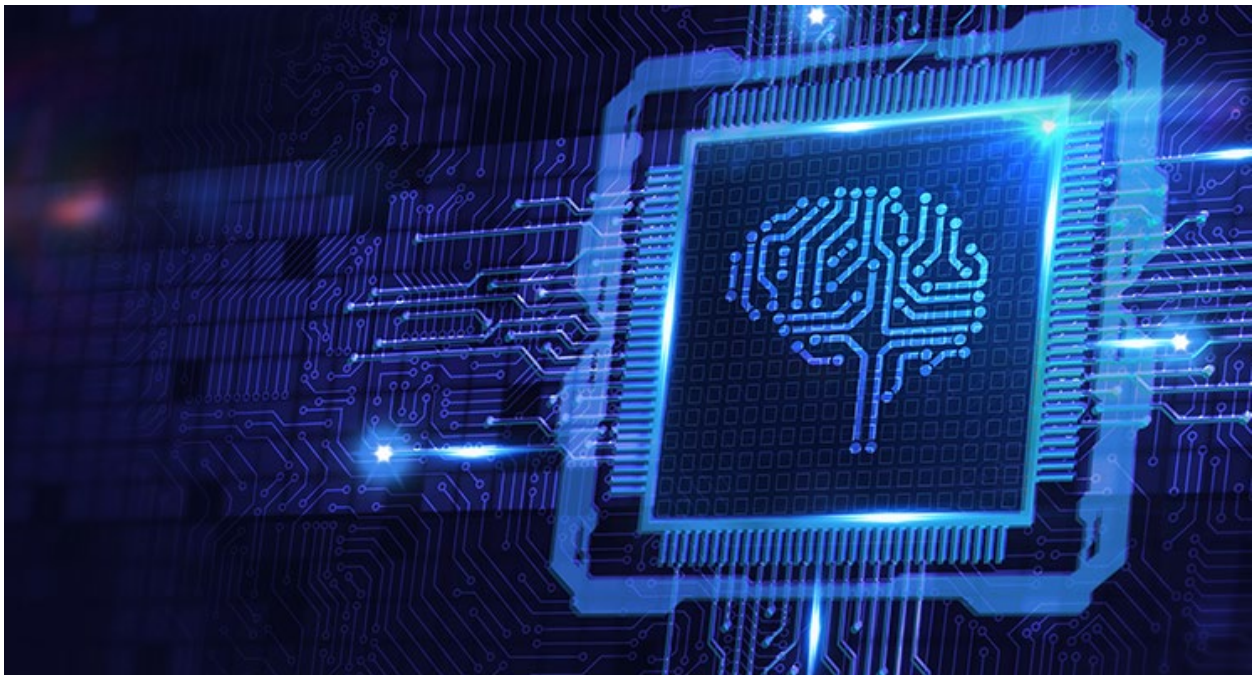


AI 将如何重新定义数据中心？

Microchip 总裁兼首席执行官

Ganesh Moorthy

AI 已经存在了一段时间，其影响也相当巨大。生成式 AI 刚开始崭露头角，对于其如何颠覆世界的种种预言已经迅速成为热门话题，影响之深广，令人深思。



这项技术已经对数据中心产生了重大影响。基于必须在云端训练和运行的模型的大小，生成式 AI 的基础设施需求预计将比早期 AI 模型高出 10 到 100 倍。事实上，所有数据中心基础设施都受到这一趋势的影响，包括电力、HVAC、网络 and 物理布局。数据中心将如何扩展规模以适应呈指数级增长的生成式 AI 大型语言模型？

生成式 AI 的成功可能取决于半导体行业对以下几个方面的关注，一是先进的 AI 处理器，二是如何存储和处理所有这些数据，最后是确保数据中心散热更佳、更高效且更安全。

我们已经解决了这类问题，以便数据中心运营商能够为 5G、云服务和新一代的 AI/ML 应用提供支持。我们的单片机、存储器、PCIe®开关、安全解决方案和其他产品提

供了各种解决方案，以应对从连接 AI 处理卡到优化数据访问、存储、安全和保护等一系列 AI 挑战，同时提供必要的带宽、端口密度和连接能力，从而处理不断增加的流量。

类似的挑战也适用于数据中心和本地服务器中的生成式 AI，其应对难度也进一步放大。生成式 AI 似乎也存在一种途径，能够扩展到物联网应用中的网络边缘，如患者监测和汽车交通违规检测。有些人认为，如果生成式 AI 产生的海量数据必须在云端而不是在边缘处理，那么这些应用便无法取得成功。

如今的边缘 AI 解决方案可能为生成式 AI 提供了一种路线图。语音识别系统和 AI 摄像头等计算机视觉产品可在边缘提供许多优势，包括消除网络延迟和将隐私控制权交还给用户。它们有助于降低功耗，因为只有在检测到不规则的变化后才会将视频流传输到云端，而不是像“傻瓜”摄像头那样连续不断地传输。可以预计，未来的边缘设备也会采用这种方案。

但是，将生成式 AI 引入到边缘将比引入传统 AI 更加困难。由于数据量更大，电力问题将变得更具挑战性。硬件需求将远远超过传统 AI 所需，并且解决方案还需要为智能边缘的电池供电、深度嵌入式设备提供必要的安全性和其他独特功能。

我们已经关注到这些挑战。例如，我们正在探索如何利用当前的 Wi-Fi®设备实现串行端口到生成式 AI 的桥接。在这种情况下，生成式 AI 并非在边缘实现，而是通过 8 位单片机的 UART 访问。单片机会产生一个实时传感器数据穿插其中的预定义提示语，将其发送到生成式 AI 解决方案，然后根据回答采取行动。目前，尚不明确端点能否在边缘独立运行生成式 AI 而不回调至数据中心，但我们正处于创新的早期阶段。

与我们服务的任何市场一样，Microchip 的目标是尽可能简化解行业一些最基本的系统设计挑战。这包括提供丰富的产品组合，以便客户能够采用全面系统解决方案的方法实现创新。我们会帮助解决在需要新功能的成熟应用中遇到的挑战，同时也为长期的安全性、保护性和可靠性要求提供支持。针对生成式 AI 等最新的新兴技术，我们也会采取同样的做法，这些技术描绘了令人期待的未来，但在兑现承诺之前，这些技术仍然面临着相同的基本系统设计挑战。