

Blackbox.ai uses SambaNova Cloud to power agentic workflows

CyberCoder autonomous agent leverages the high performance and low latency of SambaNova to automate tasks and increase developer productivity

Challenge

Blackbox.ai builds autonomous coding agents that empower their customers to quickly build custom applications. By using the Blackbox autonomous agent, CyberCoder, developers can quickly and easily edit multiple files simultaneously, a task that was previously complex and time consuming. By enabling users to build applications from scratch, edit existing code, and properly document it, Blackbox.ai is streamlining workflows and enhancing developer productivity.

To deliver on the needs of their customers, Blackbox is taking advantage of the latest Llama models from Meta. “We have done multiple experiments and have been running the Llama models since their first release” said Robert Rizk, Blackbox Co-founder and CEO, “The Llama models are very good at diverse use cases. We are able to fine tune them and get them to work really well on the tasks we want to use them for.”

Quickly growing to over 10 million monthly active users and a client base that includes Fortune 500 companies, Blackbox needed a platform that could automate and power their autonomous agent tasks and user queries as well as meet the significant performance demands of their customers today and seamlessly scale to meet their changing requirements.

Solution

The SambaNova Cloud provides developers with the fastest inference performance across the latest open-source models, including the Llama 3.1 and 3.2 models. Powered by the industry leading SambaNova DataScale system, SambaNova Cloud delivers incredible inference performance across a wide range of models, all running at full precision.

At the heart of the SambaNova technology is the SN40L Reconfigurable Dataflow Unit (RDU). With a revolutionary dataflow architecture and unique three-tiered memory architecture, the SN40L is capable of running hundreds of models at the same time and switching between them in as little as two microseconds. This capability is vital to agentic workflows, as it dramatically reduces the latency involved when switching between different models.

Results

Blackbox needed to be able to meet the rapidly growing performance demands of their customers, and their existing vendors could not meet their needs. “We reached out to SambaNova when we saw the post about SambaNova Cloud from Andrew Ng”, said Rizk, “We are using Llama 8B and 70B for easier tasks and 405B for more complex tasks. We needed a platform that could run all of those models with low latency and high performance.”



Blackbox took advantage of the SambaNova Cloud to automate CyberCoder tasks, including building applications from scratch, debugging code, and adding features to existing programs. “We are using 100s of requests per second to the SambaNova Cloud and with the 8B model we are seeing 3x-4x faster completions and are planning to increase our utilization by the end of 2024” said Rizk, “High performance and low latency are significant to us. The agentic workflow is longer by nature, but having faster inference and low latency on these models is allowing us to optimize for a better user experience.”

“It is not just the technology that is really great, the support is also really great”, said Rizk, “The speed at which the SambaNova team responded to us and supported us during the testing and the production phase is outstanding and was a real differentiator.”

To learn more about SambaNova Cloud, visit: cloud.sambanova.ai



SambaNova

To learn more about how SambaNova Systems can accelerate and transform your organization with generative AI, [schedule a meeting](#).

Learn more at SambaNova.AI

 [linkedin.com/company/sambanova](https://www.linkedin.com/company/sambanova)

 [@SambaNovaAI](https://twitter.com/SambaNovaAI)

 info@sambanova.ai

Customers turn to SambaNova to quickly deploy state-of-the-art generative AI capabilities within the enterprise. Our purpose-built enterprise-scale AI platform is the technology backbone for the next generation of AI computing.

Headquartered in Palo Alto, California, SambaNova Systems was founded in 2017 by industry luminaries, and hardware and software design experts from Sun/Oracle and Stanford University. Investors include SoftBank Vision Fund 2, funds and accounts managed by BlackRock, Intel Capital, GV, Walden International, Temasek, GIC, Redline Capital, Atlantic Bridge Ventures, Celesta, and several others. Visit us at sambanova.ai or contact us at info@sambanova.ai. Follow SambaNova Systems on LinkedIn and on X.