mbx systems

Medical AI: The Shift from Cloud to Edge Inferencing

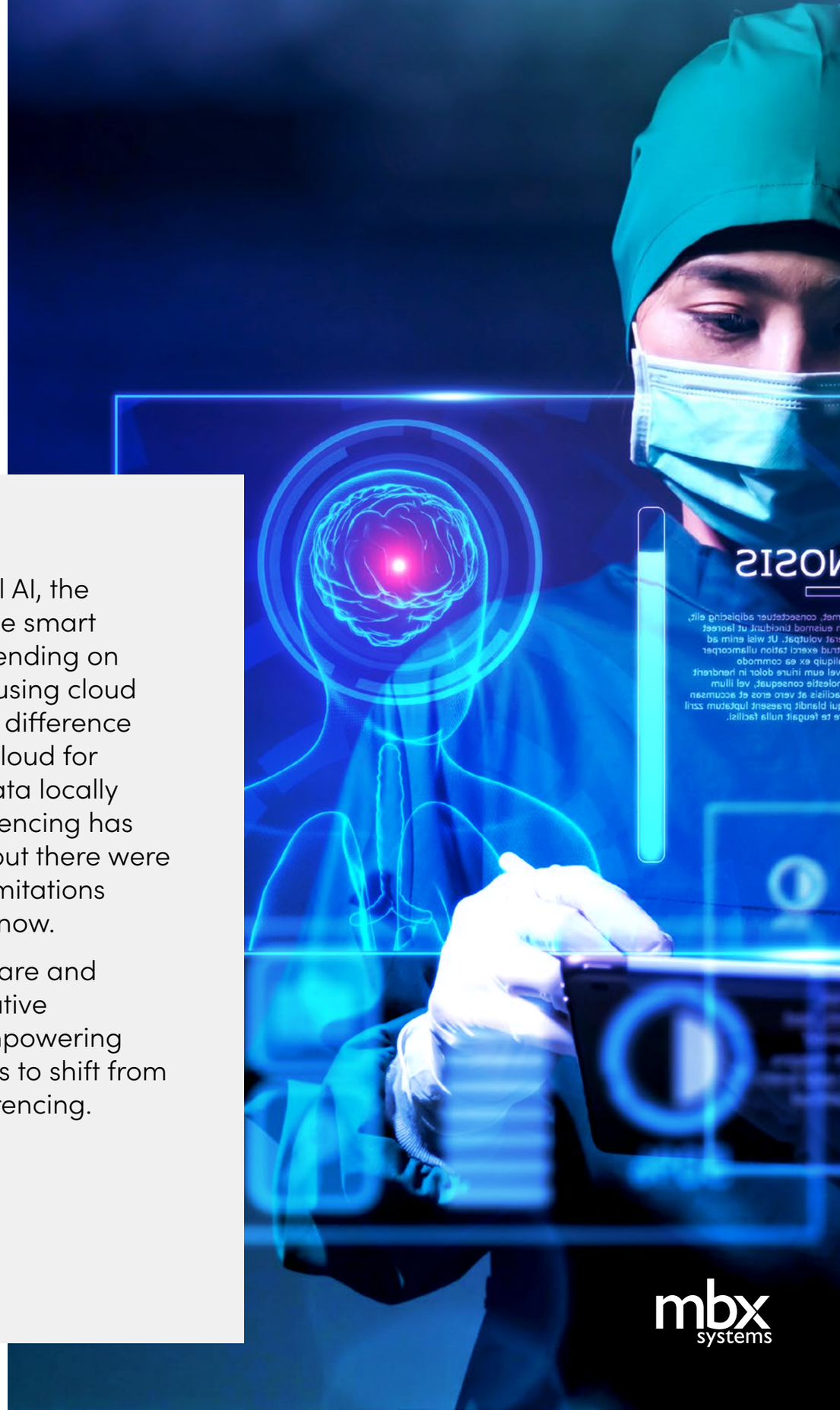
Advances in hardware and software are fueling a new wave of medical devices that leverage edge inferencing for real-time AI insights.

AI is becoming a foundational technology across nearly every industry, and despite lagging behind industrial and manufacturing, the healthcare industry is no exception. In fact, the **adoption of smart hospital technologies** has been accelerating rapidly since the start of the pandemic. Medical device companies are recognizing they can integrate AI and machine learning to harness the enormous amount of data at their disposal.

In recent years, **AI has been utilized in the healthcare industry** to deliver data-driven clinical decision support (CDS) and programmed insights for healthcare providers. This AI-guided care approach increases the accuracy and speed of diagnosis, leading to improved patient outcomes at a much lower cost. AI is also fueling a new era of clinical research that's transforming the pharmaceutical and medical industries for the better.

Despite the benefits of medical AI, the performance potential for these smart medical devices can vary depending on whether they're implemented using cloud or edge inferencing. This is the difference between sending data to the cloud for processing or analyzing this data locally where it's collected. Edge inferencing has many benefits for medical AI, but there were also too many technological limitations for widespread adoption until now.

Learn how advances in hardware and software, along with an innovative deployment approach, are empowering medical technology developers to shift from cloud inferencing to edge inferencing.



The Advantages of Edge Inferencing for Medical Devices

Many AI implementations in the healthcare industry rely on cloud inferencing, but edge AI is the key to generating real-time insights from massive datasets in healthcare and life sciences. Here are some of the advantages of implementing edge inferencing for medical devices:

01.
Reduced Latency

Cloud inferencing typically requires data to be transferred to the cloud for processing, which greatly increases the time it takes to generate AI insights. Since edge inferencing is performed locally, this reduces latency and enables medical devices with real-time capabilities for patient monitoring, diagnostics, and more. It's also possible to process massive amounts of data locally for genomics, bioinformatics, and proteomics use cases without latency and bandwidth issues.

02.
Enhanced Security

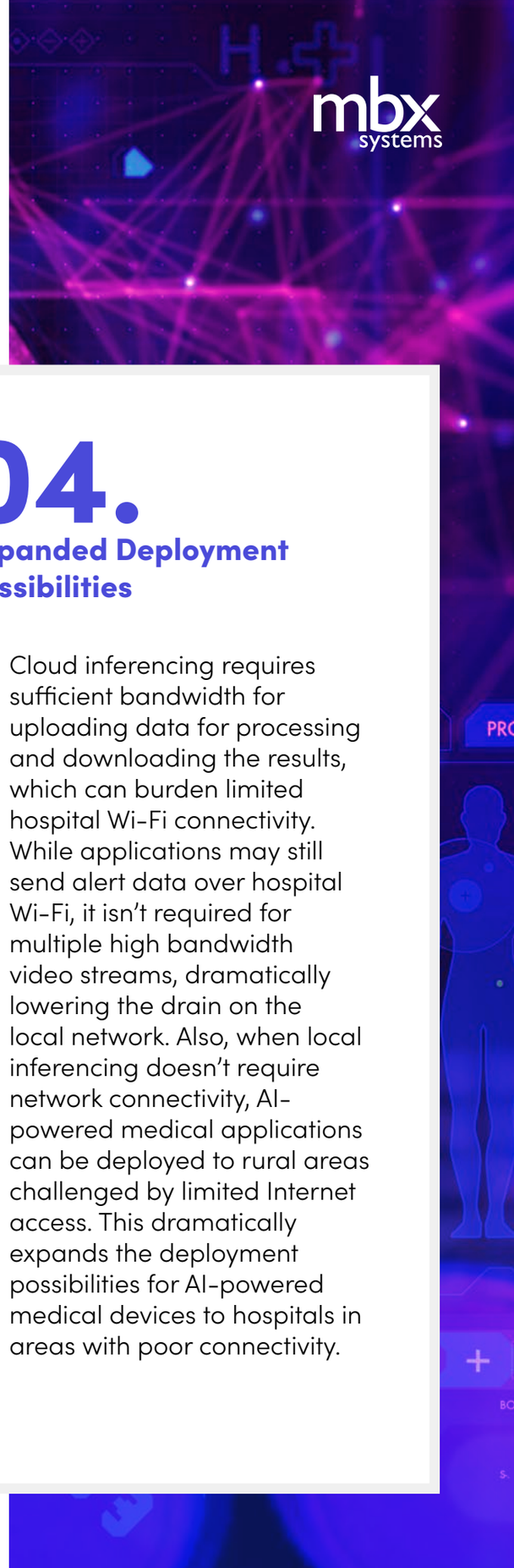
Besides the latency and bandwidth constraints with cloud inferencing, there are also security challenges associated with processing datasets offsite. Transferring data over the Internet introduces the risk of it being intercepted by unauthorized third parties. When inferencing locally at the edge, data doesn't need to leave the hospital's network. This eliminates the risk of man-in-the-middle or other cybersecurity threats that target data in transit.

03.
Increased Data Privacy

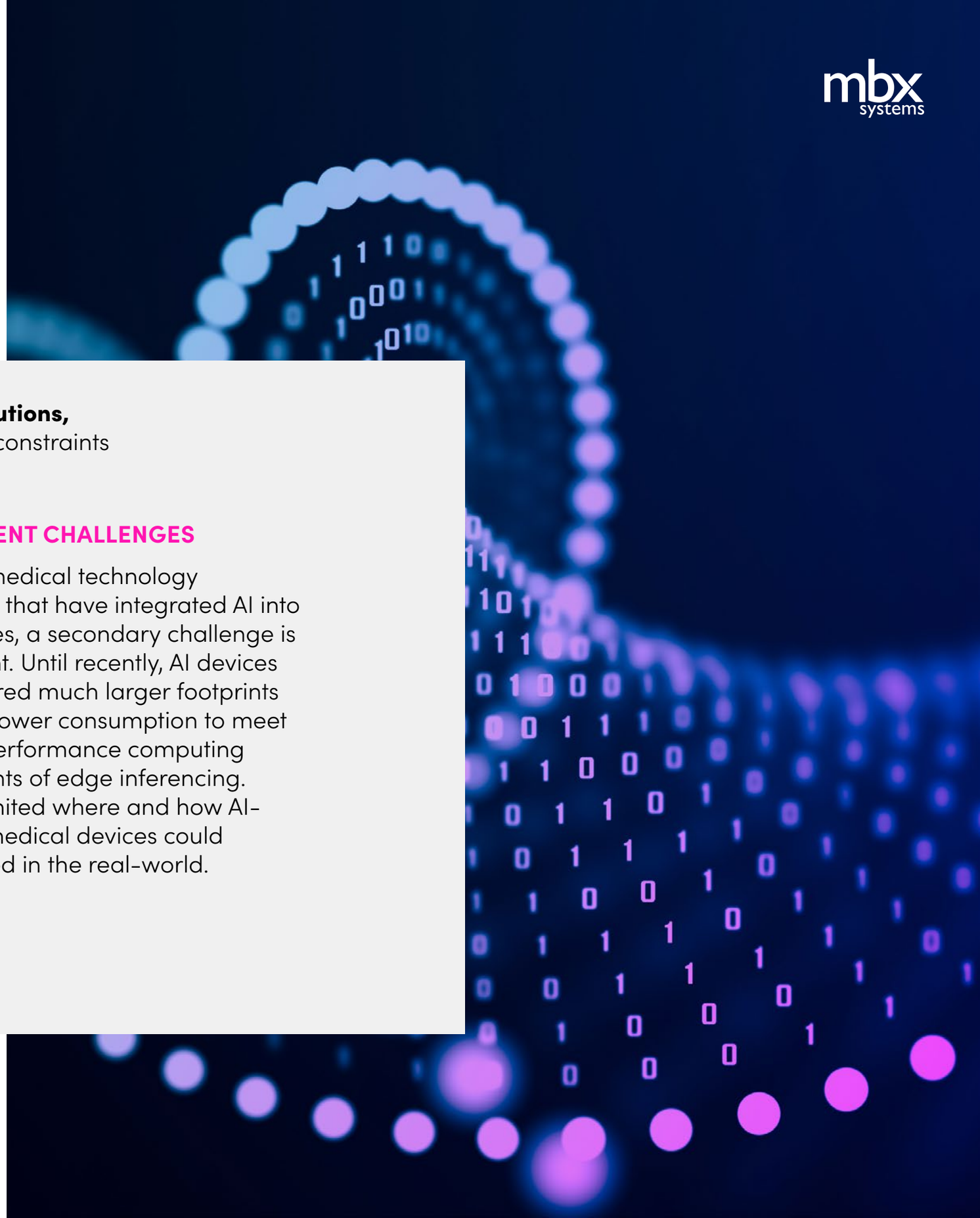
Many medical applications deal with personally identifiable information (PII) and protected health information (PHI) that need to comply with HIPAA rules, which cause data privacy concerns for hospitals. Patient data that's stored in the cloud can be more vulnerable to data breaches, which causes some healthcare providers to be hesitant to adopt cloud-based AI solutions. Using edge inferencing, AI/ML algorithms can process this sensitive data without the risks associated with storing the data in a shared public cloud or other external environments.

04.
Expanded Deployment Possibilities

Cloud inferencing requires sufficient bandwidth for uploading data for processing and downloading the results, which can burden limited hospital Wi-Fi connectivity. While applications may still send alert data over hospital Wi-Fi, it isn't required for multiple high bandwidth video streams, dramatically lowering the drain on the local network. Also, when local inferencing doesn't require network connectivity, AI-powered medical applications can be deployed to rural areas challenged by limited Internet access. This dramatically expands the deployment possibilities for AI-powered medical devices to hospitals in areas with poor connectivity.



Why Medical Application Edge Deployments Have Been Limited



While demand is growing for AI-powered medical solutions, developers have previously faced many technological constraints when implementing edge inferencing. Here's why:

PERFORMANCE & COST LIMITATIONS

AI applications are resource-intensive, especially when inferencing is performed locally on the device. The hardware costs and limited processing capabilities of most chipsets prevented widespread deployment of AI-powered medical devices at the edge. In short, the performance-per-dollar of edge devices hasn't been practical for most medical use cases.

DEPLOYMENT CHALLENGES

For those medical technology developers that have integrated AI into their devices, a secondary challenge is deployment. Until recently, AI devices have required much larger footprints and high power consumption to meet the high-performance computing requirements of edge inferencing. This has limited where and how AI-powered medical devices could be deployed in the real-world.

How Hardware & Software Advances Are Fueling Edge Inferencing

Technology is evolving at a rapid pace, and the **previous barriers to medical AI and edge inferencing are collapsing**. Advances in hardware and software will not only promote widespread adoption of edge inferencing for medical devices, but also enable AI to become a foundational technology in the healthcare industry.

For one, chipsets are getting faster and cheaper every year. In fact, AI processing performance has even outpaced general compute speeds over the last decade, **doubling every six months or so**. This has greatly reduced the cost and energy use associated with AI and computer vision workloads, and made it more economical to implement inferencing within edge devices and embedded systems.

Innovative **AI inference platforms** and pre-defined software solutions are also accelerating medical application development. For example, a machine learning platform like **NVIDIA Clara** can get you **80% of the way to a functioning app** by providing pre-trained models, transfer learning toolkits, deployment SDKs, analytics tools, and other capabilities. These predefined software models are dramatically reducing the costs associated with developing AI applications from scratch for edge deployments.

MBX Develops NVIDIA Powered Optio Product Line for Medical Edge AI

MBX is introducing the Optio series based on NVIDIA IGX Orin™ , AGX and NX platform architectures. The MBX Optio series will achieve a new level of performance per dollar that makes medical AI edge computing a reality.

OPTIO L100 & L150 FOR LARGE-SCALE PERFORMANCE:

NVIDIA IGX powered server platform that is optimized for building medical devices designed with **NVIDIA Holoscan** to dramatically shorten time to obtain 62304 medical certification. L100 comes in a small desktop form factor, while L150 is rack mountable in a 4U form factor supporting additional storage and AI processing. Ideal for handling the resource-intensive processing of genomics data or other massive data sets.

OPTIO M100 & M150 FOR MID-TIER PERFORMANCE:

AGX powered devices with a single integrated 1792-core, 56 Tensor Core / 2048-core, NVIDIA Ampere architecture GPU with 64 Tensor Core, 8 / 12 core Arm Cortex processor and 32 / 64GB LPDDR5 memory. Ideal for AI use cases that are less data-intensive but still require multiple sensors and low-latency processing such as endoscopy, ultrasound, diagnostic imaging, radiation therapy, microscopy, and others.

OPTIO S100 & S150 FOR SMALL-SCALE PERFORMANCE:

NX powered devices with an integrated embedded 1024 core NVIDIA Ampere architecture GPU with 32 Tensor Cores and 8 core Arm Cortex processor and 8 / 16GB LPDDR5 memory. Ideal for bringing AI to patient monitoring, patient positioning analytics and other solutions that require a few A/V feeds at a lower cost.

These latest building hardware blocks are part of the ongoing effort of **MBX and NVIDIA to fast-track the development of AI workflows in hospitals**. In fact, many companies have already built innovative medical devices based on **MBX Kori**, a modular mobile platform for computer vision at the edge.

Learn more about Delivering AI Edge Computing Devices for Industrial and Medical Environments.

mbx systems

Bringing Medical AI to the Edge with MBX

MBX has a team of hardware engineers that specialize in complex solutions for many industries, including medical technologies. Our innovative development approach balances pre-designed modular hardware building blocks and customization to bring time and cost efficiencies to the device manufacturing process.

Many medical technology developers choose to design a custom solution from the ground up, but this is complicated, time-consuming, and costly. A viable option MBX pioneered is to guide customers to pre-designed modular hardware as a starting point, and then leverage a compatible predefined machine learning platform to greatly accelerate the development of AI-powered medical devices. This reduces time-to-market for medical technology developers because the hardware is pre-certified, pretested, and optimized for machine learning and AI.

Whether a technology developer plans to implement cloud or edge inferencing for their AI-powered medical device, a trusted hardware solutions partner like MBX can offer strategic guidance for designing, building, and integrating the product. Join MBX in empowering medical technology developers to adapt to an AI-centric future in the healthcare industry.

ABOUT MBX SYSTEMS

MBX Systems provides purpose-built and deployment-ready hardware platforms on a foundation of customized services and interactive software tools for technology companies that deliver complex products as integrated hardware/software solutions. Building on 25+ years of experience and product deployments in 185 countries, the MBX ecosystem features MBX Hatch™, the industry's most advanced toolset for orchestrating hardware program data and action. Hardware solutions are manufactured in ISO 9001:2015 and ISO 13485:2016 certified facilities using the award-winning Forge™ infrastructure developed by MBX to automate customers' high variability manufacturing requirements for faster time to market. For more information, visit **www.mbx.com.**

mbx systems

mbx systems

mbx.com

+1 847.487.2700
SALES: 800.560.1195
SUPPORT: 888.440.1617

1200 Technology Way, Libertyville,
Illinois 60048 | MBX Systems © 2023

Contact MBX Systems and let us help you bring innovative AI-powered medical devices to market.

CONTACT US