

Electronic Products、「the 2024 Product of the Year Awards」の 受賞製品を発表

EdgeCortexの「SAKURA-II エッジAIアクセラレータ」が Electronic ProductsのデジタルIC部門で
「the 2024 Product of the Year Awards」を受賞

記事原文はこちらからご確認ください。

<https://www.electronicproducts.com/electronic-products-announces-winners-of-the-2024-product-of-the-year-awards/>



今年で49回目を迎えるこの賞は、技術またはその応用の著しい進歩、デザインの革新、価格と性能における実質的な達成、デザイン性能の向上、新しい製品デザイン／機会の可能性のいずれかの特徴を持つ優れた製品を表彰するものです。Electronic Products編集部は今回、13のカテゴリーにわたる125点以上の製品を評価しました。

EdgeCortex Inc.: SAKURA-II AI アクセラレータ (デジタル IC)

「SAKURA-II」は迅速でリアルタイムのAI推論を必要とするアプリケーション向けに設計されたエネルギー効率の高いAIアクセラレータです。小型で低消費電力のシリコンデバイスに最適化されており、小型言語モデル、大型言語モデル、そして大規模なビジョンモデルなど最新の生成AI (GenAI) タスクをエッジで処理するために設計されています。

SAKURA-IIIは業界をリードする90%の効率を実現し、わずか8Wで60TOPSを達成し、既存のGPUや他のAIプロセッサを凌駕する性能を誇ります。低消費電力と相まって、SAKURA-IIIは低消費電力のエッジAI処理に適しています。これにより、スマートシティ、農業、製造業、小売業、家電製品、さらには自動運転車、ロボティクス、航空宇宙・防衛、セキュリティ、そしてマルチアクセスエッジコンピューティングを含むさまざまなスマートエッジに特化したアプリケーションで使用される数十億のパラメータを持つモデルのエッジAI処理をサポートします。

SAKURA-II AIアクセラレータに加え、このプラットフォームにはDynamic Neural Accelerator (DNA)、MERAコンパイラプラットフォーム、およびSAKURA-II モジュールとカードが含まれています。



EdgeCortex SAKURA-II AI アクセラレータ
(出典: EdgeCortex)

DNAニューラルネットワークIPは、エッジAI向けにランタイムで再構成可能なアーキテクチャで、畳み込みネットワークとトランスフォーマーネットワークの両方をサポートしています。DNAは、EdgeCortexの特許取得済みの並列処理アーキテクチャを3つの計算ブロックにわたって使用し、モデルの処理性能を最適化し、数十億のパラメータを持つモデルを容易に処理できると、同社は述べています。多くの生成AI (GenAI) モデルが約100億パラメータに近づいているため、この機能はAI業界にとって極めて重要です。自社でチップを設計する顧客向けに、DNAは新しいカスタムシリコンデバイスに追加可能なIPコアとして利用可能です。

MERAは、エッジAI機能を既存のシステムにシームレスに統合できるヘテロジニアスなコンパイラです。Intel、AMD、Arm、RISC-Vなど、さまざまな既存のCPUをサポートしており、Hugging FaceやEdgeCortexモデルライブラリから直接入手できる、事前定義および最適化された大規模なモデルライブラリも活用できます。

SAKURA-IIIは、さまざまなPCIeカードおよびM.2モジュールとともに利用可能です。また、PCIeカードとMERAコンパイラが事前にインストールされたボックスレベルのソリューションも提供しており、システムを持っていない顧客でも、すぐにモデルのロードとテストを開始できます。