

Dataflow-as-a-Service


A transformative deep learning platform, designed to empower enterprises to train and deploy state-of-the-art models with scale, accuracy, and speed of deployment

SambaNova Dataflow-as-a-Service™ delivers the most powerful integrated deep learning platform for the training and deployment of state-of-the-art models, leveraging the most recent advances in AI. Offered as a product suite including Language, Vision, and Recommendation, Dataflow-as-a-Service leverages different forms of data for use in a diverse set of solutions. Powered by an AI infrastructure purpose-built for deep learning applications, and with out-of-the-box pre-trained deep learning models, Dataflow-as-a-Service enables organizations to accelerate and scale their AI capabilities in weeks, not years.


Dataflow-as-a-Service: An extensible platform for large deep learning models

Running end-to-end workflows for training and deployment of large deep learning models, such as GPT or computer vision models with large image sizes, is time consuming, complex, resource intensive, and in some cases technologically unachievable. In addition to requiring a team of expert engineers and data scientists, advances in AI/ML software and algorithm development created a need for computing infrastructure beyond existing technologies provided by CPUs and GPUs. These complexities have significantly decelerated the adoption of AI in enterprises. With pre-trained large language and computer vision models running on the SambaNova Reconfigurable Dataflow Architecture™ — a software-defined hardware architecture purpose-built for deep learning workloads, the platform offers unparalleled ease of use and rapid time to value eliminating years of iteration required to deploy complex AI models into production. Offered as a subscription model and fully managed by SambaNova, Dataflow-as-a-Service eliminates the heavy upfront investment into AI infrastructure set-up, IT support, commitment to monopolized cloud services' fees, and securing a large pool of AI and ML talent. Instead, it enables organizations to quickly train and deploy state-of-the-art models on a single platform purpose-built for enterprise-grade applications.


BENEFITS




Scale training and deployment of large, state-of-the-art deep learning models




Achieve value from AI in weeks, not years



Deploy state-of-the-art AI and deep learning models without the need for hard-to-find infrastructure



Elevate efficiency with unparalleled accuracy, flexibility, and performance



Leverage fully managed services and reduce complexity with easy-to-use and flexible low-code/no-code APIs

Key Features

- **An integrated platform** for training and inference offering a comprehensive catalog of turnkey pre-trained, already-configured state-of-the-art NLP, computer vision, and recommendation models including GPT 1.5B, 13B, and 175B, HuBert, RescaleNet, UNET, Retinanet, ViT, and DLRM
- **Streamlined workflows** combining pre-trained checkpoints, domain specific training and task specific training to achieve higher accuracy
- A comprehensive selection of **large language** models leveraging text and speech data for various tasks, such as name entity recognition, sentiment analysis, document classification, semantic search, question answering, text generation, speech-to-text, and intent classification
- A comprehensive selection of **vision** models supporting a collection of tasks including image classification, object detection, and 2D and 3D segmentation to support true resolution images up to 50k x 50k
- **Recommendation** models supporting large embeddings with unmatched accuracy and performance
- **Various interfaces** targeting different levels of expertise, including REST APIs, python SDK, command line, and a browser interface
- **Various functional blocks** for ML operation training and deployment workflows, job scheduling and management, and system and application monitoring and logging
- **Flexible deployment options** including on-premises, cloud deployment or colocation
- **Fully managed services** by SambaNova offering seamless integration with enterprises' infrastructure and existing services, such as identity management and storage




Data Privacy, Security, and Deployment

SambaNova Systems understands that inputs to Dataflow-as-a-Service, including our customers' data, are among the most valuable business assets. We are fully committed to protecting and securing our customers' data against unauthorized or unwanted use and adhere to best-in-class data privacy and security standards across all storage and deployment scenarios. Furthermore, our AI models and solutions do not have a memory across client systems. Therefore, none of our customers' data will be used or indirectly leveraged with other customers and it is not seen by anyone outside their organization.



Solution Options

Dataflow-as-a-Service provides organizations with a suite of solutions with flexible subscription options for natural language processing, high-res computer vision, and recommendation systems.

	 Language	 Vision	 Recommendation
Overview	Develop and deploy highly-customized language models—efficiently at scale	Unlock the full advantage of your high-resolution images beyond today's limits	Deliver more personalized recommendations than ever before
Catalog options	Pre-training, fine-tuning, inference, co-design	Training with low to high resolution images, inference, co-design	Training, real-time online inference, co-design
Use cases	Billion-parameter models	True resolution images beyond 5K ² or 128 ³	Terabyte-sized embedding tables
Features	1. Enterprise-level support 2. Pre-configured models ready to take data 3. Flexible deployment options 4. Predictable subscription pricing 5. Flexible configuration		

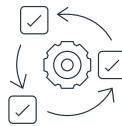
The Fastest, Easiest, and Guaranteed Path to AI

Fast track and simplify your AI adoption with fewer iterations, less investment in talent and infrastructure and guaranteed production grade solutions.

Accelerate time to value for AI innovation



Stay up to date with advances in AI hardware and software innovation



Augment your ML expertise



Next Steps

Contact SambaNova Systems to learn more about how our solutions can accelerate your AI journey.

Learn more at SambaNova.ai



linkedin.com/company/sambanova



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



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



info@sambanova.ai