Deloitte.
Insights

2025科技、传媒和电信行业预测

Deloitte.
Insights

2025科技、传媒和电信行业预测：弥合差距

德勤预测，2025年将是生成式人工智能（简称“生成式AI”）与科技、传媒和电信（TMT）行业迎来重大转折的“间隔年”，这一年将凸显出弥合关键差距以解锁当今潜力的迫切需求。

展望未来，至2025年及以后，TMT行业即将实现重大飞跃，这一飞跃在很大程度上得益于生成式AI的迅速普及。然而，要实现这一愿景，行业还需要弥合以下差距：平衡生成式AI基础设施投资与商业化进程、解决生成式AI使用中的性别差异、管理生成式AI数据中心的能耗、应对公众对深度伪造内容的信任问题、探索生成式AI在媒体和游戏领域的最佳应用、利用生成式智能体实现实时管理与行动，以及填补流媒体视频和云支出方面的缺口。同时，还有一些非差距性预测值得关注，包括搭载生成式AI芯片的新型智能手机和个人电脑、提升观众体验的新场馆与体育基础设施，以及电信运营商的整合（特别是无线运营商）。克服这些障碍对于企业和行业的繁荣发展至关重要。

标志着2025年成为TMT行业“间隔年”的关键差距

1. **生成式AI数据中心电力与可持续性差距**。拟建的生成式AI数据中心电力需求急剧增长，并寻求低碳电力，这在其需求与电网容量及公司可持续发展目标之间产生了差距。尽管全球的超大规模企业、芯片公司和公用事业公司正在努力弥合这一差距，但预计2025年这一差距仍将存在。

2. **生成式AI的性别差距**。相较于男性，女性在工作和娱乐中使用生成式AI工具的可能性较低。部分原因是缺乏信任，但在某些市场，女性对生成式AI的使用率有望在年内赶上男性。

3. **生成式AI深度伪造信任差距**。深度伪造生成式AI内容（图像、视频和音频）的泛滥导致消费者信任度下降。生成式AI生态系统需全面且不可篡改地标注内容，并可靠、准确地实时检测虚假图像。创建可信的深度伪造内容的边际成本正在不断下降，而检测成本需要以同等速度下降，以帮助弥合这一差距。

4. **制片公司使用生成式AI的差距**。许多人期望大型制片公司使用生成式AI进行内容制作，且部分公司已付诸实践，但预期与现实之间仍存在差距。许多制片公司对生成内容固有的知识产权挑战持谨慎态度，但他们渴望获得企业能力，以缩短时间、降低成本并扩大影响力。

5. **自主生成式智能体 (Autonomous Gen AI agents)差距**。能够持续可靠地完成离散任务并协调整个工作流程的自主机器人极具吸引力。2024年已启动代理式人工智能 (Agentic AI)试点项目——它们能否在2025年实现广泛应用？

6. **流媒体视频差距**。许多媒体和娱乐公司认为消费者会“购买并持有”多个订阅服务。然而，消费者正在通过捆绑自己喜爱的订阅并放弃其他订阅来降低成本。我们发现，每个家庭的服务数量不仅停滞不前，还在减少，流媒体公司越来越依赖捆绑服务来填补增长缺口，并利用第三方来聚合和分发其内容。

7. **云支出差距**。使用云的最初卖点之一是其成本更低，但实际上，支出往往是分散且难以控制的。一些买家正在利用“FinOps” （一套衡量和优化云支出的工具和战略）来弥补承诺的成本节约与当前支出之间的差距，以管理其云支出并节省数十亿美元。

今年新增内容

今年，我们推出了两个新部分，其中包含10个微型预测。在 “最新动态”部分，我们回顾了之前TMT行业预测报告中的六个主题，同时还探讨了这些主题的最新预测。另外，在“上升趋势”部分，我们关注了TMT领域的四个前沿话题，尽管这些新兴主题可能尚未成为主流预测的一部分，但我们坚信它们将成为行业即将流行的焦点。

女性与生成式人工智能：使用鸿沟快速弥合，信任差距仍然存在

要想女性真正享有生成式AI的红利，科技公司应努力增强信任，减少偏见，为其提供更多相关工作岗位。

德勤预测，到2025年底，美国体验和使用生成式AI技术的女性用户人数将与男性用户持平甚至实现超越。虽然2023年女性用户仅为男性用户的一半，但据其采用速度推测，她们很可能在明年内实现齐平，并在一到两年内在许多欧洲国家实现平等。尽管采用速度在加快，但女性对人工智能供应商是否会保护她们的数据安全表示不太信任——这可能会阻碍她们全面参与人工智能，并影响她们在人工智能方面的支出。为了克服这一问题，科技公司应加强数据安全和数据管理实践，减少人工智能偏见，并提高女性在该领域的参与度。

随着生成式人工智能功耗日增，数据中心寻求更绿色可靠的能源解决方案

科技行业应优化基础设施、创新芯片设计，并与电力提供商合作，助力数据中心实现未来可持续发展。

人工智能数据中心用电量将持续攀升，德勤预测，到2025年，数据中心用电量大约占全球用电量的2%。预计这一增长源自高密度数据中心基础设施，以支持庞大的计算能力和冷却需求。然而，监管、基础设施和成本问题正给发电和电网带来挑战，使其难以满足数据中心对全天候可靠能源前所未有的需求。技术和电力行业可以通过增加无碳能源的使用、提高生成式AI芯片和算法中的能效以及重新平衡计算密集型AI工作负载来共同应对这些问题。

雄心勃勃的体育场馆项目旨在弥合公共投资和私人投资目标之间的差距

体育场馆所有者致力于将体育场馆改造为可促进社会经济增长、推动社区参与和实现收入多元化的增长点。

体育产业已多次证明其有望成为经济和社会发展的催化剂，体育场馆大多位于社区的中心。对体育基础设施的投资呈上升趋势，原因是该等开发项目常会为公共和私人部门带来广泛的社会经济效益。政府和社区以增长为共同目标，可与体育投资者合作提供交通枢纽和社区资源等配套基础设施，助力提升体育运动的社会经济影响，提高球迷参与度，并使体育组织的收入来源多样化。我们预计，2025 年新的体育场馆开发项目将继续加快步伐，预计近50%的新建体育场馆基础设施项目将在北美和欧洲落地。

处于研发阶段的自主生成式智能体

自主生成式智能体 (Autonomous Gen AI agents)——亦称作 “代理式人工智能 (Agentic AI)”，不仅能提升知识工作者的工作效率，还可实现各类工作流程的高效运作。然而，由于代理式人工智能的 “自主性” 特征，其广泛应用尚待时日。

德勤预计，到2025年，在已部署生成式AI的企业中，将有25%的企业开展代理式人工智能的试点项目或进行概念验证。到2027年，这一比例将增至50%。代理式人工智能可以自主完成复杂的任务，提高知识工作者的生产力和效率。目前最有前景的应用包括软件开发、客户支持、网络安全和监管合规。代理式人工智能进步迅速，但与大多数新技术一样，广泛应用尚需时日。尽管如此，2025 年，在某些行业和用例中，一些代理式AI应用程序将被实际应用于现有工作流程。

深度伪造之战：网络安全的大规模挑战与深远影响

随着检测和打击虚假内容的力度持续加大，维护可信互联网的成本或由消费者、创作者及广告商共担。

随着人工智能生成越来越多的在线图像和视频，围绕内容真实性和虚假内容的潜在危害的问题变得越来越紧迫。社交平台、科技公司和媒体机构正在开展跨界合作：利用技术（通常是人工智能）检测和标记假冒内容，以及利用加密元数据确保真实媒体内容的出处。德勤预计，该市场的发展轨迹或与网络安全行业相仿，媒体公司及技术提供商将通过投资验证技术并与各方合作，以领先于不断进化的伪造手段。

云服务的精益管理：“FinOps” 让每一分钱发挥最大效益

随着企业云支出不断增加，运用FinOps策略能让每笔投入的价值回报最大化，实现成本节约、价值提升及跨部门协同增效。

2025年，全球云支出预计将达到8,250亿美元，[1]但企业高层可能难以言明具体的支出细节。不过，德勤预测，2025 年采用 “FinOps” 可为企业节省约210亿美元。企业可以先关注初步措施，采取行动减少云浪费，优化资源配置，并积极调整计算、网络和存储的规模。但是，经验丰富的公司也可以推动文化变革，例如确立跨部门的责任和财务问责制。我们的目标是创建 “云单位经济效益” 模型——确保每一笔云支出都与其产生的业务价值相对应，助力企业做出有效的IT决策。

端侧生成式人工智能能否助力智能手机市场复兴

凭借特殊芯片和移动操作系统的广泛集成，智能手机可真正实现大智慧，但用户是否愿意为智能交互的革新买单？

德勤预测，2025年全球智能手机出货量将将从 2024 年的5%年增长率小幅提升至 7%左右。部分增长量可能来自换机周期（过去两年的换机周期有所下降），另一部分增长动力是早期用户对端侧生成式AI的热情追捧。智能手机搭载生成式AI功能将验证智能助手、对话界面等功能的实用性，展现设备上运行小型模型的能力，并探索在生成式AI高资本投入下实现经济价值的商业模式。尽管生成式AI前景广阔，但其能否真正实现潜力，以及用户是否愿意接纳这种与普及度极高的消费电子产品的全新交互方式，仍有待时间验证。

生成式人工智能用于内容创作，大型制片公司犹豫不决，社交媒体积极推进

好莱坞（以及其他电影制片公司）可能会谨慎采用生成式AI技术进行内容创作，但或会率先将其用于运营和发行。

德勤预测，2025年，各影视巨头公司仍将对采用生成式AI进行内容创作持谨慎态度，在此方面只投入不到3%的制作预算。我们也预测，运营支出将增加10%，用于使用生成式AI工具执行合同和人才管理、许可与规划、营销和广告以及内容本地化和配音等工作，以此帮助扩大制片公司在全球各个市场的影响力。这种方法可以帮助制片公司减少生成式AI技术在人才和内容创作上的影响，同时加速利用该技术降本增效。

重新评估直接面向消费者（DTC）模式：转向视频聚合商

视频内容创作者或需更多经销商来扩大可触达市场规模。

德勤预测，订阅视频点播（SVOD）的“堆叠”现象——即消费者订阅多个独立视频点播服务——将在2025年减少。根据德勤的调查，在 2023 年和 2024 年，美国每个用户的订阅数约为4个，而在欧洲市场高于2个，目前来看，我们已经度过了这一峰值，在大多数市场中，独立流媒体服务的数量将缓慢下降。消费者将不再直接订购单一内容提供商的服务，而是更倾向于使用聚合服务，即从电信公司、付费电视、技术平台到流媒体本身等中间商将多个内容源整合到一个套餐中。这一趋势很可能对众多参与者而言都是有利的，有助于控制成本，并为2025年及以后的流媒体市场创造一个稳定且可持续的生态系统。

监管放宽助力，无线电信市场整合提速

在许多市场，小型无线电信公司面临增长缓慢、利润低下以及债务压力。并购活动，尤其是资产整合乃至整个面向消费者的公司合并，在获得监管机构批准的情况下，或能带来转机。

在一些市场，尤其是欧洲和亚洲，无线电信市场过于分散，一些参与者规模过小，没有能力长期投资网络，且难以为继。尽管这些市场历来保持着较多的运营商数量，但最近的讨论中出现了允许甚至鼓励整合的机会。德勤预测，尽管这一过程预计会比较缓慢，并且监管机构会提出相应条件，但从2025年开始，整合的步伐将会加快，并将持续进行，从而创建一个更具可行性和可持续性的无线生态系统，特别是在较小规模的市场中。

最新动态

今年，我们将回顾以往的六个预测，看看我们的预测结果如何，以及最新的发展动态是什么：

生成式人工智能走向企业边缘：“本地部署人工智能”盛行

搭建企业内部服务器，以构建更加私密、安全、灵活且低成本的人工智能信息技术环境。

德勤预测，2025年，尽管云端生成式AI仍将是主流选择，但全球约半数企业将在本地增设人工智能数据中心基础设施，主要是为了保护其知识产权和敏感数据，遵守数据主权或其他法规，以及节省开支。德勤2024年第二季度《企业生成式人工智能现状》调查显示，人工智能专业水平极高的企业中，有80%表示在云端人工智能上投入了更多资金……但61%的企业表示在自身硬件上投入了更多资金。企业最终可能会倾向于采用混合模式部署生成式AI，即结合云端与本地资源进行操作。

（重新）确定女子赛事的投资案例

女子赛事收入不断增长，投资者热情高涨，收入估值创下纪录。

全球女子体育运动的职业化和商业化程度日益提高，正吸引着体育迷、赞助商以及至关重要的投资者的关注。2024年，我们预测女子精英体育市场的收入将超过10亿美元。在北美，俱乐部的估值屡创新高，包括美国国家女子足球联赛的洛杉矶天使城足球俱乐部（估值2.5亿美元）[2]和美国女子职业篮球联赛的拉斯维加斯王牌队（估值1.4亿美元）[3]。在其他地区，各组织正在创建创新性的结构，以便为女子运动队注入投资，并强调战略增长、独立领导地位和商业机会。2025年，我们预计包括机构投资者、私募股权和高净值个人在内的投资者群体将进一步扩大，并对女子体育领域给予更多关注。

固定无线接入（FWA）：与普遍观点相反，FWA采用率或将持续增长

美国FWA净增用户数可能略低于去年，而部分市场的FWA净增用户数或将到2026年才会实现高速增长……无论美国或是全球，FWA净增用户数均存在未实现增长或潜在增长。

固定无线接入（FWA），即消费者通过固定蜂窝设备（主要是5G）而非电线获得家庭宽带服务，在美国5G发展中发挥着重要作用。预计到2024年底，将有超1,000万户家庭接入。然而，其增长速度正在放缓，2024年第一季度的净增用户数低于2023年第一季度，且预计2025年将进一步放缓。尽管如此，企业将越来越多地使用FWA，除美国和印度两个规模较大的FWA市场，其他市场的净增用户数仍将超数百万。德勤预测，2025年和2026年全球FWA净增用户数将继续以每年约20%的速度增长。（与德勤2022年对FWA的预测一致）。

5G独立组网进展缓慢：6G到来会否延期？

面对投资回报疑虑，电信公司重新评估5G独立组网投资，6G推进或受影响。

5G独立组网（SA）网络的部署进展比最初预期的要慢。电信公司可能因现有5G投资回报不理想而对下一代5G技术的大规模投资持谨慎态度，这使得6G的推出似乎愈发遥远。2022年，德勤全球预测，到2023年底，投资5G独立组网网络的电信运营商数量将从2022年的100多家翻倍至少200家，但这一情况并未发生：截至2024年3月，在全球已推出5G服务的585家运营商中，仅有49家部署、推出或试运行了5G独立组网网络。[4]德勤预测，到2025年，新升级到独立组网的网络数量将不足20个，5G SA在所有5G部署中的占比将保持在12%左右。

开放式无线接入网（RAN）移动网络与供应商选择：目前采用单一供应商模式，何时实现多供应商模式？

开放式RAN迈向多元化、多供应商生态系统进程缓慢，且面临错综复杂的挑战。

开放式无线接入网 (Open RAN) 旨在为构建RAN的移动网络运营商 (MNO) 提供更多选择并提高其灵活性，以实现网络民主化。2021年，德勤预测全球动态公共网络开放式RAN部署数量将从35个增至70个。我们当时过于乐观：截至2024年3月，全球正在部署和试验的公共网络开放式RAN数量为45个，仅有两个网络为多供应商开放式RAN。[5]向多元化、多供应商生态系统过渡比一些人最初预计的更加缓慢和复杂，实现真正的多供应商开放式 RAN仍需时日。德勤预测2025年将不再部署或公布其他多供应商开放式RAN网络。

量子技术起步虽慢，网络安全防御不容滞后

量子药物发现与金融建模尚待时日，但量子时代的网络防御升级却刻不容缓。

正如德勤在过往报告中预测的那样，量子计算机目前仍处于研发阶段，至少在现阶段，它们能提供计算优势的现实应用场景很少。但是，“先窃取、后解密”攻击的威胁已达到临界点，在这种攻击中，威胁行为者窃取加密数据，存储数年，然后在某个时间点使用未来具有密码学意义的量子计算机进行解密。德勤预测，与2023年相比，2025年致力于实施后量子加密解决方案的公司数量预计将增至四倍，其相关支出也将翻两番。预计2025年，后量子密码学解决方案将涵盖从企业和超大规模企业到消费者智能手机和消息服务的各个领域。

上升趋势

请密切关注这些新兴趋势。我们预测，它们很快就会成为关注的焦点，改变讨论的方向，并塑造行业的未来：

生成式人工智能与网络安全：风险与机遇并存

网络安全专业人士深知，生成式AI在带来网络威胁的同时又能用于开发网络解决方案，因此正在探索如何利用生成式AI的力量应对新兴风险，同时帮助强化技术环境。

根据2024年德勤与国家州信息主管协会 (NASCIO) 联合进行的网络安全研究显示，近四分之三接受调查的安全专家表示，人工智能带来的网络安全威胁很高。2024年利用生成式AI进行的网络攻击频发（较以往增加了一倍甚至两倍），而到2025年网络攻击频率还将持续增长，被用于编写恶意网络钓鱼邮件、深度伪造内容或恶意软件攻击的软件代码的威胁行为者将增多。开发生成式AI工具的技术公司可能会在2025年开发防护栏，以防止恶意使用。虽然威胁行为者可以使用生成式AI工具进行恶意活动，但防御者同样可以利用这些工具来帮助改进安全流程、监控和风险管理。

硅芯“化整为零”：芯粒“续命”摩尔定律

芯粒致力于为人工智能和高性能计算环境提供更加灵活、可扩展和高效的系统，同时提高良品率。

芯粒 (Chiplet)，作为一种开发和封装半导体的异构技术架构，能够实现高速数据传输、减少延迟，并有助于优化PPA（功耗、性能和面积）。德勤预测，基于芯粒的全球先进封装技术收入将从2021年的约70亿美元增至2025年的160亿美元，增幅超过一倍。芯粒已经应用到一些快速增长的市场，如AI加速器（尤其是生成式AI）、高性能计算和电信应用。它们正在推动半导体行业持续提升性能和产量。

业务/运营支持系统（B/OSS）：电信公司对其业务和运营支持系统进行现代化升级

电信公司的后端业务和运营软件市场增长缓慢，但通过采用SaaS、微服务架构、云迁移等方式实现其现代化升级，是目前软件供应商的增长热点，也是电信公司利用5G、光纤和人工智能拓宽业务的重要机遇。

电信公司一直拥有两套独立但重要的电信专用IT系统。分别为业务支持系统 (BSS)，主要用于客户订单捕获、客户关系管理和计费；以及运营支持系统 (OSS)，负责服务订单管理、网络库存管理和网络运营。这通常是两个部署在本地的独立系统，且一般是定制的硬件定义系统，主要由个性化和专业化解决方案构成。然而，到2025年，随着客户期望的不断变化和新的数字收入来源的出现，预计许多电信运营商将对这些系统进行现代化改造和整合。德勤预测，到2025年，OSS和BSS市场 (B/OSS) 的全球总收入将达到700亿美元，年增长率约为5%。基于云的解决方案和软件即服务 (SaaS) 产品的增长预计将更加迅猛，年增长率分别为22%和18%。预计未来几年，大部分的增长将来自美洲、中东、北非以及新兴的亚太地区。

硅光子：生成式人工智能实现光速通信

生成式AI要求日益提高，硅基光学器件正走出研究实验室，成为数据中心的应用焦点。

德勤预测，用于光收发器的硅光子芯片的销售额将从2023年的8亿美元增至2025年的12.5亿美元，复合年均增长率为25%。硅光子芯片可助力生成式AI数据中心实现光速通信，采用更小、更便宜的组件，能耗更低，产生的热量也少于传统替代品。2025年，2025年，预计硅光子技术主要应用于数据中心，尤其是用于运行生成式AI训练和推理——尤其是在芯片、托盘和机架之间数据需要传输10厘米至10米距离的场景中。

By

Ariane Bucaille
France

Kevin Westcott
United States

Gillian Crossan
United States

Lara Abrash
United States

尾注

1. Gartner, "*Gartner forecasts worldwide public cloud end-user spending to surpass $675 billion in 2024*," press release, May 20, 2024.

2. Angle City FC, "*Willow Bay and Bob Iger to become Angel City's new controlling owners*," July 17, 2024.

3. Josh Sim, "*Las Vegas Aces valued at US$140m as average WNBA team hits US$96m*," *SportsPro*, June 18, 2024.

4. GSA, "*5G - GSA Market Snapshot March-2024*," March 4, 2024.

5. TeckNexus, "*Current State of Open RAN – Countries & Operators deploying & trialing Open RAN*," March 10, 2024.

致谢

We wish to thank **Duncan Stewart**, **Je Loucks,** and **Paul Lee**, plus the entire team, for their work on the Predictions report.

Cover image by: **Jaime Austin**; Getty Images, Adobe Stock

Deloitte.
Insights

随着生成式人工智能功耗日增，数据中心寻求更绿色可靠的能源解决方案

科技行业应优化基础设施、创新芯片设计，并与电力提供商合作，助力数据中心实现未来可持续发展。

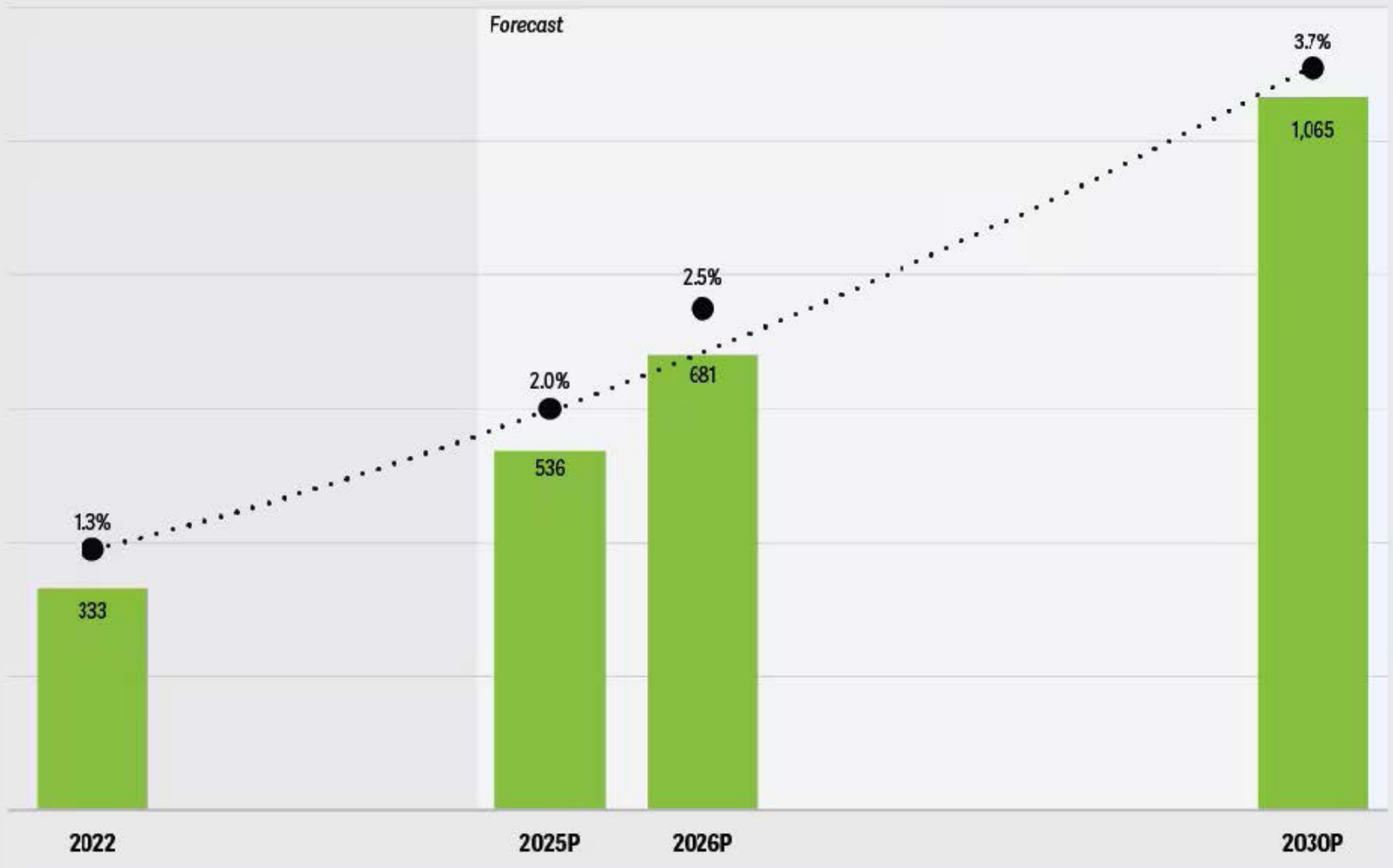
人工智能数据中心用电量将持续攀升，但事实上，数据中心用电量占全球电力需求的比重并不高。德勤预测，到2025年，数据中心用电量大约仅占全球用电量的2%，即536太瓦时（TWh）。然而，随着电力密集型生成式人工智能的训练和推理需求迅速增长，超过其他应用，预计到2030年全球数据中心的用电量将翻一番，约达1,065太瓦时（见图1）。[1]为保障数据中心供电并减少环境影响，许多公司正探索将数据中心的创新节能技术与更多无碳能源结合使用。

然而，发电和电网基础设施要满足人工智能数据中心激增的用电需求实属不易。由于电气化进程——运输、建筑和工业领域从化石燃料设备转向电力设备——以及其他因素，电力需求增长迅速。加之生成式人工智能的出现，导致电力需求增长超出预期。此外，数据中心往往对电力供应有特殊要求，需要具备高冗余度、高可靠性的全天候电力供应，并致力实现供电过程的碳中和。

由于多重变量的影响，要估算2030年及以后全球数据中心的用电量并非易事。据德勤评估，如果人工智能和数据中心的处理效率不断提升，到2030年全球数据中心的能耗水平将达约1,000太瓦时。然而，若预期的效率提升在未来几年内未能实现，则到2030年全球数据中心的能耗水平或将超过1,300太瓦时，这将直接影响到电力提供商，并阻碍气候中立目标的达成。[2]因此，未来十年里，推动人工智能创新和提高数据中心效率，将成为塑造可持续能源格局的关键。

图1

受能源密集型生成式人工智能模型的推动，预计2030年全球数据中心用电量将激增



注：P表示预测值。

资料来源：德勤基于公开资料来源以及与行业专家的讨论所做的分析。

分析方法：德勤基于美国能源信息署《2023年国际能源展望》中关于住宅、商业、工业和交通业终端用户总用电量的基础用电数据（参见表：按终端用户部门和燃料划分的能源消耗量），得出了2022年至2030年全球数据中心用电量（太瓦时）的估计值和预测值。德勤对数据中心用电量占全球总用电量比重的估计和预测，是基于对Semi Analysis、EPRI、高盛、彭博和Latitude Media等多方公开资料的分析，并通过与科技、能源和可持续发展行业专家的讨论予以进一步验证。

随着人工智能数据中心电力需求的日益增长，全球部分地区已面临发电和电网容量管理问题。[3]2023年至2026年间，全球数据中心关键组件（包括GPU 和CPU服务器、存储系统、冷却设备及网络交换机）所需电力预计将增长近一倍，到2026年将达96吉瓦（GW），其中仅人工智能运算就可能占用超40%的电力。[4]预计到2026年，全球人工智能数据中心的年用电量将达90太瓦时（约占届时全球数据中心预计用电量681太瓦时的七分之一），较2022年增长约十倍。[5]因此，生成式人工智能投资大幅推高了用电需求，例如，2024年第一季度全球人工智能数据中心的新增电力净需求约为2吉瓦，较2023年第四季度增长了25%，更是2023年第一季度的三倍多。[6]满足数据中心的用电需求颇具挑战，因为数据中心设施往往集中在特定区域（如美国），且其全天候电力需求将对现有电力基础设施造成负担。[7]

德勤预计，科技行业和电力行业将共同应对上述挑战，降低人工智能（尤其是生成式人工智能）对能源行业的影响。目前，许多大型科技公司和云服务提供商已着手进行无碳能源投资，并推动实现净零排放目标，[8]积极践行对可持续发展的坚定承诺。

超大规模云服务提供商拟大规模扩建生成式人工智能数据中心，以满足日益增长的客户需求

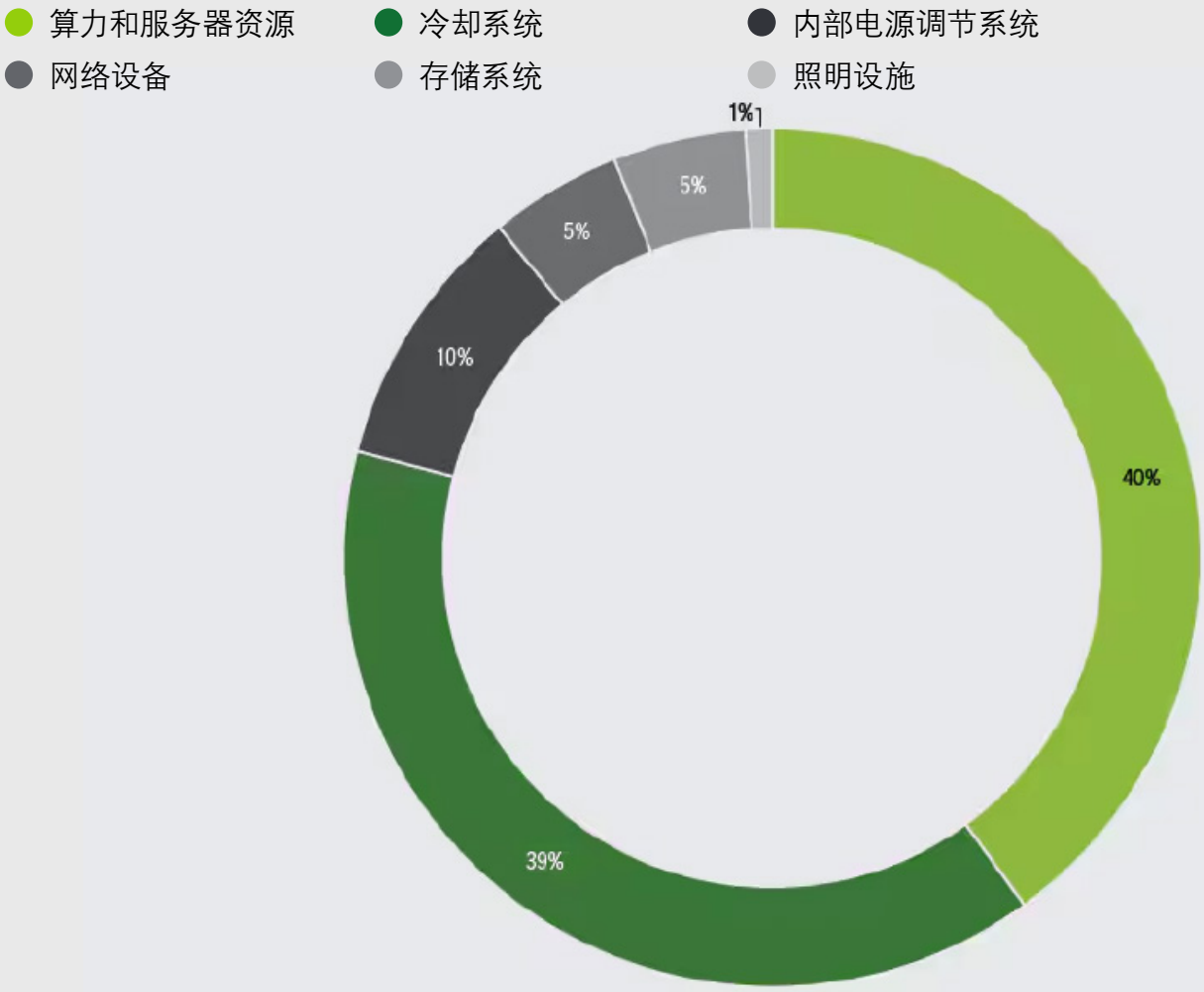
电力需求激增的主要原因在于超大规模云服务提供商计划拓展全球数据中心的容量。[9]随着人工智能（尤其是生成式人工智能）需求上升，企业和国家纷纷投入数据中心建设。各国政府也在打造主权人工智能能力，以保持其技术领先地位。[10]有数据显示，几家主要云服务提供商的数据中心建设投资已创历史新高，2024年资本支出约2,000亿美元，2025年或将超过2,200亿美元。[11]

此外，德勤调研报告《企业生成式人工智能应用现状》指出，截至目前，多数企业的人工智能应用尚处于试点和实验阶段。[12]但在探索生成式人工智能价值的过程中，受访企业已看到切实成果，因此打算在试点和概念验证之后迅速扩大其应用规模。随着生成式人工智能技术的成熟和使用量的增加，预计到2025年和2026年，云服务提供商的资本支出将继续保持高位。

数据中心的电力消耗主要集中在两大领域：算力和服务器资源（如服务器系统，约占总电力消耗的40%）以及冷却系统（约占38%~40%）。即便在人工智能数据中心，这两者亦是能耗最高的部分，持续推高整体电力消耗。此外，内部电源调节系统占8%~10%，网络通信设备和存储系统各占约5%，照明设施则通常仅占1%~2%（见图2）。[13]考虑到生成式人工智能的高电力需求，超大规模云服务提供商、数据中心运营商等数据中心提供商在设计数据中心时，应考虑采用替代能源、创新冷却技术和更节能的解决方案。目前，多项相关工作已在进行中。

图2

算力和冷却系统是人工智能数据中心能耗主因



资料来源：德勤基于ScienceDirect（2023年）和IEEE Access（2021年）等公开调研报告所做的分析。

生成式人工智能带来电力需求增长

自2023年以来，数据中心能耗因人工智能需求的激增而持续攀升。[14]部署先进的人工智能系统需要大量芯片和处理能力，而训练复杂的生成式人工智能模型更需要数千个GPU。

为此，支持人工智能和高性能计算的超大规模云服务提供商和大型数据中心运营商，必须构建高密度基础设施以保证算力。过去，数据中心主要依靠CPU，每块芯片的运行功率约在150瓦至200瓦之间。[15]2022年，用于人工智能的GPU运行功率为400瓦，而2023年用于生成式人工智能的最先进GPU运行功率已达700瓦；预计2024年，新一代芯片的运行功率将高达1,200瓦。[16]相较几年前的传统数据中心设计，新一代芯片（约8个）组成的刀片机架（每个机架10个刀片），每平方米占地面积的耗电量和发热量都更大。[17]截至2024年初，数据中心的机架功率普遍超过20千瓦。预计到2027年，单个服务器机架的平均功率密度将从2023年的36千瓦增至50千瓦。[18]

自生成式人工智能问世以来，以每秒浮点运算次数（FLOPS）衡量的人工智能总算力亦呈指数级增长。自2023年第一季度以来，全球人工智能总算力每季度增长50%~60%，并预计到2025年第一季度仍将保持这一增速。[19]但数据中心不仅以FLOPS来衡量算力，还以兆瓦时（MWh）和太瓦时（TWh）作为衡量标准。

包含数十亿参数的生成式人工智能大型语言模型及其巨额功耗

生成式人工智能的大型语言模型（LLM）日益精密，其参数（即实现人工智能学习和预测功能的变量）数量也在逐步增加。2021年至2022年间，问世的初始模型拥有1,000亿至2,000亿个参数，而到2024年中期，先进的大型语言模型已扩展至近两万亿个参数，能够解读和解码复杂的图像。[20]此外，全球正竞相发布十万亿参数级大型语言模型。由于人工智能必须经过训练和部署，更多的参数也将增加数据处理和算力需求。这将进一步加大对生成式人工智能处理器和加速器的需求，增加耗电量。

此外，大型语言模型的训练过程极为耗能。研究表明，对于参数量超过1,750亿的大型语言模型，单次训练的耗电量在324兆瓦时至1,287兆瓦时之间……而且模型通常需要多次训练。[21]

平均而言，生成式人工智能提示请求的电力消耗是普通网络搜索的10到100倍。[22]德勤预测，如果全球每天有5%的网络搜索采用生成式人工智能提示，则需要约20,000台服务器（每台服务器配备8个专用GPU）来满足这些请求，每台服务器平均耗电6.5千瓦，这意味着日均用电量可达3.12吉瓦时，年用电量达1.14太瓦时[23]——相当于约108,450个美国家庭一年的用电总量。[24]

数据中心用电需求对电力行业转型是一把双刃剑

电力行业已着手制定相应计划，以满足不断增长的用电需求。业内人士普遍预测，到2050年，部分国家的用电量将增加两倍之多。[25]但近期，由于数据中心用电需求激增，部分地区的用电量增速明显加快。多个国家曾做出预测，电气化进程推进、数据中心用电量增长以及整体经济增长将导致电力需求持续上升。但近期数据中心用电需求的激增或许仅是冰山一角，供电压力日渐凸显。[26]

随着电力公司对电网基础设施进行建设和升级，以及推进去碳化和数字化，电力行业步入长达数十年的转型期。在许多地区，电力公司还在加固设备，以应对日益严峻的气候事件，并保护网络免受与日增加的网络安全威胁。[27]部分国家的电网难以满足电力需求，尤其是对低碳或零碳电力的需求。2026年，美国数据中心的用电量预计将占全国总用电量的6% (约260太瓦时) 。[28]由于人工智能的发展，英国数据中心的电力需求或在短短十年内增长六倍。[29]到2026年，中国数据中心 (包括人工智能数据中心) 预计将占全国电力需求的6%。[30]此外，数据中心的电力需求，对于中国能源转型来说是一个新的变量，只有与清洁能源发电相结合，才有利于推动能源转型和双碳目标的实现。[31]

面对数据中心用电需求的不断增长，部分国家正在制定相关法规予以应对。例如，爱尔兰现有数据中心的用电量占全国总用电量的五分之一，随着人工智能数据中心不断涌现，该比例或将进一步提高；家庭用电甚至出现下降。[32]爱尔兰一度叫停接入电网的新数据中心建设计划，但后来改变了这一决策。[33]与爱尔兰一样，荷兰阿姆斯特丹亦暂停了新数据中心的建设，以支持城市的可持续发展。[34]新加坡则针对数据中心推出了全新的可持续发展标准，要求运营商逐步提高设施运行温度至26摄氏度或以上，以减少冷却需求并降低功耗，但这将缩短芯片的使用寿命。[35]

数据中心需求的紧迫性、地域集中度以及对全天候无碳能源的需求，使得科技公司和电力提供商面临更加严峻的挑战。电气化、制造业等领域也将产生新的用电需求。弗吉尼亚州北部是全球最大的数据中心市场，[36]本地公用事业公司Dominion Energy预计，未来15年弗吉尼亚州北部的电力需求将增长约85%，其中数据中心的用电需求将翻两番。[37]许多科技公司难以在短期内获得全天候无碳电力。电力提供商正想方设法，以满足需求并维持电力供应的可靠性和可负担性。除开发新的可再生能源和电池储能技术外，多家电力提供商还计划建设含碳的天然气发电厂，[38]这或将加大公用事业、州乃至国家实现脱碳目标的难度。[39]

尽管人工智能将消耗大量清洁能源，但亦有助于加速清洁能源转型：部分公用事业公司已开始利用人工智能改进天气预报、电力负荷预测、电网管理、可再生资产性能、风暴后恢复和野火风险评估等，从而降低电网的运行成本、提高运行效率和可靠性。[40]

数据中心冷却系统耗水量巨大

新一代CPU和GPU较上一代具有更高的热密度。与此同时，为迎合高性能计算和人工智能应用的强劲需求，部分服务器提供商在每个服务器机架上安装更多高能耗芯片。这样的高密度机架需水量更大，尤其是冷却生成式人工智能芯片。到2027年，人工智能数据中心的淡水需求量最高可达1.7万亿加仑。[41]如果一个超大规模数据中心计划采用空气冷却和饮用水蒸发冷却技术来控制过热，其年用水量将超过5,000万加仑 (约为制造14,700部智能手机的用水量[42]) ，且这些水量无法回流到含水层、水库或供水处。[43]

在普通数据中心中，仅空气冷却技术的耗电量就高达40%。因此，数据中心正在寻找传统空气冷却方式的替代方案，首选即是液体冷却技术，原因是液冷技术具有更高的热传导性能，有助于冷却高密度服务器机架，且相较于空气冷却方式，可减少高达90%的用电量。[44]液冷技术还可对服务器机架进行直接冷却，因此可支持50至100千瓦或更高功率的密集机架。[45]此外，液冷技术还有助于减少对传统冷却器的依赖。

尽管液冷技术在降低数据中心整体能耗方面颇具潜力，[46]但其应用仍处于早期阶段，尚未在全球范围内的人工智能数据中心广泛采用。[47]此外，水作为一种有限资源，其成本和供给状况预计将决定液冷技术的未来应用。

科技行业正朝着更可持续的解决方案和无碳能源的方向发展

科技巨头持续通过购电协议 (PPA) 或与可再生能源提供商签订长期合同，积极寻求可再生能源，以加速利用无碳能源为人工智能数据中心供电。[48]该等交易为可再生能源项目提供了融资支持。同时，科技公司还与电力提供商和创新企业展开合作，以帮助测试和推广有前景的能源技术，如先进的地热能、风能、太阳能和水电技术，甚至探索建立海底数据中心。

在某些地区，受当地电网的制约以及新能源与电池储能设施接入用时过长的影响，该等设施的并网进程出现延误。[49]在美国，由于用电需求旺盛且输电基础设施不足，常常导致电力供应延误，延误时间最长可达五年。因此，科技公司正积极寻求现场或离网的能源解决方案[50]，并投资于长时储能 (LDES) 和小型模块化核反应堆 (SMR) 等新技术，以应对该等挑战。同时，科技公司与公用事业公司计划开展协作，推动新型清洁能源技术的规模化应用，从而惠及更多客户，并加速电网脱碳进程。[51]其中许多研发项目、试点项目和其他清洁能源投资需要数年时间才能显现效益和商业化潜力。[52]例如，小型模块化核反应堆目前仍处于早期开发阶段，短期内可能非理想的零碳解决方案。[53]

科技行业在美国企业的可再生能源采购中始终占据主导地位。在2024年2月28日前的12个月内，美国企业达成近200项可再生能源采购交易，合同容量约19吉瓦，其中科技行业占68%以上。[54]同样，印度的超大规模云服务提供商和数据中心运营商亦日益依赖于风能和太阳能为其数据中心供电。[55]若无这些采购合同，许多可再生能源项目恐难落地。[56]

因此，科技行业在为清洁能源技术提供资金支持、推动其规模化发展方面将继续发挥重要作用。科技行业不仅直接与创新企业和可再生能源生产商合作，还与公用事业公司合作。[57]重要的是，创新企业和电力行业通常无法比肩科技行业的资金实力，因此科技公司如何注资以助推清洁能源转型，显得尤为关键。

中国通过能源转型与优化资源配置助力数据中心可持续发展

目前，中国数据中心耗电量约占全社会用电量的3%。其中，AI驱动的数据中心将成为重要的能源消耗来源，随着生成式AI技术的不断发展和大规模部署，未来数据中心的电力需求将呈现几何级增长。预计到2025年中国数据中心用电量将突破4,000亿千瓦时，占全社会用电量4.1%。因此推动清洁能源的应用以及优化数据中心的能效成为当务之急。中国的新可再生能源计划强调了在数据中心建设中逐步增加可再生能源的使用比例，支持在有冷水资源的国家枢纽节点建设数据中心，并逐步对旧基站和分散的小型数据中心进行绿色技术升级，并通过“东数西算”工程优化资源配置。未来，中国将更加积极应对平衡AI驱动的电力需求激增与能源可持续发展的挑战，进一步促进算力与电力协同，推动数据中心可持续发展。

小结

广大科技行业、超大规模云服务提供商、数据中心运营商、公用事业公司和监管机构应如何行动，以推动生成式人工智能的可持续发展？以下是超大规模云服务提供商和科技行业需考量的几点，它们与德勤全球在2021年关于云迁移预测中提出的观点不谋而合。[58]尽管市场需求驱动因素或有转变，且变化节奏加快，但基本需求大致相同，所以可持续发展的基本要求和重点依旧不变：控制数据中心不断增长的能源需求，并寻求更可持续的方式为人工智能（尤其是生成式人工智能）供电。

1. 提升生成式人工智能芯片能效：目前，新一代人工智能芯片可在90天内完成人工智能训练，耗电8.6吉瓦时，不到上一代芯片在同等情形下所需能耗的十分之一。[59]芯片公司应与半导体生态深化合作，以聚焦并提高每瓦特性能，让未来芯片能在更低能耗情况下训练出远超现有规模的人工智能系统。

2. 优化生成式人工智能应用，实现数据处理边缘化：评估数据中心与边缘设备在训练和推理方面的能耗差异，据此调整数据中心的设备配置。边缘计算不仅适用于时间敏感型应用，还能有效处理敏感数据和满足高隐私需求。边缘设备还有助于节省网络和服务器带宽，将生成式人工智能的工作负载导向本地和近地或主机托管设备，只将必要的人工智能工作负载传输到数据中心。[60]

3. 改变生成式人工智能算法，调整人工智能工作负载：我们是否应一味追求建立更大的基础模型（例如，万亿参数级模型），还是转而使用更具可持续性的小模型？目前，初创企业正在开发端侧的多模态人工智能模型，该等模型无需依赖高能耗的云端计算。[61]客户应根据实际业务需求，精准调整人工智能工作负载，并选取适当的人工智能模型（包括现成模型，仅在需要时才进行训练），以最大限度地减少能耗。此外，CPU可根据人工智能推理的具体需求（例如，实时推理和低延迟推理）充分发挥优势、提升效率。[62]

4. 建立战略合作伙伴关系，满足地方和集群级人工智能数据中心的需求：部分中小型客户（如大学）可能难以获得足够的生成式人工智能数据中心资源，因此应与专业数据中心运营商和云服务提供商开展合作，后者专注于为小型HPC GPU集群主机托管提供HPC解决方案。[63]因此，数据中心应主动监测自身使用情况和资源可用性，发掘潜在商机和需求洼地，以满足短期主机托管服务需求。

5. 与多方利益相关者和行业合作，对整体环境产生积极影响：超大规模云服务提供商及其客户、第三方数据中心运营商、主机托管服务提供商、电力提供商、地方监管机构和市政当局以及房地产公司等生态系统参与者，应围绕商业、环境和社会效益展开持续对话。[64]合作内容应涵盖多个方面，包括：确定潜在的战略主机托管服务需求（即数据中心公司向一家或多家公司出租计算和服务器资源）、评估冷却需求（如液冷系统的适当温度）、制定热能和废水管理解决方案以及回收利用策略。例如，在欧洲，已有数据中心运营商利用余热为附近泳池供暖。[65]电力提供商应与科技行业开展更密切合作，以保障数据中心的能源供给，并确定科技公司如何资助和推广新能源技术，这对推动清洁能源上网尤为重要。

长远来看，超大规模云服务提供商和电力提供商在提升数据中心（包括专为生成式人工智能而建的数据中心）的无碳能源利用比重、满足其电力需求方面所做的全面努力预计将取得成效。

By **Karthik Ramachandran**
India

Duncan Stewart
Canada

Roger Chung
China

Kate Hardin
United States

Gillian Crossan
United States

尾注

1. Deloitte analysis based on publicly available information sources and conversations with industry specialists. We used base electricity consumption data from the US Energy Information Administration's (EIA) International Energy Outlook 2023 data on total electricity usage across residential, commercial, industrial, and transportation end uses (a reference to US Energy Information Administration, "*Table: Delivered energy consumption by end-use sector and fuel*," accessed Nov. 4, 2024) to arrive at estimates and prediction values for global data centers' electricity consumption (TWh) between 2022 and 2030. Our estimates and projections for data centers' percent electricity consumption of global total are based on our research of multiple publicly available sources including SemiAnalysis, EPRI, Goldman Sachs, Bloomberg, and Latitude Media, and further validated based on our conversations with subject matter specialists in the areas of technology, energy, and sustainability. Total energy consumption by end-use sector and fuel (as noted from the aforementioned table from EIA's International Energy Outlook 2023 data), globally, is estimated and forecast at 26,787 TWh in 2025, 27,256 TWh in 2026, and 29,160 TWh in 2030— increasing from 25,585 TWh back in 2022.

2. As noted in endnote 1 above, we arrived at 2022 to 2030 data, estimates, and predictions based on a combination of in-depth secondary research of multiple publicly available sources, and validated further from our discussions with subject matter specialists. Also, see Prof. Dr. Bernhard Lorentz, Dr. Johannes Trüby, and Geoff Tuff, "*Powering artificial intelligence,*" Deloitte Global, November 2024."

3. One-fifth of Ireland's electricity is consumed by data centers, and this is expected to grow, even as households are lowering their electricity use. To read further, see: Chris Baraniuk, "*Electricity grids creak as AI demands soar*," BBC, May 21, 2024.

4. Dylan Patel, Daniel Nishball, and Jeremie Eliahou Ontiveros, "*AI data center energy dilemma: Race for AI data center space*," SemiAnalysis, March 13, 2024.

5. Ibid.

6. Data center BMO report, Communications Infrastructure, "1Q24 data center leasing: Records are made to be broken," April 28, 2024; Moreover, due to strong demand from cloud providers and AI workloads, the data center primary market supply in the United States alone was up 26% year over year to 5.2 GW in 2023, and more are under construction. See further: CBRE, "*North America data center trends H2 2023*," March 6, 2024.

7. Lisa Martine Jenkins and Phoebe Skok, "*Mapping the data center power demand problem, in three charts*," Latitude Media, May 31, 2024.

8. Based on our analysis of multiple publicly available information and reports from what companies self-report, and further validated from third-party sources.

9. For context, hyperscalers are large cloud service providers and data centers that offer huge amounts of computing and storage resources typically at enterprise scale. See: Synergy Research Group, "*Hyperscale operators and colocation continue to drive huge changes in data center capacity trends*," Aug. 7, 2024.

10. Yifan Yu, "*AI's looming climate cost: Energy demand surges amid data center race*," Nikkei Asia, June 12, 2024.

11. Data center BMO report, Communications Infrastructure, "1Q24 data center leasing: Records are made to be broken," April 28, 2024. Further, Deloitte analysis based on information from select tech companies' publicly available sources such as earnings releases and Dell'Oro Group's market research data on data center IT capital expenditure shows that if we consider the capital expenditure spending of other data center providers, including third-party operators and outsourced cloud service providers, data centers' aggregate capital expenditure spending could be at least US$250 billion in 2025. See: Baron Fung, "*Market research on data center IT capex*," Dell'oro Group, accessed Nov. 4, 2024.

12. Nitin Mittal, Costi Perricos, Brenna Sniderman, Kate Schmidt, and David Jarvis, "*Now decides next: Getting real about generative AI*," Deloitte's State of Generative AI in the Enterprise quarter two report, Deloitte, April 2024.

13. Deloitte analysis based on publicly available research reports including: Wania Khan, Davide De Chiara, Ah-Lian Kor, and Marta Chinnici, "*Advanced data analytics modeling for evidence-based data center energy management*," *Physica A* 624, 2023; Kazi Main Uddin Ahmed, Math H. J. Bollen, and Manuel Alvarez, "*A review of data centers energy consumption and reliability modeling*," in *IEEE Access* 9, 2021: pp. 152536–152563.

14. Tom Dotan and Asa Fitch, "*Why the AI industry's thirst for new data centers can't be satisfied*," *The Wall Street Journal*, April 24, 2024.

15. Noam Brousard, "*Examining the impact of chip power reduction on data center economics*," Semiconductor Engineering, March 12, 2024.

16. Based on our analysis of multiple publicly available sources including: Michael Studer, "*The energy challenge of powering AI chips*," Robeco, Nov. 6, 2023; Agam Shah, "*Generative AI to account for 1.5% of world's power consumption by 2029*," HPCwire, July 8, 2024.

17. From our study and analysis of select gen AI data center chip solutions offered by major AI chip vendors, further corroborated with publicly available third-party sources including: Beth Kindig, "*AI power consumption: Rapidly becoming mission-critical*," *Forbes*, June 20, 2024.

18. Jones Lang LaSalle, "*Data centers 2024 global outlook*," Jan. 31, 2024; Doug Eadline, "*The gen AI data center squeeze is here*," HPCwire, Feb. 1, 2024; Per IDC, besides graphics processing unit, servers, data centers also need to grapple with a corresponding growth in storage capacity, which is likely to double between 2023 and 2027 to reach 21 zettabytes in 2027. See: John Rydning, "*Worldwide Global StorageSphere forecast, 2023 to 2027: Despite decreased petabyte demand near term, the installed base of storage capacity continues to grow long term*," IDC Corporate, May 2023.

19. Patel, Nishball, and Eliahou Ontiveros, "*AI data center energy dilemma*."

20. Sean Michael Kerner, "*What are large language models?*" TechTarget, May 2024; Yu, "*AI's looming climate cost*."

21. Alex de Vries, " *The growing energy footprint of artificial intelligence*," *Joule* 7, no. 10 (2023): pp. 2191–2194.

22. Eren Çam, Zoe Hungerford, Niklas Schoch, Francys Pinto Miranda, and Carlos David Yáñez de León, "*Electricity 2024: Analysis and forecast to 2026 report*," International Energy Agency, accessed Nov. 4, 2024.

23. Deloitte analysis based on publicly available reports and sources including: de Vries "*The growing energy footprint of artificial intelligence*," pp. 2191–2194.

24. Deloitte analysis based on data related to energy use and electricity consumption in homes in the United States. See: US Energy Information Administration, "*Use of energy explained*," accessed Dec. 18, 2023.

25. Darren Sweeney, "*Utility execs prepare for 'tripling' of electricity demand by 2050*," S&P Global, April 19, 2023.

26. Robert Walton, "*US electricity load growth forecast jumps 81% led by data centers*," Utility Dive, Dec. 13, 2023.

27. Aaron Larson, "*How utilities are planning for extreme weather events and mitigating risks*," *POWER*, March 13, 2024.

28. Çam, Hungerford, Schoch, Miranda, and de León, "*Electricity 2024*."

29. Baraniuk, "*Electricity grids creak as AI demands soar*."

30. Yu, "*AI's looming climate cost*."

31. Data on China's energy use and CO2 emissions sourced from International Energy Agency, accessed September 25, 2024. See: International Energy Agency, "*China's energy use*," accessed Nov. 4, 2024; International Energy Agency, "*China's CO2 emissions*," accessed Nov. 4, 2024.

32. Baraniuk, "*Electricity grids creak as AI demands soar*."

33. Paul O'Donoghue, "*Build it and they will hum: What next for Ireland and data centers?*" *The Journal*, Sept. 2, 2024.

34. Hosting Journalist, "*City of Amsterdam puts halt to new data center construction*," Dec. 21, 2023.

35. With every 1C increase, operators could save 2% to 5% on the energy they use for cooling equipment. To read further, see: Inno Flores, "*Singapore unveils green data center road map amid AI boom that strains energy resources*," Tech Times, May 30, 2024.

36. Julie R. Peasley, "*Ranked: Top 50 data center markets by power consumption*," Visual Capitalist, Jan. 10, 2024.

37. Whitney Pipkin, "*Energy demands for Northern Virginia data centers almost too big to compute*," Bay Journal, June 18, 2024.

38. Zach Bright, "*Southeast utilities have a 'very big ask': More gas*," E&E News, Jan. 22, 2024.

39. Ibid.

40. Robert Walton, "*AI is enhancing electric grids, but surging energy use and security risks are key concerns*," Utility Dive, Oct. 23, 2023.

41. Karen Hao, "*AI is taking water from the desert*," *The Atlantic*, March 1, 2024.

42. Deloitte analysis based on publicly available information sources including: Jennifer Billock, "*Photos: How much water it takes to create 30 common items*," North Shore News, Jan. 19, 2023.

43. Hao, "*AI is taking water from the desert*"; One case in point is China—where its data centers' annual water consumption is expected to increase from around 1.3 billion cubic meter as of 2023 to over 3 billion cubic meter by 2030. To read further, see: Yu, "*AI's looming climate cost*."

44. Eadline, "*The gen AI data center squeeze is here*."

45. Diana Goovaerts, "*Data center operators want to run chips at higher temps. Here's why.*" Fierce Network, June 11, 2024.

46. Scott Wilson, "*Is immersion cooling the answer to sustainable data centers?*" Ramboll, Dec. 13, 2023.

47. David Eisenband, "*100+ kW per rack in data centers: The evolution and revolution of power density*," Ramboll, March 13, 2024; Direct-to-chip cooling (also known as cold plate liquid cooling or direct liquid cooling) cools down servers by distributing heat directly to server components, while, immersion cooling involves submerging servers and components in a liquid dielectric coolant that also helps prevent electric discharge.

48. Based on Deloitte's analysis of developments and announcements from select major cloud hyperscalers and tech companies— and information gathered from publicly available sources (time period: 2023 and first three quarters of 2024).

49. Joseph Rand, Nick Manderlink, Will Gorman, Ryan Wiser, Joachim Seel, Julie Mulvaney Kemp, Seongeun Jeong, and Fritz Kahrl, "*Queued up: 2024 edition*," Lawrence Berkeley National Laboratory, April 2024.

50. Based on Deloitte's analysis of developments and announcements from select major cloud hyperscalers and tech companies— and information gathered from publicly available information sources between the first quarter of 2023 and the third quarter of 2024.

51. Julian Spector, "*Duke Energy wants to help Big Tech buy the 24/7 clean energy it needs*," Canary Media, June 11, 2024.

52. For example, it's not easy to submerge and drop a 1,300-ton data center unit underwater, especially since it demands special equipment to withstand pressure and corrosion caused by seawater. Moreover, there are concerns related to its impact on marine life.

53. David Schlissel and Dennis Wamsted, "*Small modular reactors: Still too expensive, too slow, and too risky*," Institute for Energy Economics and Financial Analysis, May 2024.

54. Deloitte's analysis of data and information gathered from multiple reports from S&P Global Market Intelligence, published during March and August 2024.

55. Manish Kumar, "*India's data center boom opens up a fresh segment for green developers*," Saur Energy International, July 1, 2024.

56. Naureen S. Malik and Bloomberg, "*With AI forcing data centers to consume more energy, software that hunts for clean electricity across the globe gains currency*," *Fortune*, Feb. 25, 2024.

57. Based on Deloitte's analysis of developments and announcements from select major cloud hyperscalers, tech companies, and power and utility players—on publicly available information sources during 2023 and the first three quarters of 2024.

58. Duncan Stewart, Nobuo Okubo, Patrick Jehu, and Michael Liu, "*The cloud migration forecast: Cloudy with a chance of clouds*," *Deloitte Insights*, Dec. 7, 2020.

59. Wylie Wong, "*Nvidia launched next-generation Blackwell GPUs amid AI 'arms race'*," Data Center Knowledge, March 19, 2024; For instance, Nvidia notes that it can train a very large AI model using 2,000 Grace Blackwell chips in 90 days, consuming 4 MW power. In comparison, it would take as much as 8,000 of the previous generation chips to do the same work within the same time, consuming 15 MW power.

60. To read further, see section "Generative AI comes to the enterprise edge: 'On prem AI' is alive and well" in our 2025 TMT Predictions chapter on "*Updates*"; Additionally, see:
Sabuzima Nayak, Ripon Patgiri, Lilapati Waikhom, and Arif Ahmed, "*A review on edge analytics: Issues, challenges, opportunities, promises, future directions, and applications*," *Digital Communications and Networks* 10, no. 3 (2024): pp. 783–804.

61. Yu, "*AI's looming climate cost*."

62. Luke Cavanagh, "*GPUs vs. CPUs in the context of AI and web hosting platforms*," Liquid Web, Aug. 20, 2024.

63. Eadline, "*The gen AI data center squeeze is here*."

64. Goovaerts, "*Data center operators want to run chips at higher temps. Here's why*."

65. Baraniuk, "*Electricity grids creak as AI demands soar*."

致谢

The authors would like to thank **Dilip Krishna, Marlene Motyka, Jim Thomson, Adrienne Himmelberger, Thomas Schlaak, Freedom-Kai Phillips, Johannes Truby, Clement Cabot, Negina Rood, Ankit Dhameja, Suzanna Sanborn,** and **Akash Cha erji** for their contributions to this article.

Cover image by: **Jaime Austin;** Getty Images, Adobe Stock

Deloitte.
Insights

女性与生成式人工智能：使用鸿沟快速弥合，信任差距仍然存在

要想女性真正吃到生成式人工智能的红利，科技公司应努力增强信任，减少偏见，为其提供更多相关工作岗位。

德勤预测，到2025年底，美国体验和使用生成式人工智能技术的女性用户人数将与男性用户持平甚至实现超越。[1]虽然2023年女性用户仅为男性用户的一半，但据其采用速度推测，她们很可能会快速赶超，在明年内实现齐平。[2]尽管这一预测仅针对美国，生成式人工智能使用方面的性别差距却是一个全球现象：我们对欧洲国家的调研发现，男性和女性在生成式人工智能技术的使用方面存在巨大差异，但女性正在快速迎头赶上。[3]这些国家有望在未来两年内弥合技术使用层面的性别差距。这么看来，全球面临的挑战和机遇与美国的情况如出一辙。

虽然女性使用者的人数正在增加，但与男性使用者相比，她们更不相信技术提供商会保护其数据安全。[4]这一“技术信任差距”将导致女性减少对生成式人工智能技术的使用，阻碍其全面参与新的技术应用，降低其未来购买相关产品和服务的积极性。为帮助克服这一信任差距，科技公司应提高数据安全性，实施清晰明确的技术管理措施并为用户提供更大的数据控制权限。

人工智能模型偏见不利于构建信任。[5]女性在人工智能劳动力中占比不足三分之一，[6]绝大多数人工智能从业者认为，只要该领域持续呈现男性主导的局面，那么人工智能系统就将会产出带偏见的结果。[7]提高女性在该领域的参与度，不仅有助于减轻人工智能中的性别偏见，还能使女性在引领人工智能的未来发展进程中发挥更重要的作用。

生成式人工智能的使用鸿沟正在快速弥合

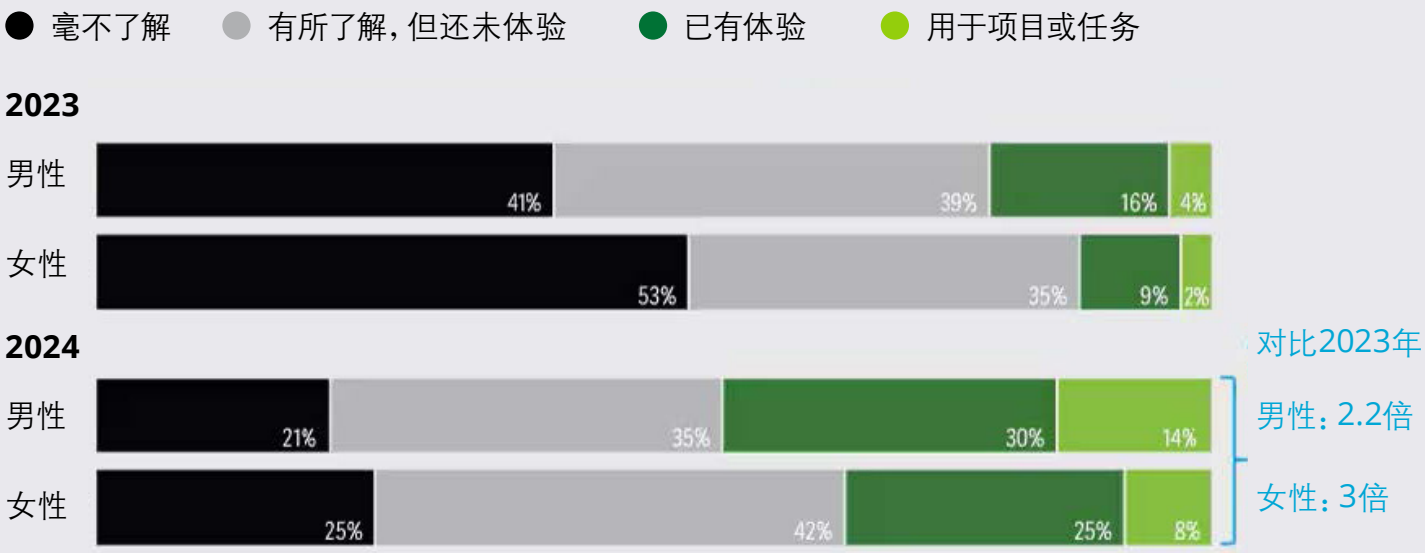
近期的德勤研究发现，各地区在使用生成式人工智能方面存在显著的性别差异。过去两年，德勤互联消费者调研在研究美国消费者的数字生活的同时，也调查了他们对生成式人工智能技术的使用情况。[8]分析表明，美国女性消费者在使用这一新兴技术方面稍显落后（图1）：2023年，生成式人工智能女性用户人数约为男性的一半（11%的女性正体验这一技术或将其正式用于项目或任务，而男性却有20%）；2024年，生成式人工智能的总体使用率已经超过了之前的两倍，但性别差距仍然存在：33%的受访女性表示正使用或体验这一技术，而男性的这一比例为44%。

不仅仅是美国，其他地区也存在这样的差距。德勤英国2024年数字消费者趋势调研发现，在英国使用生成式人工智能的消费者中，女性和男性的比例分别为28%和43%。[9]德勤英国另一项关于生成式人工智能和信任的欧洲研究也表明，在12个欧洲国家中，女性和男性在生成式人工智能的使用人数方面存在着两位数的差异。[10]

图1

美国女性消费者在使用生成式人工智能方面稍显落后，但不久就会迎头赶上

2023至2024年美国消费者对生成式人工智能系统和技术的了解和使用情况



注：美国受访的女性消费者人数：2023年1,040人，2024年1,992人；男性消费者人数：2023年962人，2024年1,841人。

资料来源：德勤《2024年互联消费者调研报告》，第5版；德勤《2023年互联消费者调研报告》，第4版。

Deloitte. Insights | deloitte.com/insights

在美国，女性使用者的人数正在快速增加，弥补使用鸿沟。去年，受访者中使用生成式人工智能技术的女性人数翻了三倍，超过男性人数2.2倍的增长速率。[11]基于对当前使用率和增长率的分析，德勤预测，到2025年底，美国体验生成式人工智能技术或将其用于任务和项目的女性用户人数将与男性用户人数齐平甚至实现超越。[12]

完全参与或许还难以实现

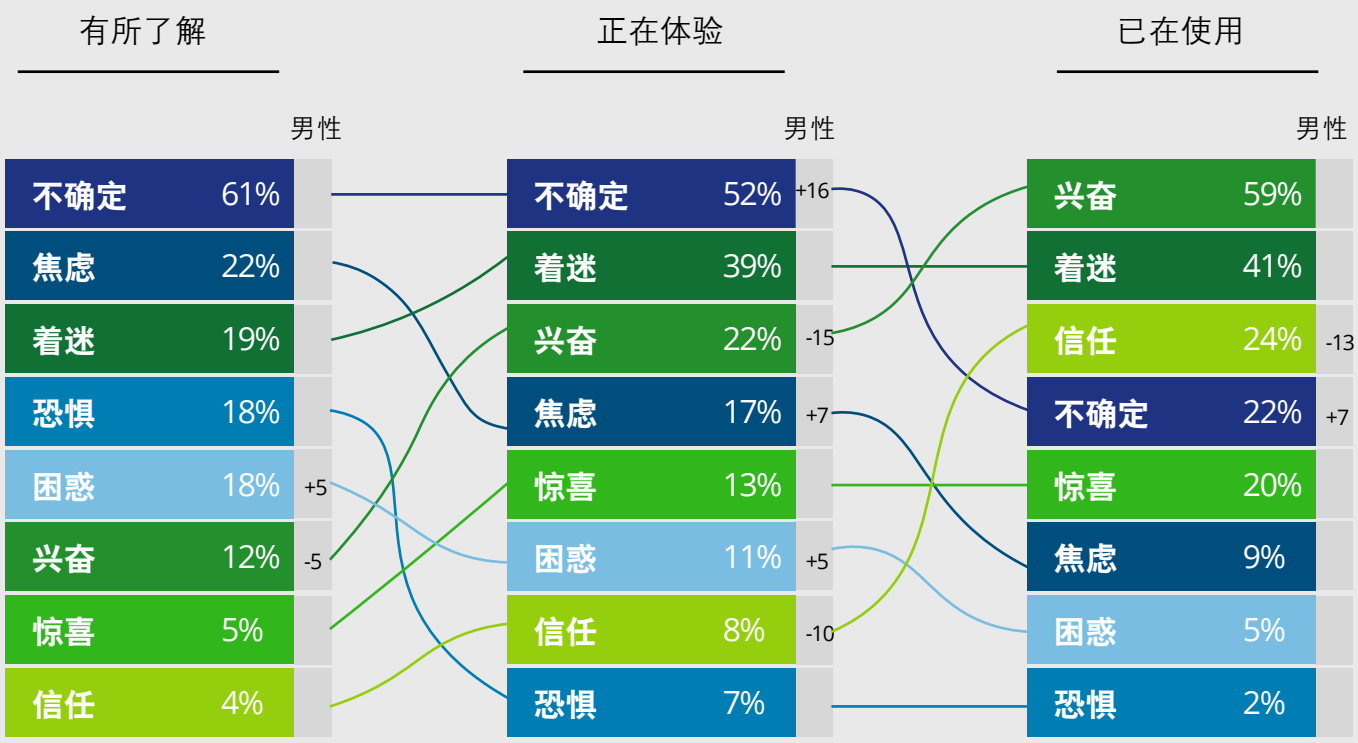
女性和男性达成相同的使用率这一趋势鼓舞人心，但这并不会自动确保女性将生成式人工智能纳入其日常工作。事实上，在德勤2024年互联消费者调研中，34%的女性表示她们每天至少使用一次生成式人工智能技术，而男性的这一比例为43%。[13]在将这一技术用于处理专业任务的用户中，41%的女性认为生成式人工智能极大提升了生产力，而男性的这一比例为61%。[14]科技公司和其他组织要想使用生成式人工智能创造效益，应注意这一性别差异并采取积极措施提高女性的参与度。

这一差异可能部分源于男性和女性用户对技术信任的不同认知。[15]女性用户从起初的对生成式人工智能有所了解发展到体验并使用这一技术，不确定、焦虑、恐惧和困惑等负面情绪逐渐消失，着迷、兴奋、惊喜和信任等积极情绪不断增长（图2）。[16]但是，在体验和使用层面，女性对技术的信任大幅低于男性，反之，不确定感较高。事实上，在声称正在体验或使用生成式人工智能的女性受访者中，只有18%表示“高度”或“极其”信任技术提供商会保护她们的数据安全，而在男性受访者中，这一数字达到了31%。[17]

图2

随着女性用户的使用经验增加，她们对生成式人工智能的负面情绪逐渐减少，信任等积极情绪不断增加，但她们对这一技术的信任程度仍低于男性。

无论女性用户对生成式人工智能处于何种了解/使用阶段，她们对该技术的主要情绪



注：在美国受访者中，对生成式人工智能有所了解的人数：女性827人，男性647人；正在体验生成式人工智能的人数：女性496人，男性547人；已在将这一技术用于项目/任务的人数：女性170人，男性259人。只有当女性与男性的差异达到或超过 5 个百分点时，才会注明。

资料来源：德勤《2024年互联消费者调研报告》，第5版。

Deloitte. Insights | deloitte.com/insights

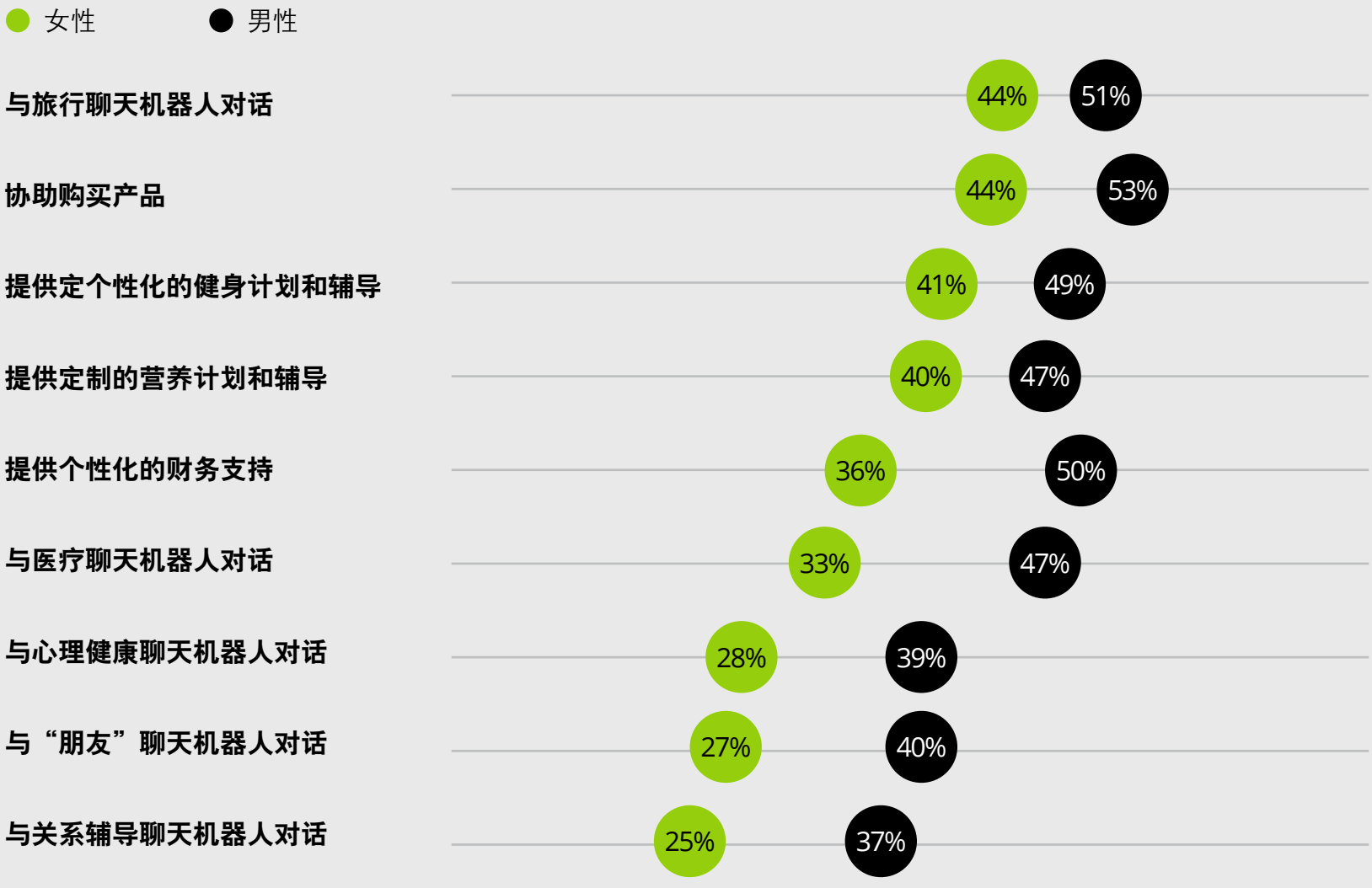
这一信任差距不仅局限于生成式人工智能这一种技术，而是出现在广泛的技术服务和交互中。在2024年互联消费者调研中，54%的女性受访者认为其从线上服务获得的便利超过其对数据隐私的担忧（2023年的比例为46%），而男性受访者的这一比例为62%。[18]在去年的调研中我们发现，女性比男性对于个人数据的使用和保护更为谨慎，这影响了她们分享数据的意愿，尤其是涉及健康和健身指标等敏感数据。[19]女性可能认为安全漏洞或数据滥用的潜在后果更为严重。[20]

生成式人工智能的日益普及可能会加剧这些长期存在的数据隐私和技术问题。[21]当用户使用这一技术时，系统可能会将用户数据反馈到人工智能模型中进行模型训练，但用户是否可以选择拒绝，专家表示这目前还不明确，甚至有些复杂。[22]当用户开始就敏感话题或个人话题与生成式人工智能进行对话，问询其建议时，数据隐私和安全性的风险随之增加。事实上，数据隐私和安全方面的信任差距可能是女性和男性对未来各种生成式人工智能体验表现出兴趣差异的原因（图3）。[23]与男性受访者相比，女性受访者与生成式人工智能交流不太敏感的话题（如旅游、购物、健身、营养）的兴趣较低，交流敏感话题（如个人财务、人际关系、医疗或精神健康状况）的兴趣更低。

图3

与男性相比，女性对生成式人工智能带来的各种体验兴趣较低

表示拥有以下生成式人工智能体验的受访者



注：对生成式人工智能有所了解、正在体验或已在使用的美国受访者人数：女性1,492人，男性1,454人。

资料来源：德勤《2024年互联消费者调研报告》，第5版。

信任差距也会导致女性对于新兴技术产品和服务的购买欲下降。科技公司开始出售内置人工智能芯片的笔记本电脑、平板和智能手机，它们的功能更加完善，例如实时总结信息、生成照片和视频、即刻语言翻译。[24]在德勤2024年互联消费者调研中，当被问到这些新的人工智能功能是否会影响其设备升级计划时，与男性受访者相比，表示会提前升级更换设备的女性受访者人数较少。[25]例如，在智能手机方面，43%男性受访者表示，嵌入式人工智能将使他们非常有可能或有一定可能提前更换手机，但只有32%的女性持相同观点（相反，58%的女性表示这一功能不会影响其换机计划，而男性的这一比例为50%）。在笔记本电脑方面，41%的男性受访者表示其很有可能或较有可能很快升级为带人工智能功能的电脑，而女性的这一比例为28%。据估计，女性控制或影响着85%的消费支出，她们对升级到搭载人工智能技术的设备热情较低，这可能会给技术供应商带来挑战。[26]

信任差距并不是阻碍女性最大限度使用生成式人工智能的唯一因素。在受访者中，61%的女性用户认为公司积极鼓励她们在工作中使用这一技术，而男性的这一比例为83%。[27]49%的女性用户称公司提供相关培训，仍低于男性用户（79%）。无论这些数字反映的是观念上的差异，还是在获得培训项目和工作场所鼓励方面的实际经历，公司都应加以重视，并努力缩小差距。

科技行业女性的使用率较高，但其代表性仍有待提升

在科技行业，生成式人工智能的使用情况则截然不同。科技行业女性可能是未来促进女性整体更多参与生成式人工智能的主要力量。毫无疑问，创造人工智能产品和服务的科技行业其员工对生成式人工智能的使用率肯定高于其他行业。在德勤2024年互联消费者调研中，科技行业中有70%的女性和78%的男性表示正在体验生成式人工智能或将其用于项目或任务，这一比例远高于其他行业的女性 (32%) 和男性 (40%) 。[28]更让人惊讶的是，科技行业女性似乎比男性同行进展更快——她们已经结束了新技术体验期，开始将生成式人工智能正式运用到项目和任务中，两个群体的比例分别为44%和33%。但无论男性还是女性，他们都对这项技术抱有极大期望：约70%预计，一年以后，生成式人工智能将“大幅提高”工作效率。[29]

此外，科技行业中也不存在因性别差异而造成的明显信任差距。相比于整体使用者，这两个群体对生成式人工智能的信任度更高：在使用或体验生成式人工智能的科技行业男性和女性中，超过40%表示信任或非常信任技术提供商会保护他们的数据安全。[30]这两个群体中，75%都认为他们从线上服务获得的便利超过了对隐私的担忧，而在科技行业以外工作的女性和男性中，这一比例分别仅为54%和60%。[31]与非科技从业者相比，科技行业女性可能更了解生成式人工智能的工作原理，而她们在专业领域对技术的大量使用也提高了她们的信心，并使她们了解如何从这项技术中获益。此外，大多数使用生成式人工智能技术的科技行业女性表示，她们的公司鼓励使用该技术 (84%) 并提供培训 (72%) ——相比之下，在其他行业的女性中，表示公司鼓励使用生成式人工智能技术 (55%) 或提供培训 (45%) 的要少得多。[32]

尽管科技行业中女性对人工智能技术的使用率较高，但在人工智能岗位上的女性却相对缺乏。女性仅占人工智能相关劳动力的30%左右，与她们在STEM (科学、技术、工程和数学) 领域的总体比例相当。[33]女性在人工智能领域的参与度不高可能会对各个领域和行业的人工智能系统的开发和部署产生严重影响。

女性在人工智能劳动力中的相对缺席会带来一项主要挑战——在人工智能应用中对女性的性别偏见有可能长期存在。[34]在各行各业中，多达44%的人工智能系统存在性别偏见，这会对人工智能系统的产出产生负面影响，使女性继续被边缘化，其权益得不到充分代表。[35]例如，人工智能中的性别偏见会导致有歧视的雇佣，[36]降低医疗质量，[37]减少妇女获得金融服务的机会。[38]德勤的研究表明，人工智能模型中的偏见会削弱员工和客户的信任。[39]让更多女性走上人工智能岗位，对于实现性别平等和确保人工智能造福社会至关重要。[40]

小结

科技公司应努力提高女性对生成式人工智能的参与度，原因如下：第一，女性掌控或影响着绝大部分的购买决策，如果不能让女性频繁使用生成式人工智能，可能导致人工智能产品和服务无法实现预期销量。第二，如果女性员工不能像男性员工一样充分使用生成式人工智能工具，公司就有可能在对技术投资后无法实现预期的生产力提升。并且，由于生成式人工智能依赖于收集和建立互动数据，女性的代表性不足可能会加剧人工智能模型的偏差。[41]最后，如果女性不能尽其所能参与新兴人工智能应用案例，这可能会阻碍她们最大限度地利用未来的技术优势（例如，使用聊天机器人进行医疗或心理健康方面的干预），并加深现有的不平等。[42]

为增强女性对生成式人工智能技术的信任，科技公司应努力解决与这一技术相关的潜在风险。德勤2024年互联消费者调研发现，建立信任的部分方法包括提高数据隐私和安全政策的透明度，以及为用户提供更大的个人数据控制权限。[43]科技公司应优先采取稳健的技术安全措施，有效展示其数据处理做法。向消费者清晰阐明数据的收集范围和使用方式，同时提供更简便的数据使用控制方法（比如在适当时候提示用户就其数据的使用做出知情选择），这样不仅可以建立信任，还可以带来竞争优势。但应该注意生成式人工智能潜在风险的不仅仅是科技公司：84%的受访者认为，政府应加大力度监管企业对消费者数据的收集和使用。[44]

在各行各业，公司如果想要充分发挥生成式人工智能的全部潜力，应鼓励其员工，不论男性还是女性，积极探索和使用这一技术。除了各种常见用例，比如文档编辑、网络搜索、材料总结和研究协助，还应探求适应本行业的使用方法。[45]此外，还需为员工制定并实施相关培训计划。

努力让消费者全面使用生成式人工智能技术无疑是个可贵的目标，但如果女性和男性没有平等地参与技术开发，这一目标可能难以实现。为了提高人工智能岗位上女性的多样性和包容性，公司应考虑重点创建能满足其员工需求的工作场所。例如，一项针对人工智能领域女性的研究指出，工作与生活的平衡是她们对工作满意度的最重要因素，其中包括拥有灵活的工作安排或能够远程工作等要素。[46]女性还表示，她们希望找到有女性担任领导职务、薪酬和晋升透明、对骚扰和虐待采取零容忍政策的工作。[47]要想吸引更多女性进入这一领域，还要为女性提供更多的教育和培训机会，让她们学习人工智能技能和能力。具体比如，创建更多的导师计划和交流项目，让女性分享经验、相互支持；提供资金扶持，让更多女性参与人工智能研究和创新项目。女性在技术开发中发挥越来越大的作用，未来才会出现越来越多的吸引所有女性都参与其中的应用和系统。

By **Susanne Hupfer**
United States

Bree Matheson
United States

Gillian Crossan
United States

Ariane Bucaille
France

Jeff Loucks
United States

尾注

1. To understand consumer attitudes toward digital life, the Deloitte Center for Technology, Media & Telecommunications surveyed 3,857 US consumers in the second quarter of 2024 and 2,018 US consumers in the second quarter of 2023. These 2024 and 2023 Connected Consumer Surveys collected data on consumers' reported adoption of generative AI, including experimentation and use for projects and tasks (beyond experimentation). By analyzing longitudinal adoption data and calculating the rate of change in adoption from 2023 to 2024 for men and women, we are able to project that women will close the adoption gap by the end of 2025; see: Jana Arbanas et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey, 5th edition,* Deloitte, December 3, 2024; Jana Arbanas, Paul H. Silverglate, Susanne Hupfer, Jeff Loucks, Prashant Raman, and Michael Steinhart, "*Balancing act: Seeking just the right amount of digital for a happy, healthy connected life*," *Deloitte Insights*, Sept. 5, 2023.

2. Ibid.

3. Our analysis was conducted from August to October 2024, based on data from Deloitte UK's 2023 and 2024 Digital Consumer Trends surveys, as well as a 2024 Deloitte UK survey of European consumers on the topic of generative AI; see: Paul Lee and Ben Stanton, "*Generative AI: 7 million workers and counting*," Deloitte, June 25, 2024; Jonas Malmlund, Frederik Behnk, and Joachim Gullaksen, "*Generative AI is all the rage*," Deloitte, 2023; Roxana Corduneanu, Stacey Winters, Jan Michalski, Richard Horton, and Ram Krishna Sahu, "*Europeans are optimistic about generative AI but there is more to do to close the trust gap*," *Deloitte Insights*, Oct. 10, 2024.

4. Analysis based on Deloitte's 2024 Connected Consumer Survey; see: Arbanas et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.*

5. Don Fancher, Beena Ammanath, Jonathan Holdowsky, and Natasha Buckley, "*AI model bias can damage trust more than you may know. But it doesn't have to.*" *Deloitte Insights*, Dec. 8, 2021.

6. World Economic Forum, "*Global gender gap report 2023*," June 2023.

7. Deloitte AI Institute, "*Women in AI*," accessed November 2024.

8. Jana Arbanas et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey, 5th edition,* Deloitte, publishing December 3, 2024; Arbanas, Silverglate, Hupfer, Loucks, Raman, and Steinhart, "*Balancing act*."

9. Deloitte, "*Generative AI: 7 million workers and counting*," accessed November 2024.

10. The Digital Consumer Trends study conducted in various countries in 2024 revealed gen AI adoption gaps of 17 points in Denmark; 12 points in Sweden, Italy, and the Netherlands; 11 points in Belgium; and 10 points in Norway. Additional analysis of a Deloitte European gen AI study revealed gen AI adoption gaps ranging from 10 to 15 points in 11 European countries studied (Belgium, France, Germany, Ireland, Italy, the Netherlands, Poland, Spain, Sweden, Switzerland, and the United Kingdom); see: Deloitte, "*Generative AI*"; Deloitte, "*Generative AI is all the rage*," accessed November 2024; Corduneanu, Winters, Michalski, Horton, and Sahu, "*Europeans are optimistic about generative AI but there is more to do to close the trust gap*."

11. Analysis based on 2024 and 2023 Deloitte Connected Consumer Surveys; see: Arbanas et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey*; Arbanas, Silverglate, Hupfer, Loucks, Raman, and Steinhart, "*Balancing act*." Deloitte, "*Generative AI*."

12. Ibid.

13. Arbanas, et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.*

14. Ibid.

15. Ibid.

16. Ibid.

17. Ibid.

18. Ibid.

19. For example, only 43% of women we surveyed in the Deloitte 2023 Connected Consumer Survey who owned smart watches or fitness trackers said that they share the data collected by those devices with their health care provider, vs. 57% of men; see: Susanne Hupfer, Jennifer Radin, Paul H. Silverglate, and Michael Steinhart, "*Tech companies have a trust gap to overcome—especially with women*," Deloitte Insights, Nov. 8, 2023.

20. These fears may be warranted. Consider that most health apps—along with the data they gather and transmit—are not covered by the Health Insurance Portability and Accountability Act, which means the data may be shared or sold to third parties; see: Steve Alder, "*Majority of Americans mistakenly believe health app data is covered by HIPAA*," The HIPAA Journal, July 26, 2023.

21. Ina Fried, "*Generative AI's privacy problem*," Axios, March 14, 2024; Federal Trade Commission, "*AI companies: Uphold your privacy and confidentiality commitments*, Jan. 9, 2024.

22. Ibid; Matt Burgess and Reece Rogers, "*How to stop your data from being used to train AI*," Wired, April 10, 2024.

23. Arbanas, et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.*

24. Baris Sarer, Ricky Franks, Cheryl Ho, and Jake McCarty, "*AI and the evolving consumer device ecosystem*," *The Wall Street Journal*, April 24, 2014; Sam Reynolds, "*AI-enabled PCs will drive PC sales growth in 2024, say research firms*," Computer World, Jan. 11, 2024; Clare Conley, "*Generative AI in 2024: The 6 most important consumer tech trends*," Qualcomm, Dec. 14, 2023.

25. Arbanas, et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.*

26. Monique Woodard, "*Unlocking the trillion-dollar female economy*," TechCrunch, May 21, 2023.

27. Arbanas, et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.*

28. Ibid.

29. Across industries, 51% of women workers using gen AI anticipate it would substantially boost their productivity at work a year from now, vs. 64% of men; see: Arbanas, et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.*

30. Tech women and men are statistically tied: Forty-two percent of tech women who use or experiment with gen AI have "high" or "very high" trust that gen AI providers will keep their data secure, and another 40% report moderate trust, while 47% of tech men report "high" or "very high" trust and another 30% report moderate trust; see: Arbanas, et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.*

31. Ibid.

32. Greater proportions of men in the tech industry who use gen AI report that their employers encourage its use (93%) and provide training (91%). While there's still a gender gap in these views among workers in the tech industry, the gap is significantly smaller than among men and women working in other industries; see: Arbanas, et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.*

33. World Economic Forum, "*Global gender gap report 2023*."

34. Deloitte, "*Generative AI*."

35. Genevieve Smith and Ishita Rustagi, "*When good algorithms go sexist: Why and how to advance AI gender equity*," Stanford Social Innovation Review, March 31, 2021.

36. Charlotte Lytton, "*AI hiring tools may be filtering out the best job applicants*," BBC, Feb. 16, 2024.

37. Carmen Niethammer, "*AI bias could put women's lives at risk - A challenge for regulators*," *Forbes*, March 2, 2020.

38. Ryan Browne and MacKenzie Sigalos, "*A.I. has a discrimination problem. In banking, the consequences can be severe*," CNBC, June 23, 2023.

39. Fancher, Ammanath, Holdowsky, and Buckley, "*AI model bias can damage trust more than you may know. But it doesn't have to.*"

40. World Economic Forum, "*Global gender gap report 2023*."

41. Smith and Rustagi, "*When good algorithms go sexist.*"

42. Hyun-Kyoung Kim, "*The effects of artificial intelligence chatbots on women's health: A systematic review and meta-analysis*," Healthcare, Feb. 23, 2024; Sheryl Jacobson and Jen Radin, "*Can FemTech help bridge a gender-equity gap in health care?*" Deloitte, Oct. 5, 2023; Karen Taylor, "*Why investing in FemTech will guarantee a healthier future for all women*," Deloitte UK, June 23, 2023.

43. Arbanas, et al., *Earning trust as gen AI takes hold: 2024 Connected Consumer Survey.*

44. Ibid.

45. Deloitte AI Institute, "*The generative AI dossier: A selection of high-impact use cases across six major industries*," April 3, 2023.

46. Women in AI, "*WAI at work: Shaping the future of work for women in AI*," 2022.

47. Ibid.

致谢

Authors would like to thank **Duncan Stewart**, **Paul Lee, Ben Stanton**, **Vipul Mehta**, **Roxana Corduneanu**, **Michael Steinhart**, **Michelle Dollinger**, **Je Stoudt**, **Catherine King**, **Elizabeth Fisher**, **Andy Bayiates**, **Prodyut Borah**, **Molly Piersol**, Deloitte Insights team.

Cover image by: **Jaime Austin**; Getty Images, Adobe Stock

Deloitte.
Insights

端侧生成式人工智能能否助力智能手机市场复兴

凭借特殊芯片和移动操作系统的广泛集成，智能手机可真正实现大智慧，但用户是否愿意为智能交互的革新买单？

智能手机是全球使用最广泛的消费技术产品。[1]在不断融合其他设备功能的同时，其先进和微型化组件也广泛应用于无数消费和工业设备。[2]其随手可得的便利性和实用性重塑了用户行为与市场竞争格局。但尽管如此，近年来的智能手机创新似乎逐渐令市场失望，都是些增量改进，缺乏变革性创新。

目前，主流移动生态系统供应商开始围绕下一代操作系统和先进芯片重新设计产品，打造以生成式人工智能为中心的智能手机体验。[3]越来越多的原始设备制造商 (OEM) 开始推出支持生成式人工智能的智能手机。[4]供应商期待搭乘生成式人工智能的快车，实现智能手机的再次复兴，但机遇往往与风险并存。

德勤预测，到2025年，全球智能手机出货量将实现约7%的小幅增长，高于2024年约5%的年增长率。[5]部分增长量可能来自因消费者升级到最新机型而重置的换机周期。另一部分增长动力可能来自技术狂热的早期使用者和开发人员，他们对即将面世的新一代智能手机充满期待，因为这些手机搭载有特殊设计的芯片，能够支持在本地设备上运行生成式人工智能。德勤进一步预测，到2025年底，具备生成式人工智能功能的智能手机出货份额将超30%。[6]

生成式人工智能的引入万众瞩目，但它能否兑现承诺的能力？**用户能否接受使用这一方式与智能手机进行新的智能交互**？[7]

生成式人工智能能否开启新一轮换机周期

短期内，领先智能手机品牌的设计师或会将生成式人工智能技术作为卖点，以此来提振其高端机型的需求。在2024年之前，智能手机销量已经连续两年下滑，[8]部分原因在于市场已经出现一定程度的饱和。据估计，目前拥有智能手机的人数近50亿，超过全球总人口的一半。[9]近年来，消费者的换机周期逐渐拉长，平均每两到三年才会更换一次手机。越来越多的家庭表示感受到通胀压力，从而限制了其可支配开支[10]。与此同时，越来越多的消费者为延长使用年限，特意选择高端机型。[11]这也对智能手机提出了更严格的市场期望，不仅要有技术上的提升，比如优质硬件，还要在用户体验上下功夫，提供更深层次的价值和使用便捷性。

2024年第一季度，由于消费者信心日益增强以及对配备生成式人工智能的高端机型表现出初步浓厚兴趣，智能手机销量出现了强劲增长。[12]这一点在德勤2024年互联消费者调研中也得到印证，调研显示，目前更多家庭认为，经济能力问题不会影响其购买互联设备。[13]这一消费复苏趋势似乎也延续到了欧洲市场，2024年第二季度，欧洲智能手机销量迎来持续增长。[14]因此，2025年有望迎来换机潮，更多消费者可能会更换其现有的智能手机，其中大部分还会升级购买具备生成式人工智能功能的高价位高端机型。

值得关注的是，虽然生成式人工智能技术可能成为消费者更换手机的一大诱因，但各区域市场和各年龄群体的换机意愿却有所不同。德勤2024年互联消费者调研显示，7%的美国受访者声称生成式人工智能技术可能会促使其提前更换智能手机，但在24至45岁的受访者中，这一比例跃升至50%，这可能是由于这一年龄段的消费者更加依赖智能手机，也更易接受新技术。[15]然而，根据德勤英国发布的《2024年数字消费者趋势报告》，只有4%的英国受访者表示每天都在使用生成式人工智能，23%认为该技术用处不大，19%表示对生成式人工智能给出的答案不满意。[16]

生成式人工智能能否进极大推动智能手机的升级换代？答案在于该项技术到底能带来多大的价值和实用性。**预计在2025年，智能手机将在实际应用中对生成式人工智能的实际效用进行全面验证**。

搭载生成式人工智能的新一代个人电脑

对用户体验、实用性和价值的考量，以及影响超大规模生成式人工智能发展的广泛压力，也同样适用于配备生成式人工智能端侧芯片的新一代个人电脑。

德勤2024年互联消费者调研表明，消费者对购买集成生成式人工智能功能的个人电脑兴趣浓厚，在计划升级笔记本电脑或个人电脑的美国受访者中，34%认同生成式人工智能芯片可能会促使他们提前购买计划。德勤认为，个人消费者贡献了约50%的个人电脑年销售额，因此他们的高购买意愿是重要增长因素。[17]对于企业客户来说，不同的个人电脑厂商提供了不同价位的各种选择，采用哪种生成式人工智能协处理器最匹配其业务需求还存在一定的不确定性。[18]

预计随着时间的推移，大多数高端个人电脑将载配特殊芯片，集成生成式人工智能功能。据估计，到2028年，80%的个人电脑将载配这种芯片。[19]另一预测数据显示，2024年第二季度将有近900万台“具备人工智能能力”的机器出货，但尚不清楚其中有多少设备载有足以运行生成式人工智能工作负载的神经处理单元 (NPU) 。[20]事实上，潜在客户或会观望一年左右，等待下一代设备的性能提升后再升级换代。

德勤预测，在2024年售出的所有个人电脑中，约30%将具备一定的本地生成式人工智能处理能力[21]，在2025年这样的电脑将近占销量的一半。

预计2024年电脑销量约为2.61亿台[22]，智能手机销量约为12.3亿台[23]，尽管电脑市场规模不如智能手机市场，但电脑的平均售价较高，其经济价值不容小觑。德勤估计，2024年电脑销售额约为2,200亿美元，只比同期智能手机5,200亿美元的销售额略显逊色[24]。

目前尚不明确集成生成式人工智能功能的电脑会对个人电脑行业产生怎样的影响。我们认为，这类设备将抬升个人电脑的平均售价，每台增加约15%的溢价。[25]但尽管如此，预计2025年的个人电脑销量将仅有个位数的增幅。[26]

对消费者而言，智能手机和个人电脑组件的发展或将重塑供应链并降低成本，以实现其在更多设备上的规模应用。生成式人工智能功能将有望在各类互联设备中变得更加普及。

生成式人工智能，实现手机的真正智能

智能手机中的“智能”通常指其可以接入网络并运行应用程序。生成式人工智能或会通过全新的交互方式，使智能手机变得更加个性化，能够感知用户的互动和意图，并通过对话界面等建立亲密关系。虽然此前的语音助手未能达到预期，但部分用户已经在尝试最新的会话式大型语言模型 (LLM) 。[27]这将可能形成一种全新的交互范式，即使用对话式人工智能与数字系统进行交互，同时，这也预示着一种新型可信赖智能代理 (Intelligent agent) 模式的出现，它们能够自主学习并有效地代表用户执行任务。

端侧生成式人工智能模型可以通过推断用户意图、读取用户日历和位置信息并规划最佳路线，来回答“下午2点有约，我应提前多久出发”这样的问题。预计端侧智能模型将专注完成较高精确性的任务，利用高性能的神经处理单元（NPU）（估计每秒至少可进行30万亿次操作[28]）来支持端侧的智能推断。该模型还可进一步识别用户问题是否超出本地处理范围，如超出，即将任务上传至大规模云端模型进行更好地解答。这种混合型的高性能移动计算方法既能实现在设备端进行即时、安全的互动，也能支持直接访问云端模型。[29]

通过在设备上运行的小型智能模型，可根据需要在本地保存并保护用户交互和数据，并支持更多可能需要快速响应的低延迟操作，如实时翻译。[30]此类功能或可帮助获取用户信任，并提供更多显见的实用价值。供应商还可从用户交互中获得新的数据飞轮，反馈到本地和云端模型中以持续优化用户体验，同时提供更深入的业务洞察。

我们预计，作为消费者交互中枢的智能手机未来将变得更加个性化和智能化，经调校可适应个人行为并预测用户需求。请参见德勤《2025科技、传媒和电信行业预测》——“处于研发阶段的自主生成式智能体”章节。这种“代理”功能可推动智能手机，并连带其经常交互和逐渐改变的设备生态系统，从单纯的“智能”进化到“智慧”。

未来一年用户将迫不及待地体验生成式人工智能，测试其早期功能的价值和可理解性。供应商有望在未来几个月内推出新功能，但他们普遍认为大规模应用仍需要一定的时间。[31]在未来一年，那些在设备端而非云端运行的小型模型将成为市场关注对象，我们会主要测验它们的运行能力和限制。假以时日，这可能会改变生成式人工智能的经济模式。如果更多的生成式人工智能任务得以从昂贵的数据中心转移到消费设备端侧，那么相关基础设施建设所需的资本投入将会减少。

行业能否通过资金投入促成生成式人工智能的实质性发展？

市场迫切要求通过前沿模型的产品适用性来验证其成本的合理性，并提升这些模型在构建和运营过程中的成本效益。[32]在这一背景下，领先的供应商已斥资数十亿美元用于开发最新的前沿模型，并继续追加数十亿投资来建设其认为满足大规模需求所必需的数据中心。[33]据估计，每年对生成式人工智能的投资达到6,000亿美元。[34]**然而，还需长期维系这样的密集资本投入才可见到经济价值回报**，而经济价值又要求更优质的产品来匹配。

想要降低模型成本，可能需要缩小模型、精简模型所需的数据量，以及根据工作负载的范围对模型进行拆分，尤其是对于可随着用户使用而扩展的推断型任务。消费者和企业用户产生的许多任务都会用到生成式人工智能，这类型任务或可通过相对便宜和更具效益的小型模型来解决，还能因此得到加强或提升。

但现在尚不明确有多少推断型任务需保留在设备端侧运行。目前对生成式人工智能的交互体验和用户期望大多由公共云端模型决定。用户可能需要时间来了解哪些任务和提示可以在本地安全、免费地运行；哪些需要联网接入云端模型。与会话式的端侧智能代理和云上智能代理的交互是一种新的智能交互范式，其对用户采用和用户行为的影响仍需时间来给出答案。

生成式人工智能的广泛应用仍面临挑战

德勤《2024年互联消费者调研报告》显示，38%的美国受访者使用过生成式人工智能，其中63%的用户表示这些技术超出预期。[35]许多使用过生成式人工智能的用户已经尝到甜头，但供应商可能仍需向更广大的消费者人群展示生成式人工智能的效用，以证明新智能手机高溢价的合理性。

对于在智能手机上使用生成式人工智能，用户可能会因要尝试新的交互方式而感到困惑，可能会犹豫是否要将自己的代理权让给智能助手，例如让其管理日程。[36]随着新技术的使用，我们会考虑电池消耗、集成公共模型的成本以及未察觉的不实信息等问题，这些可能会对高价值用例产生不利影响。**要在用户、其私人智能体和公共模型之间建立信任需要时间的累积，而摧毁信任则往往只需要一瞬间**。

供应商希望下一代前沿模型能释放出更大的价值，但目前尚不明确前沿模型的能力是否会继续增强，亦或是不再发展甚至倒退。是否有足够的数据来支持数据需求越来越大的智能训练模型？[37]由模型创建合成数据进行自我训练等解决方案可能会导致人工智能的推断质量随时间推移而下降。[38]能否在不增加数据、训练和推断成本的情况下提高人工智能性能？是否存在降低资本和数据强度但又增强模型功能的机遇？面对这些压力，投资者或会在技术收益实现之前就要求获得更高的回报收益。

监管机构也可能会实施更多举措来严格规范生成式人工智能的发展，以防范出现新的弊端，如深度伪造、错误信息和具有说服能力的类人机器人。对话型机器人可能会与用户建立更亲密的关系，从而更深刻地影响用户的想法和意识形态。[39]个性化的对话型智能代理可以与用户进行更深层次的交互，在提供更多帮助的同时，也有可能让人沉迷上瘾。[40]端侧生成式人工智能搭配第三方模型使用，有可能会产生更大范围的安全漏洞。[41]这些都可能会迫使供应商进一步增强生态系统安全防护，并促使监管机构设置更多防护规则。

生成式AI成为中国智能手机差异化亮点

近年来，中国智能手机市场的微创新，诸如对外观细节的微调、小功能的迭代，已难以激起消费者的兴趣。在此背景下，头部厂商纷纷将生成式 AI 融入智能手机，试图开辟差异化竞争新赛道。目前，生成式 AI 在国内智能手机中的应用，主要集中在实时语音文本翻译、文档处理、全局搜索、影像修图等方面。未来有望引入更多创新功能。比如，能精准感知用户习惯、提供个性化建议的更加智能的助手；或是实现多任务流畅切换、大幅提升效率的多任务处理能力，为智能手机市场带来全新变革，进一步推动端侧AI市场的发展。

小结

尽管行业都在热议“下一代智能手机”是有可能推动整体市场变革和升级的消费设备平台，但它还并未实际到来。而拥有数十亿用户的智能手机仍然占据市场主导地位，并为新的服务和用户交互提供了巨大的试验平台。2025年，生成式人工智能用户或将迎来迅猛增长，他们会通过高端智能手机和个人电脑尝试使用这一技术，了解其价值并测试其优势。如果端侧的生成式人工智能成功获得市场认可，智能手机则将再次焕发光彩，它将拓宽平台功能，支持全新用途类别和商业机遇，推动个人设备的新一轮繁荣。但这或将需要较长时间才能实现，预计未来一年中，行业将开始逐步引导用户了解并接受这一个人计算领域的新范式。

未来几年，智能手机操作系统将会进行更多的人机交互，比如下一代对话式搜索，该功能能够在本地返回更多信息汇总，而不是只提供远程链接，从而直接连接用户与服务提供商及信息源头，实现去中介化的信息检索新体验。如果用户采用更加个性化的代理式人工智能，数字交互的性质将发生改变，本地设备上的智能代理将执行更多任务，无需用户进行直接的界面操作。如此一来，**计算可以变得更加隐秘而优雅，在后台悄然执行用户任务，且更具空间感知能力**，更易感知我们周围的环境和网络交互。

供应商在努力刺激市场需求的同时，会发现其正面临着巨大的经济压力，必须加快节奏以抵消大规模训练和运行智能模型的密集资本投入与能源成本。为此，行业可开发小型模型和混合架构，并深化对生成式人工智能各类工作负载的了解，准确掌握各项工作负载所需的各项计算开销。当前，气候不确定性和生态焦虑是各行业不能绕开的课题，而生成式人工智能数据中心的建设已然增加了能源和水资源的消耗，加重了普通家庭和市政的能源负担。[42]即便生成式人工智能能够克服经济债务的困境，也还会因背负能源债务而遭遇发展瓶颈。

截至2024年底，超大规模计算服务供应商、智能手机生态系统供应商和新兴公共智能模型供应商都赌定其提供的智能优势能够转化为广泛的经济价值。但其究竟价值几何？生成式人工智能是否会步电信和早期互联网的后尘，**它们斥巨资建立的基础设施，却最终为下一代创新者做了嫁衣**？[43]

生成式人工智能的诞生、部署和广泛应用，可能是自互联网普及以来人类社会启动的一项最为宏大的实验。其如同登月计划，不论最终结果如何，其发展势必激起大量新技术、新行为和新商业模式的涌现。

By

Chris Arkenberg
United States

Duncan Stewart
Canada

Roger Chung
China

Gillian Crossan
United States

Kevin Westcott
United States

尾注

1. GSM Association, "*Smartphone owners are now the global majority, new GSMA report reveals*," press release, Oct. 11, 2023.

2. Wolfgang Bock, François Candelon, Steve Chai, Ethan Choi, John Corwin, Sebastian DiGrande, Rishab Gulshan, David Michael, and Antonio Varas, "*The mobile revolution: How mobile technologies drive a trillion-dollar impact*," Boston Consulting Group, Jan. 15, 2015.

3. IDC Corporate, "*The future of next-gen AI smartphones*," Feb. 19, 2024.

4. Counterpoint, "*Gen AI-capable smartphone shipments to grow over 4x by 2027*," April 16, 2024.

5. IDC Corporate, "*Worldwide smartphone market up 7.8% in the first quarter of 2024 as Samsung moves back into the top position, according to IDC tracker*," press release, April 15, 2024.

6. IDC anticipates a 364% compound annual growth rate in 2024 (from a low base in 2023) for global gen AI smartphone shipments, with 73% growth in 2025. Canalys expects AI-enabled smartphone market share to reach 54% by 2028. Our analysis, for reasons outlined in this paper, is less bullish than the former, and a bit more than the latter. Sources: IDC Corporate, "*The future of next-gen AI smartphones*"; Canalys, "*Now and next for AI-capable smartphones*," accessed Oct. 30, 2024.

7. Jim Fellinger, "*CTA study: Smartphones most-owned tech, 5G and wireless drive adoption*," press release, Consumer Technology Association, May 31, 2023.

8. IDC Corporate, "*Worldwide smartphone market up 7.8% in the first quarter of 2024 as Samsung moves back into the top position, according to IDC tracker*."

9. GSM Association, "*Smartphone owners are now the global majority, new GSMA report reveals*."

10. Sarah Barry James, "*Consumer checkup: Higher interest rates lead to longer tech replacement cycles*," S&P Global, March 26, 2024.

11. IDC Corporate, "*Worldwide smartphone market up 7.8% in the first quarter of 2024 as Samsung moves back into the top position, according to IDC tracker*."

12. Chris Donkin, "*Smartphone sales up again ahead of expected gen AI boost*," Mobile World Live, July 15, 2024.

13. Susanne Hupfer, Michael Steinhart et al., "2024 Connected Consumer Study," *Deloitte Insights*, publication forthcoming, 2024.

14. Counterpoint, "*Europe smartphone market recovery continues, shipments up 10% YoY in Q2 2024*," Aug. 28, 2024.

15. Susanne Hupfer, Michael Steinhart et al., "2024 Connected Consumer Study," *Deloitte Insights*, publication forthcoming, 2024.

16. Deloitte, "*Generative AI: 7 million workers and counting*," June 25, 2024.

17. The installed base of PCs is estimated to be about 2 billion, and there are about 1 billion knowledge workers, suggesting that the market is roughly half consumer and half enterprise.

18. Author interviews with enterprise chief information officers in July and August 2024.

19. Canalys, "*AI-capable PCs forecast to make up 40% of global PC shipments in 2025*," March 18, 2024.

20. Ibid.

21. Deloitte Global analysis of publicly available information for H1 2024, and extrapolation based on usual PC seasonality trends.

22. IDC Corporate, "*PC refresh cycle and tablets in emerging markets expected to spur demand in coming quarters, according to IDC*," press release, Sept. 23, 2024.

23. IDC Corporate, "*Worldwide smartphone market forecast to grow nearly 6% in 2024, driven by stronger growth for android in China and emerging markets, according to IDC*," press release, Aug. 27, 2024.

24. Based on quarterly data so far in 2024, Deloitte believes smartphone average selling price is declining and should be roughly US$425 for the year. PC average selling prices were high during the 2021 chip shortage, but are declining and Deloitte estimates them to be about US$850 for 2024.

25. Roshan Ashraf Shaikh, "*Analysts expect 15% price hike for AI PCs—60% of PCs will have local AI capabilities by 2027*," Tom's Hardware, April 26, 2024.

26. IDC Corporate, "*PC refresh cycle and tablets in emerging markets expected to spur demand in coming quarters, according to IDC*."

27. Sigal Samuel, "*People are falling in love with—and getting addicted to—AI voices*," Vox, Aug. 18, 2024.

28. IDC, "*The future of next-gen AI smartphones*."

29. Baris Sarer, Mark Szarka, Nataliia Bacchus, and Edem Isliamov, "*The world of hybrid AI*," *The Wall Street Journal* and Deloitte, July 31, 2024.

30. Malik Saadi, "*On-device generative AI unlocks true smartphone and PC value*," *Forbes*, April 17, 2024.

31. Lisa Eadicicco, "*AI is changing our phones, and it's just getting started*," CNET, April 3, 2024.

32. Goldman Sachs, "*Gen AI: Too much spend, too little benefit?*" June 27, 2024.

33. David Cahn, "*AI's US$600B question*," Sequoia, June 20, 2024.

34. Ibid.

35. Susanne Hupfer, Michael Steinhart et al., "2024 Connected Consumer Study," *Deloitte Insights*, publication forthcoming, 2024.

36. Jon Victor, "*Software firms race to beat OpenAI in AI agents*," The Information, Sept. 26, 2024.

37. Deepa Seetharaman, "*For data-guzzling AI companies, the internet is too small*," *The Wall Street Journal*, April 1, 2024.

38. Michael Peel, "*The problem of 'model collapse': How a lack of human data limits AI progress*," *Financial Times*, July 24, 2024.

39. Yuval Noah Harari, "*Yuval Noah Harari argues that AI has hacked the operating system of human civilization*," *The Economist*, April 28, 2023.

40. CBS News, "*Virtual valentine: People are turning to AI in search of emotional connections*," Feb. 14, 2024.

41. Matt Burgess, "*Generative AI's biggest security flaw is not easy to fix*," *Wired*, Sept. 6, 2023.

42. Camilla Hodgsin, "*US tech groups' water consumption soars in 'data center alley'*," *Financial Times*, Aug. 17, 2024.

43. Bryce Elder, "*Gen-AI revisited, by Goldman Sachs*," *Financial Times*, Sept. 5, 2024.

致谢

Authors would like to thank **Rohan Gupta** and **Steve Fineberg**.

Cover image by: **Jaime Austin**; Getty Images, Adobe Stock

Deloitte.
Insights

处于研发阶段的自主生成式智能体

自主生成式智能体 (Autonomous Gen AI agents)——代理式人工智能 (Agentic AI)——不仅能提升知识工作者的工作效率，还可实现各类工作流程的高效运作。然而，由于代理式人工智能的“自主性”特征，其广泛应用尚待时日。

自主生成式智能体，亦称作“代理式人工智能”，是一种软件解决方案，能够在极少或没有人工监督的情况下，完成复杂任务并实现既定目标。代理式人工智能不同于聊天机器人和人工智能助手，后者本身通常被称为“代理”。代理式人工智能拥有巨大的潜力，它能够提升知识工作者的工作效率，同时助力实现跨业务部门的多步骤流程自动化。德勤预计，到2025年，在已部署生成式人工智能的企业中，将有25%的企业开展代理式人工智能的试点项目或进行概念验证。到2027年，这一比例将增至50%。[1]到2025年（特别是下半年），在某些行业和用例中，一些代理式人工智能应用程序将被实际应用于现有的工作流程中。

这些战略举措获得了初创公司和知名科技公司的支持，此类公司致力于代理式人工智能技术开发，并且对该技术刺激收益增长的潜力持乐观态度。在过去两年里，投资者向代理式人工智能初创公司进行注资，投资总额超20亿美元，其中大部分资金流向了面向企业市场的公司。[2]与此同时，许多科技公司、云服务提供商及其他企业也在开发各自的代理式人工智能产品。此外，这些公司还开展战略并购活动，并逐渐倾向与初创公司签订代理式人工智能技术许可协议，或从初创公司招募相关人才，而非直接收购公司。[3]

代理式人工智能具有“代理功能”

生成式人工智能驱动的聊天机器人和人工智能助手是一项先进技术，它们可以直接与人类进行交互，合成复杂的信息并生成内容。然而，相较于代理式人工智能所承诺的代理功能和自主性水平，聊天机器人和人工智能助手还有所不足。尽管聊天机器人和智能体的构建均基于大型语言模型 (LLM)，但后者却需要更多的技术与工艺，才能实现独立运作，将任务分解为不同步骤，并在最少的人工监督或干预下自主完成。智能体的功能不仅限于交互，它们还能代表用户进行有效推理并采取行动。

顾名思义，代理式人工智能具有“代理功能”：能够自主采取行动并选择具体策略。[4]同时，代理功能意味着自主性，即独立行动和决策的能力。[5]当我们将这些概念延伸至代理式人工智能时，便意味着代理式人工智能拥有自主规划、执行并实现既定目标的能力——换言之，它扮演了“代理角色”。[6]虽然目标是由人类设定，但代理式人工智能能够自主决定实现目标的路径。

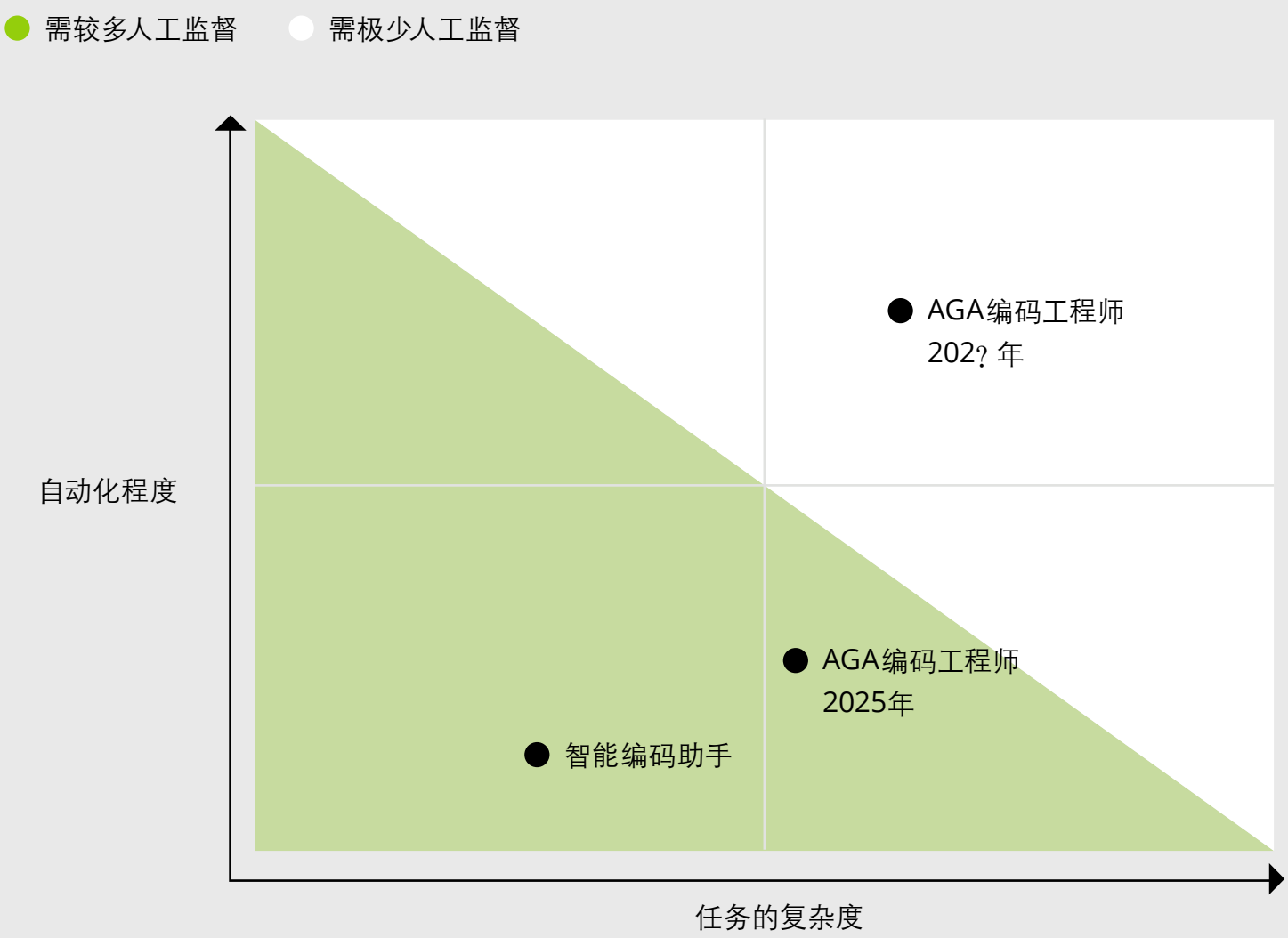
代理式人工智能与人工智能助手及聊天机器人之间的差异示例如下。人工智能助手能够在代码测试和建议方面为软件开发人员提供协助，是迄今为止最成功的生成式人工智能用例之一。[7]它提高了资深软件工程师和初级程序员的工作效率。人工智能助手可以将自然语言提示（支持多种语言）转化为代码建议，并测试代码的一致性，但它只能响应工程师的指令，缺乏自主代理能力。随着代理式人工智能的出现，“软件工程师”这一角色得到了进一步延伸。人类程序员仅需输入有关软件构思的提示，代理式人工智能“软件工程师”便能将这些构想转化成可执行代码，进而实现软件开发流程中多个步骤的自动化。

例如，Cognition软件公司在2024年3月推出了一款名为“Devin”的产品，旨在创建一款自主软件工程师，能够进行推理、规划并完成需要数千个决策的复杂工程任务。[8]Devin的设计理念是根据人类程序员的自然语言提示，独立完成编程工作，包括设计完整的应用程序、测试和修复代码库，以及训练和微调大型语言模型。[9]与此同时，其竞争对手Codeium——一家专注于企业级软件开发的公司——推出的相关产品以及Devin的开源版本，均在2024年夏季推向市场。[10]

代理式人工智能软件工程师具有类似的能力和弱点。[11]其中一项弱点在于目前的错误率较高，导致它们在没有人工监督的情况下，难以独立完成全部或部分任务。在近期的一项基准测试中，Devin成功解决了现实世界代码库中14%的GitHub问题，是基于LLM的聊天机器人的两倍，[12]但仍未实现完全自主。众多科技巨头[13]和初创公司正不懈努力，致力提高代理式人工智能软件工程师的自主性和可靠性，从而让人类程序员及其雇主能够放心地将部分工作交由代理式人工智能软件工程师处理（图1）。

图1

生成式智能体正不断发展



资料来源：德勤，2024年11月

提高工作效率

代理式人工智能软件工程师只是自主生成式智能体转变工作模式的示例之一（有关自主生成式智能体的应用前景，请参阅下文）。随着代理式人工智能的不断发展，其展现出了巨大的影响力。在美国，知识工作者的数量超1亿，而在全球范围内，这一数字更是高达10.25亿。[14]全要素生产率[15]作为衡量知识工作成效的重要指标，其增长却陷入停滞。在美国，该指标在1987年至2023年的增幅仅为0.8%，而2019年至2023年的增幅更是低至0.5%。[16]在多数经合组织国家，情况也是一样。[17]通过任务自动化提高知识工作效率的举措仅取得了部分成效。许多企业仍面临知识工作者的需求缺口。客服代表、半导体工程师等职位的人才短缺问题始终存在。此外，新员工入职后，需迅速提高工作效率。

在流程不明确或涉及多个步骤的情况下，专家系统和机器人流程自动化（RPA）可能会出现问题。基于传统机器学习的系统需进行大量训练，以应用于特定用途。相较于机器学习或深度学习，基于LLM的代理式人工智能具有更高的灵活性，能够处理更为广泛的用例。

代理式人工智能显著增强了LLM的能力，同时也证明企业投资于生成式人工智能技术的正确性。生成式人工智能工具的正式推出，迅速引起了高管们的关注，他们不难构想出其所在企业对这项技术的运用。然而，生成式人工智能所蕴含的可量化商业价值往往难以实现。数据基础、风险、治理政策以及人才缺口等方面的多重挑战，均会阻碍企业对生成式人工智能项目的规模化扩展。[18]仅30%的生成式人工智能试点项目全面投产。[19]对于生成式人工智能的产出缺乏信任，以及生成式人工智能一旦出错，将对“现实世界”所产生的潜在影响，都使高管们对该项技术持保留态度。[20]

企业在研发和部署代理式人工智能时，需考量生成式人工智能带来的挑战，以及构建具备推理、执行、协同和创造力的机器人所涉及的复杂性。最重要的一点在于，各类生成式智能体必须足够可靠，企业方能安心使用，这就意味着仅在多数情况下顺利完成任务还远远不足。2024年末，部分代理式人工智能的用例和应用在可靠性方面呈现出积极态势，表明其足够可靠，可在2025年初推广使用。

这项技术的潜在回报值得我们为之付出努力，初步成果亦令人振奋。各大公司正在探索如何将这些模型与其他人工智能技术及训练方法相结合，以提升LLM的性能。尽管构建自主且可靠的代理是生成式人工智能的目标，而逐步提高准确性与独立性将助力企业总体实现生成式人工智能在生产力和效率提升方面的初期目标。[21]随着代理式人工智能技术在应用层面的广泛与深入扩展，并拥有明确的业务目标，代理式人工智能越来越接近高管们最初设想的生成式人工智能解决方案。

深入了解何为生成式智能体

代理式人工智能能够将复杂任务拆分为一系列步骤，再逐一执行，并解决突然出现的难题。此外，代理式人工智能还能感知环境，根据不同的用例，可以是虚拟环境、物理环境，或是二者的结合。在执行任务的过程中，代理式人工智能能够自主决定采取何种行动，并向各类工具、数据库或其他代理寻求支持，最终根据人类设定的目标交付成果。

代理式人工智能是一项新兴技术，仍处于不断发展中，但已展现出一些共有特性与能力：

- **构建于基础模型之上**：LLM等基础模型赋予了代理式人工智能推理、分析并适应复杂且不可预测工作流程的能力。相较于RPA与专家系统，代理式人工智能具有更高的灵活性。目前，LLM正处于快速发展之中，取得的新进展包括推理能力增强、能够将任务拆分为细分步骤等。[22]然而，基础模型本身无法与环境交互、做出决策或执行任务。[23]它们必须通过其他技术和能力来增强。

- **自主行动**：尽管不同代理式人工智能的自主程度存在差异，但经过精心训练，它们能够基本独立完成复杂任务的规划与执行。通过引入推理标记，思维链模型能够比以前的LLM应对更为复杂的挑战。虽然思维链模型的响应速度较慢，但其在推理问题时更加慎重，不仅能够自我纠错，还能展示出为找到解决方案所采取的步骤。[24]

- **环境感知**：代理式人工智能具备感知环境、处理信息并理解任务背景的能力。[25]高级代理式人工智能能够处理多模态数据，包括视频、图像、音频、文本及数值信息。

- **工具运用**：代理式人工智能能够与各类工具和系统（如软件、企业应用程序和互联网）进行交互，从而完成任务。

- **协同作业**：代理式人工智能具备指挥其他系统及机器人共同完成任务的能力。对于多代理系统，这意味着需与其他自主生成式人工智能进行协作。

- **内存访问**：LLM是无状态的，即每次交互均为独立处理，一旦交互结束，相关信息均不予以保存。然而，代理式人工智能通过增加检索机制与数据库，能够访问短期内存，进而在执行特定任务时保留上下文，同时还能访问长期内存，从中吸取经验并进行自我完善。[26]

部分最新模型采用了思维链功能，虽然比从前的大规模模型速度更慢、更审慎，但能对复杂问题进行更高阶的推理。[27]多模态数据分析丰富了可解释和生成的数据种类，从而提高代理式人工智能的灵活性。多模态人工智能还表明，当与其他类型的人工智能技术（如计算机视觉（图像识别）、转录和翻译）相结合时，代理式人工智能可以变得更加强大。[28]与代理本身一样，多模态人工智能仍处于发展阶段。

真正的多代理系统在自主代理网络之间协作工作，目前正在开发中，一些试点项目将于2024年底启动。[29]多代理模型在分配任务方面（尤其是在复杂环境中）优于单一模型系统。[30]初创企业和大型科技公司正在开发多代理生成式人工智能系统，包括协助企业建立自有定制代理的工具。[31]

自主生成式智能体用例前景

大型科技公司和初创企业正在开发早期阶段的解决方案，可实现软件开发、销售、营销和监管合规等功能的部分自动化。以下仅为目前应用实例速览，而非详尽清单。部分实例基于概念验证和演示，虽然很有前景，但还不能用于企业部署。虽然这些实例是跨行业的，但特定行业的代理应用也在不断涌现。

客户支持：客户服务是一项必不可少的工作，但往往压力很大，每年的离职率为38%。[32]有效实现部分客户支持工作流程的自动化，可以减轻员工的压力和枯燥感，帮助公司服务更多客户。[33]相较于现下的客户支持聊天机器人，代理式人工智能能够应对更复杂的客户咨询，且能自主解决问题。例如，一家音频公司正在使用代理式人工智能辅助客户设置新设备，该过程包含多个步骤，通常需要人工代理。如需人工代理，代理式人工智能在转接客户之前，将汇编相关信息并总结问题。[34]下一代客户支持代理除了文字聊天外，还可能整合语音及视频等多模态数据。

网络安全：网络安全专家是熟练知识工作者短缺的缩影：目前全球缺口为400万。[35]与此同时，恶意行为者正在利用生成式人工智能渗透网络安全系统。新兴代理网络安全系统将人类专家的部分工作自动化，以提高其工作效率。它们可以自主检测攻击并生成报告，进而提高系统安全性，并将人类专家的工作量至多减少90%。[36]代理式人工智能还能协助软件开发团队检测新代码中的漏洞。它可以运行测试，并直接与开发人员沟通，解释如何修复问题——这正是目前人类工程师须手动完成的工作。[37]

监管合规：金融服务和医疗保健等各行业公司需定期进行合规审查。相关法规的范围和复杂性不断增加，而合规专业人士紧缺，使合规成为日益严峻的挑战。初创企业正开发代理式人工智能，能够分析法规和公司文件，并快速判断公司是否合规。代理可引用具体法规，主动向人类监管专业人员提供分析及建议。[38]当前使用生成式人工智能的公司认为，监管合规是开发及部署生成式人工智能的首要障碍，其次是缺乏人工智能技术人才及实施过程中面临的挑战。[39]监管的不明确性有一定影响，但新法规的范围和复杂性也不容忽视。更具代理性质的人工智能解决方案通过帮助企业了解并遵守已颁布的法规，从而助力企业广泛应用生成式人工智能。

代理构建器和协调器：代理式人工智能解决方案兴起，助力实现其他跨行业和特定垂直行业工作流程的自动化。不过，企业不一定要等待市场，它们可以构建自己的代理和多代理系统。借助谷歌的Vertex，公司可以使用无代码工具为特定任务创建代理，例如根据以前的营销活动制作营销材料。[40]LangChain利用开源技术协助企业构建多代理系统。例如，初创公司Paradigm推出“智能电子表格”，其中多个代理式人工智能开展协作，从不同来源收集数据、构建数据并完成任务。[41]

中国智能体发展势头迅猛

国内互联网巨头均发布了智能体相关战略，使得2025年有望成为中国“智能体年”。在应用场景方面，预计智能体将在客户服务（可以处理比当前客服聊天机器人更复杂的查询，并可以自主解决问题）、网络安全（自主检测攻击并生成报告，检测新代码漏洞）、合规（分析法规和公司文件，快速确定公司是否合规）、研发与创新（帮助研发人员进行数据分析，提高研发效率）等多方面落地。可以预见，随着技术进步与行业需求的双重推动，智能体将在中国的多个行业中逐步渗透，推动我国数字化转型和业务模式创新。

小结

代理式人工智能蕴藏着巨大潜力，通过自动化工作流程和离散任务来帮助提高知识工作者的效率。它既能作为单个代理独立行动，也能与其他代理协同合作，这使其有别于当前的聊天机器人和人工智能助手。然而，代理式人工智能还处于开发和应用的早期阶段。尽管早期的代理实例成果骄人，但也可能出错并陷入循环。在多代理系统中，“幻觉”会从一个代理传递至另一个代理；它们会说服其他代理采取错误的步骤并给出错误的答案。[42]虽然代理式人工智能主要为自主型，但通常需要人类对其做出的决策进行复核（即“human on the loop”，而非更具约束性的人机回圈“human in the loop”），使得代理式人工智能更适合当前部署。当生成式智能体陷入困境时，它们可以咨询人类专家，在专家帮助下解决难题并继续前进。在此模式下，代理式人工智能就像一名初级员工，在完成有价值工作的过程中，不断汲取经验，学习发展。[43]

虽然一些公司计划投资数十亿美元来创建稳定可靠的代理式人工智能，但何时投资，以及在什么情况下投资目前尚不明确。代理式人工智能是否将在2025年或未来五年内得到广泛应用？技术普及需要突破性创新，还是需要改进当前人工智能技术和训练方法？如果开发代理式人工智能的大型企业和初创公司取得成功，情况则将迅速发生改变。设想下，自主生成式智能体能处理多模态数据、使用工具、协调其他代理、记忆和学习，并能持续可靠地执行任务。再设想下，企业可以在“无代码环境”中，仅使用对话文本提示即可快速、轻松地开发定制代理。

代理式人工智能的发展前景广阔，且技术发展日新月异，企业应即刻做好准备。准备过程中应考虑以下方法。

优先考虑并重新设计代理式人工智能的工作流程：根据技术能力和公司的最高价值点，考虑哪些任务和工作流程适合由代理式人工智能执行。重新设计，去除不必要的步骤。确保代理式人工智能解决方案有明确的目标，并能访问所需的数据、工具和系统。虽然这些代理可以帮助其他代理适应环境，但流程冗杂且局部优化或将导致无法取得预期成果。

关注数据治理和网络安全：代理式人工智能要创造价值，须能访问重要且可能敏感的企业数据，以及内部系统和外部资源。在开始使用自主生成式智能体之前，公司应建立强大的数据治理和网络安全机制。生成式人工智能早期采用者增加信息技术投资的重点领域是数据管理（75%）和网络安全（73%）。[44]尽管进行了这些投资，仍有58%的公司高度关注在模型中使用敏感数据和管理数据安全的问题。23%的公司则表示已为管理人工智能风险和治理做好了充分准备。简而言之，目前许多生成式人工智能领导者似乎对代理式人工智能的到来毫无准备。如果还未做准备，那些仍在生成式人工智能领域观望的公司必然还需采取更多举措。

平衡风险与回报：开始使用代理式人工智能时，公司应考虑允许代理的自主程度和数据访问权限。涉及非关键数据和人工监督的低风险用例可以帮助公司为安全的代理式人工智能应用建立数据管理、网络安全和治理。一旦准备就绪，公司便应考虑使用战略数据、访问更多工具和更多自主权的高价值用例。

保持合理的怀疑态度：代理式人工智能不断发展，明年有望变得更加强大，并将应用于更多特定水平和垂直用例。预计2025年将发布出色的演示、模拟和产品。但是，我们注意到的挑战可能需要一定时间才能解决。在此之前，代理式人工智能在受控环境中的表现难以提高企业绩效。需谨慎评估并保持合理的怀疑态度。

By

Jeff Loucks
United States

Gillian Crossan
United States

Roger Chung
China

Baris Sarer
United States

China Widener
United States

尾注

1. According to Deloitte's *State of Generative AI in the Enterprise* survey, 23% of enterprises that currently use gen AI are exploring "gen AI agents" to a "large" or "very large" extent, with another 42% exploring it "to some extent." Given the high interest in agentic AI and the products and services that are being launched by startups and established tech companies, we expect this interest to turn to action, at least on an experimental scale.

2. CB Insights. Gen AI Investment Database, Aug 21, 2024. This data excludes Open AI. It includes funding to companies that are developing agentic AI with "varying degrees of autonomy."

3. Kate Clark, "*Investors undaunted by spate of AI acqui-hires*," *The Information*, Aug. 19, 2024.

4. Cambridge English Dictionary, "*Agency*," accessed Aug. 26, 2024.

5. Cambridge English Dictionary, "*Autonomous*," accessed Aug. 26, 2024.

6. For humans, agency and autonomy are moral and political concepts. In the context of gen AI agents, we are speaking only of the extent to which software-based technology has scope to design and perform tasks without human direction.

7. Faruk Muratovic, Duncan Stewart, and Prashant Raman, "*Tech companies lead the way on generative AI: Does code deserve the credit?*" *Deloitte Insights*, Aug. 2, 2024.

8. Scott Wu, "*Introducing Devin, the first AI software engineer*," Cognition Software, March 12, 2024.

9. Rina Diane Caballar, "*AI Coding is going from copilot to autopilot*," *IEEE Spectrum*, April 9, 2024.

10. Jenna Barron, "*Codeium's new Cortex assistant utilizes complex reasoning engine for coding help*," *SD Times*, Aug. 14, 2024; Aswin Ak, "*OpenDevin: An artificial intelligence platform for the development of powerful AI agents that interact in similar ways to those of a human developer*," *Marktechpost*, July 28, 2024.

11. Carl Franzen, "*Codium announces Codiumate, a new AI agent that seeks to be Devin for enterprise software development*," *VentureBeat*, April 3, 2024.

12. Cognition Software, "*SWE-bench technical report*," March 15, 2024.

13. Big tech companies continue to improve their software co-pilots to make them more like gen AI agents. For example, see: Alex Woodie, "*The semi-autonomous agents of amazon Q*," *BigDATAWire*, May 3, 2024.

14. Molly Talbert, "*Overcoming disruption in a distributed world: Insights from the Anatomy of Work Index 2021*," Asana, January 14, 2024.

15. Total factor productivity, which measures how efficiently both capital and labour are used, can be a proxy for knowledge worker efficiency. Knowledge work requires access to capital-intensive technology, and effectively designed processes.

16. US Bureau of Labor Statistics, "*Table A. Productivity, output, and inputs in the private nonfarm business and private business sectors for selected periods, 1987-2023*," March 3, 2024.

17. Organisation for Economic Co-operation and Development, "*Multifactor productivity*," accessed Oct. 30, 2024.

18. Jim Rowan, Beena Ammanath, Costi Perricos, Brenna Sniderman, and David Jarvis, *State of gen AI in the Enterprise*, Q3 report, Deloitte, August 2024.

19. Ibid.

20. Ibid.

21. Ibid.

22. James O'Donnell, "*Why OpenAI's new model is such a big deal*," *MIT Technology Review*, Sept. 17, 2024.

23. Janakiram MSV, "*AI agents: Key concepts and how they overcome LLM limitations*," *The New Stack*, June 11, 2024.

24. "OpenAI, "*Learning to Reason with LLMs*," Sept. 12, 2024.

25. Anna Gutowska, "*What are AI Agents?*" IBM, July 3, 2024.

26. Janakiram MSV, "*AI agents: Key concepts and how they overcome LLM limitations*."

27. Simon Willison, "*Notes on OpenAI's new o1 chain-of-thought models*," Simon Willison's Blog, Sept. 12, 2024.

28. Hamidou Dia, "*So much more than gen AI: Meet all the other AI making AI agents possible*," Google Cloud Blog, Aug. 20, 2024.

29. Vivek Kulkarni, Scott Holcomb, Prakul Sharma, Edward Van Buren and Caroline Ritter, "*How AI agents are reshaping the future of work*," Deloitte AI Institute, November 2024.

30. *The Economist*, "*Today's AI models are impressive. Teams of them will be formidable*," May 13, 2024.

31. CB Insights, "*The multi-agent AI outlook: Here's what you need to know about the next major development in genAI*," Aug. 30, 2024.

32. Mike Desmarais, "*The call center burnout problem*," SQM Group, Feb. 24, 2023.

33. It's important to balance the work of human agents. When they get only the most complicated and difficult cases, it can lead to burnout. See Sue Cantrell, et al., "*Strengthening the bonds of human and machine collaboration*," *Deloitte Insights*, Nov. 22, 2022.

34. Sierra, "*Sonos elevates the listener experience*," Feb. 13, 2024.

35. Michelle Meineke, "*The cybersecurity industry has an urgent talent shortage. Here's how to plug the gap*," World Economic Forum, April 28, 2024.

36. Ken Yeung, "*Dropzone AI gets $16.85M for autonomous cybersecurity AI agents that reduce manual work by 90 percent*," *VentureBeat*, April 25, 2024.

37. Simon Thomsen, "*Software development cybersec startup Nullify banks $1.1 million pre-seed round*," *Startup Daily*, June 26, 2023.

38. Kyt, Dotson, "*Norm Ai raises $27M to help businesses handle regulatory compliance with AI agents*," *SiliconANGLE*, June 26, 2024.

39. Rowan, *State of Generative AI in the Enterprise*, Q3 report.

40. Ron Miller, "*With Vertex AI Agent Builder, Google Cloud aims to simplify agent creation*," *TechCrunch*, April 9, 2024.

41. Iris Coleman, "*Paradigm utilizes LangChain and LangSmith for advanced AI-driven spreadsheets*," *Blockchain.News*, Sept. 5, 2024.

42. *The Economist*, "*Today's AI models are impressive*."

43. Maria Korolov, "*AI agents will transform business processes — and magnify risks*," *CIO*, Aug. 21, 2024.

44. Rowan, *State of Generative AI in the Enterprise*, Q3 report.

致谢

Authors would like thank **Chris Arkenberg**, **Duncan Stewart**, and **Ankit Dhameja**.

Cover image by: **Jaime Austin**; Getty Images, Adobe Stock.

Deloitte.
Insights

深度伪造之战：网络安全的大规模挑战与深远影响

随着检测和打击虚假内容的力度持续加大，维护可信互联网的成本或由消费者、创作者及广告商共担。

深度伪造内容，即看似真实却是由人工智能工具生成的图片、视频和音频片段，加剧了公众对于网络信息的信任危机。随着人工智能生成内容的数量和质量不断提升，网络多媒体资源更易被不法分子利用以散布虚假信息和实施欺诈。社交媒体平台充斥着此类伪造内容，引发了公众的疑虑与担忧。[1]

根据德勤《2024年互联消费者调研报告》，有半数受访者表示，相较于去年，他们对网络信息的准确性与可靠性持更加怀疑的态度。在了解或使用生成式人工智能的受访者中，68%表示担忧合成内容可能被用于欺骗或欺诈目的，59%表示难以辨识人类创作与人工智能生成的内容。此外，高达84%了解生成式人工智能的受访者赞同，生成式人工智能生成的内容应始终注明其来源。[2]

标识是媒体机构与社交媒体平台向用户提示合成内容的一种方式。然而，随着深度伪造技术运用更先进的模型来生成合成内容或篡改既有媒体素材，可能需要采取更复杂的策略来检测虚假内容并助力重建公众信任。

分析人士预计，全球深度伪造检测市场——在科技、传媒和社交网络巨头的推动下——年增长率料将达42%，市场规模将从2023年的55亿美元增至2026年的157亿美元。[3]德勤预计，该市场的发展轨迹或与网络安全行业相仿。媒体公司及技术提供商或将通过投资于内容验证解决方案和建立联盟合作，以领先于不断进化的伪造手段。这对消费者、广告商乃至创作者而言，创作或获取可信内容的成本可能会增加。[4]

目前打击虚假内容的手段主要分为两类：一是检测虚假内容，二是确立内容来源。

检测虚假内容

科技公司通常使用深度学习、计算机视觉等方法来分析合成内容，寻找虚假或篡改痕迹，并利用机器学习模型来识别深度伪造内容中的模式和异常。[5]这些工具还能检测出音视频内容中的不一致之处，如与人类唇部细微动作或语音语调的不符之处。[6]

部分生成式人工智能工具包含检测某段内容是否由其协助制作的功能，但它们可能无法检测出由其他模型生成的深度伪造内容。[7]一些虚假内容检测工具会寻找生成式人工智能工具的篡改痕迹或“指纹”，[8]一些工具采用“白名单”和“黑名单 ”方法（即维护可信任来源和已知造假者的名单），而还有一些工具则寻找人类特征（而非伪造证据），如自然的血液流动、面部表情和语调变化。[9]

目前的深度伪造检测工具据称准确率超过90%。[10]然而令人担忧的是，不法分子可能正在利用开源的生成式人工智能模型来生成能够规避这些检测工具的媒体内容。例如，生成式人工智能工具的高效内容生成能力可能会让现有检测系统难以及时识别，此外，该等工具根据用户提示对输出进行的细微调整也可能被用来掩盖虚假内容。[11]

社交媒体平台本身也经常使用人工智能工具帮助检测图像或视频中的问题内容，并按相对程度对其进行评分，然后将最可疑的内容转交审核人员进行最终判定。但这种方法既耗时又昂贵，目前各大平台正利用机器学习加速这一流程。[12]

如果这听起来让人联想到网络安全领域的发展，那可能事实如此。正如具有安全意识的公司采用多层防护措施来保护数据和网络安全，德勤预计，新闻机构和社交媒体公司亦或需要多种工具以及内容来源验证方法，来帮助判断数字内容的真实性。

确立内容来源并构筑信任

部分公司正在探索另一类方法，即在媒体文件创建时添加加密元数据（或数字水印）。这些随附于媒体文件的数据，能够详细说明文件的来源并保留所有的修改记录。[13]

社交平台正与媒体机构、设备制造商及科技公司开展跨界合作，共同推动内容真实性标准的建立。包括德勤在内的多家科技和媒体公司已加入内容来源和真实性联盟（C2PA），并承诺实施C2PA元数据标准，以更便捷地验证人工智能生成的图像。[14]C2PA技术通过创建详尽的变更和修改日志，能够记录图片生命周期（从创建到编辑过程）的每个阶段。[15]凭借可查询的C2PA记录，内容发布机构和用户得以检验视觉素材的来源，并评估其可信度。

为进一步区分由真人运营的账号，一些社交媒体平台开始向创作者推出实名认证选项。这可能需要创作者提交身份证明材料，并支付一定认证费用。此外，平台还可能将实名认证作为参与某些收益分享计划的先决条件，以鼓励创作者完成认证。[16]

随着人工智能生成内容的普及，验证真人运营账号的真实性将有助于平台提升可信度和公信力。[17]平台可能需要考量，将认证成本转移至创作者、广告商或用户是否具有长期可行性。

有待立法出台

尽管部分政府已实施了内容真实性监管措施，[18]但构建更全面且全球统一的立法可能更具成效。此外，加强公共宣传教育也十分关键，可以帮助用户认识深度伪造技术的风险，并掌握辨别媒体内容真伪的方法。

美国已提出一项法案，要求人工智能生成的内容必须添加数字水印，目前该法案正在参议院商业、科学和交通委员会审议中。[19]加利福尼亚州正在审议AB-3211法案，该法案要求设备制造商更新固件，以便为照片附加来源元数据，并要求在线平台公开网络内容的来源元数据。如果获得通过，该法案将于2026年生效。[20]其他一些州也已通过类似立法，将未经同意制作和传播、旨在散布虚假信息的深度伪造行为定为犯罪。[21]美国联邦贸易委员会 (FTC) 正在制定新规，旨在禁止模仿个人的深度伪造内容的创建和传播。[22]

欧盟《人工智能法案》(AI Act) 的修订重点强调了透明度要求，规定必须对人工智能生成及深度伪造内容作出明确标识。此举旨在推进人工智能技术发展的同时，保障用户对接触内容性质的知情权。欧盟委员会设立了人工智能办公室，旨在促进人工智能的发展与应用，并倡导对人工生成或合成内容进行有效标识。[23]

深度伪造技术的迅猛发展要求监管框架兼具灵活性和适应性，能够随着技术发展不断演进。

小结

图像、视频或音频片段的真实性可以通过分析并验证其来源来确定。随着生成式人工智能不断用于创建各类合成内容，加之不法分子通过调整模型和输出以规避检测，媒体公司和社交网络很可能将加大对这两类方法的投入。

随着生成式人工智能变得日益强大且用途广泛，领先于不法分子以防止技术滥用变得尤为重要。利用更先进的技术，如血液体积检测和面部分析，能够有效鉴别内容的真假。然而，与网络安全工具一样，这些技术的运用应将对最终用户和消费者的干扰降至最低，即在确保内容完整性的同时，不影响用户体验。数字水印等技术可在无需牺牲内容质量和使用实时计算资源进行分析的情况下，帮助验证内容的真实性。[24]

对于使用训练有素的机器学习模型（或委托第三方）检测虚假内容的公司而言，采纳一项领先实践十分必要：优先选用拥有多元化、高质量图像及音视频数据集的工具和供应商。这些数据集应涵盖各类人口统计群体，以确保检测的公正性并最大限度地减少准确性偏差。[25]

科技公司与媒体公司应积极开展跨界合作，[26]共同制定并推广深度伪造检测和内容认证的标准。例如，当设备制造商与媒体机构对内容的创作与发布进行联名认证时，数字水印技术的作用将更加显著。此类合作能形成更完备且广受认可的行业实践，进而提升数字内容整体的安全性和可靠性。

在企业安全方面，各行业公司需警惕，生成式人工智能或提升社会工程攻击效率，并削弱部分身份验证机制。[27]因此，有必要增设额外验证层级，尤其是在以视频和音频为主的流程中。应鼓励最终用户向可靠信息源求证信息，并采用多因素身份验证，以降低深度伪造带来的风险。鉴于技术态势的持续演变，用户教育（如网络安全意识培训）亦成为公司不得不重视的关键措施。

这些策略不仅能防范深度伪造技术所带来的威胁，还有助于科技公司和媒体公司在维护数字内容完整性和可靠性方面建立领导地位。在这关键时刻，企业应着力构建高度可信的内容领域，并在不确定性日增的数字环境中，稳固自身作为可靠信息源的权威性。

By **Michael Steinhart**
United States

Bree Matheson
United States

Ankit Dhameja
India

Gillian Crossan
United States

尾注

1. Margaret Talev and Ryan Heath, "*Exclusive poll: AI is already great at faking video and audio, experts say*," Axios, accessed Oct. 28, 2023.

2. Susanne Hupfer, Michael Steinhart, et.al, "2024 Connected Consumer Survey," Deloitte, December 2024

3. Vivaan Jaikishan, Cameron D'Ambrosi, Jennie Berry, and Stacy Schulman, "*The rising threat of deepfakes: Detection, challenges, and market growth*," Liminal, May 7, 2024.

4. Ian Shepherd, "*Human vs. machine: Will AI replace content creators?*" *Forbes*, April 26, 2024.

5. Analytix Labs, "*Detecting deepfakes: Exploring advances in deep learning-based media authentication*," Medium, January 4, 2024.

6. For example, see: Intel, "*Trusted media: Real-time FakeCatcher for deepfake detection*," accessed Oct. 28, 2024.

7. Cade Metz and Tiffany Hsu, "*OpenAI releases deepfake detector to disinformation researchers*," *The New York Times*, May 2024.

8. Danial Samadi Vahdati, Tai D. Nguyen, Aref Azizpour, and Matthew C. Stamm, "*Beyond deepfake images: Detecting AI-generated videos*," Drexel University, accessed Oct. 28, 2024.

9. Alex McFarland, "*5 best deepfake detector tools & techniques* (October 2024)," Unite.AI, Oct. 1, 2024.

10. Konstantin Simonchik, "*Deepfake detection: Accuracy of commercial tools*," LinkedIn, February 2024

11. Jiansong Zhang, Kejiang Chen, Weixiang Li, Weiming Zhang, and Nenghai Yu, "*Steganography with generated images: Leveraging volatility to enhance security*," *IEEE Transactions on Dependable and Secure Computing* 21, no. 4 (2024): pp. 3994–4005; see also: Mike Bechtel and Bill Briggs, "*Defending reality: Truth in an age of synthetic media*," *Deloitte Insights*, Dec. 4, 2023; and, Loreben Tuquero, "*AI detection tools for audio deepfakes fall short. How 4 tools fare and what we can do instead*," Poynter, March 21, 2024.

12. Barbara Ortutay, "*Content moderation in the AI era: Humans are still needed across industries*," Fast Company, April 23, 2024; also see: Meta, "*How review teams work*," Jan. 19, 2022.

13. Glenn Chapman, "*Meta wants industry-wide labels for AI-made images*," *AFP News*, Feb. 6, 2024; also see: Nick Clegg, "*Labeling AI-generated images on Facebook, Instagram and Threads*," Feb. 6, 2024; Sasha Luccioni et al., "*AI watermarking 101: Tools and techniques*," Hugging Face, Feb. 26, 2024; and Partnership on AI, "*Building a glossary for synthetic media transparency methods, part 1: Indirect disclosure*," Dec. 19, 2023.

14. Ryan Heath, "*Inside the battle to label digital content as AI-generated media spreads*," Axios, accessed Oct. 28, 2024.

15. Demian Hess, "*Fighting deepfakes with content credentials and C2PA*," CMSWire, March 13, 2024.

16. Andrew Hutchinson, "*X will require ad revenue share participants to confirm their ID*," Social Media Today, May 22, 2024.

17. Guy Tytunovich, "*The future of trust and verification for social media platforms*," *Forbes*, May 22, 2024.

18. Amanda Lawson, "*A look at global deepfake regulation approaches*," Responsible Artificial Intelligence Institute, April 24, 2023.

19. US Congress, "*S.2765—Advisory for AI-Generated Content Act*," Sept. 12, 2023.

20. California Legislative Information, "*Assembly Bill 3211—California Digital Content Provenance Standards*," Aug. 24, 2024.

21. Kevin Collier, "*States are rapidly adopting laws regulating political deepfakes*," *NBC News*, Aug. 7, 2024.

22. Federal Trade Commission, "*FTC proposes new protections to combat AI impersonation of individuals*," Feb. 15, 2024; also see: Michelle M. Graham, "*Deepfakes: Federal and state regulation aims to curb a growing threat*," Thompson Reuters, June 26, 2024.

23. Melissa Heikkilä, "*Five things you need to know about the EU's new AI Act*," *MIT Technology Review*, Dec. 11, 2023.

24. Deloitte, "*How to safeguard against the menace of deepfake technology*," accessed Oct. 28, 2024.

25. AI Index Steering Committee, "*The AI Index 2024 Annual Report*," accessed Oct. 28, 2024.

26. AI Election Accord, "*A tech accord to combat deceptive use of AI in 2024 elections*," accessed Oct. 28, 2024.

27. Stu Sjouwerman, "*The growing threat of AI in social engineering: How business can mitigate risks*," Fast Company, April 8, 2024.

致谢

The authors would like to thank **Je Loucks, Susanne Hupfer, Duncan Stewart, Je Stoudt, Jason Williamson, Tim Davis, Gopal Srinivasan, Shreeparna Sarkar,** and **Andy Bayiates** for their contributions to this article.

Cover image by: **Jaime Austin; Getty Images, Adobe Stock**

Deloitte.
Insights

重新评估直接面向消费者（DTC）模式：转向视频聚合商

视频内容创作者或需更多经销商来扩大可触达市场规模

德勤预测，订阅视频点播（SVOD）的“堆叠”现象——即消费者订阅多个独立视频点播服务——将在2025年减少。各市场“堆叠”用户的平均订阅数量或将达到峰值，各市场有所不同，在美国，每个用户的订阅数约为4个，而在欧洲市场略高于该数字的一半[1]。这意味着，每个市场的独立订阅总量可能会下降，即使由于提价、打击密码共享和捆绑销售，SVOD的收入仍可能增加。

上述峰值是对视频行业市场生存能力重新评估的预期结果，该行业市场主要由数十家直接面向消费者（DTC）的视频订阅服务提供商组成[2]，每个家庭都会购买多种订阅服务，而非单一的付费电视订阅服务。然而，该行业的发展方向或将回到不同服务提供商的内容聚合。这一付费电视提供商的传统方式已然过时。

从中期来看，我们预计视频行业最终可能会由各国市场少数几家（多为两家或三家）独立SVOD服务商和聚合商组成。聚合公司预计包括：传统付费电视公司、电信公司、技术平台或最大的SVOD服务商。在英国，截至2024年9月，受访的SVOD用户中有43%通过另一方（付费电视、电信公司或技术平台）购买了至少一项服务。在受访的小型SVOD服务提供商中，近半数的订阅用户通过聚合商购买；在大型提供商中，约四分之一的订阅用户间接购买[3]。

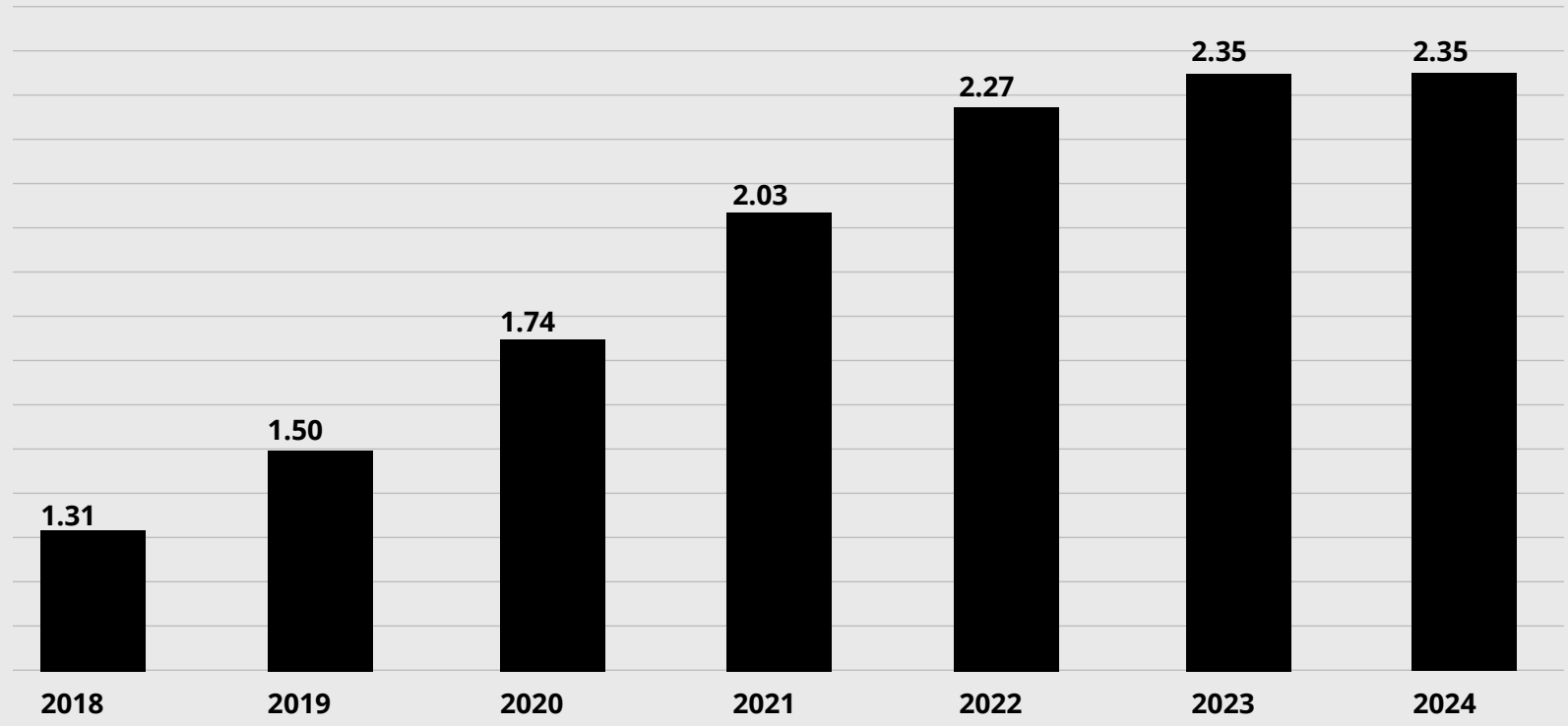
根据我们与业内人士的交流，德勤预计每家聚合商都将提供一种付费电视模式，包括以下部分内容（有时是全部内容）：单一账户和账单、标准和可选内容频道、12个月（或更长）合同、显示所有可看内容的电子节目指南、广告销售和播放以及集中营销。2025年，有望加速回归聚合模式，但可能几年后才能完成。

行业正步入SVOD堆叠高峰期

2010年代有两大应用趋势：使用SVOD服务的家庭数量持续增长；使用的服务数量稳步增长。德勤研究显示，欧洲市场的平均订阅数从2018年的1.3个稳步上升，到2023年和2024年稳定在2.35个（见图10，资料来源在图下方）。据德勤《2024年数字媒体趋势》报告，自2020年以来，美国SVOD服务的平均订阅数一直稳定在4个[4]。

图1

欧洲多个市场SVOD平均订阅数



资料来源：德勤数字消费者趋势，2018-2024年。该平均值所代表的国家包括：英国、丹麦、爱尔兰、挪威、瑞典、德国、比利时、意大利、荷兰。这些市场共代表2.65亿人口。

付费电视模式复苏的原因

独立SVOD的基本优点是消费者可以掌控内容选择和合同期限，且内容提供商能够绕过经销商，鉴于此，重新采用数量更少、规模更大、价格更高、期限更长的捆绑模式看似是一种倒退。

但是，回归聚合模式或许能在消费者及供应商的需求之间取得平衡。

对于消费者而言，理想化的独立SVOD可能包括多种相对低价的视频服务，每种服务都可通过直观的应用程序随时访问，且每种服务都可提前几周通知取消，但遗憾的是，这可能不具备商业可行性。近期面临的现实情况是，订阅价格不断攀升，密码共享遭受打击[5]，内容选择铺天盖地[6]，用户界面流畅度参差不齐。[7]在制作内容并按多年期合同销售给经销商方面，内容提供商拥有数十年的盈利经验，但他们可能很难轻松地转向运营端到端DTC业务的各个方面，例如从设置信用卡支付到管理建立广告视频点播（AVOD）层级所需的合规性[8]。

部分订阅用户可能会重新选择付费电视模式。

实现重新聚合的途径

有多种途径可以实现视频平台和内容的重新聚合，并最终形成新的市场结构。

捆绑服务：此类服务将SVOD订阅与付费电视、电信或金融服务合同打包，提供比单独购买每项服务更优惠的价格。作为交换，用户需承诺最短合同期限，通常至少一年。SVOD公司加入捆绑服务的一个主要原因就是为降低一直居高不下的客户流失率：在美国市场约为40%，英国市场约为20%。[9]据行业分析估计，2023年仅美国就有1.393亿次退订。[10]在取消服务的用户中，约四分之一为“频繁退订者”，在过去两年中至少取消了三到四次服务，而2019年这一比例仅为3%。[11]五分之一的频繁退订者在24个月内取消过七次或更多次。[12]

例如，将SVOD与18个月的传统付费电视合同进行绑定（可换取总价的折扣），可以延迟用户可能发去的取消行为，减少周期性客户流失。超过一半的美国消费者愿意为了折扣而选择一年的订阅[13]，但截至2024年初，美国仅4%的SVOD合同期限为12个月。[14]对于目前经营独立VOD的公司来说，加入第三方捆绑服务也是为了外包客户获取、计费（和坏账管理）、客户支持和广告销售等一系列职责。

截至2024年，捆绑服务已蔚然成风，到2025年及以后，可能会进一步普及。付费电视公司捆绑SVOD产品，在不同市场会收获不同的效益，但共同的出发点是将热门内容套餐固定嵌入其产品组合，以帮助降低自身的客户流失率。对于那些核心服务增长不够强劲的公司来说，增加SVOD还能提高总收入。[15]

在英国，截至2019年，Sky所有付费电视套餐均默认包含Neftlix的广告套餐[16]，所有订阅内容均可使用统一搜索栏。在法国，付费电视频道Canal+为所有用户提供Disney+和Paramount+，还有各种提供多个SVOD品牌的套餐。[17]在中欧，据估计，25%的SVOD通过付费电视或电信运营商间接订阅。[18]在美国，Xfinity宽带用户可以添加一项名为Streamsaver的捆绑服务，其中包含Apple TV+、带广告的Netflix Standard和带广告的Peacock Premium，可节省30%的费用。[19]预计到2028年，全球25%的在线视频订阅将通过电信公司完成，而2023年这一比例仅为20%。[20]

对于电信公司而言，以优惠价格引入热门的SVOD服务也有助于提高用户留存率，尤其是在用户认为运营商彼此网络性能差异不大的市场。[21]

一些银行也将SVOD纳入其订阅服务。截至2024年8月，英国巴克莱银行向Bank Account+ Blue Rewards客户免费提供Apple TV+服务。[22]

部分小型SVOD服务正从独立的DTC转向由聚合商分发的附加频道，或完全退出某些市场 。

媒体聚合：服务通常通过捆绑销售进行聚合，其中最典型的做法是以相对于单独购买更优惠的价格来出售原先各自独立的多项服务。例如，在美国市场，Disney+、Max和Hulu捆绑服务的折扣高达38%。[23]此类捆绑服务可能是纯视频，也可能是视频加其他媒体（音乐、游戏或新闻）。随着捆绑服务日渐流行，单项服务可能会被取消，或以高定价劝退单独购买。

视频服务聚合应提高易用性，例如，通过提供统一的搜索栏及电子节目指南，根据观众的订阅情况，为其提供所有可看内容。相比之下，如果使用独立服务，用户可能会在两项服务切换时遇到操作上的不顺畅，尤其是当使用处理器性能较差的旧款电视或经济型电视进行观看。德勤的研究发现，在受访的美国消费者中，几乎一半表示如果流媒体平台的内容检索更加便捷，他们会在该流媒体服务上花费更多时间。[24]在受访的美国Z世代和千禧一代消费者中，约四分之三希望在可访问的所有服务平台之间进行无缝搜索。[25]德勤英国的研究表明，取消订阅的主要原因之一是找不到心仪的观看内容，但其实当前节目资源空前丰富，这两者实在矛盾。[26]

捆绑SVOD服务有助于减少用户流失。根据行业分析，订购捆绑服务（Disney+、Hulu和ESPN+）的用户比单独订购Disney+的用户其流失可能性低59%[27]。德勤的研究表明，消费者对进一步涨价的承受程度可能已至临界点：在2023年第四季度的调查中，近一半的美国消费者表示，如果SVOD服务每月价格上涨5美元，他们将取消该服务。[28]

永久流失：对于一些家庭而言，其他VOD服务可能会取代付费SVOD，例如在欧洲盛行的免费广播VOD（BVOD），或YouTube等视频共享服务（绝大部分免费）。SVOD用户流失的一大关键因素是成本。近年来，订阅成本已越来越成为影响用户决策的重要因素，在取消SVOD服务的英国受访者中，24%的人选择“订阅太贵”作为取消原因。到2024年，这一比例上升到31%[29]。

过去只提供订阅服务的公司也许将提供更多FAST（免费广告支持型电视）服务。例如，2008年开始提供SVOD的亚马逊于2022年在美国推出了FAST频道Freevee。[30]Crunchyroll自2010年代中期以来主推动画SVOD，[31]于2023年推出了FAST服务。[32]此外，目前对用户上传且通常免费的服务平台（如YouTube）的使用量也有所增加。2024年至2029年期间，通过电视机观看YouTube的时间预计将大幅增加90%，每日观看时长从12分钟增至22分钟。[33]其中部分增长可能来自曾经的SVOD用户。

小结：电视业务再度演变

从广播到基于网络的传播，整个电视行业正在经历长达数十年的转变。

回顾过去，独立SVOD市场的兴起和发展是一个显著的里程碑。在2010年代，SVOD产品新颖，竞争对手较少，订阅价格亲民且通常可与家人朋友共享使用，因此SVOD市场实现了飞速增长。

但到了2020年代，情况不容乐观——用户流失率居高不下，增长难度超出预期。在英国这一相对成熟的市场，近来SVOD的收视份额增长乏力，从2023年的15.8%微升至2024年的16.4%，而广播公司电视旗下内容和视频共享网站的份额显然更胜一筹。[34]独立的SVOD市场难以占据主导地位，类似传统付费电视的聚合服务模式有望在未来几年内蓬勃发展[35]。

独立的SVOD播放平台应思考自身在这种新模式下所扮演的角色。少数几家播放平台已经具备了足够的规模和能力（从计费到用户界面设计再到视频压缩），因此它们有可能继续致力于提供全方位的SVOD服务（即由一家公司提供从内容制作到广告销售再到客户管理的全面整合服务）——但在第三方分销更具商业意义的市场中，情况或许有所不同。

一些SVOD播放平台除了自有内容外，可能还想率先聚合其他播放平台的内容资源：它们可以打造以自有内容为核心的捆绑产品，还可以像传统付费电视一样营销和托管其他公司的频道。然而，许多播放平台还是决定专注于一个直接目标：将内容卖给出价最高的买家——这是过去一直支持其成功运营的策略[36]。

本节的要点在于，纯粹的DTC模式往往充满挑战，特别是对于那些可能需要增强技术能力和调整企业文化的公司。几十年来长期专注于内容创作，将分销交给第三方完成的公司，可能无法即刻在DTC领域取得成功。将DTC作为一种独立商业模式的难点，实际上是一种普遍限制，它困扰着大多数行业，并非视频行业独有。能够全面转型到DTC模式的大规模公司寥寥无几。分销商在大多数大型消费品牌的供应链中占据着核心地位，这是近乎不变的信条。

广播电视公司在众多市场中把控着绝大部分观众的收视市场，并且独立的SVOD未必会成为行业主流模式，但这并不意味广播电视公司可以高枕无忧。在欧洲 (42个市场) ，2023年传统电视的平均收视时长为每天3小时16分钟；但年轻观众群体仅贡献1小时12分钟[37]。广播公司应共同努力，各广播电视公司应携手打造优质内容平台，确保其节目和服务对年轻受众具有持久且不断增长的吸引力。

By **Paul Lee**
United Kingdom

Eliza Pearce
United Kingdom

Rupert Darbyshire
United Kingdom

Kevin Westcott
United States

尾注

1. Kevin Westcott, Jana Arbanas, Chris Arkenberg, and Jeff Loucks, "*Streaming video at a crossroads: Redesign yesterday's models or reinvent for tomorrow?*" *Deloitte Insights*, March 20, 2024; Deloitte, "*Generative AI: 7 million workers and counting*," June 25, 2024.

2. Ampere Analysis, "*Analytics - SVoD*," accessed November 2024.

3. This is based on a nationally representative survey of 4,000 respondents ages 16 to 75, commissioned by Deloitte UK and undertaken in August 2024.

4. Westcott, Arbanas, Arkenberg, and Loucks, "*Streaming video at a crossroads.*"

5. David Pierce, "*Streaming services keep getting more expensive: all the latest price increases*," The Verge, Sept. 23, 2024; Emma Roth, "*Disney's password-sharing crackdown starts 'in earnest' this September*," The Verge, Aug. 7, 2024.

6. As of June 2023, Nielsen's Gracenote reported that the number of individual titles available on streaming services was 2,346,171, up from 1,882,401 in July 2021. These numbers span the range of content available in the United States, the United Kingdom, Canada, Mexico, and Germany; see: Nielsen, "*Data-driven personalization: The future of streaming content discovery*," accessed November 2024.

7. Selome Hailu and Jennifer Maas, "*From 'glitchy' HBO Max to 'overwhelming' Amazon Prime Video, Hollywood insiders spill on their (least) favorite streaming interfaces*," Variety, April 11, 2023.

8. Andrew Blustein, "*How GDPR, ad fatigue and content costs are complicating the now-global streaming wars*," The Drum, Sept. 4, 2019.

9. Westcott, Arbanas, Arkenberg, and Loucks, "*Streaming video at a crossroads*"; Deloitte, "*Generative AI*."

10. Antenna, "*The rise of the show chaser*," accessed November 2024.

11. Ibid.

12. Ibid.

13. Westcott, Arbanas, Arkenberg, and Loucks, "*Streaming video at a crossroads.*"

14. Antenna, "*Understanding the relationship between annual plans and promotions*," accessed November 2024.

15. IDC Research, "*IDC forecasts slower growth for global telecommunications market: Could AI help telcos to maintain healthy margins*," May 3, 2024.

16. Sky, "*Official website*," accessed November 2024.

17. Canal Plus, "*S'abonner à Disney+ avec les offres CANAL+*," accessed November 2024; Georg Szalai, "*Paramount+, Canal+ expand partnership in France*," The Hollywood Reporter, Aug. 21, 2024.

18. Omdia, "*Omdia unveils surging SVOD growth in CEE through strategic pay TV and telco partnerships*," June 12, 2024.

19. Xfinity, "*Streaming services: Stream one, stream all*," accessed November 2024.

20. Bango, "*Super bundling: What telco leadership needs to know about securing a wider role in the subscriptions market*," accessed November 2024.

21. Barclays, "*Current accounts: Barclays Bank Account + Blue Rewards*," accessed November 2024.

22. Viaplay Group, "*Q2 2024 interim report January-June*," press release, July 18, 2024.

23. Disney Plus, "*New Disney+, Hulu, Max bundle is now available in ad-supported and ad-free plans*," July 25, 2024.

24. Westcott, Arbanas, Arkenberg, and Loucks, "*Streaming video at a crossroads.*"

25. Ibid.

26. Deloitte, "*Generative AI.*"

27. Ampere Analysis, "*Subscribe, cancel, repeat: 42% of US consumers are SVoD 'resubscribers'*," July 8, 2024.

28. Westcott, Arbanas, Arkenberg, and Loucks, "*Streaming video at a crossroads.*"

29. Deloitte, "*Generative AI.*"

30. Christian de Looper, "*Amazon Freevee: Everything you need to know about the free streaming service*," Amazon UK, May 10, 2023.

31. Janko Roettgers, "*Chernin, AT&T Set Brand for New Online Video Venture: Ellation (Exclusive),*" Variety, Aug. 3, 2015.

32. Crunchyroll News, "*Crunchyroll launches 24/7 anime channel in the US*," Oct. 11, 2023.

33. Enders Analysis, *https://www.endersanalysis.com/reports/video-viewing-forecasts-slowdown-change*

34. Share of viewing is calculated based on 12-month rolling averages from October 2022 to September 2023 through August 2023 to July 2024. All data is from Barb's monthly viewing summary; see: Barb, "*Monthly viewing summary*," accessed November 2024. Barb's methodology for capturing viewing patterns is explained in: Barb, "*What is the Barb panel and why is it important?*" accessed November 2024.

35. Richard Waters, *The next phase of the streaming wars*, Financial Times, June 6, 2024 (subscription required)

36. Etan Vlessing, "*Sony CFO: Without a streaming platform, we're free to sell films and shows 'to the highest bidder',*" The Hollywood Reporter, March 6, 2023; Diane Haithman, "*Why Sony's streaming deals with Netflix and Disney make sense for everyone*," The Wrap, May 10, 2021.

37. *Audience Trends Television 2024*, European Broadcasting Union Media Intelligence Service, August 2024 (registration required)

致谢

The authors would like to thank **Stacy Hodgins, Beth Rae Rosenstein, Helen Rees, Ben Stanton,** and **Duncan Stewart** for their contributions to this article.

Cover image by: **Jaime Austin; Getty Images, Adobe Stock**

Deloitte.
Insights

生成式人工智能用于内容创作，大型制片公司犹豫不决，社交媒体积极推进

好莱坞（以及其他电影制片公司）可能会谨慎采用生成式人工智能技术进行内容创作，但或会率先将其用于运营和发行。

用于生成图像、音频和视频的生成式人工智能模型在不断进步，可制作出更逼真、更富有创意和更长期可控的内容。制片公司可能早就开始尝试使用生成式人工智能进行内容创作，但要将其投入全面内容制作，他们表现得十分犹疑，因为这一工具目前尚不成熟，以及公共模型下的内容创作可能引发法律责任和知识产权保护问题。**但越来越多制片公司相信**，这一技术若应用于各项业务流程，则**能够帮助降低成本**，**提高盈利能力**。

事实上，各大制片公司都面临着成本压力，只有极少数实现盈利。[1]他们的营收数字虽然可观，但运营费用以及制作、营销和广告的成本也在不断攀升，超过了营收增长速度。[2]而对于许多流媒体制片公司来说，情况往往是其投入了资金发展流媒体服务但却无法盈利，同时反而因有线电视订阅量和广告收入下降而营收大减。如今，电影制片公司一方面要面临因通货膨胀、银行加息以及新冠疫情影响而抬升的成本，另一方面还要与社交媒体、用户生成内容（UGG）和视频游戏争夺市场流量和收入。

德勤预测，到2025年，各影视巨头公司（尤其是美国和欧盟的制片公司）仍将对采用生成式人工智能进行内容创作持谨慎态度，在此方面只投入不到3%的制作预算，[3]但会拨出约7%的运营支出，用于使用生成式人工智能工具执行合同及人才管理、许可与规划、营销和广告以及内容本地化和配音等工作，以此帮助扩大制片公司在全球各个市场的影响力。

如此一来，制片公司既能在人才和内容创作方面少受这一新兴技术干扰，又能降低成本并提高业务效率。然而，独立内容创作者和社交媒体平台正在积极拥抱生成式人工智能技术，迅速将其应用至工作流和内容创作中。这有可能催生出新的媒体形式，从而使与其争夺流量的传统制片公司更加处于下风。[4]

生成式人工智能工具尚不能支持好莱坞级别的内容创作

低成本且方便易用的大型语言模型 (LLM) 和扩散模型 (Diffusion Model) 的出现，使得电影制片公司能够对剧本、对话和故事元素进行快速原型设计，并对角色和场景设计进行早期可视化和发掘。[5]一些电影制片公司使用生成式工具对演员进行去老化处理或生成数字双胞胎，以用于商业广告或其离世后的作品制作。[6]在这种情况下，制片公司可直接在演员合约中添加保护条款，以控制潜在的法律责任风险。未来一年或会出现更多的第三方制作机构向制片公司出售相关服务和工具，以满足他们对于先进技术能力的需求。

生成式人工智能能够在内容创作的前期阶段激发新颖的创意，但其水平还不足以制作出好莱坞级别的作品。[7]目前最强大的视觉扩散模型能够生成栩栩如生的图像，但其输出结果反而因为过于逼真而显得“不真实”。[8]技术领先的视频模型能够生成短片，但无法制作篇幅更长、更连贯的故事。[9]视频生成模型发展迅速，却还不够成熟，要想集成到现有工具和制作流水线中仍需时日。

不过，对于通常需要快速产出内容以及高频率发布更新的社交媒体创作者而言，这些限制可能并无大碍。尽管流行趋势一直在变，但当前快节奏的剪辑手法广受青睐。[10]社交媒体视频的时长通常较短，因此可能无需太担心相关责任风险问题。有部分内容创作者很早就尝试采用生成式模型和工具制作内容，并定期在社交媒体上发布，展现了借助第三方解决方案而得以快速进步的视频模型的无限潜力。[11]

预计在未来一年，独立创作者将会使用生成式人工智能引领内容创作潮流。制片公司可在静观技术能力演化的同时，巧妙规避相关风险。但这也有可能让用户生成内容平台率先抢占市场流量，使传统媒体在与其的激烈竞争中失去先机。

大型制片公司顾虑公有模型带来的法律责任风险

大型制片公司还担心采用生成式人工智能内容工具可能会引致知识产权及相关法律责任风险，或是利用该工具生成的内容无法被认定为原创作品享受版权保护。[12]部分功能强大的公有模型是基于公共数据（如来自其他创作者的图像和视频）训练而成，这使得这类模型的输出结果从根本上成为衍生内容。[13]如有制片公司将此类公有模型的输出内容作为商用，而该模型的训练数据包含了他人的版权作品，那么该公司则有可能承担侵权责任。模型训练集可能收录了数十亿部作品，使得追查侵权行为极为困难，如同大海捞针。但**这种不确定性或就足以吓退以创建、保护和捍卫自己的知识产权为主营业务的制片公司**。目前已有公有模型因涉嫌侵权遭到独立艺术家和创作者的起诉[14]，此外出版商[15]和音乐厂牌[16]也纷纷提起类似诉讼。

公有模型也可能使电影制片公司难以或无法维护自己的知识产权。美国《版权法》(Copyright Act) 规定，作品必须包含足够的“人类创作内容”（即“充分性要求”）才可获得版权保护。[17]在近期的审议案例中，美国版权局指出，在包含人工智能生成内容的作品中，人类创作内容的多少将视具体情况而定，只要符合“充分性要求”，即可获得版权保护。换言之，电影制片公司采用生成式人工智能辅助创作的作品也可获得版权，但作品中的人工智能生成内容不得超过一定比例，不得以生成模型为主要创作手段。随着该领域的不断发展，相关争议和讨论将持续存在，但由于缺乏准确的定义，相关不确定性和风险将只增不减。

由于迫切需要更多的数据来丰富其训练集，领先的生成式人工智能供应商正积极争取电影制片公司的支持，鼓励其授权内容库。[18]但制片公司要么出于保护知识产权的考虑拒绝授权，要么可能会向生成式人工智能公司收取高昂授权费用，这将增加后者本已沉重的运营成本。制片公司甚至可能集体拒绝提供训练集数据，希望以此抑制前沿模型，即正被编码和训练以用于生成文本、音频、图像和视频的先进算法的发展。

此外，制片公司必须与行会和工会合作，尤其是美国的制片公司更要遵循行会和工会的要求，而这些组织对于采用生成式人工智能表现出强烈抵制，并要求制片公司严格限制其使用。[19]英国[20]和欧盟也出现了类似的发展阻力，该地区的制片公司也需遵守欧盟《人工智能法案》(Artificial Intelligence Act) 关于人工智能模型安全管理的规定，以及《通用数据保护条例》(GDPR) 关于训练数据收集和存储的相关规定。[21]

完全的私有模型对于制片公司来说又可能过于昂贵

在德勤2024年的相关预测文章中，我们讨论了私有生成式人工智能模型的兴起可以帮助避免公有模型的应用挑战，并让模型的输出内容更加可控。[22]制片公司可以利用自有知识产权训练自己的私有模型，从而避免公有模型带来的法律责任和版权风险。

然而训练生成模型的成本非常昂贵，据估计，训练一个前沿模型大约需耗费1,000亿美元[23]，此外的推理和再训练成本还会随着模型的使用而增加。尽管开源解决方案（更准确地说，是“开放权重”模型）可能会降低部分成本，但其训练集不透明，且成本依然不低。[24]此外，制片公司可能难以吸引到有能力构建此类模型的高端技术人才，这类人才往往更倾向于加入能够提供高薪的超大规模云服务公司。再者，由于模型的快速开发节奏，当前投资开发的模型可能在半年内就需要更新升级。而要建设更高效的私有模型，制片公司和投资者或须在思维和行为上都向科技公司看齐，与技术供应商建立并维护生态合作关系，向其支付相关费用。出于如上种种原因，除非拥有雄厚的经济实力，制片公司不太可能会训练自己的私有模型。

但在未来一年，制片公司与智能服务供应商可能会开展一系列合作，以更加合理地分摊成本。[25]在这种合作模式下，由第三方提供预先训练好的模型和界面，再使用制片公司的内容进行进一步训练和定制。例如，经过进一步训练的定制模型可以产出符合制片公司美学风格或包含其标志性角色和场景片段的内容。此外，制片公司利用自身原创内容生成的衍生作品能够控制潜在的知识产权风险。而通过这种合作，智能服务供应商也将能够继续开发更加成熟的工具和生成内容。

使用生成式人工智能工具帮助优化制片公司业务

预计在未来一年，电影制片公司将尝试使用生成式人工智能技术进行内容创作，但其大概率会更先利用该项技术来支持和优化其他业务。生成式人工智能或可帮助实现多项业务工作的自动化和增强，例如合同谈判、人才和劳动力管理、财务与会计，以及本地化、营销和推广、存储和发行等媒体业务。

制片公司可能会通过其现有软件和SaaS解决方案来尝试部分生成式人工智能功能。越来越多的小制片公司也在利用这一技术来处理前期制作过程中的耗时和高成本工作。**生成式人工智能可以加快剧本评析**，拆分剧本并将各剧本元素**分配到对应拍摄和制作计划**，**甚至可以“侦察”剧本拍摄可能需要的场景地点**。[26] 生成式人工智能还可帮助释放内容库的价值，例如，通过“观看”老电影，标记演员、主题和氛围基调以对内容库进行多种分类，从而便于流媒体平台对旧内容进行动态重放和变现，满足更多个性化推荐或热点追踪需求。[27]

为帮助加快和扩大内容传播，一些制片公司目前在利用语言和语音模型来增强译制工作，使作品触达更多全球观众。[28]用户只需微调，这些模型工具便可提供极具表现力和感染力的高保真度配音。[29]对于面向全球市场的内容创作者和发行商而言，无论其是向全球市场推出内容，还是从全球市场引进内容，这都是一大福音。领先的用户生成内容创作平台也拓展了此类功能以供其创作用户使用。[30]

生成式人工智能配音和翻译还有助于实现更广泛的文化传播交流，让有些只能在本地流行的内容成为全球大热作品。德勤《2024年数字媒体趋势调研报告》显示，66%的受访美国人喜欢观看有助于了解不同于本国文化的电视节目或电影。[31]生成式人工智能不仅能帮助媒体公司提高营收和效率，还可促进全球观众间的文化交流和理解。

小结

与大多数公司一样，电影制片公司、流媒体平台和创意人才对生成式人工智能的能力既着迷又担忧。对于计划在未来一年探索生成式人工智能内容创作的制片公司来说，看重的正是该技术超乎寻常的创造力，相信前沿模型能融合人类创造力进而生成全新的创意内容。此外，好莱坞之外新媒体形式的兴起，也是推动制片公司探索生成式人工智能的重要动力。

越来越多的独立内容创作者正利用最新的合成媒体技术快速进入市场。好莱坞电影公司曾一度把控着内容和发行渠道的稀缺资源，但如今这些资源已变得更加丰富和大众化。[32]迫近的内容颠覆趋势或将在未来一年愈演愈烈。

前沿模型几乎每月都在更新迭代，其能力也在快速发展中逐渐接近人类的智力、创造力和洞察力。一年前就有人预言，到2030年，一部大片几乎可完全由人工智能制作完成，[33]而到2025年，这一观点似乎将更具说服力。

与此同时，内容所有者也将竭力维护其知识产权的竞争优势，或会对公有模型提起更多侵权诉讼，并加强监管力度。监管机构也会要求领先模型供应商证明其训练集没有对现有内容造成侵权。如有足够的经济实力，大多数大型电影制片公司可能会拒绝人工智能模型供应商递出的橄榄枝，不会将其内容库授权用于公有模型训练，而更愿意与小公司合作，打造围绕自有知识产权的更加定制化和受保护的私有模型。

从宏观层面来看，生成式人工智能需要巨额投资，若不能在未来一年内显现出广泛的经济价值，其增长速度或将放缓。[34][欲了解更多相关动态信息和分析，请参阅德勤今年关于端侧生成式人工智能的预测报告]但若新一代前沿模型能够克服现有挑战，其能力有望快速提升。预计行业也将努力降低模型的训练和运行成本，并减少训练所需的数据量。

大型电影制片公司作为大型企业，可能会更多地利用生成式人工智能来帮助削减成本、优化业务、提高生产力以及加速扩大客户基础。德勤《2024年企业生成式人工智能应用现状》的调查显示，42%的受访高管表示，效率、生产力和成本降低是其使用人工智能获得的最重要收益；58%的受访高管表示，他们也从中获得了额外的收益，如促进创新、改进产品和服务以及增强客户关系。[35]现代企业似乎越来越乐于接受人工智能工具。

俗话说，“水涨众船高”。生成式人工智能工具有或能帮助更多小公司和创作者达到以往只有大型公司才能达到的生产力和质量水平。小型制片公司和独立创作者迎来大展身手的机会，同时还能规避大型制片公司的风险和成本开销。而大型制片公司不仅要与其他制片公司竞争，还要与用户生成内容平台、社交媒体和游戏争夺市场份额，想要在如此激烈的竞争中突出重围，则可能需要进一步降低成本，加快产品上市速度。制作和发行资源可能不再稀缺，但市场流量仍是有限资源。

By

Chris Arkenberg
United States

Danny Ledger
United States

Ricky Franks
United States

Kevin Westcott
United States

尾注

1. George Szalai, "*Studio profit report: A year of major transition*," *The Hollywood Reporter*, April 24, 2024.

2. George Szalai, "*Studio profit report: Disney dives as Sony soars, Paramount rises*," *The Hollywood Reporter*, Feb. 24, 2024.

3. This prediction is based on our analysis of earnings reports from leading streaming video providers and other available industry information.

4. Chris Arkenberg, "*Will generative AI challenge authenticity in social media?*," *Deloitte Insights*, Dec. 8, 2023.

5. Hannah Murphy, "*Media groups look to AI tools to cut costs and complement storytelling,*" *Financial Times*, March 26, 2024.

6. David Smith, "*'We're going through a big revolution': How AI is de-ageing stars on screen*," *The Guardian*, Feb. 6, 2023.

7. Ibid; Murphy, "*Media groups look to AI tools to cut costs and complement storytelling*."

8. Alon Yaar, "*What's next for AI video generation*," AI Business, Aug. 6, 2024.

9. Lauren Leffer, "*Everything to know about OpenAI's new text-to-video generator, Sora*," *Scientific American*, March 4, 2024.

10. Taylor Lorenz, "*The 'Beastification of YouTube' may be coming to an end*," *The Washington Post*, March 30, 2024,

11. Dennis Ortiz and Kenny Gold, "*Gen AI and the creator economy: How creators are looking to leverage AI and what this means for brands*," Deloitte, accessed Oct. 30, 2024.

12. Paul Sweeting, "*Hollywood's AI concerns present new and complex challenges for legal eagles to untangle*," *Variety*, April 17, 2024.

13. Jennifer Wolfe, "*What would have to happen for gen AI to take over Hollywood? So glad you asked.*," NAB Amplify, July 5, 2024.

14. Ibid; Sweeting, "*Hollywood's AI Concerns Present New and Complex Challenges for Legal Eagles to Untangle*."

15. Baker & Hostetler, "*Case tracker: Artificial intelligence, copyrights and class actions*," accessed Oct. 30, 2024.

16. Natalie Sherman, "*World's biggest music labels sue over AI copyright*," *BBC News*, June 25, 2024.

17. US Copyright Office, "*Copyright and artificial intelligence*," March 16, 2023.

18. Lucas Shaw, "*Alphabet, Meta offer millions to partner With Hollywood on AI*," Bloomberg, May 23, 2024.

19. Erin Degregorio, "*Hollywood is back to work after strikes, but AI remains in the spotlight*," *Fordham Law News*, Jan. 29, 2024.

20. Daniel Thomas and Cristina Criddle, "*UK shelves proposed AI copyright code in blow to creative industries*," *Financial Times*, Feb. 4, 2024.

21. Lindsey Wilkinson, "*EU passes AI Act, places first binding rules on generative AI*," CIO Dive, March 13, 2024.

22. Chris Arkenberg, Baris Sarer, Gillian Crossan, and Rohan Gupta, "*Taking control: Generative AI trains on private, enterprise data*," *Deloitte Insights*, Nov. 29, 2023.

23. Jowi Morales, "*AI models that cost $1 billion to train are underway, $100 billion models coming — largest current models take 'only' $100 million to train: Anthropic CEO*," Tom's Hardware, July 7, 2024.

24. Red Hat, "*What is an open-source LLM?*," July 1, 2024.

25. Kyle Wiggers, "*Generative AI startup Runway inks deal with a major Hollywood studio*," *TechCrunch*, Sept. 18, 2024.

26. Lauren Forrestal, "*Filmustage leverages AI to break down film scripts, create shooting schedules and more*," *TechCrunch*, March 20, 2023; Lauren Forrestal, "*Avail rolls out its AI summarization tool to help Hollywood execs keep up with script coverage*," *TechCrunch*, Dec. 7, 2023.

27. Emma Cosgrove, "*Nvidia, Amazon, Microsoft, and Paramount execs discuss the use of AI in Hollywood. Here are 9 startups they're watching.*," *Business Insider*, July 24, 2024.

28. *The Economist*, "*The dawn of the omnistar*," Nov. 9, 2023.

29. Audrey Shomer, "*The state of generative AI in Hollywood: A special report*," *Variety*, June 3, 2024.

30. Andrew Hutchinson, "*YouTube announces expansion of auto-dubbing to more creators and languages*," SocialMediaToday, Sept. 19, 2024; Julia Walker, "*How Meta's AI dubbing breaks down language barriers*," PR Week, Sept. 30, 2024.

31. Jana Arbanas, Jeff Loucks, Brooke Auxier, Kevin Westcott, Chris Arkenberg, and Bree Matheson, *2024 Digital Media Trends*, *Deloitte Insights*, March 20, 2024.

32. Michael D. Smith, "*Lessons from Hollywood's digital transformation*," *Harvard Business Review*, Dec. 16, 2021.

33. Jackie Wiles, "*Beyond ChatGPT: The future of generative AI for enterprises*," Gartner, Jan. 26, 2023.

34. David Cahn, "*AI's $600 billion question*," Sequoia, June 20, 2024.

35. Deloitte, *The State of Generative AI in the Enterprise*—Moving from potential to performance, June 2024.

致谢

Authors would like to thank **Howie Stein** and **Ankit Dhameja**.

Cover image by: **Jaime Austin**; Getty Images, Adobe Stock

Deloitte.
Insights

雄心勃勃的体育场馆项目旨在弥合公共投资和私人投资目标之间的差距

体育场馆所有者致力于将体育场馆改造为可促进社会经济增长、推动社区参与和实现收入多元化的增长点

体育场馆、比赛场地和训练设施等体育基础设施投资呈上升趋势，原因是该等开发项目通常会为公共和私人部门带来广泛的社会经济效益。近年来，随着北美、欧洲和亚太等地区的体育团队大力投资于基础设施建设，相关投资趋势再次备受瞩目。政府和社区以增长为共同目标，可与体育投资者合作提供交通枢纽和社区资源等配套基础设施，助力提升体育运动的社会经济影响。多个基础设施项目有望于明年落地，将增加社区经济收益，并进一步推进体育与文化和社会的深度融合。

德勤预测，到2025年，全球将翻新或新建体育场馆300余个。根据德勤对在建体育基础设施开发项目的分析，预计近50%的新建体育场馆基础设施项目将在北美和欧洲落地。欧洲各国加大对体育场馆的投资，其中以足球场馆作为重点投资项目，以试图吸引新一波球迷，并为启动该等投资项目的组织提供多元化收入途径。如此一来，体育场馆开发有助于私人投资者实现投资回报最大化、加速公共部门实现社会经济目标。随着以体育设施为重点的翻新项目落地、球迷对场馆内外创新接触点的需求日益增加，全球多个地区的体育场馆投资有望实现增长。

场馆建设以社区为中心

体育组织可以促进社区团结、提高公民自豪感和凝聚力，并进一步丰富城市的文化活动。实施以体育场馆为依托、以体育运动为主导的翻新项目，通常需要与政府和其他主要利益相关者开展合作，推动战略项目落地实施，促进社区参与，并实现可持续繁荣发展，满足人们对居住和旅游的需求。

体育场馆建设不再仅仅只为满足单一俱乐部的利益需求。有关新建和改建体育场馆的各项决策应考虑社区利益。

2024年4月，现英格兰足球甲级联赛伯明翰城俱乐部所有者Knighthead Capital Management宣布，俱乐部计划建设以新的世界级体育场为中心的体育区。[1]主席Tom Wagner阐述了他对这一雄心勃勃的项目的愿景，即将体育场、男女训练场和学院团队都集中在距市中心步行可达的地点[2]。此外，他还提到了与酒店和其他商业实体的沟通情况，该等实体有意入驻该场地，并参与伯明翰东部重建项目。[3]Wagner表示，“蓝军”（Blues）将借此机会充分融入伯明翰城，并成为世界公认的“卓越灯塔”[4]。该体育区项目预计耗资20亿至30亿英镑，旨在推动西米德兰兹郡的社会经济长期发展[5]。Knighthead所有者团队就一系列战略优先事项与政府和公共部门开展合作[6]。

隶属于美国职业棒球大联盟的坦帕湾光芒队（Tampa Bay Rays）于2024年7月与佛罗里达州圣彼得斯堡就新建一座棒球场达成协议[7]。新球场的开发团队承诺建造1,250套经济适用房、创造30,000个建筑工作岗位和7,000个长期工作岗位，其中部分岗位留给当地居民和弱势居民[8]。项目负责人重申，其致力于通过新建新球场缩小周边社区的代际贫富差距，如未达到此目标，该项目则被视作失败[9]。

满足各代人的喜好，提高球迷参与度

不同年代球迷的体育消费方式有所不同。根据德勤英国发布的《2024年未来体育：“抓住时机”》，84%的受访全球体育领导者表示，他们预计不同的体育消费偏好在未来五年内将成为最具影响力的新一代趋势之一。体育组织应在秉承提供赛日体验这一核心传统与Z世代和Alpha世代对娱乐的更高期望之间取得平衡[10]。

各组织在打造体育场馆体验时，首先应考虑以下基本要素：舒适和安全、视野、优质的场内设施以及激动人心的氛围。该等要素对许多球迷至关重要，因此在提前制定任何计划之前应加以了解。

在确定上述基础要素之后，部分组织可能会在赛前、赛中和赛后为球迷提供端到端的娱乐选择，从而打造与众不同的体验。如此一来，不仅可以让球迷在体育场馆中投入更多时间与金钱，还可以帮助增强体育组织的社区意识。通过将社区文化融入体育场馆结构，有助于打造独具特色的比赛日体验。

为提升球迷体验，多伦多蓝鸟队（Toronto Blue Jays）开始对其位于罗渣士中心的体育场馆进行翻新，工程分为两个阶段。第一阶段于2023年完成，在球场看台内推出本地美食和娱乐等五个不同的外场“区域”，以满足球迷不同的体验需求，且各区域提供社交空间[11]。翻新工程还包括数字技术升级，例如“Tap N Go”（一种新的食品和饮料自动化市场服务）和“Walk Thru Bru”（可提升服务效率的饮料自动售货台）[12]。第二阶段包括实施以球迷为中心的调整举措，例如将座位向本垒板方向倾斜，以改善视线[13]。

智慧体育场馆

新一代球迷倾向于优先以数字化方式进行体育消费，比赛日体验亦是如此[14]。部分体育组织正在设计“智慧体育场馆区”，以整合先进技术，为球迷提供个性化体验[15]。全球智慧体育场馆市场不断增长，2024年市场规模超80亿美元，预计到2033年将超过380亿美元[16]。

由于Z世代和Alpha世代更喜欢简短、动态的内容，球迷参与需求正在发生变化[17]。随着球迷日益渴望获得价值，并愿意花钱买体验而非实物，体验感逐渐成为各组织的差异化因素[18]。新建场馆将融入游戏、商品销售业务等元素，其设计还考虑到了“第二屏效应”，即大多数球迷在观看体育比赛时通常会看第二屏幕。根据德勤研究，77%的受访体育迷表示，他们在观看体育赛事之余还参与了一项或多项与体育比赛相关的活动，包括查询数据、使用社交媒体或投注比赛[19]。新建体育场馆将利用集成技术来转播该等内容，使更多球迷将注意力集中在体育场馆内[20]。

美国国家篮球协会 (NBA) 洛杉矶快船队 (Los Angeles Clippers) 在其新球馆Intuit Dome举行了揭幕仪式，该球馆以打造独特的球迷体验为首要考量。新球馆的一大亮点是定制的 “光环板 (HaloBoard) ”，它优化了所有座位的视线，并优先考虑上层看台的观看体验。悬挂于中场上方的双面视频板包含比赛转播、提供深层统计数据的 “教练角”、即时回放、Steve摄像头 (跟拍洛杉矶快船队老板Steve Balmer) 、包括照片和其他个人宣传信息 (例如球员基本信息) 在内的球员资料功能等。为提升球迷体验，Halo Board还利用T恤大炮 (t-shirt cannon) 创新了令人垂涎的掷T恤游戏 (T-Shirt Toss) ，使上层看台的球迷也能获得商品。Intuit Dome计划为球迷的欢呼声予以奖励，为每个座位配备游戏机，供球迷在比赛日娱乐使用，进一步增加球迷的游戏热情，从而将球迷的参与度提高至体育界前所未有的水平[21]。

下一代球迷未来可能希望获得个性化、无缝的按需体验。体育场馆区可鼓励球迷延长场内逗留时间，享受美食、音乐和文化等不同服务以及为各类球迷提供的社交空间。体育组织正在建造该等体育场馆区，以便为主场球迷和客场球迷提供优质体验，并为场馆区内球迷创造新的接触点[22]。新技术简化了购买流程，使得购买更加快捷，例如点击即可获得商品或食品和饮料，以及为每位球迷提供个性化信息的票务软件[23]。此外，球迷的入场方式有所创新，这往往有助于为球迷带来更便捷的体验。梅赛德斯-奔驰体育场与达美航空 (Delta) 合作创建了 “飞行通道”（Fly Through Lanes) ，利用面部识别技术让球迷快速进入体育场馆[24]。

利用基础设施和技术，助力实现收入来源多元化

体育场馆所有者利用增强的基础设施和数字技术进一步实现收入来源多元化。

由于全球体育组织历来严重依赖转播收入来弥补球员工资和其他成本支出，体育组织的成功之道有所改变[25]。通常而言，中央转播收入增长将导致北美体育联盟的工资上限提高，并将导致部分成本控制法规所允许的欧洲体育组织的支出上限提高[26]。

各体育组织似乎均认识到，过度依赖单一收入来源将导致盈利潜力受到制约，并易受到新冠肺炎疫情等带来的市场冲击。由于精英体育俱乐部拥有文化和商业资本，部分组织目前正在利用其商业资产增加收入[27]。鉴于票务和转播收入通常受到容量限制或不在单个俱乐部的谈判范围内，而商业收入则是体育组织可在其控制范围内实现收入增长的杠杆，但未得到充分利用[28]。体育场馆发展为更广泛的娱乐区，有助于体育组织拓展服务范围，扩大商业版图。例如，英超托特纳姆热刺队通过在新体育场馆举办美国国家橄榄球比赛 (NFL) 和音乐会等非足球活动，其商业收入从2016/17赛季的7,200万英镑增至2022/23赛季的2.27亿英镑[29]。

2023年，欧洲足球巨头皇家马德里足球俱乐部 (Real Madrid) 对其圣地亚哥·伯纳乌球场 (Santiago Bernabeu stadium) 进行了翻修，使得2022至2023赛季俱乐部收入创下新高，各业务线均实现了增长 (除转播权以外，俱乐部正与转播商协商合同签订事宜) [30]。2024年7月，俱乐部宣布其收入超10亿欧元，创下足球俱乐部收入的最高纪录[31]。本财年下半年，俱乐部开展了以举办大型活动和推出尊享贵宾体验为重点的新商业活动，营业收入突破10亿欧元大关[32]。例如，哥伦比亚流行歌手Karol G在该体育场举办的四场演唱会为皇马带来了1,800万欧元的收入[33]。俱乐部有望在新赛季完成球场翻新工程，并力争在未来几年增加非足球收入[34]。

体育基础设施的新兴趋势

从满足女子体育的增长需求到推进体育产业的可持续发展，体育组织致力于满足球迷的新体验需求和优先事项。体育基础设施演变旨在为体育产业创造一个更具包容性、创新性和责任感的未来：

女性体育基础设施

随着女子体育的关注度不断上升、商业估值与日俱增，体育组织开始加大对女子体育专用基础设施的关注。在美国顶级职业女子足球联赛 (NWSL) 中，堪萨斯城潮流队 (Kansas City Current) 为其耗资1.17亿美元的河滨体育场CPKC体育场举行了揭幕仪式，该体育场被公认为首个专为女子职业运动队建造的体育场[35]。为进一步扩大体育场的社会经济影响，一项在足球场附近建造囊括公寓、酒店、餐馆和零售商的综合开发项目计划已获批准，该项目将耗资6.5亿美元[36]。体育场预计每年将为堪萨斯城带来近5,000万美元的经济收益[37]。

在美国国家女子篮球联盟 (WNBA) 中，新的训练设施可提升球场成绩，进而提高特许经营权的整体估值。拉斯维加斯王牌队于2023年启用了其训练设施，西雅图风暴队、菲尼克斯水星队和芝加哥天空队亦紧随其后[38]。

英格兰女子足球超级联赛 (WSL) 布莱顿和霍夫阿尔比恩足球俱乐部 (Brighton & Hove Albion) 计划为其女队建造一座专用球场[39]。曼彻斯特城女足 (Manchester City Women) 也已获准在曼彻斯特城市足球学院 (City Football Academy) 所在地建造一座专用训练设施[40]。

高端、个性化的接待服务

接待服务已不再是传统的企业服务，如今往往被用作为所有人群创造更易获得的差异化体验的工具。体育组织设法与该领域的高端品牌合作，提供从名厨到礼品袋等更多优质接待服务[41]。例如，一级方程式锦标赛在围场俱乐部 (Paddock clubs) 提供高端接待服务，接待社交媒体影响者、名人和品牌合作伙伴。体育组织在翻新和新建体育场馆时，可将接待空间设计地更加灵活，以适应不同类型的活动需求[42]。

可持续发展

更多体育基础设施新项目将可持续发展原则纳入规划。就以体育为主导的翻新项目而言，注重可持续发展能够展示积极的环境和社会实践，助力释放公共资金要素[43]。可持续发展对社区有益，将公共资金要素纳入基础设施亦有助于减少负面影响，并逐渐产生效益，包括削减能源账单。同时，这亦有助于提高品牌亲和力，创造更多合作机会，提升球迷参与度。体育产业与气候变化息息相关，体育产业与气候变化息息相关，既是气候变化的促成者，也是气候变化的受害者。体育场馆新建项目有助于推动大型建筑项目和交通运输的发展，而这却是全球碳排放的两大来源[44]。然而，体育行业也将受到气候变化带来的影响，热浪等极端天气条件或将对比赛、主办地和运动员福利产生负面影响[45]。

体育组织在商讨房地产开发项目规划时，应审慎考量可持续发展实践和战略。

小结

全球体育组织都致力于推进基础设施建设，以帮助提高球场容量，提升球迷的终身价值。体育场馆区可为私人投资者和所有者提供多元化收入途径，使其充分利用体育场馆的全年使用率（而不仅仅是比赛日的使用率），从而提高企业价值。数字接触点亦可为体育组织提供丰富的球迷数据，助其提供更加个性化、有针对性的产品。

对于公共投资者和政府而言，推进体育基础设施项目建设有助于造福广大社区。体育组织应努力培养社区意识、改善健康和福利状况，并吸引客流量。

精英[46]体育已成为经济和社会发展的强大助推器，可协调公共和私人投资议程。在不久的将来，体育组织可利用其体育场馆帮助突破体育产业的边界，进入更广泛的娱乐和数字产品领域。

By

Jennifer Haskel
United Kingdom

Pete Giorgio
United States

Alice John
United Kingdom

Kevin Westcott
United States

尾注

1. Alex Dicken, "*Tom Wagner reveals timeline for new Birmingham City stadium as Knighthead pledge billions*," Birmingham Live, April 9, 2024.

2. Ibid.

3. Ibid.

4. Alex Dicken, "*Another reason for Tom Wagner's Birmingham City takeover has now become clear*," Birmingham Live, Sept. 26, 2023.

5. Dicken, "*Tom Wagner reveals timeline for new Birmingham City stadium as Knighthead pledge billions*."

6. Dicken, "*Another reason for Tom Wagner's Birmingham City takeover has now become clear*."

7. Hines, "*Hines and Tampa Bay Rays gain approval of new ballpark, historic gas plant district development*," press release, July 31, 2024.

8. FOX 13 News Staff, "*Tampa Bay Rays, city of St. Pete sign deal to build new ballpark, keeping team in town for 30 years*," FOX 13 News, July 31, 2024.

9. Ibid.

10. Jamie Pugh and Zoe Burton, The Future of Sport 2024, Deloitte, Sept. 2, 2024.

11. Major League Baseball, "*Blue Jays showcase all-new 100 level seating bowl at Rogers Centre, as part of multi-year renovations*," April 4, 2024.

12. Populous, "*Blue Jays unveil completed outfield district of Rogers Centre renovations, designed by Populous*," April 6, 2023.

13. Toronto Blue Jays, "*100 level renovation*," accessed Nov. 5, 2024.

14. Pete Giorgio, David Jarvis, Brooke Auxier, Hannah Bobich, and Kat Harwood, "*2023 sports fan insights: The beginning of the immersive sports area*," *Deloitte Insights*, June 26, 2023.

15. Katelin Kharrati, "*Global smart stadium market size likely to expand at a compound annual growth rate of 22.5% by 2033*," press release, Custom Market Insights, June 28, 2024.

16. Ibid.

17. Ed Dixon, "*Study: Nine in 10 Gen Z sports fans use social media to consumer content as consumption habits shift*," SportsPro, June 28, 2023.

18. Giorgio, Jarvis, Auxier, Bobich, and Harwood, "*2023 sports fan insights*."

19. Ibid.

20. Gary Drenik, "*Stadium of the future: Emerging game day technologies for engaging fan experience*," *Forbes*, Aug. 18, 2022.

21. Ohm Youngmisuk, "*Storms, stats, and T-shirt cannons: LA Clippers' Halo Board goes all out*," ESPN, Aug. 16, 2024.

22. Deloitte, *2024 Sports Industry Outlook*, March 12, 2024.

23. Drenik, "*Stadium of the future*."

24. Mercedes-Benz Stadium, "*Delta Fly-Through Lanes*," accessed Nov. 5, 2024.

25. Deloitte Sports Business Group, *Annual Review of Football Finance 2024*, June 2024.

26. Bryan Toporek, "*The NBA's new TV deals are poised to send the salary cap skyrocketing*," *Forbes*, May 30, 2024.

27. Deloitte Sports Business Group, *Annual Review of Football Finance 2024*.

28. Sports Business Institute Barcelona, "*Commercial revenue: Increasing financial power of football clubs and leagues*," July 11, 2024.

29. Deloitte Sports Business Group, *Annual Review of Football Finance 2024*.

30. Gavin Hamilton, "*Real Madrid announces record turnover as stadium rebuild nears completion*," SportBusiness, July 18, 2023.

31. Guillermo Rai, "*Real Madrid surpass €1bn in revenue for 2023 to 2024 season*," *The Athletic*, July 23, 2024.

32. Real Madrid, "*Real Madrid becomes the first football club to exceed 1 billion euros in revenue*," July 23, 2024.

33. Conor Laird, "*The staggering sum Real Madrid earned from four Karol G concerts*," Yahoo Sports, July 24, 2024.

34. Ibid.

35. Kansas City Current, "*CPKC Stadium and University of Kansas Health System Training Center*, accessed Nov. 5, 2024.

36. Kevin Collison, "*Port KC approves massive project next to KC Current Stadium*," Flatland, April 23, 2024.

37. Ibid.

38. Women's National Basketball Association– Las Vegas Aces, "*Home sweet home! Aces take up residence in first-of-its-kind Women's National Basketball Association practice facility and team headquarters*," press release, April 29, 2023.

39. Morgan Ofori, "*Brighton ready to spark revolution with women's football stadium*," *The Guardian*, Oct. 29, 2023.

40. Simi Iluyomade, "*Manchester City Women are building a new £10 million training facility*," Versus, May 15, 2024.

41. Samuel Agini and Alice Hancock, "*Sports hospitality shifts focus to fans as UK demand for 'experiences' grows*," *Financial Times*, Aug. 13, 2021.

42. Georgina Yeomans, "*How Formula One transformed its hospitality product*," BlackBook Motorsport, Jan. 6, 2022.

43. Deloitte, *The Future of Sport 2023*, April 2023.

44. Center for Climate and Energy Solutions, "*Global emissions*," accessed Nov. 5, 2024.

45. Directorate-General for Climate Action, "*Sport—a key player in climate action?*" European Union, July 26, 2024.

46. Elite sports are defined as the highest level of competition, which may or may not be classified as "professional" sports where participants are paid for their performance.

致谢

The authors would like to thank **Tim Bridge, Je Harris, James Savage, Brooke Auxier**, and **Dhruv Garg** for their contributions to this article.

Cover image by: **Jaime Austin**

Deloitte.
Insights

监管放宽助力，无线电信市场整合提速

在许多市场，小型无线电信公司面临增长缓慢、利润低下以及债务压力。并购活动，尤其是资产整合乃至整个面向消费者的公司合并，在获得监管机构批准的情况下，或能带来转机。

德勤预计，2025年及以后，在欧盟的引领下，将有更多的电信行业内合并交易获得批准。[1]在许多地区和国家，无线电信市场较为分散，一些参与者规模较小。监管机构向来鼓励尽可能多的面向消费者的参与者参与市场竞争，以最大限度地促进竞争，进而降低消费价格。然而，越来越多的咨询专家向监管机构建议，允许市场内部进行整合可能更有利于未来网络的增长、功能、安全性和韧性的保持。

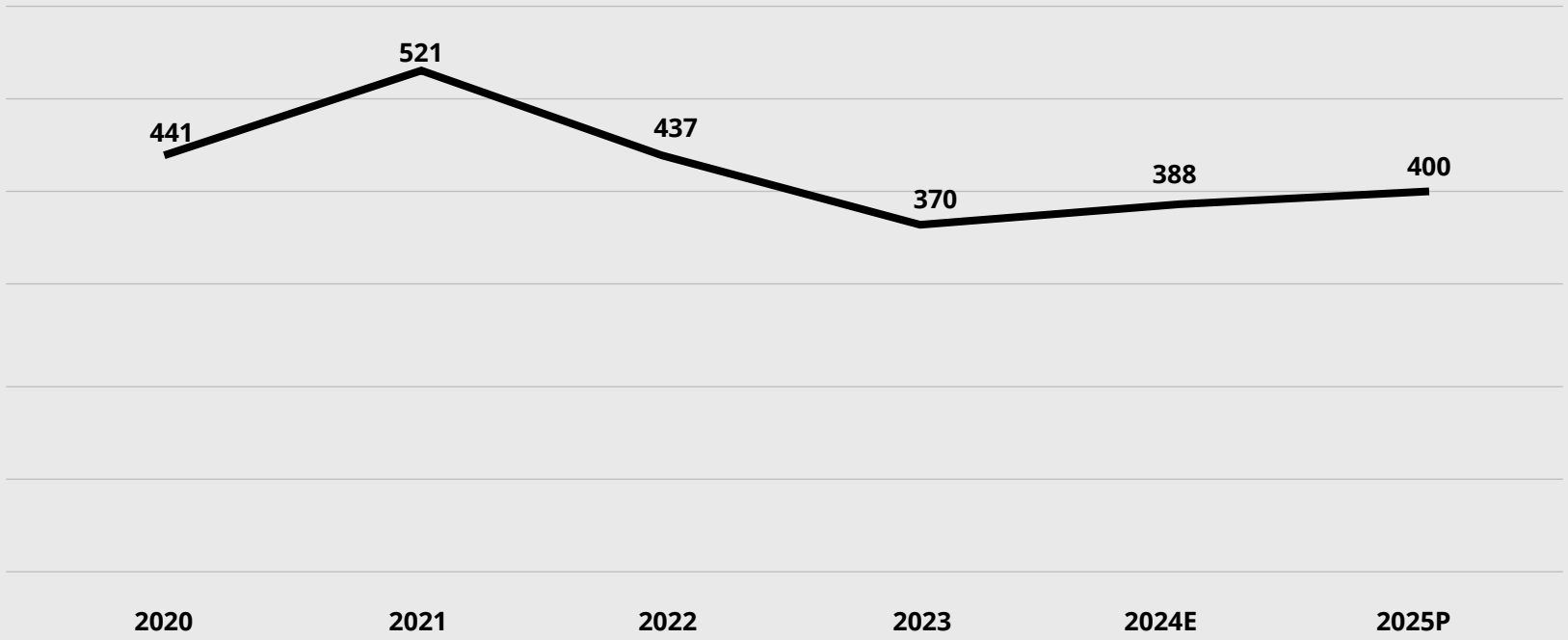
德勤预计，2025年将有约400项电信并购交易，这与过去五年的交易量基本持平（图1）。[2]这或许不足为奇，但值得注意的是——实际运营商之间的整合交易料将增加。

电信行业的并购交易类型繁多（图2），但总体来看，并未出现某一种交易类型占据主导地位的情况。无论是无线网络交易还是有线网络交易，各交易类型的占比随时间的推移保持相对一致，不过近期数据中心交易的活跃度有所上升，可能是受到人工智能数据中心相关活动的助推。[3]

图1

2020-2025年电信行业并购交易量

平均交易量



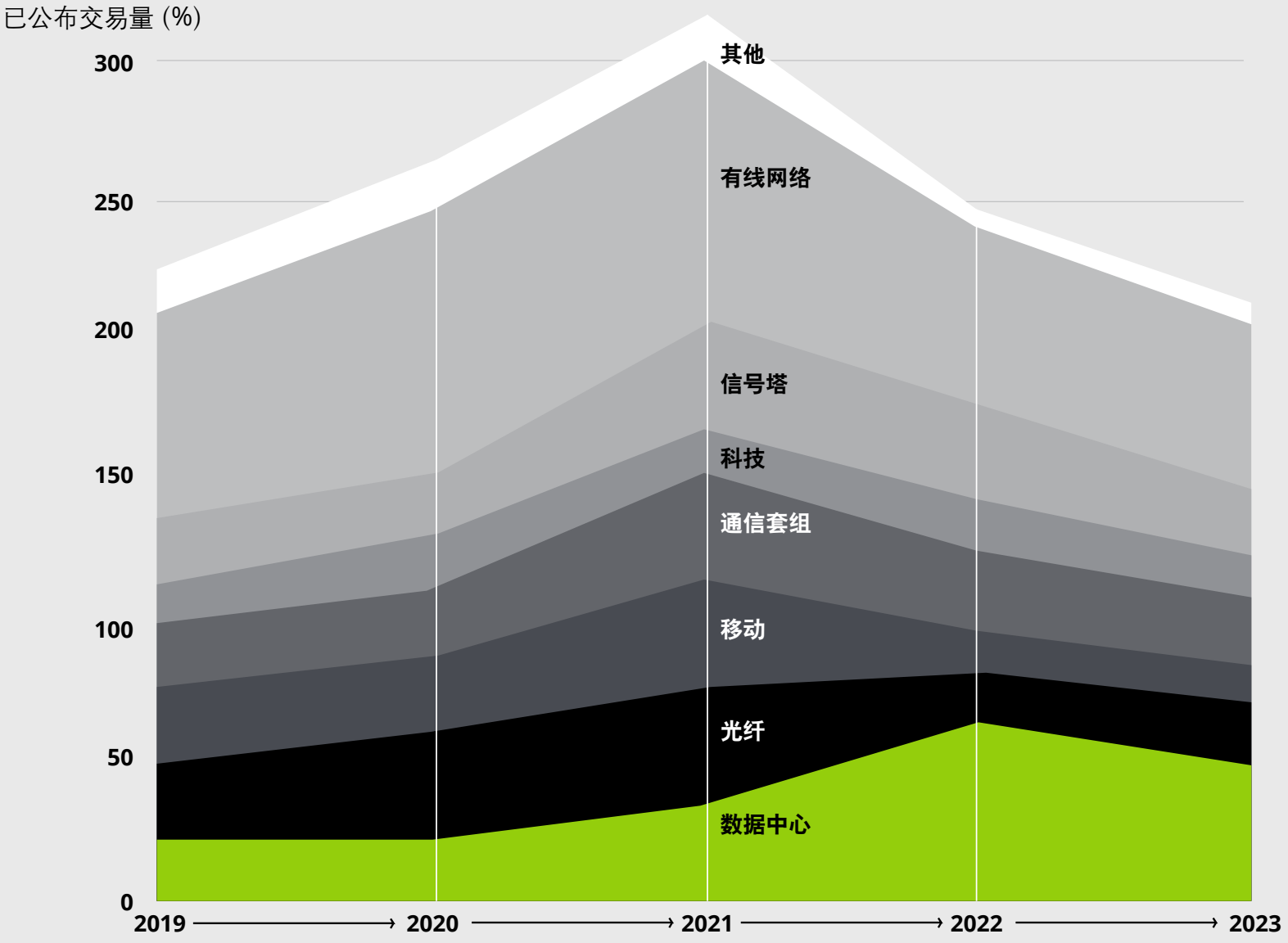
注：E表示当前年度的估计值，P表示德勤预测值

资料来源：德勤根据标普全球市场财智 (S&P Global Market Intelligence) ——标普Capital IQ和CB Insights的数据绘制的图表。
上图展示了电信行业并购交易的平均数量（2020年1月1日至2023年7月31日）。

图2

2019-2023年电信运营商并购交易量

已公布交易量 (%)



资料来源：Omdia刊载于CSI Magazine的文章，2024年7月23日。

部分形式的合并或分拆已持续多年，其增长阶段渐趋尾声：例如，截至2023年，美国和墨西哥97%的手机信号塔是由独立的信号塔公司而非电信公司运营（这一比例高于2016年的65%），而在欧洲，信号塔公司的市场份额从2016年的36%增至2023年的70%，几乎翻了一番。[5]

同样，多年来，有线网络（铜缆、光纤和同轴电缆）和后台软件系统（如计费和运营系统、现场服务车队和数据中心）也一直在经历整合和并购活动。[6]与此同时，各种无线网络的整合活动也在日益增多。

以下是一些无线网络整合的实例：自2009年起，加拿大两大无线网络运营商开始共享无线接入网（RAN）。[7]在马来西亚，原本有三家独立的无线网络，但政府于2021年决定建立全国统一的5G网络（尽管目前该国决定再建第二个全国性5G网络）。[8]在文莱，虽有三家移动公司为消费者与企业提供服务，但它们均使用统一国家网络有限公司（UNN）提供的同一无线网络。[9]2024年，澳大利亚两家运营商达成协议，共享4G和5G RAN。[10]

然而，无论具体情况如何，“零售商”（向消费者和企业提供电信服务的公司）的数量基本保持不变：对于电信服务的购买者而言，依然有三家甚至更多的公司在相互竞争。[11]

德勤预测的核心观点是：全球各国和地区政府以及监管机构正在放宽对电信企业合并的限制。自2020年以来，已有13项电信合并案或合资案获得了批准或正处于审批阶段，这导致面向客户的参与者数量有所减少：

- **美洲**：六项合并案（美国3项，加拿大、智利和哥伦比亚各1项）。[12]
- **亚太地区**：五项合并案（印度尼西亚、马来西亚、泰国、中国台湾和澳大利亚）。[13]
- **欧洲**：两项合并案（荷兰和西班牙）。[14]

观察人士正密切关注英国监管机构对沃达丰英国（Vodafone UK）与Three UK合并案的最终裁决。[15]值得注意的是，近年来意大利和丹麦的合并案均未能获得批准。[16]

此外，意大利前总理Enrico Letta于2024年4月向欧盟提交报告，明确呼吁加强电信行业整合。[17]他的建议部分基于欧盟的一份白皮书，其中讨论了电信公司在高度分散的欧洲市场上获得投资回报的困难。[18]有数据表明为何欧洲可能在批准合并方面走在前列：在欧洲，每家移动运营商的平均用户数为450万，而在美国为9,500万，印度为3亿，中国大陆为4亿。[19]意大利前总理、欧洲央行前行长Mario Draghi于2024年9月提交了一份69页的报告，其中也谈及对电信市场整合的支持。[20]

在大多数国家和地区，按收入和用户数量计算，排名前两位的无线电信运营商通常财力雄厚，排名第三的运营商往往经济实力较弱，排名第四（或第五或第六，视市场而定）的运营商则可能更弱。这些排名靠后的运营商常常预警称，未来可能无力继续投资网络建设。在这些市场中，有支持者向监管机构提出，相较于两强争霸，三足鼎立的市场格局更有利于保护消费者权益、激发企业活力和促进市场的公平竞争。

因此，欧洲等地区的无线电信市场或将迎来面向客户的整合。

监管立场转变的一个重要原因是网络连接选择的多样性。正如德勤2024年及2023年电信展望报告所述，近年来消费者和企业的选择激增，包括但不限于：

- **固定无线接入（FWA）对家庭宽带服务的竞争**。德勤预计，2025年将有超过3,000万户家庭采用FWA服务，同比增长20%。[21]

- **低轨卫星互联网对家庭宽带服务的竞争，尤其在农村和偏远地区**。目前，全球已有超过300万户家庭通过低轨卫星互联网实现了网络连接，预计2025年和2026年将有多个新网络推出。[22]值得注意的是，低轨卫星互联网在人口密度高、地理条件不复杂的国家（如欧洲部分国家）的作用较小，而在人口密度较低的市场，或有山脉、沙漠或许多小岛的市场，如亚太地区、非洲和美洲，发挥的作用则更大。[23]

- **3G、4G和5G网络仍在投入使用**。在某些市场，多个网络的共存为消费者提供了更多选择并形成了竞争。[24]

- **移动虚拟网络运营商（MVNO）正面临变革**。拥有25年发展历程的MVNO近期迎来发展拐点，开始在新增移动用户市场中占据更大的份额。[25]例如，截至2024年，美国有线MVNO服务的无线用户数已达约1,400万。这一成就的取得，部分归功于Wi-Fi技术的广泛应用。家庭Wi-Fi和城市公共Wi-Fi热点帮助美国有线电视公司分担了87%的无线数据流量消耗。[26]

小结

在未来几年，电信运营商维持强大且具竞争力无线网络的成本可能降低。对于发达国家的许多运营商而言，5G网络建设中成本高昂的部分，如设备和频谱购置，已大致完成。RAN建设支出在2022年达到峰值后，预计将在可预见的未来以两位数的速度下降。[27]大多数最初建立了5G非独立组网（NSA）网络的全球电信公司，在升级至5G独立组网网络时并未大幅增加投资。[28]而且，没有迹象表明6G会在2030年前到来。因此，德勤预计，2025年至2029年，电信行业的年度资本支出比例（2022年达到十年峰值17.8%）[29]将进一步降至15%~16%。这对网络运营商来说是个好消息，但对RAN原始设备制造商（OEM）则构成了挑战。

电信公司在5G等服务（除FWA之外）方面的盈利挑战依然存在。正如2024年预测报告所述，消费者对更高网速的追求不再强烈，大多数人不愿为此支付更多费用。[30]而诸如虚拟现实或增强现实高级服务、企业专用5G网络、自动驾驶汽车或远程外科手术等潜在收入来源，目前市场规模仍然有限。一些电信公司开始探索与电信相关的增值服务领域（如医疗、农业技术、安全等），但这些服务目前对公司收入的贡献尚显微薄。一些电信公司考虑涉足生成式人工智能数据中心业务以创造额外的收入和利润，但这通常是大型企业的选择，对于可能被合并的排名第三和第四的小型企业而言，并非合适之路。[31]此外，多数增值服务往往需形成一定规模才能取得成功，而如前所述，由于市场分散，欧洲大多数电信公司和亚洲较小的电信公司都缺乏这种规模。

在监管层面，全球范围内通常有两类监管机构负责无线通信合并案的审批。一类是行业监管机构（如英国通信管理局、美国联邦通信委员会以及欧盟的欧盟层面和国家级行业监管机构）[32]，另一类是竞争监管机构（如英国竞争和市场管理局、美国联邦贸易委员会、欧盟的竞争总司以及国家级竞争管理局）。亚太地区大部分国家也设有类似的两类监管机构。[33]

此外，从总体来看，行业监管机构通常对国内无线运营商的合并持更开放的态度，而竞争监管机构则表现得更为谨慎。预计这一现状在某些地区可能会有所改变，特别是受到欧洲近期发布的鼓励小型无线运营商合并的政策鼓舞。

尽管如此，监管机构在处理合并案时可能会更加审慎：最近的多项合并案耗时24到36个月才完成。[34]不过，德勤预计，未来获批的合并案比例可能会增加。

监管机构在批准合并时，有时会无条件通过，有时则会附加条件，如资产剥离、定价保证、未来投资承诺或提供5G覆盖。[35]

By **Duncan Stewart**
Canada

Dan Littmann
United States

Jack Fritz
United States

Ariane Bucaille
France

尾注

1. Deloitte analysis of recent events in Europe, specifically recent letters and white papers from Draghi and Letta (see endnotes 16 and 19).

2. Deloitte analysis of historical merger and acquisition trends, combined with early publicly available signs of deal activity.

3. *CSI Magazine*, "*Telecom consolidation: Over 500 M&A deals in five years*," July 23, 2024.

4. Duncan Stewart, Dan Littmann, Girija Krishnamurthy, and Matti Littunen, "*Telecoms tackle the generative AI data center market*," *Deloitte Insights*, Sept. 16, 2024.

5. Stephanie Price, "*For sale: Canadian read-throughs from global telecom asset sales*," CIBC Capital Markets, Aug. 14, 2024.

6. ArdorComm News Network, "*Telecom M&A activity witnesses surge: 514 deals from 2019 to 2023*," ArdorComm Media Group, July 25, 2024.

7. Sue Marek, "*Marek's take: 5G network sharing may be the answer*," Fierce Network, May 14, 2021.

8. Affandy Johan, "*5G in Malaysia – Single wholesale network driving regional leadership*," Ookla, March 17, 2024.

9. Digital Regulation Platform, "*Changing the operating model: the creation of UNN in Brunei Darussalam*," Aug. 27, 2020.

10. Australian Competition & Consumer Commission, "*Optus Mobile Pty Ltd and TPG Telecom Limited proposed network and spectrum sharing*," Sept. 5, 2024.

11. Megan Emfosi Meena and Jiaying Geng, "*Dynamic competition in telecommunications: A systematic literature review*," *Sage Open*, April 26, 2022.

12. Detecon Spotlight, "*Telco mergers and acquisitions*," 2022; M&A Community, "*12 significant telecom mergers and acquisitions over the last 20 years*," Sept. 17, 2024; Henri Capin-Gally, Sergio Márquez García Moreno, and Bill Parish, "*Energy and telco deals power Mexican M&A*," White & Case, March 6, 2024; Geusseppe Gonzalez, "*Access Alert: Colombian telecoms industry faces shakeup with potential Millicom-Telefonica merger*," Access Partnership, Aug. 8, 2024.

13. Julber Osio, "*Asia-Pacific telcos consolidate to compete with market leaders*," S&P Global, May 25, 2023; Tom Leins, "*Deal-making Down Under: Oceania's wave of telecom M&A*," TeleGeography, March 24, 2021.

14. Jan Frederik Slijkerman, "*Telecom Outlook: Will we see more mergers and buyouts in 2022?*" ING, Jan. 28, 2022; Jan Frederik Slijkerman and Diederik Stadig, "*Telecom tycoons on the move?*" ING, Feb. 1, 2024.

15. Competition and Markets Authority, "*How we are investigating the Vodafone / Three potential merger*," Sept. 13, 2024.

16. Slijkerman and Stadig, "*Telecom tycoons on the move?*" ING, Feb. 1, 2024.

17. Enrico Letta, "*Much more than a market*," Consilium, April 2024.

18. European Commission, "*White paper - How to master Europe's digital infrastructure needs?*" Feb. 21, 2024.

19. Hamish White, "*Europe's looming mobile crisis*," Technative, April 2024.

20. Mario Draghi, "*EU competitiveness: Looking ahead*," European Union, Sept. 9, 2024.

21. See section "Fixed wireless access: Contrary to popular opinion, adoption may continue to grow" in chapter *Updates: Past TMT Predictions' greatest hits and (near) misses*.

22. Ongoing Deloitte analysis of current and proposed low Earth orbit satellite networks.

23. Deloitte author conversations with communications providers in North America, Europe, Africa, and Asia.

24. Deloitte analysis of developing world wireless networks.

25. Piran Partners, "*Seizing the future of telecoms: The continuing ascendancy of MVNOs*," April 12, 2024; Puneet Takyar, "*The journey of MVNOs*," Comviva, April 1, 2021.

26. Jeff Baumgartner, "*Cable snared nearly half of US mobile line adds in Q3 – analyst*," Light Reading, Nov. 16, 2023.

27. Juan Pedro Tomás, "*Global RAN market faces challenging scenario in Q2: Dell'Oro*," RCR Wireless News, Aug. 19, 2024.

28. Deanna Darah, "*5G NSA vs. SA: How do the deployment modes differ?*" TechTarget, July 25, 2024.

29. Matt Walker, "*Telco capital intensity hits 10 year peak in 2Q22*," MTN Consulting, Sept. 6, 2022.

30. Paul Lee, "*No bump to bitrates for digital apps in the near term: Is a period of enough fixed broadband connectivity approaching?*" *Deloitte Insights*, Nov. 29, 2023.

31. Stewart, Littmann, Krishnamurthy, and Littunen, "*Telecoms tackle the generative AI data center market*," *Deloitte Insights*, Sept. 16, 2024.

32. DataHub, "*Regulatory authority - Institutional structure*," accessed October 2024; Policies, "*Telecommunications national regulatory authorities*," European Commission, Jan. 11, 2023.

33. Lynn Robertson, "*Interactions between competition authorities and sector regulators – contribution from business at OECD (BIAC)*," Organisation for Economic Co-operation and Development, Nov. 18, 2022.

34. Research Notes, "*5G rollout slowed while mobile operators await merger approval. Case in point: the slow rollout in UK*," Strand Consult, Oct. 5, 2024.

35. Deloitte author's assessment of multiple approved mergers in North America, Europe, and Asia in the period of 2015 to October 2024.

致谢

The authors would like to thank **Dieter Trimmel, Hugo Santos Pinto, Prashant Raman, Pankaj Bansal,** and **Akshay Jadhav** for their contributions to this article.

Cover image by: **Jaime Austin; Getty Images, Adobe Stock**

Deloitte.
Insights

云服务的精益管理：“FinOps”让每一分钱发挥最大效益

随着企业云支出不断增加，运用FinOps策略能让每笔投入的价值回报最大化，实现成本节约、价值提升及跨部门协同增效。

2025年，全球云支出预计将达到8,250亿美元，但企业高层可能难以言明具体的支出细节。[1]不少企业对于云支出要么不甚清楚，要么无法给出合理解释。

随着企业对云服务的依赖程度日益加深，制定一套有效的云投资管理战略变得至关重要。“FinOps”（即“Finance”和“DevOps”的合成词）正是这样一套助力企业监测和优化云支出的策略。德勤预计，仅在2025年，采用FinOps工具与实践就可为公司节省210亿美元，且这一数字在未来几年还会增长。部分公司在实施FinOps后，其云成本有望降低40%。展望未来，预计那些尚未建立FinOps团队的公司将会迅速行动起来，开始组建团队，而已具备FinOps丰富经验的公司则会进一步制定更高阶的优化策略。

云服务十分复杂，往往导致资源浪费

云服务已成为现代企业运作不可或缺的一部分。启动新的云环境只需简单几步操作，而构建私有物理基础设施则需要经过繁琐的采购和服务器安装过程，耗时可达数周甚至数月。云服务的便捷性和可扩展性赋能企业无需大量高端人才就能快速推进创新，助力视频点播、共享出行、挑战者银行和远程医疗等颠覆性行业的发展，[2]同时也支持数据分析、远程办公和人工智能等应用的实现。

如今，越来越多的企业投资于软件工程，现成的商业产品已难再满足它们的需求。例如，戴姆勒等汽车公司已组建开发团队，为电动汽车构建软件平台。[3]即使是传统行业（如木制托盘分销），对定制软件和专业知识的需求也在不断增长。[4]这些因素共同推高了云支出。

然而，云服务的复杂性也在加剧。企业日益倾向于采用私有计算资源与公有云服务相结合的混合云基础设施。目前，已有73%的公司采用了这一模式。此外，过半数（53%）的公司从多个云服务提供商处采购云服务，以利用促销活动、特定功能或避免供应商锁定。[5]各部门（如财务、人力资源或市场营销部门）常常在中央工程团队不知情的情况下采购云软件应用程序。这些情况均加剧了企业数据整合、合规性和安全性等方面的复杂性。

总体来看，企业在云预算管理方面表现欠佳。调查显示，去年有半数受访企业出现云预算超支，平均超支率达15%。[6]按需付费模式导致的成本波动，是造成预测难度加大的因素之一。有时甚至会出现某些极端情况，比如云工程师在一夜之间产生数千美元的费用。[7]

云服务并不便宜，其成本可能高于等效的私有基础设施，[8]并且正迅速成为企业最大的IT单项开支。例如，可口可乐公司 (Coca-Cola) 近期签署了一项价值11亿美元的云服务协议。[9]但更重要的是，据云服务相关负责人指出，27%的云支出实则被浪费。[10]意识到这一点，目前已有半数企业成立了专门的FinOps团队，另有20%企业计划在一年内成立FinOps团队。[11]

启用FinOps势在必行

FinOps是一门财务管理学科。其既涉及技术层面的内容 (例如重构云工作负载、审查适合长期保存的低频访问内容) ，也涉及技术含量偏低的内容 (例如折扣和信贷谈判) 。不过，FinOps的深远影响在于推动文化变革，其核心是确立跨部门的责任和财务问责制，确保每一笔云支出都与其产生的业务价值相对应。

启用FinOps的关键在于规划：首先，要审视现行战略，评估标记和警报结构，然后定义关键绩效指标 (KPI) 。[12]第一步是提高可见性，即对现有云资源进行清点，并探索如何与组织需求对接。为此，云服务提供商提供了资源监控工具和成本管理工具，或推出了一系列第三方FinOps平台，以提供更精细的指标。

然而，解读仪表板数据需要FinOps专家和专业人员，这类人才往往紧缺。此外，与多个云服务提供商合作的公司可能需要为每个提供商配置独立的仪表板。由于各提供商的数据馈送各不相同，因此建立统一的集成门户颇具挑战。最后，FinOps工具成本高昂 (占比高达云支出的3%~5%) ，公司在部署前应了解其带来的云经济效益。

FinOps应用入门：初步措施梳理

刚开始使用FinOps的企业可能会重点关注初步措施，以减少浪费、优化资源配置、审阅合同，并充分获得潜在的信贷和折扣。

浪费和消耗：实施FinOps的首要步骤可能是消除浪费。FinOps工具和仪表板可助力公司确定未充分利用或闲置的资源，以便及时调整或关闭，从而实现成本节约。例如，过大的虚拟机、冗余存储实例、孤立资源和重复数据等均属于资源浪费。熟练使用FinOps的公司或将采用预测分析来预估资源使用情况，并通过自动化脚本以动态调整容量。该项工作通常交由中央云工程团队迅速完成。

结构和层次：云服务并非一刀切。计算和存储实例的质量和价格各不相同。公司应评估其云资源配置的有效性，及其是否充分满足应用程序的需求。某些应用程序可能更适合成本效益更高的实例。例如，为了应对季节性波动，活动票务网站可能会选择一个在非高峰时段无需CPU持续满负荷运行、但在流量高峰期时能够短时间内提供足够资源的实例。[13]

激励措施：云平台提供的折扣计划可节省大量费用。有些云平台允许用户通过承诺稳定的资源使用量来换取更低的费率。对于某些公司，直接与云服务提供商重新洽谈可能是获取折扣的有效方法。云服务提供商往往乐于以折扣费率换取多年合同承诺。

FinOps应用进阶：全局视角下的精细化管理

2025年，经验丰富的FinOps实践者预计将继续深化FinOps应用，同时也会将其成本可视化和控制方法推向一个新阶段。

问责制：云服务对许多企业部门都至关重要，因此各部门（和团队）都应对云支出承担财务责任。通过Chargeback模式（直接向部门收费）或Showback模式（向部门展示其成本），让各部门监督并负责各自产生的云支出。[14]因此，企业需要实施可靠的标记策略，将资源成本分配给相应团队或项目，最好是根据预设规则自动标记。理想情况下，这有助于培养一种企业文化，让各个团队都参与到降低云支出的行动中。

本地部署：FinOps社区，如FinOps基金会，正在鼓励企业将本地基础设施（通常不为用户所了解）纳入整体成本管理方案。[15]企业应该考虑其整个IT资产的成本，但此项工作十分复杂，因为中央云团队需要与分支机构和基础设施站点沟通协调，而这些地方可能使用了各种硬件和软件工具来满足本地需求。为了降低本地部署成本，可以考虑取消不必要的许可授权或延长硬件的使用寿命。

可持续发展：FinOps还与日益兴起的“GreenOps”概念密切相关。GreenOps是一套推动可持续发展的云管理策略。FinOps报告工具提供的细化指标，有助于衡量能源消耗、碳排放等可持续发展目标。[16]随着欧盟《企业可持续发展报告指令》（CSRD）等主要报告法规的出台，跟踪并改进能源和碳排放指标，将成为企业投资FinOps获得的显著附加价值。

FinOps成就不凡

FinOps实践对许多力求降低云支出的公司至关重要：

- **爱彼迎（Airbnb）**[17]：旅游住宿应用Airbnb成功节省了6,350万美元的云服务成本。采取的策略之一是将数据存储迁移到成本更低的服务层，并用云服务提供商的解决方案代替了原有的自建备份系统。

- **Sky集团**[18]：传媒娱乐公司Sky在短短6个月内便耗尽了全年的云预算。该公司通过部署自有FinOps工具，节省了150万美元，并引入可视化仪表板，这在随后一年又为公司节省了380万美元。

- **Home Depot**[19]：家居装饰零售商Home Depot于2022年建立了一支专门的云成本管理团队，并较上一年节省了“数千万美元”。

- **Lyft**[20]：共享出行应用Lyft在6个月内将单次行程的云计算成本降低了40%，整个公司都可以使用基于电子表格的软件工具来跟踪计费数据。因此，公司内部掀起一股资源优化浪潮，为不同工作负载分配适当类型和规模的资源。

- **WPP**[21]：广告公司WPP仅在实施FinOps的前三个月就节省了200万美元，最终使其年度云支出降低了30%。该公司使用了自动生成的资源规模推荐等一系列工具和技术。

目前，许多企业都在积极投资FinOps。例如，FinOps基金会（致力于推广云成本管理最佳实践的非营利组织）的成员就包括了沃尔玛（Walmart）、万事达卡（Mastercard）和美国航空公司（American Airlines）。[22]

小结: 云的单位经济效益

FinOps的兴起，体现了企业对于提高云支出透明度、改进预算管理以及主动控制云支出的迫切需求，这类需求随着企业对云服务依赖程度的加深而不断增长。

展望未来，受数字化转型和人工智能的驱动，全球IT支出预计将持续攀升，到2025年将超过5.1万亿美元[23]。此外，目前企业约有半数的工作负载依靠私有基础设施，若将这些负载迁移至公有云，云支出或将显著增加。另外，相较于过去十年，当前的高利率环境促使企业更加重视盈利能力和成本削减，尤其是希望消除成本的波动性。换言之，FinOps的发展已是大势所趋。

FinOps不应仅仅被视为一个短期应急方案，而应当作为一项长期战略实践，成为企业运营战略的重要组成部分。尽管FinOps的起点是降低成本，但其终极目标是将云支出转变为战略资产和业务增长的驱动力。

部分领先企业的最终目标可能是创建“云单位经济效益”模型，以此量化每个云服务单位（包括各类应用程序、工作负载或已处理的千兆字节数据）的相关成本，并将其与相应的业务指标（例如收入、单次交付成本、单次预订成本和单次乘车成本）相对应。这种深入分析可助力企业基于全局业务视角做出有效的IT决策，确保每笔支出都与最终利润紧密相关。

对于部分企业，通过FinOps节省下来的成本可以重新投入到新的增长机遇中，例如通过新的云服务扩大业务规模，或加速制定产品路线图。

云服务固然复杂且成本高昂，但运用FinOps的企业能够充分利用其优势，实现利润的空前增长。

By **Ben Stanton**
United Kingdom

Adam Gogarty
United Kingdom

Paul Lee
United Kingdom

Gillian Crossan
United States

尾注

1. Gartner, "*Gartner forecasts worldwide public cloud end-user spending to surpass $675 billion in 2024*," press release, May 20, 2024.

2. Yury Izrailevsky, Stevan Vlaovic, and Ruslan Meshenberg, "*Completing the Netflix cloud migration*," Netflix, Feb. 12, 2016.

3. Douglas Busvine, "*Daimler to hire 1,000 programmers in Germany*," Reuters, April 18, 2021.

4. CHEP, "*CHEP uses 'track and trace' technology on its reusable pallets*," press release, April 8, 2022.

5. Flexera, "*2024 State of the cloud report*," 2024.

6. Ibid.

7. Parshv Jain, "*Avoiding costly cloud mistakes: Lessons learned from a $72K bill*," *Medium*, June 12, 2023.

8. Owen Rogers, "*Reports of cloud decline have been greatly exaggerated*," Uptime Institute, Jan. 18, 2023.

9. The Coca-Cola Company, "*The Coca-Cola Company and Microsoft announce five-year strategic partnership to accelerate cloud and generative AI initiatives*," press release, April 23, 2024.

10. Flexera, "*2024 State of the cloud report*," 2024.

11. Ibid.

12. Nikhil Roychowdhury, Nik Jethi, Farhan Akram, and Rishabh Kochhar, "*Optimizing the value of cloud: A guide to getting started*," Deloitte, March 30, 2023.

13. Amazon Web Services, "*TicketSwap tames demand ups and downs with AWS*," 2021.

14. FinOps Foundation, "*Invoicing & chargeback*," accessed Nov. 4, 2024.

15. For example: The Linux Foundation, "*FinOps across public cloud and on-prem*," accessed Nov. 4, 2024.

16. Meredith Shubel, "*What is GreenOps? Putting a sustainable focus on FinOps*," *The New Stack*, Sept. 22, 2023.

17. Magda Puzniak-Holford, Adithya Subramoni, and Simon Brennan, "*EU Corporate Sustainability Reporting Directive (CSRD) - Strategic and operational implications*," Deloitte, Sept. 8, 2023.

18. Belle Lin, "*Airbnb details road map to lower cloud costs*," *The Wall Street Journal*, Nov. 7, 2022.

19. James Ma, "*How Sky saved millions with Google Cloud*," Google Cloud Blog, July 19 2021.

20. Angus Loten and Isabelle Bousquette, "*Amazon warns of weaker cloud sales as businesses cut spending*," *The Wall Street Journal*, April 13, 2023.

21. Amazon Web Services, "*Lyft uses AWS Cost Management to cut costs by 40% in 6 months*," 2020.

22. IBM, "*How the world's largest ad company optimizes FinOps*," accessed Nov. 4, 2024.

23. FinOps Foundation, "*FinOps Foundation Members*," accessed Nov. 4, 2024.

24. Gartner, "*Gartner forecasts worldwide IT spending to grow 8% in 2024*," press release, Oct. 18, 2023.

致谢

The authors would like to thank **Nikhil Roy Chowdhury, Mitesh Gursahani, Nik Jethi, Rebecca Wood, Avishek Swain, Sophia Atkinson,** and **Vipul Mehta** for their contributions to this article.

Cover image by: **Jaime Austin**

Deloitte.
Insights

最新动态：回顾过往

本章节重新审视了过往对企业边缘计算、5G通信、女子体育以及后量子加密的预测。

2025年我们推出全新的系列短文，内容聚焦以往TMT行业预测的主题。历史证明，我们的预测成果颇为可观，但没有人能做到百分之百准确。欢迎您继续关注我们的文章，了解这些经典议题的新进展。

生成式人工智能走向企业边缘：“本地部署人工智能”盛行

搭建企业内部服务器，以构建更加私密、安全、灵活且低成本的人工智能信息技术环境。

德勤预测，2025年，尽管云端生成式人工智能仍将是主流选择，但全球约半数企业将在本地增设人工智能数据中心基础设施，其中一个例子是企业边缘计算。此举的部分原因是为了帮助企业保护知识产权和敏感数据，遵守数据主权或其他法规，同时也是为了帮助企业节省开支。这与我们在2024年底分析的情况一致：一种生成式人工智能芯片中，约45%流向超大规模企业，55%流向消费者、互联网和企业。[1]根据德勤2024年《企业生成式人工智能应用现状》第二季度报告，人工智能专业水平极高的企业中，有80%表示在云端人工智能上投入了更多资金……但61%的企业表示在自身硬件上投入了更多资金。[2]我们预测，到2025年，企业在本地部署的人工智能服务器市场规模将接近1,000亿美元。[3]

2021和2023年《科技、传媒和电信行业预测》报告中讨论了企业将如何投资边缘人工智能解决方案，以更快地执行和运行计算任务。[4]延迟（人工智能收到请求到给出响应所需的时间）是早期企业边缘应用的主要驱动因素。[5]而今年，延迟似乎不再是一个重要驱动因素：大多数生成式人工智能请求的处理时间为数千毫秒，因此周转时间通常不是问题。[6]

相反，对私有、主权和安全生成式人工智能的需求正在推动企业边缘技术迎来新一轮发展浪潮。

随着企业扩大对生成式人工智能的投资和投入，[7]云提供商、超大规模企业、电信公司以及人工智能和科技公司正在构建数据中心，为满足激增的需求提供支持，预计2025年在芯片和数据中心方面的投入将超过2,000亿美元。[8]但许多全球性企业正在采用一种混合方法：既使用第三方云解决方案，又投资硬件，在其本地进行部分训练和推理，构建更加安全、可控、自主和灵活的信息技术环境。[9]与使用外部云基础设施相比，在本地处理数据可使企业减少响应时间，并解决与生成式人工智能实施相关的隐私和安全问题——德勤在《企业生成式人工智能应用现状》第三季度报告中有所阐述。[10]

各种企业用例和机会使边缘生成式人工智能具有可行性和相关性。例如，银行和金融服务公司通常倾向于将大量敏感数据保存在本地，以解决数据安全问题，并加强对生成式人工智能模型的控制。[11]媒体和娱乐（M&E）公司已开始使用自然语言人工智能解决方案来激发动画和内容创作（如编写剧本初稿或文章）、游戏和娱乐（如利用电影字幕内容中的模式来增强面向最终用户的推荐引擎）领域的创造力。[12]

2025年生成式人工智能的企业边缘是什么样？企业用于本地生成式人工智能的部分支出将用于员工智能手机和个人电脑等设备，这些设备将配备更多专门的生成式人工智能芯片。[13]但还有其他选择，它们会影响企业在本地解决方案上的支出，以及企业可能需要在数据机柜或数据中心进行的调整，因为新版本往往耗电量更大、外观尺寸更大，有时还需要液冷技术（一种相对较新的解决方案）。

2024年，许多企业都配备了生成式人工智能盒子（市面上有多种选择和供应商），其大小如家用打印机，重约300磅，功耗约10千瓦，售价近50万美元，具有约30 PetaFLOPS的处理能力。[14]到2025年，一些公司可能会继续采购类似的设备，但也有公司会购买更大、功能更强的机型。这类机型比大型冰箱还要大，重逾3,000磅，功耗达160千瓦，需配备液冷系统，售价超过300万美元。[15]

需要明确的是，企业并非自行制造这些设备。生成式人工智能芯片制造商和多家服务器原始设备制造商均提供这些机架式服务器，并将为客户进行现场安装。

花费300万美元购置一台百万兆级的生成式人工智能超级计算机，乍一看似乎是笔巨额投资，但事实上却未必。各行各业的大型企业的IT预算动辄数十亿美元。而私有的大型语言模型（LLM）训练成本亦在一百万到一亿美元不等。[16]至于云端的生成式人工智能算力，无论是用于训练还是推理，也绝非免费。

企业不太可能仅依赖本地部署的生成式人工智能解决方案。在多数应用场景中，企业将不可避免地使用云计算……甚至，云计算很可能在企业人工智能计算中占据更大比重。遵循IT架构的一般趋势，企业最终或会采用混合云服务架构：在此架构下，本地部署在安全性、延迟、韧性和隐私方面发挥优势，而云服务则在选择性、灵活性、可扩展性和实验性方面提供便利。

但对于那些希望建立自有硬件的企业而言，从投资回报率的角度来看，购买本地部署的人工智能解决方案是否划算？答案往往是肯定的：非人工智能领域的本地计算成本普遍低于云计算，[17]预计生成式人工智能的本地计算亦具有相似的成本优势。即便“硬性”投资回报不够显著，但本地部署在知识产权归属、隐私保护、安全性和韧性等方面的优势，同样可以证明其价值。此外，新型的训练和推理技术（如拆分学习和拆分推理）可以将生成式人工智能的工作负载分散至边缘设备，进而优化计算需求、降低延迟，并解决隐私和安全问题。[18]

小结

有些公司大概认为，虽然他们短期内不需要本地部署的生成式人工智能计算，但从长远来看，这种部署是必然的趋势。因此，他们当前投入在学习如何最有效地使用本地计算资源上的时间和资金，作为混合云和本地部署策略的一环，将来或能证明是一次合理的硬件投资。

谈谈可能出现的生成式人工智能泡沫问题。众多企业纷纷投资硬件，担忧投资不足的风险超过投资过度，这可能在短期内造成严重的产能过剩。如果出现这种情况，已经投资本地生成式人工智能服务器的企业会如何应对？他们很可能更偏向于利用自有硬件（既已购买且在折旧），而非在云端的生成式人工智能计算上增加开支。

大部分预测都是基于这样的设想：例如，大型企业可能拥有自己的本地生成式人工智能IT基础设施，以支持其部分流程和业务（如银行、汽车制造商、医疗机构、政府部门等）。这就是我们（主要）所讨论的边缘计算市场。但也有人将电信边缘计算视为另一个边缘计算市场。根据我们九月发布的文章，全球已有超过15家电信运营商宣布正在建设生成式人工智能数据中心，旨在自用或向企业客户提供“生成式人工智能即服务”。[19]因此，企业可以在采购云服务提供商提供的本地硬件与“生成式人工智能即服务”之间作出选择。

最后，谈谈可持续性及其固有的权衡问题。企业边缘部署的生成式人工智能服务器在很多方面与超大规模数据中心的服务器类似。不过，在一千家企业中部署一千台服务器可以分散电力负荷，相比在单个场所集中部署一千台（或一万台）服务器，对电网造成的压力更小。[20]另一方面，超大规模数据中心的运营商往往能效更高，其电源使用效率（PUE）值普遍在1.2到1.3之间（且他们通常能够大规模采购低碳能源，而小型企业难以做到）。[21]与此同时，一般企业数据中心的PUE平均值介于1.67到1.8之间，这意味着在千家企业分散部署一千台生成式人工智能服务器，相较于在一个超大规模数据中心集中部署等量服务器，可能产生更高的碳足迹。[22]

（重新）确定女子赛事的投资案例

女子赛事收入不断增长，投资者热情高涨，收入估值创下纪录。

全球女子赛事的职业化和商业化程度不断提高，体育迷、赞助商及投资者热情高涨。德勤在去年的报告中预计，2024年女子精英赛事收入将突破十亿美元大关，预计收入将比2021年市场估值高出三倍。[23]

收入增长使得市场估值创下纪录，且流入该领域的资本规模不断扩大[24]，有助于提高精英[25]女子赛事的知名度、监管标准和赞助创新。

2025年，预计越来越多的投资者（包括机构投资者、私募股权和高净值人士）将关注这一领域。

德勤英国发布了《2024年未来体育报告：“抓住时机”》（2024 Future of Sport report: “Seizing the moment），（65%的全球受访体育领导者指出，女子赛事是该领域最大的增长机遇。女子赛事正在迅速发展，其获得的关注、观众、收入和投资都超过了我们最初的预测[26]。

虽然预计未来该领域将呈增长态势，但并不保证一定会实现增长。

受结构性因素影响，估值上升

包括美国国家女子篮球联盟 (WNBA) 和美国国家女子足球联盟 (NWSL) 在内的北美体育联盟正在为球队估值设定高标准。

2024年，“Caitlin Clark效应”席卷了WNBA，为联盟吸引了新的球迷和赞助商[27]。据报道，随着收视率、应用程序下载量和参与度的增长，WNBA达成一份价值22亿美元的新媒体转播权协议，是此前签订协议的三倍多。[28]

随着知名度提高，球迷参与度提高，一些投资者正在利用这一发展机遇。2024年8月，达拉斯飞翼队 (Dallas Wings) 在联盟中排名垫底，但在两位投资者以208万美元买下该队1%的股份后，其估值飙升至2.08亿美元[29]。在这笔交易之前，拉斯维加斯王牌队 (Las Vegas Aces) 以1.4亿美元的估值摘得最有价值球队的桂冠。[30]王牌队于2021年被Mark Davis以200万美元的价格收购，在对球队进行大量投资，包括投资4,000万美元建造球队专用的训练设施后，他的球队现在的价值可能是他收购价的70多倍。[31]支持这一投资论点的部分原因是，他相信WNBA球迷人数会持续增长，而目前的逆风形势看起来是有利的。金州女武神队 (Golden State Valkyries) 是一支计划于2025年开赛的扩军球队，目前已售出创纪录的17,000张预订季票，这表明在可预见的未来，需求有望持续增长。[32]

2024年9月，WNBA宣布扩军，将在俄勒冈州波特兰市新增设一支球队，该球队将于2026年开始参加联赛。[33]这支球队将由Raj Sports拥有和运营，由Lisa Bhathal Merage和Alex Bhathal领导，他们也是NWSL波特兰荆棘队的拥有者[34]。集团为这支球队支付了1.25亿美元，比为2026年参赛的金州队和多伦多队支付的5,000万美元大幅增加。[35]

NWSL的各俱乐部也享受到新签转播协议带来的好处。联盟与ESPN、CBS、Amazon Prime Sports和Scripps's ION Network新签了一份四年期媒体转播权协议，协议自2024年起生效，据报道每年价值6,000万美元 (总价值2.4亿美元) 。[36]该协议较上一轮与哥伦比亚广播公司 (CBS) 签订的三年期、总价值为450万美元的协议有所增长。[37]

联赛赞助水平有所提高，投资更趋成熟，推动联赛转播权价值增长。2024年，多家俱乐部易主，全年估值不断攀升。6月，私募股权公司Carlyle与西雅图海湾人足球俱乐部的所有权集团合作，完成了对NWSL俱乐部Seattle Reign的收购，据报道收购价格为5,800万美元[38]。前所有者OL Groupe在2019年收购该俱乐部的收购价格为约350万美元[39]。今年7月，南加州大学院长Willow Bay和她的丈夫、迪斯尼公司首席执行官Bob Iger宣布成为天使城足球俱乐部 (Angel City FC) 的新控股人，该俱乐部的投资交易价值为2.5亿美元，是有史以来估值最高的女子职业运动队。[40]据报道，Angel City FC在2022年加入NWSL时支付了200万美元的扩军费。[41]

欧洲女子足球在过去几年中不断发展壮大。2022年至2023年，欧洲一些顶级俱乐部的收入比上一赛季增长了61% 。[42]在欧洲，投资独立女子足球实体的机会通常极少，目前的合并结构限制了对女子足球的集中投资，因为潜在投资者需要投资男子足球队。[43]

然而，一些投资者正挑战现状，投资部分欧洲精英女子球队。2024年，Michele Kang完成了对奥林匹克里昂女子足球队52.9%股份的收购，该球队曾八次获得欧足联女子冠军联赛冠军。[44]在该笔开创性的交易中，Kang同意支付50年的奥林匹克里昂女子足球队的品牌使用许可费。[45]此次收购增加了Kang的投资组合，继2022年收购NWSL的华盛顿精神队 (Washington Spirit) ，2023年收购英国为数不多的独立女子俱乐部伦敦城雌狮队 (London City Lionesses) 之后，Kang的多俱乐部所有权模式在全球足球领域不断发展壮大。[46]

在英格兰女子超级联赛 (WSL) 中，切尔西女足也宣布了一项新的战略增长计划，该计划将重新定位切尔西女足，使其在俱乐部的整体业务结构中与男足并驾齐驱，而不是处于男足之下[47]。这样可以直接向女子足球队进行资本投资，而非通过男子球队进行投资，其他附属女子球队也有可能效仿这一做法，以实现直接投资渠道。

小结

女子赛事有望在2025年继续保持增长势头。去年，一些激进组织和投资者进军该市场。更多的利益相关者可能需要效仿，以便抓住机遇，推动市场突破一次性投资的局限。部分女子赛事资产的交易估值倍数高于整个行业的一般估值倍数，但随着收入的增长，未来交易的估值倍数可能会降低。明年，更多转播权持有者可能会对其投资结构进行评估，因为随着收入和球迷参与度的提高，市场需求也将随之增长。

固定无线接入 (FWA)：与普遍观点相反，FWA采用率或将持续增长

美国FWA净增用户数可能略低于去年，而部分市场的FWA净增用户数或将到2026年才会实现高速增长……无论美国或是全球，FWA净增用户数均存在未实现增长或潜在增长。

过去几年里，固定无线接入 (FWA) ——即消费者通过固定蜂窝设备 (主要是5G) 而非电线获得家庭宽带服务——在美国5G发展中发挥着重要作用。预计到2024年底，将有超1,000万户家庭接入。[48]极具竞争力的价格和高速网络吸引着广大消费者，[49]而对运营商而言，在人口密度较小的地区铺设光纤不具成本效益，因而在此类地区提供宽带服务更经济实惠。[50]

但就固定无线接入 (FWA) 而言，美国2024年第一季度的净增用户数 (又称净增量，或季度新增用户数减去取消订阅的用户数) 低于2023年第一季度，[51]而印度FWA市场 (净增用户数有望大幅增长的另一个全球市场，预计到2030年将有1亿用户) 仍处于起步阶段，许多人预计2025年FWA增速将有所放缓。[52]

根据德勤《2022科技、传媒和电信行业预测》报告，FWA年增长率接近20%，到2023年全球FWA用户 (主要是4G用户) 数量将低于1亿。[53]FWA实际增长与预期基本会一致，到2024年底，全球FWA用户总数有望超过1.5亿 (其中约30%为5G用户) 。[54]

德勤预测，受以下三大趋势的推动，2025年和2026年全球FWA净增用户数将继续以每年约20%的速度增长。

1. **部分市场FWA增长未引起全球关注**。美国和印度均是规模较大的FWA市场。美国FWA的季度净增用户数接近100万。[55]随着印度FWA数量的增长，季度净增用户数或将随之增加。[56]但在其他市场，FWA也有所增长，例如，意大利FWA的季度净增用户数为100,000，这并未引起意大利以外市场的关注，[57]但以家庭为单位，意大利的FWA增长率仅略低于美国。[58]相比欧洲其他国家、拉丁美洲、东南亚和非洲的大多数国家，意大利的FWA增长率较高，[59]但即便如此，预计2025年和2026年美国和印度以外市场的净增用户数仍将超数百万。

2. **企业越来越多地选择使用FWA**。迄今为止，美国FWA增长主要来自消费者。近期，越来越多的企业（主要是中小型企业）开始使用FWA连接。[60]FWA提供了单点联系、统一计费且安全性更高，部分企业对FWA的兴趣日益提高。[61]德勤预测，2025年和2026年，美国FWA企业用户数将分别超过100万。[62]

3. **新技术提高了各市场5G用户数的上限**。过去几年里，一些人认为美国FWA自然存在一个上限：鉴于现有5G技术和可用频谱（主要包括2.5GHz和3.5GHz频段），在不影响固定和移动无线用户体验的情况下，美国5G FWA用户数很难超过约1,000万。[63]人们认为，在短期内，某些地区运营商将被迫停止扩大FWA用户规模。[64]不过，即使FWA用户数趋近1,000万，用户的下行速度也在不断提高。[65]一些5G技术不断升级（主要是围绕无线电技术的先进应用），表明高达2,000万户美国家庭（在约1.06亿户美国宽带家庭用户中占19%）[66]可使用现有频谱连接到FWA，这意味着目前美国每年约300-400万FWA新用户的运行速率可能至少还能持续几年。

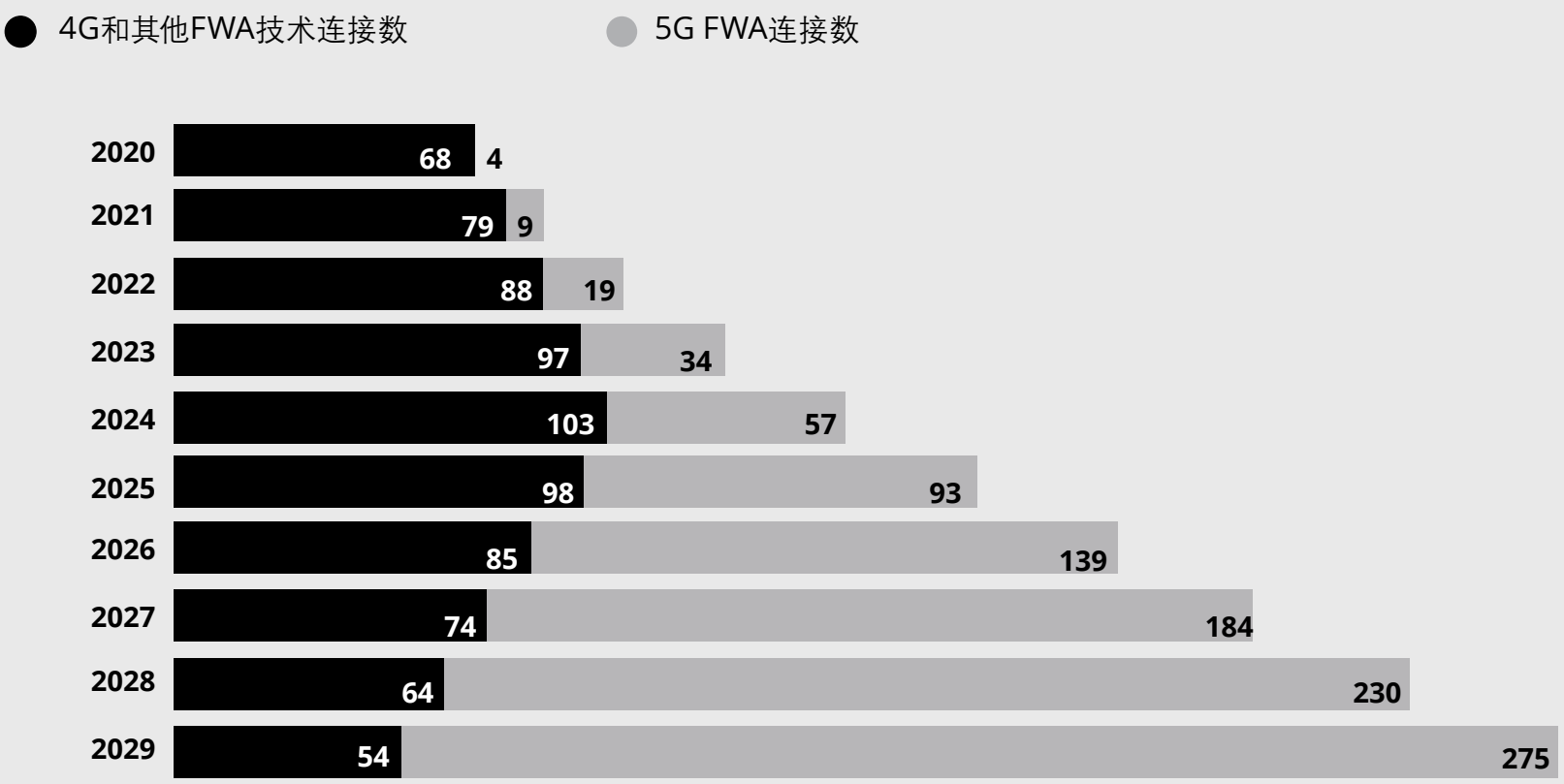
小结

FWA的持续增长或有利于电信公司实现5G投资货币化。截至2025年，除FWA以外，大多数5G服务收入增长潜力大，但规模较小。[67]举例而言，美国FWA的平均价格约为每月50美元，到2025年，FWA用户数增加400万，增量收入将增加24亿美元，此外还能通过捆绑服务减少移动用户流失。[68]

另一方面，提供其他类型宽带接入（同轴电缆、铜缆DSL、光纤）的公司或将继续面临来自FWA的竞争，恐将造成这些公司用户流失或难以维持或提高价格。此类宽带接入在多个市场上始终面临来自FWA的竞争，但FWA的发展不似提供其他类型宽带接入的公司所期望的那样接近尾声。

图1

FWA连接数



资料来源：《2024年爱立信移动市场报告》固定无线接入展望

5G独立组网进展缓慢：6G到来会否延期？

面对投资回报疑虑，电信公司重新评估5G独立组网投资，6G推进或受影响。

5G独立组网网络的部署进展比最初预期的要慢。[69]电信公司在是否大力投资下一代5G方面显得犹豫，部分原因是现有5G投资回报不理想。目前看来，6G的推出似乎愈发遥远。

截至2024年3月，在全球已推出5G服务的585家运营商中，仅有49家部署、推出或试运行了5G独立组网（SA）网络，占比8%。[70]这一较低的采用率反映出，面对资本回报的不确定性，运营商短期内可能不愿在独立组网上投入资金。此外，5G SA商业用例的缺失也对推广构成挑战。运营商正在放缓5G SA部署步伐，由于缺乏明确的收益途径，他们对5G SA基础设施的大规模投资仍保持观望。[71]因此，德勤预计，到2025年，新升级到独立组网的网络数量将不足20个，5G SA在所有5G部署中的占比将保持在12%左右。[72]

5G于2018年面世之初，大多数运营商选择了非独立组网架构（NSA），即在现有的4G基础设施之上构建5G网络，以加快服务的推出。[73]采取这一策略的主要原因是当时5G SA设备尚不完备，同时也希望充分利用现有设施。[74]因此，NSA的实施提高了5G服务的部署效率，也更具成本效益，使运营商能够提供更快的网速和更佳的连接性，尽管它不具备5G SA的全部功能。[75]然而，德勤全球曾在2022年预测，投资5G SA网络（包括试验、计划部署或实际推出）的移动网络运营商数量将翻一番，从2022年的超100家增至2023年底的至少200家，但这一预测并未实现。[76]

当前的4G和5G NSA网络已足以高效支持最常用的消费者及企业应用，这可能降低了运营商大力投资5G SA的紧迫性。[77]在企业应用层面，尽管5G具有颠覆各个行业的巨大潜力，但由于基于SA的新兴5G方案的开发和推广速度低于预期，导致运营商在5G SA部署上更为谨慎。[78]

5G SA基础设施对于测试和验证6G的基础技术至关重要。[79]它支持5.5G和6G的关键应用场景，例如低延迟、高可靠性和网络切片。[80]5G SA部署的延迟可能会阻碍这些技术的开发和测试，从而影响6G的推进。

由于成本高昂、商业化进程迟缓、用户接受度低下，电信公司在实现5G投资回报方面遭遇难题。[81]此外，5G网络应用编程接口 (API) 等其他5G服务的预期营收也尚未实现。[82]由于6G用例仍在开发中，移动网络运营商可能会专注于最大限度地发挥5G的潜能并收回投资，此举可能会进一步推迟6G的开发和部署。虽然预计相关的研究和标准制定工作将在前期展开，但普遍观点认为， 6G的大规模商用将在2030年左右实现。[83]

小结

移动网络运营商 (MNO) 在布局5G SA时，务必考虑持续优化现有5G NSA网络的性能。针对需求较高的地区或行业，采取分阶段投资的策略部署5G SA，能够更有效地控制成本并集中资源。此外，探索新的收入模式，如将5G SA作为一项服务提供给特定行业，或将其与云计算、人工智能和边缘计算等服务打包销售，也能开辟新商机。与企业客户合作，协助开发针对制造、医疗、物流等行业的定制解决方案，也将有助于推动5G SA的普及和对该技术的进一步投资。

各国政府和监管机构应认识到，SA部署是一个持续的过程，耗时可能超出预期。这可能会对5.5G和6G的发展产生影响，他们或需重新考虑规划下一代网络需求和频谱分配的策略。在2021年5G投资高峰之后，预计本世纪20年代末至30年代初将迎来新一轮投资高峰，在两轮高峰之间的预期投资低谷，电信设备制造商在一定程度上期望通过部署SA网络来获得持续收入。[84]如果SA依旧发展缓慢，这一低谷可能会比预期的更深、更长，从而对收入和盈利能力产生潜在不利影响。

开放式无线接入网（RAN）移动网络与供应商选择：目前采用单一供应商模式，何时实现多供应商模式？

开放式RAN迈向多元化、多供应商生态系统进程缓慢，且面临错综复杂的挑战。

开放式无线接入网 (OpenRAN) 旨在为构建RAN的移动网络运营商 (MNO) 提供更多选择并提高其灵活性，以实现网络民主化，从而有望以更低价格提供优质网络服务。尽管人们对这一系统寄予厚望并给予高度认可，但事实证明，向多元化、多供应商生态系统过渡比一些人最初预计的更加缓慢和复杂。[85]德勤预测2025年将不再部署或公布其他多供应商开放式RAN网络，因此实现真正的多供应商开放式RAN仍需时日。[86]

无线接入网 (RAN) 是蜂窝网络的重要组成部分，用于管理移动设备与核心网络之间的无线通信。无线接入网 (RAN) 过去作为封闭的专有解决方案：整个网络由一家供应商提供，其他供应商的设备无法与之配合使用。[87]开放式RAN作为电信行业的一个变革性概念应运而生，旨在实现网络组件设计和实施的标准化和民主化。[88]2018年随着O-RAN联盟成立，旨在推动搭建更加开放和可互操作的RAN架构，开放式RAN在21世纪10年代末的发展势头日益强劲。[89]

德勤预测，2021年全球动态公共网络开放式RAN部署数量将翻一番，从35个增至70个。[90]该预测过于乐观：截至2024年3月，全球正在部署和试验的公共网络开放式RAN数量为45个，仅有两个网络为多供应商开放式RAN。[91]

开放式RAN的一大目标是为运营商提供更开放的选择，以促进市场竞争和多样性。[92]然而，目标尚未实现。许多运营商仍然从同一个供应商处购买用于任何特定站点的无线电和基带产品，迄今为止，单一开放式RAN公司的市场影响力微乎其微：五大RAN供应商目前占据了约95%的市场份额。[93]

预计2024年，开放式RAN的市场规模仅占整个RAN市场的7-10%，即整个RAN市场份额预计为350亿美元，开放式RAN约为25-35亿美元。[94]此外，在不久的将来，单一供应商解决方案仍将比开放式RAN应用更为广泛。到2028年，单一供应商开放式RAN解决方案预计将占RAN总收入的15%-20%，而多供应商开放式RAN解决方案预计市场占有率仅为5%-10%，传统RAN的市场占有率仍为80%-85%。[95]尽管许多业内人士仍然认为多供应商开放式RAN具有巨大的长期潜力，[96]但这一愿景的实现仍需时日。

如要推动开放式RAN发展，须整合各种硬件，以实现高容量性能和成本效益。运营商历来倾向于从传统RAN供应商处进行一站式采购，原因是传统RAN供应商负责处理一切事务并承担单一来源采购主体责任。[97]多供源采购尤其会对小型运营商构成挑战。

然而，开放式RAN市场是美国广泛战略的一部分，旨在帮助加强经济安全、减少对外国供应链的依赖，尤其是电信等关键领域。[98]其中，开放式RAN技术对实施该战略至关重要，旨在打破目前主要由少数几个非美国公司主导的市场集中格局。[99]地缘政治紧张局势导致其中一些公司被排除在美国和欧洲等重要市场之外，从而限制了供应商选择，这凸显了供应链多元化的必要性。[100]开放式RAN通过加强国内市场竞争和创新有助于解决这一问题。利用开放式RAN将业务迁回本国成为一个战略机遇，有助于推动电信设备的国内生产，并保持其在全球电信基础设施市场的领导地位。[101]

小结

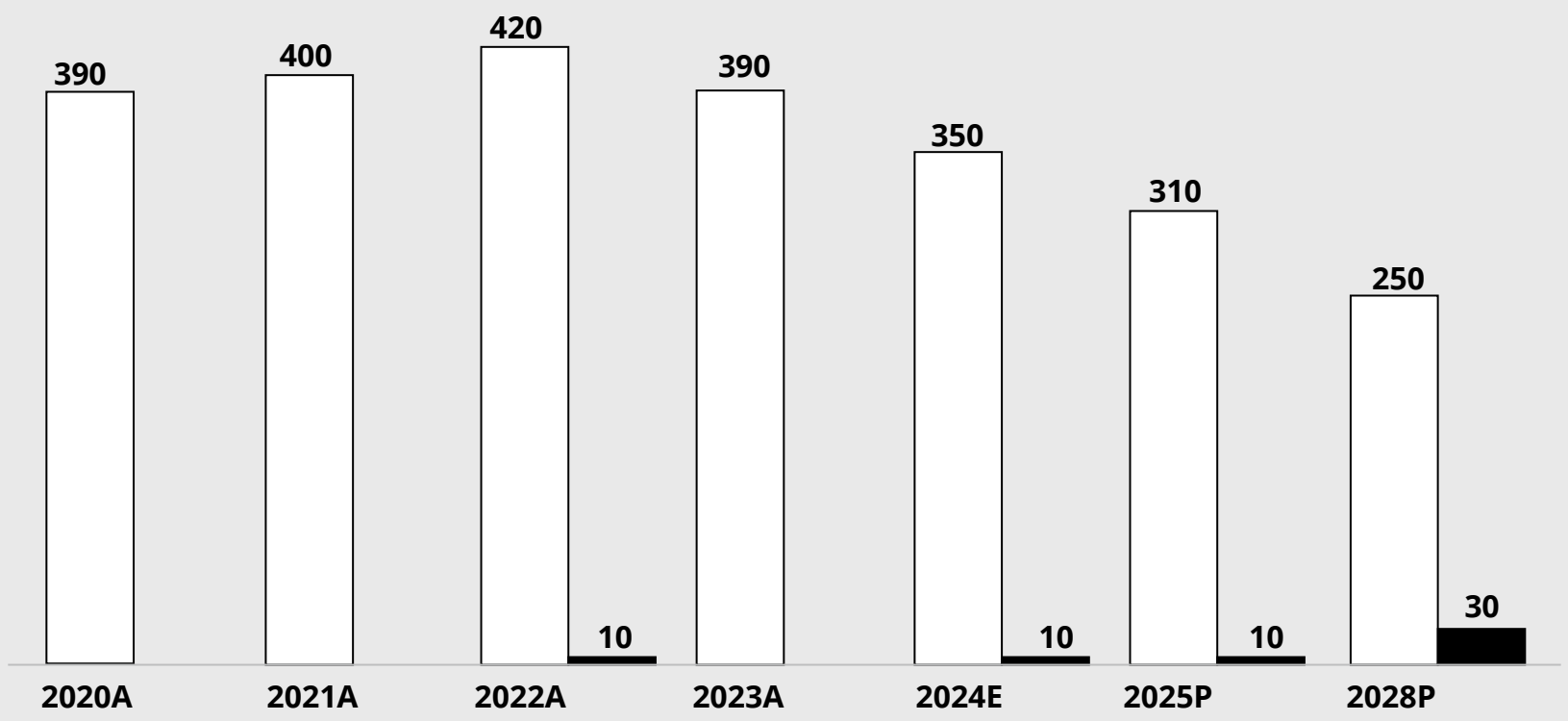
大多数移动网络运营商已将开放式RAN列入其规划议程：此议题已讨论多年。[102]部分运营商已开始采取以下措施；其他运营商预计将在未来一两年内开始采取相关措施。移动网络运营商可将其组件逐步集成到其网络的非关键区域，以低风险和实验性的方式探索开放式RAN。与其他运营商和供应商或联盟进行合作测试，有助于在部署多供应商解决方案时解决相关技术挑战。与新兴的开放式RAN供应商以及已退出市场的RAN供应商展开密切合作，亦有助于开发满足特定网络需求的解决方案。此外，针对开放式RAN技术的内部能力和专业知识提升进行投资，无论是通过员工培训还是聘用专家，均有助于应对多供应商环境的复杂性。

现有大型RAN供应商希望在“开放”之间取得平衡：从某种程度上讲，开放式RAN可能会被视作对其现有业务的威胁，但如果开放式RAN已成必然趋势，大型RAN供应商或许会考虑先发制人，尽早进行变革。原始设备制造商 (OEM) 已开始专注于开发可与现有网络基础设施无缝集成的互操作性产品。[103]他们很可能会继续与开放式RAN领域的小型企业和初创公司合作。与行业联盟保持交流有助于原始设备制造商和新兴的开放式RAN供应商确保其产品与不断发展的标准和趋势保持一致。

多供应商开放式RAN的一个益处是可增加供应商的多样性，尤其是欧洲和亚洲以外的供应商。[104]如果开放式RAN目前仍缓慢发展，则供应商的地域平衡在短期内不大可能发生实质性变化，因此可能需要其他工具来帮助实现现有RAN供应链多元化。

图2

2020-2028年全球RAN和多供应商开放式RAN收入（单位：亿美元）



资料来源：德勤分析和Dell'Oro。A为实际值，E为当前年度的估计值，P为德勤预测值。

量子技术起步虽慢，网络安全防御不容滞后

量子药物发现与金融建模尚待时日，但量子时代的网络防御升级却刻不容缓。

我们在2019年和2022年的科技、传媒和电信行业预测报告中探讨了量子计算，[105]并在2022年的预测中提到了随之而来的加密威胁。[106]其中部分预测已经成真，例如目前的量子计算机尚不成熟，无法满足实际应用需要；[107]网络安全领域受到高度关注。[108]此外，还如德勤在2023年12月的预测所言，今年网络安全标准的各方面建设都取得了成果。[109]例如，美国国家标准与技术研究院（NIST）近期发布了后量子加密标准[110]。（NIST发布的标准被广泛认为是网络安全和加密领域的黄金标准，尤其在加密和数字签名算法、协议以及框架的开发与应用方面，旨在确保数据通信安全和交易保护。）此外，肖尔算法（1994年提出的一种特定算法，可利用量子效应快速破解公钥加密方案[111]）有朝一日若在量子计算机上得以实现，将对网络安全构成严重威胁。

德勤预测，相比2023年，2025年致力于实施后量子加密解决方案的公司数量预计将增至四倍，其相关支出也将翻两番。该预测是基于2025年至2035年联邦系统迁移成本的估算[112]，以及金融服务业为降低上述风险所做的持续努力，此为保守估计数据。[113]

务必周密考虑扩展量子用例与构建量子网络安全防御的时间线。关键并不在于市场如何以及何时使用量子计算机来实施积极用例或肖尔算法，而在于各组织需要多长时间才能采纳并实施最新发布的NIST标准，以确保依靠云服务巨头、专用安全组件以及“自研”应用程序来利用加密库构建保密性和信任等安全功能？为寻找答案，已有组织开始分析潜在的网络威胁，规划与企业使命契合且满足运营需求的升级路径，并研究未来采购和合同条款，以迎接量子网络时代的到来。

据称，已有国家或其支持的行为者（以及其他行为者）在窃取加密数据，以备日后技术成熟时使用量子计算机对其进行解密（也称为“先窃取，后解密”攻击），这一行为加剧了时间线的不确定性。[114]且看企业将如何应对此类“沉睡的威胁”，以及会否主动采用NIST标准来防范未来的数据泄露风险，尤其当这类风险威胁到个人信息等需要长期保护的数据时。

值得关注的是，已有大型供应商开始在其平台引入后量子加密技术。[115]预计目前提供端到端加密功能的其他消息服务平台也将在2025年及以后实施后量子加密技术。此外，超大规模云服务提供商也在推出相应服务，以支持客户进行基准测试、原型设计，以及评估后量子加密技术对云服务性能的影响。[116]

中国持续加码量子技术相关政策与监管

当前，中国已制定和实施多项政策来应对量子技术可能导致的泄密风险。例如2024 年 发布了 《量子密钥分发 (QKD) 网络安全技术要求》和2023年发布《关于促进数据安全产业发展的指导意见》，规定了 QKD 网络的安全技术要求，包括量子密钥分发设备、网络节点、通信链路等方面的安全技术指标和安全防护措施。中国科技企业也正在积极研发量子技术加密技术，例如，2025年1月3日，国内一家科技企业发布了最新的量子加密通信系统，采用256位的量子加密方式，核心优势在于其高达256位的量子加密方式。未来，中国政府将进一步加强与企业的合作，持续提升量子技术防护能力，保障境内数据与网络安全。

小结

量子优势是指量子计算机在解决实际问题方面比传统计算机更具有优势，但如短期内没有重大突破，这种优势可能仍需数年才能显现价值（图3）。随着下图蓝色阴影区域的临近，企业在用例开发与商业化之间的过渡阶段，应考虑从风险防控的角度出发，采取对策来应对加密威胁。企业应了解该风险相比其他风险的紧迫性，并据此积极寻求降低相应风险。[117]

即便有组织成功开发并能够操控量子计算机用以解密，也大概率不会进行大肆宣传。因此风险的发生可能不会有提前预警，而尽早采取行动就显得更加重要，可避免组织在应对相关威胁事件时缺乏资源或导致系统关停。[118]

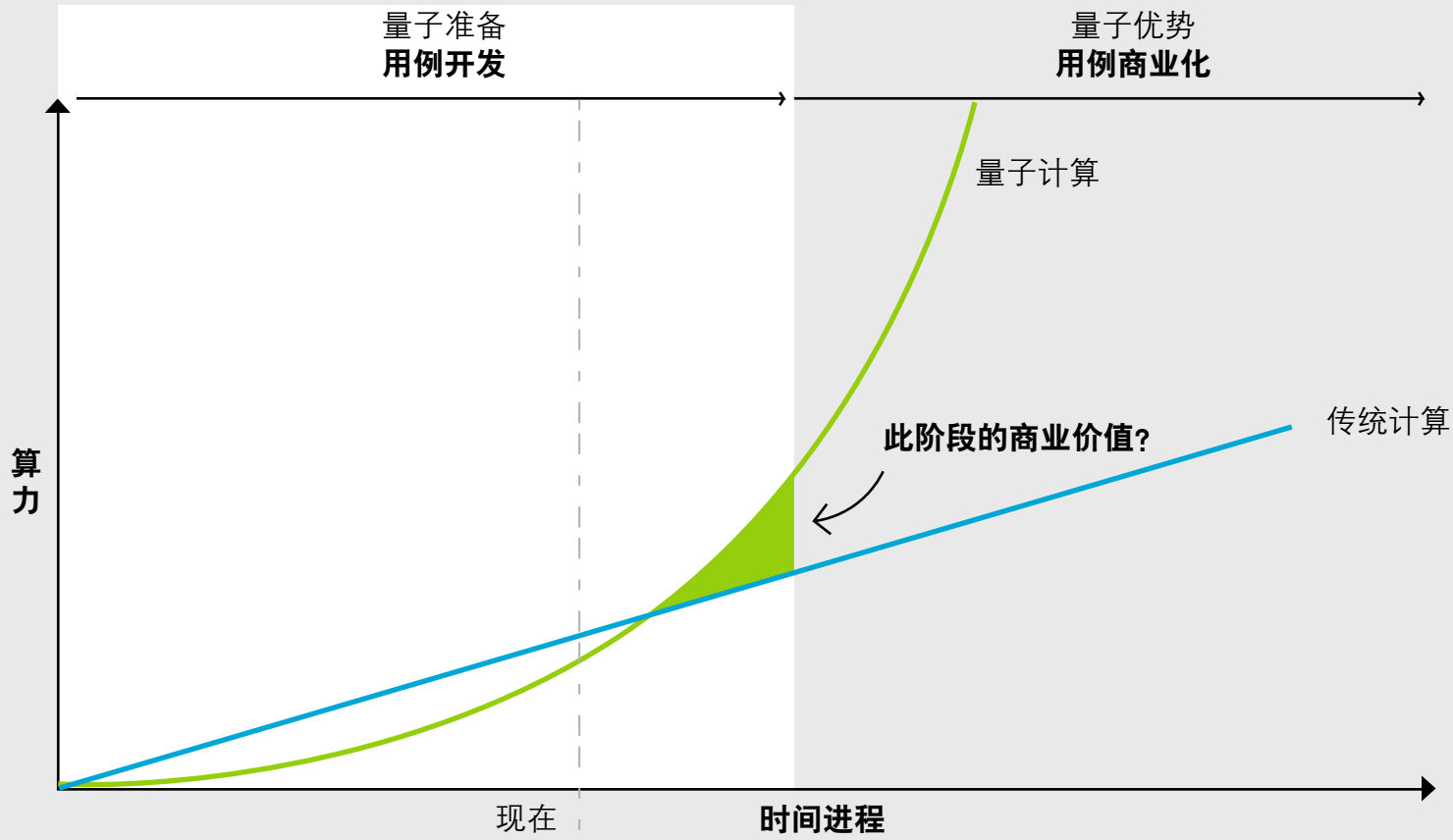
尽管研究后量子加密技术是2025年量子计算领域的最紧迫趋势，但对其他量子研究领域的探索也不应停下脚步，二者并行不悖，共同为迎接未来量子计算机带来的量子优势做好准备；此外各组织还可深入研究其他应用场景，为金融、医疗等众多行业带来积极效益。

量子技术有望推动更有效的药物发现，为金融市场或环境变化提供更精准的预测建模，为解决长期挑战带来希望。量子计算拥有推动人类进步的无限潜能，不应任由伴随而来的网络安全威胁让其蒙尘，因此要做好充分准备，开发足以抵御威胁的先进加密技术，并有条不紊、循序渐进甚至系统性地加以实施。

图3

量子计算成熟曲线

我们正在积极探究当前成果，展望明日创新突破，追寻未来潜藏机遇。



资料来源：德勤分析

Deloitte. Insights | deloitte.com/insights

By **Duncan Stewart**
Canada

Karthik Ramachandran
India

Roger Chung
China

Prashant Raman
India

Jennifer Haskel
United Kingdom

尾注

1. Investing.com, "*Earnings call: NVIDIA posts record revenue, bullish on data center growth*," Aug. 29, 2024.

2. Nitin Mittal et al., "*Now decides next: Getting real about generative AI*," Deloitte's State of Generative AI in the Enterprise Q2 report, April 2024.

3. Deloitte global analysis based on current market share of gen AI chip sales, combined with our survey data, and a US$200 billion to US$300 billion gen AI server market.

4. Jack Fritz et al., "*Battle for the enterprise edge: Providers prepare to pounce on the emerging enterprise edge computing market*," Deloitte 2023 Global TMT Predictions, November 30, 2022; Chris Arkenberg et al., "*Gaining an intelligent edge: Edge computing and intelligence could propel tech and telecom growth*," Deloitte 2021 Global TMT Predictions, December 7, 2020.

5. Sending to and from the cloud took too much time for real-time visual inspection of soft drink bottles on a high-speed line, for example, requiring AI decision-making on-prem, or at least within a few kilometers: milliseconds matter.

6. Dr. Lance B. Eliot, "*Speeding up the response time of your prompts can be accomplished via these clever prompt engineering techniques*," *Forbes*, June 12, 2024.

7. Rowan et al., "*Now decides next: Moving from potential to performance*."

8. See our related 2025 Global TMT Predictions chapter on "Gen AI, data centers, and the electricity grid." Further, see: Datacenter BMO report, Communications Infrastructure, "1Q24 data center leasing: Records are made to be broken," April 28, 2024; CBRE, *North America data center trends H2 2023*, March 6, 2024.

9. Fifty-five percent of organizations reported avoiding certain generative AI use cases because of data-related issues. Top data-related concerns include using sensitive data in models and managing data privacy and security. To read further, see: Deloitte's State of Generative AI in the Enterprise Q3 report, August 2024.

10. Organizations were much more worried about using sensitive data (for example, customer or client data) including IP. Using sensitive data in models (58% had at least a high level of concern), data privacy issues (58%), and data security issues (57%) were the top three issues that execs noted in the survey. To read further, see: Deloitte's State of Generative AI in the Enterprise Q3 report, August 2024.

11. Andy Lees et al., "*Financial services: Scaling gen AI for maximum impact*," Deloitte, accessed October 2024; Bijit Ghosh, "*The rise of small language models—efficient & customizable*," *Medium*, November 26, 2023.

12. Jana Arbanas et al., *2024 media and entertainment outlook: Generative AI*, Deloitte, 2024; Ghosh, "*The rise of small language models—efficient & customizable*."

13. See our related 2025 Global TMT Prediction on the topic of "Gen AI in consumer devices."

14. Deloitte analysis of publicly available specifications for various gen AI small servers, as well as conversations with companies globally.

15. Deloitte analysis of publicly available specifications for proposed rack-scale gen AI servers likely to be sold in the first half of 2025, although some may be available in limited quantities in late 2024.

16. Jennifer L, "*The carbon countdown: AI and its 10 billion rise in power use*," Carboncredits.com, February 28, 2024.

17. Diana Goovaerts, "*Could AI drive a new cloud repatriation wave?*," *Fierce Network*, June 6, 2024.

18. Ramesh Raskar, *Split Learning and Inference*, MIT Media Lab's Split Learning Project, accessed September 24, 2024.

19. Duncan Stewart et al., "*Telecoms tackle the generative AI data center market*," Deloitte Insights, September 2024.

20. See the related 2025 TMT Prediction chapter on "Gen AI, data centers and the electricity grid."

21. Jacqueline Davis, "*Large data centers are more efficient, analysis confirms*," Uptime Institute, February 7, 2024.

22. Duke Robertson, "*What Is a hyperscale data center? Overview and comparisons*," *Enconnex Blog*, March 27, 2023.

23. Deloitte, "*Breaking the billion-dollar barrier: Women's elite sports to generate more than $1 billion in revenue in 2024*," press release, December 1, 2023.

24. Priya Oberoi, "*Investors have their eyes on women's sports as profitably soars*," *Forbes*, June 21, 2024.

25. Elite sports is defined as the highest level of competition, which may or not be classified as "professional" sport where participants are paid for their performance.

26. Deloitte, "*Breaking the billion-dollar barrier: Women's elite sports to generate more than $1 billion in revenue in 2024*."

27. Brandon Drenon, "*The Caitlin Clark Effect has made women's basketball the hottest ticket around*," BBC News, April 5, 2024.

28. Doug Feinberg, "*WNBA announces landmark 11-year media rights deal with Disney, Amazon Prime and NBC*," AP News, July 24, 2024.

29. Alex Schiffer, "*$2M cash injection sends WNBA's worst team to league-leading valuation*," *Front Office Sports*, August 12, 2024.

30. Josh Sims, "*Las Vegas Aces valued at US$140m as average WNBA team hits US$96m*," *SportsPro*, June 18, 2024.

31. Abhimanyu Chaudhary, "*Tom Brady's WNBA investment sees dizzying 6,900% appreciation as Las Vegas Aces' valuation touches $140,000,000*," *Sportskeeda*, June 18, 2024.

32. Dee Lab, "*WNBA expansion team Golden State Valkyries breaks season-ticket record*," *Just Women's Sports*, September 17, 2024.

33. Doug Feinberg, "*WNBA awards Portland an expansion franchise that will begin play in 2026*," AP News, September 18, 2024.

34. Ibid.

35. Shawn Medow, "*Tanenbaum's KSV to shell out $50m for WNBA expansion franchise in Toronto*," *SportBusiness*, May 13, 2024.

36. Cesar Hernandez, "*NWSL announces new 4-year rights deal with ESPN, CBS, Prime and Scripps*," ESPN, November 9, 2023.

37. Ibid.

38. Meg Linehan, "*Seattle Sounders and Carlyle ownership group completes purchase of Seattle Reign in NWSL*," *The New York Times*, June 17, 2024.

39. Ibid.

40. Angel City, "*Willow Bay and Bog Iger to become Angel City's new controlling owners*," press release, July 17, 2024.

41. Kurt Badenhausen, "*NWSL team value 2024: Angel City, KC lead, average up 57% to $104M*," *Sportico*, September 25, 2024.

42. Timothy Bridge et al., "*Deloitte Football Money League 2024*," Deloitte UK, January 12, 2023.

43. Samuel Agini, "*English women's football clubs hope new league will kick start investment*," *Financial Times*, January 26, 2024.

44. Charlotte Harpur, "*Lyon women's team bought by Washington Spirit owner Michele Kang*," *The Athletic*, February 9, 2024.

45. Meg Linehan, "*Spirit owner Michele Kang joins OL Groupe to create new global women's football organization*," *The New York Times*, May 16, 2023.

46. ESPN, "*Washington Spirit owner Kang buys London City Lionesses*," December 15, 2023.

47. Chelsea Football Club, "*Chelsea Women announces strategic growth plan*," May 29, 2024.

48. Deloitte extrapolation based on 7.8 million as of Q1 2024, and 2–3 million additional subscribers for the balance of the year. Masha Abarinova, "*Fixed wireless continues to climb US broadband charts—Parks*," *Fierce Network*, June 13, 2024.

49. 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?*," *Deloitte Insights*, November 29, 2023.

50. Ziply Fiber, "*Fiber internet in rural areas: When will it be here?*," January 18, 2024.

51. Abarinova, "*Fixed wireless continues to climb US broadband charts—Parks*."

52. Gagandeep Kaur, "*Why 5G is failing to gain momentum in India*," *Light Reading*, April 25, 2024.

53. Naima Hoque Essing et al., "*Fixed wireless access: Gaining ground on wired broadband*," *Deloitte Insights*, December 1, 2021.

54. Ericsson, *Ericsson mobility report*, June 2024.

55. Nick Ludlum, "*5G home broadband continues to bring real competition to cable*," *CTIA Blog*, January 31, 2024.

56. Jolanta Stanke, "*Global broadband subscriber growth in Q4 2023 slowest since 2019*," Point Topic, April 22, 2024.

57. Andrey Popov, "*5G FWA is a game-changer for broadband services in Italy*," Opensignal, May 16, 2024.

58. Deloitte calculation based on 2024 population and household estimates.

59. Deloitte analysis of publicly reported information from wireless network operators.

60. Jericho Casper, "*Fixed wireless subscriber growth solid in Q2*," *Broadband Breakfast*, August 5, 2024

61. Bob Wallace, "*Exploring telco enterprise fixed wireless access (FWA) services*," *Network Computing*, August 1, 2024.

62. Based on 2024 publicly reported business FWA additions from the major wireless providers, combined.

63. Robert Wyrzykowski, "*5G fixed wireless access (FWA) success in the US: A roadmap for broadband success elsewhere?*," Opensignal, June 6, 2024.

64. Ibid.

65. Ibid.

66. Abarinova, "*Fixed wireless continues to climb US broadband charts—Parks*."

67. Ericsson, "*FWA is the largest 5G use case after mobile broadband*," February 2023.

68. Ericsson, *Ericsson mobility report*.

69. It was initially anticipated that at least 200 operators would have launched standalone (SA) 5G networks by the end of 2023. However, as of March 2024, only 49 operators have successfully deployed 5G SA networks.

70. GSA, "*GSA Market Snapshot March-2024*," March 2024.

71. James Kirby, "*5G standalone networks: How vendors can accelerate adoption*," Analysys Mason, September 2023.

72. Based on lead author conversations with multiple operators globally between January and October 2024.

73. Deanna Darah, "*5G NSA vs. SA: How do the deployment modes differ?*," *TechTarget*, July 25, 2024.

74. Ibid; GSMA, *The 5G guide: A reference for operators*, April 2019.

75. Darah, "*5G NSA vs. SA: How do the deployment modes differ?*"

76. Naima Hoque Essing et al., "*5G's promised land finally arrives: 5G standalone networks can transform enterprise connectivity*," *Deloitte Insights*, November 30, 2022.

77. Darah, "*5G NSA vs. SA: How do the deployment modes differ?*"

78. Ibid.

79. Philippe Poggianti and Pratik Das, "*It's time for 5G to standalone*," Qualcomm, July 6, 2023; Gavin Horn, "*6G foundry: Make the migration from 5G to 6G a rewarding experience*," Qualcomm, May 14, 2024; Roger Billings, "*What is 5G standalone? 5G SA means network slicing, security, and automation*," *Ericsson Enterprise Wireless Blog*, June 27, 2023.

80. GSMA, *The state of 5G 2024*, February 2024; Sylwia Kechiche, "*Will 5G Advanced deliver on the 5G promise?*," Opensignal, April 4, 2024.

81. Andrew Wooden, "*The telecoms industry's biggest problem? Failure to monetise 5G*," *Telecoms.com*, March 14, 2024.

82. Mike Dano, "*A deeper dive into the 5G network API opportunity*," *Light Reading*, April 1, 2024.

83. Global 6G Conference, "*Three 3GPP chairs clarify 6G standard release timeline at Global 6G Conference*," April 23, 2024.

84. *Communications Today*, "*5G investment cycle tapering off and another one not on the horizon*," March 30, 2024; Mary Lennighan, "*5G growth comes with new operator spending requirements—GSMA*," *Telecoms.com*, February 29, 2024.

85. Caroline Gabriel, "*OpenRAN: Increased collaboration and a realistic view of timing will be needed to deliver the full benefits*," Analysys Mason, April 2023.

86. Based on the authors' review of public announcements and multiple conversations with network operators globally.

87. Naima Hoque Essing et al., "*The next-generation radio access network: Open and virtualized RANs are the future of mobile networks*," *Deloitte Insights*, December 7, 2020.

88. Zineb Gdali, "*The future of mobile networks: Exploring the benefits of Open RAN 5G*," Firecell, December 15, 2023.

89. Parallel Wireless, *Everything you need to know about Open RAN*, 2020.

90. Essing et al., "*The next-generation radio access network: Open and virtualized RANs are the future of mobile networks*."

91. TeckNexus, "*Current state of Open RAN—countries & operators deploying & trialing Open RAN*," March 10, 2024.

92. O-RAN Alliance, *About* page, accessed October 2024.

93. Dan Jones, "*Dell'Oro: Don't expect 5G-Advanced to fuel a RAN resurgence*," *Fierce Network*, July 26, 2024.

94. Ray Le Maistre, "*Single vendor solutions to dominate Open RAN sales—Dell'Oro*," *TelecomTV*, February 7, 2024.

95. Ibid.

96. Ibid.

97. Iain Morris, "*Telcos doubt open RAN challengers will have a role*," *Light Reading*, March 15, 2024.

98. Dan Oliver, "*Open RAN: Everything you need to know*," *5Gradar*, July 23, 2021; Mike Dano, "*How American 5G operators learned to love open RAN*," *Light Reading*, February 12, 2024.

99. Wes Davis, "*The US government makes a $42 million bet on open cell networks*," *The Verge*, February 12, 2024.

100. Ibid.

101. Oliver, "*Open RAN: Everything you need to know*."

102. David Debrecht, "*Building a smarter 5G future through Open RAN development*," CableLabs, November 29, 2023.

103. Ibid.

104. Hosuk Lee-Makiyama and Florian Forsthuber, "*Open RAN: The technology, its politics and Europe's response*," European Centre for International Political Economy (ECIPE), October 2020.

105. Scott Buchholz et al., "*Quantum computing in 2022: Newsful, but how useful?*," *Deloitte Insights*, December 1, 2021; Duncan Stewart, "*Quantum computers: The next supercomputers, but not the next laptops*," TMT Predictions 2019, *Deloitte Insights*, 2018, pp. 96–103.

106. Deborah Golden et al., "*Preparing the trusted internet for the age of quantum computing*," *Deloitte Insights*, August 6, 2021.

107. Alex Wilkins, "*Useful quantum computers are edging closer with recent milestones*," *New Scientist*, September 30, 2024.

108. Filpe Beato et al., *Transitioning to a quantum-secure economy*, World Economic Forum, September, 2022.

109. Nancy Liu, "*Deloitte predicts 2024 will be a breakthrough year for post-quantum cryptography*," SDxCentral, December 14, 2023.

110. National Institute of Standards and Technology (NIST), "*NIST releases first 3 finalized post-quantum encryption standards*," August 13, 2024.

111. P.W. Shor, "*Algorithms for quantum computation: Discrete logarithms and factoring*," *Proceedings 35th Annual Symposium on Foundations of Computer Science* (1994): pp. 135–42.

112. Executive Office of the United States President, *Report on post-quantum cryptography*, July 2024.

113. FS-ISAC PQC Working Group, *Building cryptographic agility in the financial sector*, October 2024.

114. John Potter, "*Deloitte: Companies face harvest now, decrypt later quantum threat*," *IoT World Today*, September 27, 2022.

115. Apple Security Engineering and Architecture (SEAR), "*iMessage with PQ3: The new state of the art in quantum-secure messaging at scale*," *Apple Security Research Blog*, February 21, 2024.

116. AWS, "*Post-quantum cryptography*," accessed September 24, 2024.

117. Katherine Noyes, "*NIST's postquantum cryptography standards: 'This is the start of the race'*," Deloitte's *CIO Journal* for *The Wall Street Journal*, June 11, 2024.

118. Ibid.

致谢

Authors would like to thank **Hugo Pinto**, **Dan Li　mann**, **Jack Fritz**, **Paul Lee**, **Itan Barmes**, **Casper Stap**, **Ben Shapiro**, **Adnan Amjad**, **Mark Nace**, **Emily Mossburg**, **Deborah Golden, Joe Mariani**, **Adam Routh**, **Ankit Dhameja**, **Zoe Burton**, and **Lizzie Tantam.**

Cover image by: **Jaime Austin;** Getty Images, Adobe Stock

Deloitte.
Insights

新兴趋势：值得关注新一代技术

本章节探讨人工智能对网络安全防御的影响、用于提高半导体产量的芯粒、电信运营商采用基于云的系统进行现代化改造，以及硅光子技术在人工智能数据中心的应用。

2025年，我们推出了一系列聚焦新兴技术的短篇文章。这些技术趋势通常较我们传统的预测主题更早被采用，尽管当前市场规模相对有限，但具备显著的增长潜力，有望在未来1-2年内发展为主流技术方向。

生成式人工智能与网络安全：风险与机遇并存

网络安全专业人士深知，生成式人工智能在带来网络威胁的同时又能用于开发网络解决方案，因此正在探索如何利用生成式人工智能的力量应对新兴风险，同时帮助强化技术环境。

当前网络攻击日益高发，其中与人工智能（包括生成式人工智能）相关的网络威胁与日俱增。据2024年的一项调研显示，在美国，71%的首席信息安全官认为人工智能的威胁等级为“非常高”或“较高”。[1]人工智能正带来一系列挑战，包括监管变化、当前解决方案的有效性遭到破坏，以及对立方人工智能水平的提高，而企业加快应用人工智能又使问题进一步复杂化。

德勤预计，2024年利用生成式人工智能进行的网络攻击频发（较以往增加了一倍甚至两倍），而到2025年网络攻击频率还将持续增长。[2]生成式人工智能可用于发起多种方式的网络攻击，例如编写恶意网络钓鱼邮件。截至2024年第一季度，这类攻击较2023年同期增长了856%。[3]威胁行为者现已利用生成式人工智能工具编写恶意软件攻击代码。[4]

但生成式人工智能工具也可作为一股积极力量，帮助抵御或减少新一代人工智能带来的网络威胁。

一些网络行业从业人士担心，生成式人工智能在带来各种益处的同时，也会增加网络风险，因为其作为一种新的攻击载体，增加了攻击面。[5]不法分子可利用生成式人工智能发起多种方式的网络攻击，例如用于生成复杂、高容量的文本型网络钓鱼攻击，以及生成冒充首席执行官或其他C级高管的深度伪造图像和视频。根据调研数据显示，61%的受访企业在去年遭遇过深度伪造攻击，其中75%是冒充高管的网络攻击。[6]对此，多家生成式人工智能解决方案提供商也在设置技术防护栏，以防止其工具被用于生成此类文本和视频攻击，并嵌入数字水印，以便检测、标记或拦截生成式人工智能图像或文本（参阅《2025科技、传媒和电信行业预测》中关于水印和人工智能检测的相关讨论）。

此外，以科技行业为首的众多行业越来越广泛地使用生成式人工智能编码工具，这些工具能够提升程序员的编码效率。[7]但生成式人工智能工具创建的代码可能存在安全问题，根据2023年底的调研数据显示，半数以上 (56%) 的受访开发人员表示，代码安全问题时有发生，甚至会经常发生。[8]另外，调研还显示，部分开发人员经常绕过公司的编码工具使用政策，往往对生成代码的安全性过于自信，且并未仔细检查所有生成代码以排查安全问题。[9]尽管可能存在安全问题，生成式人工智能编码和安全大语言模型 (LLM) 也在帮助加快形成成熟、高效率和高效用的安全流程，例如安全信息和事件管理 (SIEM) 技术中的监控规则自动生成、身份和访问管理领域围绕访问工作流、配置和第三方风险管理的相关用例。[10]

目前，生成式工智能与网络安全交叉融合，监管法规和地缘政治局势随之发生变化。例如，欧盟《人工智能法案》(AI Act) 第15条对高风险人工智能系统的网络安全问题做出了明确规定。[11]此外，自2022年以来，针对已实施有关人工智能 (尤其是生成式人工智能) 的各项技术存在出口限制，例如用于生成式人工智能模型训练和推理的先进节点半导体、先进芯片制造设备以及设计工具等。[12]

中国持续加强生成式AI相关法规与监管

深度伪造是生成式AI领域面临的重大风险问题，其中具有代表性的 “Deepfake” 深度伪造技术，尽管在特定领域对社会有益，但也在经济、政治和社会等诸多领域遭到非法滥用，被世界经济论坛发布的《2024年全球风险报告》评为 “未来两年全球十大风险” 之首。未来防伪大模型将会成为中国深度伪造治理关键。防伪大模型是指利用大模型技术手段，围绕伪造、伪冒检测问题，所构建的复杂模型。相较于传统专家模型，防伪大模型在参数量以及训练数据规模方面展现出显著优势，拥有更为庞大的体量。2024年9月，国内金融领域首个针对Deepfake检测的技术规范《虚假数字人脸检测金融应用技术规范》的推出，为银行业的身份认证和交易验证场景注入了新的安全屏障。2025年，Deepfake检测技术有望在更多行业中得到广泛应用，推动数字安全技术的跨领域标准化与协同合作，成为构建全面数字安全生态的关键力量。

小结

随着生成式人工智能技术与企业的进一步融合，人工智能解决方案提供商应继续专注为最终用户提供安全的产品。不仅要确保产品本身的安全，公司在向第四方 (通常是LLM服务提供商) 提供自己或其他公司客户数据时也应保持谨慎。

随着欧盟《数字市场法案》(Digital Markets Act) [13]、《数字服务法案》(Digital Services Act) [14]和新法《人工智能法案》(AI Act) 的出台，监管变得更加复杂。[15]科技公司不仅是人工智能的领先开发者，也是广泛使用人工智能模型的部署者。因此，大众可能对科技公司抱有更高期望，认为科技公司应发挥更大和更积极的作用，确保其开发并向企业销售的产品和解决方案中实施的生成式人工智能的可信度和安全性。威胁行为者对生成式人工智能技术的使用和滥用，尤其是在风险加剧、地缘政治分裂、换届选举、战争频发等时期，或将成为2025年及以后日益重要的防御和战略考虑因素。(参阅今年关于人工智能以及公司常用工具信任度的预测文章)

硅芯“化整为零”：芯粒“续命”摩尔定律

芯粒致力于为人工智能和高性能计算环境提供更加灵活、可扩展和高效的系统，同时提高良品率。

德勤预测，基于“芯粒”（当今最先进系统级封装（SiP）的构建模块）的全球先进封装技术收入将从2021年的约70亿美元增至2025年的160亿美元，增长超过一倍。[16]与依赖印刷电路板（PCB）上独立互连芯片的传统架构相比，芯粒可实现高速数据传输，降低延迟，优化PPA（功耗、性能和面积），甚至扩展摩尔定律。[17]芯粒通常在一些高速增长的市场中进行应用和探索，如人工智能加速器（特别是生成式人工智能）、高性能计算（HPC）和电信应用。

什么是芯粒（chiplet）？

芯片与芯粒有何区别？芯片制造商一般使用直径300毫米的硅晶片（约70,000平方毫米）来制造单片晶粒（Die），其封装之后我们称之为“芯片”。高端先进芯片的尺寸一般为20毫米*20毫米（或400平方毫米），因此每个300毫米晶圆可切割晶粒数约为175个。但芯粒并非单片芯片——它是一种异构架构，其中更小的晶粒以类似单片晶粒的方式封装在一起工作。另外，这些晶粒和模块可以来自不同的芯片制造商。[18]

为何当前芯粒如此重要？芯粒早在上世纪80年代就已经出现。但直到过去四五年间，行业才开始重新关注芯粒，并开始大规模进入该领域，其中主要原因在于前沿制造节点亟需提高产量。[19]随着行业发展日益接近摩尔定律的物理极限，先进芯片的制造变得越来越具挑战性。但芯粒的出现推动了半导体的微型化发展，采用系统级封装技术（SiP）的芯片，其性能可与使用单片晶粒设计的传统系统级芯片（SoC）相媲美。[20]

芯片越小、越复杂（如5纳米和3纳米的先进制程节点），在300毫米晶圆上的缺陷率就越高，进而可能影响良率。[21]800平方毫米左右的晶粒用于制造3纳米或5纳米工艺的最先进人工智能芯片，在采用传统单片方法组装和封装时可能只有50-55%的良率，缺陷密度为每平方厘米0.1个缺陷。[22]在此情况下，成熟的90纳米和130纳米半制程节点的正常良率在90-95%左右。[23]为解决这一难题，芯粒将多个尺寸更小、良率更高的芯片集成为一个整体运行系统，即将180平方毫米尺寸（良率达95%）的小晶粒使用基于芯粒的架构进行封装，能够实现以更低的成本制造出更高效、更强大的人工智能处理器，同时提高产品/功能的灵活性和可配置性，满足不断变化的市场需求。[24]

随着芯粒应用日益广泛，业内参与者正在寻找更多创新方法来改进设计制程，提高连接速度和带宽，改善能效。例如，业界正在研究数字孪生技术，以逐步模拟和可视化复杂的设计制程，包括芯粒的移动或交换技术，以衡量和评估多芯粒系统的性能。[25]部分公司引入了一系列在芯片上组装和堆叠分立元件的互连技术，提高了传统大尺寸单片设计效率。[26]此外，玻璃是一种更灵活、可扩展的有机基板，且在热传导性和每瓦性能方面更加优越，因此业内还在持续研发，探索将玻璃用作芯粒封装基板以用于满足高性能计算和人工智能环境的需求。[27]行业甚至在探索利用光子学（利用光进行数据传输）作为提供光输入/输出（I/O）的互连解决方案，特别是用以满足HPC和AI工作负载的需求。该项技术可进行高能效和高速度的数据传输与处理（详见“硅光子”的相关预测章节）。[28]

但与此同时，芯粒架构仍面临着独特的挑战。例如，通过薄基板连接的多个晶粒堆叠会产生热管理问题，导致潜在的电路故障和功率损耗。[29]此外，随着更多的知识产权集成到这些复杂的封装中，从各地区不同供应商处采购组件或会增加网络攻击的风险，并使底层系统面临新的安全威胁。[30]

小结

为帮助释放芯粒的商业价值，半导体价值链中的行业参与者应考虑协同合作，缩小差距，共同应对挑战，探索更多增长途径：

设备制造商、代工厂、集成器件制造商、无晶圆厂 (fabless) 公司以及外包半导体组装和测试供应商可进一步加强从晶圆厂到封装产链的合作伙伴关系，推进共同研发工作。电源和逻辑集成电路制造商以及设计人员应将热管理相关的细微差别纳入考量。[31]

各业内公司还应考虑在行业早期成果（如《通用芯粒互连技术》（Universal Chiplet Interconnect Express）标准、高带宽内存协议（High Bandwidth Memory Protocol）以及线束（Bunch of Wires）互联技术）的基础之上，转向制定芯粒互连和数据互操作标准，推动早期成果的进一步发展。[32]

电子设计自动化 (EDA) 公司、芯片设计人员和安全专家可设计开发内置功能的方法，以实现在芯粒层级感知潜在的知识产权盗窃和网络侵权行为，并与供应链的其他环节合作，帮助应对可能影响芯粒的更多威胁和攻击参数。此外，设计人员还应与EDA及其他计算机辅助设计和计算机辅助工程公司合作，加强针对混合系统和复杂异构系统的设计、模拟、验证和确认工具及能力，包括将人工智能技术应用于芯片设计。[33]

业务/运营支持系统（B/OSS）：电信公司对其业务和运营支持系统进行现代化升级

电信公司的后端业务和运营软件市场增长缓慢，但通过采用SaaS、微服务架构、云迁移等方式实现其现代化升级，是目前软件供应商的增长热点，也是电信公司利用5G、光纤和人工智能拓宽业务的重要机遇。

电信公司一直拥有两套独立但重要的电信专用IT系统。分别为业务支持系统（BSS），主要用于客户订单捕获、客户关系管理和计费；以及运营支持系统（OSS），负责服务订单管理、网络库存管理和网络运营。[34]这通常是两个部署在本地的独立系统，且一般是定制的硬件定义系统，主要由一系列针对特定服务线（固定/移动）或技术领域（如接入、核心和传输）的个性化和专业化解决方案构成，形成了电信公司分散而复杂的基础设施。[35]预计在2025年及以后，电信公司将引入更高水平的自动化和智能化以对这些系统进行现代化升级，从而加快增长速度。从长远来看，BSS和OSS甚至可能整合为一个平台。

为何亟需开展现代化升级？按需获取服务和产品的市场发展要求企业重塑客户体验、重新定义产品服务、变革业务模式并重新布局销售渠道。在BSS领域，特别是计费板块，为应对不断变化的客户期望和新的数字收入流，需要开发新的功能来支持以产品和/或客户为中心的计费模式。这种影响或将贯穿整个B/OSS生命周期。

根据德勤分析师的预测，预计到2025年，OSS和BSS市场（B/OSS）的全球总收入将从2023年的630亿美元增至约700亿美元，年增长率约为5%。[36]表明电信公司期望抓住5G独立组网、光纤等带来的潜在创收新机遇。此外，遗留基础设施的维护成本（电信公司花费在集成和定制遗留B/OSS系统方面的IT预算可高达80%）[37]可能会促使电信公司加快对B/OSS软件的现代化升级。预计B/OSS的软件即服务产品的年增长率约为18%，而云迁移（又称“云化”）业务的年增长率约为21%，这表明BSS/OSS现代化升级下的各种子类型的增长速度是整个BSS/OSS软件行业5%增长率的三到四倍。[38]

随着现代化升级步伐的加快，业内可通过API和微服务等工具，并利用基于云的软件定义解决方案（这类解决方案符合现有的标准配置并可提供模块化功能）实现BSS和OSS更有效的集成。此外，系统集成有助于电信公司降本增效、创造新的收入来源，提升网络韧性并加强网络和运营安全，以及为未来利用生成式人工智能技术合并系统奠定基础。“以服务为中心”或成为发展关键，因为下一代OSS系统将在每个服务层面（而非每个技术领域层面）上对供应、执行和保证进行协调。这有助于使OSS系统的流程更加“以服务为中心”，同时对技术和支持软件进行横向整合，即将各个技术平台和支持软件集成到一个统一系统，并主要部署在云端。

许多欧洲电信公司在过去数年对B/OSS软件系统进行了现代化升级，并取得了一定的经济效益。 但未来新技术部署的目标或将以服务为中心。预计未来几年的大部分增长将来自美洲、中东和北非以及新兴的亚太地区。[40]此外，BSS系统（特别是客户管理系统）近年呈现出向云端迁移的趋势，而OSS的迁移速度则较慢，部分原因在于电信公司对将重要功能转移到新兴系统持谨慎态度，不过，这种情况似乎正在发生改变。[41]

小结

或需首先理清“由谁来提供这些新服务？”这一问题。过去，此类服务由各BSS或OSS解决方案提供商和集成商提供，或由公司在内部构建自己的解决方案。随着B/OSS的现代化升级，知名企业软件供应商和超大规模云服务商也在尝试推出相关产品。[42]他们要想取得成功，或需采用云计算/人工智能的思维模式进行现代化系统集成。

计费转型（B/OSS现代化升级的一个子类型）会产生重大影响，因为流经遗留计费系统的费额高达数十亿美元。部分电信公司高管担心开展计费转型会使该收入置于风险之中。[43]在开展计费转型的同时，需审慎地平衡这一财务收入基石、调整业务目标并最大限度减少业务中断。我们在《驾驭计费转型的复杂性》一文中，进一步探讨了应对这一难题的考虑因素和选择方案。[44]

在B/OSS的现代化升级过程中，电信公司应设法开源节流。降低成本是现代化商业案例的一个重要目标，但也应抓住网络即服务或融合产品等带来的创收机遇，有助于增强电信公司固定无线接入服务的变现能力。

此外，B/OSS现代化升级还需电信公司整合内部的OSS和BSS系统，采用行业标准应用程序编程接口（API），[45]注重研发运维（DevOps）以控制成本，并引入新兴的人工智能和机器学习（AI/ML）技术。

最后，随着电信公司寻求从相对分散的B/OSS环境转型为更加统一的模式，公司的治理模式也应随之改变。B/OSS曾是工程部门的专属范围，但在其现代化升级的各个阶段节点，应有更多的利益相关方包括人力资源、IT和财务部门参与其中。

硅光子：生成式人工智能实现光速通信

生成式人工智能要求日益提高，硅基光学器件正走出研究实验室，成为数据中心的应用焦点。

德勤预测，用于光收发器的硅光子芯片的销售额将从2023年的8亿美元增至2025年的12.5亿美元，复合年均增长率为25%。[46]相较2026年全球芯片的预期销售额6,870亿美元，该等硅光子芯片的销售额所占比重较小，[47]但与传统替代方案相比，硅光子芯片可助力生成式人工智能数据中心（与其他数据中心相比，生成式人工智能数据中心需应对海量数据的高速传输需求）实现光速通信、使用更小组件、降低成本、减少能耗、减少热量产生（热管理）。[48]

硅芯片可应用于电气领域，使用通过电线传输的电信号与其他芯片进行通信，或者需连接或结合外部激光器和调制器，利用光子通过光纤电缆传输信号。光纤的带宽通常高于铜线；光信号能够在较少的能量消耗下传输更远的距离。此外，光纤电缆不受电磁干扰，而铜线则容易受电磁干扰的影响。相比铜线，光缆通常更难被窃听或拦截，因此更加安全。不过，传统光子技术存在局限性（主要在成本和尺寸方面），而硅光子技术有望克服该等局限性。[49]

2025年，光子器件制造逐渐呈以下趋势：

- 采用与许多电子芯片相同的材料——硅。
- 采用与许多电子芯片相同材料的基板——硅。
- 采用与许多芯片相同的制造工艺。
- 采用成熟的硅光子生态系统，涵盖设计、制造、代工、测试、封装和组装流程，应与目前硅芯片制造的同类生态系统兼容。

上述趋势使得芯片公司能够在单个芯片上集成电子和光子元件。随着时间的推移，可能会对诸多不同用例产生影响。不过，2025年，预计硅光子技术主要应用于数据中心，尤其是用于运行生成式人工智能训练和推理。多数数据中心的芯片、托盘和机架之间的通信速度低于100G（每秒100千兆字节），而生成式人工智能设备则需以更快的速度传输更多数据——速度要求达到400G，甚至800G——光子技术是最佳解决方案。[50]

需了解一些必要的数据中心背景资料。生成式人工智能数据中心内有许多服务器机架。标准服务器机架宽度为24英寸（600毫米）、深度为42英寸（1066.80毫米）、高度为73.6英寸（1866.90毫米）：这就是所谓的42U服务器机柜（1U代表一个标准机架单元的高度，相当于1.75英寸【44.45毫米】）。[51]各种芯片和机架需以不同的速度在不同的距离间相互通信，通信距离和速度部分取决于机架尺寸。

因此，硅光子技术在2025年及以后应用前景广阔。不同技术有不同的“甜蜜点”，这主要取决于元件之间的距离，硅光子技术元件之间的距离有一个最佳区域值（大于10厘米且小于10米），与铜或传统光子技术相比，硅光子技术在短期内拥有更大优势，因此近期可能拥有更多收益机会。

托盘上芯片到芯片互连：生成式人工智能机架式服务器配置包含2个GPU和1个CPU托盘，托盘高度为1U或2U，具体取决于所选的冷却技术。2025年，芯片间托盘（距离小于10厘米）通信为电子通信，但随着时间的推移或将发展为光通信。由于可用空间有限（高度为2到4英寸），同时为了降低成本，可能需要使用集成硅光子技术，而非分立光子器件。不过，鉴于芯片间托盘距离极短，2025年，电信号或许已足以实现通信。

机架上托盘到托盘互连：一个服务器机架可装置18个高度为1U的托盘。这是最密集的配置。每个托盘均需与其他所有托盘进行通信，通信距离不超过垂直方向一到两米的距离。[52]据估计，到2025年初，每个机架的光学成本将达到约144,000美元。[53]到2025年下半年或2026年初，硅光子器件在这一领域的应用将逐步得到推广。

机架到机架互连，距离较近：由于各种原因（电力、冷却、成本），许多配对服务器机架的密度减半，相邻机架紧密靠拢，需在一两米的距离内进行通信。两个服务器机架之间在一定程度上几乎完全可以实现光学通信，这可能是硅光子技术在2025年迎来的最大机遇。

机架到机架互连，距离较远：在超大规模数据中心，每个服务器机架（或配对服务器机架）均需与其他所有机架服务器（以及各种存储器和处理器）进行通信：可能需要数十米甚至数百米长的光纤电缆。硅光子技术可实现高带宽和长距离传输，并且由于光子器件集成度高，可降低成本和功耗。[54]虽然成本较高是一个考虑因素，但预计硅光子技术在短期内不会取代传统光子技术在该应用领域中的地位。

关于硅光子技术的另一个预测：并购。如果生成式人工智能数据中心持续增长，尤其是对高速度和低功耗的需求（均有可能）持续增长，且硅光子技术被视为一项日益重要的新兴技术，则大型企业或将斥资数十亿美元收购在硅光子技术领域处于领先地位的硅光子技术初创企业、公司或其他公司的部门。[55]

尽管本文聚焦生成式人工智能数据中心在加速硅光子需求方面的重要性，但必须指出的是，硅光子技术在其他领域亦有潜在的应用前景。近期最显著的应用机会是为高级驾驶辅助系统（近期）和自动驾驶功能（长期）制造片上激光雷达装置。[56]

By

Duncan Stewart
Canada

Karthik Ramachandran
India

Prashant Raman
India

Roger Chung
China

尾注

1. Srini Subramanian and Meredith Ward, "*2024 Deloitte-NASCIO Cybersecurity Study*," *Deloitte Insights*, Sept. 30, 2024.

2. The Deloitte authors make this prediction based on what they are seeing in the market and what their clients are telling them.

3. Duncan Riley, "*Generative AI services have driven a huge surge in phishing attacks*," *Silicon Angle*, May 22, 2024.

4. Michael Crider, "*Hackers are now using AI-generated code for malware attacks*," *PCWorld*, Sept. 25, 2024.

5. Tiernan Ray, "*Generative AI is new attack vector endangering enterprises, says CrowdStrike CTO*," *ZDNet*, June 30, 2024.

6. Ian Barker, op. cit.

7. Faruk Muratovic, Duncan Stewart, and Prashant Raman, "*Tech companies lead the way on generative AI: Does code deserve the credit?*," *Deloitte Insights*, Aug. 2, 2024.

8. Snyk, *2023 Snyk AI-generated code security report*, accessed Aug. 11, 2024.

9. Ibid.

10. Mandy Andress, "*Generative AI for cybersecurity: Is it right for your organization?*," *Fast Company*, June 17, 2024.

11. EU Artificial Intelligence Act, "*Article 15: Accuracy, Robustness and Cybersecurity*," accessed Aug. 28, 2024.

12. Christie Simons et al., *2024 global semiconductor outlook*, Deloitte, Jan. 22, 2024.

13. European Commission, "*The Digital Markets Act: Ensuring fair and open digital markets*," accessed Oct. 6, 2024.

14. European Commission, "*The Digital Services Act*," accessed Oct. 6, 2024.

15. European Commission, "*AI Act*," accessed Oct. 6, 2024.

16. Deloitte analysis based on data and chart presented in "*Semiconductor – A treasure trove for private equity investors*,"(June 2024, p. 11). We used chiplet packaging baseline market share for 2021 (24% of the US$30 billion total market) and applied 22% CAGR to arrive at the 2025 predicted value of US$16 billion.

17. Moore's Law notes that the number of transistors on an integrated circuit would double every two years but with a smaller increase in cost—translating into nearly twice the superior performance over the previous generation (because of doubling of transistors by shrinking the linewidths) at a marginal additional cost. However, shrinking transistor linewidths is reaching its physical limit. To read further, see: AIchip's "*Moving from SoCs to chiplets could help extend Moore's Law*," Sept. 26, 2022.

18. Deloitte's analysis of multiple publicly available sources including product information published by chip companies, as well as articles from sources such as *EE Journal*, *Semiconductor Engineering*, *EE Times*.

19. Based on our analysis of chiplet-based new product announcements and launches from major semiconductor companies (including IDMs, fabless, and chip design players). Chiplets allow multiple functionalities such as GPU, CPU, and memory components to be densely packed on a single chip. Moreover, chiplets have helped deal with the complexity involved in integrating the diverse components with varying manufacturing and packaging technologies coming in from IDMs, foundries, and other component manufacturers from various regions worldwide. To read further, see: Dr. Uwe Lambrette et al., "*Semiconductor – A treasure trove for private equity investors*," Deloitte, June 2024.

20. *TrendForce*, "*[News] Understanding 3DIC, heterogeneous integration, SiP, and chiplets at once*," March 19, 2024; AIchip, "*Moving from SoCs to chiplets could help extend Moore's Law*."

21. Max Maxfield, "*Are you ready for the chiplet age?*," *EE Journal*, July 27, 2023.

22. Yinxiao Feng and Kaisheng Ma, "*Chiplet actuary: A quantitative cost model and multi-chiplet architecture exploration*," Institute for Interdisciplinary Information Sciences (Tsinghua University, China), April 9, 2024.

23. "*Test & reliability challenges in advance semiconductor geometries*" (presentation at 2013 Semiconductor Wafer Test Conference), June 9, 2013. Data for 130 nm and 90 nm based on the chart on page 22 titled "Dramatic rise in systematic yield issues."

24. Deloitte analysis based on our conversations with subject matter experts in the areas of advanced packaging, as well as data and research from publicly available sources, including Feng et al., "*Chiplet actuary: A quantitative cost model and multi-chiplet architecture exploration*"; Maxfield, "*Are you ready for the chiplet age?*"

25. Ann Mutschler, "*Digital twins gaining traction in complex designs*," *Semiconductor Engineering*, June 27, 2024.

26. Eric Beyne, "*Chiplet interconnect technology: Piecing together the next generation of chips*," *3D InCites*, July 3, 2024.

27. Bilal Hachemi, "*Glass Core substrates: The new race for advanced packaging giants*," Yole Group, June 17, 2024; Anton Shilov, "*Intel's glass substrates advancements could revolutionize multi-chiplet packages*," *Tom's Hardware*, Sept. 18, 2023.

28. See section Silicon photonics: Gen AI communicates at light speed.

29. Karen Heyman, "*Thermal challenges multiply in automotive, embedded devices,*" *Semiconductor Engineering*, July 2, 2024.

30. Saman Sadr and Richard Lin, "*Securing the new frontier: Chiplets & hardware security challenges,*" Universal Chiplet Interconnect Express, Feb. 7, 2024; Nitin Dahad, "*Chiplets are the latest buzz, but many challenges lie ahead*," *Embedded*, March 10, 2024.

31. Thermal and heat management are noted as one of the major roadblocks to commercializing 3D ICs. To read further, see: Brian Bailey, "*Why there are still no commercial 3D-ICs*," *Semiconductor Engineering*, Jan. 29, 2024.

32. As noted in *Deloitte Global's 2024 semiconductor industry outlook*, not only the traditional OSATs but even major IDMs, foundries, fabless companies, EDA vendors, and startups are making the moves and ramping up solutions based on chiplets architectures to push the bar on advanced packaging technologies. Also, see: Ann Mutschler, "*Chiplet IP standards are just the beginning*," *Semiconductor Engineering*, March 6, 2024; Majeed Ahmad, "*A sneak peek at chiplet standards*," EDN, Sept. 4, 2023.

33. Ann Mutschler, "*Chip design digs deeper into AI*," *Semiconductor Engineering*, June 3, 2024.

34. Andrew Wooden, "*The evolution of BSS and OSS in the telecoms sector*," *Telecoms.com*, Aug. 15, 2023.

35. Ibid.

36. Deloitte analysis, based on Alex Bilyi, "*CSPs' spending on telecoms-related OSS/BSS software and services will reach USD80 billion by 2028*," Analysys Mason, Nov. 13, 2023.

37. Nia Batten, "*The hidden costs of legacy tech*," *Data Centre Review*, Sept. 1, 2023.

38. Deloitte analysis, based on Alex Bilyi, "*CSPs' spending on telecoms-related OSS/BSS software and services will reach USD80 billion by 2028*."

39. Chris Silberberg and Chris Barnard, "*How telcos are transforming in Europe: Technology, services and customers*," IDC, Sept. 1, 2022.

40. Deloitte analysis, based on Alex Bilyi, "*CSPs' spending on telecoms-related OSS/BSS software and services will reach USD80 billion by 2028*."

41. Mark Mortensen, Andy He, and John Abraham, *Market pulse: Digital transformation of BSS/OSS to the cloud & DevOps*, Analysys Mason, Jan. 2018.

42. Ryan, "*OSS/BSS in the clouds*," Passionate about OSS, July 20, 2020; Anjali Mishra, "*OSS/BSS market players are building robust systems to support next-gen networks*," Global Market Insights (GMI), May 6, 2022.

43. Amit Kumar Singh et al., "*Navigating the complexities of billing transformation*," *Deloitte Insights*, 2024.

44. Ibid.

45. TM Forum, "*Introduction to Open APIs*," accessed Sept. 24, 2024; GSMA, "*GSMA Open Gateway API descriptions*

46. Deloitte analysis and interpolation of *Light Trends Newsletter*, "*Sales of silicon photonics chips will reach $3 billion by 2029*," LightCounting, May 2024.

47. WSTS, "*WSTS Semiconductor Market Forecast Spring 2024*," press release, June 4, 2024.

48. Adam Carter, "*Silicon photonics key to unlocking AI's full potential*," *EE Times*, Aug. 18, 2023.

49. Deloitte analysis based on publicly available third-party sources, including Karen Heyman, "*Transitioning to photonics*," *Semiconductor Engineering*, April 13, 2023; Maxime Fazilleau, "*What makes optical fibre immune to EMI?*," Tiny Green PC, Jan. 23, 2017.

50. FiberStamp, "*Driving the future of high-speed data transfer: The role of PAM4 and silicon photonics in the age of AI*," *Medium*, Nov. 8, 2023.

51. Christopher Tozzi, "*A guide to server rack sizes for data centers*," Data Center Knowledge, Jan. 8, 2024.

52. Mary Zhang, "*Data center racks, cabinets, and cages: An in-depth guide*," *Dgtl Infra*, Sept. 28, 2023; Tozzi, "*A guide to server rack sizes for data centers*."

53. Dylan Patel and Daniel Nishball, "*Nvidia's optical boogeyman – NVL72, Infiniband Scale Out, 800G & 1.6T Ramp*," *SemiAnalysis*, March 25, 2024.

54. M. Duranton, D. Dutoit, and S. Menezo, "*3 - Key requirements for optical interconnects within data centers*," in *Optical Interconnects for Data Centers*, Tolga Tekin et al. (eds) (Sawston, UK: Woodhead Publishing, 2017), pp. 75–94.

55. Contributions from Deloitte subject matter specialists in July and August 2024.

56. Eric Walz, "*Stellantis invests in lidar startup SteerLight*," *Automotive Dive*, April 2, 2024.

致谢

The authors would like to thank **Jack Fritz, John Levis, Stephen Winsor, Sandy Lawrence-Morgan, Essaki Velusami, Kannan Ramakrishnan, Nina Zhang, Gautham Du** , and **Dan Hamling** for their contributions to this article.

Cover image by: **Jaime Austin; Getty Images, 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合伙人
电邮：leoschen@deloittecn.com.cn

濮清璐
科技、传媒和电信行业华东区主管合伙人
电邮：qlpu@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

胡新春
科技、传媒和电信行业咨询业务合伙人
电邮：tonyhu@deloittecn.com.cn

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

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

谢似君
科技行业主管合伙人
科技、传媒和电信行业咨询业务合伙人
电邮：trxie@deloittecn.com.cn

王易
体育行业主管合伙人
电邮： cryswang@deloittecn.com.cn

张森
科技、传媒和电信行业华北区主管合伙人
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

叶勤华
科技、传媒和电信行业审计合伙人
电邮：jiip@deloittecn.com.cn

张耀
电信行业执行总裁
电邮：yaozhang@deloittecn.com.cn

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

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。

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。

CQ-001CN-25