WHAT IS AN AI PC?

THE PERSONAL AI REVOLUTION IS NOW UNDERWAY

INTRODUCTION

In the late 1990s, the computer industry radically changed the PC platform by introducing the Internet PC, shifting the focus from standalone devices to those optimized for the World Wide Web. This shift opened up new applications, including e-commerce, search, and mass adoption of e-mail.

It also led to refocusing the underlying chips and software to support faster communications (especially Wi-Fi and Ethernet), graphics, and video processing to support richer online content. Over time, the PC has become more of a portal into cloud-based applications and other Internet-based functions, largely reserving onboard processing for advanced compute-intensive activities such as content creation, gaming, and offline productivity applications.

On May 20, 2024, Microsoft and its semiconductor and OEM partners announced a new category of AI PCs called Copilot+ PCs. This new paradigm, which may change the PC platform even more significantly than the Internet PC did, is nothing short of a ground-up reimagining of both the silicon and software stack needed to bring PCs fully into the age of AI. Copilot+ PCs shift the emphasis back to local processing on the PC platform to create an intelligent personal device that can co-create, mentor, and advise individuals in almost every aspect of their lives.

WHAT DEFINES AN AI PC?

As the name suggests, an AI PC is a laptop or desktop computer that has a set of hardware components capable of processing AI models locally on the device. The first AI PCs started shipping in early 2024; these included both a discrete neural processor (NPU) to offload AI workloads and a Microsoft Copilot keyboard key to qualify them under Microsoft guidelines as AI PCs. The systems shipped with fairly mundane NPU capabilities powered by either Intel or AMD processors. These ranged in speed from 8 to 15 TOPS and provided minimal additional OS enhancements, mostly in Windows Studio Effects functions and beta versions of some third-party software.

This brings us to Microsoft's Copilot+ PC. At the core of this next-generation AI PC is a more robust set of minimum hardware required to run the newest version of Windows

11 and perform more significant AI functions. This includes a significantly higher minimum TOPS threshold of 40 for the dedicated NPU, along with a minimum 16GB of RAM and 250GB of storage. The increase in required TOPS is meant to ensure a seamless user experience by offloading most AI functionality to the power-efficient NPU and leaving the CPU and GPU free to process other functions in parallel. This eliminates lag for the user and creates the ability for the NPU to run more advanced AI models extremely quickly — and with advanced privacy features keeping sensitive AI data contained on the device.

While labeling earlier models as "AI PCs" may have been intended to build market momentum for PCs with AI capabilities, it has now created potential consumer confusion between lower-powered devices and the new Copilot+ PCs, which allow AI-enabled applications to take advantage of more than 40 on-device AI models built into the new version of Windows 11.

There is a simple way to understand whether a PC meets the Copilot+ criteria and will properly run the AI-native version of Windows: the processor brand. For the next several months, only Qualcomm's Snapdragon X Elite and Snapdragon X Plus processors, both stated to run at 45 TOPS, will meet the minimum 40 TOPS requirement for the NPU. This puts Qualcomm in a favorable position relative to Intel and AMD, at least for now. Intel and AMD's next-generation processors, which are expected to climb above 45 TOPS, will arrive later this year, with at least some anticipated for the 2024 holiday season. However, Microsoft has not confirmed a launch date for the updated Copilot+ runtime required to support those x86-based processors.

AN AI PC NEEDS AN AI OS: THE BREAKTHROUGH OF COPILOT+

With more clarity now on what an AI PC is, what can it really do? This is where the industry — and end users — must confront a chicken-or-egg dilemma between hardware capabilities and the investments made by software developers. For many developers, it is hard to justify prioritizing investment in brand-new functionality such as on-device AI without a large installed base of devices that can capitalize on it. In the case of Copilot+ PCs, there have not yet been the necessary NPUs available to get the most out of the new AI features in Windows 11.

However, Microsoft's dominant share in the PC OS market has allowed it to set the bar for the necessary hardware while strategically committing itself to integrating on-device AI via its Windows Copilot Runtime. That solves the chicken-or-egg problem by assuring software developers a large enough target market for new AI-enabled wares.

Moor Insights & Strategy (MI&S) believes that many developers will be quick to respond to what should be a substantial market opportunity by producing versions of their apps that make the most of Copilot+ PC specs and Windows 11 AI functionality. Developers and ISVs that have not already started down that road may find themselves lagging the market by late 2024 or early 2025.

Windows Copilot Runtime includes building blocks at every layer of the OS for developers to integrate AI capabilities more easily into their applications. These include toolchains and frameworks such as PyTorch and WebNN, along with the Copilot Library of more than 40 AI models bundled into the Windows OS. These two layers remove much of the friction so developers can more easily create the top layer of AI applications and experiences.

This will not only improve software already in use, but will likely inspire new types of applications that maximize AI functionality on the PC. One notable example of this is the Recall feature in the new version of Windows 11. Through this feature, Copilot+ PCs will automatically take screenshots every five seconds, enabling users to come back later with natural language queries to find anything ever done on those PCs. While Recall is impressively simple to use based on the demo shown by Microsoft, it is built with some serious engineering under the hood. Recall runs Microsoft's proprietary Phi Silica small language model in conjunction with the new integrated Windows Semantic Index that accesses the screen captures. These captures are then indexed into a vector database for use by the built-in retrieval-augmented generation (RAG) capabilities. All of this happens in the background, primarily using the power-efficient NPU.

Recall could prove to be especially powerful, even transformational, because it enables the PC to be an extension of the user's memory, providing instant and accurate recall of information gathered, communications delivered, photos stored, and work done. Given the sensitive nature of what an individual does on their PC, Recall is a feature that highlights the privacy concerns associated with AI and the importance of having all this information encrypted, stored, and processed privately on the device.

Another impressive application of the Windows Copilot Runtime is its ability to co-create content. In its demo, Microsoft showed how a user can take advantage of new AI-driven functionality in the built-in Paint program to augment a basic sketch. As the image was taking shape, the user described what they intended to paint and asked the Copilot assistant for help to boost the level of artistic augmentation, which the AI applied instantly. This took the image from a more abstract, lower-quality representation to one that was more realistic and detailed.

AN AI OS NEEDS BROAD AI APP SUPPORT

The AI features in Windows alone likely won't be enough to ensure the mass success of Copilot+ PCs. There must also be a broad array of reimagined software made more useful by leveraging on-device AI. Many of today's most popular applications are already starting to integrate AI functionality, but they are doing this principally through their cloud-based offerings.

Take Adobe Firefly as an example. As of March 2024, Adobe Creative Cloud users had made more than 7 billion images using Firefly's cloud-based generative AI (GenAI) model. At first glance, this is exciting news for Adobe because it represents great traction for this new functionality. However, GenAI in the cloud requires massive compute resources, estimated at about 10x the operating cost of traditional compute for cloud applications. Yet if Adobe can leverage the local AI compute power of an AI PC — especially a Copilot+ PC — to run the Firefly model on-device, Adobe's costs go from 10x historical cloud costs to zero. This is an enormous economic incentive for Adobe (and any similarly situated software maker) to invest development resources into tuning its models to run locally on AI PCs, even ahead of mass deployment of those AI PCs in the market.

We are already seeing this trend underway, with several content-creation applications as early movers. Luminar Neo, for example, has implemented on-device AI in its photo editing software that dramatically speeds up current features. Whereas a non-AI PC can take two minutes to sharpen an image, a Qualcomm X Elite–powered AI PC finishes it in eight seconds. Another example is DaVinci Resolve video editing software, which includes more than 100 AI-accelerated features in its latest release. Popular features such as Magic Mask can now run faster and more efficiently on the NPU as compared to a GPU, making advanced video editing possible on a battery-powered AI PC laptop.

WHY CONSUMERS NEED AN AI PC—SPECIFICALLY A COPILOT+ PC

While many of the applications announced so far are interesting, none reach the bar of a "killer app" that drives insatiable demand for an AI PC. However, the inevitability of the best new features and applications requiring a powerful NPU on-device will likely make an AI PC an absolute requirement for PC replacement purchases going forward. While Copilot+ PCs' cool new features will draw many consumers, one major factor in consumer behavior around technology is the fear of obsolescence. After all, computers are still a fairly significant investment, typically between $1,000 and $2,000, making them a highly considered purchase for most buyers.

This reality played out 25 years ago in the platform transition to the Internet PC. That transition came before the ubiquitous availability of high-speed Internet or Wi-Fi, but the rollout of broadband was inevitable, and PC purchasers feared that their computers would not be capable of connecting to, and making the most of, services that required an Ethernet port. Tens of millions of Internet PCs were sold with Ethernet ports that remained mostly unused, sometimes for years, until high-speed Internet was universally available.

The Copilot+ PC is similarly positioned, but with a much larger potential scale of transformation in terms of user experience. It also has a major advantage in the immediate usefulness of the NPU, the OS, and enough early applications to make it a checklist item for nearly all new PC purchases. With this in mind, consumers will need to scrutinize any NPU's performance as measured in TOPS and prioritize having the fastest NPU available at the time of purchase. At this writing, there is broad support from PC brands — including market leaders Dell, HP, and Lenovo — which will offer more than 20 models that meet Copilot+ PC specifications in mid-June 2024; all of these will ship with some version of the Qualcomm Snapdragon X family of processors.

These systems claim to be up to 100x faster at AI processing than previous AI PCs and [to outperform in most CPU, GPU, and battery-life benchmarks](#) against almost every current notebook in their class, including the much-lauded Apple M3-based MacBook Pro. Qualcomm's long history in designing Arm-based processors drives much of the raw performance and power efficiency.

But these benefits do come with a possible caveat. While Windows has supported the Arm instruction set for many years, most Windows-based software is natively supported only by x86 architecture processors, which are primarily supplied by Intel and AMD. To ensure compatibility, hundreds of today's leading software applications have been updated to support Arm natively. In fact, Microsoft claims that 87% of total minutes spent on apps today are on apps with native Arm versions.

It's also important to note that there is a solution to compatibility issues even for apps that lack Arm-native versions. For example, many video games — along with a huge library of older apps — have never been ported into Arm versions. Yet most of these can be handled seamlessly by the new version of Microsoft's Prism emulator, which allows nearly every traditional game and application to run on Arm with minimal or no performance impact. Microsoft has invested significant resources into speeding up this emulation mode over the past few years, achieving a 2x to 3x improvement over the previous version. In fact, benchmark tests have shown that Prism emulation of Arm is at

What Is an AI PC?

least on par with Apple's Rosetta Windows emulation, if not better in some cases. With this in mind, it is reasonable to believe that most users will never experience any negatives — and possibly never notice anything different at all — while using Arm-based systems versus x86 alternatives.

FINAL THOUGHTS

The Internet PC radically changed the use cases for personal computing by enabling native access to the Internet. The next generation of the AI PC — the Copilot+ PC — will have an even more transformative impact on PC use cases. This is made possible by the complete reimagining of the PC for the AI era, from a purpose-built discrete AI processor in the form of an NPU to an all-new Windows OS that has AI interwoven at every layer. The result is that, for the first time, the PC will be truly *personalized* to the user. It will even begin to develop a "persona" uniquely crafted for each individual and adapted to their particular needs. It can analyze, reason, and help users in the most personalized ways to do nearly whatever is desired. It will even forget the things a user doesn't want it to remember.

The entire industry is rallying around the new category of Copilot+ PCs. For consumers ready to dive in, there are choices available from nearly every leading PC brand; all of these feature the Snapdragon X series, whose 45 TOPS handily exceeds the minimum required AI processing power while also delivering impressive overall compute performance — and some of the longest battery life ever for a Windows laptop. Indeed, some consumers may buy Copilot+ PCs this year even without a specific interest in AI functionality, simply because these laptops should deliver high performance and great efficiency in extremely thin form factors.

In some ways, the Copilot+ PC may be the true fulfillment of what many envisioned the "personal computer" would someday become 40-plus years ago when the category was just in its infancy.

IMPORTANT INFORMATION ABOUT THIS PAPER

CONTRIBUTOR(S)
Michael Rubin, Chief Strategy Advisor

PUBLISHER
Patrick Moorhead, CEO, Founder and Chief Analyst at Moor Insights & Strategy

INQUIRIES
Contact us if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

CITATIONS
This paper can be cited by accredited press and analysts but must be cited in-context, displaying author's name, author's title, and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

LICENSING
This document, including any supporting materials, is owned by Moor Insights & Strategy. This publication may not be reproduced, distributed, or shared in any form without Moor Insights & Strategy's prior written permission.

DISCLOSURES
Qualcomm commissioned this paper. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

DISCLAIMER
The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.

©2024 Moor Insights & Strategy. Company and product names are used for informational purposes only and may be trademarks of their respective owners.