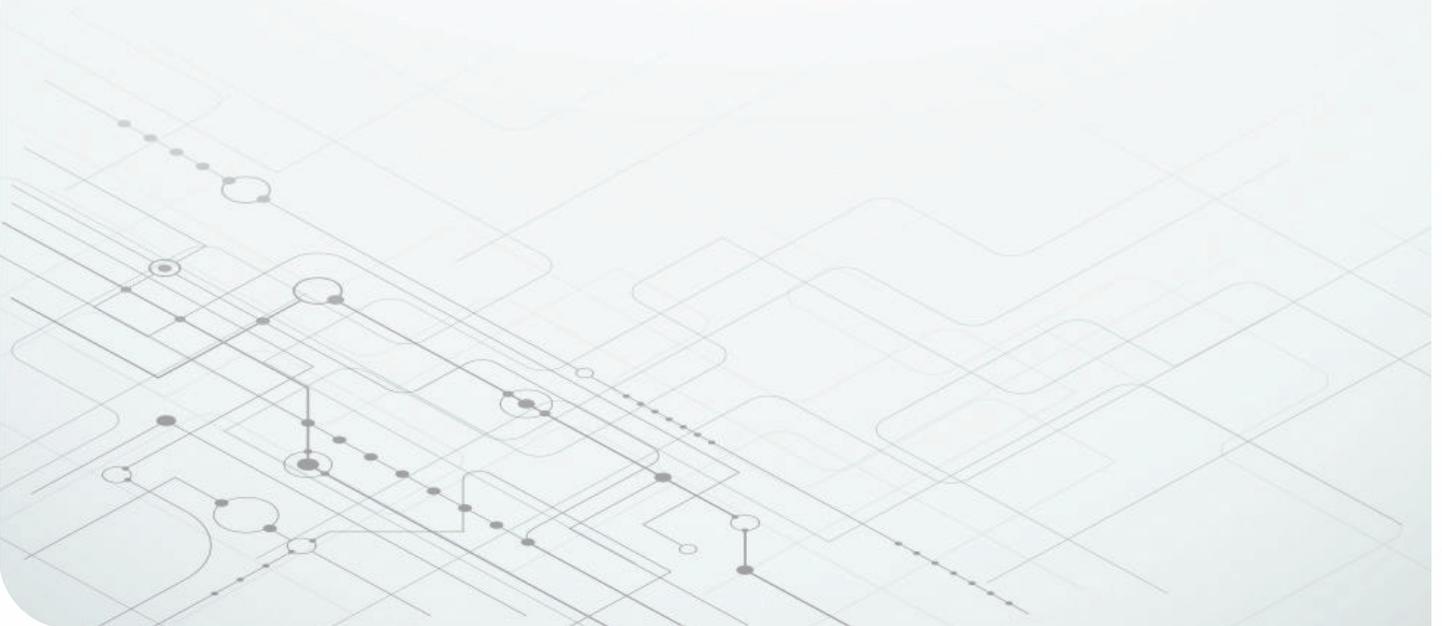


# 利用嵌入式处理器提高边缘智能化程度



**Alec May**  
Systems Manager  
Jacinto™ High-Performance Computing Processors



# 内容概览

## 1 简介

本白皮书探讨了边缘人工智能 (AI) 的用途和优势，以及嵌入式处理器和软件的进步如何使 AI 比以往更容易在各种应用中实现。

## 2 边缘 AI 有哪些优势？

了解边缘 AI 的优势及其在电子设备中启用本地推理的能力。

## 3 AI 如何向边缘发展

了解嵌入式硬件和软件工程师面临的挑战，以及 TI 如何应对这些挑战。

## 4 边缘 AI 的可扩展性

探索 TI 为应对可扩展性和可重用性挑战而提供的硬件和软件。

### 简介

基于云的 AI 解决方案的广泛可用性和用户友好性使几乎所有人都能更容易地使用为 AI 设计的模型和工具。

但是，并非所有 AI 创新都在云中发生。随着嵌入式处理器设计的技术进步，AI 功能正在逐渐应用于笔记本电脑和手机等消费类产品以及其他产品：电池供电应用，如可视门铃、汽车系统中的视觉处理以及能源基础设施和工业系统的电机。

边缘 AI 能够在本地（靠近数据源）运行 AI 模型，从而提高产品的响应能力、效率、可靠性和安全性。实现这种云到边缘转型的嵌入式处理器集成了用于数字信号处理 (DSP) 的专用内核等元件，并由基于 GUI 的易用工具提供支持，从而更大限度地减少将 AI 推向边缘所需的时间和专业知识。

在本白皮书中，我将探讨边缘 AI 的发展和优势，以及实现边缘 AI 的硬件和软件方面的进步。

### 什么是 AI？

如今，当大多数人想到 AI 时，他们通常会想象文字和图像生成器。但在字面意义上，即使是最简单的算法，在技术上也是 AI 的示例。

AI 的广泛性及其多种用例已经产生多个子领域，包括机器学习和深度学习，如 **图 1** 所示。

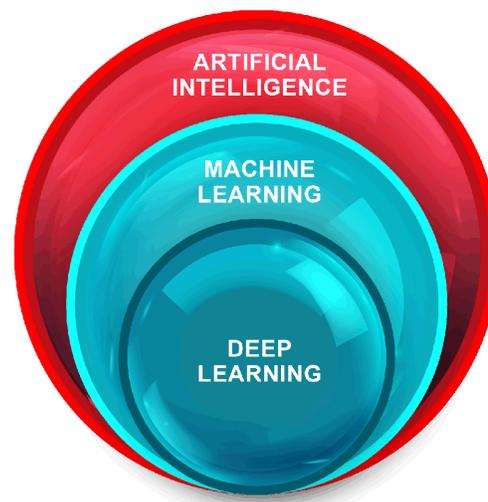


图 1. 不同 AI 子领域之间的关系。

嵌入式应用中使用的大多数 AI 都是机器学习，即机器和算法“学习”如何通过数据解决问题的子领域；例如，车辆通过分析图像数据中的常见模式来识别行人和障碍物。机器学习模型从训练数据中学习，训练数据可以用基础真实信息（即经过验证的准确数据）标记，以更好地识别要从中学习的模式。通过这种训练过程，机器学习模型可以辨别数据中的模式，并据此做出未来的推断。

在机器学习领域，深度学习虽然需要大量的计算资源，但由于能够准确解决高度复杂的问题，因此已经成为非常受欢迎的实现方式之一。深度学习使用多层神经网络，这是一种受人脑神经元启发的数据模型。神经网络使开发人员能够解决模式过于复杂而无法辨别或编写自定义规则的问题。

### 边缘 AI 有哪些优势？

AI 及其子领域通常可以在云或本地服务器上执行处理。基于云的 AI 历来较为常见，因为在大型服务器之外，要实现

有影响力的 AI 所需的计算能力并不容易。然而，随着嵌入式处理器的计算能力和能效的提高，边缘 AI 越来越受欢迎。

图 2 展示了边缘 AI 和云 AI 在接收和处理数据以及与基于云的资源交互方面的差异。

在开发过程中，边缘 AI 通常使用云或桌面资源来训练模型。将模型部署到嵌入式器件后，就可以在本地独立地使用新数据做出模型推断和决策。

直到最近，最有意义的 AI 示例才要求具备普通消费类电子产品所能提供的处理能力。这意味着机器学习模型通常在基于云的资源上进行训练和实施。虽然基于云的实施通过尽可能减少硬件投资提供了便利性，但也限制了 AI 的采用。基于云的 AI 实现无法在没有云访问权限（即网络连接）的任何应用中使用。此外，与基于云的 AI 相比，边缘 AI 还可以提高安全性和响应能力。

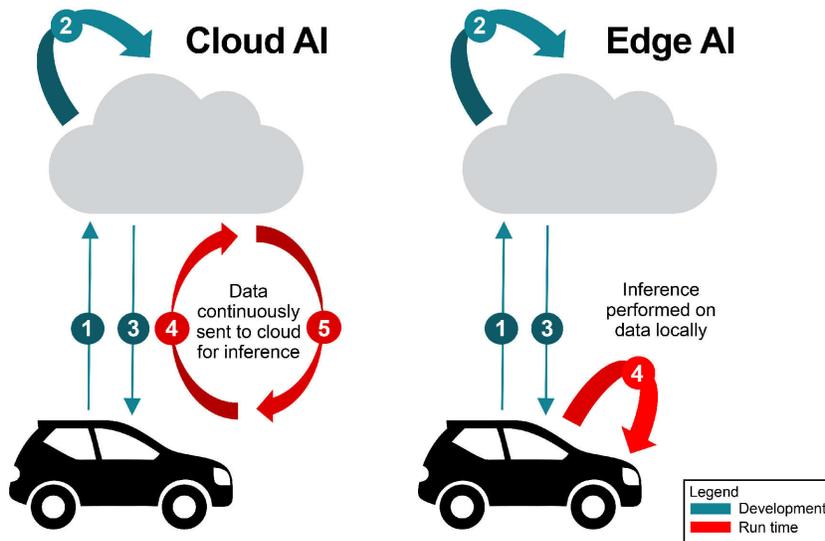


图 2. 基于云的 AI 和边缘 AI 的比较。

随着半导体技术的进步和 AI 工具链的改进，直接在嵌入式处理器和微控制器 (MCU) 中实现 AI 解决方案将 AI 推向边缘。将 AI 推向边缘意味着在传感器提供数据的位置执行计算和 AI 推理，考虑到电子产品收集的传感器数据不断增加，这一点非常重要。不断增长的数据量使云资源变得不那么实用，因为将大量数据传输到云端以及从云端传输到设备可能成本高昂、过程复杂，而且会出现单点故障。

在网络边缘运行 AI 模型通常可以减少基于传感器数据进行推理和决策的延迟；例如，车辆中用于碰撞检测的摄像头传感器。借助边缘 AI 功能，车辆可以更快地进行推理、实时响应刺激，而无需等待来自云的推理。通过物理 AI（如图 3 所示）将此本地推理转化为物理 AI 操作。这个术语指的是能够通过实际的行动来感应、解释和响应的系统，例如移动工厂车间箱子的机器人或自动刹车的车辆。



图 3. 人形机器人中边缘 AI 和物理 AI 的简单比较

这些系统可以使用边缘 AI 将高级感知与机械驱动相结合，使机器能够与人类一起安全协作。

与基于云的 AI 相比，边缘 AI 还有其他几个优势，包括减少对网络连接的依赖。边缘 AI 可用于无法访问云的应用，并尽可能缩短因网络中断而导致的潜在停机时间。此外，由于基于云的 AI 需要网络连接，因此访问可能会产生经常性的服务费用，这在设计消费类产品时可能是一种具有挑战性的商业模式。

## AI 如何向边缘发展

嵌入式处理器的处理能力和功耗限制，以及高级的内部编程专业知识和资源，限制了边缘 AI 的广泛应用。能够满足 AI 计算性能要求的嵌入式器件往往尺寸过大、功耗过高、发热过多。

出现了专用的硬件解决方案，能够更好地加速实现边缘 AI 所需的计算操作，但一些权衡因素限制了它们在边缘应用中的进一步广泛采用。图形处理单元 (GPU)、现场可编程门阵列 (FPGA) 和应用特定集成电路 (ASIC) 等专用硬件解决方案实现了出色的性能。不过，这些解决方案通常会受到限制，要么是高功耗（尤其是 GPU 和 FPGA），要么是灵活性有限 (ASIC)。

集成神经处理单元 (NPU) 已成为一种将 AI 功能直接引入嵌入式系统的解决方案。与通用处理器不同，NPU 专门用于执行构成现代神经网络主干的矩阵乘法、卷积运算和激活功能。通过从主 CPU 分流计算密集型任务，NPU 极大地提高了推理速度，同时降低了功耗，这是边缘部署的两项基本要求。

为了解这些元件的影响，我们来看一下德州仪器 (TI) 为边缘 AI 市场不同领域设计的两款 NPU：适用于高性能应用的 C7™ NPU 和适用于低功耗、可扩展器件的 TinyEngine™ NPU。

### C7 NPU

C7 NPU 是一款集成到 TDA54-Q1 和 TDA4VE-Q1 片上系统 (SoC) 中的高性能、高效 AI 加速器。为了满足视觉应用的计算需求，TI 微处理器和 SoC 提供了适用于多代处理器的 C7 NPU 版本。此 NPU 来源于 TI 的长期 DSP，能够在不牺牲成本或功耗的情况下高效提供 AI 解决方案。

C7 NPU 还使 TI 处理器能够处理多个并发 AI 工作负载，这对于需要同时处理来自摄像头、雷达、激光雷达和其他传感器的数据的系统来说是一项重要功能。

TDA54-Q1 使用 C7 NPU 在高级驾驶辅助系统、信息娱乐系统和机器人中实现边缘 AI。

### TinyEngine NPU

TinyEngine NPU（如图 4 的简化方框图所示）是一款适用于 MCU 的专用硬件加速器，可优化深度学习推理操作，从而在资源受限器件（包括电池供电产品）中处理 AI 工作负载时降低延迟和功耗。

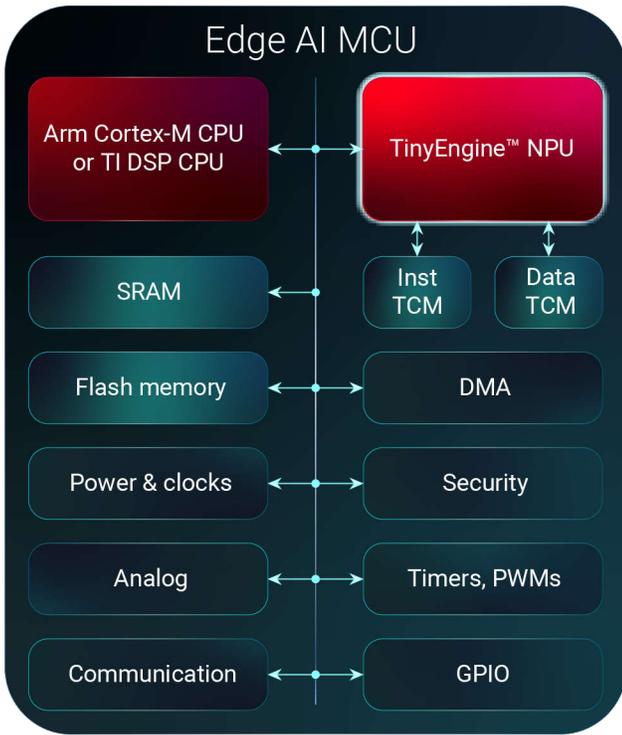


图 4. 集成 TinyEngine NPU 的 TI 边缘 AI MCU 简化框图

此 NPU 执行机器学习算法，与运行应用代码的主 CPU 并行。相较于未配置加速器的同类 MCU，配备 TinyEngine NPU 的 MCU 在运行 AI 模型时，每次推理的延迟最多降低至 1/90，能耗最多减少至 1/120。

**TMS320F28P550SJ** C2000™ MCU 使用 TinyEngine NPU 进行电机轴承和太阳能电弧故障检测，让主 CPU 处理实时电机控制。**AM13E23019** 将 TinyEngine NPU 与先进的实时控制架构（集成多达四个电机）相结合，适用于电器、机器人和工业系统中的自适应控制和预测性维护。基于 **MSPM0G5187** Arm®Cortex®-M0+ 的 MCU 使用其专用 TinyEngine NPU 来独立于主 CPU 执行深度神经网络模型，从而在可穿戴健康监护仪、电器和工业系统的预测性电机维护中实现边缘 AI 功能。

## 边缘 AI 软件创新

除了嵌入式器件中高效 AI 计算取得的硬件进步外，开源社区和半导体制造商也使测试和部署 AI 模型变得更加容易，只需极少的编程专业知识即可。使 AI 更易于使用（某些情况下基于 GUI）有助于减少投入额外资源或培训的需要。

对于更熟悉 AI 模型的设计人员而言，PyTorch 和 TensorFlow 等开源工具可以为其自定义数据集训练模型架构，并将模型导出为嵌入式友好格式，例如 ONNX 或 LiteRT（以前称为 TensorFlow Lite）。然后，该模型在器件上使用等效的开源运行时软件运行。

这些开源工具通过对嵌入式平台的细节进行抽象化处理来促进边缘 AI 开发，提供一致的接口，以便访问硬件加速后端（也称为代理）。这些后端可提供更多配置，使设计人员能够更好地控制将模型委派给硬件加速器的过程。

**CCStudio™ Edge AI Studio** 一系列网页端工具，旨在使用远程 TI 硬件和图形用户界面，简化和加快 TI 嵌入式器件上边缘 AI 应用的开发。这些工具包括模型编写器、模型分析器、模型选择工具和模型制作器，可帮助设计人员快速评估模型及其性能，而无需实际连接到评估板。

## 边缘 AI 的可扩展性

在使用嵌入式微控制器或微处理器开发产品时，务必考虑产品如何随时间演变和扩展。工程师们不希望花几个月的时间在一个微处理器上开发解决方案，然后在产品升级到更高性能的处理器时又不得不从头开始。

制造这些嵌入式器件的半导体制造商需要开发在功能、性能和成本方面具有可扩展性的产品系列。这种方法有助于确保在用于 AI 的各种嵌入式处理器之间制定一个无缝迁移策略，使开发人员可以尽可能简单地在不同器件上重复使用之前的工作成果。

边缘 AI 也不例外。例如，制造家用机器人的设计人员可能希望同时生产配备三个摄像头用于环视的高端版本，以及仅配备一个前置摄像头的入门级版本。可扩展的边缘 AI 加速器件组合可将软件从高端型号移植到入门级型号，从而更大限度地减少生产这两种产品所需的资源。可扩展性还允许开发人员随着产品的发展，将研发投资从一个平台转移到下一个平台。

## 结语

尽管边缘 AI 仍然相对较新，但它重塑我们日常生活的潜力正逐渐成为人们关注的焦点，尤其是它为几乎所有应用带来更高的响应速度和性能的能力。随着低功耗、经济高效的嵌入式处理器以及直观软件和模型训练工具的进步，任何经验水平的设计人员都可以轻松入门。我们可以预见，随着边缘 AI 设备和关键元件（例如，用于检测、供电和连接的半导体）的不断发展，这种情况将继续下去，这些关键元件负责管理我们所使用和依赖的电子产品的运行和数据收集。

## 其他资源

- 探索 [TI 的边缘 AI 处理产品系列和设计资源](#)。
- 阅读以下技术文章，了解 TI 的边缘 AI 加速 MCU
  - [边缘 AI 加速的 Arm® Cortex®-M0+ MCU 如何为电子产品注入更强智能](#)
  - [在工业自动化和家用电器设计中实现支持边缘 AI 的电机控制](#)
- 在产品概述 [TI 的 TinyEngine™ NPU 在更多嵌入式系统中释放边缘 AI 加速能力](#) 中了解有关 TinyEngine NPU 的信息及其为支持边缘 AI 加速的嵌入式设计带来的优势。

**重要声明:** 本文所提及德州仪器 (TI) 及其子公司的产品和服务均依照 TI 标准销售条款和条件进行销售。建议客户在订购之前获取有关 TI 产品和服务的最新和完整信息。TI 对应用帮助、客户的应用或产品设计、软件性能或侵犯专利不负任何责任。有关任何其它公司产品或服务的发布信息均不构成 TI 因此对其的认可、保证或授权。

所有商标均为其各自所有者的财产。

## 重要通知和免责声明

TI“按原样”提供技术和可靠性数据（包括数据表）、设计资源（包括参考设计）、应用或其他设计建议、网络工具、安全信息和其他资源，不保证没有瑕疵且不做任何明示或暗示的担保，包括但不限于对适销性、与某特定用途的适用性或不侵犯任何第三方知识产权的暗示担保。

这些资源可供使用 TI 产品进行设计的熟练开发人员使用。您将自行承担以下全部责任：(1) 针对您的应用选择合适的 TI 产品，(2) 设计、验证并测试您的应用，(3) 确保您的应用满足相应标准以及任何其他安全、安保法规或其他要求。

这些资源如有变更，恕不另行通知。TI 授权您仅可将这些资源用于研发本资源所述的 TI 产品的相关应用。严禁以其他方式对这些资源进行复制或展示。您无权使用任何其他 TI 知识产权或任何第三方知识产权。对于因您对这些资源的使用而对 TI 及其代表造成的任何索赔、损害、成本、损失和债务，您将全额赔偿，TI 对此概不负责。

TI 提供的产品受 [TI 销售条款](#)、[TI 通用质量指南](#) 或 [ti.com](#) 上其他适用条款或 TI 产品随附的其他适用条款的约束。TI 提供这些资源并不会扩展或以其他方式更改 TI 针对 TI 产品发布的适用的担保或担保免责声明。除非德州仪器 (TI) 明确将某产品指定为定制产品或客户特定产品，否则其产品均为按确定价格收入目录的标准通用器件。

TI 反对并拒绝您可能提出的任何其他或不同的条款。

版权所有 © 2026，德州仪器 (TI) 公司

最后更新日期：2025 年 10 月