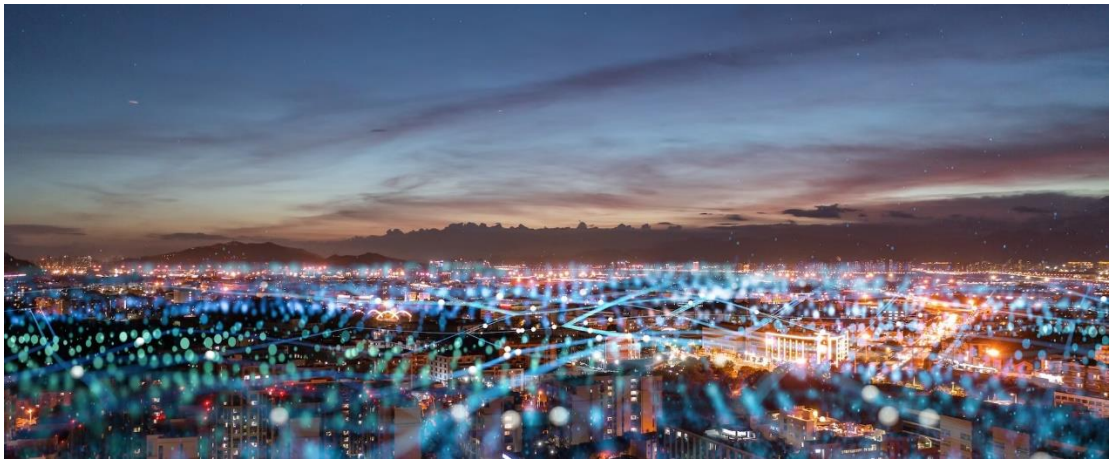

AI 边缘的网络机会

我们正在迅速迈向一个高度互联的数字世界，人工智能可以帮助我们以创新的方式管理、货币化和物化数据。本文重点介绍运营商在帮助经济高效且可靠地交付 AI 应用方面可以发挥的作用以及可以解锁的增收机会。

AI 的赢利点藏在哪里?

AI 技术正在快速发展，但在许多方面仍处于初期。苹果、阿里巴巴、亚马逊、谷歌、Meta 和 Microsoft 等超大规模企业正在投资数十亿美元来构建他们训练 ChatGPT、Gemini、Llama 和 Qwen 等大型语言模型所需的大型数据中心（和发电厂）。这就是对通用人工智能（AGI）的追求：构建更智能的数字代理，可以反映人类的认知能力，例如阅读、写作、听力、语音、学习、推理、操作机器和执行复杂任务。

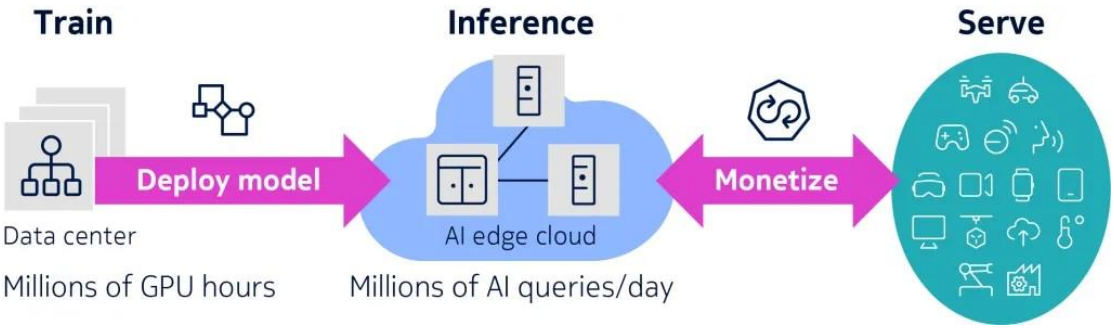

图 1.AI 推理对于从 AI 训练的投资中获得盈利至关重要

虽然 AGI 对于人机交互以及 AI 聊天机器人和虚拟助手等应用至关重要，但最大的回报将来自无数的应用，这些应用将在所谓的“推理”中将这些预先训练的 AI 模型用于特定任务、功能和数据查询。

照片编辑和智能家居自动化等消费类应用越来越依赖 AI 来解释私人数据并推断关键决策。商业、金融、医疗保健、制造、运输和公共安全领域的众多应用也是如此。这些推理应用偏爱更专业或聚焦的 AI 模型，其针对特定任务、环境和数据集进行了优化。这些聚焦的模型所占用的资源较小，在某些情况下甚至可以独立于云进行运行。

弥合与云的差距

目前，AI 推理逻辑要么驻留在数据中心，要么驻留在用户设备或本地。在用户设备和数据中心之间来回传输数据需要时间和金钱，并且会带来风险。除了能源和空间限制造成的实际扩展限制外，我们不希望数据中心变得大到不能停止服务。在用户方面，有数十亿台高度多样化且广泛分布的设备以及数以千计的组织可以从 AI 中受益。但是，这些组织可能并不总是拥有必要的硬件或 IT 资源，或者某些应用或功能（即混合云和拆分推理）仍必须依赖外部 AI 计算和存储资源。

在网络边缘、集中式数据中心和用户设备之间托管 AI 推理工作负载，将弥合这一差距并解决以下挑战：

- 通过将数百万次日常用户交互的负载分配到大量地理冗余的 AI 边缘计算服务器上，减少对集中式数据中心的依赖。
- 通过使 AI 推理逻辑更接近用户设备和数据源，降低网络带宽成本和延迟（数据包往返时间）。
- 通过将数据流量保持在单个托管网络的边界内，减少网络拥塞、网络安全和数据隐私或主权风险。

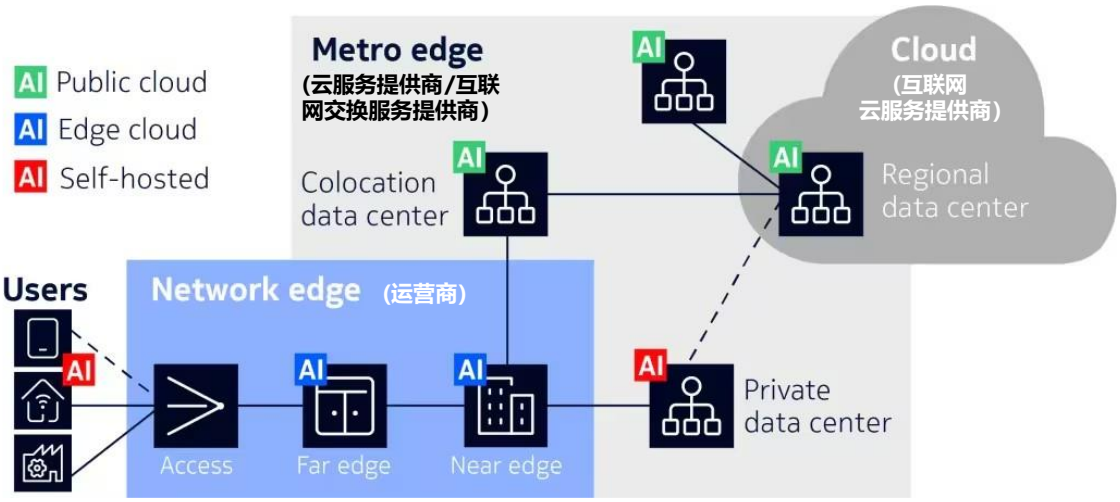

图 2.通过 AI 边缘推理弥合用户与云之间的差距

与必须提前仔细规划和确定规模以满足预期需求的核心数据中心不同，构建 AI 边缘云在很大程度上可以由需求驱动。此外，AI 边缘计算可以通过预处理和管理原始数据输入来卸载数据中心和网络流量。但问题是，谁来建造和运营它？

处于超级互联世界的边缘

互联网云服务提供商无法自行轻松弥合其云数据中心和最终用户之间的差距。他们拥有技术，但最典型的是依赖 Equinix 等数据中心主机托管交换提供商来托管其服务器设备，以便将其云业务扩展到大型大都市和城市。由于数据主权法以及在全球范围内收购、配备、运营和维护合适的边缘站点所面临的无法克服的成本和物流挑战，再走得更远将是一个太远的桥梁。

运营商实际上位于这个超级连接的数字世界的边缘。他们可以利用其本地业务、房地产资产、网络基础设施和专业服务，使云构建商和数字基础设施提供商能够横向扩展 AI 边缘云。运营商可以托管 AI 服务器基础设施，供云构建者、数字基础设施合作伙伴、大型企业和他们自己的私人使用。

根据他们的能力和舒适区，运营商可能会考虑支持一系列增值服务，例如，

- 提供通过光纤接入的主机托管服务给模块化数据中心，其所在的商业园区、购物中心、机场、医院和其它商业聚集地需要本地 AI 计算服务。
- 出租地面空间或机架空间，用于在运营商机房托管 AI 计算服务器，并提供电源、冷却、安装和维护服务、安全（量子安全）光纤连接和虚拟专用网络（VPN）服务。
- 代表批发合作伙伴和客户以及 AI RAN 和 AI NetOps 等内部电信应用提供 AI 计算和软件托管服务。

通过 AI 边缘推理服务推动收入增长

AI 边缘推理为运营商提供了巨大的增长机会。连接是 AI 时代的关键推动因素，但仅凭这一点不足以释放其对超级连接数字世界的全部价值。积极参与构建具有增值托管服务的 AI 边缘云将使运营商能够更有效地将从连接服务中盈利。它使运营商能够从“管道”转变为增值服务的促进者，从而创造新的收入来源、增强客户体验并激发忠诚度。