THE RACE TO AI/ML VALUE

SCALING AI/ML AHEAD OF YOUR COMPETITION

SambaNova®
SYSTEMS

TABLE OF CONTENTS

INTRODUCTION

The AI/ML Opportunity

We’re on the cusp of an AI/ML-driven revolution in business. IDC forecasts that the global AI/ML market will grow more than 18% year over year in 2022. Many experts believe that AI/ML will be the next disruptive technology to refactor the Fortune 500, just as the internet has done over the past several decades.

To come out ahead in this revolution, companies need to use AI/ML for more than just gaining efficiencies or streamlining operations. Organizations have to drive innovation, changing how people live and how business is done. That means understanding the AI use cases that will drive business outcomes, then developing the capability to deploy AI/ML at scale across their organizations. The companies that rise to this challenge will be the winners in a new AI/ML-driven economy.

To take the pulse of where companies are in their AI/ML maturity, SambaNova surveyed 600 AI/ML, data, research, customer experience and cloud infrastructure leaders at the director level and above. Almost all respondents (94%) utilized high-performance computing infrastructure and applications.

We found that while these technical leaders are optimistic about AI/ML’s potential, they face pain points as they attempt to scale, including a lack of trained talent and the limits of computing architecture itself. These issues will become more pressing as organizations invest in compute-intensive deep learning applications and use AI/ML to drive more of their business operations. Overcoming barriers to scale will be key to unlocking AI/ML’s true promise to revolutionize business.

The global AI/ML market will grow
more than 18%
year over year in 2022.

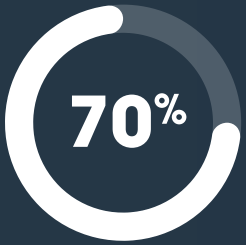
94%

Almost all respondents utilized high-performance computing infrastructure and applications.

INSIGHT 1

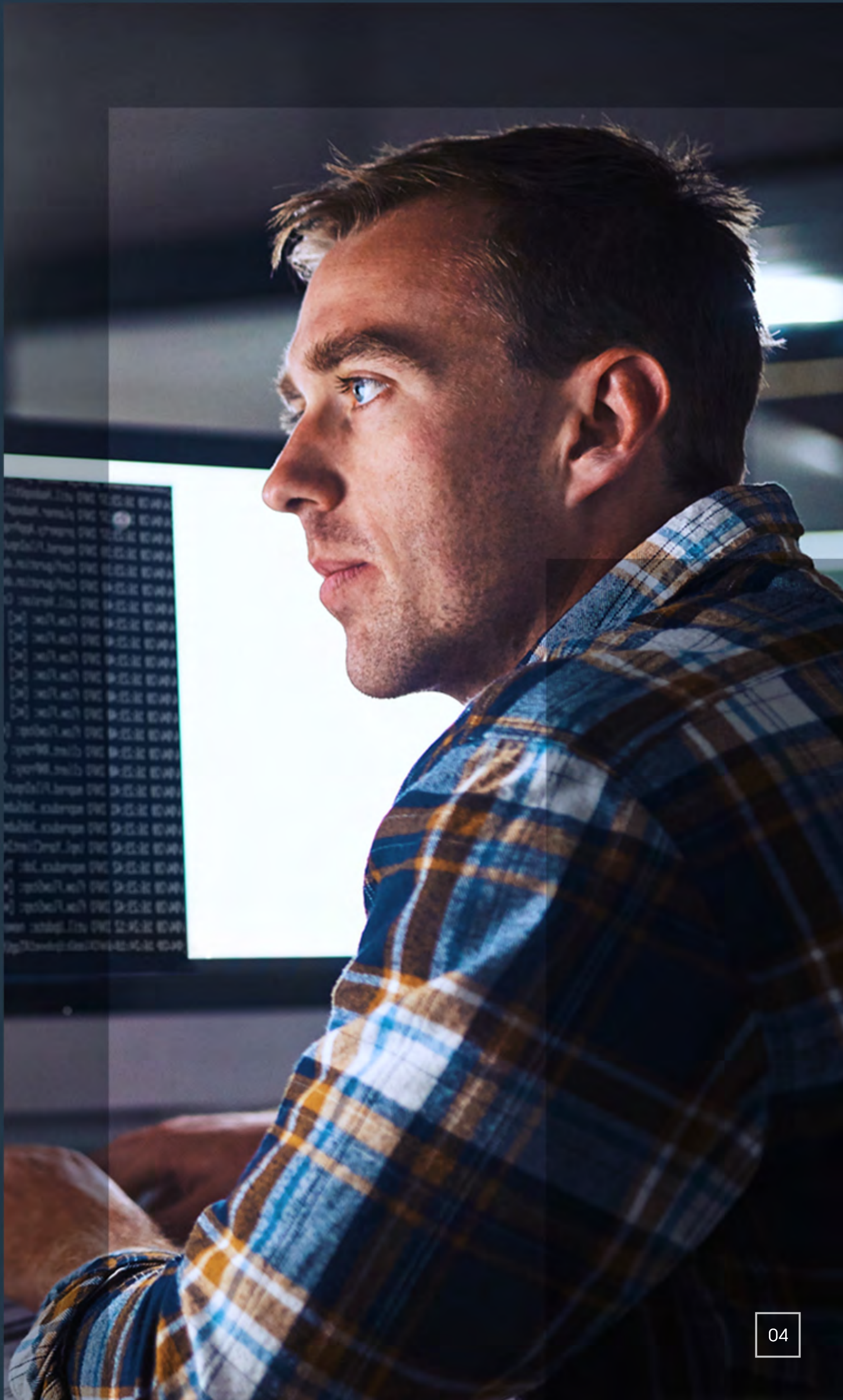
Organizations Have High Aspirations for Their AI/ML Initiatives.

Most respondents reported that their organizations are scaling up their investments in strategic technology, especially AI/ML. Over two-thirds (70%) of respondents said their organizations plan to allocate more than $100 million of IT budget toward strategic technology goals, and almost one-third (32%) said more than 20% of their IT budget is dedicated to AI/ML. Two-thirds of respondents said their organization plans to significantly increase their investment in AI/ML in the next five years.



Over two-thirds (70%)

of respondents said their organizations plan to allocate more than $100 million of IT budget toward strategic technology goals.

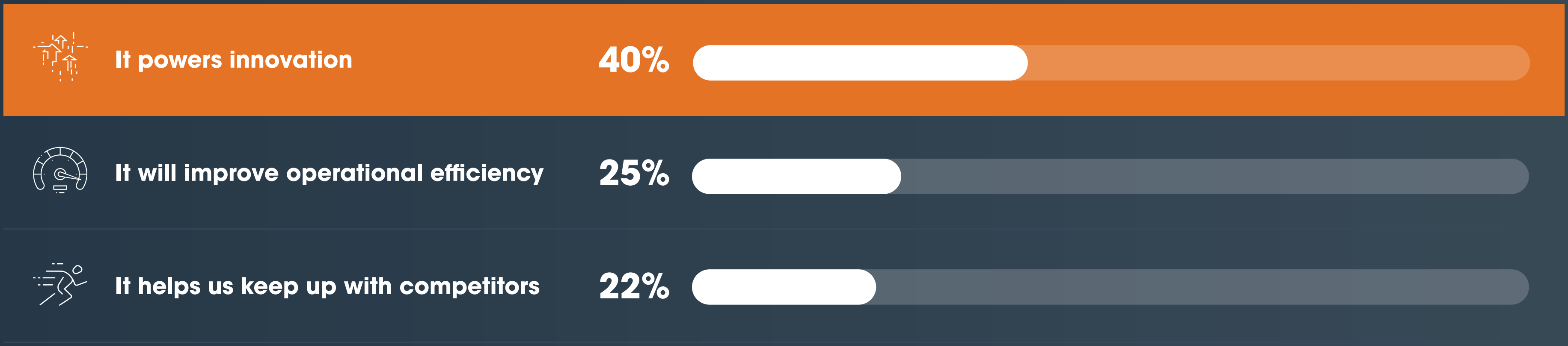


Organizations are looking to do more with their AI/ML investments than simply automate some tasks to gain efficiency. They know they need to innovate — creating new products and services and new lines of business — in order to stay competitive in a rapidly evolving market. Unsurprisingly, powering innovation is the top reason respondents say their companies invest in AI/ML, by a large margin (see Chart 1).

Most respondents understand that ultimately, AI/ML needs to drive revenue and business goals, not just cut costs. **More than three-quarters (78%) say that AI/ML is very important for driving revenue at their organization,** while more than two-thirds (68%) said their organization's AI/ML strategy was very aligned with business goals.

CHART 1: TOP 3 REASONS TO INVEST IN AI/ML

Which of the following best describes the main reason why your organization is investing in AI/ML?



However, it's clear that these statements are somewhat aspirational. Just because AI/ML is important to driving revenue in theory doesn't mean it is driving revenue in reality. Organizations are still in the early stages of AI/ML adoption at an enterprise scale, and they face multiple barriers to effective implementation including a skills gap and insufficient infrastructure (see Insight 3).

Cost savings is the top KPI used to measure AI/ML initiatives' success (see Chart 2). This shows that most organizations are still applying AI/ML more toward increasing efficiency rather than innovation, which will drive additional revenue streams.

CHART 2: TOP 5 KPIS FOR MEASURING SUCCESS OF AI/ML INITIATIVES

Which KPIs does your organization use to measure the success of your AI/ML initiatives?

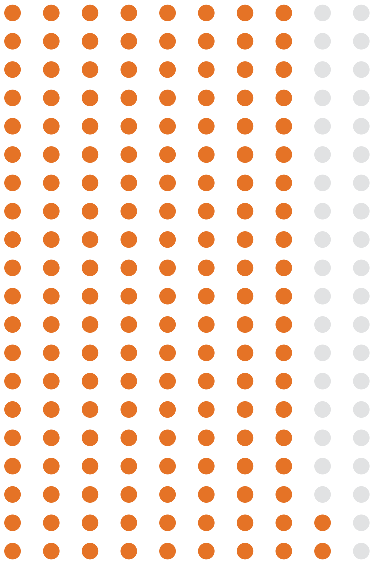
Cost savings	**72%**
Revenue growth	**67%**
Time savings	**60%**
New product development	**56%**
Time to insight	**52%**

INDUSTRY CALLOUT

The Financial Industry Is Investing Heavily in AI/ML

Almost one-third (31%)

of financial services respondents say their organization spends more than a quarter of its IT budget on AI/ML.



81%

plan to significantly increase their investments in AI/ML, the highest percentage of any industry.

INSIGHT 2

Organizations Are Looking to Innovate With Deep Learning.

When it comes to driving innovation, one subfield of AI/ML is particularly important: **deep learning**. Deep learning uses artificial neural networks to ingest and process unstructured data like text and images. Common use cases for this powerful technology include:

Natural Language Processing (NLP)

Computer Vision

Recommendation Algorithms

Natural language processing (NLP) enables a machine to understand spoken and written language, and also produce language of its own. Common use cases include automated customer service fulfillment with intelligent discussions, as well as fraud detection. By watching daily transactions and customer accounts, an NLP-enabled solution can generate alerts for any suspicious customers or transactions identified on an account.

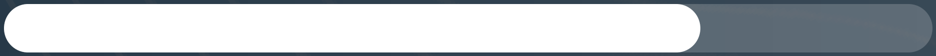
Computer vision enables a machine to classify and process digital images, and is used in everything from image search functions to the navigation systems of self-driving cars. In factories and production facilities, this technology will eventually make it unnecessary to manually examine the quality and integrity of packaging, eliminating hundreds of hours of mundane and repetitive human labor. In the medical field, computer vision models will detect lesions, pneumonia, aneurysms and other issues on radiology scans such as x-rays, MRIs and CT scans, as well as performing other visual medical functions.

Recommendation algorithms learn a user's preferences over time and use that insight to surface content tailored to their needs, whether that's a new show on a streaming service or a recommended next step on a project. Use innovations for this model include high-quality content recommendation on news feeds as well as tailored product recommendations customized to each individual.

Three-quarters of respondents (75%) say improving access to deep learning is very important for fostering competition and innovation in their industry. Since driving innovation is the No. 1 reason organizations invest in AI/ML, it's unsurprising that many organizations are exploring applications of deep learning (see Chart 3).

75%

of respondents say improving access to deep learning is very important for fostering competition and innovation in their industry.



However, just as with respondents' statements about AI/ML driving revenue, their statements about the importance of deep learning to their organizations are likely somewhat aspirational. Real-life deep learning workloads are very compute- and data-heavy, and can present a challenge for all but the most advanced computing infrastructure. And as we'll see in the next section, **insufficient infrastructure is already holding the scale of AI/ML initiatives back.**

There is also the education gap. Successful AI/ML initiatives are driven by the desire to achieve defined business outcomes, which requires a clear understanding of specific use cases. Relatively few business leaders understand what deep learning is or grasp its transformative potential for business, and this lack of awareness can prevent deep learning initiatives from receiving the support they need. With improved education on the business side, organizations can unleash the potential of AI/ML to unlock innovation, achieve scale and drive business outcomes across the enterprise.

CHART 3: TOP 3 USES FOR DEEP LEARNING

Which of the following use cases does your organization use deep learning for?

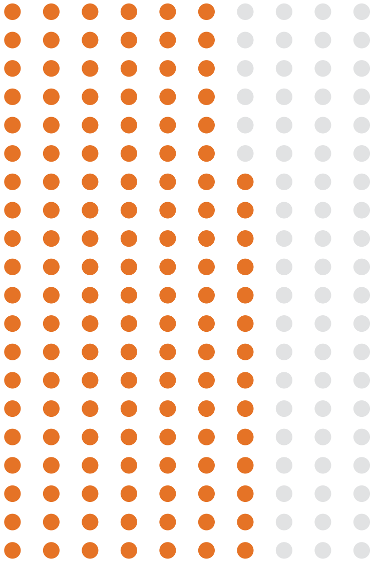
Natural Language Processing (NLP)	**81%**
Computer Vision	**68%**
Recommendation	**55%**

INDUSTRY CALLOUT

Top Uses of Deep Learning Vary

NLP is the most common use case in every industry except one:

In the public sector, computer vision is the top use case with 71%.



67%

of financial services respondents use deep learning for recommendations — significantly more than other industries.

INSIGHT 3

Organizations Face Multiple Barriers to AI/ML Scale.

For AI/ML initiatives to drive innovation and boost revenue, **they need to scale dramatically.** They face a number of key challenges in doing so.

1. **The difficulty of customizing models**

2. **Insufficient infrastructure**

3. **The AI/ML skills gap**

1. The difficulty of customizing models

To drive business outcomes and gain a significant competitive advantage with AI/ML, you must tailor models to use cases specific to your business. Training proprietary AI/ML models requires a significant investment of time and expertise — expertise that many organizations struggle to recruit (see No. 3). No wonder the difficulty of customizing models is the top challenge in scaling AI/ML initiatives, according to respondents (see Chart 4).

Organizations can't just set it and forget it, either. Data is ever-changing, and models should be too. They must be updated to avoid issues like concept drift that can reduce their accuracy over time, adding to the challenge of developing effective AI/ML.

2. Insufficient infrastructure

The compute-heavy workloads associated with AI/ML, particularly deep learning, pose a challenge to organizations looking to scale up. With the end of Moore's Law, we can no longer expect the number of transistors on a chip to double every two years, as has been the case for the past five decades. A majority of respondents (53%) strongly agree that they'll run out of computing power in the next decade without new computing architecture.

And without access to more efficient infrastructure — such as new, AI/ML-specific chip architectures — organizations are quickly running out of sufficient space for their compute needs. Almost two-thirds of respondents (65%) already struggle with limited space for server racks to a significant extent. As AI/ML workloads increase, that struggle will only worsen.

53%

of respondents strongly agree that they'll run out of computing power in the next decade without new computing architecture.

42%

of respondents say they either lack enough AI/ML engineers on staff, lack an adequate pool of potential applicants, or both.

3. The AI/ML skills gap

Deploying AI/ML workloads efficiently and at scale requires expertise in multiple computing architectures as well as a deep understanding of training and tuning models. The rarity of these skills makes AI/ML talent acquisition difficult and expensive. Almost half of respondents (42%) say they either lack enough AI/ML engineers on staff, lack an adequate pool of potential applicants, or both.

This talent shortage exacerbates the barriers to scale described in the first two challenges. A lack of access to trained talent is the No. 3 challenge in scaling AI/ML (see Chart 4). It may not be possible to solve the skills gap. If organizations want to harness the full transformative power of AI/ML, they need to find a way around it.

CHART 4: TOP CHALLENGES WITH SCALING AI/ML

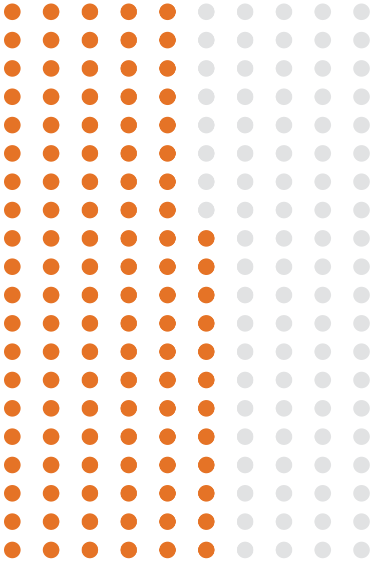
What are your organization's biggest challenges in scaling your AI/ML efforts?

Difficulty of customizing models to our unique needs	**50%**
Complexity of working around restrictive computing architectures such as CPUs or GPUs	**35%**
Not enough compute to analyze the amount of big data	**28%**
Lack of access to trained talent	**28%**
Lack of buy-in/trust from company leadership	**25%**
Cost of powering additional servers	**25%**
Limited space for servers	**22%**

INDUSTRY CALLOUT

Retail and the Public Sector Suffer a Worse AI/ML Skills Gap

While the AI/ML talent shortage is painful everywhere, some industries have been hit harder than others.



56%

More than half of retail & ecommerce respondents (56%) complained about a talent shortage, as did half of public sector respondents.

CONCLUSION

Solving the AI/ML Scale Dilemma

As the AI/ML revolution pushes forward, many organizations face a dilemma. On the one hand, they understand the growing need to innovate, increase revenue and drive operational efficiency with AI/ML. On the other hand, they face limitations on scale, including a skills gap and insufficient infrastructure to handle increasingly compute-intensive AI/ML

However, two key trends offer a path forward:

1. Customization at scale through a partner:
As the AI/ML talent shortage continues, more companies are turning to outside partners to train AI/ML models for their businesses. Unlike many off-the-shelf AI/ML products, which have limited flexibility, a good AI/ML partner will customize models to your organization's unique needs and deploy them in the location of your choice. By outsourcing the technical details of AI/ML training, tuning, infrastructure and maintenance, you'll free up your time to focus on innovating with the resulting insights.

2. More efficient chip architectures:
Despite the end of Moore's Law, not all chips are created equal where performance on AI/ML workloads is concerned. You can extract more efficiency and even reduced power consumption from a chip architecture that's tailored to AI/ML, similar to how the architecture of GPUs is tailored to graphics processing workloads. If you plan to work with a partner, look for one that takes advantage of innovative chip architectures to boost efficiency — they'll be able to scale up with you as your AI/ML compute needs grow.

If you want your company to harness the full power of AI/ML to drive innovation and revenue, the time to solve these scale issues is now. By working with knowledgeable AI/ML partners and taking advantage of more efficient AI/ML-specific infrastructure, you'll set yourself up to stay ahead of disruption.

Methodology

In August 2021, SambaNova surveyed 600 full-time AI/ML, data, research, experience and cloud infrastructure leaders at the director level and above. The survey captured 100 responses from each of six key industries: financial services, healthcare and life sciences, manufacturing and auto, retail and e-commerce, public sector and oil and gas.

SambaNova®
SYSTEMS