Synthesis.ai

ADAPT OR BE LEFT BEHIND:

89 Percent of Tech Execs See Synthetic Data As a Key to Staying Ahead

SYNTHETIC DATA

A Cost and Resource Saving Tool for AI

Synthetic data has been used throughout the AI industry since its very inception, from synthetic line drawings for early vision systems, through synthetic video feeds for the first self-driving neural network ALVINN in the 1980s, to the pioneering works of Harvard statistics professor Donald B. Rubin in the early 1990s about the statistical properties of synthetic datasets.

According to Gartner, in the coming years, the data used to create AI models will be primarily synthetic and "generated by rules, statistical models, simulation, and other techniques" and will usurp the use of real data obtained from direct measurements. In other words, companies that do not embrace and integrate emerging synthetic data technologies will be left behind. It's possible you simply won't be able to build a competitive AI model without synthetic data.

AI is driven by the speed, diversity, and quality of data. Today's systems leverage 'supervised learning' approaches in which humans label attributes in image data to then train AI models. This approach is fundamentally limited as humans do not scale and, more importantly, cannot label key attributes (e.g., 3D position, interactions, etc.) necessary to enable emerging industries such as AR/VR, autonomous vehicles, robotics, and more.

Synthetic data is emerging to be an essential element in building accurate and capable AI models, as it provides developers with vast amounts of perfectly labeled data on-demand. A single dataset may contain tens of millions of elements. To manually collect and label data of this magnitude is time-consuming and costly for organizations, not to mention prone to human errors. Synthetic data aims to simulate real-world scenarios to train AI systems virtually. This approach reduces the time and resources needed to build these models by delivering vast amounts of perfectly labeled data to organizations in a matter of hours. Also, since synthetic data isn't generated from real-world sources, privacy and bias issues are reduced.

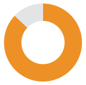
Synthetic data has the potential to reshape and transform several industries. From improving the safety of vehicles to helping workers have more productive video calls, the use-cases of synthetic data are broad.

As the world begins to understand synthetic data and its applications, how are companies viewing this emerging technology? Do they understand its potential? What barriers do early adopters see on the horizon?

This report, conducted by Synthesis AI in conjunction with Vanson Bourne, presents findings and takeaways from a survey of 100 senior technology executives on their perceptions of synthetic data, potential benefits and barriers of implementation, and what industry leaders think it will take to continue driving the adoption of synthetic data.

Gaps in Knowledge in Tech Company Leaders

Synthetic data adoption is increasing, but its usage and understanding of the technology vary across the board.



87% of organizations use techniques to enhance their data, including image augmentation, bootstrapping, generative models, etc.

just **43%** use it whenever possible

and **44%** have only just started using the technology

Resource Savings with Synthetic Data

Synthetic data could be a solution to the time-consuming and cost-prohibitive nature of supervised learning.

ON AVERAGE, data labeling costs organizations

$2.3 MILLION ANNUALLY

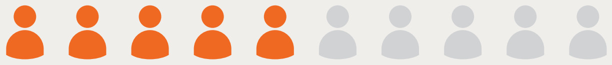
16 WEEKS

▸ THE AVERAGE LENGTH OF TIME SPENT TO CONDUCT SUPERVISED LEARNING ON A NEW PROJECT.

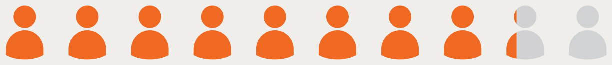
More Knowledge Equals More Execution and Confidence

Despite recognizing the importance of data enhancement, only half (51%) of the respondents aligned with the explicit technical definition of state-of-the-art synthetic data approaches indicating a critical knowledge gap.

Of those who selected the correct definition:



50% believe a critical benefit of synthetic data is overcoming limited labels provided through supervised learning/human annotation.



82% recognize their organization is at risk when they collect "real-world" data.

Respondents who were knowledgeable of state-of-the-art synthetic data technologies expressed confidence in the technology's ability to address critical issues utilizing "real-world" data. Reducing the knowledge gap in the enterprise will lead to a better understanding of synthetic data benefits.

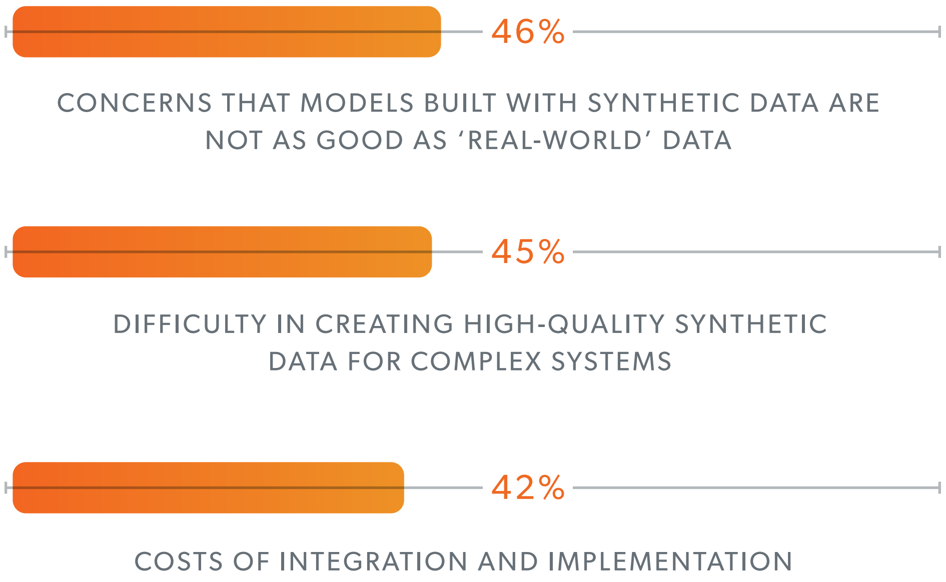
Barriers to Overcome in Further Adoption

While most respondents appear to understand synthetic data, our research found that may not be true with their colleagues. Prominent barriers to entry when using synthetic data include a lack of organizational knowledge and slow buy-in from colleagues. Buy-in from colleagues and decision-makers will be critical for synthetic data to be accepted.

67% agree that their organization lacks the knowledge and understanding when it comes to implementing synthetic data.

67% agree that users in their industry will not accept synthetic data until they see the benefits for themselves.

Respondents reported the following as being the most challenging aspects of utilizing synthetic data within their organization:



A Bright Future for Synthetic Data

More than half (59%) of decision-makers believe that their industry will utilize synthetic data either independently or in combination with ‘real-world’ data within the next five years. This suggests that many organizations are only just starting to experiment with it. Synthetic data will be critical to the future of many industries and organizations and will lead to widespread change, especially among those who use vision data.

89% ***AGREE***

THAT SYNTHETIC DATA IS A

new and innovative technology,

THAT WILL TRANSFORM THEIR INDUSTRY.

Those who don’t employ synthetic data are at risk of falling behind the curve. Nearly nine in ten (89%) of those who use vision data believe organizations that fail to adopt synthetic data in training their internal systems will lag behind.

OF THE COMPANIES THAT USE VISION DATA,

95% ***BELIEVE***

SYNTHETIC DATA IS A NEW AND INNOVATIVE TECHNOLOGY.

The growth potential is evident, and those who work with vision data are best placed to take advantage. Among those working with vision data that don’t use or have only started using synthetic data, only three in ten (30%) respondents cite a lack of tools to create and manage synthetic data as a barrier to broader utilization.

SYNTHETIC DATA

The Fast Track for Safe, Unbiased, Reliable AI Deployment

Synthetic data is just beginning its cycle of adoption and value to the enterprise. Many industries and companies are only just beginning to experiment with the technology. Still, synthetic data shows promise to cut down on the cost, improve access, and reduce the time it takes to build AI models in traditional ways.

That doesn't mean there aren't barriers to broader adoption. A key to further implementation is educating colleagues throughout the entire organization, not just the C-suite, as there is confusion and a lack of understanding among many groups. Organizations already using vision data are positioned to lead this charge, as they understand the value of vision data and how it can benefit their industry. According to Synthetic Data for Deep Learning, new research is starting to provide proof points around the utility of synthetic data across use-cases, including robotics, autonomous vehicles, smart homes, consumer products, manufacturing, logistics, healthcare, and more.

When further education and adoption are achieved, the benefits of implementing synthetic data technology are abundant. Synthetic data can help improve access to high-quality data. Synthetic data can also enable more capable models through new and more accurate data labels.

Synthetic data can also help reduce bias in AI models. Bias is often a result of unbalanced training data that does not properly represent the real-world data distribution. For instance, it is vital to have a dataset that covers all gender and identities in face recognition. By supplementing training data with synthetic data, data distributions will better reflect key demographics resulting in more balanced and fair AI systems. By lowering the barrier and cost, difficult-to-obtain datasets become more available, opening doors for enterprises of all sizes to build state-of-the-art models. Millions of dollars and months of work could be saved, paving the way to create more models in a fraction of the time with fewer resources.

Most technology industry leaders agree that synthetic data will be an essential enabling technology and key to staying ahead.

Synthesis.ai