

Artificial Intelligence
Partner Sales Guide

Selling Intel® AI Hardware: A Conversation Guide

Customer conversations that initiate and guide AI investments

Artificial intelligence (AI) is transforming nearly all aspects of business, helping to accelerate innovation, improve efficiency, automate and optimize workflows, and reduce costs. Enterprises that see their AI ideas through to fruition stand to gain numerous benefits as well as a competitive advantage.

Your customers are likely feeling pressure to get their enterprise AI initiative up and running. However, they may not know how to start such a complex project when it comes to their hardware. That's where your expertise, the comprehensive Intel® AI hardware portfolio, and this guide can help.

Goal: Discover customer needs

Begin your customer engagement with an exploratory conversation to uncover business challenges and desired outcomes. The insights gathered will be critical in determining the amount of compute power required and which AI technologies can address their needs.

Use these open-ended questions to guide conversations that can help you gain an understanding of your customer's unique situation and inform right-fit hardware and AI technology recommendations.

CUSTOMER QUESTIONS REASON FOR ASKING

What are you hoping to achieve with AI?

Desired AI outcomes

This is an essential starting point for understanding customer technology requirements. For example, the customer may want to launch a dynamic GenAI enterprise-class large language model (LLM) application on their website that will use massive datasets and intricately trained algorithms. This type of outcome will require powerful hardware that delivers low latency and fast response times, which is different than what is needed to launch a customer chatbot.

What type of AI workload do you want to deploy?

AI workload types

The computational requirements for training and deploying AI models vary greatly based on several factors, including the type of AI workload being used. For example, if your customer is deploying a classical machine learning or deep learning workload with relatively smaller compute demands, a CPU with integrated AI accelerators—rather than a discrete GPU—may be able to address their needs.

What size of AI workloads do you anticipate processing?

AI workload size

Each AI workload type, from machine learning to deep learning, large language models, or GenAI, is made possible from a dataset. That dataset can vary in size and number of parameters—from small scale to large scale. The larger and more complex the dataset, the more specialized and powerful the AI hardware will need to be.

Will you be performing dedicated (one function) or mixed (operational and analytical) workloads?

AI processing capabilities

AI can be useful in completing a variety of tasks relative to business targets, from automating routine work to performing complex computations to synthesizing and analyzing massive amounts of data. If your customer mentions they plan to work with mixed AI workloads, meaning both operational and analytical, they will need flexibility in their infrastructure. If your customer is deploying only GenAI workloads, they will need powerful, high performance compute.

Where do you plan to deploy your AI workloads?

Deployment plans

Deployment gets AI projects into production for real-world use and application. Technically, customers have multiple options for deploying their workloads—in the data center, at the edge or far edge, on-premises, in the cloud, or some combination of environments. Your customers' chosen deployment path can influence their hardware investment.

Do you have in-house expertise to build your own AI infrastructure?

Build vs. buy

This question helps you determine if the customer can build a solution or if they will need to buy and customize a prepackaged solution. Their answers will be pivotal in determining the optimal infrastructure for their environment. Depending on their level of in-house expertise, your customer may be able to build what they need themselves with reference designs, or you may need to offer a ready-made solution.



Setting customer expectations across key AI workflows

Every stage of the AI pipeline has its own unique challenges and compute requirements, so it's vital to understand your customers' needs.



Preparing data for migration or use:

The first natural step to success for any AI project begins with data preparation and can be more complex than your customer may understand. Intel has a broad selection of programs, utilities, and resources that can be used to support customers with their data prep initiatives.



Training, fine-tuning, or inferencing:

Use the information gathered during your conversation, such as the desired business outcome, size and type of workload, and AI model, to determine the hardware recommendations.



Building applications from scratch or customizing a ready-made reference implementation:

It's important to learn the customers' technical capabilities, resources, and commitment to developing an AI solution from the ground up or to customize an existing solution.



Deploying solutions at scale:

As AI models get bigger, power efficiency becomes a critical factor in driving productivity, with a wide range of complex workload functions from data preprocessing to training and inference.

Intel® hardware technologies deliver throughout AI projects

Since all Intel hardware technologies can be applied during these stages of the AI pipeline, having the pertinent project details will be key.

No matter where you need to extend AI capabilities, Intel offers an optimized hardware solution for your customers' AI applications. [Explore our flexible, end-to-end portfolio for AI acceleration.](#)

Intel delivers the architecture that powers AI performance and innovation

From training to deployment, take advantage of the complete Intel® AI processor and hardware portfolio to accelerate client, edge, data center, and cloud applications.



Intel® Core™ Ultra processors

Powering AI PCs with dedicated hardware to handle AI tasks locally >

New-generation PC devices with a built-in central processing unit (CPU), graphics processing unit (GPU), and neural processing unit (NPU) deliver supercharged speed, efficiency, privacy, and enhanced security for AI workloads.



Intel® Xeon® Scalable processors

Accelerating performance across fast-growing workloads >

Intel Xeon Scalable processors are purpose-built to deliver powerful computing performance to help customers grow and unlock new opportunities. They can be used for a wide range of AI workloads across the entire AI pipeline and feature integrated AI accelerators. From edge to cloud, boost performance for machine and deep learning training and inferencing without using specialized hardware. If your customers have more-advanced, more-demanding workloads, Intel Xeon Scalable processors can be augmented with a discrete GPU.

Intel® Accelerator Engines are integrated features in Intel Xeon Scalable processors that help boost performance, reduce costs, and improve power efficiency for demanding workloads in the data center, in the cloud, and at the edge.



The AI PC by Intel

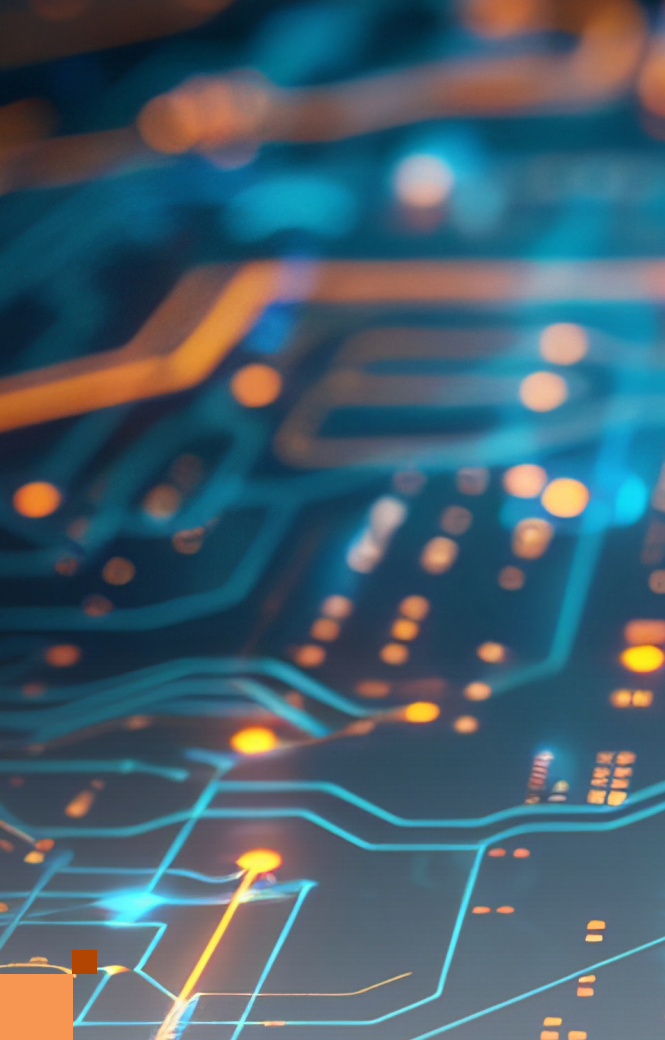
AI built into a new generation of PCs >

Built to run the next wave of advanced AI workloads locally and more efficiently, AI PCs with Intel® Core™ Ultra processors feature dedicated CPU, GPU, and NPU engines to help unlock the power of AI, create immersive graphics experiences, and enable high-performance, low-power processing.

Featured AI accelerator engine

Intel® Advanced Matrix Extensions (Intel® AMX) is ideal for:

- Machine learning training and inference
- Intermittent deep learning training and cost-effective deep learning inference
- GenAI inference model fine-tuning and less-demanding training workloads



Intel® Gaudi® AI Accelerators

Large-scale deep learning training and inferencing >

Intel Gaudi AI Accelerators are architected and built to handle the deep learning training and inference demands of large-scale era AI. With these high-performance, high-efficiency accelerators, your customers can now realize the advantages of large-scale deep learning, which—until recently—wasn't attainable for many enterprises and organizations. Another advantage Intel Gaudi AI Accelerators bring to customers is flexible scalability. Deployments can scale from a single accelerator to clusters of accelerator-enabled AI servers.



Intel® Data Center GPU Flex Series

High throughput and low TCO for AI visual inference workloads at scale >

The Intel Data Center GPU Flex Series is a general-purpose GPU for customers exploring options to maximize the amount of inference throughput per dollar spent. With two SKU options—the Intel® Data Center GPU Flex Series 140 and the Intel® Data Center GPU Flex Series 170—you can tailor your recommendation based on the power level the customer needs. The Flex Series 140 is ideal for lighter models, such as object detection, whereas the Flex Series 170 is optimal for more-complex AI models, including object detection or multiple classification models.



Intel® Max Series Product Family

Designed for data center computing models used in AI and HPC >

Intel Max Series processors (CPUs and GPUs) are architected for the future compute demands of AI and HPC workloads. Customers have a choice to maximize bandwidth with the Intel® Xeon® CPU Max Series, the first and only x86-based processor with high-bandwidth memory (HBM). Or maximize impact with the Intel® Data Center GPU Max Series—Intel's highest-performing, highest density, general-purpose discrete GPU.



Intel® Arc™ GPU

Fast AI inference at the edge >

Discrete Intel Arc GPUs can help enhance customer engagement with interactive kiosks and provide future-ready, high-efficiency AI, visual computing, and media processing at the edge.



Pro tip:

Share the Intel® Tiber™ Developer Cloud with customers so they can experience the advantages of running training and inference workloads on the Intel® Gaudi® platform.

Staying secure >

Customers can responsibly deploy solutions and protect AI data in compliance with industry regulations by leveraging the confidential compute and hardware-based security features integrated into Intel® technologies.



Next steps

Learn more about how to sell the latest Intel AI hardware with these partner-ready tools, training, and resources.

Watch Intel® Cloud TV

[The AI PC Opportunity >](#)

[5th Gen Intel® Xeon® Scalable Processors Overview >](#)

[Your GenAI Opportunity with Intel® Gaudi® AI Accelerators >](#)

Become an Intel® Certified Pro

[Principles of AI Training >](#)

Visit the Intel® Partner Alliance website

[Activation Zones on AI PC, GenAI, and Edge AI >](#)

[Partner Sales Enablement Resources >](#)



Notices and disclaimers

Intel® technologies may require enabled hardware, software, or service activation. No product or component can be absolutely secure.

Your costs and results may vary. Performance varies by use, configuration, and other factors.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

See our complete legal notices and disclaimers at [edc.intel.com/content/www/us/en/products/performance/benchmarks/overview/](https://www.intel.com/content/www/us/en/products/performance/benchmarks/overview/).

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

1024/AK/CMD/PDF