

Bringing AI processing to the edge requires companies to clearly understand device requirements and constraints and choose their vendor partners carefully.

Bringing AI Edge Solutions to Market: Challenges and Solutions

April 2023

Written by: Nina Turner, Research Director, Enabling Technologies: Artificial Intelligence Edge Processor Architectures

Introduction

Edge devices are beginning to use AI as a tool to enable more capable and intelligent functions. Device designers and manufacturers must decide where the AI computation will occur. AI inference can occur in the cloud, at the edge infrastructure, or on the edge system itself. The decision of where the AI compute will reside depends on many factors, including the complexity of the AI inference, data security, privacy, latency requirements, connectivity, the power envelope, and physical limitations of the device.

One of the biggest challenges for device designers and manufacturers in implementing AI is the lack of AI expertise in the company itself. Core competency in many companies does not reside with AI. The shortage or lack of AI expertise can lead companies to define incorrect product specifications or requirements, slow product development, increase development costs, or create failed products that do not meet the expectations of the company.

Choosing the right platform and partnering with vendors that have the right level of AI compute; experience in the AI functionality, product area, or vertical industry; and a proven track record of implementation can speed time to market, help companies clarify what can and cannot be implemented within the limitations of the device itself, and bring the AI expertise that the device manufacturer lacks.

Benefits

The key benefit of bringing AI functionality to the edge device is that it reduces the latency of the compute by processing the data locally. This means decisions can be made in real time and is especially important for critical functions that involve safety. Latency is also important in the user interface, as delays in data processing can lead to poor user experiences. The processing of data on the device enhances data security and privacy, as sensitive data stays at the device level and is not

AT A GLANCE

WHAT'S IMPORTANT

AI at the edge will be a growing competitive differentiator, but successful development and implementation pose several challenges.

KEY TAKEAWAYS

- » System designers and manufacturers must understand the requirements and limitations of the system they are adding AI to.
- » Companies must determine whether they have the AI expertise to create the product in-house or if partnering with AI vendors that have experience in implementing the desired AI functionality can create a product successfully.

transmitted across the network. An increasing number of security incidents, more government regulations, and the resultant customer awareness of what data is transmitted are also driving forces for moving compute toward the edge. Edge AI can help lower cloud storage costs by ensuring that only the most essential data is stored in the cloud.

Other leading AI functions moving to the edge today include computer vision, audio analysis, and autonomous devices. Computer vision can improve and quickly enable functions such as object detection, classification, and tracking in real time. Companies whose systems work with unstructured data can potentially improve their systems' performance and capabilities by deploying AI at the edge. Companies are also combining AI at the edge with cloud analytics, which delivers on data processing performance augmented by cloud analytics or services that can also lead to new services and new revenue streams, potentially adding on recurring revenue sources beyond the system itself.

Partnering with an experienced AI semiconductor or platform vendor can speed up device and AI development, shorten time to market, and potentially deliver a first-mover advantage. Device manufacturers must decide whether the AI expertise resides in-house or if they want to invest in developing in-house talent while considering the training time required. It's important for device manufacturers to clearly define what capabilities should reside in-house versus what capabilities can be developed utilizing an AI vendor's expertise when they select their AI vendor. This choice can lead to cost savings in product development, as the incorrect choice can lead to wasted time or mistakes.

Device companies must also think beyond single-device development and consider their road map when determining which AI vendor to work with. The AI vendor product and platform road map should be clear and be in line with the device company road map. This will allow companies to leverage the investments they make in AI into other products or create a family of products. AI vendors that take a flexible platform approach allow device manufacturers to determine how much of the AI vendor's solutions they want to use while also enabling rapid product extensions.

Considerations

As companies determine how to include AI on their product road maps, they must keep in mind several considerations, including device form factor, device function, and power constraints. The device form factor is a key limiter on choice of semiconductor, as the device size or shape can constrain what can be used. The device may be battery operated, or the cooling requirements of the semiconductor could limit how much compute can be included on the device.

Device manufacturers must also determine whether adding AI to the device will bring real benefits or whether AI is not useful for the device application. The manufacturer must determine whether the benefit of adding AI brings compelling new product functionality that will deliver returns on its investment. Other considerations before adding AI to devices are what experience the device manufacturer has in AI software and hardware integration and how much of an investment to make in AI development capability. A misunderstanding or an underestimation of the amount of expertise involved in AI development could lead to implementation challenges and product development delays.

Device companies must also think beyond single-device development and consider their road map when determining which AI vendor to work with.

Trends

Device manufacturers must choose between an integrated processor and an AI accelerator when moving AI compute to the edge. The utilization of AI accelerators can speed time to market, as the accelerator allows the system designer to continue to use familiar main processors while offloading some functions that utilize AI. Accelerators can improve performance for specific tasks, increase energy efficiency compared with a general processor, and reduce costs. An integrated processor can bring benefits for some applications such as efficient use of space, lower overall power consumption, and improved general performance.

Device manufacturers are partnering with companies that have AI software and model libraries that have the functionality they are looking for to fast-track device development. Edge devices demand efficient AI inference models, and companies are looking to their AI vendors to utilize AI model efficiency tools and techniques to size models correctly for their device application.

Conclusion

Before deciding to add AI functionality to their systems, designers and manufacturers must do a cost-benefit analysis, as not all systems will benefit significantly from AI. Device manufacturers must clearly specify the processing and compute requirements as well as the system constraints.

Companies must evaluate which processor or accelerator suppliers to use and partner with a vendor that provides the correct balance of power and performance. In addition, they must understand their in-house AI development capabilities when determining whether to develop in-house or partner for AI function integration. By taking these factors into account, device manufacturers will be more likely to develop successful AI-capable edge systems.

About the Analyst



Nina Turner, Research Director, Enabling Technologies: Artificial Intelligence Edge Processor Architectures

Nina Turner is a Research Director on IDC's Enabling Technologies and Semiconductor team. Her core research coverage includes AI edge processors and automotive semiconductors. With more than 20 years of experience covering the semiconductor industry, she has extensive experience in various technologies, product development and management, and market and technology strategic assessments.

MESSAGE FROM THE SPONSOR

About Syntiant

Founded in 2017 and headquartered in Irvine, Calif., Syntiant Corp. is a leader in delivering hardware and software solutions for edge AI deployment. The company's purpose-built silicon and hardware-agnostic models are being deployed globally to power edge AI speech, audio, sensor and vision applications across a wide range of consumer and industrial use cases, from earbuds to automobiles. Syntiant's advanced chip solutions merge deep learning with semiconductor design to produce ultra-low-power, high performance, deep neural network processors. Syntiant also provides compute-efficient software solutions with proprietary model architectures that enable world-leading inference speed and minimized memory footprint across a broad range of processors. The company is backed by several of the world's leading strategic and financial investors including Intel Capital, Microsoft's M12, Applied Ventures, Robert Bosch Venture Capital, the Amazon Alexa Fund and Atlantic Bridge Capital. More information on the company can be found by visiting www.syntiant.com or by following Syntiant on Twitter [@Syntiantcorp](https://twitter.com/Syntiantcorp) or [LinkedIn](https://www.linkedin.com/company/syntiant).



The content in this paper was adapted from existing IDC research published on www.idc.com.

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2023 IDC. Reproduction without written permission is completely forbidden.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
[idc-insights-community.com](https://www.idc-insights-community.com)
www.idc.com