而视频播客则需要更专注的观看。44%的美国视频播客观众表示他们在观看时从不分心做其他事情，相比之下，只有29%的音频播客听众表示他们在收听时从不分心。¹⁵

这种对内容更加集中的关注，可能会带来更高的用户参与度和订阅用户增长——从而吸引更多广告商和赞助商的关注与投资。根据最新的2025年秋季《数字媒体趋势》数据，目前约有四分之一的美国播客观众和听众（以及超过三分之一的Z世代和千禧一代播客观众和听众）表示，他们经常购买在播客中听到广告的产品或服务。¹⁶更多的广告商和赞助商意味着收入增加，这将推动影响力的扩大、业务增长以及创新发展。简而言之，视频播客的兴起很可能将定义播客行业演进的下一个阶段。

本地化、多语言播客内容：即将登陆您身边的市场

最初主要以美国为中心的音频格式，正在迅速发展成为一种充满活力的全球性媒介。¹⁷尽管各地区每周收听播客的人数比例差异较大，但全球平均水平约为22%，其中印度尼西亚（42.6%）和墨西哥（41.8%）在听众比例方面处于领先地位。¹⁸推动新兴市场播客收听量激增的因素有多个：移动网络连接不断扩展、流媒体音频平台的全球投资持续增加，以及本地化和多语言播客内容与视频播客的日益普及。

老牌音频流媒体平台正在全球范围内扩张，并通过投资本土语言内容来推动播客发展。

全球移动互联网接入的扩展使连接和内容消费变得更加普及。在印度、尼日利亚和巴西等国家，价格实惠的智能手机和数据套餐已使数百万人接入网络。¹⁹例如，过去十年中，尼日利亚的移动数据费用下降了约97%：2014年1GB移动数据费用为11.15美元，到2023年已降至0.39美元。²⁰低成本设备和资费方案的普及，使得按需音频和视频内容（如播客和视频播客）能够被更多地区的人们所获取。

主流音频流媒体平台正通过投资本地语言内容，向全球扩张并推动播客增长。例如，Spotify正在拉丁美洲、非洲和亚洲资助内容创作者，并建立独家合作关系，以打造符合当地需求的节目。²¹其他平台正在授权热门本地播客节目、制作原创内容，并开发工具以支持本地人才发展。²²尽管整个行业仍以英语为中心，但人们越来越意识到，多语言和具有文化针对性的节目内容可能是维持全球播客持续增长的关键。与此同时，在黎巴嫩、印度和尼日利亚等国²³，新的播客平台正逐渐成为本土内容的枢纽，并与全球巨头建立合作关系以扩大其影响力。

新的内容形式——如视频播客——也在推动用户参与度的提升，因为它们吸引了新兴市场中习惯数字环境的年轻受众，这些市场的年龄结构通常比更发达经济体更为年轻。²⁴

播客的全球崛起对盈利策略产生了深远影响——平台正着力推动内容发现与消费，尤其是非英语内容的传播。²⁵越来越多跨国品牌倾向于在地区性节目中投放广告，或是结合贴合文化背景的故事化内容，触达多元且活跃的受众——尽管新兴市场的每千次展示成本（CPMs）仍偏低，令创作者面临创收挑战。虽有阻碍，市场全球化仍在助推本地化广告、品牌内容及跨境创作者合作的激增。²⁶

小结：未被开发的全球受众和新的增长机遇

视频播客与播客正渗透进客厅场景，并向全球新兴市场扩张——这给媒体与娱乐行业的多个参与者带来了机遇与挑战。

这类平台可能需要聚焦技术能力建设或升级，以实现视频播客在APP内的无缝流传输——但这需要投入基础设施、技术系统与人力成本。对于已具备相关能力的平台，下一步可探索在视频播客广告中加入动态可购物元素（例如直接点击视频广告即可购物或下单），这是与广告商锁定丰厚合作及赞助、推动行业变现的关键。

成功的关键在于：能否本地化内容、与多元受众建立信任，以及应对日益复杂的竞争格局。

同时，这些平台应积极拓展新兴市场，策略可能包括：投资本地/区域内容与创作者IP，以及利用生成式AI实现音视频内容的自动翻译与唇形同步²⁷——以此吸引新市场用户、扩大受众基数。尽管前景向好，但全球化布局需要对区域偏好、监管环境与变现模式有细致入微的理解。成功的关键在于：能否本地化内容、与多元受众建立信任，以及应对日益复杂的竞争格局。此外，流媒体音频服务商还应探索提供内容离线访问功能、文件压缩技术及低比特率流媒体模式，尤其是在带宽成本仍较高、网络不稳定的地区。

流媒体视频点播（SVOD）提供商或许也能抓住视频播客热潮的独特机遇——最显著的是通过推出伴随式视频播客，在季与季之间保持观众参与度，同时扩充内容清单。²⁸流媒体平台也可将播客视为新故事与新人才的低成本孵化器，通过“播客到屏幕”的转化漏斗挖掘现有粉丝基础并降低开发风险。²⁹同样，与已拥有忠实粉丝群、懂得如何引发社交热议的成熟创作者合作，可快速实现文化相关性。这些策略并非追逐视频播客的热度——它们将“音频优先”的故事讲述转化为驱动用户留存、深化粉丝参与及实现可持续长期增长的战略引擎。

Brooke Auxier

United States

Gillian Crossan

Global

Duncan Stewart

Canada

Akash Rawat

India

Tim Bottke

Germany

Wenny Katzenstein

United States

ENDNOTES

1. Based on Deloitte analysis; Brooke Auxier, Bree Matheson, Duncan Stewart & Kevin Westcott, "[Shuffle, subscribe, stream: Consumer audio market is expected to amass listeners in 2024, but revenues could remain modest](#)," Deloitte Insights, Nov. 29, 2023.
2. Spotify Newsroom, "[Spotify unveils uninterrupted video podcasts, audience-driven payments, and the new Spotify for Creators platform](#), Nov. 13, 2024; Wondery, "[Now playing: Video podcasts on the Wondery app for Wondery+ subscribers](#)," accessed Oct. 23, 2025; Angela Yang, "[Podcasts are taking over TV screens as video formats grow increasingly popular](#)," NBC News, Dec. 23, 2024.
3. Spotify Newsroom, "[From audio to video, Spotify's \\$100 million payout fuels creator success stories](#)," Apr. 28, 2025.
4. Ariel Shapiro, "[YouTube launches a dedicated page for podcasts](#)," The Verge, Aug. 23, 2022.
5. Todd Spangler, "[YouTube says it now has more than 1 billion monthly viewers of podcast content](#)," Variety, Feb. 26, 2025; Zach Vallese, "[YouTube launches weekly top podcast list to rival Spotify and Apple](#)," CNBC, May 15, 2025.
6. Alex Schiffer, "[Taylor Swift draws 1.3 million live viewers in 'New Heights' appearance](#)," Front Office Sports, Aug. 13, 2025; Vicki Newman, "[Taylor Swift earns podcast record with appearance on boyfriend Travis Kelce's New Heights, Guinness World Records](#)," Aug. 26, 2025.
7. Based on Deloitte analysis of publicly available data.
8. Data from Deloitte's Fall 2025 Digital Media Trends 19 survey.
9. Edison Research, "[YouTube is the preferred podcast listening service](#)," Oct. 23, 2024.
10. Ellie Hammonds, "[Vodcasts: Is it the future of podcasting?](#)" The Media Leader, Aug. 28, 2025.
11. Molly Fuard, "[Visual podcasting is now a thing and here's what advertisers should know](#)," Adweek, accessed Oct. 23, 2025.
12. Lloyd George, "[Why social media is a game-changer for growing your podcast](#)," Acast, accessed Oct. 23, 2025.
13. Alexander Lee, "[Podcast consumption shifts towards connected TVs](#)," Digiday, May 7, 2025.
14. Eve Upton-Clark, "[Netflix is eyeing video podcasts as it expands beyond TV and film](#)," Fast Company, April 21, 2025.
15. Data from Deloitte's Fall 2025 Digital Media Trends 19 survey.
16. Ibid.
17. Sara Fischer, "[Axios media trends](#)," Axios, April 22, 2025.
18. Simon Kemp, "[Digital 2025: The essential guide to the global state of digital](#)," Meltwater, Feb. 5, 2025.
19. Global System for Mobile Communications Association, "[The mobile economy 2025](#)," accessed Oct. 23, 2025.
20. Paula Gilbert, "[Nigeria's 1GB data price has dropped 75% over five years](#)," Connecting Africa, June 5, 2020; Peter Oluka, "[\\$0.39 \[604 NGN\] per 1GB: Nigeria among countries with cheapest data rates](#)," Tech Economy, Jan. 10, 2025; Bruno Venditti, "[The cost of 1GB of mobile data worldwide](#)," Visual Capitalist, Oct. 21, 2024.
21. Spotify Newsroom, "[Get to know the 13 podcast grantees of Spotify's new Africa podcast fund](#)," Oct. 24, 2022; Blueprint Magazine, "[Spotify and the pod network enters a new era of Filipino podcasting with the launch of their state-of-the-art studio](#)," April 28, 2025.

22. Spotify Newsroom, "The Spotify partner program expands to nine new markets, giving more creators new ways to monetize their content," March 27, 2025.
 23. IndustryPods, "Podcast distribution on international platforms," December 2024; The Storiez, "How Anghami is dominating the music streaming market globally," Sept. 14, 2024; Peerzada Abrar, "Kuku FM raises \$25 mn from investors; aims to expand content, improve tech," Business Standard, Sept. 20, 2023; Samuel Viavonu, "The podcast boom in Nigeria: an era of noise or knowledge?" Afrocritik, Feb. 19, 2025.
 24. Acast, "The video podcast opportunity," June 10, 2025; Devan Kaloo and Robert Gilhooly, "Demystifying emerging markets," Aberdeen Investments, Sept. 8, 2023.
 25. BeMultilingual, "What are the most popular languages on YouTube?" July 26, 2025; David R. Gonzalez, "The state of podcasting in Latin America," PodNews, Feb. 15, 2024.
 26. Aaron Chow, "Nike Japan launches 'NIKELAB RADIO,'" HypeBeast, July 28, 2021.
 27. Burt Helm, "How AI for lip dubbing could change the film industry," Fast Company, November 2023.
 28. The New York Times Style Magazine: Australia, "Forensic fandom and the age of the companion podcast," Feb. 27, 2025.
 29. Damion Taylor, "How podcasts are becoming Hollywood's new development pipeline," Forbes, Jan. 30, 2025.
-

ACKNOWLEDGMENTS

The authors would like to thank **Matt Varraveto**, **Abel Sun**, and **Kenny Gold** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

自主新纪元：引领技术主权征程

各国和地区集团正竞相建设自身的主权技术和AI基础设施。这将带来哪些影响，全球企业又该如何应对？

随着全球地缘政治环境日益复杂和不确定，企业和政策制定者正敦促各自国家和地区加强对数字基础设施的直接控制，尤其是与AI相关的部分。Gartner®预测，“到2028年，全球65%的政府将提出某种技术主权要求，以提升独立性并防范域外监管干预”。¹

技术主权基于国家和地区联盟独立开发、控制、监管及资助数字技术的能力，这些技术包括云计算、量子计算、人工智能、半导体和数字通信基础设施。²它可能包含关于数据流动、物理设施所在地、由谁拥有、由谁治理、由谁运营，以及由谁提供支撑这些技术的硬件、软件和服务等方面的具体地理、法律和监管要求。

对主权的追求并非新现象，但向技术主权的转变可能会在2026年进一步加速。在未来的十年中，大量投资将涌入云计算、半导体、数据中心、AI模型、连接技术以及卫星通信等领域。在相互关联的世界中，任何国家或地区都难以实现完全的主权，但许多国家和地区正致力于至少实现更大的自主性。

由于AI被普遍视为推动经济发展和国家竞争力的下一个主要动力，其生态系统目前正受到广泛关注。这种紧迫性尤为明显，因为计算能力（也称为“算力”）等先进的AI能力目前由极少数国家和公司掌控。

牛津互联网研究所的研究发现：“只有34个国家拥有公共AI算力；其中仅有24个国家能够获得用于模型训练级别的算力；而大多数国家依赖由少数外国实体控制的云计算或芯片基础设施。”³同一项研究发现，全球90%的AI计算由美国和中国公司管理。⁴

德勤预测，到2026年，更多国家将具备获取AI算力的能力，主权AI算力建设投入将突破1,000亿美元。到2030年，由美国和中国以外公司管理的AI算力在全球总容量中的占比，预计将从当前的10%实现翻倍。作为AI与加速计算平台领军企业，英伟达预计其2025年在主权数据中心市场的AI芯片销售额将达200亿美元，同比增长100%。这正是上述AI算力格局演变趋势的典型印证。⁵

欧洲正在引领这一进程

2024年9月，欧盟发布《德拉吉报告》，就提升欧洲整体经济竞争力提出建议。⁶该报告部分内容聚焦如何推动本土科技产业发展，涉及增强创新能力、加速技术应用、提升劳动者生产率等关键议题。在此报告基础上，逾200家欧洲企业与机构联合发起“欧洲科技栈倡议”（EuroStack Initiative），呼吁围绕技术主权采取“彻底行动”，⁷具体包括：优先采购欧洲技术、优化现有资源整合与盘活、推动研发成果向产品化转化、保障充足资金供给以及保护欧洲云用户数据。欧盟委员会已任命专职的技术主权事务委员统筹相关工作。欧盟追求技术主权的努力由来已久，基于《通用数据保护条例》（GDPR）、《数字服务法案》（DSA）和《人工智能法案》等法规体现的价值观与原则，其坚信主权技术方案最能契合欧盟理念。

尽管2025年初的狂热预期近期随《欧盟国际数字战略》的发布有所缓和——该战略更侧重在人工智能、半导体、量子计算及网络安全领域开展国际合作——但关于最佳战略路径的辩论仍在持续。⁸未来五年内，欧洲预计将在云计算、AI数据中心、AI企业、半导体及卫星通信领域投入逾1000亿欧元的公私联合资金。

云计算

本地欧洲云服务供应商仅占整体市场很小的份额（不到20%）。⁹它们需要投入大量资金和时间，才能真正成为全球超大规模云服务商的竞争对手。更可能发生的情况是，全球主要厂商将越来越多地提供针对欧洲市场的特定适配能力。亚马逊云服务（AWS）已宣布，将投资近80亿欧元在德国建立欧洲主权云（European Sovereign Cloud）。该项目的目标包括使客户能够将数据保留在欧盟境内，实现自主可控，并由欧盟公民主导、运营、保障安全和进行治理。¹⁰微软还宣布了一系列针对欧洲的承诺——特别是在AI、网络安全、隐私保护、韧性建设以及经济竞争力方面——并推出了微软主权云平台及相关解决方案。¹¹

AI模型与数据中心

政府和商业部门均有多个旨在提升整体AI能力的举措。欧盟委员会的“AI大陆行动计划”（AI Continent Action Plan）旨在利用现有的超级计算基础设施，在欧洲各地建立一系列AI工厂和超级工厂，并通过InvestAI计划推动全新的投资。¹²该计划将为最多五个新的AI超级工厂提供200亿欧元的资金支持，以推动创建被称为“主权前沿模型”（Sovereign frontier models）的尖端AI模型。该行动计划还旨在提升AI模型的数据可用性、促进AI应用的使用，并加强技能和劳动力发展。在商业方面，NVIDIA与Perplexity正携手合作，共同训练并广泛推广开源的本地化AI模型。¹³NVIDIA与投资公司MGX、Mistral AI及其他合作伙伴共同支持，计划到2028年耗资85亿欧元建成欧洲最大的AI数据中心。¹⁴此外还有“星际之门英国”（Stargate UK）项目，这是一项分阶段推进的计划，旨在在全国范围内建设AI基础设施，并加速国内AI技术的应用。¹⁵

半导体领域

与美国类似，欧洲正着力推动半导体制造本土化、增强供应链韧性、完善本土产业生态并培育欧洲企业。为此，《2023欧盟芯片法案》（EU Chips Act (2023)）设立了专项基金，建设试验性试点产线，打造协同设计平台与能力中心，并为量子芯片研发提供资源——预计到2030年总投资规模达430亿欧元。¹⁶目前已有多项重要的商业投资正在进行，包括专注于FinFET（鳍式场效应晶体管）的纯代工厂、“智能功率晶圆厂”以及碳化硅芯片制造厂等。¹⁷

卫星通信领域

欧洲另一关键举措是构建自主卫星通信星座以降低对域外服务商的依赖，确保军事、政府及商业应用获得安全可靠的服务。目前主要的两项计划包括“卫星互联与安全”（IRIS²）星座和Eutelsat OneWeb。IRIS²由近300颗分布在多个轨道上的卫星组成，耗资约110亿欧元。¹⁸ Eutelsat 正寻求加快建设与增强 OneWeb 低地球轨道（LEO）卫星互联网星座的步伐，目前该星座已有超过 630 颗在轨卫星。¹⁹他们最近获得了英国和法国政府的新一轮投资以加速建设。²⁰在竞争激烈的市场环境中，这两大星座要形成有效竞争力并完全满足欧洲需求仍需时日（IRIS²计划于2031年完成）及持续巨额投入。²¹

世界其他地区呢？

尽管欧洲正在大力推进大量技术主权相关活动，但其他国家和地区也在探索各自独特且创新的路径——其中大多数努力集中在AI领域。这并非要面面俱到，而是为了展示全球相关活动的广泛性与深入程度：

- 韩国：韩国希望基于本国语言并结合其文化，发展自主的AI能力。²²例如，Kakao与OpenAI合作推出新的个性化数字服务。²³为加强国内基础设施建设，SK集团与AWS宣布将联合投资约50亿美元，于2029年前建成韩国最大的AI数据中心。²⁴
- 日本：该国正致力于通过Rapidus计划（一家专注于2纳米技术的新公司）以及到2030年投入650亿美元的政府投资计划，提升其AI能力并重塑国内半导体产业。²⁵
- 非洲：非洲首个由NVIDIA AI和加速计算平台技术支持的AI工厂将落户南非Cassava Technologies的数据中心，并计划扩展至非洲其他地区，包括埃及、肯尼亚、摩洛哥和尼日利亚。²⁶
- 印度：印度在技术栈的各个层级都大力推动自力更生，政府项目如印度半导体任务（India Semiconductor Mission）和印度AI计划（IndiaAI）正致力于满足这些需求。²⁷印度在开发其AI模型时正面临一些独特的挑战，包括计算资源的可获得性、需要支持多种语言，以及缺乏高质量的训练数据。²⁸印度为“印度技术栈”（India Stack）所开发的强大国内数字能力，可以进一步扩展并输出到其他国家，以打造一个全新的竞争性数字生态系统。²⁹
- 加拿大：加拿大政府的主权AI计算战略致力于提升私人投资、公共基础设施以及获取计算资源的资金支持。³⁰它还宣布与国内AI公司Cohere建立合作伙伴关系，以探索双方如何共同提升加拿大的整体技术能力。³¹此外，多家加拿大电信公司正计划建设主权AI数据中心，包括TELUS、SaskTel和Bell。³²

- 中东：该地区许多国家正在加大在主权云和AI数据中心方面的投资，包括诸如阿联酋Stargate项目这样的重大工程，该项目是一个1吉瓦的AI集群。³³沙特公共投资基金于2025年成立了HUMAIN公司，该公司致力于在全球合作伙伴（如AWS、NVIDIA等）的支持下，开发端到端的AI基础设施。³⁴目前宣布的合作伙伴相关投资已达230亿美元。³⁵
- 中国：未来中国在技术主权领域的发展将呈现多维度、系统化的趋势。首先，核心技术自主化将全面加速，聚焦半导体、人工智能、基础软件等环节，通过国家科研专项与市场机制结合，突破关键瓶颈。其次，数字主权体系将进一步深化，藉由数据本地化储存、网络安全审查及自主协议标准，构筑安全可控的数字生态。同时，中国坚持开放创新与自主可控并行的策略，在参与全球科技合作的基础上，强化本土产业链的韧性与竞争力。此外，中国将积极争夺新兴技术规则主导权，在6G通讯、量子科技等前沿领域推动标准制定，提升国际话语权。为应对外部不确定性，供应链韧性强化成为重点，通过国内替代、国际多元布局及战略储备，降低关键技术断供风险。最后，科技治理与伦理框架将同步完善，建立符合国家发展需求的监管体系，平衡技术创新与安全伦理，形成可持续的技术主权路径。

可能带来的影响

如果如预期那样，大多数政府在不久的将来推行强有力的技术主权政策和计划，会发生什么？加强对端到端技术能力的控制将带来诸多潜在益处。这些益处包括经济方面的收益，例如增加税收收入和私人资本投资、为公民创造更好的就业机会，以及为本土科技企业的发展提供更大空间。通过提高自给自足能力，可以提升整体韧性，加强隐私与安全，并减少受外部潜在政治干扰的风险。此外，在AI领域，如果基础模型由本国研发，将能更好地反映本地语言、习俗和数据集。

我们也可能看到一些挑战的出现，例如：

- 投资流向的转变：外国直接投资（FDI）、并购和合资企业可能会面临越来越多的限制和要求。风险投资（VC）也可能发生调整。风险投资公司是否会将国家战略性利益置于更具全球性的机会之上？³⁶
- 碎片化加剧：采取更加封闭、零和的方式可能导致合作水平下降、国际关系破裂以及学术合作减少。还可能导致跨境数据流动减少、专有通信基础设施增多，以及标准和法规数量的增加。
- 劳动力影响：随着各国更加重视本国的技术能力，高技能人才的全球流动格局可能发生改变。这种变化在AI、网络安全和芯片设计等关键领域尤为明显。同时，各国也可能会加大投资，以全面提升本国整体工作能力。³⁷
- 环境影响：晶圆厂、实验室、数据中心及其相关配套基础设施的大规模建设将对资源造成压力。一些国家的电网已接近满负荷运行，而新建数据中心需要额外数千兆瓦的电力，可能与居民用电需求产生竞争。³⁸此外，还面临着利用无污染、低碳或零碳电力来源的挑战。³⁹
- 新的合作伙伴关系：并非所有人都能独自成功。未来，我们可能会看到更多的双边协议、区域合作框架以及非传统的技术联盟，旨在利用彼此的优势实现共赢。⁴⁰
- 产能过剩：市场能够支撑多少基础AI模型？目前，全球在AI基础设施方面的资金支出巨大，预计到2028年将接近3万亿美元，但所有这些投入最终都能获得投资回报吗？⁴¹长期需求可能难以匹配异常高的预期，而新的技术革新或将使现有方案面临淘汰风险。⁴²

小结：为更加自给自足的未来做好准备

到2026年，对技术主权的追求将持续推进，引发更多讨论、政府行动和投资活动。尽管这一趋势背后的动机及其最终结果尚存争议，但相关行动已经展开，并且还将继续推进。许多人认为，本国及所在区域集团未来的繁荣面临重大挑战。

- 审查全球依赖关系：识别并评估所有关键依赖项——数据流、公共云、供应商、供应链、财务和监管。建立新的合作并加强现有伙伴关系，以提供最大程度的全球灵活性。确保能够清晰阐释企业全球运营情况。
- 预判监管复杂性：为快速变化的监管环境做好准备。预计在数据本地化、网络安全、并购以及资本流动方面将出台新规定。明确企业业务面临风险最高的领域，并立即制定应对不同情景的预案。加强合规体系建设。
- 重新审视云战略：思考公有云与私有云能力之间的平衡。采用多云或主权云模式，以增强弹性和合规性。优先考虑跨环境的可移植性、互操作性和控制能力。确保供应商支持数据存储、处理和传输的自动合规。制定应急预案，以在地缘政治变化面前保持敏捷性。
- 增强人才韧性：需明确企业关键人才的来源渠道，并预判供应中断可能造成的影响。应制定多元化的人才获取策略，同时借助政府劳动力计划及高校合作项目来培育专业技能储备。

David Jarvis

United States

United Kingdom

Girija Krishnamurthy

Global

Duncan Stewart

Canada

Gillian Crossan

Global

Roger Chung

China

Nick Seeber

Tim Bottke

Germany

ENDNOTES

1. Gartner, "Gartner reveals top technologies shaping government AI adoption," press release, Sept. 9, 2025; Gartner is a registered trademark and service mark of Gartner Inc. and its affiliates in the United States and internationally and is used herein with permission. All rights reserved.
2. Sean Fleming, "What is digital sovereignty and how are countries approaching it?" World Economic Forum, Jan. 10, 2025.
3. Zoe Hawkins, Vili Lehdonvirta, and Boxi Wu, "AI compute sovereignty: Infrastructure control across territories, cloud providers, and accelerators," SSRN, June 24, 2025.
4. Adam Satariano and Paul Mozur, "AI computing power is splitting the world into haves and have-nots," The New York Times, June 21, 2025.
5. Yahoo Finance, "NVIDIA Corporation (NVDA) Q2 FY2026 earnings call transcript," Aug. 27, 2025.
6. Mario Draghi, "The Draghi report on EU competitiveness," European Commission, Sept. 9, 2024.
7. EuroStack, "Building Europe's digital future," accessed Oct. 30, 2025; EuroStack, "Open letter: European industry calls for strong commitment to sovereign digital infrastructure," March 14, 2025; Natasha Lomas, "European tech industry coalition calls for 'radical action' on digital sovereignty—starting with buying local," TechCrunch, March 16, 2025.
8. European Commission, "The international digital strategy for the European Union," July 8, 2025.
9. Diana Goovaerts, "Europe's cloud market poised for 24% growth," Fierce Network, July 28, 2025.
10. Amazon, "AWS plans to invest €7.8 billion into the AWS European Sovereign Cloud," May 15, 2024; Amazon, "Built, operated, controlled, and secured in Europe: AWS unveils new sovereign controls and governance structure for the AWS European Sovereign Cloud," June 3, 2025.
11. Brad Smith, "Microsoft announces new European digital commitments," Microsoft, April 30, 2025; Judson Althoff, "Announcing comprehensive sovereign solutions empowering European organizations," Microsoft, June 16, 2025.
12. European Commission, "Commission sets course for Europe's AI leadership with an ambitious AI Continent Action Plan," press release, April 9, 2025.
13. Belle Lin, "Nvidia and Perplexity team up in European AI push," The Wall Street Journal, June 11, 2025.
14. Amiya Johar, "Nvidia, MGX lead €8.5B project to build French AI data center," Mobile World Live, May 20, 2025.
15. OpenAI, "Introducing Stargate UK," Sept. 16, 2025; Tom Bristow, "US tech firms pour £30B into UK as Trump lands," Politico, Sept. 16, 2025.
16. European Commission, "European Chips Act: The Chips for Europe Initiative," Nov. 4, 2024; European Commission, "European Chips Act," accessed Oct. 30, 2025.
17. Jingyue Hsiao, "TSMC breaks ground on EUR10 billion semiconductor fab in Dresden," Digitimes Asia, Aug. 21, 2024; Infineon, "German government issues final funding approval for new Infineon fab in Dresden," press release, May 8, 2025; Adrià Calatayud and Mauro Orru, "Apple supplier STMicroelectronics to build \$5.4 billion chip plant in Italy," The Wall Street Journal, May 31, 2024.
18. Jeff Foust, "Europe signs contracts for IRIS² constellation," SpaceNews, Dec. 16, 2024.
19. Eutelsat, "High-speed, low-latency connectivity," accessed Oct. 30, 2025.
20. Jason Rainbow, "French government to lead Eutelsat's \$1.56 billion capital boost," SpaceNews, June 19, 2025; Rachel Jewett, "UK to join Eutelsat's capital raise with \$105M investment," Via Satellite, July 10, 2025.

21. Margherita Stancati, Matthew Dalton, and Vera Bergengruen, "Europe scrambles to break its dependence on Musk's satellites," The Wall Street Journal, April 13, 2025.
22. Byun Hee-won and Kim Mi-geon, "South Korea to pour \$735 bn into developing sovereign AI built on Korean language and data," The Chosun Daily, June 17, 2025.
23. Zinnia Lee, "Korea's Kakao teams up with OpenAI to develop AI products," Forbes, Feb. 4, 2025.
24. Zinnia Lee, "Billionaire Chey's SK Group partners with Amazon to build a \$5 billion AI data center in Korea," Forbes, June 23, 2025.
25. Dylan Butts, "Japan is ramping up efforts to revive its once dominant chip industry," CNBC, Nov. 13, 2024; Rapidus, "Rapidus Corporation," accessed Oct. 30, 2025.
26. Cassava Technologies, "Cassava to upgrade its data centres with NVIDIA supercomputers to drive Africa's AI future," accessed Oct. 30, 2025; Nell Lewis, "Africa's first 'AI factory' could be a breakthrough for the continent," CNN, April 3, 2025.
27. INDIAai | Pillars; Government of India, "India semiconductor mission," accessed Oct. 30, 2025.
28. Shadma Shaikh, "Inside India's scramble for AI independence," MIT Technology Review, July 4, 2025.
29. India Stack, "India Stack," accessed Oct. 30, 2025.
30. Government of Canada, "Canadian sovereign AI compute strategy," Oct. 1, 2025.
31. Government of Canada, "Canada partners with Cohere to accelerate world-leading artificial intelligence," press release, Aug. 19, 2025.
32. Telus, "TELUS to launch Canada's leading sovereign AI factory, powered by NVIDIA to drive the nation's AI future," March 19, 2025; Bell, "Increasing sovereign AI capacity: Introducing Bell AI Fabric," May 28, 2025; SaskTel, Deloitte Canada and SaskTel announce strategic alliance to bring Artificial Intelligence (AI) capabilities and solutions to market, advancing Canada's AI vision," press release, Sept. 23, 2025.
33. OpenAI, "Introducing Stargate UAE," May 22, 2025.
34. Amazon, "AWS and HUMAIN announce a more than \$5B investment to accelerate AI adoption in Saudi Arabia and globally," May 13, 2025; Nvidia, "HUMAIN and NVIDIA announce strategic partnership to build AI factories of the future in Saudi Arabia," press release, May 13, 2025; PIF, "HRH Crown Prince launches HUMAIN as global AI powerhouse," press release, May 12, 2025.
35. Natasha Turak, "Saudi AI firm Humain is pouring billions into data centers. Will it pay off?" CNBC, Aug. 27, 2025.
36. Chris Metinko, "Defense tech venture funding gains traction," Crunchbase News, Feb. 12, 2025.
37. 3MTT, "Shaping the future of Nigeria's digital workforce," accessed Oct. 30, 2025; European Commission, "Commission to invest €1.3 billion in artificial intelligence, cybersecurity and digital skills," press release, March 28, 2025.
38. Goldman Sachs, "AI to drive 165% increase in data center power demand by 2030," Feb. 4, 2025; Felicity Barringer, "Thirsty for power and water, AI-crunching data centers sprout across the West," Stanford University, April 8, 2025.
39. Karthik Ramachandran, Duncan Stewart, Kate Hardin, Gillian Crossan, and Ariane Bucaille, "As generative AI asks for more power, data centers seek more reliable, cleaner energy solutions," Deloitte Insights, Nov. 19, 2024.
40. ECDPM, "Von der Leyen in India: A tech sovereignty partnership in the making," Feb. 28, 2025; Nii Simmonds and David Timis, "How Europe and Africa can unlock tech opportunities through stronger collaboration," World Economic Forum, Aug. 18, 2025.

41. Rolfe Winkler, Nate Rattner, and Sebastian Herrera, "Big tech's \$400 Billion AI spending spree just got Wall Street's blessing," The Wall Street Journal, July 31, 2025; Financial Times, "What'll happen if we spend nearly \$3tn on data centres no one needs?" July 30, 2025.
 42. Caiwei Chen, "China built hundreds of AI data centers to catch the AI boom. Now many stand unused," MIT Technology Review, March 26, 2025.
-

ACKNOWLEDGMENTS

The authors would like to thank **Richard Nunan, Michael Greco, Julia Tavlas, Paul Lee, Ben Stanton, Manel Carpio, and Mosche Orth** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

生成式AI视频完美适配社交媒体，或成社交媒体公司威胁

最新的生成式人工智能（*Generative AI*）视频模型正接近好莱坞水准，似乎正在推动独立视频的快速发展，但也可能引发针对社交媒体平台更严厉的监管反应。

大脚怪自拍冒险。童话公主的爱情建议。真假难辨的突发新闻。生成式视频正以富有创意却时而令人担忧的方式充斥社交媒体，争夺着人们的注意力。现实世界似乎正面临比以往更加激烈的竞争。

生成式视频技术可能使独立创作者以更低的成本产出更多内容，同时增强社交媒体平台提供引人入胜的短视频娱乐的能力，从而获得更大的数字广告市场份额。然而，同样的能力也可能使观众不堪信息过载，削弱内容的真实性，并促使监管机构试图遏制生成式视频可能带来的负面效应。

当任何人都能制作出逼真的视频，并将其作为“新闻”、品牌内容、同人作品等发布给潜在的数百万受众，或利用其进行诈骗、胁迫或故意传播虚假信息时，这种滥用的潜在风险可能会加剧监管机构遏制新媒体的呼声。

德勤预测，到2026年，生成式视频可能在美国引发监管回应，或将在更多州推动年龄验证措施，并重新引发针对《通信规范法》1996年确立的第230条款保护的联邦层面挑战。¹并要求对在社交平台上发布的AI内容进行标注。此类监管举措在美国一些州（如纽约、田纳西和犹他州）已经启动。²美国最高法院拒绝受理针对密西西比州社交媒体使用年龄验证法的异议。³欧盟的《数字服务法案》还包括关于“有效的年龄验证方法”的规定。⁴2026年是美国的选举年，社交平台可能将被迫利用其AI和数据能力，以更好地管理生成式内容。一些平台已在积极推进此类解决方案。⁵

生成式AI可能会导致视频内容大量涌现，同时也有望实现对海量内容更有效的审核。监管机构可能会关注这些措施在应对网络有害信息方面的实际成效。平台方同样会密切关注，同时也会留意用户参与度下降、变现能力减弱以及合规方面面临的挑战。

一些独立内容创作者正借助生成式工具获得能力提升

生成式视频模型能够生成高质量的音视频短片段，这些片段几乎与“真实”内容难以区分。⁶其相对易用性与成本优势，正赋能部分创作者充分释放创造力——他们能以低得多的失败风险尝试各类创意，甚至能在社交视频的超竞争市场中快速验证创意想法。

尽管它们可能无法制作30分钟的电视剧或完整的两小时电影，但生成式视频工具完全能够制作出适合社交媒体的引人入胜的视频内容，例如将网络迷因生动呈现的高调广告。⁷事实上，生成式视频在好莱坞推广中所面临的感知局限性，反而为一些创作者和社交视频平台带来了机遇。在这些平台上，短小精悍、快速剪辑和自拍内容更容易走红，观众对其所获得的免费、开源娱乐内容也可能不那么挑剔。

对于独立创作者而言——也许很快也将适用于所有媒体制作——生成式工具的重点可能不在于替换整个制作流程以生成完全合成的内容，而在于消除成本高昂的微小任务，缩短创作时间，并使小型团队能够完成更多工作。

生成式AI和视频工具正在推动更低成本、更快速的内容创作，进一步消除了生产、分发和衡量过程中的大量微小任务。⁸这可以扩大产出，有助于跟上快速的内容发布节奏，而这种节奏通常是吸引关注者并在社交媒体平台的算法信息流中脱颖而出所必需的。

许多工具专注于节省时间和金钱的捷径，例如根据脚本快速生成视频以及“一键”生成片段。⁹这有助于创作者快速测试不同版本，以确定哪些方式更能吸引特定受众并适应热门算法趋势。其他工具则使创作者能够生成自己的AI虚拟形象，既能减轻疲劳，又能持续吸引观众，甚至实现大规模的个性化互动。¹⁰相同的功能正在扩展到生成式广告中。¹¹

生成式工具还可以通过更快的编辑来支持非生成式内容，例如删除“呃”、“嗯”等语气词、静音片段和不理想的镜头，修复画面抖动，以及自动移除无效空白区域。¹²多语言配音工具可帮助触达外语受众，扩大用户参与度并提升广告收入潜力。¹³

凭借这些能力，创作者工作室能够更快地进行创意构思、生成内容、精准触达受众、评估效果并不断迭代。这不仅可能颠覆内容创作的经济模式，还可能导致内容数量呈指数级增长。内容供给的增加可能加剧创作者之间的竞争压力，从而激发更多富有创意的内容产出。

生成式视频或对好莱坞和社交媒体平台构成威胁

正如德勤在《2025传媒及娱乐展望》报告中所说，一些大型影视工作室和出版商一直在探索生成式视频技术，但在将其整合到实际制作中时仍持谨慎态度。这种谨慎态度部分可能源于担心合成媒体会削弱其高端内容的竞争力，同时也受到来自演艺人员方面的阻力。2023年美国演员工会（SAG/AFTRA）的罢工行动中，就包含了限制在影视制作中使用生成式AI的条款。¹⁴然而，好莱坞影视公司常常面临高昂的制作成本，因此可能希望生成式AI能最终能够减轻这一负担。¹⁵

与此同时，传统影视公司和流媒体平台在广告收入方面面临着更激烈的竞争。¹⁶尽管一些好莱坞影视公司正努力遏制因传统线性电视业务下滑而造成的广告收入损失，并将广告业务转向联网电视和流媒体视频服务，但许多社交媒体平台却攫取了更多的数字广告支出。广告主在这些平台上的投入增长显著高于其他数字媒体，例如流媒体视频服务。¹⁷

生成式AI既能快速生成内容，又能预测哪些用户群体和个人会参与互动的能力，似乎正在重塑数字广告领域。¹⁸通过简单提示，社交平台可自动生成数千条仅有细微差异的广告，并即时测试哪些变体表现最佳。¹⁹这使得广告主能减少广告制作投入，将更多预算用于测试回报率最高的广告变体，进而强化了社交平台的竞争优势。²⁰

然而，社交平台在开展自身生成式视频业务的同时，也可能面临更大的风险。它们或将面对需要提供和管理的视频内容数量进一步激增，其中部分内容可能涉及侵犯版权，甚至落入更严重的类别。平台还可能面临用户被“AI垃圾内容”淹没的风险，进而导致推动社交媒体增长的核心要素——真实性——迅速贬值。

一年前，我们询问了美国民众对生成式媒体的感受。²¹64%的受访者认为社交媒体上的生成式AI具有危险性；76%的受访者认为网络内容创作者应公开说明在何时何地使用了生成式AI；53%的受访者认为使用生成式AI的网络内容创作者缺乏真实性。

如今，一年过去了，生成式视频的技术能力及其在平台上的数量都显著增长。生成式视频似乎正迅速接近真实画面的水平，很快可能进入危险领域，既可能引起监管机构的关注，也可能被不法分子利用。随着模型的发展，人们担心不法分子可能利用AI冒充他人，造成潜在的欺诈风险。²²伴随着大脚怪自拍视频的出现，很可能还会带来影响性宣传运动、诈骗、政治虚假信息以及阴谋论的喧嚣言论。如果视频证据变得不可信，甚至可能影响法律诉讼进程。然而，若缺乏监管监督、用户大规模流失或惩罚性赔偿，平台通常缺乏动力去加以遏制。

小结：社交平台既能保护真相，也能维护自身利益

与社会面临的挑战相比，合成媒体、AI垃圾内容以及被打乱的商业模式可能都显得微不足道——当任何人都能制作并传播逼真视频，而视频证据不再是一种可靠的真实形式时，情况尤为严峻。观察当前生成式视频技术的前沿发展——尤其是那些试图揭示虚假新闻、名人出没、栽赃陷害和政治失言风险的视频——我们很难低估即将到来的这场颠覆性浪潮，它似乎正迅速逼近。

为了防范这些风险，社交平台应致力于开发并整合水印技术、AI标注功能，以及追踪和披露所有上传至其服务或由其服务生成的内容（包括广告）来源的方法。面对似乎不可避免的政治操纵和消费者欺骗现象，监管机构应与平台合作，为生成式内容建立更严格的规范，例如强制要求进行标注和添加水印。

在美国，如果开放平台不提前应对这些日益加剧的风险，《通信规范法》第230条可能会受到挑战，而该条款一直保护着这些平台免于为其托管的内容承担法律责任。²³欧洲监管机构已表现出强烈意愿，将对美国的社交平台和数据收集企业进行监管。²⁴开发更强大的合规自动化机制，例如能够监控其他生成式工具输出内容的合规代理，可使平台能够大规模快速应对违规行为。

尽管社交媒体在连接性、透明度和弘扬人性方面成就显著，但它也加剧了信息碎片化、媒体去监管化以及恶意行为者的活动能力。如果社交媒体平台不大力遏制这些能力，生成式视频技术可能将这一状况大幅加剧，进一步使社会脱离对基本事实的共同认知。

Chris Arkenberg

United States

Gillian Crossan

Global

Tim Bottke

Germany

ENDNOTES

1. U.S. Congress, Senate, "S. 314 – A bill to protect the public from the misuse of the telecommunications network and telecommunications devices and facilities," accessed Oct. 22, 2025.
2. Mayer Brown LLP, "Children's online privacy: recent actions by the states and the FTC," Feb. 25, 2025.
3. Ella Lee, "Supreme Court — Mississippi social-media law and minors' access," The Hill, Aug. 14, 2025.
4. European Commission, "Commission press corner detail: IP/25/1820," press release, July 14, 2025.
5. James Beser, "Extending our built-in protections to more teens on YouTube," YouTube News & Events Blog, July 29, 2025.
6. The New York Times, "AI video deepfakes – quiz and playground," June 29, 2025.
7. Bill Chappell, "AI video ad, Kalshi advertising NBA finals," NPR, June 23, 2025.
8. Thomas H. Davenport and Nitin Mittal, "How generative AI is changing creative work," Harvard Business Review, Nov. 14, 2022.
9. Torin Anderson and Shuo Niu, "Making AI-enhanced videos: Analyzing generative AI use cases in YouTube content creation," In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, pp. 1-7. 2025.
10. Collectively Inc., "How content creators are embracing generative AI and AI avatars: insights from our latest survey," Jan. 14, 2025.
11. Jess Weatherbed, "TikTok ads may soon contain AI-generated avatars of your favorite creators," The Verge, June 17, 2024
12. Anderson and Niu, "Making AI-enhanced videos: Analyzing generative AI use cases in YouTube content creation."
13. Yael Malamatinas, "7 of the best AI dubbing tools to translate videos into different languages," Vimeo, blog, April 28, 2025.
14. Screen Actors Guild – American Federation of Television & Radio Artists, "SAG-AFTRA statement on the use of artificial intelligence and digital doubles in media and entertainment," March 17, 2023.
15. Katie Kilkenny, "Higher costs are hitting film and TV producers even as studios keep trimming budgets," The Hollywood Reporter, April 17, 2025.
16. Chris Arkenberg, Jeff Loucks, Kevin Westcott, Danny Ledger, and Doug Van Dyke, "2025 media and entertainment outlook," Deloitte Insights, April 23, 2025.
17. Interactive Advertising Bureau, "Digital ad revenue surges 15% YoY in 2024, climbing to \$259 B," April 17, 2025.
18. Ryan Browne, "AI is disrupting the advertising business in a big way — industry leaders explain how," CNBC, June 15, 2025.
19. Charles James, "Generative AI for retail ad campaign variants and A/B testing automation," ResearchGate, Nov. 9, 2024.
20. Interactive Advertising Bureau, "Nearly 90% of advertisers will use Gen AI to build video ads, according to IAB's 2025 video ad spend & strategy full report," July 15, 2025.
21. China Widener, Jana Arbanas, Doug Van Dyke, Chris Arkenberg, Bree Matheson, and Brooke Auxier, "2025 digital media trends: Social platforms are becoming a dominant force in media and entertainment," Deloitte Insights, March 25, 2025.
22. Clare Duffy, "OpenAI's Sam Altman warns of an AI 'fraud crisis,'" CNN, July 22, 2025.

23. Paris Martineau, "Exclusive: Section 230 may finally get changed — lawmakers prep new bill," The Information, accessed Oct. 22, 2025.
 24. Dawn Carla Nunziato, "The Digital Services Act and the Brussels Effect on platform content moderation," Chicago Journal of International Law 24, no. 1 (2024): pp. 1-37.
-

ACKNOWLEDGMENTS

The authors would like to thank **Rohan Gupta** and **Jana Arbanas** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

当公共媒体遇上流媒体巨头：传统电视可持续发展的破局之道

公共服务广播机构机构正积极在社交媒体平台发布内容，与流媒体服务商联合制作节目，并与大型视频发行商建立合作关系。它们可为美国营利性媒体公司提供可借鉴的经验。

在美国，传统电视业务虽仍能盈利，但已呈衰退之势。在流媒体大战、社交视频和互动娱乐的冲击下，内容库正纷纷转向流媒体视频服务，版权也在重新谈判，线性电视业务则被拆分出售。¹然而，许多传统媒体集团仍在行动并投入大量资金，试图围绕自身全新且昂贵的流媒体视频服务，重建电视行业的黄金盈利时代。但到目前为止，这一黄金时代仍未回归。

在美国以外，公共服务广播机构（PSBs）正在涌现出更多适应性强的成功模式。欧洲的公共服务广播机构拥有经受过考验的内容制作能力，并能将内容输出给更广泛的受众。它们通过与流媒体平台达成联合制作协议来加强制作和分发，通过在社交平台发布和推广内容来触达年轻观众，并通过在自有服务平台与全球合作伙伴之间尝试错开发布窗口来扩大其覆盖范围。²这些创意策略可以为面临类似转型压力、力求保持竞争力的美国小型电视网提供宝贵的经验借鉴。

2025年，在短短几个月内，广播公司与流媒体平台之间达成了三笔引人注目的交易，合作进程明显加速。³德勤预测，到2026年，还将有数笔广播公司与流媒体平台的交易达成。此外，我们预计将出现更多联合制作项目及其他举措，且这些活动仍将由公共服务广播机构牵头。这或将为流媒体视频服务及社交平台额外带来数万小时的广播公司内容，同时有望提升广告收入分成及全球观看时长。

目前，公共服务广播机构在达成此类交易方面行动更为迅速，其动力更多源于拓展受众范围而非追求利润，而流媒体交易似乎已成为一条有效途径。有趣的是，一些商业广播公司也在达成类似的流媒体交易，但究竟它们是特例，还是预示着其他商业电视广播公司也将步其后尘，仍有待时间验证。无论如何，广播业界都在密切关注，看这些交易最终能否带来“皆大欢喜”的结局。

公共广播机构正积极寻求变革，并探索创新发展路径

在全球范围内，许多公共服务广播机构都是大型文化机构，在代表公众、触达公众以及塑造娱乐、新闻、教育和文化方面发挥了重要作用。⁴然而，它们也面临着与冲击传统线性电视相同的人口结构和行为变化所带来的威胁。⁵

然而，商业广播机构与公共服务广播机构之间存在重要差异。公共服务广播机构通常肩负着制作服务于公众利益的电视节目、电影、新闻和纪录片的使命，而不论其资金来源模式如何。如果国内观众不收看它们的线性电视频道或流媒体服务，公共服务广播机构可能无法有效服务公众，从而加剧了关于它们是否应继续享有资金支持和优先受众渠道的争议。

公共服务广播机构依赖公民和政府的资金支持，同时被赋予履行公共使命的职责，因此可能面临资金不足的问题，但也往往比私营机构更专注于服务受众和传播文化，而非追求利润。这一特点使其能够比私营同行更迅速、灵活甚至更大胆地进行创新，因为后者通常将风险承受能力与利润和股东利益挂钩。

以下示例为其他公共媒体机构在现代媒体环境中履行其使命提供了借鉴，也为那些可能因领先流媒体平台的成功而蒙蔽双眼、同时又对涉足更具创新性和颠覆性机遇持犹豫态度的私营传统媒体公司提供了参考。当然，机遇中也伴随着风险，如果与营利性供应商的关系损害了公共媒体的价值和使命，就会带来问题。

在YouTube上吸引年轻观众：ARD和ZDF的破局尝试

2016年，德国公共广播机构ARD和德国电视二台ZDF已经意识到年轻观众的关注点所在，以及他们正在远离哪些内容。为此，他们推出了“funk”这一大胆的数字内容计划，旨在通过年轻人偏爱的社交媒体平台与他们建立联系。funk工作室并非发布现有电视节目的完整剧集或短视频片段，而是专门为YouTube、Instagram和TikTok等社交平台制作视频内容。⁶Funk推出了数十种专为数字原生代设计的原创节目形式，包括简洁明了的科普内容、风格鲜明的喜剧以及短纪录片。例如，Funk旗下的科学科普频道maiLab通过通俗易懂地讲解化学和COVID-19研究而广受欢迎。

到2026年，近41%的30岁以下德国人现在每周至少观看一次funk的节目内容。⁷在两年期间，Funk相关内容在YouTube上获得了约22亿次观看和1.73亿小时的观看时长。随着年轻观众转向以短视频为主的新平台，Funk也随之迁移。如今，TikTok和Instagram在扩大Funk影响力方面发挥着重要作用。事实上，过去两年中，Funk内容在TikTok上的观看量甚至超过了YouTube，达到约23亿次，反映出短视频片段日益流行的趋势。

在年轻人聚集的地方与他们建立联系，对于ARD和ZDF履行公共使命至关重要。通过在那些由网红和算法信息流主导的空间中提供专业的公共资助内容，他们为纯粹商业化的社交媒体提供了另一种选择。

加拿大广播公司（CBC）的YouTube战略大转向

直到最近，拥有90年历史的加拿大公共广播公司CBC的重点一直放在其自身的流媒体应用CBC Gem上。但在2023年，CBC的数字战略团队开展了一项试验：他们将CBC过往节目的完整剧集甚至整季内容上传至YouTube，将该平台视为一种新的发行渠道。⁸他们采取了“测试并学习”的策略，一旦发现某些内容从CBC自有平台分流了过多观众，便立即下架。YouTube平台非但没有削弱CBC Gem，反而成为了一个补充性渠道，作为一个营销渠道，吸引了更多年轻的新观众观看CBC的内容。许多观众通过YouTube发现了这些节目，随后便前往CBC Gem寻找更多剧集，从而形成了良性循环的“飞轮效应”。

CBC目前管理着超过50个YouTube频道，涵盖新闻、喜剧、儿童节目等多个领域。⁹短视频，如喜剧小品和热门新闻片段，通常能获得数百万次观看。加拿大广播公司新闻的YouTube频道目前拥有超过430万订阅者和总计26亿次观看量。¹⁰CBC还发布了完整的20多分钟的戏剧、纪录片和儿童节目的剧集，这些内容几乎占到了所有观看时间的一半。¹¹虽然短视频能带来点击量，但真正让观众持续关注频道的还是有完整长度的节目。

到2024年底，CBC在YouTube上的尝试已逐渐吸引观众。其各频道的总观看时长增长了65%，超过了团队设定的25%增长目标。¹²YouTube扩大了CBC触及的人群范围，这些人群是传统电视难以有效覆盖的。这一策略使CBC能够更充分地发挥其内容的价值，为过往的节目内容注入新的活力，并创造新的收入来源。

法国广播公司借助顶级发行商之力

2025年7月，PSB法国电视台与亚马逊Prime Video达成了一项“历史性发行协议”。¹³根据该协议，法国的Prime Video订阅用户可免费观看法国电视台（France Télévisions）频道的直播内容，以及其点播目录中的2万个节目。¹⁴Prime Video的主界面现在设有一个专门的france.tv专区，可在亚马逊的界面中展示该广播公司的内容。¹⁵事实上，亚马逊的流媒体服务已成为承载法国公共频道的一个新的虚拟有线电视运营商。

对于法国电视台而言，其好处在于能够更有效地接触年轻且不再使用传统有线电视的观众群体。在观众观看习惯日益碎片化的背景下，出现在热门流媒体平台的菜单中，是一种保持影响力的方式。一些欧洲的私营广播公司似乎也认同这一点。

法国最大的私营广播公司TF1与Netflix签署了一项类似协议。到2026年，这一合作（Netflix在全球范围内的首个此类合作）将使法国的Netflix用户无需离开Netflix应用程序即可观看TF1的直播节目。¹⁶法国正在进行的这项实验将受到欧洲各地媒体高管的密切关注。毕竟，如果你无法击败最大的流媒体平台，那么加入它们或许就是次优选择。

BBC和Channel 4借流媒体合作伙伴之力实现全球成功

曾经，电视剧上出现BBC（英国广播公司）或Channel 4（第四频道）的标志意味着这完全是一档本土制作的节目，但如今它也可能与Netflix、亚马逊Prime视频或HBO Max等平台合作出品。

通过借助流媒体平台雄厚的资金实力、广泛的国际发行网络以及对英国优质叙事内容的需求，合拍片使公共广播机构能够开展原本可能超出其财政承受能力的项目。¹⁷正如一个行业贸易组织指出的那样，第三方资金（通过联合制作协议、国际预售、税收抵免等方式）目前每年为英国公共服务业节目制作提供约4亿英镑的资金支持。¹⁸实际上，Netflix或亚马逊可能会承担大部分制作费用，以换取在全球范围内流媒体播放该节目的权利。这种合作模式降低了英国广播公司的风险，同时有助于确保一部本土热门节目能够传播到英国以外更广阔的市场。

BBC积极寻求此类合作，尤其是在制作豪华电视剧集方面。根据菲利普·普尔曼小说改编的奇幻史诗剧《黑暗物质》便是BBC与HBO联合制作的剧集，据报道，其第一季的制作成本约为5,000万英镑。¹⁹HBO的资金使BBC得以实现真正电影化的创作愿景，而作为回报，HBO则获得了一部现成的、具有高声誉的剧集，可用于美国市场。²⁰即使是典型的英国时代剧也从中受益：阴郁的维多利亚时代犯罪剧《浴血黑帮》由BBC与Netflix联合制作（Netflix在后续几季中接管了美国地区的发行），使一部原本局限于BBC Two的小众剧集成为全球热门作品。Channel 4也通过与全球流媒体平台合作，取得了其他多部国际成功的剧集。²¹

通过与Netflix、亚马逊等公司的合作，BBC和Channel 4正在确保英国公共服务内容不仅能在21世纪的媒体环境中生存下来，而且能够蓬勃发展，受到全球观众的喜爱、分享和追看。²²与CBC类似，Channel 4也通过在YouTube上发布完整剧集，实现了各项业务的逐步增长。²³2025年7月，Disney+与英国公共广播公司ITV宣布建立合作关系，让双方观众可以体验彼此的内容。根据协议，ITV的流媒体服务（ITVX）将轮换上线一系列Disney+的热门剧集，而英国地区的Disney+也将同步推出精选的ITV热门节目。双方均表示，这是一次互利共赢的尝试，也表明流媒体竞争正逐步转向战略联盟。

陷阱与风险：公共广播机构在合作中可能面临的风险

与全球平台结盟虽能提升知名度、获取资金支持，但也存在重大风险。若管理不善，合作关系可能削弱广播机构的自主性、稀释其品牌价值，并损害其公共服务使命。当公共服务广播机构涉足科技巨头与巨额资本的领域时，应审慎评估以下风险：

- 控制权的丧失与依赖性：**当内容分发与收入来源均依赖外部平台时，公共服务广播机构可能受制于算法调整或企业战略转向。单一合同终止即可能导致其失去受众触达渠道或内容版权。若平台算法降低某广播机构视频的推荐权重，其传播范围可能在一夜间锐减且难以补救。
- 直接受众联结的弱化：**通过其他平台获取公共服务广播机构内容的受众可能不再访问其自有平台，导致品牌曝光度下降，且难以获取关键收视数据。对于依赖广告收入的公共服务广播机构而言，这还意味着变现潜力降低。

- **内容导向与使命偏离风险：**为追求点击量、流媒体资金支持或全球吸引力，公共广播机构可能偏离其核心职责。为迎合算法或平台商业目标而定制新闻或纪录片，可能损害编辑独立性及本土文化特色。
- **编辑独立性与国家主权风险：**与外国流媒体巨头的合作已引发政治关注，部分制作方与立法者担忧美国企业借此影响他国公共内容。²⁴过度依赖外部商业资金支持，也可能削弱公众对执照费或纳税人资金支持的认同。
- **财务依赖性与可持续性风险：**若流媒体平台调整战略或收缩投入，高度依赖其资金的公共广播机构将陷入困境。其预算编制可能缺乏稳定性，甚至旗舰节目亦难幸免。

尽管公共广播服务机构正在不断创新，以跟上受众行为、参与方式和资金模式的变化，但它们在与流媒体平台合作时，也面临着在控制权、身份认同和可持续性方面的潜在风险。公共广播机构在一定程度上正以部分独立性和直接接触受众的能力，换取短期的资金和观众增长。这种做法是一种博弈，其前提是它们能够妥善管理这种权衡——即能够驾驭这头猛兽而不被甩下或吞噬。

为了管理这种权衡，公共广播机构应：

1. **保护品牌和可见性：**确保在第三方平台上标识和署名保持显著位置，以维持信任和认知度。
2. **确保数据获取和比例收入分成：**协商获取观看数据的权限，并争取按比例获得报酬，以保持议价能力。
3. **建立联盟：**诸如英国Freely平台之类的联合举措有助于公共广播服务机构在全球流媒体竞争中保持竞争力。
4. **坚持公共价值导向的内容：**持续投资于本地新闻、教育、文化以及少数语言节目，即使这些内容未必成为全球热门。
5. **以目标为导向进行创新。**通过合作，借鉴全球平台的技术，并将这些经验应用于提升内部数字化服务能力。

美国新兴流媒体频道的启示

对于美国媒体公司而言，它们正努力将日渐衰退的传统电视业务转型为具有竞争力的流媒体服务，而英国和欧盟的公共广播机构则提供了更具灵活性和创新性的路径。在综合考虑上述诸多风险的情况下，公共广播机构的发展历程为那些正面临更强大竞争对手的美国流媒体公司提供了一份可借鉴的行动指南：

通过建立战略合作伙伴关系，利用传统和小众内容扩大覆盖范围：与大型平台合作不仅不会侵蚀现有观众群或削弱知识产权（IP），反而能够唤醒沉寂的受众，并将品牌和知识产权带给那些原本不会订阅小众服务的全新受众群体。

利用优质内容作为核心，建立有价值的合作关系，以扩大覆盖范围：广播公司已将其本地内容作为谈判筹码，以获得全球发行渠道。²⁵商业工作室可利用珍贵的知识产权（IP）吸引新订阅用户。例如，ITV借助本土热门节目《恋爱岛》（Love Island）将《熊家餐馆》（The Bear）引入其平台；Disney+则利用《曼达洛人》（The Mandalorian）的粉丝群体，吸引ITV观众转向Disney+。商业电视台还可以像公共广播服务机构那样，更频繁地与全球流媒体平台联合制作内容。

保护品牌与数据，但需务实。与公共广播机构类似，商业媒体公司正在逐步认识到：必须追随受众以保持相关性，需要通过合作实现规模效应，并且即使在其他平台上运营，也必须重视维护自身身份认同。随着线性收视率和广告收入的下降，以及许多流媒体服务仍处于亏损状态，这些合作关系可能将从战术性尝试演变为核心战略。公共广播机构的经验表明，若运作得当，联盟关系能够实现价值叠加，并具备财务上的明智性。

小结：公共广播公司与美国流媒体的关键考量因素

许多公共服务广播机构非但没有被流媒体革命和社会化视频所淘汰，反而正借此实现自我重塑——将内容推向社交媒体平台，与顶级流媒体创作者联合制作，甚至允许流媒体平台转播其频道。如果实施得当，这或将催生一个更加丰富的媒体生态系统，使公共服务内容与商业内容在各个平台上共存，从而为全球性频道注入一定的本土化和伦理平衡。

尽管公共服务广播机构在面对营利性合作时，面临履行和捍卫其公共使命的重要挑战，但此处展示的创新案例同样适用于面临相同适应压力的美国第二梯队及小众制片公司和流媒体服务商。网络流媒体已经解构并颠覆了传统电视，而社交视频平台正将观众从电视和流媒体服务中分流。最大的视频分发平台仍在持续重塑和重新定义电视。无论是公共广播机构还是私营媒体，几乎别无选择，只能不断尝试和探索和适应。

Jeff Loucks

United States

Tim Bottke

Germany

Chris Arkenberg

United States

Duncan Stewart

Canada

ENDNOTES

1. Chris Arkenberg et al., "2025 media and entertainment outlook," Deloitte Insights, April 23, 2025.
2. Ofcom, "Transmission critical: The future of public service media," July 21, 2025.
3. Elsa Keslassy, "How streamers and broadcasters' cross carriage deals could disrupt the TV business in Europe," Variety, July 11, 2025.
4. Knight Foundation, "Public broadcasting: Its past and its future," accessed Oct. 29, 2025.
5. Once hugely profitable, in most countries the audience for traditional "linear" TV is getting smaller and older. Globally, multiple public service traditional TV broadcasters (PSBs) are seeing their viewership erode, especially among younger audiences. In the United Kingdom, for example, less than half (48%) of 16 to 24-year-olds watched any broadcast TV in a given week in 2023, down from 76% five years earlier [CSI, "Gen Z abandons traditional broadcast TV: Ofcom," July 31, 2024], and only 55% of children between 4 and 15 tuned in weekly, down from 81% in 2018. Still in the United Kingdom, young adults who do watch traditional TV spend barely half an hour per day on it, versus 93 minutes on video-sharing platforms like YouTube and TikTok. Even overall reach is shrinking: the weekly audience for any broadcast TV fell to 75% of Britons in 2023 (down from 79% the year prior)—the steepest decline on record. The same pattern is found in nearly every advanced country's media market and is generally true for both PSB and commercial broadcasters.
6. Funk, "Funk Bericht 2024," Dec. 13, 2024.
7. Ibid.
8. Evan Shapiro and Marion Ranchet, "TESTING & LEARNING: The CBC Case Study," The Media Odyssey, audio podcast episode, April 24, 2025.
9. Ibid.
10. Social Blade, "CBC News YouTube channel statistics," accessed Oct. 29, 2025.
11. Shapiro and Ranchet, "TESTING & LEARNING."
12. Ibid.
13. K.D. with AFP, "'Un accord historique': après TF1 et Netflix, France Télévisions s'associe à Prime Video pour diffuser ses contenus sur Amazon," BFM Tech & Co, July 3, 2025.
14. Ibid.
15. Ibid.
16. AFP, "Netflix breaks new ground with global launch of French TV content," ForbesIndia.com, June 19, 2025.
17. Mark Sweeney, "BBC and ITV slash big-budget TV spend as US streamers pour money into UK," The Guardian, Feb. 16, 2025.
18. Pact, "Submission to Ofcom consultation on the proposals for the new Channel 4 licence," February 2024.
19. BBC, "His Dark Materials: Critics heap praise on 'ravishing' dramatisation," Nov. 4, 2019.
20. Sheena Scott, "'His Dark Materials' is BBC's most expensive series and promises to be a faithful adaptation," Forbes, Oct. 31, 2019.
21. BBC, "How The End of the F***ing World became a cult TV phenomenon," Nov. 4, 2019; Daniel D'Addario, "'It's a Sin' is a transporting and tragic tale of the AIDS epidemic: TV review," Variety, Feb. 21, 2018.
22. Travis Clark, "8 great Netflix original TV series that show how well its British strategy is working," Business Insider, April 2, 2019.

23. John Moulding, "[ITV and C4 happy to let viewers watch long-form content on YouTube](#)," *The Media Leader*, March 13, 2025.
 24. Max Goldbart, "[Streamers Will Not Be Regulated Fully In UK For Another Two Years](#)," *Deadline*, Feb. 26, 2025.
 25. Lucas Manfredi, "[Netflix, France's TF1 strike landmark distribution deal](#)," *TheWrap*, June 18 2025.
-

ACKNOWLEDGMENTS

The authors would like to thank **Santosh Anoo, Dan Littmann, Jack Fritz, Sai Tarun Dronamraju, Girija Krishnamurthy, Michael Greco, Kathryn Walby, Paul Lee, and David Jarvis** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

下一代卫星互联网正在全球范围内改变定价、容量及监管格局

卫星连接技术在直连设备（D2D）领域持续增长，但常面临变现难题；与此同时，低地球轨道（LEO）数据扩展和技术进步正在重塑部署方式与系统韧性，并带来复杂的监管挑战。

卫星连接的发展似乎正以前所未有的速度加速。直连设备（D2D）的卫星数量常呈激增之势，但普遍难以实现盈利；与此同时，低地球轨道（LEO）宽带星座的数量持续增长，要求电信运营商把握机遇并应对颠覆。伴随这些进展，技术进步正在重塑整个行业，助力实现更快的部署、更强的韧性和更低的成本。监管挑战和频谱管理也日益成为关键因素，有助于确保行业的可持续增长及其与地面网络的融合。

一些分析师预计，LEO卫星星座在2026年的年收入将达到约150亿美元。¹德勤预测，到今年年底，全球用户数将超过1500万。²我们进一步预测，到今年年底，LEO通信卫星的数量将增至五大星座，³由超过1.5万至1.8万颗卫星组成。⁴我们还预测，2026年D2D卫星容量的支出将达到60亿至80亿美元，到年底将有超过1,000颗具备D2D能力的卫星在轨运行……尽管目前D2D的变现方式和商业模式尚不明确，因此我们暂未对D2D收入进行预测。

D2D与LEO卫星服务密切相关。自2019年初以来，多颗卫星被发射至低地球轨道，为地面小型卫星天线提供数据服务，使消费者和企业能够在地面网络覆盖有限或无覆盖的地区，以合理的价格获得低延迟的宽带服务。⁵然而，如果没有比萨盘大小的碟形天线，就无法接收到这些信号。⁶2023年，主要位于低地球轨道的新卫星配备了新设备、新天线，并利用新的监管许可，实现了与智能手机等设备的直接连接。与通过LEO卫星和碟形天线服务提供的50 Mbps甚至更高速率的连接不同，2023年的D2D通信仅支持低比特率的简单消息传输。⁷未来，D2D可能提供更高的连接速度，但仍不及碟形天线的速度快。⁸LEO和D2D在所使用的轨道方面可能存在重叠，甚至在某些使用的卫星上也可能存在重叠，⁹但它们并不相同。2025年9月，一家主要的LEO运营商购买了用于设备到设备通信的5G频谱区块。¹⁰这很可能不会在2026年实现：新智能手机需要配备可在该频谱上收发信号的新芯片，还需发射新的卫星以使用这些频段，同时尚不清楚该服务能支持多少用户同时使用，以及其在室内环境中的表现如何。但到2028年左右，智能手机用户或许将能够直接通过太空向手机流式传输视频。¹¹

D2D能够连接未被连接的用户，但盈利模式仍然难以捉摸

D2D技术有助于实现卫星与智能手机等标准消费设备之间的直接通信，绕过传统的地面基础设施，提供基本的、通常为低带宽的连接服务。¹²在缺乏地面蜂窝网络覆盖的偏远或农村地区（包括海洋区域），这一能力可能尤为重要。

2023年，德勤预测D2D卫星通信市场在网络基础设施（主要是卫星）上的支出将达到约30亿美元。¹³实际支出超过预期，2024年达到约40亿美元。¹⁴当前各公司的路线图及已公开宣布的投资计划显示，2026年的总资本需求约为60亿至80亿美元。其中，我们预计约85%至90%将用于资助新卫星的部署，其余10%至15%则用于替换现有卫星。¹⁵许多手机制造商和芯片供应商已将卫星连接功能集成到智能手机中：德勤预计，2024年将售出超过2亿部支持卫星功能的手机（带动约20亿美元的专用芯片市场）。¹⁶大多数主要智能手机制造商都推出了支持卫星通信的旗舰设备。¹⁷

其次，移动网络运营商（MNO）与卫星公司之间的一系列合作推动了D2D服务的普及。¹⁸这些合作有助于在没有蜂窝网络覆盖的服务不足地区实现基本连接（紧急短信、低带宽数据），从而覆盖大量尚未联网的人群。

在许多人均GDP较低的市场中，农村和偏远地区对地面电信运营商而言通常缺乏商业吸引力。这些运营商通常优先发展城市中心，因为较高的收入水平能够支撑更高的用户平均收入（ARPU）。在一些国家的低收入、人口稀少地区，每个地面蜂窝网络基站的收入大约比城市低十倍，而资本支出和运营成本却高出两到三倍。¹⁹因此，许多电信公司倾向于避免在这些地区投资基础设施。²⁰截至2024年底，估计有3.5亿人（占全球人口的4%）居住在基本没有移动互联网网络覆盖的偏远地区，这凸显了D2D通信的发展机遇，但这些地区居民收入较低，可能导致许多D2D或LEO数据服务对消费者而言难以负担。²¹

第三，全球监管机构和行业标准组织已迅速采取行动，支持非地面网络的发展，通过分配频谱并完成5G非地面网络（NTN）规范，使普通手机能够无缝连接至卫星。²²

LEO卫星星座：快速发展、经济实惠的连接以及新兴的竞争格局

LEO通过位于特殊（低）轨道上的专用卫星与专用地面设备相结合，提供了高速、低延迟的宽带服务，并持续发展壮大：每周都有更多LEO卫星被发射，这些卫星组成的星座数量不断增加，用户数量和收入也在持续增长。

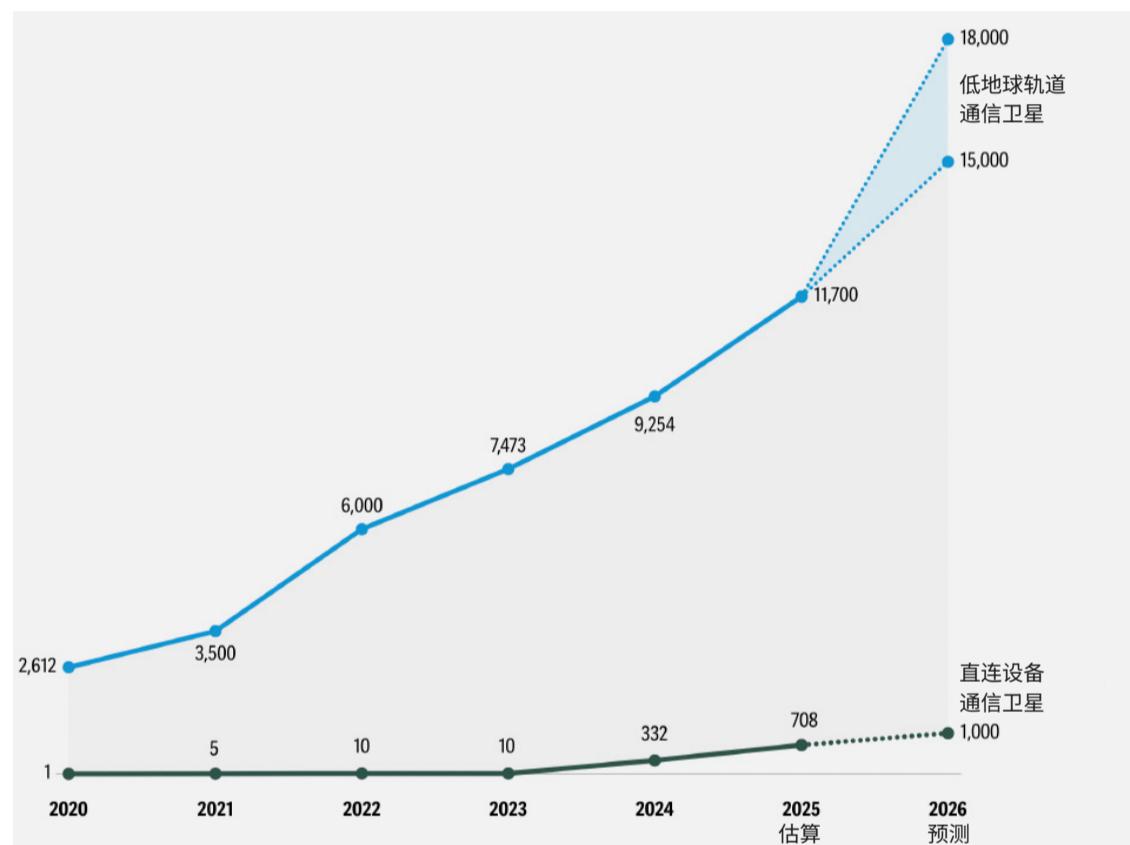
2026年，亚马逊的柯伊伯计划将进入市场，²³可能以低于其他低地球轨道解决方案的价格，在发展中国家市场与地面宽带服务提供商直接竞争。来自中国国网宽带巨型星座的其他低地球轨道星座，以及千帆（G60/太空帆）项目，²⁴加拿大的Telesat Lightspeed，²⁵欧洲IRIS²项目将于2026年或将未来几年内发射卫星进入轨道。²⁶一些区域性举措也正在涌现，例如总部位于阿联酋的Orbitworks项目。²⁷欧洲的成熟运营商（如Eutelsat OneWeb）正在升级其卫星星座，以帮助扩大容量、降低延迟，并实现直接连接设备的功能。²⁸

2022年时，德勤预测到2023年底，将有超过5,000颗宽带卫星位于低地球轨道，为近百万用户提供高速互联网服务。²⁹实际情况来看我们当时的预测过于保守，截至2023年底，已有约7,473颗活跃的具备宽带能力的LEO卫星。³⁰

与传统的地球静止轨道卫星相比，LEO卫星能够提供更低的延迟和更快的连接速度，部分LEO用户数量的增长是以地球静止轨道互联网服务提供商的市场份额为代价的。³¹LEO主要面向缺乏传统地面连接选项的用户，尽管迄今为止其价格通常高于同等的地面宽带服务。³²

LEO卫星服务的分销模式呈多样化，部分供应商直接向客户销售，或通过地面电信合作伙伴销售，亦或采用混合模式。一些供应商可能从2026年开始，通过提供更具价格优势的订阅模式，尤其是在新兴市场，直接与地面电信运营商展开竞争。³³

图1：低地球轨道（LEO）及直连设备（D2D）通信卫星估算



来源：德勤对公开可用的卫星行业数据以及2020 - 2025年公布的部署里程碑进行分析；这些估算数据反映了从行业追踪数据中得出的LEO活跃卫星的年度数量，以及基于已披露的发射活动和服务演示的D2D通信的卫星数量。

Deloitte. deloitte.com/insights

供应商的差异化营销策略

随着LEO宽带和D2D卫星市场的发展，德勤预测将出现两种截然不同的分销策略：合作与竞争。LEO服务提供商经常在特定地区与本地电信运营商合作，例如，我们在日本、澳大利亚和菲律宾的D2D市场中也看到了类似的合作伙伴关系。³⁴

来自LEO服务商的未来竞争

或者，德勤预测，部分卫星运营商将采取直接竞争策略，尤其是在发展中地区。这些运营商将以显著低于地面服务提供商的价格提供服务，旨在通过激进的定价策略和简化的服务方案来抢占未被充分覆盖的市场细分领域。尽管目前有多个新的LEO星座正处于规划或建设阶段，但预计亚马逊的Kuiper将成为下一个开始提供大规模服务的系统，他们正在开发一种低成本的月度定价模式。³⁵ Kuiper计划直接瞄准大量未覆盖或覆盖不足的地区，这可能对地面电信运营商构成挑战。³⁶

话虽如此，并非所有地面宽带市场都同样容易受到太空服务供应商的冲击。所选发达市场的平均宽带价格每月在33至80美元之间，而一些发展中国家市场的宽带价格则低于每月10美元。³⁷ 这表明，即使成本相对较低的卫星，在这些价格更实惠的市场中也难以获得显著的市场份额。与此同时，其他新兴市场如巴西或南非，其月费价格范围在21至48美元之间，³⁸ 可能看到更高的采用率，特别是如果LEO服务提供商将价格定在较低水平，或通过补贴价格来吸引客户，或提供低成本的连接服务，以便消费者能够更好地使用购物或流媒体等其他服务。需要注意的是，用于LEO的地面站终端每台成本为200至500美元。³⁹ 在发展中国家，即使每月费用相对较低（约为15美元），这种成本相对较高的消费设备对许多人来说仍然难以负担。⁴⁰ 因此，降低地面终端的价格也将是一个重要因素。

改造卫星容量可能是实现下一代连接的关键

单颗卫星及其所属星座的容量对于确保基于卫星的通信（包括LEO宽带服务和D2D的基本连接）的有效性、可靠性和商业可行性具有至关重要的作用。预计到2025年底，全球卫星数据流量将增长20倍，这对卫星容量带来了重大挑战。⁴¹ 提升卫星容量对于实现广域覆盖、支持高速数据传输以及在偏远或服务不足的地区为多个同时使用的用户提供连接至关重要：当前的网络通常能够很好地连接偏远地区数十个用户，但在人口密度稍高的区域往往面临挑战。LEO需要新技术来实现进一步的提升。⁴²

卫星容量的可用性受到多种技术和监管因素的影响。从技术角度看，容量取决于部署的卫星数量、各卫星的性能能力及其轨道位置。⁴³ 尽管理论模型表明，在理想条件下，近地轨道最多可容纳1,000万至1,200万颗卫星，⁴⁴ 由碰撞风险、跟踪误差和监管延迟带来的实际限制，可能将可持续运行的活跃卫星数量限制在约10万个左右。⁴⁵ 此外，监管要求，特别是与Ku波段和Ka波段等频段的频谱可用性相关的规定，可能会限制卫星服务提供商扩展其容量的能力。⁴⁶ 例如，尽管Starlink具有全球覆盖能力，但偶尔仍面临网络拥塞问题，导致服务可用性出现暂时中断，这凸显了充足容量的重要性。⁴⁷ 他们还需要限制某些地区（如英格兰东南部）的用户数量。⁴⁸

许多LEO公司正在采用更先进的技术解决方案，例如自适应波束成形、动态频谱共享、星间链路以及由人工智能驱动的网络优化。⁴⁹ 此外，投资于更先进、容量更高的卫星，改进卫星架构，以及协调监管工作以实现有效的频谱分配，将在平衡卫星容量与不断增长的用户需求和技术进步方面发挥重要作用。

对于那些没有其他选择的用户而言，LEO卫星网络可能非常有帮助，但在大多数发达国家市场，它不太可能成为地面现有网络运营商的重大竞争者。例如，在英国部分地区，用户密度已接近其极限，约为每平方公里0.35名客户，有分析师指出，目前的星链（Starlink）网络仅能支持大约20万个英国家庭（约占市场份额的0.7%）。同一份报告指出，利用星链现有基础设施所能实现的市场渗透率在德国约为0.4%，在西班牙则为1.4%。即使对其卫星星座进行全面升级至V2版本（而非当前混合使用的V1.5和V2卫星），在英国的渗透率也仅会适度提升至约1.4%（若实现计划中完整的1.5万颗卫星星座，有望达到

3%至4%的极限目标）。尽管未来V3版本卫星星座可能实现8%至10%的渗透率，但这可能需要超过十年时间以及重大的技术进步。截至2025年8月，尚无V3卫星在轨运行。⁵⁰

另一个对提升LEO和D2D整体容量以及特定区域内容量至关重要的因素是地面站，也称为网关。地面站是基础设施的关键组成部分，负责在大型数据中心与卫星之间中继数据、管理卫星网络并向卫星发送信号。到2025年，已有超过100个用于近地轨道的地面站，未来还需要新增约100个，以支持多个卫星星座的运行。⁵¹尽管许多LEO卫星开始采用激光通信，使在轨卫星能够与其他在轨卫星直接通信（而无需通过地面站转发所有信息），但这并不意味着对更多地面站的需求将会消失。⁵²最后，地面站应通过光纤连接到数据中心，以帮助最大化容量并最小化延迟，这可能为陆地光纤服务提供商（通常是电信公司）带来收入机会。

应对频谱管理中的监管考量

随着卫星通信市场的增长，频谱分配相关的监管问题将变得愈发重要。

德勤预测，提供D2D服务的LEO卫星网络将面临重大的监管挑战，主要原因在于这些网络需要在已分配给地面蜂窝服务的频段内运行。这种复杂性在美国和欧洲等地区尤为突出，这些地区的国家和区域监管机构对蜂窝频谱的分配实施严格管控，以防止干扰并确保频谱的公平使用。⁵³在美国，联邦通信委员会（FCC）实施了诸如“空间补充覆盖”（Supplemental Coverage from Space）框架等举措，旨在将卫星运营商与地面网络相整合，以促进D2D的连接。⁵⁴此外，美国国家电信和信息管理局（NTIA）针对425亿美元宽带公平、接入和部署（BEAD）计划发布的政策通知，标志着一种转变，将扩大对LEO卫星服务提供商的资金支持机会。⁵⁵这种技术中立的方法消除了对光纤的偏好，确立了基于性能的标准，使LEO卫星与传统宽带技术处于同等竞争地位，可能将对LEO卫星的资助从约40亿美元增加至100亿至200亿美元。⁵⁶

在欧洲，监管管理较为分散，各国家监管机构在欧盟和欧洲邮政与电信管理大会（CEPT）建立的框架内负责频谱分配。⁵⁷CEPT正在积极评估将卫星服务与地面移动网络集成所面临的技术和监管挑战。⁵⁸

在亚洲，类似的监管动态也存在，但由于各国政策多样以及基础设施发展水平不同，情况更为复杂。印度、中国和日本等国正在积极评估监管框架，以协调地面与卫星频率的使用，确保在促进创新与竞争的同时实现无干扰共存。例如，印度正通过印度电信管理局（TRAI）制定全面的指南，以有效管理频谱分配。⁵⁹在中国，正在实施重大的监管改革以适应卫星通信的发展。工业和信息化部（MIIT）已积极制定政策，以简化频率分配、管理频谱干扰，并推动卫星通信领域的创新。MIIT近期推出的举措包括一系列综合框架，旨在促进卫星服务与地面移动基础设施的融合，支持中国实现广泛数字化连接的战略目标。⁶⁰同样，日本正通过总务省（MIC）不断完善其监管框架。⁶¹

小结

需要考虑的D2D和LEO增长带来的一个影响涉及资本支出，不仅对航天公司如此，对地面连接服务提供商也是如此。德勤预测，到2026年底，对D2D卫星和LEO宽带星座的累计投资将达到约100亿美元⁶²——其中一些星座的某些卫星将具备D2D能力。自2019年以来，这100亿美元的投资已分多年投入，但即使将其全部集中在一年内支出，与截至2025年每年约3000亿美元的全球电信运营商资本支出相比，这一数额也显得微不足道。⁶³

对许多地面电信运营商而言，D2D和LEO合作模式之所以重要，是因为它们能以较低的资本支出（capex）应对持续不断的压力，即无论人口居住地多么偏远或地处农村，都需实现100%的网络覆盖。若采用有线或无线技术为这些地区提供服务，所需成本将比与基于太空的解决方案提供商合作（后者无需资本支出）甚至直接投资这些企业高出几个数量级。例如，AST SpaceMobile已从沃达丰（Vodafone）、AT&T和Verizon等全球巨头处募集资金，而其融资总额仅占这些公司年度资本支出的极小一部分。⁶⁴

值得注意的是，随着星座的老化，以及LEO卫星平均4到5年的使用寿命，资本支出可能会在较长时间内保持高位，每年需要更换星座中20%至25%的卫星。⁶⁵

基于卫星的宽带正成为某些传统地面服务的有力替代方案，尤其是在发展中地区。例如，在尼日利亚，一家LEO卫星服务提供商在进入市场仅两年后，已成为该国第二大互联网服务提供商。⁶⁶似乎有可能，单个LEO服务提供商，或更可能的是LEO服务提供商整体，会成为许多当前地面宽带连接水平较低的新兴市场中的最大服务提供商。

监管格局可能会发生重大变化，各国政府将在推动创新和市场竞争的同时，平衡国家安全与主权方面的关切。围绕频谱分配、轨道碎片管理和网络安全等方面，新的国际法规和标准有望陆续出台，以应对近地轨道环境快速扩张所带来的复杂挑战。公共应急通信系统以及应急服务与公共安全要求已经存在，并且因国家而异。设备制造商将需要遵守不同国家的各种应急通信和公共安全监管框架。

LEO运营商将需要与地面移动网络运营商（MNO）协商复杂的协议，采用频谱共享或租赁等策略，并在设计用于防止干扰的条件下开展合作。关键的监管问题包括采用动态频谱分配和地理波束成形等先进的干扰管理策略。⁶⁷在平衡地面运营商权利与卫星连接增强功能的同时，可能需要加强监管监督，并建立卫星运营商与地面服务提供商之间的强有力协作模式。

尽管我们关注的是消费者LEO宽带市场，但值得注意的是，未来几年很可能会出现一个规模庞大且稳健的企业市场，到2030年企业用户数量预计将增长近十倍，达到340万。⁶⁸尽管这一用户数量少于当前消费市场，但企业客户的月收入可能远高于个人消费者，且流失率更低。

Prashant Raman
India

Gillian Crossan
Global

Jody McDermott
Canada

Duncan Stewart
Canada

Tim Bottke
Germany

Ben Stanton
United Kingdom

ENDNOTES

1. Gartner, "[Gartner forecasts LEO satellite communications services spending to hit \\$14.8bn globally in 2026](#)," press release, July 30, 2025.
2. Deloitte analysis of publicly available market research and forecasts, combining current adoption trends, planned service launches, and demand in underserved regions to assess the feasibility of future subscriber growth.
3. Deloitte analysis of global low Earth orbit (LEO) satellite deployment trends indicates five major constellations—Starlink, Kuiper, Guowang, Honghu-3, and G60—will account for a significant proportion of the estimated 15,000 to 18,000 LEO satellites expected in orbit by the end of 2026. This projection aggregates operator-specific deployment targets, launch rate trends, and industry growth forecasts, referencing broker research and company filings.
4. This is based on a Deloitte analysis of publicly available industry data and forecasts, including current deployments as of mid-2025, announced launch schedules from major operators, and long-term projections from leading research providers. Estimates were derived by combining existing satellite counts with confirmed launch plans and aligning them with independent analysts' projections.
5. Yarnaphat Shaengchart and Tanpat Kraiwanit, "[Starlink satellite project impact on the Internet provider service in emerging economies](#)," Research in Globalization, May 4, 2023.
6. Nick Cowell, "[Satellite-based internet connectivity LEO Satellite Broadband](#)," Fujitsu, May 22, 2023.
7. Karen L. Jones and Audrey L. Allison, "[The great convergence and the future of satellite-enabled direct-to-device](#)," Center for Space Policy And Strategy, September 2023.
8. Joe Madden, "[The difference between NTN/D2D and satellite broadband – Madden](#)," Fierce Network, Jan. 16, 2024.
9. Christopher Baugh, "[Satellite direct-to-device: The characteristics of D2D constellations will limit SpaceX's ability to dominate](#)," Analysys Mason, July 22, 2024.
10. Mike Robuck, "[Musk outlines SpaceX D2D spectrum strategy](#)," Mobile World Live, Sept. 10, 2025.
11. Ibid.
12. These services include emergency messaging, basic data transmission, and sometimes voice calls.
13. David Jarvis, Duncan Stewart, Raghavan Alevoor, and Kevin Westcott, "[Signals from space: Direct-to-device satellite phone connectivity boosts coverage](#)," Deloitte Insights, Nov. 29, 2023.
14. Deloitte analysis of publicly available data on satellite industry investments for 2023–2024; investment amounts reflect disclosed funding rounds, commercial agreements, and reported capital commitments related to direct-to-device satellite communication.
15. Deloitte analysis of global direct-to-device satellite communication capital requirements, based on company filings, investor presentations, earnings call transcripts, government announcements, press releases, research reports, and expert interviews.
16. David Jarvis, Duncan Stewart, Raghavan Alevoor, and Kevin Westcott, "[Signals from space](#)."
17. Aamir Siddiqui and Andrew Grush, "[Android and iPhone satellite connectivity: What is it and what are your options right now?](#)" Android Authority, Feb. 11, 2025.
18. Arun Menon, "[Satellite industry trends to watch in 2024](#)," TM Forum, Jan. 31, 2024.
19. GSMA, "[Open consultation for the council working group on international internet related public policy issues](#)," August 2020.
20. Ibid.

21. GSMA, "New GSMA report shows mobile internet connectivity continues to grow globally but barriers for 3.45 billion unconnected people remain," press release, Oct. 23, 2024.
22. 5G Americas, "New developments and advances in 5G and NT," February 2025.
23. Amber Jackson, "Project Kuiper explained: Australia's bid to improve internet access with Amazon," Capacity, Aug. 5, 2025.
24. Ling Xin and Victoria Bela, "China launches first satellites for GuoWang project to rival SpaceX's Starlink," South China Morning Post, Dec. 16, 2024.
25. Mark Holmes, "Telesat's Lightspeed is now fully funded, MDA to build constellation," Via Satellite, Aug. 11, 2023.
26. Connectivity and Secure Communications, "ESA confirms kick-start of IRIS² with European Commission and SpaceRISE," Dec. 16, 2024.
27. SatNews, "Loft Orbital and Marlan Space to create the Middle East's first private manufacturing space company of commercial satellite constellations for LEO," Aug. 26, 2024.
28. Reuters, "Eutelsat announces contract with Airbus for 100 satellites," Dec. 17, 2024.
29. David Jarvis, Duncan Stewart, Kevin Westcott, and Ariane Bucaille, "Too congested before we're connected? Broadband satellites will need to navigate a crowded sky," Deloitte Insights, Nov. 30, 2022.
30. CCIA, "Low earth orbit (LEO) satellite broadband facts and stats," March 5, 2025.
31. Rick Mur, "Low-earth orbit (LEO) networks in your global connectivity strategy," GNX, Jan. 22, 2025.
32. Ibid.
33. Garinder Shankrowalia, "Amazon's ambitions: Project Kuiper and the complex future of satellite broadband," Omdia, May 20, 2025.
34. Rakuten.Today, "Moshi moshi? Space calling: Rakuten Mobile and AST SpaceMobile achieve Japan first satellite-to-mobile video call," May 2, 2025; Cameron Page, "Australia's TPG completes first D2D satellite trials with Lynk," TelcoTitans, May 8, 2025; John Tanner, "Globe kicks off Lynk Global D2D SMS tests in Zambales," Developing Telecoms, Oct. 7, 2024.
35. Nadine Hawkins, "Amazon to launch Project Kuiper satellites next week," Capacity Media, April 3, 2025.
36. Amazon, "Here's how Project Kuiper's satellite network can help telecom partners like Vodafone and Vodacom enhance reliability and extend reach," Sept. 5, 2023.
37. World Population Review, "Internet cost by country 2025," accessed Oct. 30, 2025.
38. Ibid.
39. Jack Kuhr, "Starlink Mini Impact and Rapid Terminal Iteration: Payload Research," Payload, June 26, 2024.
40. Hawkins, "Amazon to launch Project Kuiper satellites next week."
41. Elton Chang, "Satellite network capacity and scalability," TelecomWorld101, Jan. 17, 2025.
42. Ibid.
43. Kim Larsen, "The next frontier: LEO satellites for internet services," techneconomyblog, March 12, 2024.
44. Andrea D'Ambrosio, Miles Lifson, and Richard Linares, "The capacity of low earth orbit computed using source-sink modeling," arxiv, June 10, 2022.
45. Harry Baker, "How many satellites could fit in earth orbit? And how many do we really need?" LiveScience, May 30, 2025.

46. Kelly Hill, "FCC revisits satellite spectrum power levels," RCR Wireless News, May 1, 2025.
47. Dan Heming, "Starlink waitlists return, network congestion on the rise and finally, a customer support phone #," Mobile Internet Resource Center, Nov. 21, 2024.
48. Mark Jackson, "Starlink's satellite broadband hits capacity limit in South East England," ISPreview, Dec. 31, 2024.
49. Marcin Frąckiewicz, "Artificial intelligence in satellite and space systems," Tech Stock 2, June 12, 2025. Luis Manuel Garcés-Socarrás et al., "Artificial Intelligence implementation of onboard flexible payload and adaptive beamforming using commercial off-the-shelf devices," arXiv, May 3, 2025.
50. James Ratzer, "Starlink: What impact might it have on the telcos?" New Street Research, June 9, 2025.
51. Stella Linkson, "Starlink ground stations: What they are and how they work," Starlink Info, March 21, 2025; Shankrowalia, "Amazon's ambitions."
52. Linkson, "Starlink ground stations: What they are and how they work."
53. Kim Larsen, "Will LEO satellite direct-to-cell networks make terrestrial networks obsolete?" techneconomyblog, January 20, 2025.
54. Federal Communications Commission, "FCC advances supplemental coverage from space framework," March 15, 2024.
55. K. C. Halm, John C. Nelson Jr., and Kasey McGee, "NTIA revamps federally funded \$42.5 billion broadband deployment subsidy program," Davis Wright Tremaine LLP, June 12, 2025.
56. David Shepardson, "US Senate panel advances Trump nominee to oversee \$42-billion government internet fund," Reuters, April 9, 2025.
57. European Conference of Postal and Telecommunications Administrations, "An introduction to the European regulatory environment for radio equipment and spectrum," Feb. 5, 2024.
58. Commission for Communications Regulation, "Radio spectrum management operating plan for 2025–2028," Dec. 13, 2024.
59. ITU-APT Foundation of India, "Recommendations on telecom regulatory authority of India consultation paper on assignment of spectrum for space-based communication services," April 6, 2025.
60. Cetecom Advanced, "China introduces first regulatory framework for radar radio management," March 24, 2025.
61. Ministry of Internal Affairs and Communications, "Progress on the WX promotion strategy action plan," May 29, 2025.
62. Deloitte analysis of publicly reported or analyst-modelled 2026 investment details of major companies in the LEO space.
63. Peter Chahal, Avinash Naga, Courtney Munroe, Bruno Teyton, and Nikhil Batra, "Worldwide telecommunications capex forecast, 2025–2029," IDC Research, June 2025.
64. Newsroom, "AST SpaceMobile secures strategic investment from AT&T, Google and Vodafone," Business Wire, Jan. 18, 2024; Hema Kadia, "Verizon's \$100 million investment in AST SpaceMobile for satellite connectivity," TeckNexus, May 29, 2024.
65. Inside GNSS, "The case for LEO GNSS at C-Band," Feb. 3, 2025.
66. Damilare Dosunmu, "How Starlink took over Africa's largest internet market," Rest of world, April 15, 2025.
67. Larsen, "Will LEO satellite direct-to-cell networks make terrestrial networks obsolete?"
68. Pablo Tomasi, "Space to grow: Enterprise LEO forecast 2025–30," Omdia, Sept. 9, 2025.

ACKNOWLEDGMENTS

The authors would like to thank **Santosh Anoo, Dan Littmann, Jack Fritz, Sai Tarun Dronamraju, Girija Krishnamurthy, Michael Greco, Kathryn Walby, Paul Lee, and David Jarvis** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

部分移动用户消费新取向：赠礼狂欢 vs 千兆孤寂

发达国家市场的部分消费者难以察觉到网络性能的提升。电信公司应考虑推出更具创意的服务以扩大市场份额。

德勤预测，到2026年，在发达市场，移动运营商的奖励计划对主流消费者的重要性可能会与网络性能相当，甚至更胜一筹。在本十年剩余的时间里，随着网络升级持续进行，非网络福利在吸引用户或抑制用户流失方面的作用可能会愈发关键：一份玛格丽特披萨片可能比一份独立的5G（5G标准的完整版）套餐更具吸引力。¹前者实实在在，而后者往往超出了主流消费者的理解范畴。

这种奖励趋势似乎反映了发达市场移动网络日益成熟。尤其是从网络速度和延迟（网络响应速度）的角度来看，用户需求已基本得到满足。网络覆盖通常并不完美——存在无覆盖区域（无信号）和过度繁忙的热点区域（相对于可用容量而言，用户过多），但对于可能缺乏工具、理解力和耐心的消费者来说，对不同网络运营商的覆盖范围进行全面比较往往是一项过于艰巨的任务。

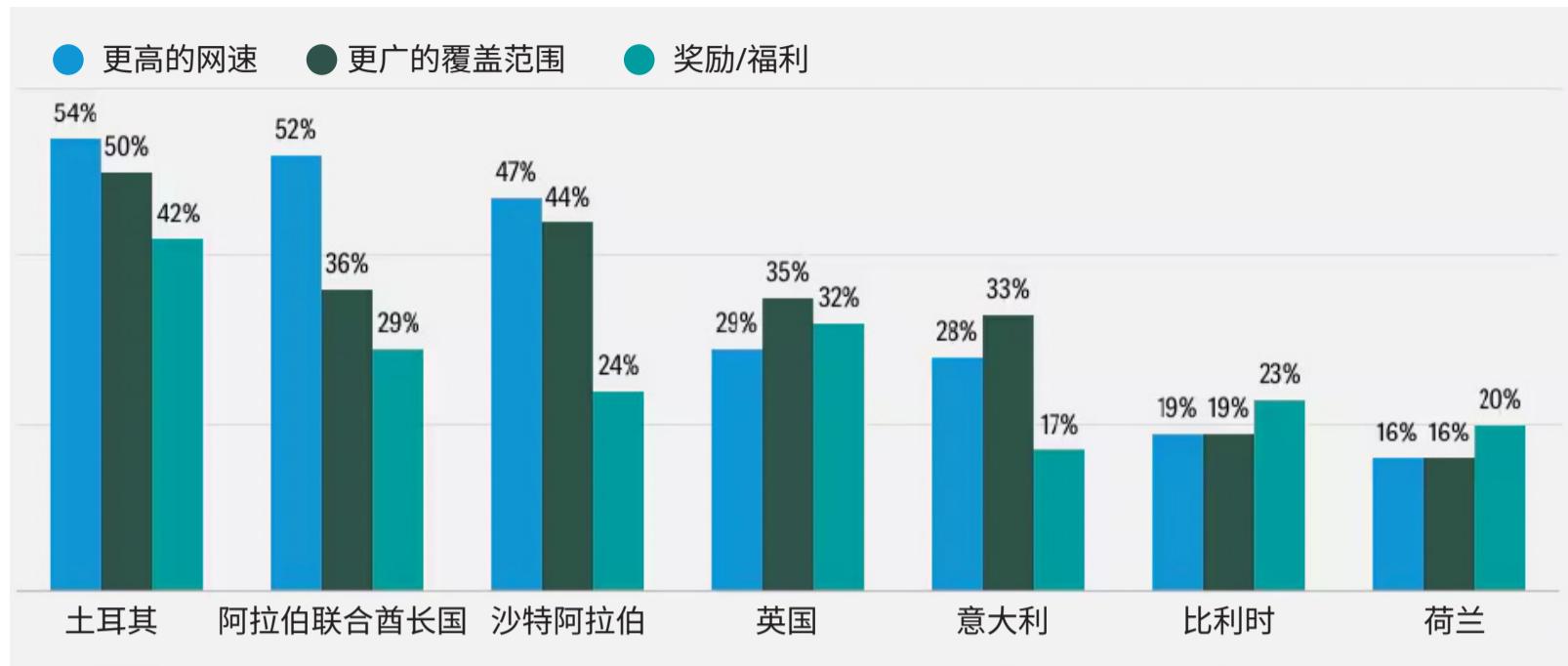
因此，以更高下行或上行速度、更低延迟为卖点的网络升级，可能对用户网络忠诚度的影响越来越小，因为许多用户既无法感知也无法评估此类升级的价值。同样，尽管停用2G和3G网络并将频谱重新分配给4G和5G网络具有重要意义，但用户可能也难以理解此举能给他们带来什么好处。

从网络升级转向基于奖励的差异化策略

德勤认为，各个市场在实现基于奖励的差异化进程中可能处于不同的阶段（图1），但大多数市场的发展方向可能是一致的。据报道，截至2024年，奖励机制在荷兰和比利时是导致用户流失的首要因素，在英国则位列第二（需注意的是，本研究未将价格因素纳入考量，而通常情况下，价格会是导致用户流失的首要宣称因素）。然而在其他市场，更高的网速或更广的覆盖范围则更为重要。²德勤预测，在此期间（至2030年），通过奖励等非网络方式实现差异化可能会变得越来越重要。

图1：多个市场中的消费者可能会为了获取奖励而更换运营商

2024年多个市场中促使受访消费者更换移动网络的因素及选择各因素的人数占比



问题：以下哪些因素（若有）会促使您更换移动网络运营商？

加权基数：拥有手机或智能手机、年龄在18至75岁之间的所有受访者，英国（3866人），荷兰（1944人），意大利（1913人），土耳其（973人），比利时（978人）；沙特阿拉伯（874人）和阿联酋（915人）的受访者年龄为18-50岁。来源：德勤英国

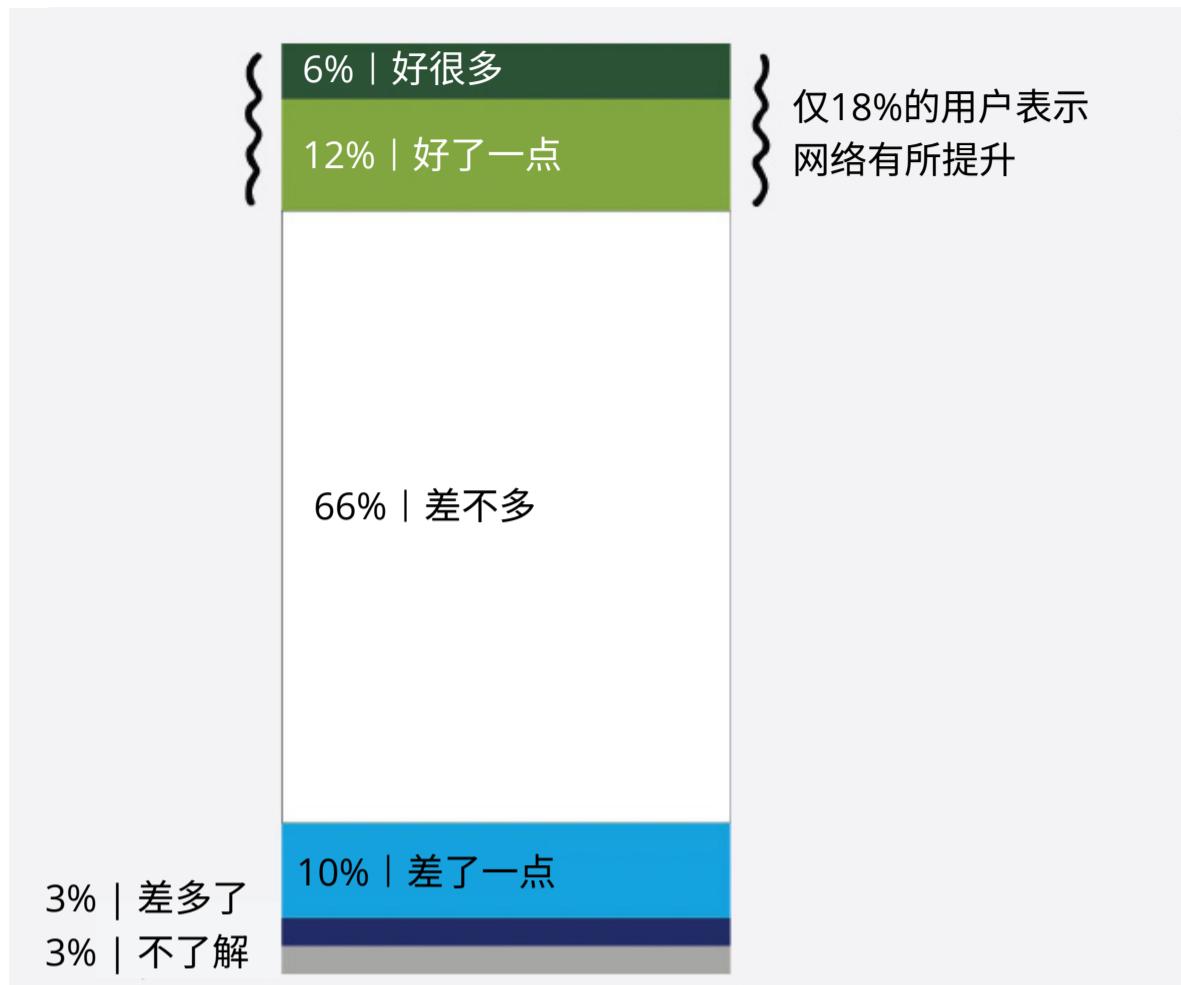
Deloitte. deloitte.com/insights

在某些情况下，网络升级可能会超出实际需求，市场上的所有运营商可能都会提供用户认为大致相当的网络性能。这与自20世纪70年代末以来长期存在的历史情况形成对比，当时几乎每一次代际升级都具有重要意义且显而易见。³例如，在2010年代初期，相比任何3G网络，4G升级带来了立竿见影的性能提升。⁴该技术在用户外出时实现了与“Wi-Fi”相当的网速和延迟（响应时间），使得原本在3G网络下表现不佳的搜索或导航等应用在4G网络下得以蓬勃发展。⁵

然而截至2025年底，几乎没有主流应用程序仅能在5G网络上运行。⁶因此，与4G时代相比，用户转向另一个声称拥有更先进5G网络的运营商的动机可能会大大降低。一些逐年改进的效果可能并不明显。例如，2024年至2025年间，英国移动网络的延迟仅改善（下降）了0.7毫秒，降至18.2毫秒。⁷（一毫秒等于千分之一秒）。0.7毫秒的差异人类无法察觉；即使是精英运动员，反应时间也大约需要140毫秒。⁸此外，几乎没有任何主流应用会从中受益（语音通话中150毫秒的延迟几乎难以察觉）。⁹例如，实时应用（如网络电话）需要100毫秒的延迟；而在英国，性能最慢的网络技术3G在2025年的平均延迟为42.3毫秒。¹⁰德勤英国分部当年的研究发现，英国受访的移动客户中有三分之二表示，他们感觉自己的网络服务与前一年相比没有任何差异（图2）。

图2：尽管网络有所改善，但受访用户中注意到这一改善的不足三分之一

三分之二（66%）的受访移动用户表示，在过去12个月中未察觉到自身网络存在差异



问：在过去12个月中，您认为您的移动互联网服务质量是变好了、变差了，还是基本保持不变？

加权基数：在英国拥有手机或智能手机、年龄在16至75岁之间的受访者（4023人）。

来源：德勤《2025数字消费者趋势》

Deloitte. deloitte.com/insights

此外，部分消费者可能难以比较本地市场的移动网络，这可能会削弱鼓励消费者转投性能更佳网络的营销活动的效果。

大多数用户的网络使用情况各不相同，出行模式和偏好的应用程序也各不相同。移动网络覆盖有相应的覆盖范围地图，但它们并未反映每个地点在各个时间点上的需求强度。¹¹用户可以通过同时使用两张SIM卡来并排比较两个网络，但这对大多数用户来说可能过于繁琐。

个人连接设备的巅峰时代或将近在咫尺

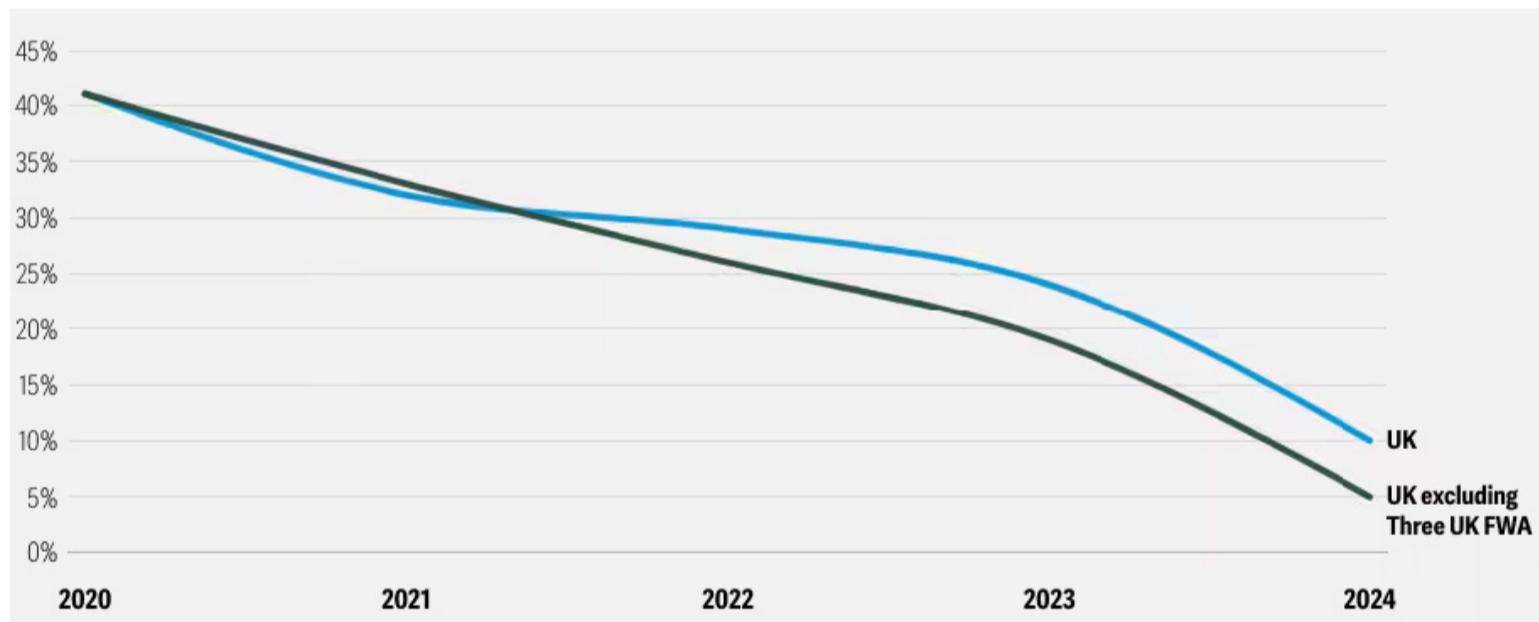
满足这一需求已耗时四十多年，但消费者连接方式的变革或许已接近尾声。¹²

虽然预测永远不应绝对，但未来中期（未来五年，即至2030年）内，不太可能再出现连接移动网络的全新根本性革命设备。同样，这些网络上可能也不会出现任何变革性的应用程序——尽管可能性不大，但主流用户向元宇宙迁移的情况仍有可能发生。最后，主流应用程序对连接的需求可能会保持稳定或有所下降。¹³

数据使用趋势似乎表明，使用模式正趋于稳定和可预测。在过去五年里，许多主要市场的每张SIM卡流量（以GB为单位）增速均有所放缓。截至2024年，在十个发达市场中，每张SIM卡流量的增速已降至个位数水平；¹⁴而在增速仍保持两位数的地区，这通常归因于通过专用固定无线接入（FWA）设备或智能手机热点共享功能，将蜂窝移动连接用于家庭宽带的用户数量实现了适度增长。例如，在英国，截至2024年，移动数据消费量的同比增幅已降至10%；然而，若排除专用固定无线接入设备的影响，则同比增幅降至5%（图3）。

图3：英国带宽消耗量的增速已显著放缓，其他市场也呈现出这一趋势

英国市场移动数据使用量同比增幅(单位:GB/月)



来源：德勤分析，基于恩德斯分析（Enders Analysis）的数据，该数据源自运营商官方数据及估算值。

Deloitte
Insights deloitte.com/insights

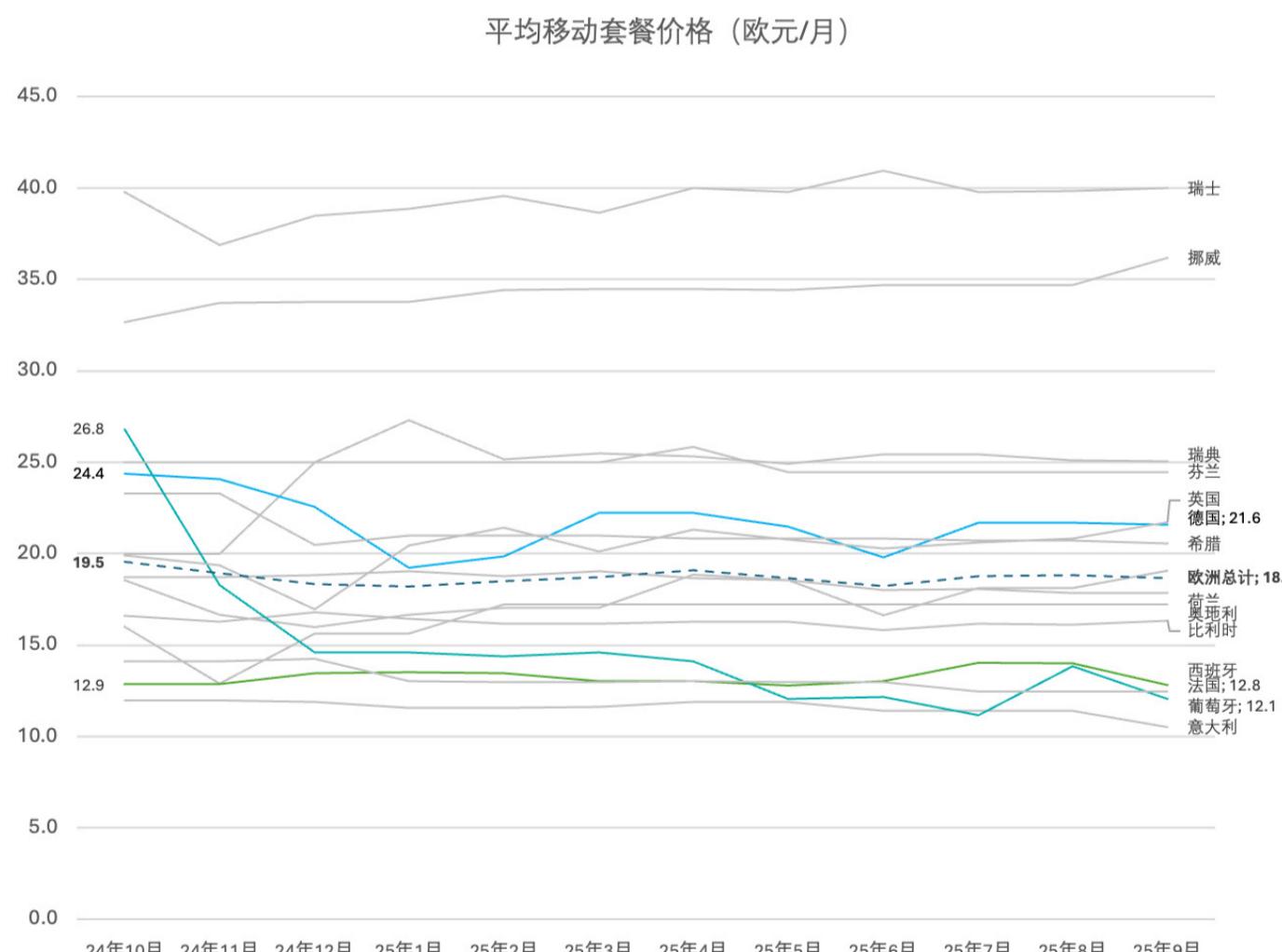
6G时代的差异化

如果消费者难以感受到5G的优势，那么推广6G的某些特性可能会更加困难。每一代新的移动网络通常遵循一个经验法则，即性能提升十倍或更高。¹⁵这将包括容量，可能在特定时间和地点（例如最大的音乐节或最繁忙的购物季）需要的容量。这也可能降低每千兆字节传输的成本，正如5G与4G之间的情况一样。¹⁶但它还将包括更高速度等因素。6G的规范可能在2026年最终确定，但相关技术的测试已经实现100 Gbit/s的传输速度。¹⁷约是5G峰值速度5 Gbit/s的20倍（这是每个小区的总速率，将在该小区内的用户之间共享。）

尽管6G可能提供更高的峰值速率，但需求可能保持稳定。向智能手机传输的典型高清视频流，每个连接通常需要低于5 Mbit/s的带宽。在未来几年中，由于压缩技术及其他因素降低了平均比特率，这一数值可能保持不变，甚至更有可能进一步下降。如果需求保持稳定，那么大规模升级至6G网络的资本回报可能会面临挑战，除非升级的主要目的是降低运营成本。

加强对奖励关注的另一个原因，可能是为了减少用户主要基于价格（业内也称为资费）的比较，例如每月10GB套餐的价格。在2024年至2025年间，某些市场的移动通信价格下降了高达50%（图4）。¹⁸如果一个套餐还包含免费咖啡和披萨等元素，则可能使得同类比较变得不太可行。

图4：部分市场的移动资费同比降幅扩大



来源：欧洲资费追踪机构（European Tariff Tracker）、新街研究（New Street Research），2025年

小结：会员忠诚度奖励计划或成破局关键

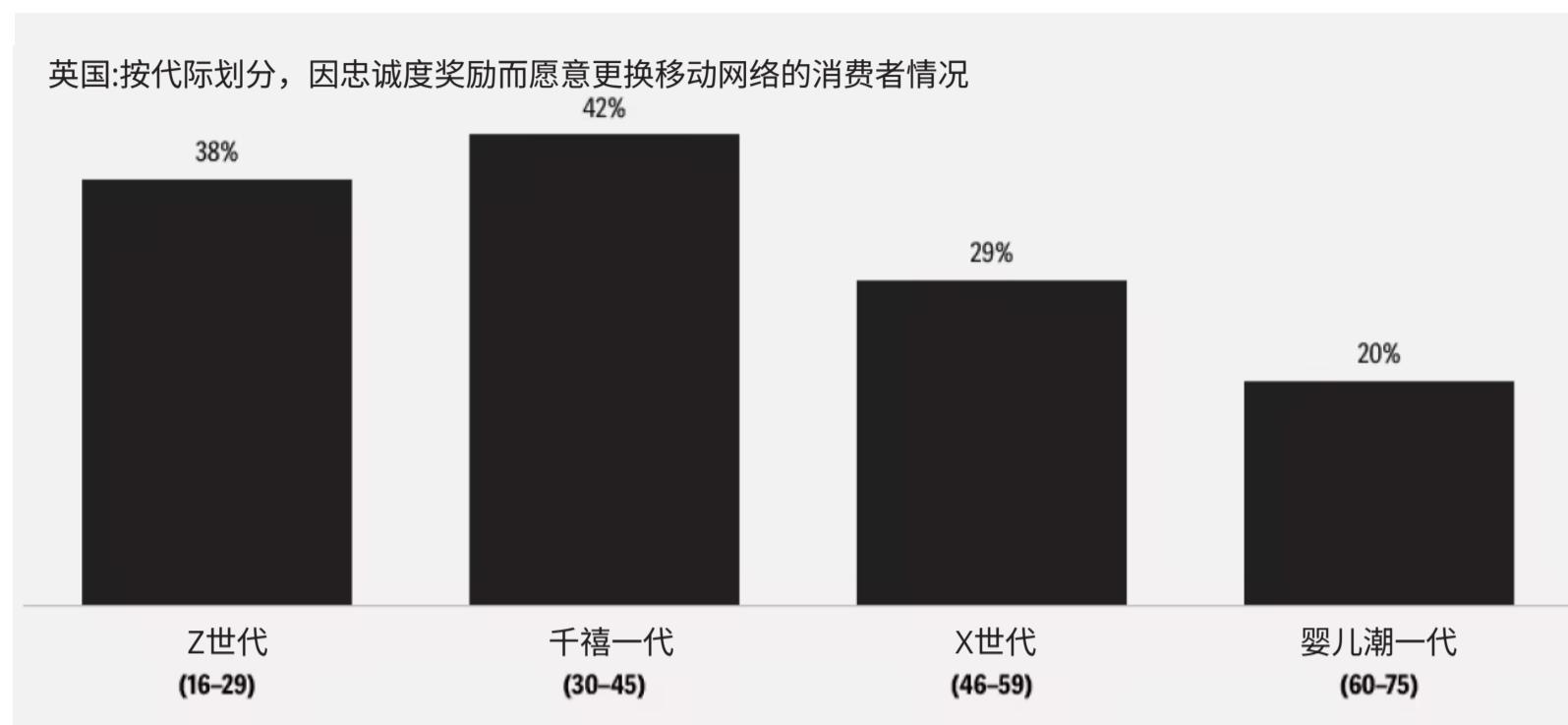
电信运营商的领导者应思考未来其网络将如何被选择，并审视这是否与过去存在重大差异。如果确实如此，则应据此作出调整。资本配置始终至关重要，而未来十年电信公司的局面可能大不相同。目前，投入资本回报率为7.3%，而加权平均资本成本为6.9%。¹⁹因此，从经济角度来看，电信投资几乎只能勉强保本。固定和移动数据使用量的放缓可能反而是件好事，使电信运营商能够避免在可能带来更高资本回报的网络升级上投入大量资金。

电信公司应注意，随着核心业务趋于成熟，许多其他行业已将奖励机制作为差异化竞争的重要手段。以航空业为例，该行业一度将超音速飞行视为增值亮点，如今却已大幅转向将奖励计划作为销售工具。航空公司的忠诚度计划估值已超过1000亿美元，仅三家航空公司的计划估值就超过200亿美元。²⁰自2019年以来，美国超过90%的普通信用卡消费均发生在带有奖励机制的信用卡上。²¹

电信运营商在投资非网络类福利时，应谨慎开展市场营销。一些电信运营商已采用高额的线上推广活动，通过电视、平面、广播、社交媒体和广告牌等渠道全面展示其会员权益。T-Mobile美国公司已庆祝用户累计领取了10亿份“感谢礼”，这些礼品包括食品、电影、汽油和旅行等。²²沃达丰英国公司通过其VeryMe计划累计发放了1.75亿项奖励。²³O2英国公司声称，其客户通过Priority计划在一年内节省了2300万英镑。²⁴

运营商应考虑到，相较于网络性能，Z世代和千禧一代用户可能更倾向于接受附加福利类的优惠。一位二十多岁的用户可能并未体验过3G网络（2000年代最先进的技术）的迟缓，其主要使用的是4G网络，且对5G的感知差异不大；而一位四十多岁的用户可能从未经历过在2G网络（1990年代最先进的网络）上浏览网页的困难。因此，不同用户群体在选择服务时可能更关注不同的差异化优势。根据德勤英国的研究，受访的Z世代和千禧一代用户因忠诚度奖励而更换网络运营商的倾向性高于年长群体（图5）。

图5：据报道，在受访的消费者中，与其它世代相比，Z世代和千禧一代中更倾向于看重额外福利而非性能表现的比例更高



问题：以下哪些因素（如有）会促使您更换移动网络运营商？[忠诚度奖励/福利]

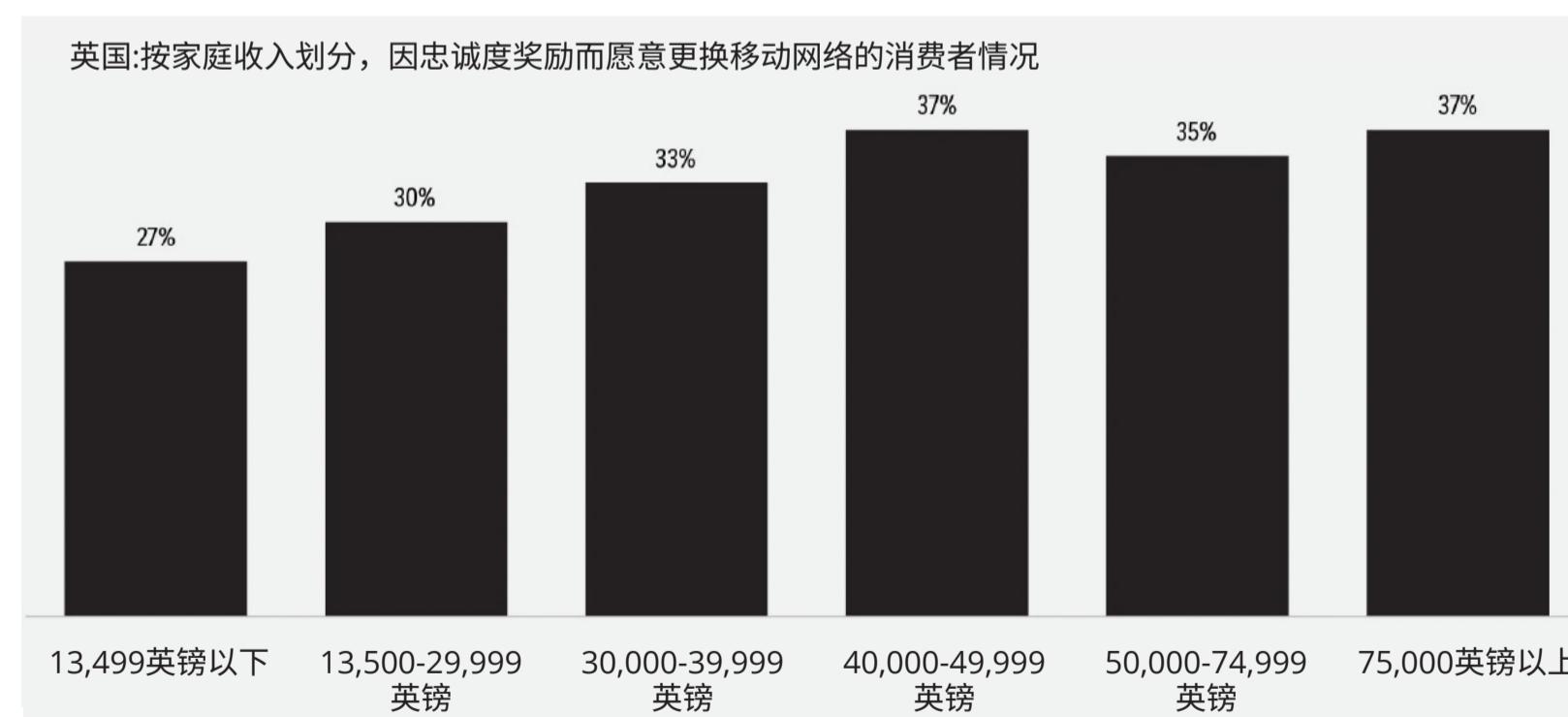
加权基数：在英国拥有手机或智能手机、年龄在16至75岁之间的所有受访者（4023人）。

来源：德勤《2025数字消费者趋势》

Deloitte. Insights deloitte.com/insights

运营商应注意，奖励措施可能对高消费用户更具吸引力。收入较高的用户相比低收入用户，更倾向于为更优惠的条件而更换运营商（图6）。提供“赠品”可能会引发强烈甚至略带非理性的积极情绪反应。²⁵

图6：高收入消费者可能更倾向于选择额外福利



问题：以下哪些因素（若有）会促使您更换移动网络运营商？【忠诚度奖励/福利】

加权基数：英国16至75岁拥有手机或智能手机的所有受访者（4023人）。

来源：德勤《2025数字消费者趋势》

Deloitte. Insights deloitte.com/insights

如果大型电信运营商采取类似的策略，一个风险可能是奖励将变得商品化，就像连接服务一样。此外，随着银行和公用事业等其他行业建立各自的计划，市场可能进一步饱和。²⁶消费者对于愿意免费享用的咖啡数量可能存在上限，而如果多家服务提供商提供相同的优惠，其吸引力可能会被削弱。因此，打造独特且差异化的方案将对吸引客户和降低流失率起到关键作用。这意味着现场活动、音乐会和体育赛事可能成为极具吸引力的资源。²⁷但这些优惠的覆盖范围可能有限，仅惠及数千万客户中的数万名客户数百万。

Paul Lee

United Kingdom

Tim Bottke

Germany

Dieter Trimmel

Germany

Ben Stanton

United Kingdom

Jody McDermott

Canada

Jack Fritz

United States

ENDNOTES

1. GSMA, [5G Network Slicing](#), accessed October 2025.
2. Paul Lee and Ben Stanton, [Deloitte Digital Consumer Trends 2025, UK Edition](#), Deloitte LLP, June 2025; Adrie Cronje et al., [Digital Consumer Trends 2024, Netherlands Edition](#), Deloitte LLP, December 2024; Vincent Frosty and Vincent Pirard, [Digital Consumer Trends 2024, Belgium Edition](#), Deloitte Belgium, 2024.
3. Vodafone UK, [2G](#), accessed October 2025.
4. Simon Thomas, "What is the difference between 3G and 4G?," 4G.co.uk, September 29, 2014.
5. Ivor Nicholls, "LTE vs 4G: Understanding the difference between LTE and 4G," UCtel, February 10, 2025.
6. Andrew Wooden, "The telecoms industry's biggest problem? Failure to monetise 5G," Telecoms.com, March 14, 2024.
7. Ofcom, [Mobile matters](#), July 17, 2025.
8. Espen Tønnessen, Thomas Haugen, and Shaher Ahmmad Ibrahim Shalfawi, "Reaction time aspects of elite sprinters in athletic world championships," *Journal of Strength & Conditioning Research* 27, no. 4 (2013): pp. 885–92.
9. IR, "Network latency—Common causes and best solutions," accessed October 2025.
10. Ofcom, [Mobile matters](#).
11. Ofcom, "Improving your mobile phone reception," May 27, 2022.
12. William Webb, "It's time to rethink 6G," IEEE Spectrum, February 10, 2025.
13. Netflix, "Netflix-recommended internet speeds," accessed October 2025; Paul Lee, Dieter Trimmel, and Eytan Hallside, "No bump to bitrates for digital apps in the near term: Is a period of enough fixed broadband connectivity approaching?," *TMT Predictions 2024*, Deloitte, November 29, 2023.
14. Tefficient, "The demand for additional mobile data is weaker than ever—ARPU growth softens," July 31, 2025.
15. Michael Irving, "'Ultrabroadband' 6G chip clocks speeds 10 times faster than 5G," ScienceAlert, September 3, 2025; 4G.co.uk, "How fast are 4G and 5G?," accessed October 2025.
16. ETTelecom.com, "5G will make cost of GB lower than 4G: Experts," July 31, 2020.
17. NTT DOCOMO, "DOCOMO, NTT, NEC and Fujitsu develop top-level sub-terahertz 6G device capable of ultra-high-speed 100 Gbps transmission," press release, April 11, 2024.
18. New Street Research, [Europea Tariff Tracker](#), accessed October 2024.
19. Jennifer Johnson, "Peak data growth is a quiet win for telcos," Reuters, June 2, 2025.
20. Evert de Boer and Xiao Yao Chin, Top 100 most valuable airline loyalty programs, On Point Loyalty, January 2023.
21. Consumer Financial Protection Bureau (CFPB), "CFPB takes action on bait-and-switch credit card rewards tactics," news release, last modified December 18, 2024.
22. Mike Sievert, "The power of appreciation: Taking customer loyalty to the next level," Un-carrier blog, T-Mobile, February 13, 2024.
23. Vodafone UK, "Spin for a chance to win £1,000 each day with VeryMe Rewards," press release, June 9, 2025.
24. Virgin Media O2, "Priority from O2 launches 'Blue Mondays' with millions of unmissable rewards, prizes and experiences for customers," April 28, 2025.

25. CI Group, "[The psychology of freebies: Why small rewards yield big returns](#)," accessed October 2025.
 26. Octopus Energy, "[Octoplus, our rewards programme for smart meter customers](#)," accessed October 2025.
 27. Vodafone UK, "[Music festivals](#)," accessed October 2025.
-

ACKNOWLEDGMENTS

The authors would like to thank **Matt Roberts, Pedro Goncalo Sanguinho, Dan Littman, James Brass, Michele Gabriel, Jolyon Barker, Dan Adams, Matt McDermott, Giles Warner, Jan-Piet Nelissen, Jonas Malmlund, Duncan Stewart, Gizem Bozdag, and Ralf Esser** for their contributions to this article.

Cover image by: **Jaime Austin**; Adobe Stock

德勤中国联系人

程中

科技、传媒和电信行业主管合伙人
电信、传媒及娱乐行业主管合伙人
电信、传媒及娱乐行业咨询业务主管合伙人
电邮：zhongcheng@deloittecn.com.cn

陈颂

半导体行业主管合伙人
科技、传媒和电信行业审计及鉴证合伙人
电邮：leoschen@deloittecn.com.cn

王佳

科技、传媒和电信行业税务与商务咨询主管合伙人
电邮：jeswang@deloittecn.com.cn

陈耀邦

科技、传媒和电信行业华南区主管合伙人
电邮：ybchan@deloitte.com.hk

钟昀泰

科技、传媒和电信行业研究总监
电邮：rochung@deloittecn.com.cn

李艳

科技、传媒和电信行业助理经理
电邮：lavli@deloittecn.com.cn

廉勋晓

科技行业主管合伙人
科技、传媒和电信行业审计及鉴证主管合伙人
电邮：mlian@deloittecn.com.cn

张森

科技行业咨询业务主管合伙人
科技、传媒和电信行业华北区主管合伙人
电邮：jasonzhang@deloittecn.com.cn

濮清璐

科技、传媒和电信行业华东区及华西区主管合伙人
电邮：qlpu@deloittecn.com.cn

李宝芝

电信、传媒及娱乐行业华南区主管合伙人
电邮：pollee@deloitte.com.hk

周立彦

科技、传媒和电信行业高级经理
电邮：liyzhou@deloittecn.com.cn

关于德勤

德勤中国是一家立足本土、连接全球的综合性专业服务机构，由德勤中国的合伙人共同拥有，始终服务于中国改革开放和经济建设的前沿。我们的办公室遍布中国31个城市，现有超过2万名专业人才，向客户提供审计、税务、咨询等全球领先的一站式专业服务。

我们诚信为本，坚守质量，勇于创新，以卓越的专业能力、丰富的行业洞察和智慧的技术解决方案，助力各行各业的客户与合作伙伴把握机遇，应对挑战，实现世界一流的高质量发展目标。

德勤品牌始于1845年，其中文名称“德勤”于1978年启用，寓意“敬德修业，业精于勤”。德勤全球专业网络的成员机构遍布150多个国家或地区，以“因我不同，成就非凡”为宗旨，为资本市场增强公众信任，为客户转型升级赋能，为人才激活迎接未来的能力，为更繁荣的经济、更公平的社会和可持续的世界开拓前行。

Deloitte（“德勤”）泛指一家或多家德勤有限公司，以及其全球成员所网络和它们的关联机构（统称为“德勤组织”）。德勤有限公司（又称“德勤全球”）及其每一家成员所和它们的关联机构均为具有独立法律地位的法律实体，相互之间不因第三方而承担任何责任或约束对方。德勤有限公司及其每一家成员所和它们的关联机构仅对自身行为承担责任，而对相互的行为不承担任何法律责任。德勤有限公司并不向客户提供服务。请参阅www.deloitte.com/cn/about了解更多信息。

德勤亚太有限公司（一家担保责任有限公司，是境外设立有限责任公司的其中一种形式，成员以其所担保的金额为限对公司承担责任）是德勤有限公司的成员所。德勤亚太有限公司的每一家成员及其关联机构均为具有独立法律地位的法律实体，在亚太地区超过100个城市提供专业服务，包括奥克兰、曼谷、北京、班加罗尔、河内、香港、雅加达、吉隆坡、马尼拉、墨尔本、孟买、新德里、大阪、首尔、上海、新加坡、悉尼、台北和东京。

本通讯中所含内容乃一般性信息，任何德勤有限公司、其全球成员所网络或它们的关联机构并不因此构成提供任何专业建议或服务。在作出任何可能影响您的财务或业务的决策或采取任何相关行动前，您应咨询合资格的专业顾问。

我们并未对本通讯所含信息的准确性或完整性作出任何（明示或暗示）陈述、保证或承诺。任何德勤有限公司、其成员所、关联机构、员工或代理方均不对任何方因使用本通讯而直接或间接导致的任何损失或损害承担责任。

© 2026。欲了解更多信息，请联系德勤中国。

CQ-017-CN-25



这是环保纸印刷品