

Musings for the CIO: Enter the "AI PC"

Author

Marc Einstein

Senior Research Advisor
Focus Network



Musings for the CIO: Enter the "AI PC"

One of the biggest buzzwords which is already starting to infiltrate the IT industry in 2024 is the "AI PC". At CES 2024, many of the world's leading chip, PC and smartphone manufacturers touted the capabilities of Generative AI when processed at the hardware level. According to them, on-device AI will usher in a new era of computing by bringing AI capabilities closer to the end user and will unlock a range of benefits which will in theory make our interactions with AI much more efficient.

While it is not yet completely clear what these devices will do exactly, the concept of an "AI PC" is an interesting one. There are certainly many issues which have emerged from using cloud based LLMs, with perhaps the most important one being privacy and security at the enterprise level. Power consumption is another issue, given increased workloads datacenters will need, to generate content. CIOs will soon need to evaluate if these devices are worth purchasing hence, they should be on the CIO's radar already.

What Exactly Is an AI PC?

While vendors haven't been exactly clear about what AI PCs will do (nor how much they will cost including GenAI service usage), they have dropped a few hints which may give a look into their future. Key features proposed by vendors thus far include:

- **NPU**s - One aspect which will certainly define an AI PC is that it will need significantly more processing power to handle workloads locally. For this reason, Neural Processing Units or NPUs will become more common in the next generation of computing devices. Current PCs rely on Computing Processing Units (CPUs) and Graphics Processing Units (GPUs) to process data but NPUs are built to mimic neural activity and reduce the burden on CPUs and GPUs. NPUs are expected to become a standard feature in PCs (and other devices like smartphones) which will result in much more robust computational devices.
- **More Memory** - Running an LLM locally on a device as opposed to using the cloud will increase the amount of memory devices need. For most PCs, 4GB or 8GB are sufficient for tasks such as web browsing and document creation. However, to run Generative AI well, the minimum specification may increase to 16GB or even 32GB, which today is used for devices which are used for gaming and intensive graphic design. This means that users who want to heavily use Generative AI may need a hardware upgrade in the near future.



- **Dedicated Button** - One of the headlines at CES 2024 in Las Vegas was the announcement from Microsoft that the next generation of PCs will have a dedicated CoPilot button. This is the first change to Windows keyboards in 30 years since the introduction of the Windows key. This move is indicative of how Microsoft sees its Generative AI service as integral to the future of computing.
- **New Software Subscriptions** - Something which will generate less enthusiasm in the industry is to access the cutting-edge features of an AI PC may require a new software subscription. Microsoft for example recently announced that PC users will need to pay an extra A\$37 per month to access CoPilot Pro on its computers. The service, however, may well be worth it for some enterprise users as CoPilot is integrated with Word, Excel and PowerPoint and can write emails, make slides, and generate graphs, saving considerable amounts of employee time.

Exhibit: Other 2024 Trends Impacting the CIO

AI PC Features

AI PC Feature	Change from Standard PC	Impact
NPU	Previously CPU & GPU Based	New Services
Increased Memory	Less than 16GB Needed	Higher Performance
Dedicated Button	No Dedicated Button	Increased Ease of Use
Subscription	No Subscription Previously Required	Increased Cost

Source: Focus Network.

What Impact Will the AI PC Have in the Enterprise?

Regardless of the exact specifications of AI PCs, running Generative AI locally will bring benefits to the enterprise. Some of the potential uses include the following:



- **Cybersecurity** – AI PCs will help to make the enterprise more secure by utilising Generative AI locally. Many functions of the technology such as coding should not be done in a cloud environment (some very large tech companies have had their source code uploaded into ChatGPT) and therefore tools such as low code/zero code application creation will become more widespread in the industry.
- **Healthcare** – Very sensitive information such as individual healthcare information should also not be uploaded in a cloud environment due to liability and often regulatory concerns. Therefore, using AI PCs to examine healthcare information such as analysing patient X-rays on-device both reduces risk and also potentially can provide faster and more accurate results.
- **Content Creation** – Creating and editing video content is heralded as a key use case for Generative AI, but doing this in the cloud consumes a lot of time and data. Migrating this to an on-device environment will result in both lower costs and faster processing time.
- **Transportation** – One of the key challenges with autonomous driving for the past several years has been that a cloud environment cannot respond fast enough to real world conditions. Having on-device AI could change this once these computing devices can be integrated into vehicles, drones, and autonomous robots.

Exhibit: AI PC Benefits

AI PC Benefits

Industry/Segment	AI PC Application	Impact
Cybersecurity	Coding	Increased Security
Healthcare	Medical Imaging	Increased Privacy
Content Creation	Video Editing	Lower Costs
Transportation	Autonomous Driving	Faster Response Time

Source: Focus Network.



Summary & Conclusions

2023 will likely be remembered as the year of Generative AI, and 2024 may well be remembered as the year of the AI PC. While it remains unclear as to exactly what an AI PC is, it is already apparent that PCs (and eventually smartphones and other devices) will have the ability to natively process Generative AI applications. In order to do this, PCs will need to be much more robust in terms of performance. The world's leading chip and device manufacturers have already indicated that this is the direction where they are heading.

AI PCs will also bring considerable benefits to the enterprise, as on-device AI will offer several benefits including better security, lower bandwidth costs and faster delivery times. Specific industries where these issues are particularly challenging are good prospects for early adoption. AI PCs will undoubtedly first arrive with premium pricing, and hence the CIO will need to decide in the near future when and if these next-generation devices should be introduced to their organisation.



About Focus Network

Focus Network is a data-driven networking, research and advisory hub for senior executives to share their insights and accelerate their learning, across the Asia Pacific region. We enable business and technology leaders to implement best practices and optimise their investments, by providing unparalleled insights and reporting, complemented with local, country-level, perspectives.

www.focusnetwork.co

Disclaimer

Whilst reasonable efforts have been made to ensure that the information and content of this product was correct as at the date of first publication, neither Focus Network nor any person engaged or employed by Focus Network accepts any liability for any errors, omissions or other inaccuracies.

Readers should independently verify any facts and figures as no liability can be accepted in this regard - readers assume full responsibility and risk accordingly for their use of such information and content.

Any views and/or opinions expressed in this product by individual authors or contributors are their personal views and/or opinions and do not necessarily reflect the views and/or opinions of Focus Network.



www.focusnetwork.co