AI PC 101

Executive Summary

AI PCs represent a transformative shift in the way organizations harness artificial intelligence, combining high-performance local computing with seamless integration into hybrid AI environments. This paper explores AI PC foundations and offers a comprehensive overview of how they fit into modern enterprise technology strategies. AI PCs and their role in a Hybrid AI ecosystem open new ways for organizations to prioritize operational efficiency, real-time decision-making, and enhanced security. This integration of on-device AI processing with the scalability of cloud infrastructure opens new possibilities for businesses to optimize workflows, scale their AI-driven operations, and reduce costs.

As AI PCs continue to evolve, they will play an integral role in future AI innovations. Projections suggest that by 2027, nearly 60% of all PC shipments will be AI-enabled, with AI PCs expected to dominate enterprise IT footprint as organizations increasingly rely on real-time processing at the edge[1]. This surge in adoption is being driven by the growing demand for low-latency AI applications, the integration of AI accelerators, and the ability of AI PCs to manage workloads autonomously. Independent Software Vendors (ISVs) will also play a pivotal role in this shift, by developing AI-optimized software that leverages the hardware accelerators in AI PCs and unlocking new use cases that maximize the value of enterprise AI investments.

This paper provides IT decision-makers (ITDMs) with an initial overview of the AI PC ecosystem, by introducing AI PC advantages, use cases, and a technology stack deep dive. It also explores how organizations can measure the value of AI PCs and evaluate whether they should invest in this compelling technology that has started and will continue to reshape enterprise computing.

Introduction to AI PCs

AI PCs are the latest generation of computing devices, designed to perform specialized AI tasks locally (i.e., on the device itself). Unlike traditional systems, which primarily rely on cloud servers for processing, AI PCs incorporate specialized hardware, such as AI-enhanced processors (e.g., NPUs), to manage these operations. AI powered software, such as Copilot type applications, enable the on-device processing that transforms the PC from a passive endpoint into an intelligent agent capable of dynamically managing and executing complex tasks with greater efficiency.

AI PCs are also valuable in how they enhance cloud infrastructure. Processing data locally leads to lower latency, faster insights, and real-time decision-making. This complement to the cloud also gives business more autonomy in managing AI workloads and strengthens data security by keeping sensitive information on the device. Cloud computing however remains essential for large-scale data processing and storage, and AI PCs enhance these capabilities by offloading certain tasks locally.

Advantages of AI PCs

In a world where speed, security, and flexibility are paramount, AI PCs offer enterprises the tools they need to address pressing challenges like real-time customer personalization, advanced predictive maintenance and heightened data protection. These advantages align to three key pillars[2]:

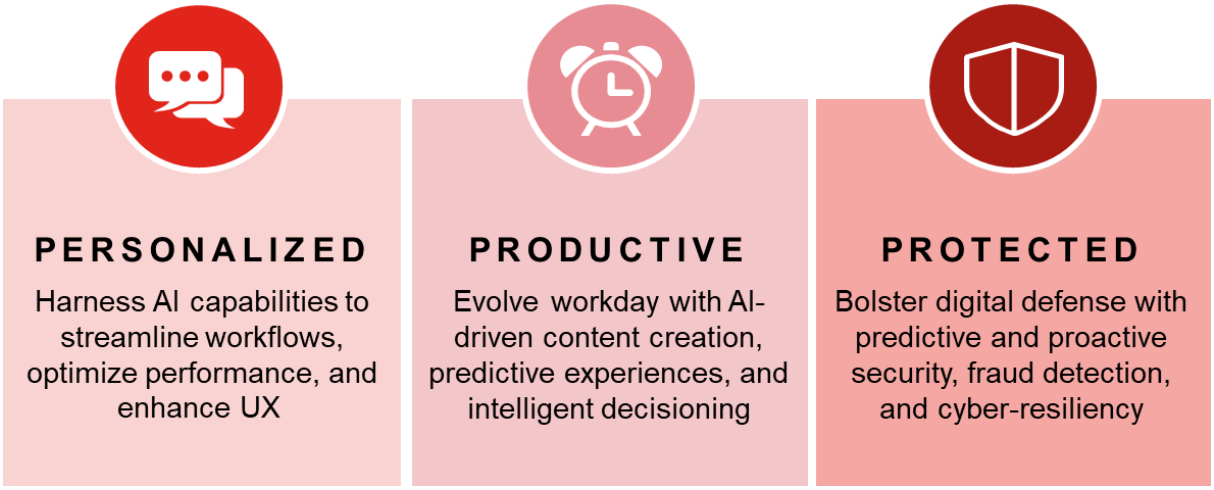


Figure 1: AI PC Ecosystem Play

- **Personalization:** AI PCs offer highly customized, user-specific computing experiences through advanced machine learning and large language models (LLMs) that adapt to individual behaviors, preferences, and work styles. By leveraging LLMs and Copilot-style capabilities, AI PCs can support users with

- contextually relevant responses, proactive insights, and intelligent recommendations, all of which enhance productivity and decision-making. For example, real-time Natural Language Processing (NLP) built into AI PCs facilitates understanding of user inputs, while Copilot+ features provide contextual assistance across workflows, enabling seamless access to insights and personalized suggestions. Examples of Personalization include Personalized Responses, Contextual Search, Meeting Assistant, Personal Chatbot, Smart Notifications, Intelligent Automation & Reporting.
- **Productivity/Efficiency:** AI PCs significantly boost productivity by enhancing operational efficiency with local AI processing, reducing reliance on cloud infrastructure and minimizing latency. With built-in AI accelerators like GPUs and NPUs, these devices handle AI inference tasks up to 10x faster than traditional CPUs, enabling quicker decision-making and more responsive workflows. By processing AI workloads directly on the device, AI PCs reduce data transfer costs and dependency on expensive cloud resources, delivering substantial savings. This local computation model not only enhances speed but also achieves cost-effective scalability, allowing businesses to support AI workloads without incurring high cloud expenses. The energy efficiency of NPUs further reduces power consumption, contributing to lower operational costs while maintaining high performance. Examples of productivity-enhancing capabilities include Design & Writing Assistants, Smart Business Intelligence, Enhanced Audio/Video Collaboration, and Summarization.

- **Protection:** AI PCs offer enhanced security by performing data processing locally, which reduces the risk of breaches during cloud transmission. In industries with stringent compliance requirements like financial services or healthcare, keeping sensitive data on the device (or on the edge) is crucial for maintaining privacy and security. AI-driven anomaly detection algorithms embedded in AI PCs can monitor real-time system behavior, quickly identifying potential security breaches or vulnerabilities. Lower latency is another key advantage of AI PCs, particularly when it comes to detecting and responding to threats. This ensures faster detection and mitigation of security risks. Additionally, local processing improves data privacy, making it easier to comply with regulations such as GDPR or HIPAA. Examples of AI-driven security features include
  - Proactive Security, Anomaly Detection, Advanced Threat Detection, Better Cyber Resiliency, and PC Lifecycle Management.

AI PC Use Cases

AI PCs are transforming industries through localized, real-time processing, secure decision-making, and efficient AI workload management. While many of these use cases represent future possibilities, some capabilities are already live today, with ongoing advancements set to expand the potential of AI PCs over the next 12–18 months. Here, we break down current and future use cases across sectors:

Current Use Cases (Live Today):

- **Co-Pilot for Knowledge Workers:** AI PCs currently enable tasks such as summarization, Q&A, and content generation. Knowledge workers can automate routine activities, such as document creation, email management, and

- scheduling, with minimal latency. On-device processing ensures operations happen instantly without relying on the cloud, enhancing both speed and privacy. With heterogeneous computing (CPUs, GPUs, NPUs), AI PCs optimize task execution while providing robust security features like on-device encryption, safeguarding sensitive data.
- **IT Support:** AI PCs enhance IT support by managing real-time diagnostics, troubleshooting, and automated helpdesk tasks, such as password resets and software updates. With on-premises processing, AI PCs improve security and privacy, making them ideal for sensitive industries while reducing cloud-related costs. Current capabilities already include faster response times and improved operational efficiency for IT teams, allowing them to focus on more complex challenges.
- **Code Generation and Development Support:** AI PCs empower developers by supporting complex coding tasks, leveraging local AI capabilities to enhance productivity in code generation, debugging, and software development. With AI-driven tools—whether integrated as Copilot+ features or other developer assistants—Lenovo hardware enables efficient, responsive development workflows. By performing demanding computations locally, AI PCs reduce latency and reliance on cloud resources, providing developers with fast, reliable support for coding, testing, and deployment.

Future Use Cases (Live in 12-18 Months):

- **Healthcare Diagnostics (In Development):** While the use of AI PCs to process complex medical imaging data (e.g., MRI, CT scans) locally shows immense potential, this capability is still in the early

- stages of adoption. In the near future, hospitals may leverage NPUs for rapid on-device image analysis, enabling early detection of conditions like tumors without the need for cloud uploads, significantly reducing latency and enhancing patient privacy.
- **Financial Services (In Development):** Although AI PCs are not yet widely used for fraud detection or real-time trading analytics, they hold great promise in these areas. Financial institutions could leverage AI PCs to process high volumes of transaction data and run predictive analytics models locally, enhancing security by minimizing cloud exposure. The combination of NPUs and GPUs would enable fast, on-device analysis of financial transactions and market trends, delivering real-time insights in a secure and cost-effective manner.

Visionary Future Use Cases (Live 18+ Months):

- **Smart Manufacturing:** AI PCs can complement existing edge server infrastructure in smart factories by enabling efficient, localized data processing for targeted tasks such as predictive maintenance and real-time equipment monitoring. These capabilities allow AI PCs to work alongside edge servers, helping to reduce downtime and lower maintenance costs through selective, on-the-spot data processing. With energy-efficient NPUs, AI PCs also optimize power consumption for continuous, focused monitoring tasks, contributing to more efficient and resilient manufacturing operations over time.
- **Enhanced AI Assistants for Knowledge Workers:** AI PCs are set to boost productivity by hosting more sophisticated AI assistants directly on the device. With advanced hardware capabilities, such as integrated GPUs and NPUs, AI PCs can support real-time contextual understanding, learning from user behaviors, and predictive capabilities—all locally, reducing latency

- and reliance on cloud resources. As AI models evolve, these PCs will facilitate complex, automated decision-making processes on-device, allowing knowledge workers to focus on strategic, high-level tasks with faster, more reliable assistance.
- **Portable AI Processing for Engineers:** In a future where AI PCs become highly portable, engineers might offload heavy AI processing tasks to their PC during breaks. This concept represents the growing flexibility of AI PCs, enabling users to work on demanding tasks from virtually anywhere, taking advantage of localized AI processing without being tethered to a desk or relying on cloud infrastructure.

AI PC Technology Stack

AI PCs represents a sophisticated convergence of advanced hardware and software, designed to manage complex AI workloads directly on the device. To deliver their full potential, AI PCs integrate specialized hardware accelerators and AI-enhanced processors, while also relying on a robust software framework that connects them with larger Hybrid AI ecosystems. This comprehensive architecture makes AI PCs a vital tool for organizations looking to leverage the power of AI at the edge while maintaining the flexibility to scale across hybrid cloud environments. In this section, we break down the core components of AI PCs with a double click on those that compose the Edge Device itself (in green below).

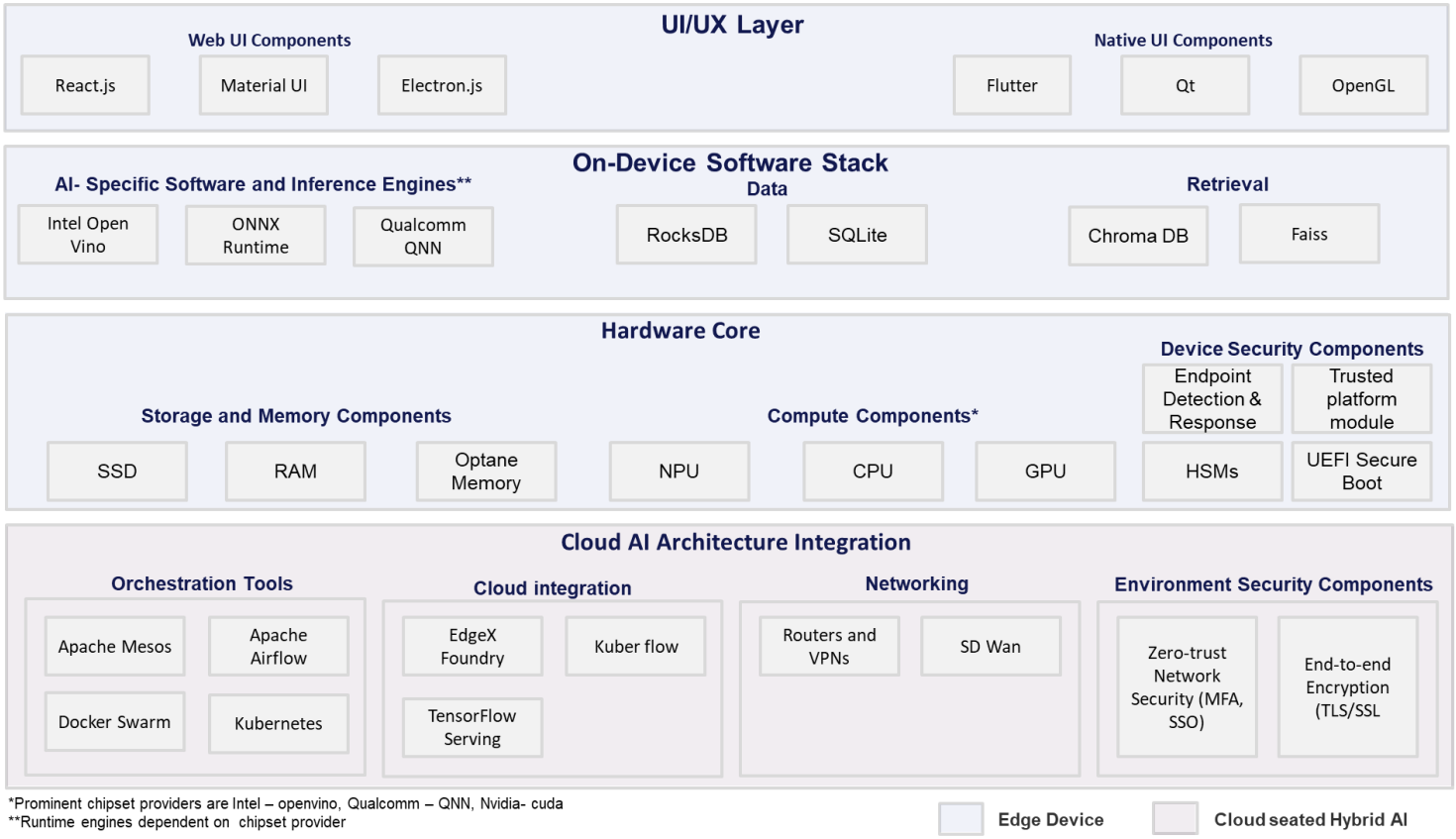


Figure 2: AI PC Infrastructure outline

AI PC Hardware Core

The integration of advanced compute units and robust storage components is what allows AI PCs to deliver highly personalized experiences by processing user-specific data locally, enhancing responsiveness and customization. Additionally, this localized processing strengthens data protection by minimizing the need for external data transfers, while simultaneously boosting efficiency through rapid, on-device computation. The reduced reliance on cloud services unlocks cost savings.

1. Compute Components

- **Neural Processing Unit (NPU):** NPUs are optimized for sustained AI workloads, offering significant power efficiency. Unlike general-purpose processors, NPUs specialize in neural network computations, reducing the overall power consumption and heat output while maintaining high performance. This makes NPUs ideal for continuous on-device AI tasks, such as real-time data processing, without requiring constant cloud connectivity. By offloading AI tasks from CPUs and GPUs, NPUs enable AI PCs to execute more complex models directly on the device, enhancing responsiveness and reducing latency in critical operations. NPUs represent the cornerstone of AI PCs, transforming them from traditional computing systems into devices capable of real-time AI processing at the edge. Their architecture is purpose-built for AI, enabling high throughput for tasks like image recognition, speech processing, and complex deep learning models. Moreover, NPUs excel at parallelism, processing thousands of operations simultaneously, something CPUs and even GPUs struggle to achieve as efficiently. For developers, leveraging NPUs means access to AI inference acceleration and lower latencies, making AI PCs more autonomous and capable of delivering powerful insights locally, without needing to offload tasks to cloud infrastructures.

- **Graphics Processing Unit (GPU):** Manages high-complexity tasks like deep learning model training, parallel processing, and video analysis. GPUs play a crucial role in large-scale AI computations.
- **Central Processing Unit (CPU):** The CPU coordinates general system tasks, managing the interaction between GPUs, NPUs, and other hardware components, ensuring smooth operation and efficient task distribution for AI workloads. CPUs can also conduct AI workloads, and with advancements in new CPUs, they can perform these tasks quite rapidly. There are numerous open-source libraries available that allow the use of CPUs for AI and generative AI computations, making it possible to leverage existing hardware for AI tasks without the need for specialized accelerators.

2. **Storage & Memory Components**

- **Random Access Memory (RAM):** RAM provides temporary storage for AI data and models during inference and training. DDR5 RAM significantly improves memory bandwidth, ensuring that AI PCs can manage large datasets in real-time.
- **Solid State Drives (SSDs):** High-speed SSDs, particularly NVMe SSDs, provide the necessary read/write speeds for storing and retrieving AI models and datasets efficiently.
- **VRAM:** VRAM, found in GPUs, is crucial for managing the large datasets and complex computations involved in AI tasks. High-capacity VRAM allows for efficient parallel processing and rapid data access, making it essential for training and inference in deep learning models. Modern GPUs with substantial VRAM ensure that AI PCs can manage intensive workloads without bottlenecks, significantly enhancing performance.
- **Optane Memory:** Acts as a high-speed bridge between SSDs and RAM, speeding up

- access to frequently used data. It minimizes latency during AI tasks, providing persistent memory solutions ideal for real-time processing.

1. **On-Device Security**

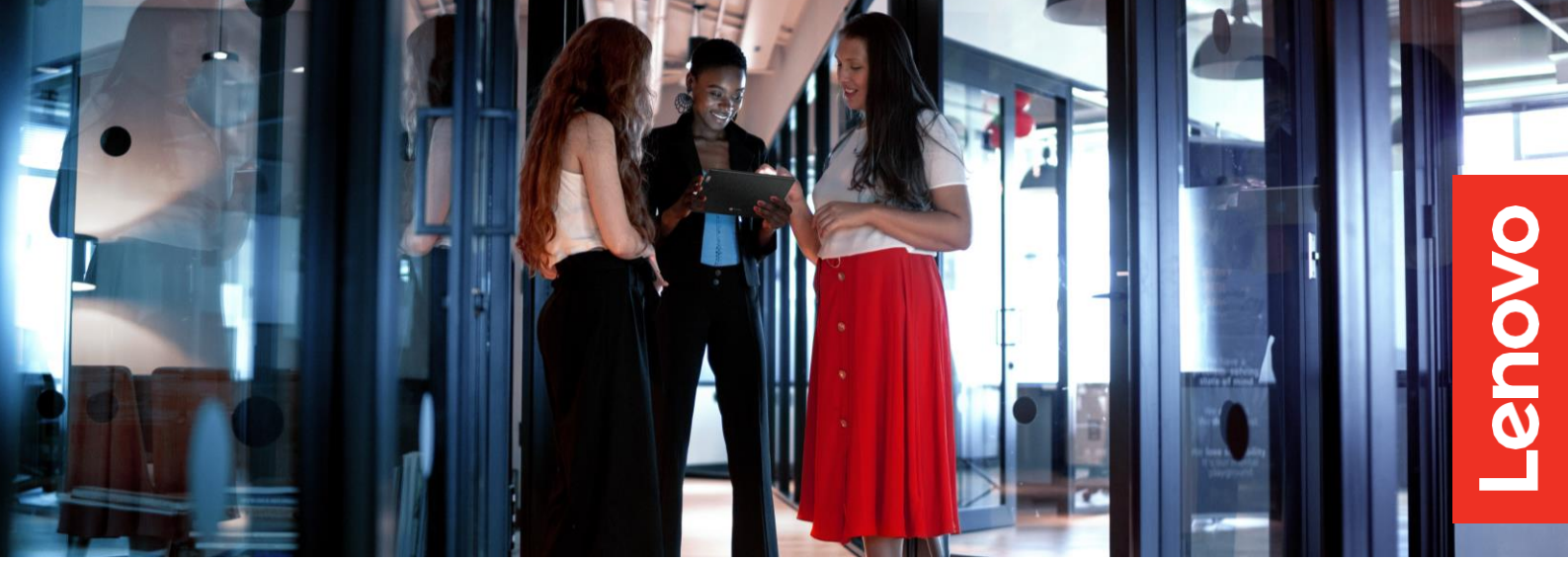
- **Endpoint Detection & Response (EDR):** Continuously monitors the device for potential security threats, identifying unusual patterns or behaviors that may indicate a breach. EDR provides real-time threat detection and rapid response capabilities, helping to safeguard sensitive data on the device.
- **Trusted Platform Module (TPM):** A dedicated security chip that ensures hardware-based protection by securely generating and storing cryptographic keys. TPM enhances overall system integrity, enabling secure cryptographic operations and protecting sensitive data against unauthorized access.
- **Unified Extensible Firmware Interface (UEFI) Secure Boot:** A firmware feature that validates and authorizes each step in the boot process, ensuring that only trusted software is loaded. This protects the system from malicious code injection during startup, reinforcing a secure foundation for all on-device operations.
- **Hardware Security Modules (HSMs):** Specialized hardware components designed to protect and manage cryptographic keys with high security. HSMs support secure encryption, authentication, and key management, safeguarding sensitive operations and enhancing the security of data processed on the device.

On-Device Software Stack

The software stack in AI PCs plays a crucial role in unlocking the full potential of the hardware. It provides the framework for building, deploying, and running AI models on the device, ensuring that the processing power of GPUs, NPUs, and other components is fully utilized. This software stack enables personalized experiences by allowing AI models to adapt to

individual user interactions via inference engines, enhances protection through advanced security protocols, and boosts productivity via optimized resource management, all of which contribute to cost efficiency.

- **AI-Specific Software Frameworks:** These frameworks provide a standardized environment for building, training, and deploying AI models on AI PCs. They allow developers to easily offload complex AI tasks to dedicated hardware like NPUs and GPUs, optimizing performance without requiring extensive customization. Many software vendors can leverage these frameworks to extend their AI capabilities onto AI PCs, enabling seamless integration with existing AI tools and ensuring high efficiency. By offering out-of-the-box support for various hardware accelerators, these frameworks reduce the development time and allow vendors to quickly optimize their software for AI PCs.
- **Inference Engines:** These streamline the execution of AI models on hardware accelerators like NPUs and GPUs, ensuring that real-time results are delivered efficiently. The integration between hardware and software is essential for handling real-time AI processing tasks, such as image recognition or natural language processing, distinguishing an ordinary chipset ecosystem (PC) from an AI PC ecosystem. Each of these “inference engines” includes a programming language based on C/C++ for programming hardware and an assembly language that other programming languages can use as a target, allowing AI models to fully utilize the hardware. Additionally, they come with a software development kit (SDK) that includes libraries, various debugging, profiling, and compiling tools, and bindings that let CPU-side programming languages invoke GPU/NPU-side code. This enables developers to use high-level languages like Python or low-code



- platforms to achieve powerful optimization steps. Examples of these inference engines include NVIDIA's CUDA, Qualcomm's QNN, and Intel's OpenVINO, which are pivotal in transforming a standard PC ecosystem into a robust AI PC ecosystem capable of handling complex AI tasks with enhanced performance.
- **Retrieval Databases (Chroma DB, Faiss):** The retrieval layer enables fast, efficient data access for AI-driven applications on AI PCs. Chroma DB and Faiss support quick indexing and searching of complex data, essential for applications like contextual search and recommendation systems. By handling data retrieval locally, these components reduce cloud dependency, lower latency, and improve responsiveness, enhancing the real-time performance of AI PCs.
- **Data Management (RocksDB, SQLite):** Data management components handle local storage of structured and semi-structured data, providing fast access and efficient organization for AI models. These databases support the persistence of user interactions, preferences, and context, enabling adaptive, personalized AI experiences without relying on cloud-based data storage. Local data management also enhances privacy and reduces latency, key benefits for on-device AI processing.

UI/UX Layer

The UI/UX layer in AI PCs is engineered to harness on-device AI capabilities, ensuring responsive, low-latency interactions and enabling advanced real-time processing for a seamless user experience.

- **Native UI Components (Flutter, Qt, OpenGL):** Native components directly access the device's hardware, including GPUs, NPUs, and sensors, allowing for high-performance, low-latency processing critical for AI-intensive applications. By leveraging direct hardware integration, native components enable real-time

- feedback and processing, supporting applications like Copilot+ for advanced contextual assistance, natural language processing, and predictive analytics.
- **Web UI Components (React.js, Material UI, Electron.js):** Web components ensure cross-platform functionality and facilitate device-level inferencing by offloading AI workloads from the cloud to local hardware. Current experiments with Google Chrome, for example, aim to enable browsers to utilize on-device inference, reducing latency and enhancing processing efficiency. This shift allows web applications to deliver faster, AI-enhanced interactions directly on the AI PC, without cloud dependency.

The bottom most layer in the diagram is cloud AI architecture integration layer. This facilitates seamless interaction between on-device capabilities and cloud-based resources, providing scalability, flexibility, and enhanced processing power. By integrating orchestration tools (e.g., Kubernetes, Apache Airflow), cloud services, and networking components, this layer enables AI PCs to offload complex tasks to the cloud when necessary, ensuring a balance between local processing and cloud resources. This hybrid approach supports high-performance AI applications, reduces latency for real-time tasks, and allows for continuous model updates and large-scale data processing.

Additionally, environment security components, such as zero-trust network security, end-to-end encryption, and secure SD-WAN, ensure that data transfers between the device and cloud are protected against unauthorized access. This secure integration enhances the AI PC's ability to handle both sensitive and large-scale workloads, making it suitable for enterprise environments where data security and efficient resource management are critical.

Evaluating AI PC Fit

AI PCs are an ideal solution for organizations that need localized, real-time AI processing and data security while reducing dependency on cloud infrastructure. Based on their unique operational needs, several types of organizations stand to benefit the most from adopting AI PCs:

1. **Archetype 1: Real-Time Decision Makers** - Organizations that rely on real-time decision-making, such as healthcare providers, manufacturing plants, and retailers, can significantly benefit from AI PCs. In these environments, immediate data processing is critical to optimize operations, whether it's detecting anomalies in a production line, providing personalized retail recommendations, or diagnosing medical conditions in real-time.
2. **Archetype 2: Privacy-Centric Organizations** - Companies managing sensitive data, such as those in financial services, legal sectors, and healthcare, can leverage AI PCs to keep data processing local, enhancing privacy and security. These organizations need strict data governance to comply with regulatory standards like GDPR, HIPAA, or CCPA, and AI PCs allow for edge processing without exposing sensitive data to external cloud environments.
3. **Archetype 3: Resource-Constrained or Decentralized Operations** - Organizations with limited cloud connectivity or that operate in distributed environments, such as remote operations, smart cities, and logistics companies, will find AI PCs a cost-effective alternative to cloud-based solutions. AI PCs reduce bandwidth requirements, latency, and cloud service costs while providing autonomous AI capabilities at the edge.

The framework below can help ITDMs assess how well AI PCs align with their technical needs, operational goals, and long-term strategy.

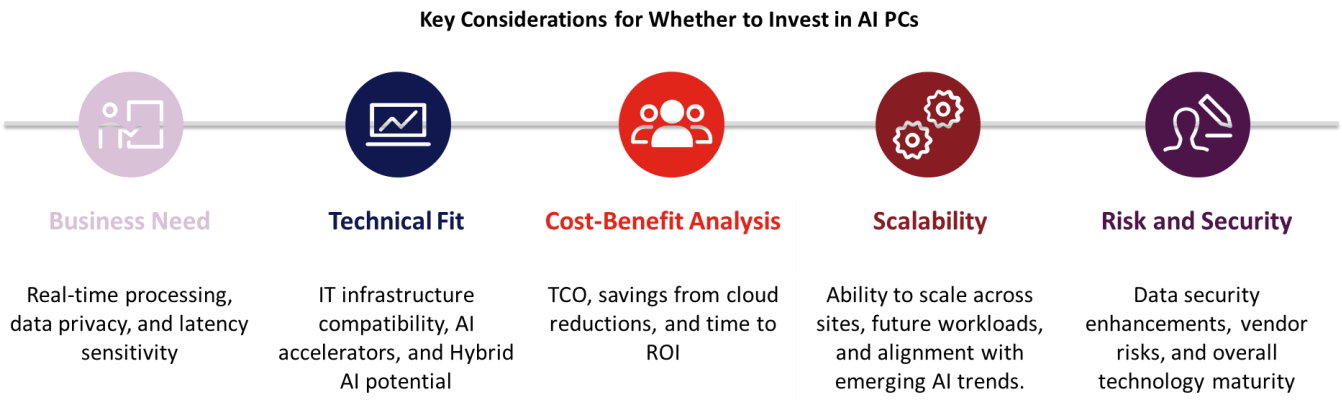


Figure 3: Considerations for investing in AI PCs

Topic	Core Question	Key Considerations
Business Need Assessment	Do the organization's operational needs demand real-time processing, enhanced data security, or edge AI capabilities?	• Identify processes or workflows that could benefit from real-time AI insights (e.g., real-time decision-making, predictive maintenance). • Evaluate the need for data privacy (e.g., compliance with regulations like GDPR or HIPAA) and whether local AI processing would improve data governance. • Determine the value of low-latency processing in use cases such as retail personalization, automated manufacturing, or smart healthcare systems.
Technical Fit	Does the organization have the infrastructure to support AI PCs and benefit from edge computing capabilities?	• Assess the organization's current IT infrastructure and its ability to integrate AI PCs. This includes network capabilities, data pipelines, and cloud resources. • Examine the need for AI accelerators (e.g., NPUs, GPUs) within the AI PC environment and whether specialized hardware can improve existing workloads. • Consider if Hybrid AI (combining cloud and edge processing) can optimize workflows and reduce latency.
Cost-Benefit Analysis	Is the ROI of AI PCs over a 3-year period higher than continuing with existing infrastructure or cloud-based AI solutions?	• Compare CAPEX (initial AI PC investment, AI accelerators, infrastructure upgrades) with OPEX (ongoing cloud costs, maintenance, and data transfer fees). • Calculate the total cost of ownership (TCO), including the initial cost of AI PCs, the cost of AI accelerators, and potential ongoing maintenance. • Analyze potential cost savings from improved efficiency, faster insights, and reduced cloud dependency. • Estimate the time to ROI based on gains in productivity and operational efficiency.
Scalability and Futureproofing	Can AI PCs scale with the organization’s growth and evolving AI needs?	• Evaluate how AI PCs fit into the organization’s long-term strategy. Can the solution scale across multiple sites or departments? • Consider how future AI workloads might grow and if AI PCs can manage the increasing complexity. • Assess whether AI PCs are aligned with emerging trends like federated learning, edge AI frameworks, or augmented reality use cases. • Analyze the vendor roadmap for AI PCs and potential integration with future technologies like 5G and advanced AI accelerators.

Topic	Core Question	Key Considerations
Risk and Security Evaluation	Will AI PCs reduce security risks or introduce new vulnerabilities?	• Analyze the organization's data security requirements. Does local processing through AI PCs reduce exposure of sensitive data to cloud services? • Review the risk of adopting innovative technology. Consider factors such as vendor lock-in, the maturity of AI PC technology, and compatibility with existing systems. • Evaluate how cybersecurity frameworks are integrated into AI PCs and whether these provide additional layers of protection compared to cloud-based solutions.

By evaluating these factors, organizations can make an informed decision on whether AI PCs align with their current operational goals and long-term technological strategy.

Measuring AI PC Value

AI PCs deliver tangible business value across multiple dimensions, making these devices a compelling investment priority for ITDMs.

These devices are expected to deliver a strong multiyear business case and ITDMS can expect the following KPIs:

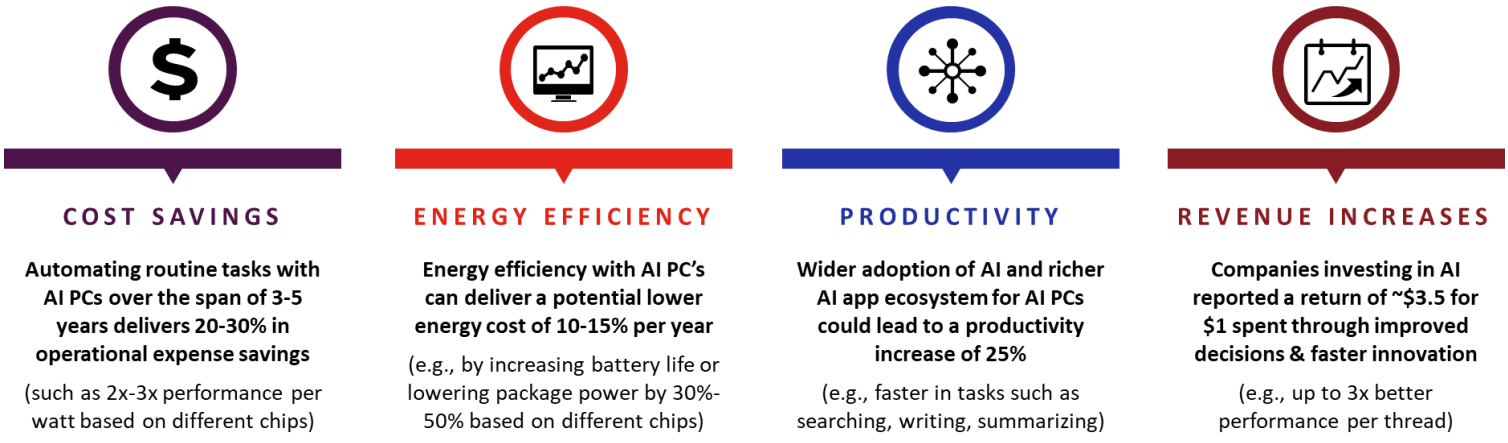


Figure 4: KPIs of AI PCs

To effectively measure this expected value, ITDMs can tailor the metrics below to their specific needs and workflows.

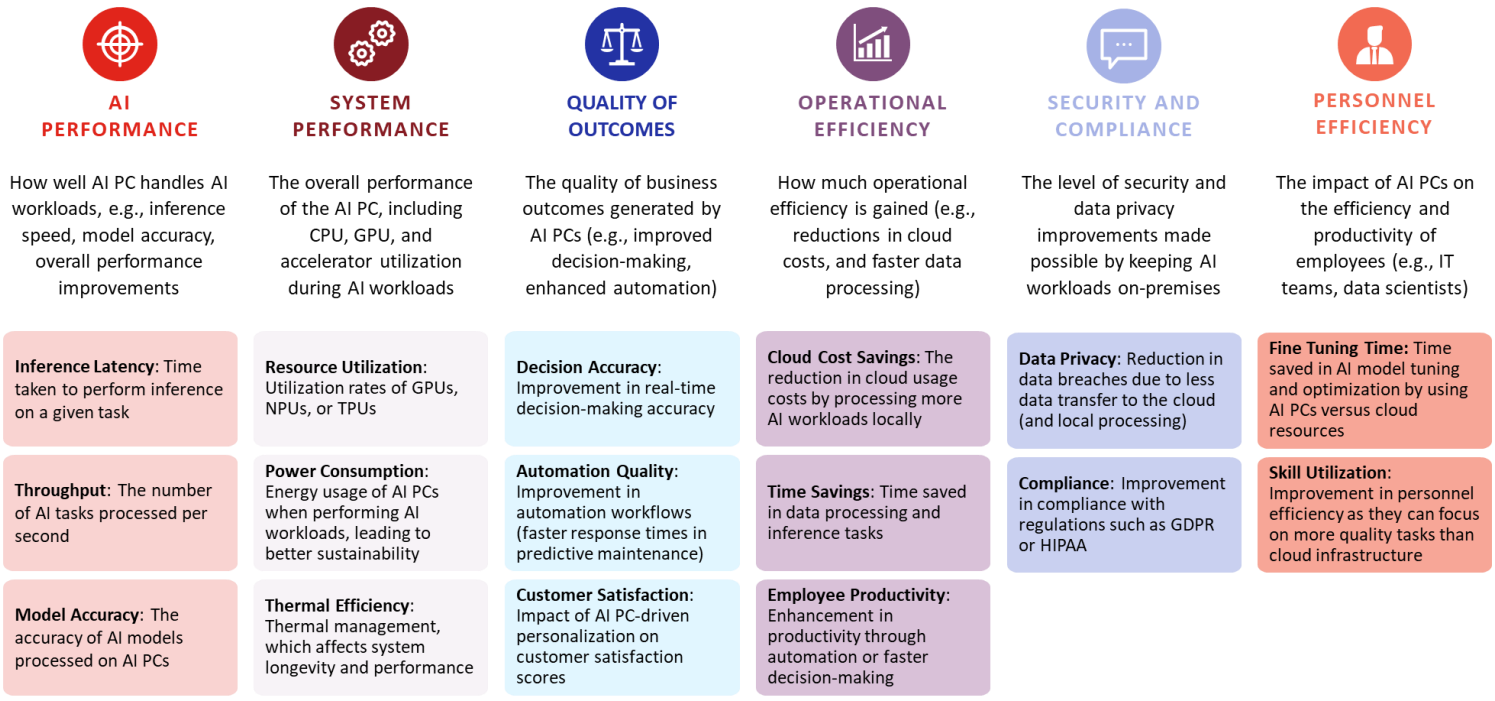


Figure 5: Measuring the value of AI PCs

Deploying AI PCs with Hybrid AI

An AI PC fits seamlessly into a hybrid AI strategy by acting as the bridge between local (edge) computing and centralized cloud environments. In this strategy, AI PCs manage critical, real-time AI tasks at the edge, while the cloud takes on more resource-intensive functions like model training, storage, and large-scale data analytics. The result is an optimized balance of performance, latency reduction, and security.

Hybrid AI plays a crucial role in ensuring that AI processing is distributed efficiently between the edge and the cloud, leveraging the unique strengths of both environments. Here is how it achieves this:

1. **Distributing AI Workloads for Optimal Performance:**

- **Edge Processing with AI PCs:** AI PCs process AI tasks locally, closer to the data source, enabling real-time responses and minimizing latency associated with cloud-based processing. This

- approach is ideal for applications requiring immediate feedback, such as predictive maintenance and autonomous functions. While AI PCs enhance edge capabilities in localized, user-specific contexts, edge servers support larger-scale private cloud deployments and can be deployed at the far edge for industrial applications like manufacturing. Together, AI PCs and edge servers create a complementary ecosystem that optimizes both individual and enterprise-level workloads.
- **Cloud for Heavy Processing:** For tasks that require significant compute power or access to vast data lakes, the cloud is leveraged to train, update, or retrain complex AI models. Once the model is trained in the cloud, it can be deployed on AI PCs for inferencing at the edge. This distribution of tasks reduces the load on AI PCs and ensures cloud resources are used only when necessary. By splitting these tasks based on their nature—real-time inferencing at the edge and computationally heavy tasks in the cloud—hybrid AI ensures optimized resource allocation and overall better performance.

2. **Reducing Latency for Real-Time Applications:** One of the primary reasons for integrating AI PCs into a hybrid AI strategy is to reduce latency. Many AI applications, such as autonomous driving, industrial robotics, or healthcare diagnostics, demand immediate processing and decision-making. Sending data back and forth to the cloud for these tasks would introduce unacceptable delays.
   - **Local AI inferencing:** With AI PCs, data can be processed at the source in real-time, eliminating the latency caused by cloud data transfers. The hybrid AI approach ensures that time-sensitive tasks stay local, while non-urgent or bulk processing can be offloaded to the cloud.

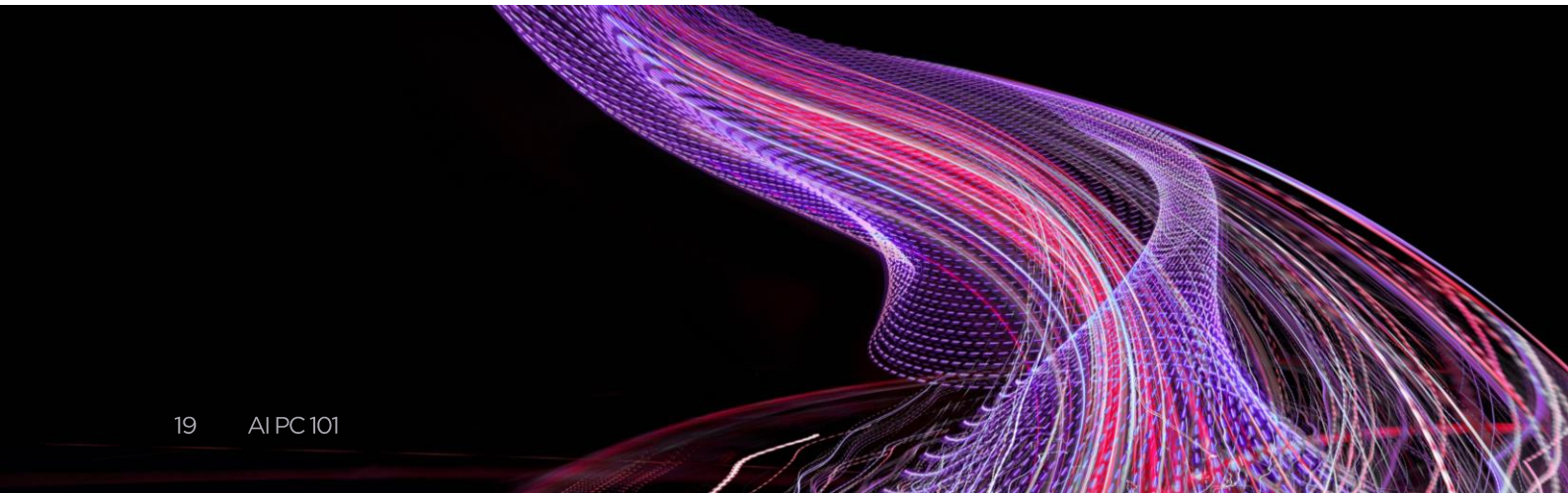
- **On-demand cloud integration:** For processes that are not latency-sensitive, AI PCs can send data to the cloud as needed, allowing businesses to retain flexibility in how they manage AI tasks while benefiting from lower latency for critical processes.

3. **Enhancing Security and Data Privacy:** Hybrid AI strategies also address security concerns by keeping sensitive data local to the edge, where AI PCs process it without needing to send the data to the cloud. This reduces the risks associated with data transfer and ensures that sensitive information is managed in compliance with privacy regulations.

- Local data processing: For industries like healthcare, finance, or government sectors, where data security and privacy are paramount, AI PCs offer the advantage of processing sensitive information locally. This limits the amount of data that must be transmitted over the network, reducing the risk of data breaches or cyberattacks during transfer.
- **Cloud for broader analytics:** While AI PCs manage sensitive or time-critical data at the edge, anonymized or less sensitive data can be sent to the cloud for deeper analytics or long-term storage. Hybrid AI strategies help ensure that only necessary data is transferred to the cloud, aligning with regulatory requirements and best practices in data security.

4. **Scalability and Flexibility:** AI PCs in a hybrid AI setup offer scalability by dynamically distributing tasks based on computing demands. This ensures optimal usage of available resources.

- **Scalable processing:** AI PCs can manage inferencing tasks or lightweight AI models locally, while the cloud is scaled up for more intensive model training or data processing. This balance allows organizations to scale their AI operations



The Role of ISVs

ISVs play a pivotal role in shaping the value proposition of AI PCs. By developing software that is optimized for AI PC hardware, ISVs enable businesses to leverage advanced AI capabilities more effectively. Partnerships between AI PC manufacturers and ISVs foster an ecosystem where innovative applications and tools can be integrated directly into the hardware, driving more innovation across industries. ISVs contribute to the creation of specialized software solutions, such as intelligent productivity tools and better content creation apps. These solutions enhance the utility of AI PCs by offering tailored functionalities that align with business objectives. Moreover, as AI PCs evolve, ISV partnerships will be critical in driving advancements in AI software frameworks, enabling more efficient deployment of AI models and further enhancing performance.

ISVs will continue to drive innovation by optimizing AI applications for specific industries and enhancing AI model deployment and performance. Key areas of focus for ISVs will include:

- **New Tools for AI Model Optimization:** ISVs will develop applications that improve the deployment, training, and inference of AI models directly on AI PCs. These tools will enable enterprises to run advanced machine learning models at the edge, reducing the need for cloud-based processing and allowing for faster, localized AI computations.
- **AI-Driven Security Solutions:** Security-focused applications, such as BufferZone and CrowdStrike, will safeguard AI PCs against emerging cyber threats. These solutions ensure that data processed on AI PCs remains secure without relying solely on cloud-based protections, which is critical for industries with strict compliance requirements, such as healthcare, financial, legal sectors.

- **Advanced Collaboration Tools:** ISVs will enhance AI PCs with smarter collaboration tools like Microsoft Teams Studio Effects and Lenovo Device Management, providing businesses with intuitive solutions for real-time communication, device optimization, and smarter decision-making. These applications will be tailored to specific verticals, ensuring that industries like education, healthcare, and retail benefit from targeted collaboration enhancements.
- **Verticalized Offerings:** ISVs will also focus on developing vertical-specific AI applications, offering tailored solutions to industries such as healthcare, manufacturing, retail, and financial services. For example, ISVs may create diagnostic tools optimized for medical imaging on AI PCs in healthcare, predictive analytics applications for inventory management in retail, or financial modeling tools for the banking sector. These verticalized offerings will drive more precise use cases, helping organizations achieve better outcomes and operational efficiency.
- **Scalable Front-End Design for On-Device Inferencing:** As AI PCs and other AI-enabled devices reach critical mass, ISVs will need to adapt their applications to leverage browser APIs and other tools for inferencing offload. This approach will ensure that web applications are optimized for local processing, achieving scalability and responsiveness without relying on cloud infrastructure.

By focusing on these areas, ISVs will expand the horizons of AI in enterprise computing, driving new use cases and unlocking the full potential of AI PCs.

The Future of AI PCs: AI PC Trends

As AI technology continues to evolve, the AI PC ecosystem will become more sophisticated, driven by the integration of innovative hardware, software, and emerging use cases. We will see a significant increase in computational power as more advanced AI accelerators (e.g., GPUs, NPUs) are developed. These components will further enhance edge processing capabilities, enabling AI PCs to manage increasingly complex AI models and applications in real time. The ecosystem will also evolve toward greater interoperability between devices and platforms. AI PCs will seamlessly integrate with other edge devices, cloud systems, and AI solutions, forming an interconnected hybrid ecosystem where AI tasks are dynamically distributed between edge and cloud environments based on latency and performance requirements. This shift will redefine workflows, as companies can leverage localized AI processing for immediate insights while offloading more complex or non-urgent tasks to the cloud. Additionally, the broader ISV ecosystem will further catalyze use case development and ultimately AI PC ROI. Greater adoption will accelerate the development of more ISV solutions and use cases which will in turn drive improved economics and greater ROI of the AI PC.

Several technologies will shape the future of AI PCs, making them pivotal for real-time, scalable AI applications across industries. Some of these include:

	Description	*Benefits*
HYBRID AI	Hybrid AI enables AI PCs to process tasks locally while offloading complex workloads to the cloud. This balance optimizes performance by reducing latency and enabling real-time decisions at the edge.	Fast, low-latency insights for applications like autonomous vehicles and smart manufacturing.
AI ACCELERATORS	AI PCs leverage specialized hardware like NPUs, GPUs, and FPGAs to accelerate AI tasks. These components enable faster inference and model execution.	High-performance processing of complex AI workloads with lower energy use.
CLOUD-EDGE INTEGRATION	Technologies like edge orchestration, 5G networks, and containerization support seamless interaction between AI PCs and cloud infrastructure, allowing flexible workload distribution.	Enables efficient data exchange and scalability for applications like real-time analytics and automation.
FEDERATED LEARNING	AI PCs can train models locally, sending only updates to a central server, enhancing privacy while reducing data transfer to the cloud.	Greater data privacy, especially for sectors like healthcare and finance.
EDGE AI FRAMEWORKS	Frameworks like TensorFlow Lite and ONNX Runtime optimize AI model deployment on AI PCs, making edge computing efficient and scalable.	Streamlined model deployment and reduced complexity for edge-based AI solutions.
AI MODEL COMPRESSION	Techniques like quantization and pruning shrink AI models, making them faster to execute on AI PCs without sacrificing accuracy.	Improved performance and energy efficiency for AI workloads on local devices.

Figure 6: The evolution of AI PC Trends and Technologies

As these trends and technologies converge, AI PCs are set to revolutionize computing by integrating powerful AI capabilities directly into everyday devices, driving efficiency, personalization, and real-time decision-making across industries.

As more organizations adopt AI PCs, the business case for these devices will grow stronger, driving economies of scale and positioning AI PCs as a standard tool for organizations seeking to enhance their AI capabilities while optimizing costs.

About Lenovo

Lenovo is a US$57 billion revenue global technology powerhouse, ranked #248 in the Fortune Global 500, and serving millions of customers every day in 180 markets. Focused on a bold vision to deliver Smarter Technology for All, Lenovo has built on its success as the world's largest PC company with a pocket-to cloud portfolio of AI-enabled, AI-ready, and AI-optimized devices (PCs, workstations, smartphones, tablets), infrastructure (server, storage, edge, high performance computing and software defined infrastructure), software, solutions, and services. Lenovo's continued investment in world-changing innovation is building a more equitable, trustworthy, and smarter future for everyone, everywhere. Lenovo is listed on the Hong Kong stock exchange under Lenovo Group Limited (HKSE: 992) (ADR: LNVGY).

You can access additional white papers covering a range of AI topics here. These resources will help you fuel your AI journey and explore themes like Hybrid AI, AI PC 101, Measuring AI Value, Accelerating from POC to Production, The Role of Human Intervention in AI, AI Security Considerations, and The Power of AI-Driven Storytelling.