

Power Your AI Transformation: 5 Reasons Why Intel® Xeon® 6 Processors with P-Cores Excel at AI

Efficiently take on growing AI needs alongside your existing general-purpose workloads by deploying systems with Intel Xeon 6 processors.

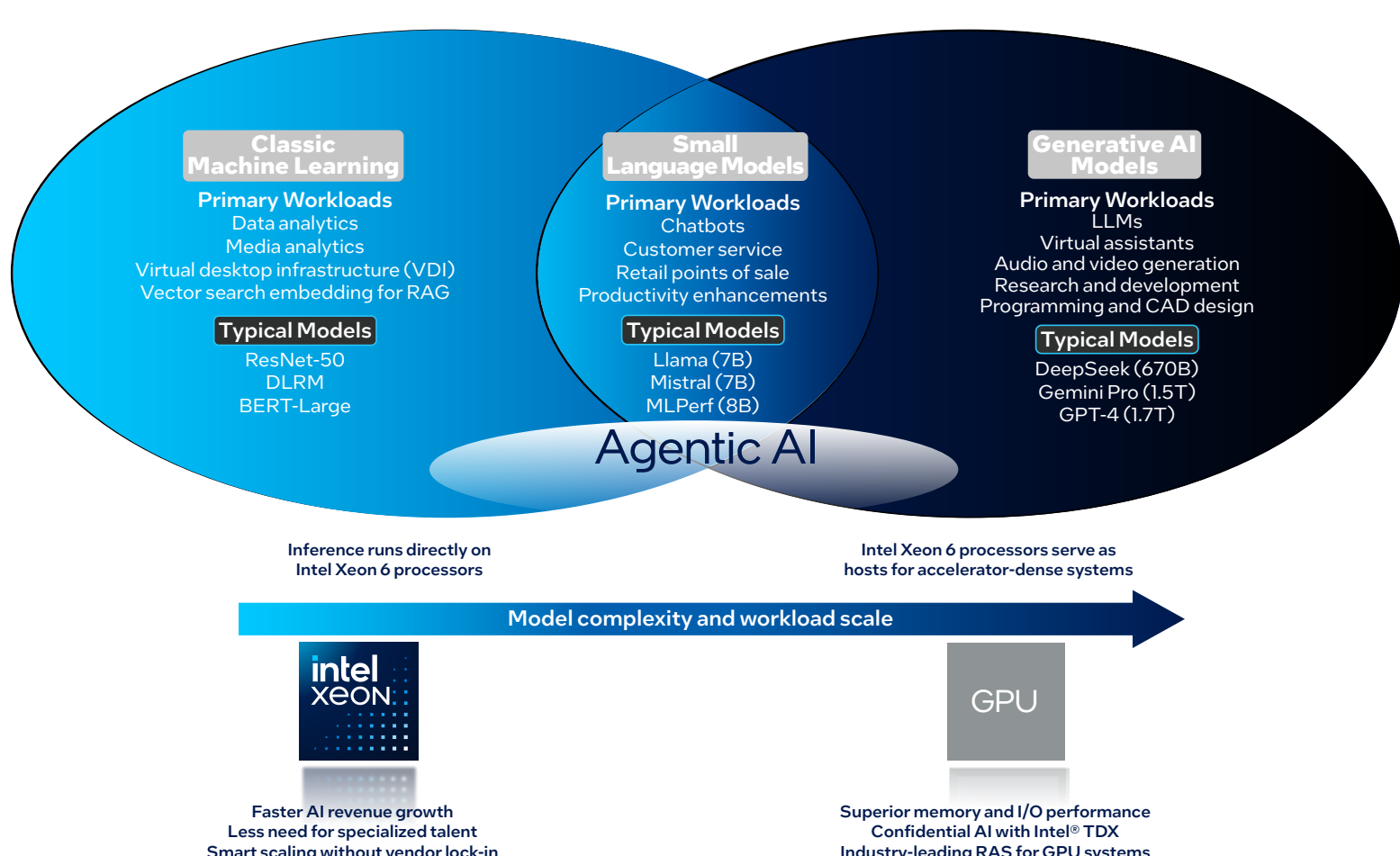
Costly, dedicated systems with GPUs aren't always needed in the data center.

With exceptional performance and efficiency, Intel Xeon 6 processors can help you reduce total cost of ownership (TCO) by consolidating servers and reducing power consumption in the data center.

But Intel Xeon 6 processors with Performance-cores (P-cores) also excel at running the majority of AI workloads used in the enterprise today.

It's no secret that AI development and spending are growing rapidly. Non-generative AI workloads will continue to be a significant part of overall AI growth. You don't need a dedicated GPU for all AI, because Xeon CPUs can efficiently handle most non-generative AI tasks and excel in inference for traditional and generative AI. With support for any precision type, Intel Xeon 6 processors with P-cores are here to help.

Choose Compute Based on Model Complexity



Why choose Intel Xeon 6 processors with P-cores?

Here are five reasons to choose Intel Xeon 6 processors with P-cores as the right CPUs to support your AI workloads.

1

More cores, memory bandwidth, and cache

Higher CPU core counts and greater memory bandwidth translate to better AI performance, directly from your Intel Xeon processor.

Innovative Multiplexed Rank DIMMs (MRDIMMs) deliver improved memory bandwidth and up to 504 MB low-latency last-level cache (LLC), which significantly boosts performance for memory-bound AI and high-performance computing (HPC) workloads.

Up to **128 cores per CPU** deliver 2x more cores per socket than 5th Gen Intel Xeon processors

Up to **30% better AI performance** compared to DDR5-6400 DIMMs¹

Up to **2.3x higher AI performance** with MRDIMMs, compared to 5th Gen Intel Xeon processors²

2

Integrated AI acceleration

Intel Xeon 6 processors with P-cores include Intel® Advanced Matrix Extensions (Intel® AMX) and Intel® Advanced Vector Extensions 512 (Intel® AVX-512) acceleration in every core to boost AI and HPC workloads.

Intel AMX includes support for INT8, BF16, and now FP16 data types. Optimizations are also integrated into the mainstream distributions of popular frameworks like TensorFlow, PyTorch, Llama CPP, vLLM, and others.

Integrated acceleration helps **eliminate costs and data bottlenecks** inherent when using discrete accelerators

Up to **42% better performance with Intel AMX** compared to the prior generation³

3

Scaled power efficiency and server consolidation

Address growing power usage and space constraints by refreshing aging infrastructure. Intel Xeon 6 processors with P-cores bring improved energy efficiency that scales with utilization.

Consolidating servers powered by Intel Xeon 6 processors reduces server space requirements and energy consumption for a lower TCO while maintaining exceptional performance for AI workloads.

Up to **1.9x better performance per watt** at typical 40% utilization compared to 5th Gen Intel Xeon processors⁴

Up to **44% lower TCO** running a BERT-Large LLM workload compared to running on an AMD EPYC processor⁵

4

Exceptional AI performance

Intel Xeon 6 processors with P-cores deliver exceptional compute power to support a wide variety of workloads, including small to medium LLMs and generative AI models for inferring, fine-tuning, and retrieval-augmented generation (RAG) use cases.

Up to **2x better AI inference performance** compared to AMD EPYC processors⁶

Up to **1.5x better AI performance with 33% fewer cores** compared to AMD EPYC processors⁷

5

Leading data security and open software ecosystem

Intel Xeon 6 processors are trusted and secure, with the most comprehensive portfolio for confidential computing, including Intel® Software Guard Extensions (Intel® SGX) and Intel® Trust Domain Extensions (Intel® TDX), leading to unmatched data security.

Intel has teamed with industry partners and the open source community to provide a rich ecosystem of validated technologies and seamless integration with common operating systems, compilers, libraries, and frameworks. With this shared software stack and a global ecosystem of hardware and software vendors, solutions can be matched to every business need. Intel actively contributed reference implementations to Open Platform for Enterprise AI (OPEA).

Up to **75% fewer firmware vulnerabilities than AMD EPYC processors⁸**

The Intel ecosystem includes **ready-to-use, enterprise AI applications** from priority software vendors.

Learn how you can power all of your AI goals with Intel AI solutions.

Explore how Intel Xeon 6 processors serve as powerful and efficient host CPUs for AI-accelerated systems.

intel
xeon

¹For MRDIMMs compared to DDR5-6400 RDIMMs.
²See [9A6] at [intel.com/processorclaims](https://www.intel.com/processorclaims): Intel Xeon 6. Results may vary.
³See [A16] at [intel.com/processorclaims](https://www.intel.com/processorclaims): 5th Generation Intel Xeon Scalable Processors. Results may vary.
⁴See [9G2] at [intel.com/processorclaims](https://www.intel.com/processorclaims): Intel Xeon 6. Results may vary.
⁵See [9T22] at [intel.com/processorclaims](https://www.intel.com/processorclaims): Intel Xeon 6. Results may vary.
⁶See [9A22] at [intel.com/processorclaims](https://www.intel.com/processorclaims): Intel Xeon 6. Results may vary.
⁷See [7A220] at [intel.com/processorclaims](https://www.intel.com/processorclaims): Intel Xeon 6. Results may vary.
⁸In hardware root-of-trust vs. AMD. Source: Intel, "2024 Intel Product Security Report," February 2025.

Performance varies by use, configuration and other factors. Learn more at www.intel.com/PerformanceIndex.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See configuration disclosure for additional details.

No product or component can be absolutely secure.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

AI features may require software purchase, subscription or enablement by a software or platform provider, or may have specific configuration or compatibility requirements. Details at [intel.com/ai/cp](https://www.intel.com/ai/cp).

© 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.