Talent Identification with SambaNova

World record performance and accuracy for secure generative AI

Challenge

The US military needs a robust system to identify and place individuals who exceed standards in appropriate roles. This will ensure efficient talent deployment and leadership development.

Objective

Deploy the SambaNova generative AI platform to analyze employee performance data, career trajectories, and skill sets from the existing evaluation databases for both enlisted personnel and officers. Custom, pre-trained trained large language models will be used to identify patterns and predict high-potential individuals for leadership roles and career advancement, ensuring optimal talent utilization across the agency.

Distinguishing Factors

- Multi-source/multi-format parsing support
- World record performance and accuracy
- Data security & privacy
- Model/data ownership
- Full stack solution
- On-site/cloud services

Justification

The SambaNova Composition of Experts generative AI model leverages "1.3T parameters" and "92 experts" to analyze employee performance data, ensuring a comprehensive understanding of individual capabilities. This level of analysis is critical for accurately identifying high-potential individuals for leadership roles and career advancement.

Solution

SambaNova Suite, combined with the Samba-1 Composition of Experts (CoE) model, delivers the highest combination of performance and accuracy for generative AI. A complete hardware/software platform, SambaNova Suite can be deployed as a fully configured rack-level solution either on-premises, including air-gapped environments, or as a cloud-based solution. The Samba-1 model provides state-of-the-art accuracy across a wide range of use cases.

Key Features

Performance, Accuracy, and Efficiency
Only SambaNova delivers performance of over 1000 tokens/s at full precision, on as few as 16 sockets. This is an unrivaled combination of performance and accuracy with a small footprint, dramatically reducing power consumption.

Total AI Stack Control
With options for on-premises and air-gapped deployments, SambaNova ensures complete data privacy and security, enabling the use of sensitive data for model training without external exposure.

Model Ownership
Once a model is fine-tuned, it becomes the property of the customer in perpetuity, eliminating model ownership concerns.

Configurable Access and Permissions
The Samba-1 model enables flexible access controls, allowing precise management of model permissions to align with organizational roles, enhancing security and ensuring compliance.

SambaNova Suite delivers the most accurate, integrated full stack generative AI platform, optimized for enterprise and government organizations. Delivered through SambaNova's integrated AI platform, which can be deployed on-premises or through the cloud, SambaNova Suite seamlessly integrates into existing business processes to deliver transformative capabilities. With the capacity to be further refined using customer data, these models can deliver unparalleled accuracy while providing enterprise grade security and data governance.

Learn more