

Are you ready for the AI PC?

How can you best leverage AI PCs & smartphones?
Here are my dozen tips in alphabetical order.

A

*Are you ready for the AI PC?
Probably smarter than us, you see.
It can learn, create, and solve,
And maybe resolve, and evolve,
Sooner or later, it won't need you or me.*

If that limerick made you wink, this trend should make you think: There's this new kid in town, the AI PC, with specific SoC (system-on-a-chip) capabilities designed to run generative AI (GenAI) tasks locally. From almost zero sales as at Q4 2023, nearly 50 million AI PCs will be sold by Q4 2024, and more than 167 million by Q4 2027, according to IDC (International Data Corp) estimates.

What does "locally" mean? It means running tasks on-premise on the PC, as opposed to running them on the cloud. Chip manufacturers have now introduced AI-specific silicon to their SoCs – called NPUs (neural processing units) – that can run AI and ML (Machine Learning) tasks more efficiently.

Why does this matter? Because traditionally, running workloads – especially AI tasks locally on a PC – needed to be done on the CPU, the GPU, or a combination of the two. However, this might have a negative impact on the PC's performance and battery life since these chips are not optimized to run AI workloads efficiently.

NPU & TOPS

What's the need now? The hype around GenAI is high, and the PC industry is rushing to capitalize on

By Raju Chellam



the potential benefits of bringing AI capabilities from the cloud to the client. "Promises around enhanced user productivity via faster performance, lower inferencing costs and the benefit of on-device privacy and security have driven strong end-user interest in AI PCs," says Tom Mainelli, IDC's group vice president for consumer research. "We will see AI PC shipments begin to ramp up this year, and over the next few years we expect the tech to move from niche to majority."

IDC has identified three types of NPU-enabled AI PCs:

- Hardware-Enabled:** These run below 40 Tops (tera operations per second) and allow specific AI features within apps to run locally. The chips come from Qualcomm, Apple, AMD, and Intel.
- Next-Gen:** These AI PCs run between 40 and 60

 **APEEJAY EDUCATION**
Soaring High is My Nature
W: www.apeejay.edu
E: aes@apeejay.edu

Apeejay Styx Advantage

Quality education from pre-nursery to doctoral level

 **50+**
Years of excellence
in education

 **85+**
Programmes to
choose from

 **40,000+**
Students

 **26**
Educational institutions
across the country

 **65,000+**
Strong alumni
network

 **5,000+**
Educators &
staff members

Tops on an OS (operating system) that enables persistent and pervasive AI capabilities in both the OS and apps. The chips will likely debut later this year from Qualcomm, AMD, and Intel. Microsoft may likewise offer updates to Windows 11.

- **Advanced AI:** These are expected to run above 60 Tops in NPU performance. While no silicon vendors have announced advanced AI PCs, IDC expects them to appear by 2027.

Meanwhile, Gartner says global shipments of next-gen devices will reach an impressive 295 million units (including 240 million GenAI smartphones and 54.5 million AI PCs) by end-2024 – up from just 29 million units in 2023. Gartner defines AI PCs as ones equipped with NPUs and APUs (accelerated processing units) or TPUs (tensor processing units), designed to optimize AI tasks on the device.

“The rapid adoption of on-device GenAI capabilities and AI processors will eventually become a standard requirement for tech vendors,” says Ranjit Atwal, a Gartner senior director analyst. “This ubiquity will pose challenges for vendors in differentiating themselves from competitors, making it harder to create unique selling points and drive increased revenues.”

What’s the underlying catalyst for AI PCs? Smaller versions of LLMs (large language models) specifically tailored for smart devices. “This evolution will transform smartphones into even more intuitive companions capable of comprehending and responding to human language and visual cues, which will elevate the overall user experience to new heights,” Atwal says.

Is running workloads on the device better than on the cloud? Running AI tasks on the device and on the edge reduces the burden of running AI workloads on DCs (data centers) or on the cloud. Why does this matter? Because the annual global AI DC costs could reach US\$76 billion by 2028. If just 20% of GenAI processing workloads could be offloaded by running them on-premise, the global AI DC costs would

decline by as much as US\$15 billion, notes a study by Tirias Research.

PC & PHONE

That’s where the AI PC comes in. While the majority of GenAI development has been focused on the cloud, GenAI is rapidly evolving to be run directly on end-devices, including smartphones, PCs, vehicles, mixed reality and IoT devices, Wi-Fi access points, and others. This trend could also boost GenAI’s potential to accelerate digital transformation.

“One AI-generated image created in the cloud requires as much power as charging a smartphone,” a post at the World Economic Forum in Davos states. “To illustrate the energy efficiency of running AI on mobile devices, we tested a commercial smartphone and were able to generate more than 400 images on a single battery charge using an optimized AI model.”

Moreover, with on-device GenAI, apps can run continuously, learn more about their users, their preferences, behaviors, and also utilize complementary external data. This can enable focused and more relevant personalized responses. What about the flip side? Privacy and PII (personally identifiable info) could be compromised – and malware that fishes out PII will benefit hackers.

Which companies are the early birds in the AI PC and smartphone race? Here are the first four:

- **Intel:** With Alder Lake, the 12th generation of Intel Core processors. They’re designed to optimize AI workloads on PCs, such as speech recognition, image enhancement, and content creation. Intel is also partnering with Microsoft to enable AI capabilities on Windows 11 devices.
- **Google:** With its Gemini Nano smartphone powered by a Tensor chip that enables advanced GenAI features, such as voice synthesis, natural language understanding, and image generation. It also offers a dual-screen design that allows users to interact with two AI models simultaneously.



APEEJAY EDUCATION

Soaring High is My Nature

W: www.apeejay.edu
E: aes@apeejay.edu

Apeejay Styx Advantage

Quality education from pre-nursery to doctoral level

 50+
Years of excellence
in education

 85+
Programmes to
choose from

 40,000+
Students

 26
Educational institutions
across the country

 65,000+
Strong alumni
network

 5,000+
Educators &
staff members

- **Samsung:** With its Galaxy S22 series that features an Exynos 2200 chipset that integrates an AMD GPU and an NPU. Offers AI-enhanced photos, video editing, gaming, and multitasking on Galaxy S22 devices. Samsung's GenAI platform, Sam, can generate personalized content too.
- **Apple:** With its M1 Max and M1 Pro chips with 10-core CPUs, up to 32-core GPUs, and 16-core Neural Engine that can perform up to 15.8 Tops. The chips can enable AI-powered features, such as FaceTime Portrait Mode, Live Text, and Universal Control on the new MacBook Pro models.

TIPS & TRICKS

How can you best leverage AI PCs and smartphones? Here are my dozen tips in alphabetical order:

- **Authenticate:** By using devices that support encryption and authentication for your data and applications, primarily to prevent unauthorized access or tampering with your data, especially if you lose your device or connect to public networks.
- **Backup:** Your data frequently to a secure location, such as an external hard drive or a cloud service. This can help you recover your data in case of device failure, theft, or loss.
- **Choose:** Devices with dedicated AI processors or cores, such as NPUs, APUs, or TPUs. These can optimize AI workloads on the device, reducing the need for bandwidth and cloud resources.
- **Delete:** Unused or unwanted apps and services from your device to free up storage space and memory, thereby reducing the risk of data leakage or security breaches.
- **Enable:** Biometric and password protection for your devices and apps so that it adds an extra layer of security and prevents unauthorized use of your device or data.
- **Firewall:** Sensitive apps or data. Use a VPN or a firewall when connecting to the Internet to

encrypt your data and hide your online activity, especially when using public networks.

- **Get:** Your devices on auto-update mode with the latest software and security patches to protect your devices from malware, viruses, or hackers that may exploit vulnerabilities in your system.
- **Hide:** Workloads in devices under privacy mode or a sandbox feature so that you can run AI apps or models in a hidden environment without impacting your main system or data.
- **Invest:** In devices that have feedback or a rating system for AI apps. This can enable you to evaluate the accuracy and reliability of the AI outputs, and report issues to the vendors.
- **Juxtapose:** The permissions and settings for your apps and services to help you control which data your apps can access, collect, or share, and how they use your device's resources.
- **Kill:** Data or apps when required. Use devices that have a kill switch or a remote wipe feature to permit you to disable or erase your device remotely in case of theft, loss, or emergency.
- **Learn:** How AI decides by buying devices that have a transparency or an explainability feature to understand how the AI works, which data it uses, and the factors that influence its outputs.

And finally, since we started with a scary limerick on the AI PC, let's end with a cool one:

*The world is changing. Are you?
Ready to explore what's false, or true?
AI can help you illuminate,
Even if it does hallucinate,
The world's cool colors other than blue.* 

Raju Chellam is a former Editor of Dataquest and is currently based in Singapore, where he's the Editor-in-Chief of the AI Ethics & Governance Body of Knowledge, and Chair of Cloud & Data Standards.
mailto:dqindia@cybermedia.co.in





APEEJAY EDUCATION

Soaring High is My Nature

w: www.apeejay.edu
E: aes@apeejay.edu

Apeejay Styx Advantage

Quality education from pre-nursery to doctoral level

 50+ Years of excellence in education	 85+ Programmes to choose from	 40,000+ Students
 26 Educational institutions across the country	 65,000+ Strong alumni network	 5,000+ Educators & staff members