On-device AI: Paving the way for the next AI revolution

Published 10 June 2024

Baoqi Zhu
Associate Director, Quantitative Research & Multi Asset Solutions

Key Takeaways

- On-device AI is implemented directly on local devices without accessing remote servers, enhancing data privacy and reducing latency and operational costs.
- On-device AI lowers the barrier for individual customers to access AI, which could transform the AI ecosystem, expanding AI applications from mainly business to daily life and making AI more accessible.
- AI-related companies across semiconductors, AI electronics, and AI application providers stand to benefit from this trend.

Microsoft recently introduced its new Surface laptops integrated with Artificial Intelligence (AI), reigniting interest in AI PCs in the market. Besides Microsoft, other major PC manufacturers have also announced plans to release AI PCs integrated with Microsoft's new Copilot Plus. The AI PCs with system-on-a-chip (SoC) capabilities are designed to run generative AI tasks locally, also known as 'on-device AI'. A forecast[1] shows that AI PCs with specific SoC capabilities designed to run generative AI tasks without accessing the cloud will grow from nearly 50 million units in 2024 (while the worldwide market is 250 million units) to more than 167 million in 2027.

What is on-device AI?

On-device AI refers to the implementation and execution of AI models directly on local devices such as smartphones, tablets, laptops, etc., without accessing remote servers. To deploy AI on these devices, the trained AI models are compressed to make them suitable for resource-constrained environments. The models are then typically converted to neural network graphs to ensure they are optimised for execution on specific hardware and AI frameworks.

Features of on-device AI

While cloud-based generative AI, represented by ChatGPT, has demonstrated impressive capabilities, on-device AI still holds unique potential due to its distinct features:

- **Data privacy:** With on-device AI, users can implement AI models without accessing remote servers and data is processed locally on the device. This eliminates the need to send sensitive information to the cloud, reducing the risk of data breaches and significantly enhancing user privacy and data security.
- **Low latency:** Since the processing occurs on the device, the time taken to transmit data between devices and remote servers is avoided. This feature provides faster response times, which is critical for real-time applications.
- **Low operational cost:** For cloud-based AI providers, increasing energy and bandwidth costs are vital factors for pricing. By eliminating data transport and transferring processing to local devices, AI service providers can offload part of the data centre's energy consumption and reduce bandwidth costs.

Is on-device AI a game changer?

As on-device AI technologies develop, more digital devices will be integrated with real-time AI, allowing users to access AI solutions through smartphones, laptops, wearable devices, smart home devices, AR/VR headsets, etc. AI will be more accessible for individuals due to its advantages in privacy, latency, and offline functionality. This will greatly expand AI's application scenarios from mainly business to daily life, significantly increasing the number of potential AI users and potentially changing the consumer electronics ecosystem.

Moreover, AI service providers will have more pricing flexibility due to the lower operational cost. By balancing AI services between on-device AI and cloud-based AI, providers can optimise their business models – using cloud-based AI for services requiring more scalability and complexity and on-device AI for services needing low latency and privacy. The current data centre model could be partially replaced by a more flexible business model, enabling smaller companies focused on specialised AI applications to gain more opportunities.

Companies benefitting from the on-device AI trend

- **Leading chip manufacturers:** On-device AI requires high-performance processors. Neural Processing Units (NPUs) are hardware accelerators designed to execute AI tasks efficiently. They play a crucial role in enhancing the performance and efficiency of on-device AI. The increasing demand for NPUs could benefit several chip manufacturers. **Qualcomm** is a leader in this area with its Snapdragon X Elite and X Plus processors, which are the only processors that meet Microsoft's performance requirements (>40 trillion operations per second) for deploying new Copilot Plus on laptops. Many laptop makers, including Lenovo, HP, and Dell, have chosen Qualcomm's chips for their laptops that are integrated with new Copilot Plus. Although **Intel** and **AMD** are lagging behind Qualcomm in NPU performance, they are catching up. Intel plans to launch its next-generation processor, Lunar Lake, in Q3 2024, which is expected to compete with Qualcomm's current lead. AMD also plans to launch a processor with higher NPU performance later this year. The competition among these chip giants in the NPU space is worth monitoring.

- **IP cores and EDA tool providers:** Companies providing Intellectual Property (IP) cores and Electronic Design Automation (EDA) tools could also benefit from the on-device AI trend. EDA tools streamline the entire design process from initial concept to final implementation, ensuring that the NPU is efficient, reliable, and manufacturable. IP cores offer pre-designed, tested, and optimised components that accelerate development, enhance performance, and reduce risks. **Synopsys** and **Cadence Design Systems** are two leading companies in these areas and could be boosted by the competition in NPUs.
- **Consumer electronics giants: Apple** and **Microsoft** are pioneers in integrating on-device AI into their main products. Apple also designed its own NPU-accelerating processor, M4, to enhance product performance.

Conclusion

On-device AI is an emerging area poised to revolutionise AI applications. Its key features, including data privacy, low latency, and low operational cost, differentiate it from cloud-based AI, which holds the potential to transform the AI ecosystem. In the short term, this trend will benefit AI-related companies across semiconductors, AI electronics, and AI application providers. Looking further ahead, it will expand AI application scenarios in daily life, significantly lower the barrier for individual customers to access AI, and potentially make AI available to everyone.

[1] IDC: IDC Forecasts Artificial Intelligence PCs to Account for Nearly 60% of All PC Shipments by 2027.

Important Risks Related to this Article

IMPORTANT INFORMATION

Marketing communications issued in the European Economic Area ("EEA"): This document has been issued and approved by WisdomTree Ireland Limited, which is authorised and regulated by the Central Bank of Ireland.

Marketing communications issued in jurisdictions outside of the EEA: This document has been issued and approved by WisdomTree UK Limited, which is authorised and regulated by the United Kingdom Financial Conduct Authority.

WisdomTree Ireland Limited and WisdomTree UK Limited are each referred to as "WisdomTree" (as applicable). Our Conflicts of Interest Policy and Inventory are available on request.

For professional clients only. Past performance is not a reliable indicator of future performance. Any historical performance included in this document may be based on back testing. Back testing is the process of evaluating an investment strategy by applying it to historical data to simulate what the performance of such strategy would have been. Back tested performance is purely hypothetical and is provided in this document solely for informational purposes. Back tested data does not represent actual performance and should not be interpreted as an indication of actual or future performance. The value of any investment may be affected by exchange rate movements. Any decision to invest should be based on the information contained in the appropriate prospectus and affer seeking independent investment, tax and legal advice. These products may not be available in your market or suitable for you. The content of this document does not constitute investment advice nor an offer for sale nor a solicitation of an offer to buy any product or make any investment.

An investment in exchange-traded products ("ETPs") is dependent on the performance of the underlying index, less costs, but it is not expected to match that performance precisely. ETPs involve numerous risks including among others, general market risks relating to the relevant underlying index, credit risks on the provider of index swaps utilised in the ETP, exchange rate risks, interest rate risks, inflationary risks, liquidity risks and legal and regulatory risks.

The information contained in this document is not, and under no circumstances is to be construed as, an advertisement or any other step in furtherance of a public offering of shares in the United States or any province or territory thereof, where none of the issuers or their products are authorised or registered for distribution and where no prospectus of any of the issuers has been filed with any securities commission or regulatory authority. No document or information in this document should be taken, transmitted or distributed (directly or indirectly) into the United States. None of the issuers, nor any securities issued by them, have been or will be registered under the United States Securities Act of 1933 or the Investment Company Act of 1940 or qualified under any applicable state securities statutes.

This document may contain independent market commentary prepared by WisdomTree based on publicly available information. Although WisdomTree endeavours to ensure the accuracy of the content in this

document, WisdomTree does not warrant or guarantee its accuracy or correctness. Any third party data providers used to source the information in this document make no warranties or representation of any kind relating to such data. Where WisdomTree has expressed its own opinions related to product or market activity, these views may change. Neither WisdomTree, nor any affiliate, nor any of their respective officers, directors, partners, or employees accepts any liability whatsoever for any direct or consequential loss arising from any use of this document or its contents.

This document may contain forward looking statements including statements regarding our belief or current expectations with regards to the performance of certain assets classes and/or sectors. Forward looking statements are subject to certain risks, uncertainties and assumptions. There can be no assurance that such statements will be accurate and actual results could differ materially from those anticipated in such statements. WisdomTree strongly recommends that you do not place undue reliance on these forward-looking statements.

WisdomTree Issuer ICAV

The products discussed in this document are issued by WisdomTree Issuer ICAV (“WT Issuer”). WT Issuer is an umbrella investment company with variable capital having segregated liability between its funds organised under the laws of Ireland as an Irish Collective Asset-management Vehicle and authorised by the Central Bank of Ireland (“CBI”). WT Issuer is organised as

an Undertaking for Collective Investment in Transferable Securities (“UCITS”) under the laws of Ireland and shall issue a separate class of shares ("Shares”) representing each fund. Investors should read the prospectus of WT Issuer (“WT Prospectus”) before investing and should refer to the section of the WT Prospectus entitled 'Risk Factors' for further details of risks associated with an investment in the Shares.

WisdomTree Artificial Intelligence UCITS ETF

Nasdaq® and the Nasdaq CTA Artificial intelligence Index are registered trademarks of Nasdaq, Inc. (which with its affiliates is referred to as the “Corporations”) and are licensed for use by WisdomTree Management Limited. The WisdomTree Artificial Intelligence UCITS ETF (the “Fund”) has not been passed on by the Corporations as to its legality or suitability. Shares in the Fund are not issued, endorsed, sold, or promoted by the Corporations. THE CORPORATIONS MAKE NO WARRANTIES AND BEAR NO LIABILITY WITH RESPECT TO THE FUND.

Notice to Investors in Switzerland – Qualified Investors

This document constitutes an advertisement of the financial product(s) mentioned herein.

The prospectus and the key investor information documents (KIID) are available from WisdomTree's website: **https:** //www.wisdomtree.eu/en-ch/resource-library/prospectus-and-regulatory-reports

Some of the sub-funds referred to in this document may not have been registered with the Swiss Financial Market Supervisory Authority (“FINMA”). In Switzerland, such sub-funds that have not been registered with FINMA shall be distributed exclusively to qualified investors, as defined in the Swiss Federal Act

on Collective Investment Schemes or its implementing ordinance (each, as amended from time to time). The representative and paying agent of the sub-funds in Switzerland is Société Générale Paris, Zurich Branch, Talacker 50, PO Box 5070, 8021 Zurich, Switzerland. The prospectus, the key investor information documents (KIID), the articles of association and the annual and semi-annual reports of the sub-funds are available free of charge from the representative and paying agent. As regards distribution in Switzerland, the place of jurisdiction and performance is at the registered seat of the representative and paying agent.

For Investors in France

The information in this document is intended exclusively for professional investors (as defined under the MiFID) investing for their own account and this material may not in any way be distributed to the public. The distribution of the Prospectus and the offering, sale and delivery of Shares in other jurisdictions may be restricted by law. WT Issuer is a UCITS governed by Irish legislation, and approved by the Financial Regulatory as UCITS compliant with European regulations although may not have to comply with the same rules as those applicable to a similar product approved in France. The Fund has been registered for marketing in France by the Financial Markets Authority (Autorité des Marchés Financiers) and may be distributed to investors in France. Copies of all documents (i.e. the Prospectus, the Key Investor Information Document, any supplements or addenda thereto, the latest annual reports and the memorandum of incorporation and articles of association) are available in France, free of charge at the French centralizing agent, Societe Generale at 29, Boulevard Haussmann, 75009, Paris, France. Any subscription for Shares of the Fund will be made on the basis of the terms of the prospectus and any supplements or addenda thereto.

For Investors in Malta

This document does not constitute or form part of any offer or invitation to the public to subscribe for or purchase shares in the Fund and shall not be construed as such and no person other than the person to whom this document has been addressed or delivered shall be eligible to subscribe for or purchase shares in the Fund. Shares in the Fund will not in any event be marketed to the public in Malta without the prior authorisation of the Maltese Financial Services Authority.

For Investors in Monaco

This communication is only intended for duly registered banks and/or licensed portfolio management companies in Monaco. This communication must not be sent to the public in Monaco.