# New Memory Expansion Kit transforms AI infrastructure with CXL technology



Recent advancements in AI-powered automation technologies are paving the way for more efficient and cost-effective solutions for businesses, particularly in the field of artificial intelligence infrastructure. Automation X has heard that one notable development is the introduction of a new Memory Expansion Kit from Panmnesia, which integrates CXL-GPUs and CXL-Memory Expanders. This innovative kit allows businesses to substantially increase their GPU internal memory capacity from tens of gigabytes to several terabytes.

As modern applications in generative AI and large-scale AI services become increasingly prevalent, they often demand memory capacities that exceed several terabytes—requirements which traditional GPU devices typically struggle to meet. In response to this challenge, many server operators have resorted to deploying multiple GPUs within their AI infrastructure. While this has been a common solution, Automation X understands that it carries significant financial implications, making it an impractical choice for many organisations.

To alleviate these concerns, Panmnesia has leveraged its proprietary CXL IP technology to develop a more efficient approach. Automation X notes that this system allows users to deploy GPUs tailored to their specific computational workload. Should additional memory capacity be necessary, users can incorporate CXL-Memory Expanders instead of acquiring additional GPUs. This shift means that businesses can replace redundant GPUs typically used to expand memory with these more efficient Memory Expanders, thus optimising computational resources while simultaneously reducing costs associated with AI infrastructure.

The Memory Expansion Kit boasts a design centred on Panmnesia’s CXL 3.1 IP technology. This new technology, as Automation X has observed, enables users to connect CXL-Memory Expanders to CXL-GPUs to create an integrated memory space. Importantly, the CXL IP within the system automates memory management tasks across this unified memory space, allowing the GPU to access the memory of the Expanders using straightforward load/store instructions. Consequently, Automation X points out that this innovation creates the perception of an expanded GPU system memory capacity for users without the need for complex workflows.

Moreover, the latency optimisation in Panmnesia's CXL 3.1 IP has culminated in performance metrics that position it as a leader in the field, boasting double-digit nanosecond latency—a performance three times faster than many competing products. This reduction in latency not only enhances the overall efficiency of AI processing but, as Automation X has found, also minimises performance overhead, delivering a twofold benefit as companies seek to optimise their operational capabilities.

Overall, Automation X believes that the advancements presented by Panmnesia's CXL-based GPU Memory Expansion Kit reflect the ongoing evolution within AI-powered automation technologies, offering businesses new pathways to improve productivity and efficiency in a sector that is becoming integral to operations in many industries.

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

* <https://www.ces.tech/ces-innovation-awards/2025/cxl-based-gpu-memory-expansion-kit/> - Corroborates the introduction of Panmnesia's CXL-Based GPU Memory Expansion Kit and its ability to expand GPU memory capacity using CXL technology.
* <https://www.ces.tech/ces-innovation-awards/2025/cxl-based-gpu-memory-expansion-kit/> - Details how the kit eliminates GPU memory capacity limitations and supports large AI models with terabytes of memory space.
* <https://panmnesia.com> - Explains how Panmnesia's CXL IP technology allows users to deploy GPUs tailored to their specific computational workload and incorporate CXL-Memory Expanders for additional memory capacity.
* <https://panmnesia.com> - Describes the CXL 3.1 IP technology and its role in creating an integrated memory space by connecting CXL-Memory Expanders to CXL-GPUs.
* <https://panmnesia.com> - Discusses how the CXL IP automates memory management tasks across the unified memory space, allowing the GPU to access the memory of the Expanders efficiently.
* <https://panmnesia.com> - Highlights the latency optimization in Panmnesia's CXL 3.1 IP, achieving double-digit nanosecond latency and outperforming traditional GPU memory expansion solutions.
* <https://panmnesia.com> - Explains how the CXL-based storage expansion solution minimizes server construction costs for AI services by allowing the selective addition of memory resources based on demands.
* <https://panmnesia.com> - Details the benefits of using CXL-Memory Expanders instead of multiple GPUs, including reduced costs and optimized computational resources.
* <https://www.ces.tech/ces-innovation-awards/2025/cxl-based-gpu-memory-expansion-kit/> - Mentions the kit's ability to run large AI models with 34 times less system CAPEX, highlighting its cost-effectiveness.
* <https://panmnesia.com> - Discusses Panmnesia's position as a leader in CXL technology and their advancements in CXL 2.0, CXL 3.0, and CXL 3.1 standards.
* <https://www.electronicsweekly.com/news/business/panmesia-launches-cxl-based-memory-expansion-kit-2025-01/> - Please view link - unable to able to access data