

How to Make AI Deployment Easier



Co-created by

Emerj Artificial Intelligence and **SambaNova Systems**



Introduction

Recent advances in AI technology are driving increasing investments in the latest machine learning (ML) innovations in areas such as natural language processing, computer vision, and recommendation. Enterprises pursue these leading-edge technologies to unlock new capabilities and research the innovations that will drive tomorrow's competitive landscapes. To the victors will go the spoils, but only if they succeed in easing the deployment of AI and freeing up the time, effort, and resources needed to speed its promised innovations post-implementation and beyond.

Across disparate industries, leaders search for the expertise, experience, and resources to develop and deploy AI solutions for their most pressing problems:

- **Financial services** firms pursue the latest AI technologies to add personalization to their offerings and build more responsive fraud detection and risk management functions.
- **Healthcare and life sciences** companies look to AI to enhance patient care and outcomes while advancing personalized medicine and improving the patient experience.
- **Manufacturers and automakers** turn to AI to deliver on their Industry 4.0 objectives and pursue accelerations and efficiencies that deliver time and cost savings.
- **Retail and e-commerce companies** are leveraging AI to transform the customer journey and embrace digital acceleration to understand their customers and remain relevant amidst rapid competitive evolution.
- **Oil and gas companies** invest in AI to drive improved decision-making and innovate the way they explore, produce, distribute, and deliver today's energy products.
- **Public sector** institutions and companies look to overcome legacy systems and talent scarcity to deploy AI solutions that better support and advance government missions.

“What we’re seeing are three areas that are driving interest in the market,” says Marshall Choy, VP, Product at SambaNova Systems. “Those are computer vision, language processing, and recommendation systems.”

Digitally aware companies face many strategic decisions and challenges on the road to AI adoption. Indeed, as companies increasingly embrace the digital age, how they select and architect their AI adoption path becomes even more critical. One of the first decisions they will make determines how they will acquire and stand up their solution.

Will they pursue the do-it-yourself approach (e.g., leveraging GPUs) and go at it alone, acquiring the people, materials, and resources needed to construct, deploy, and support their AI solution? Or will they embrace the new platform approach and partner with an AI industry expert such as SambaNova Systems whose Dataflow-as-a-Service™ model can cut the time and resources needed to get the solution online and delivering on its objectives?

In this paper, we’ll meet with Marshall Choy, VP, Product at SambaNova Systems to explore the realities and considerations of each approach for companies who face the challenges of AI deployment as they embark on developing and deploying models to accelerate their AI workloads.

The Do-It-Yourself Approach

As companies trace their own paths to AI adoption, some will choose the traditional do-it-yourself (DIY) route where they walk the road alone. These companies ‘do it themselves,’ building their models by acquiring the people, materials, and resources they need on their own. According to Marshall Choy:

“The do-it-yourself approach is something we see used by companies that are focused more on R&D activities and experimentation. They want to do a lot of that R&D-type of work on the infrastructure itself as well as on the higher levels of the stack. We do have a certain class of folks who really want to build custom, one-off models. Maybe they view that as a competitive advantage. Maybe they view that as a core competency. That’s not for everybody.”

When a company chooses the do-it-yourself approach to develop its AI infrastructure and models, the choice is often born out of necessity. Traditionally, IT and tech industry vendors have sold their customers components and piecemeal parts that the end-user has assembled and deployed, forced to play the role of the final integrator.

“There are very few other industries that do that,” says Choy. “You don’t buy cars, bicycles, or homes in parts. You buy them as complete units. In AI, we’re seeing more and more of a shift toward people wanting that.”

There will always be companies who want to use the do-it-yourself approach and build their AI models from scratch. Let's look at some of the reasons why. This approach appeals to companies who seek:



Control over the design of the solution

Companies may see this as a competitive strength and strategic advantage when they want a custom solution for a one-off function.



Security for sensitive or regulated data

Companies may seek to keep all their proprietary, sensitive, and confidential data in-house to reduce the risks of data theft, piracy, and misuse.

With this approach, whether your firm happens to be a bank, a pharma, or even a cloud services provider, what you are actually providing to your end-users is infrastructure-as-a-service. While this approach offloads many tasks to your end-user, it often results in difficult work related to AI and machine learning.

The core challenges surface around software and infrastructure and, in particular, models. With neural network models, companies choosing this approach need to select which model to run, tune, optimize, and maintain.



"Not everyone has the hundreds of ML practitioners to go stand all that up," says Choy, "And even if they do, maybe they don't want to continue having to do that. There's got to be a better way."

The Do-It-Yourself Approach: The Cons



The DIY approach comes with the detail and complexity of the entire model development and deployment process. The process can last many months, sometimes years, requiring hundreds or thousands of person-hours while the company:

- Selects the hardware
- Architects the solution
- Performs the planning and design
- Benchmarks its requirements against other models
- Sizes the solution
- Tunes and optimizes its functioning and output
- Deploys the solution
- Maintains the solution indefinitely

Confronted with the challenges of going it alone, many who choose the DIY route face a longer time-to-value (TTV) and higher opportunity costs.

Those higher costs and delayed TTV materialize in a variety of ways. Companies that opt for a DIY approach must:

- Keep up with AI innovations and technology
- Overcome AI talent scarcity and recruit ML experts
- Pay the high upfront costs of model development long before realizing any return
- Address the limitations of a DIY infrastructure such as memory latency
- Face the challenge of supporting their unique AI solution

“We talk to people who do both things,” says Choy. “So, there are different horses for different courses.” While many companies have turned to the DIY approach in the past, they now have a new option available.

Companies can now consider the platform approach, which delivers everything they need to create and deploy their AI model with the convenience and speed of a simple subscription format.

Keeping Up with the Pace of Technological Change

As the adoption rate for machine learning models continues to increase, research into new innovations grows in tandem, with new papers released every day. Keeping up with the rate and pace of model development has emerged as one of the biggest challenges facing AI practitioners.



Every new paper effectively means a new way to do something, differently and better. “Most people just don’t have time or the desire to keep up with it,” says Choy.

Many turn to their vendors in the AI and ML space and rely on their experience to keep up with the latest industry innovations. This approach puts economies of scale to work—bolstering a new appetite in the market for Platform-as-a-Service models. It also allows companies to rely on experts to chase the latest and greatest AI technologies.

Choosing Your Right Path

Very large technology companies overwhelmingly choose the do-it-yourself approach because they have large pools of time and resources to draw from. That has driven the market to support a very small number of providers who offer complete systems and an even smaller subset offering complete integrated systems.

“It’s about choosing the right level of abstraction,” says Choy. “Some people want to cobble together the solution with a plethora of subsystem components. They effectively build their own hardware based on components.”

For companies with all but the largest budgets, raising the abstraction away from subsystem components and the systems themselves frees up time and lessens the opportunity cost of tied-up resources.

“When companies have broader infrastructure solutions that encompass not just compute, but also network and storage,” observes Choy, “that minimizes system integration headaches that are at a very low and tactical level. That enables you to focus more on your performance, power efficiency, scale, and ease of use. You can turn the keys over to the developers more quickly in terms of time-to-value.”

How do you determine what the right path is for your organization? Choosing the right path comes down to comparing the value you add against the inputs you have invested. Consider the following questions:

- **How costly is the effort relative to the value of the overall application?**
- **How does the effort compare to the value you would derive by adding services and products to the organization?**



Once you have examined your current approach based on the criteria above, if you can't discern the benefits—or they are difficult to quantify—that is a clear indicator that you are not getting the value that you require. In this case, you may want to look into alternative approaches.

With Dataflow-as-a-Service, companies can realize development and deployment savings by eliminating the need to build, hire, and experiment on their own. They can focus on being first to market to capitalize on hypergrowth opportunities—with the time they save by not reading vast quantities of AI and ML research papers.

Privacy and Control

Dataflow-as-a-Service models are thoughtful to the concerns of those considering the do-it-yourself approach when it comes to data privacy and control. “The other consideration is if you want to be making your API calls to a hosted or public file environment,” says Choy. “What we’re seeing with regulated customers with sensitive datasets is that they want to make API calls to a system behind their firewall.”

Dataflow-as-a-Service offerings stand ready to help. Functioning like a cloud or CRM subscription, they can calm compliance-team nerves by making API calls to machines that sit behind firewalls and belong to the company instead of other vendors.

How to Make AI Adoption Easier

We all want to make AI adoption easier. Dataflow-as-a-Service unloads the hardest of AI design, development, and deployment to outsourced experts. This can yield significant benefits to your company and its bottom line, such as:

- Faster time-to-value
- New revenue streams
- Speed and efficiency over a DIY approach
- Professional-grade AI offerings

AI adoption via the Dataflow-as-a-Service approach can put a system behind your firewall, thereby making any compliance concerns about AI easier to resolve.

Moving much of the heavy lifting to the vendor, an API-driven approach effectively delegates away the selecting, training, fine-tuning, and maintaining of the models, freeing up much of your time and effort for other internal initiatives.

The Dataflow-as-a-Service approach provides you with the expertise you need to augment your ML capabilities without facing a talent-scarce labor force alone. That is a value-add that the vendors who offer component and DIY approaches struggle to provide.

Conclusion

On the road to AI adoption, enterprises face an important choice. They can walk the road to digital transformation alone, following the do-it-yourself approach. Alternatively, they can choose the new platform approach and partner with an AI industry expert in getting their solution developed, deployed, and online faster.

Each approach comes with important decisions that organizations must consider in designing their own AI strategies. Perhaps the biggest value of the Dataflow-as-a-Service model comes in its ability to generate time savings, new revenue streams, and speed to market in developing new products and services. All of these advantages go a long way toward creating competitive advantage in a market rife with new entrants and emerging technologies.

Partnering with the right vendor who walks your road to AI with you can augment your ML skills. It also can free you from many of the challenges that competitors face in a tight labor market.

“With an API-driven Dataflow-as-a-Service offering, this is for organizations that don’t have 300 or 3,000 data scientists and ML experts that the big hyperscalers have,” says Choy. “When you have three people and your plan is to go to five in the next 24 months, that’s who this model is for. It augments your existing ML capability and enables you to have near-hyperscaler-type AI expertise and capabilities rapidly.”

SambaNova’s goal is to support customers regardless of their preference for different types of deployment and procurement models. SambaNova can help with your ML projects and implement them in a fraction of the time required for a DIY approach. SambaNova does the work, sparing you the need to acquire additional ML experts. They’ve been thoughtful about implementing these advanced ML capabilities built into the platform itself. This enables the customer to think about their business and how they build upon the platform rather than building the platform itself.



Marshall Choy

VP, Product at SambaNova Systems

Marshall has overseen the go-to-market of dozens of industry-leading enterprise hardware and software products. Previously, he was Vice President of Systems Product Management and Solutions Development at Oracle.

About SambaNova Systems



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™, along with open standards and user interfaces. 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.



Contact us at:
info@sambanova.ai



Visit us at:
www.sambanova.ai



Follow us on LinkedIn
[SambaNova](#)

About Emerj Artificial Intelligence Research



Emerj Artificial Intelligence Research is a market research and advisory company focused exclusively on the business impact of AI.

Companies that thrive in AI disruption run on more than just ideas. They leverage data and research on the AI applications delivering return in their industry today and the AI capabilities that unlock true competitive advantage into the future - and that's the focus of Emerj's research services.

Leaders in finance, government, and global industries trust Emerj to cut through the artificial intelligence hype, leverage proven best-practices, and make data-backed decisions about mission-critical priorities.

Visit: www.emerj.com

Contact: research@emerj.com