AI Performance Starts at the Processor

Generative AI is leading a sea change in productivity that will add trillions to the economy. Small wonder, then, that AI has become the biggest spend item for almost 50 per cent of tech executives surveyed. That's more than double the number of leaders that list the cloud as their priority, and that has been a rapid shift. Cloud has been the dominant priority for many years, but now two-thirds of tech execs report they are accelerating their investments into AI.

AI is poised to play a role in just about every facet of business. This includes removing bias from job descriptions and screening for candidates, creating images for marketing and assisting the customer support team in addressing customer needs. The question is, given such a breadth of opportunity and so much hype behind AI, how organisations can start formulating an AI strategy to take advantage of the opportunity.

A Platform For Innovation

According to AMD Senior Director of Commercial Client and Workstation, Matthew Unangst, more than half of IT managers (52 per cent) say they lack the proper technology infrastructure to run AI workloads.[1]

To help address this deficiency, AMD has developed solutions that place AI at the chip level. AMD Ryzen™ AI, an integrated AI engine available on select systems, helps to dynamically boost system performance and enable organizations to take advantage of AI experiences. Built on industry-leading 4nm technology and featuring up to 8 "Zen 4" cores and RDNA™ 3 integrated graphics, the AMD Ryzen PRO 7040 Series processors enable excellent levels of performance and efficiency for business applications, with the Ryzen 7 PRO 7840U processor offering an average 17% increase in performance over the x86 competition.[2] The AMD Ryzen™ 7 PRO 7840U processor offers up to 18% higher CPU performance[3] and up to 15% better battery life while video conferencing, compared to an Apple M2 Pro processor.[4]

Immediately, organizations will be able to leverage AI to enhance the experience of video conferencing applications like Microsoft Teams. Microsoft also has whole-PC AI-driven experiences, such as Windows Copilot, that will significantly change the productivity dynamics on PCs. With it, users will be able to ask their devices to rewrite, summarise, or analyze the content that they're working with.

However, without a significant boost to the processing power of the device, AI tools such as these may slow the user experience, resulting in a decrease in productivity and causing frustration for the user. This is particularly true when users are trying to run multiple AI-enhanced applications simultaneously. This is why establishing the platform for AI with AI-assisted hardware, like AMD Ryzen AI, is so critical now. It enables a seamless user experience while working with these increasingly demanding applications, while also processing efficiently and helping to preserve battery life.

AMD has worked closely with Microsoft to develop a cohesive vision for AI, so user devices are able to better support the AI-driven applications that will soon dominate the Windows experience. "Microsoft and AMD have worked closely to enable Ryzen AI in prominent Windows features like Windows Studio Effects," said Pavan Davuluri, corporate VP of Windows silicon and system integration at Microsoft.

Some of the enhancements that Ryzen AI brings to Windows Studio Effects include Automatic Framing which automatically pans and zooms the video to focus on you as you move, Eye Contact Correction which compensates for eyes looking away from the center of the camera, and Advanced Background Effects which enables high-quality background blurs.

From device to data center

The AMD focus on integrating AI-powered accelerators into laptop PCs is a direct response to the changing dynamics of work. Remote and hybrid work environments are now the norm for many companies around the globe. At the same time, workplaces have become more collaborative than ever. To enable success, businesses need to provide immersive and efficient mobile computing experiences. At the same time, workplaces are more collaborative than ever, so the experience when remote needs to be immersive and rich to succeed.

Workforce decentralization also means that much of the computing power needs to be hosted locally. Workers need to be able to collaborate and be equally productive from anywhere while working over Wi-Fi® 4G or 5G connectivity.

Meanwhile, the acceleration of AI has been rapid at the data center too, and AMD is committed to bringing AI and innovation here, and assisting organisations deliver their higher-level AI ambitions. The company acquired Xilinx in 2022 for its range of high-performance and adaptive computing capabilities – including AI – and the first products that come as a result of the acquisition have been released this year.

Meanwhile, in June AMD announced new generations of data center hardware, specifically designed to capitalise on the potential for AI. For example, the AMD MI300A, an APU accelerator for AI and HPC workloads, allows for hyper-efficient sharing of resources between the CPU and GPU, and fully deliver on the processing requirements of AI. Additionally, the MI300X are GPUs designed to handheld the large language models (LLMs) that underpin generative AI.

In short, these data center solutions will allow organisations to scale their use of AI and apply it at a whole-of-enterprise level, centralising much of the AI experience and ensuring consistency and quality with the user experience.

IT leaders are now needing to expect the unexpected and are looking for solutions that have some kind of "future-readiness," behind them. Having an excess of power and performance on the devices and in data centers now means that the technology will continue to handle the next few waves of AI innovation, even if no one can be sure what those will be right now.

Finally, at a time when power prices are rapidly accelerating, inefficiencies in technology can quickly become very expensive. One of the clear goals with implementing AI at the chipset is that it cannot only deliver on performance, but also scale to draw lower levels of power. Tallied across a fleet of devices all running AI applications could enable considerable ROI in power savings alone.

AI is set to transform endpoint user devices like never before. As part of a true generational shift, it will be critical to have devices with high-performance chips and AI embedded at the hardware level to ensure that the technology is able to fully realise the productivity gains on offer.

For more information on how you can advance AI from Cloud to Edge to Endpoints, [click here](.).

---

[1] Online survey conducted by Edelman Data & Intelligence from May 3 to May 25, 2023, among n=2,500 IT Decision Makers in the U.S., U.K., Germany, France, and Japan.

[2] Testing as of 5/31/23 by BOXX Technologies, commissioned by AMD, utilizing an HP EliteBook 845 G10 with Ryzen PRO 7840U processor, Integrated Radeon Graphics, 16GB RAM, 1TB NVMe SSD, Windows 11 Pro versus a similarly configured Dell Latitude 5440 with an Intel Core i7-1365U processor, Intel Integrated graphics, 16GB RAM, 512GB NVMe SSD, and Windows 11 Pro , using a composite Geomean average of the scores of the following benchmark tests: Geekbench v5 Single Core, Passmark 11 CPU Mark and PCMark 10 Benchmark. Laptop manufacturers may vary configurations yielding different results. PCMark® is a registered trademark of Futuremark Corporation. PHXP-28

[3] HP EliteBook 845 G10 with AMD Ryzen™ 7 PRO 7840U processor offers up to 18% higher CPU performance in Passmark 11 CPU Mark benchmark when compared to a Apple MacBook Pro 14 with M2 Pro (10 core) processor. (PHXP-30)

[4] The HP EliteBook 845 G10 with AMD Ryzen™ 7 7840U processor and 65WH battery offers up to 15% better battery life than an Apple MacBook Pro 14 with M2 Pro 10 core CPU and 69.6 WH battery, while video conferencing using Microsoft Teams. (PHXP-29)