

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受访美国受访者中，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 and 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

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 Fellingner, “*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 Hodgson, “[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

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。

© 2025。Deloitte Development LLC版权所有保留一切权利。

CQ-001CN-25