

DISCOVER WHAT'S
POSSIBLE WITH
NEXT-GENERATION AI



”

This pre-AI to post-AI transition that the world is going through is going to be as big, if not bigger, than the internet. The magnitude of changes that the internet brought to every industry will pale in comparison to the transformation driven by AI.

RODRIGO LIANG, CO-FOUNDER AND CEO



CONTENTS

SECTION 1

Paving the Way to
Next-Gen Innovation

4

SECTION 2

The Advent of
Deep Learning

7

SECTION 3

Roadblocks
to Progress

10

SECTION 4

A Flexible New
Design Architecture

15

SECTION 5

From Zero to
State-of-the-Art
AI Overnight

19

SECTION 6

Customer
Testimonials +
Conclusion

23



SECTION 1

PAVING THE WAY TO NEXT-GEN INNOVATION



Disruptive Innovations vs Sustaining Innovations

History is the story of what Harvard Business School professor Clayton Christensen dubbed “Sustaining innovations” and “Disruptive innovations.”

From carriages to automobiles, floppy disks to hard drives, mainframes to handheld devices, disruptive innovations have been the catalyst for societal flourishing.

However, most companies produce sustaining innovations, developing incrementally better products that yield short-term profits at the risk of being left behind by more nimble upstarts.



Technological Changes Come in a Blink of an Eye

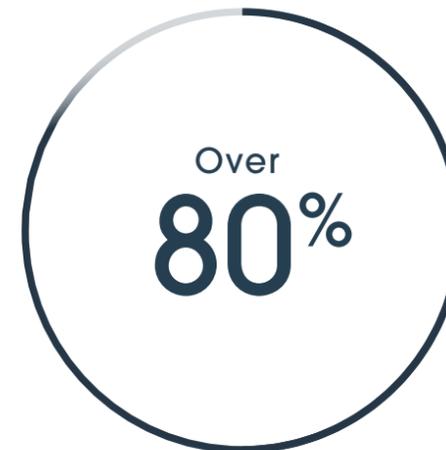
Technological change comes fast, displacing existing incumbents who aren't prepared. By the time companies realize their predicament and start playing catch-up, new leaders are born and established players risk extinction.

Case in point, over 50% of the companies in the Fortune 500 in 2000 no longer exist today. And over 80% of companies who have tried to adopt digital transformation have failed.

The determining factor was whether companies embraced change and were early adopters of new technology, or just wait-and-see bystanders.



Companies in the Fortune 500 in 2000 that no longer exist today.



Companies that have tried to adopt digital transformation but have failed.

Sources: (Top chart) Business Apocalypse: Fifty-two Percent of Fortune 500 Companies from the Year 2000 Are Extinct <https://ryanberman.com/glossary/business-apocalypse/> (Bottom chart) Couchbase: Digital Transformation Projects Continue to Be at Risk, Couchbase Research Finds <https://www.couchbase.com/press-releases/digital-transformation-projects-continue-to-be-at-risk-couchbase-research-finds>

SECTION 2

THE ADVENT OF DEEP LEARNING



Deploying Neural Networks

In 2012, researchers took decades-old techniques called neural networks and demonstrated that they could train deeper and more efficiently, utilizing hardware once reserved for computer graphics, the GPU.

Beating all previous benchmarks on the famous ImageNet dataset, neural networks now encapsulated under deep learning have become ubiquitous as the state-of-the-art in areas like Computer Vision, Natural Language Processing, and Recommender Systems.

From chatbots that can converse on the finer points of finance, to cancer detection in MRIs, to self-driving cars, applications that were once considered science projects are now seeing the light of day in real-world deployments.



The introduction of billion-parameter models, such as GPT-3, have redefined what we thought possible and surpassed human performance in many domains.

However, the rise of these more powerful models and exponentially growing datasets have strained legacy systems and have necessitated new software and hardware architectures. While GPUs ushered in this first wave of deep learning advancement, we are already pushing up against the hardware's limitations.

To stay ahead of the innovation curve, the solution to date has been to create specialized teams of deep learning experts to run training workloads on thousands of GPU cores.

This approach creates a whole set of challenges, such as physical space constraints, spiraling costs, and specialized knowledge that together exclude all but the largest enterprises from reaping the rewards of AI.

Challenges to Innovation:



Physical space constraints



Spiraling costs



Specialized knowledge

SECTION 3

ROADBLOCKS TO PROGRESS



”

If you look at those who can develop these systems, they're in the hands of the few. The large companies that have the data, the computation, and the talent to develop these sorts of algorithms have used these systems to become the most valuable companies in the world—Google, Apple, Amazon, Facebook, and the like.

KUNLE OLUKOTUN, CO-FOUNDER AND CHIEF TECHNOLOGIST

Obstacles in Putting AI Strategies Into Action

Companies of all sizes face seemingly insurmountable obstacles when it comes to putting AI strategies into action. To start, AI and ML talent is expensive and hard to find, leaving most organizations with critical skills gaps.

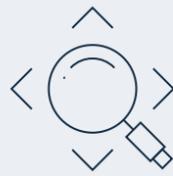
Once an organization manages to procure the talent they require, they'll need expertise in different computing architectures to build integrated systems and make models performant, meeting SLAs and other requirements.

Lastly, organizations find it difficult to stay up to date on the right solution and keep it relevant with constantly changing data.



Roadblocks to AI

1.



Talent scarcity in AI/ML

Leveraging popular deep learning frameworks and software ecosystems requires technical expertise that is expensive and in short supply.

Where once building and deploying deep learning models was limited to hyperscale enterprises that could afford to hire hundreds of skilled technical developers, today's organizations want the same AI superpowers as hyperscalers, with only a fraction of the budget.

2.



Volume of Data

The demand for compute can no longer keep up with today's massive amounts of data.

Data scientists are constantly looking for workarounds either by decreasing the resolution on their datasets, working with subsamples, or deliberately selecting less powerful models trading accuracy for viability.

These trade-offs compromise the quality of end results and prevent businesses from fully realizing the promise of AI.

3.



The DIY Approach

When it comes to AI, going with the "build your own" model can take 18-24 months, delaying access to insights and new revenue streams, with no guarantee of success. In fact, 85% of AI projects deliver erroneous outcomes, presenting daunting odds and costly risk on the hard road to delivering real value.*

Optimizing new systems and keeping models up to date requires access to ever-evolving and scarce AI expertise.

*Source: Gartner: Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence <https://www.gartner.com/en/newsroom/press-releases/2018-02-13-gartner-says-nearly-half-of-cios-are-planning-to-deploy-artificial-intelligence>

There is a Better Way

To solve the many obstacles and roadblocks, SambaNova Systems has developed an innovative hardware and software platform that provides access to state-of-the-art machine learning solutions to large and small organizations alike.

Built by one of the world's leading chip teams and software leaders from across academia and industry, SambaNova's DataScale® and Dataflow-as-a-Service™ enable digital transformation through disruptive AI innovation.



SECTION 4

A FLEXIBLE NEW DESIGN ARCHITECTURE

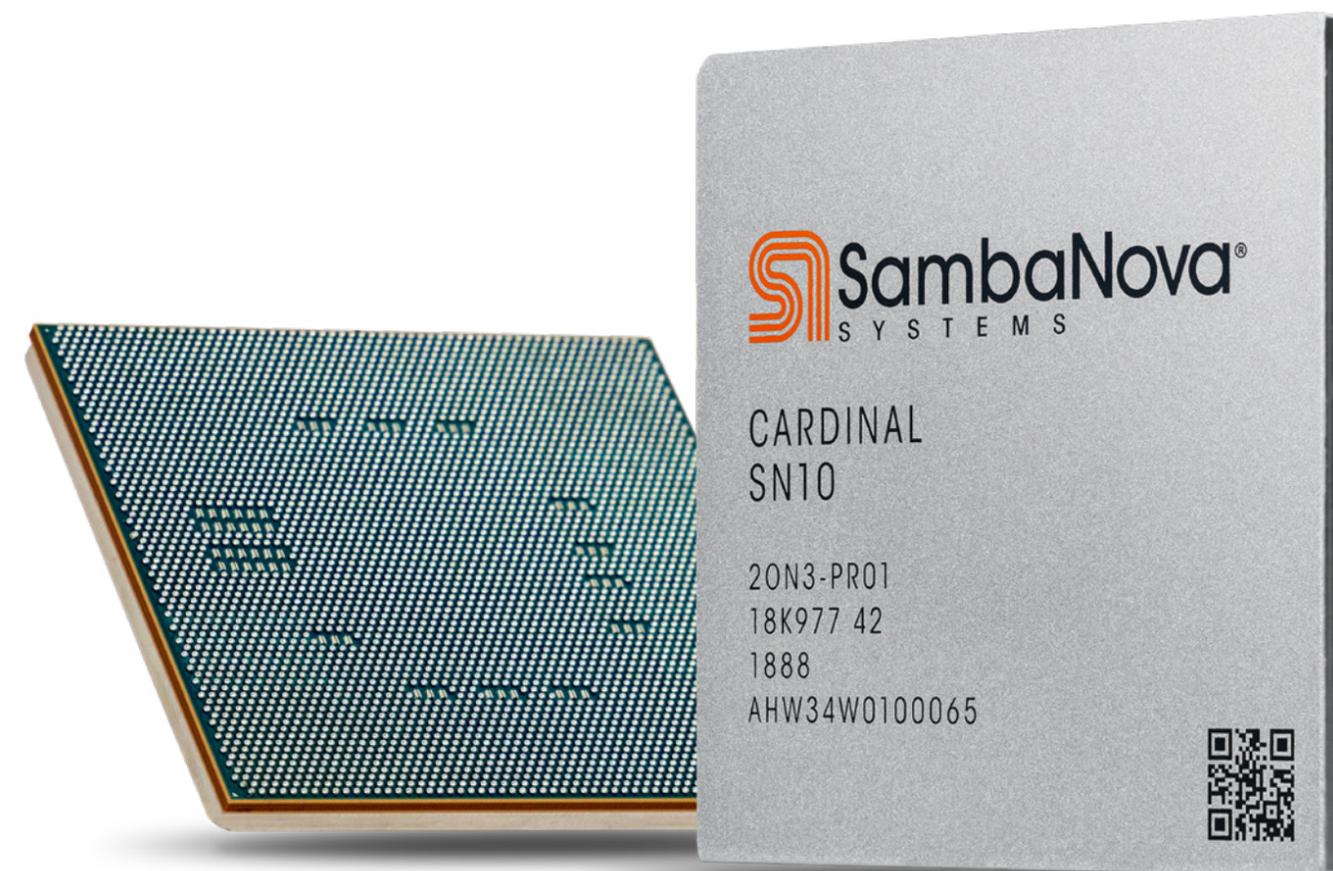


Reconfigurable Dataflow Architecture

With its flagship DataScale product, SambaNova delivers a complete, full-stack solution that incorporates innovations at all layers, including algorithms, compilers, system architecture, and state-of-the-art silicon—the Reconfigurable Dataflow Architecture™ (RDA).

DataScale provides a flexible execution model that pipelines operations and decomposes dataflow graphs into parallel patterns—key parts of all deep learning model architectures—resulting in acceleration in model training and inference, thus reducing time to deployment.

RDA enables programmable data access patterns and minimizes excess data movement found in fixed, core-based, instruction set architectures (ISAs). Instead of fixed traditional architectures, RDA is programmed specifically for each model resulting in a highly optimized, application-specific accelerator.



Future-Proof Your Infrastructure for Next-Gen AI Workloads

DataScale helps users achieve unparalleled efficiency and performance across a broad range of applications. Built with open standards and interfaces, it is designed to avoid disruption and seamlessly integrate into existing on-prem infrastructures or cloud-based environments.

SambaNova's Reconfigurable Dataflow Unit (RDU) is the industry's next-generation processor purpose-built to support AI/ML workloads. RDUs are designed to allow data to flow through the processor as it should—freely and without bottlenecks.

RDUs eliminate constant data caching and excess data movement inherent to today's core-based architectures. This unlocks significant silicon utilization to unleash more compute power than any other solution available today.



For the first time, there is a commercially-available alternative to NVIDIA's GPU for deep learning.

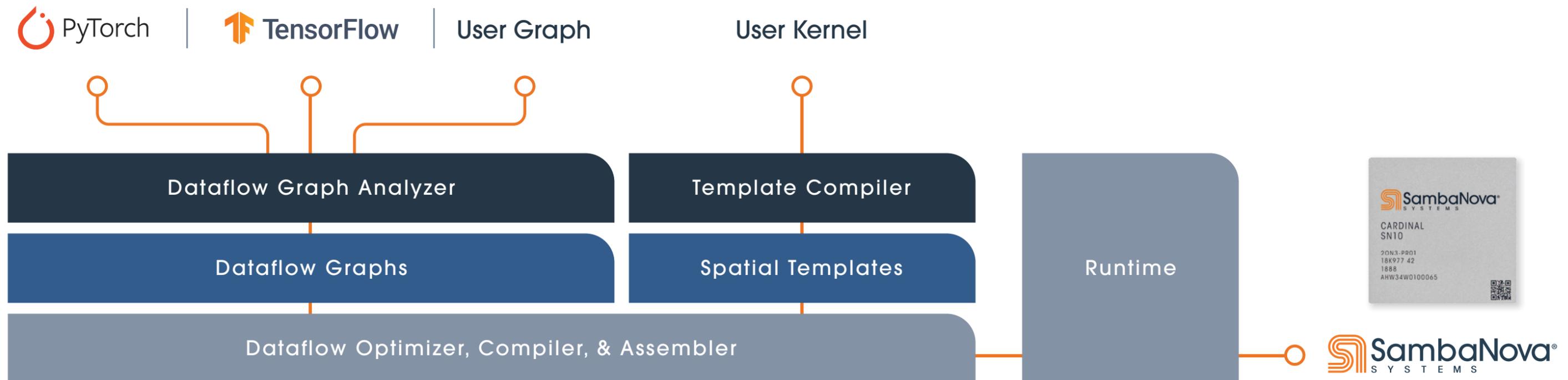
RODRIGO LIANG, CO-FOUNDER AND CEO

The Complete Software Stack

The complete software stack—SambaFlow—is designed to take input from standard machine-learning frameworks, such as PyTorch and TensorFlow. It automatically extracts, optimizes, and maps dataflow graphs onto RDUs, allowing high performance without the need for low-level kernel tuning.

The result of all these innovations is a reimagination of what could be done in areas such as Computer Vision, Recommendation Systems and Natural Language Processing.

To borrow from Henry Ford, SambaNova hasn't raised a faster horse (GPU-based solutions), it has built the first car (Dataflow-as-a-Service)—redefining transportation (AI) as we know it today.



SECTION 5

FROM ZERO TO STATE-OF-THE-ART AI OVERNIGHT





A new phenomenon has started to emerge in the last few years: machine learning has begun to eat software. Machine learning is radically changing how one builds, deploys, and maintains software.

CHRIS RE, CO-FOUNDER

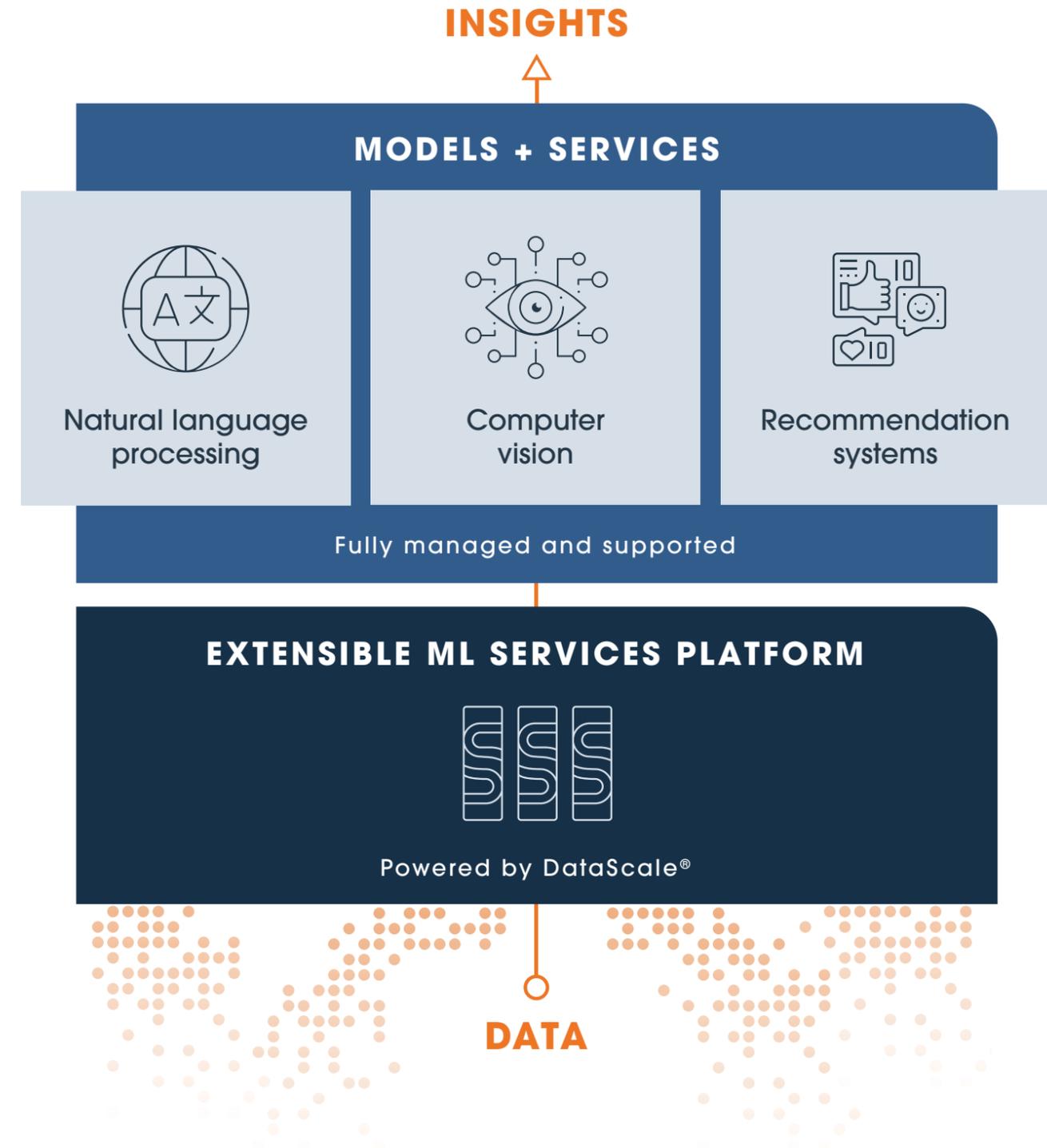


Dataflow-as-a-Service

The past decades were defined by imperative programming, where engineers and scientists explicitly and meticulously crafted algorithms by hand. Today, the revolution in computing is defined by Software 2.0, or dataflow, where data has become the electricity of the digital age.

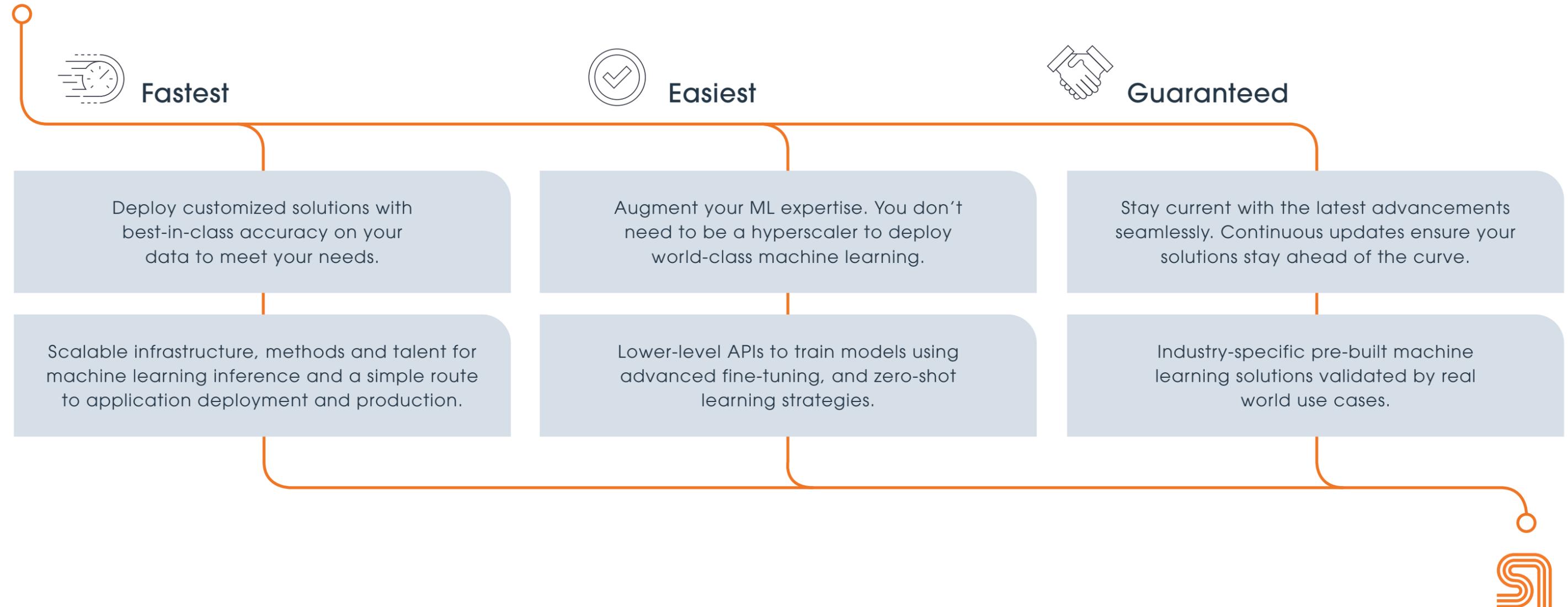
Powered by the DataScale hardware engine, SambaNova has released Dataflow-as-a-Service to provide everyone with access to state-of-the-art machine learning models.

With Dataflow-as-a-Service, organizations outside of Big Tech can offload complexity to stay focused on their core competencies while deploying AI solutions with confidence.



Augment Your Machine Learning Expertise

A Dataflow-as-a-Service subscription gives companies a path to AI.



SECTION 6

CUSTOMER TESTIMONIALS + CONCLUSION



Customer Testimonials

Lawrence Livermore National Laboratory



The support for running multiple models built into the hardware was a significant aspect of our selection of SambaNova Systems. We’re seeing roughly a 5x improvement compared to a comparable GPU running the same models.

BRONIS DE SUPINKSI, CTO

Lawrence Livermore National Laboratory (LLNL) uses large-scale systems for various physics simulations. Leveraging AI to achieve significant performance improvement for scientific applications, LLNL actively explores cognitive simulation using SambaNova’s technology. Data is moved from supercomputers to the SambaNova system over LLNL’s high-speed network and ingested into its models and integrated into its overall simulation processes.



Customer Testimonials

Argonne National Laboratory



We're interested in technology that can train models faster than individual GPUs, that has more scalability and has more performance, and so the SambaNova DataScale System met that criteria.

RICK STEVENS, ASSOCIATE LABORATORY DIRECTOR

Argonne National Laboratory uses AI for cancer and Covid-19 research. In particular, its AI models are used for drug development to automate multiple parts of the development pipeline. It is also used in cosmology to look for evidence of dark matter and in environmental sensing. The time it takes to build and train models is critical for research. Argonne sought out technology that was faster than individual GPUs, had more scalability, and could deliver more performance for its deep learning models. On both small and large-scale models. Argonne achieved orders of magnitude performance improvement over GPUs through SambaNova's naturally scalable system.



Level the Playing Field With Next-Generation AI

1

SambaNova, through its flagship DataScale hardware architecture, has optimized dataflow from silicon to algorithms and has achieved record-breaking performance and accuracy.

2

Dataflow-as-a-Service opens up a new realm of digital transformation available to everyone. Now all organizations can access accurate and powerful AI solutions at less cost and complexity.

3

With state-of-the-art accuracy, performance, scale, and ease of use, SambaNova has positioned itself as the definitive leader in the next generation of AI.

LEARN HOW YOUR ORGANIZATION CAN HARNESS DATAFLOW
TO REDEFINE YOUR INDUSTRY — AND THE WORLD.

Visit us on the web at SambaNova.AI



SambaNova Systems is an AI innovation company that empowers organizations to deploy best-in-class solutions for computer vision, natural language processing, recommendation systems, and AI for science with confidence. SambaNova's flagship offering, Dataflow-as-a-Service, helps organizations rapidly deploy AI in days, unlocking new revenue and boosting operational efficiency. SambaNova's DataScale® is an integrated software and hardware system using Reconfigurable Dataflow Architecture (RDA), along with open standards and user interfaces. Headquartered in Palo Alto, California, SambaNova Systems was founded in 2017 by industry luminaries, 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. For more information, please visit us at sambanova.ai or contact us at info@sambanova.ai. Follow SambaNova Systems on LinkedIn.