

STATE OF AI **2022**

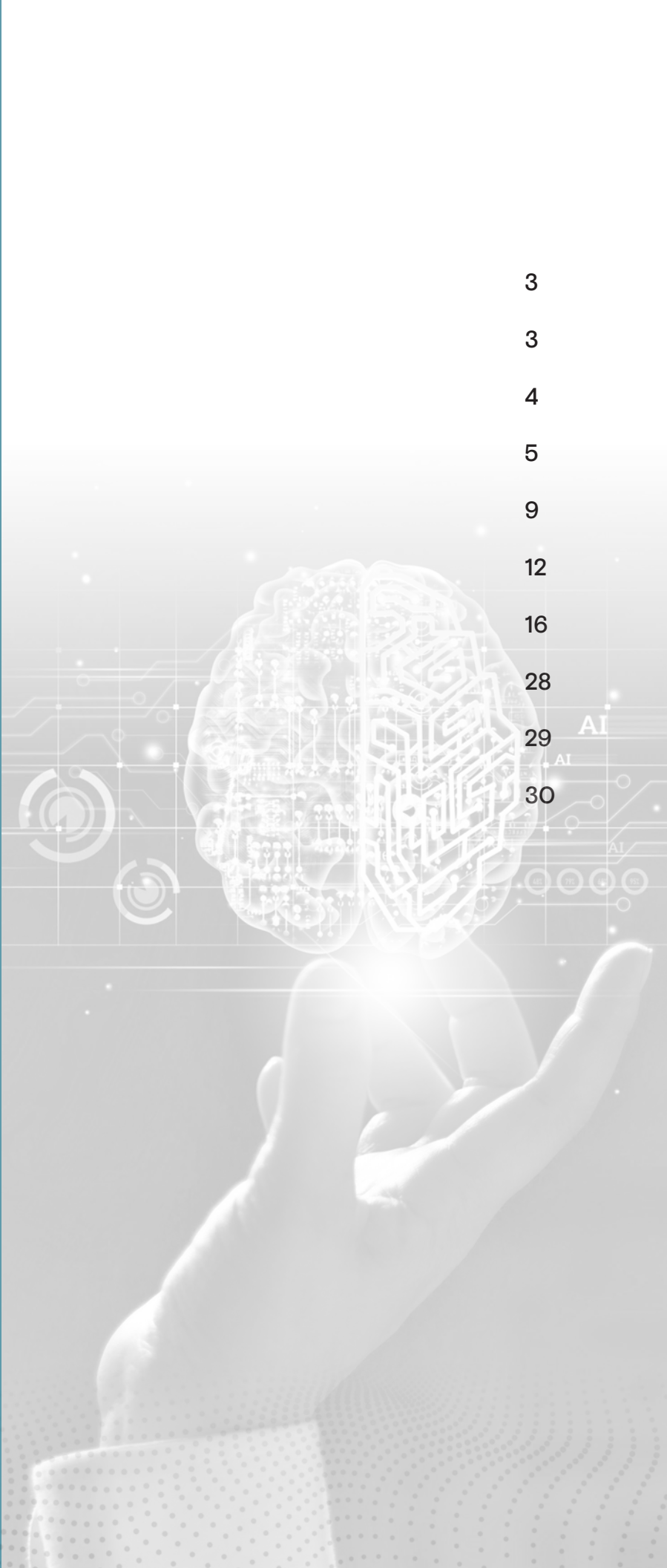
and Machine Learning Report

Managing Data for the AI Lifecycle



Table of Contents

Introduction	3
Executive Forward	3
Key Takeaways	4
Sourcing	5
Quality	9
Evaluation	12
Adoption	16
Conclusion	28
Methodology	29
About Appen	30



Introduction

The State of AI and Machine Learning is a cross-industry effort aimed at providing an overview of the status of artificial intelligence and machine learning through input from businesses and their senior decision makers and technical practitioners. For our 8th annual edition, Appen partnered with The Harris Poll to expand our survey to 504 participants in North America and Europe. The goal of this report is to help us understand AI adoption, the maturity in data management across the AI lifecycle, and the value placed on responsible AI. These responses enable us to understand how artificial intelligence is changing and adapting to the needs of a post-COVID world.

The pandemic brought accelerated growth to the AI industry, and we continue to experience an evolution. The demand for data that reflects the “new normal” caused a spike in the demand for quality data, fast. Human behavior changed, and therefore, machines had to learn to mirror this new normal in order to stay relevant. Now that the world is slowly readjusting post-pandemic, this urgent need is leveling off, but demand remains at an incline over 2020 and pre-pandemic trends. This opens an ocean of opportunity, for those who see it, to develop new innovations in AI and ML.

There’s a higher value placed on data for the AI lifecycle, with leaders and practitioners showing an understanding of the four key stages: Data Sourcing, Data Preparation, Model Training and Deployment, and Model Evaluation. As we establish greater market maturity, we’re not surprised to learn that organizations are seeking a partner in launching and maintaining AI initiatives. Data sourcing and data preparation can be a daunting task and businesses learned quickly there are efficient and effective resources to alleviate their data scientists of that burden.

Responsible AI is an important part of that maturity and a majority report it’s the foundation of all AI and machine learning projects. While decision-makers are fairly aligned, business leaders are more likely than technologists to site ethics as a key factor when thinking about responsible AI.

The results of our survey are indicative of changing times, following the initial surge in AI implementation to adapt to a global pandemic that affected all aspects of business and life. For more details on the methodology of our survey, see page 29.



Building on our 26 years of expertise, Appen is proud to partner with The Harris Poll to deliver the current state of the AI industry.

As the leader of data for the AI lifecycle, we are pleased to learn decision makers in technology, finance, healthcare, and retail understand the value of managing data across every stage. The findings and input from these technologists and business leaders provide guidance on challenges, needs, and where the industry is going. We are encouraged to see data scientists are spending less time sourcing and preparing data and that companies are more confident in their advancement on AI initiatives. It is my prediction that within 10 years, every single business application built will leverage AI in order to stay competitive.

Sujatha Sagiraju
Chief Product Officer, Appen

Key Takeaways

1

SOURCING

Considered a challenging step of the AI lifecycle, data sourcing remains an obstacle.

42% of technologists say the data sourcing stage of the AI lifecycle is very challenging. However, business leaders were less likely to report data sourcing as very challenging (24%).

2

QUALITY

Business leaders and technologists report a gap in the ideal vs. reality of data accuracy.

More than half of respondents say data accuracy is critical to the success of AI, but only 6% reported achieving data accuracy higher than 90%.

3

EVALUATION

AI will not be replacing humans any time soon.

There's a strong consensus around the importance of human-in-the-loop machine learning with 81% stating it's very or extremely important and 97% reporting human-in-the-loop evaluation is important for accurate model performance.

4

ADOPTION

Perceptions regarding the prominence of AI in business may be shifting.

Technologists are split on whether their organization is ahead or even with others in their industry. Respondents in the US are more likely than their European counterparts to say their organizations are ahead of others in their industry at adopting AI.

5

ETHICS

Responsible AI is the foundation of all AI projects.

93% of respondents agree that responsible AI is a foundation for all AI projects within their organization.

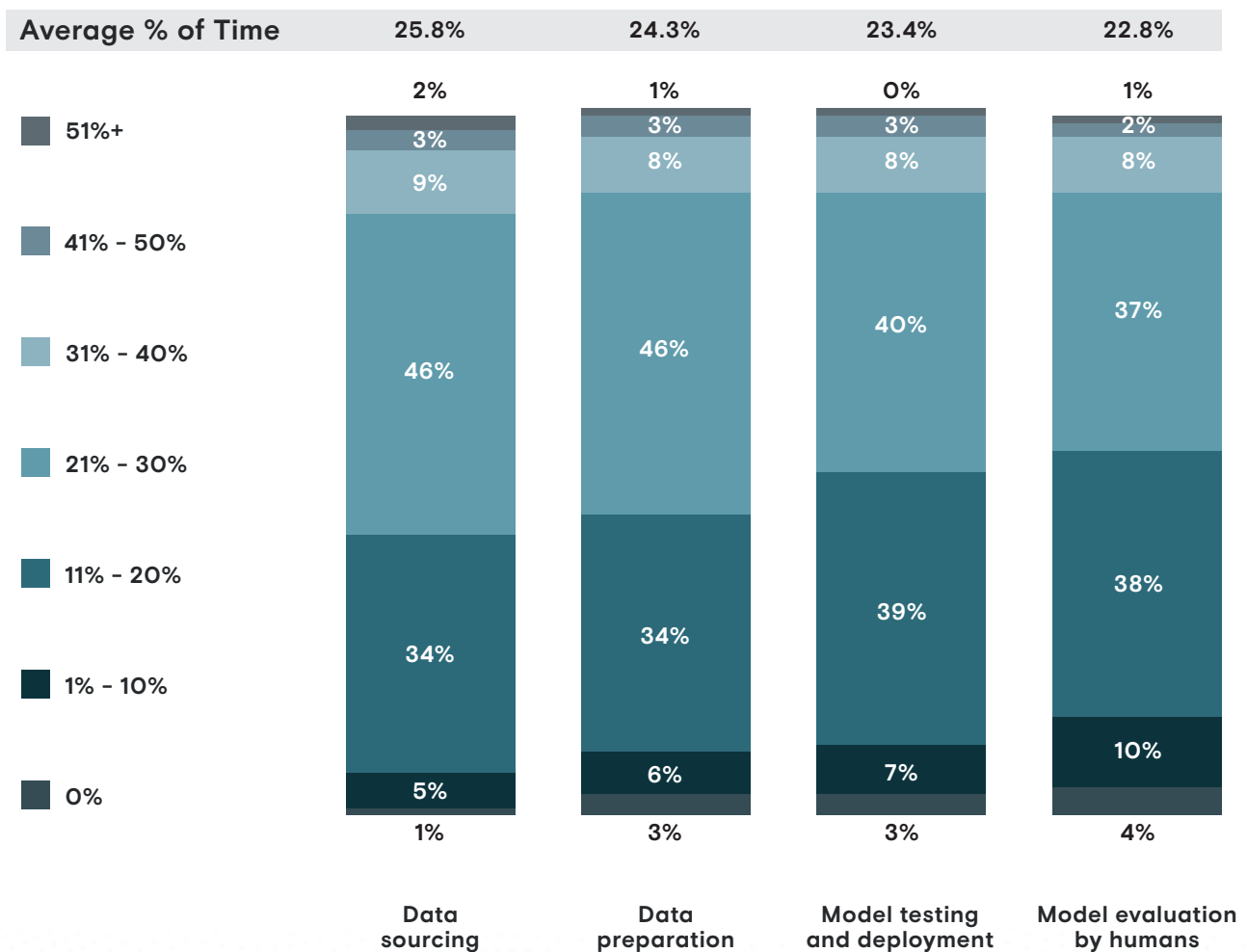
Sourcing

Data sourcing continues to be a major bottleneck for teams building artificial intelligence applications.

Many factors are in play like a lack of sufficient data for a specific use case, new machine learning techniques that require greater volumes of data, or teams don't have the right processes in place to easily and efficiently get the data they need.

Looking at data for the AI lifecycle, there's strong consensus that leadership understands the value of managing the full lifecycle (90% agree) and it's changing the way their company does business (87% agree). Decision makers split their time evenly managing data across the four stages of the AI lifecycle, and 7 in 10 (71%) say their organization struggles with the many phases of the AI lifecycle.

Percentage Of Time Allocated By Stage





“The gap between data scientists and business leaders is slowly narrowing year over year when it comes to understanding the challenges of AI. The emphasis on how important data, especially high-quality data that match with application scenarios, is to the success of an AI model has brought teams together to solve for these challenges.”

Mingkuan Liu
VP of Data Science - Appen

Though the majority surveyed (88%) feel their organization has the necessary internal resources in place to manage data across each stage, 42% of technologists find the data sourcing stage of the AI lifecycle very challenging. Business leaders, however, were less likely to report data sourcing as very challenging (24%). This shows there are still gaps between technologists and business leaders when understanding the greatest bottlenecks in implementing data for the AI lifecycle. This results in misalignment in priorities and budget within the organization.

Figure 1:
For each of the following stages, how challenging is it for you/your organization to complete the stage?

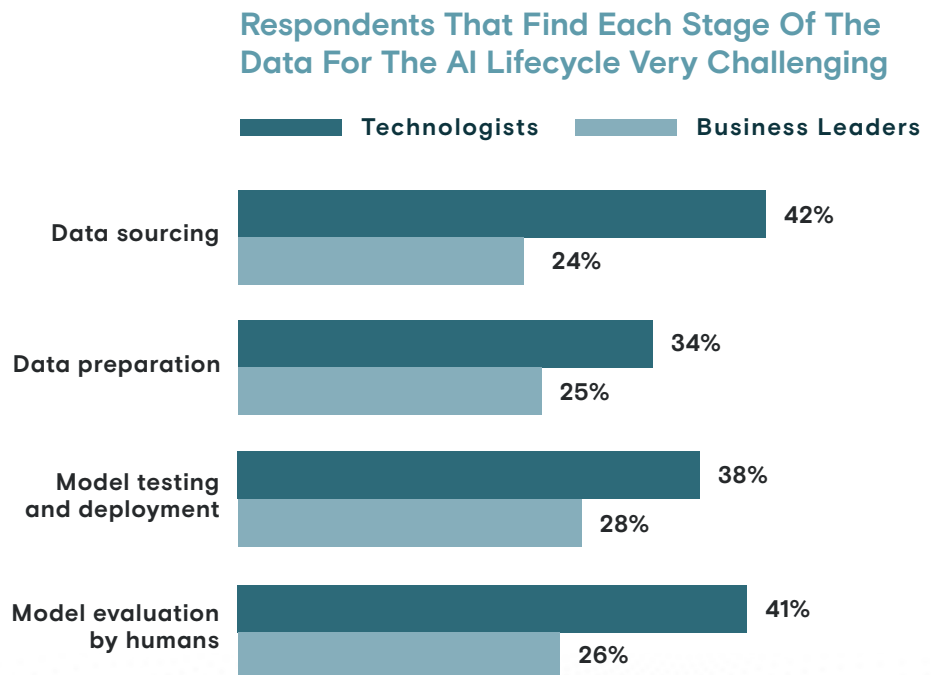
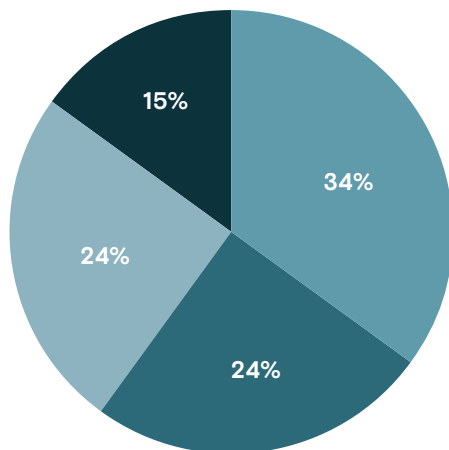


Chart reflects segment of responses identifying stages as very challenging.

Among those surveyed, there's an agreement that the data used for AI models should be sourced responsibly, contain less bias, and remain highly accurate.

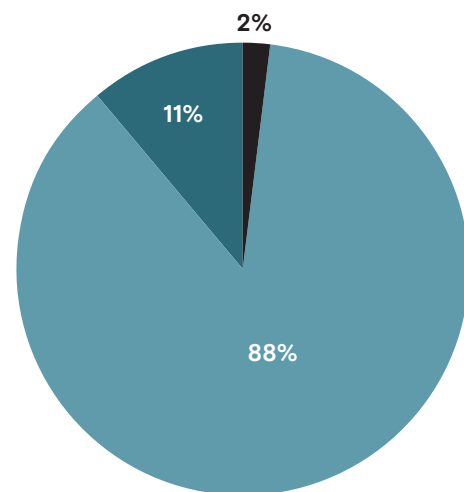
In addition to being identified as very challenging by many, the data sourcing stage also has the greatest budget allocation, with more than a third claiming it's the stage that requires the most budget.

Most Budget Allocation By Stage



- Data sourcing
- Data preparation
- Model testing and deployment
- Model evaluation by humans

Use External AI Training Data Providers



- Use external AI training data providers
- Do not use external AI training data providers
- Unsure



“With the rise of large language models (LLM) trained on multilingual web crawl data, companies are facing yet another challenge. These models oftentimes exhibit undesirable behavior due to the abundance of toxic language, as well as racial, gender, and religious biases in the training corpora. Such behaviors cannot be easily amended universally since acceptable language use varies substantially across context and cultures. While there are a number of potential ways to address this, including adjusting the way models are trained, filtering training data and model outputs, and learning from human feedback and testing, additional research is required to establish a trustworthy human-centric LLM benchmark and evaluation methodology.”

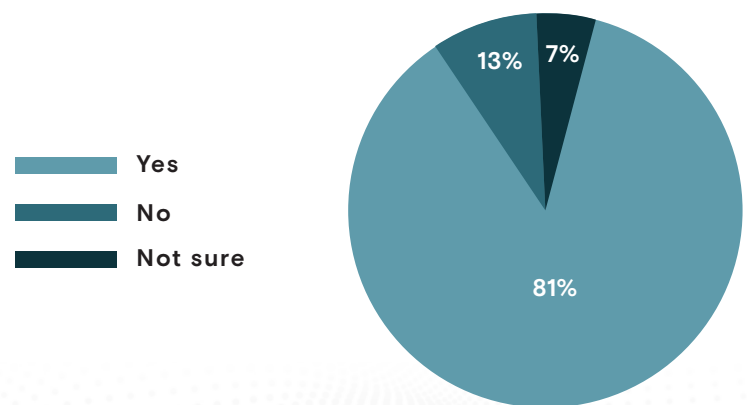
Iliia Shifrin

Sr. Director of AI Specialists - Appen

For AI solutions to properly function, massive volumes of quality data are required to train the underlying neural networks. A great example is multilingual natural language processing (NLP) that relies on millions of human speech inputs for each language prepared and delivered in formats ML models can ingest.

While 4 out of 5 of our survey respondents say they have the right amount of data to support an AI project (81%) and have access to the tools they need to do their AI-related job (90%), the majority of them are still struggling with low data quality. This typically results in underperforming systems. This becomes even a bigger challenge when integrating multimodality in NLP or connecting multiple individual NLP solutions that support several languages and content types.

Do you believe you have the right amount/enough data to support an AI initiative/project?



Quality

It's no secret that the world is changing—with more smart devices, the use of multiple screens, and new digital tools collecting information—

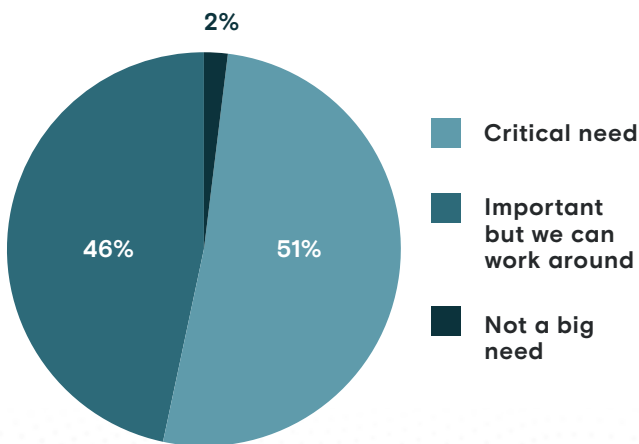
data volumes generated by the global digital footprint are so vast, that accurately structuring, annotating, and labeling data is more important than ever.

51% of our survey participants agree that data accuracy is critical to their AI use case and 46% agree it's important but can work around it. Only 20%, however, reported achieving higher than 80% data accuracy and only 6% reported higher than 90%.

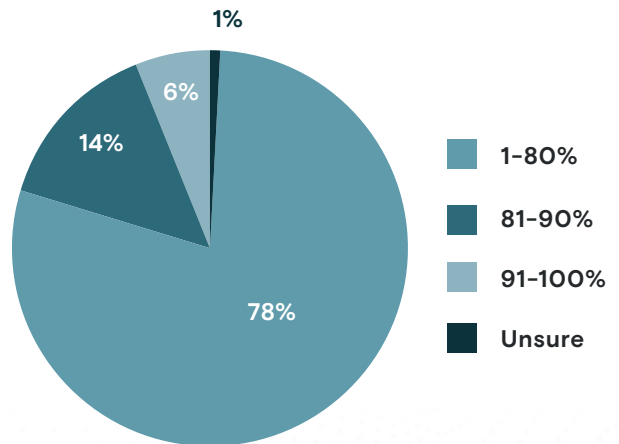


“Data accuracy is critical to the success of AI and ML models as qualitatively rich data yields better model outputs and consistent processing and decision-making. For good results, datasets must be accurate, comprehensive, and scalable.”
Wilson Pang CTO - Appen

How Critical Is Data Accuracy?



What Level Of Data Accuracy Do You Achieve?



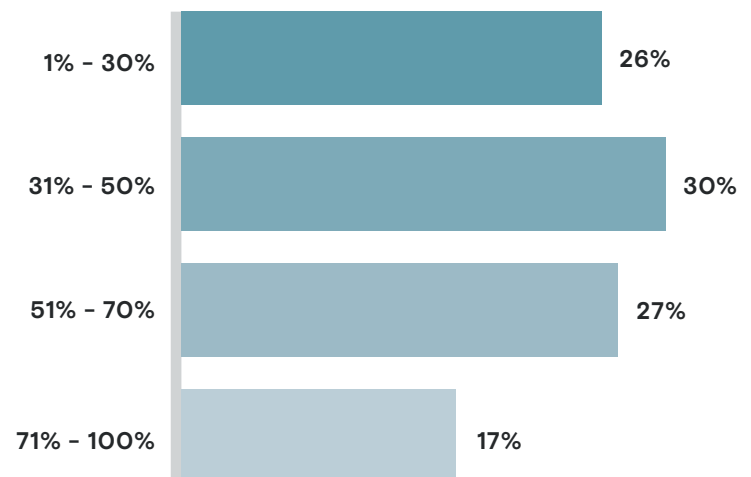
Graphs reflect rounded numbers

THE AVERAGE PROPORTION OF TIME SPENT MANAGING AND PREPARING DATA IS TRENDING DOWN, ON AVERAGE 47.4% OF TIME COMPARED TO 53% IN 2021.

Using the right data at the beginning of the lifecycle drives greater results through later stages. The average proportion of time spent managing and preparing data is trending down, on average 47.4% of time compared to 53% in 2021. With a large majority of respondents using external data providers, it can be inferred that by outsourcing data sourcing and preparation, data scientists are saving the time needed to properly manage, clean, and label their data.

On average, what percentage of your team's time is spent managing, cleaning and/or labeling data?

Percentage Of Time Spent On Data Management, Cleaning And Labeling



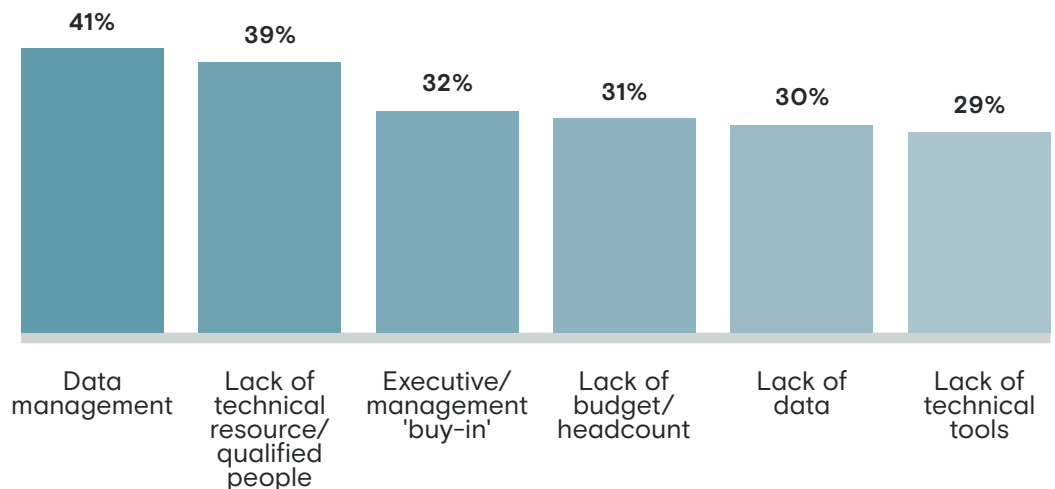
The greatest hurdle for AI initiatives is data management.



The greatest hurdle for AI initiatives is data management, with 41% indicating it as the biggest bottleneck. Right behind, 39% of respondents reported a lack of qualified talent--data scientists and technologists, data architects and engineers are scarce. 31% indicated a lack of budget for adequate headcount, adding to the challenge of properly staffing data management teams. This shortage of qualified data scientists and technologists emphasizes the importance of ensuring critical talent is focused on activities that require their valuable skills. To remedy this, companies look to external data providers to reduce their workload in areas such as data sourcing, freeing up scientists' time for other AI initiatives.

Biggest Bottlenecks For AI Initiatives

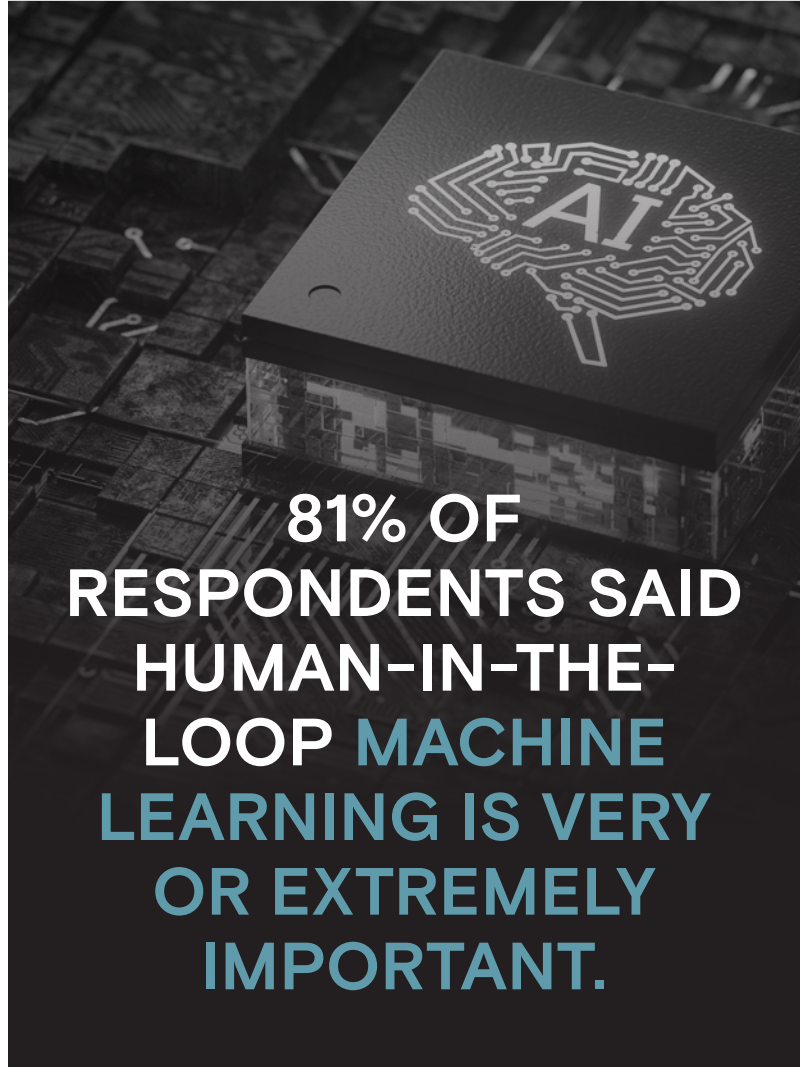
What do you consider the biggest bottleneck to any of your AI initiatives or projects?



Evaluation

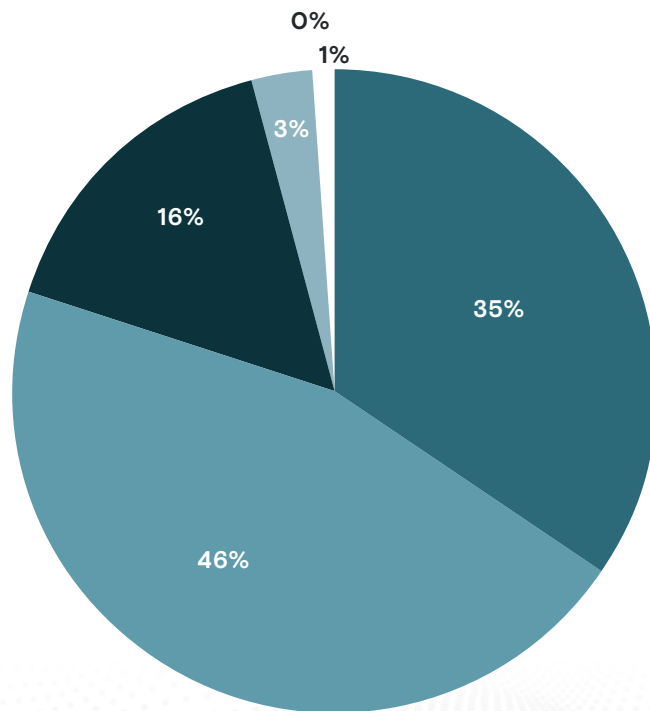
Machine learning models require continuous monitoring and adjustments to make sure they will deliver accurate, relevant information.

While the model is mostly autonomous once it's deployed, validation and re-training require human-in-the-loop operations. There's extremely strong consensus around the importance of human-in-the-loop machine learning with 81% of respondents stating it's very or extremely important and 97% reporting that human-in-the-loop evaluation is important for accurate model performance.



Importance Of Human-In-The-Loop Machine Learning

- Extremely Important
- Very Important
- Fairly Important
- Somewhat Important
- Not at all Important
- Unsure



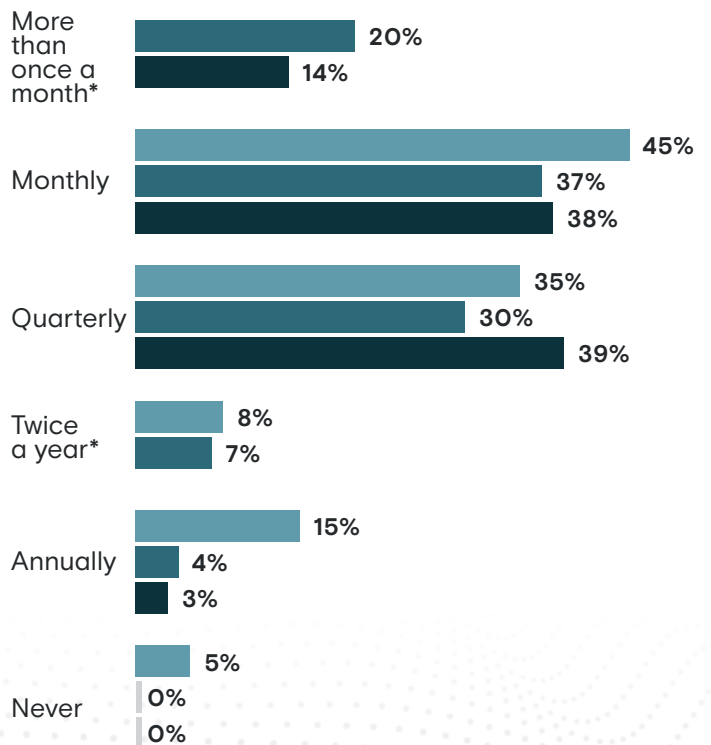
91% OF ORGANIZATIONS UPDATE THEIR MACHINE LEARNING MODELS AT LEAST QUARTERLY.

The AI Lifecycle is an ongoing process—with new data inputs and model outputs to be sourced, prepped, and evaluated constantly. This is reflected in the high number of organizations leveraging external data providers (88%), as well as in the data points we measured regarding the need to update models on an ongoing basis. Last year, 86% of organizations updated their models at least quarterly, with a directional increase to 91% this year.

How often are you retraining/ updating your machine learning models?

- 2020 (n=290)
- 2021 (n=501)
- 2022 (n=404)

*Response only applicable in 2021/2022

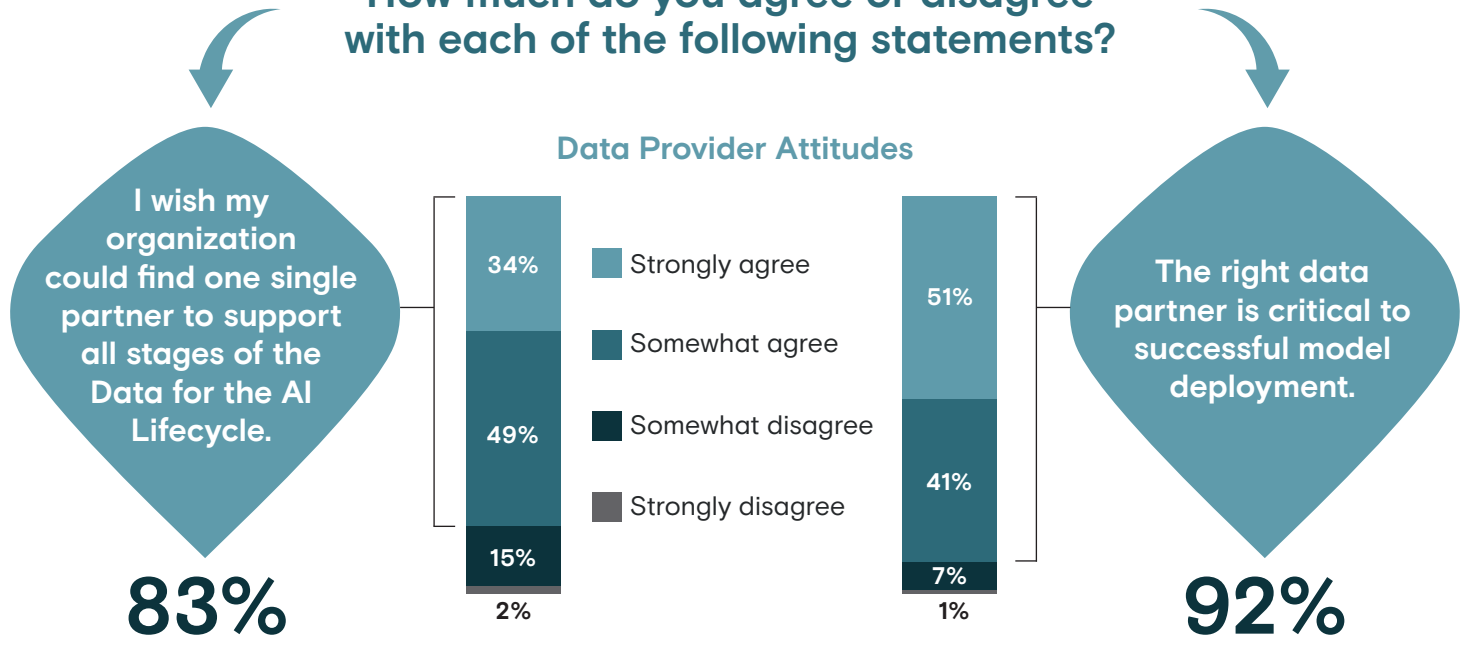




“Our unique ability to support all data-centric stages of the AI lifecycle for all data modalities positions Appen as the ideal external data provider.”
Sujatha Sagiraju,
Chief Product Officer - Appen

With those timely updates and moving toward working with external data providers, finding the right partner is important. 92% agree that finding the right data partner is critical to successful model deployment and validation, and most (83%) wish they could find a single partner to support all stages of the lifecycle. The importance of continuous validation of the model's performance is critical to successful model outputs.

How much do you agree or disagree with each of the following statements?



Finding the right partner with the technology and expertise is crucial to successful high-quality results. 93% agree that technology and expertise at each stage of the lifecycle are important to get high-quality results with 51% stating strongly agree.

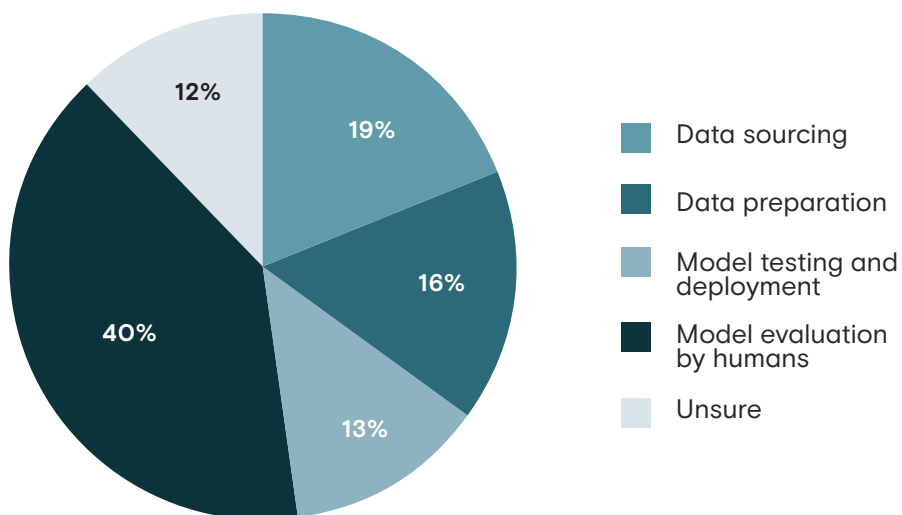
How much do you agree or disagree with the following statement?

Technology and expertise at each stage of the lifecycle are important to get high-quality results.



Least Budget Allocation By Stage

Model evaluation by humans largely has the least amount of budget allocated with 40% reporting they allocate the least budget to the final stage of the lifecycle. There is a gap between budget allocation and the importance of human-in-the-loop. Model evaluation is critical to ensuring that the AI model is accurate and reduces the need to source more data. By spending more budget on human-in-the-loop up front, companies will save money, time, and will be less likely to require re-evaluations in the future.



Adoption

AI adoption continues to grow in 2022, with benefits and applications fueled by innovation and a desire to find efficiencies and increase productivity within businesses.

As AI use becomes more commonplace, tools and best practices to improve AI have also become more advanced.

Following the race to launch during the pandemic, perceptions regarding the level of advancement of AI in organizations may be shifting. Our data shows a reduction in the belief that the organizations of the surveyed respondents are ahead in their industry (for the US market, 66% in 2021 vs 55% in 2022), likely due to the influx of AI utilization through the pandemic and the elevation of AI use cases across industries. While few feel their organization is lagging on AI adoption, business leaders are split down the middle on whether their organization is ahead of (49%) or even with (49%) others in their industry.

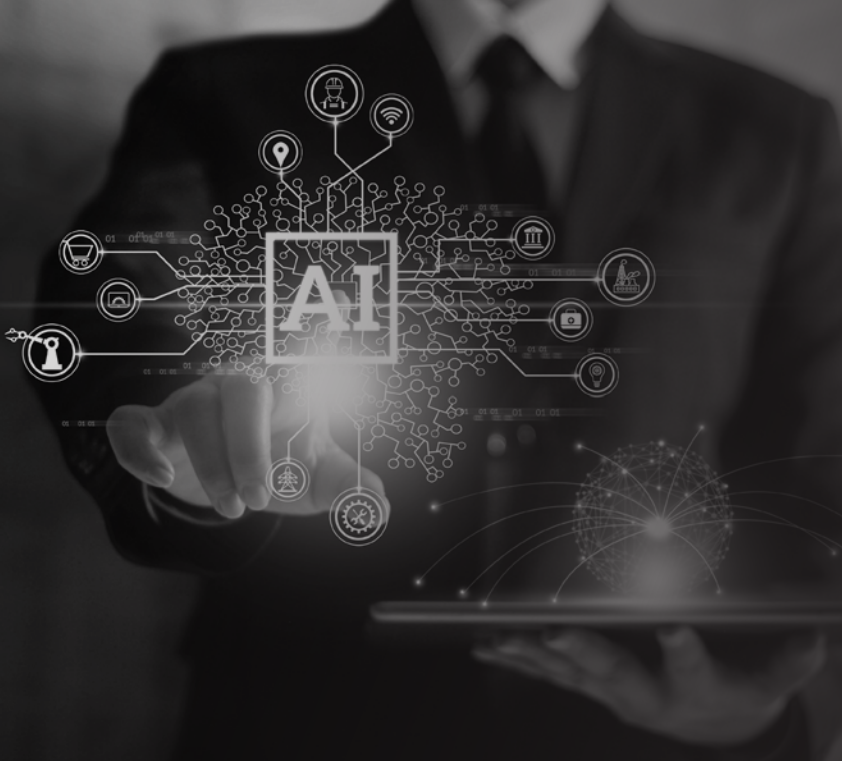
When compared to organizations in Europe, organizations in the US are more likely to say they are ahead of others in their industry at adopting AI (55% vs 44%).



“Our clients tell us that AI is a priority for their business because it is integral to their digital transformation programmes, many of which were kick started during, or because of, Covid-19.

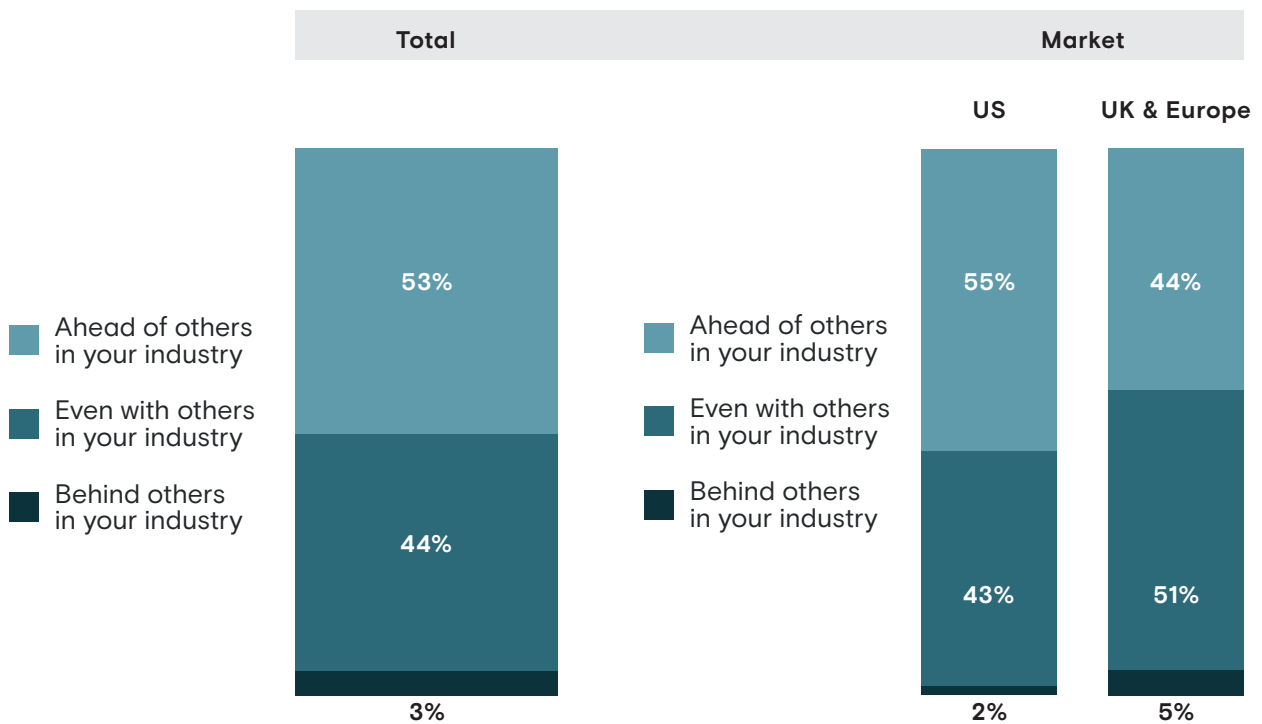
These AI initiatives within the UK and Europe have been established to fast track and improve business processes and help deliver sustainable growth in the tough economic climate of today. Our clients report that AI is fundamental to the success of these initiatives, with some saying it is now fundamental to commercial survival. They tell us that their AI strategies are displacing and disrupting elements of traditional business models, but that AI is being introduced in order to bring about business benefits, cost savings, resilience as well as innovation strategies for growth.”

**Sarah Lowe
VP of Business Development
in EMEA and APAC - Appen**



THE FOCUS ON PERFECTING SELECT PRODUCT LINES/ FUNCTIONS VERSUS ROLLING AI OUT WIDESPREAD SHOWS THE EMPHASIS OF ROI WITH A REPORTED AVERAGE OF 52.2% OF DEPLOYED PROJECTS SHOWING MEANINGFUL ROI.

When It Comes To AI Adoption Your Organization Is:

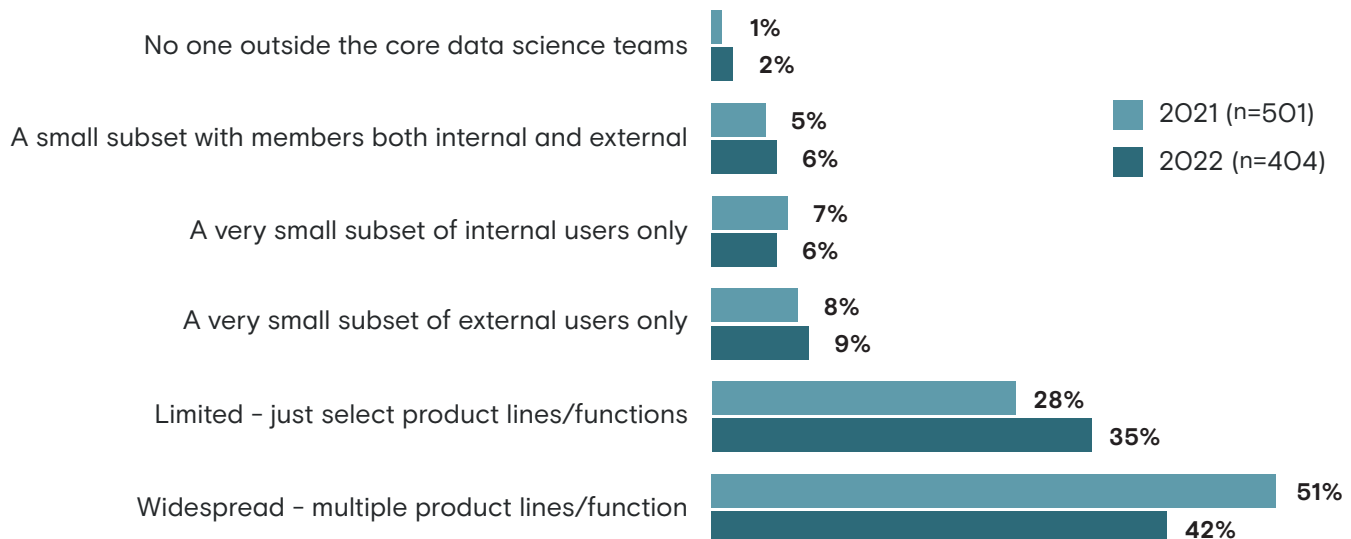


AI BUDGET IS HIGHLY CORRELATED WITH COMPANY SIZE AND COMPANIES PARTNERED WITH AN EXTERNAL TRAINING DATA PROVIDER HAVE LARGER BUDGETS.



From customer support to improving processes and production, businesses across sectors have exposed themselves to AI in recent years. Unsurprisingly, we saw this interest spike over the pandemic, and we are now reporting a shift in new initiatives. AI rollout in the US is somewhat more limited than in 2021 with an increase from 28% to 35% reporting just select product lines/functions and 51% to 42% reporting multiple product lines/functions. The focus on perfecting select product lines/functions versus rolling AI out widespread shows the emphasis of ROI with a reported average of 52.2% of deployed projects showing meaningful ROI.

Who Is AI Rolled Out To Right Now? (US Only)

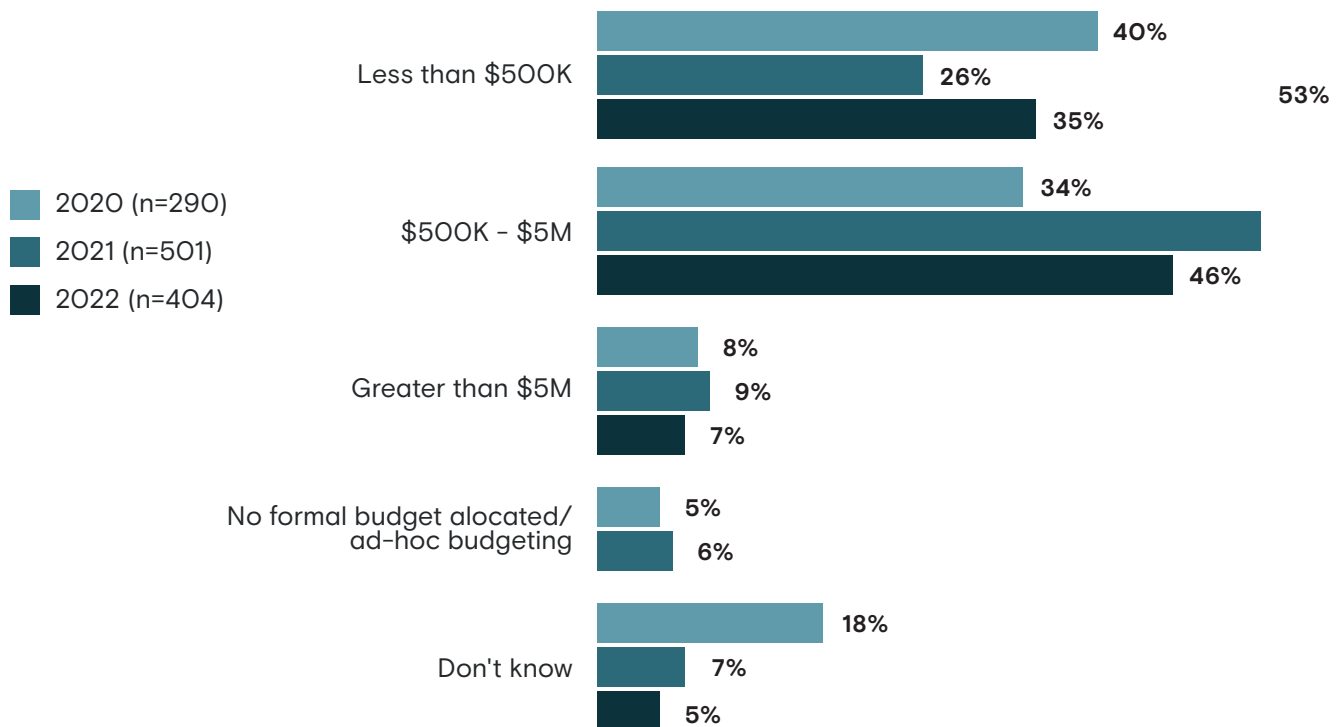


Survey respondents in the US indicated that allocated budgets for AI initiatives are similar in 2022, with 46% reporting budgets in the \$500k to \$5 million range (compared to >50% in 2021). Companies allocating budgets from \$500k to \$5M decreased 13% in 2022, with only 7% reporting budgets greater than \$5M.

This leads us to believe that while the industry continues to grow, early adopters of AI technology are now seeing the return on their initial investments as they move from project launch to maintenance.

Does your company have budget formally allocated for any AI initiatives and if so, how much?

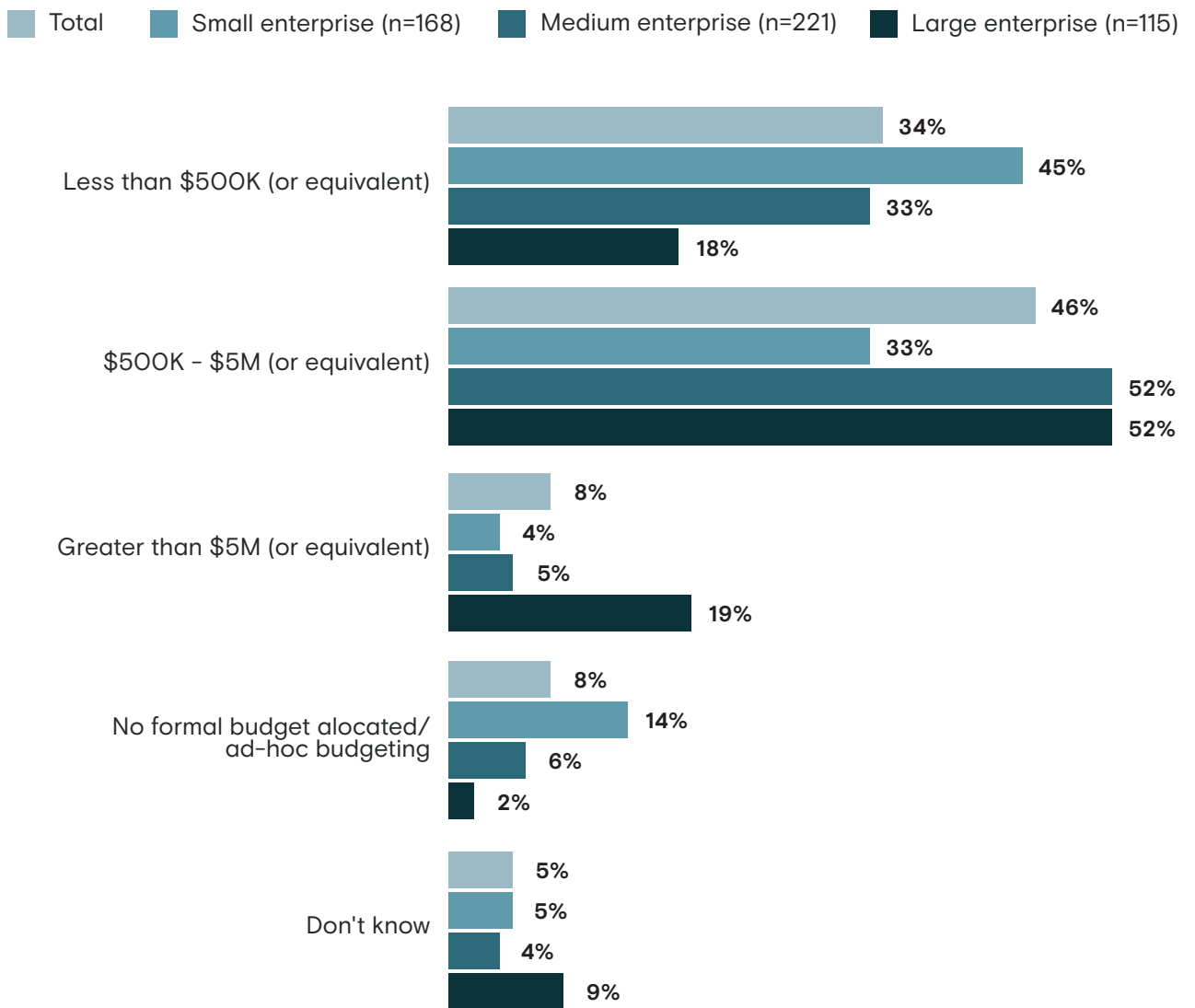
Trended AI Budget Allocation (US-Only)



Not surprisingly, AI budget is highly correlated with company size, with large organizations dominating the \$5 million+ range. Notably, companies partnered with an external training data provider have larger budgets in the \$500k to \$5 million range compared to the ones that don't work with a data provider (49% vs. 26%).

Does your company have budget formally allocated for any AI initiatives and if so, how much?

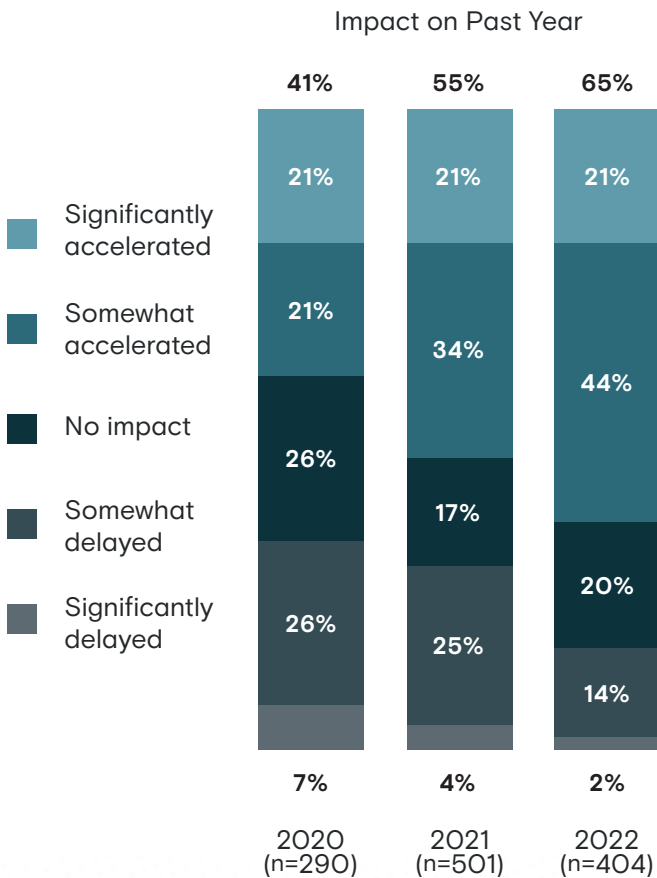
2022 AI Budget Allocation By Enterprise Size (In USD)



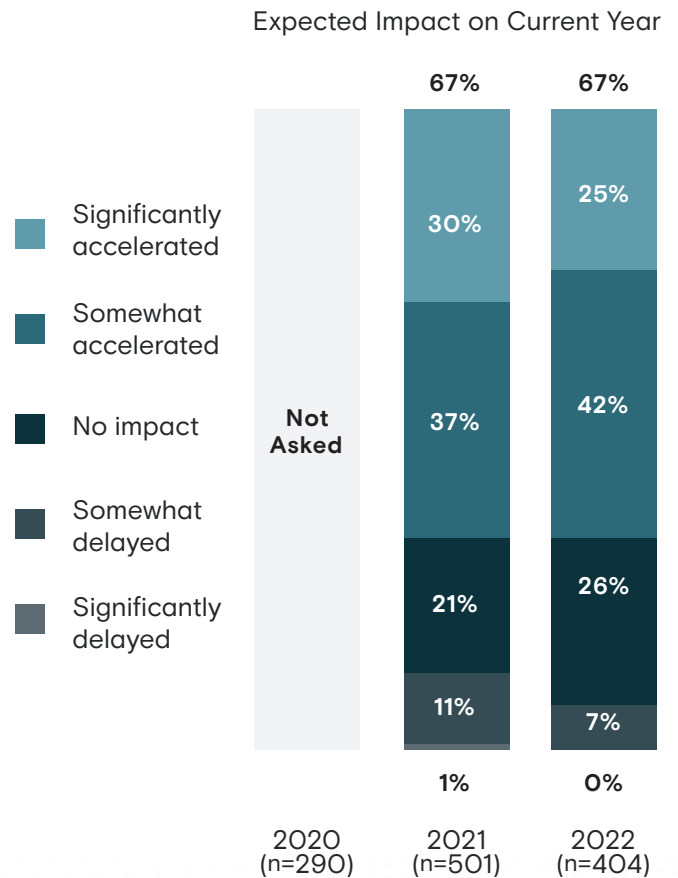
Last year, we predicted the pandemic would continue to drive accelerated growth and adoption of AI solutions across industries. With remote work and virtual interactions becoming a more accepted part of everyday life, businesses continue to seek AI solutions to meet customer demands. In the US, 65% of organizations stated COVID-19 accelerated their AI strategy in 2021 with 67% stating it will continue to accelerate in 2022 that is an increase from the 55% who stated COVID-19 accelerated their AI strategy in 2020 and even with the 67% who predicted it would continue to accelerate.

Trended COVID-19 Impact On AI Strategy (US Only)

To what extent has COVID-19 impacted your AI strategy?

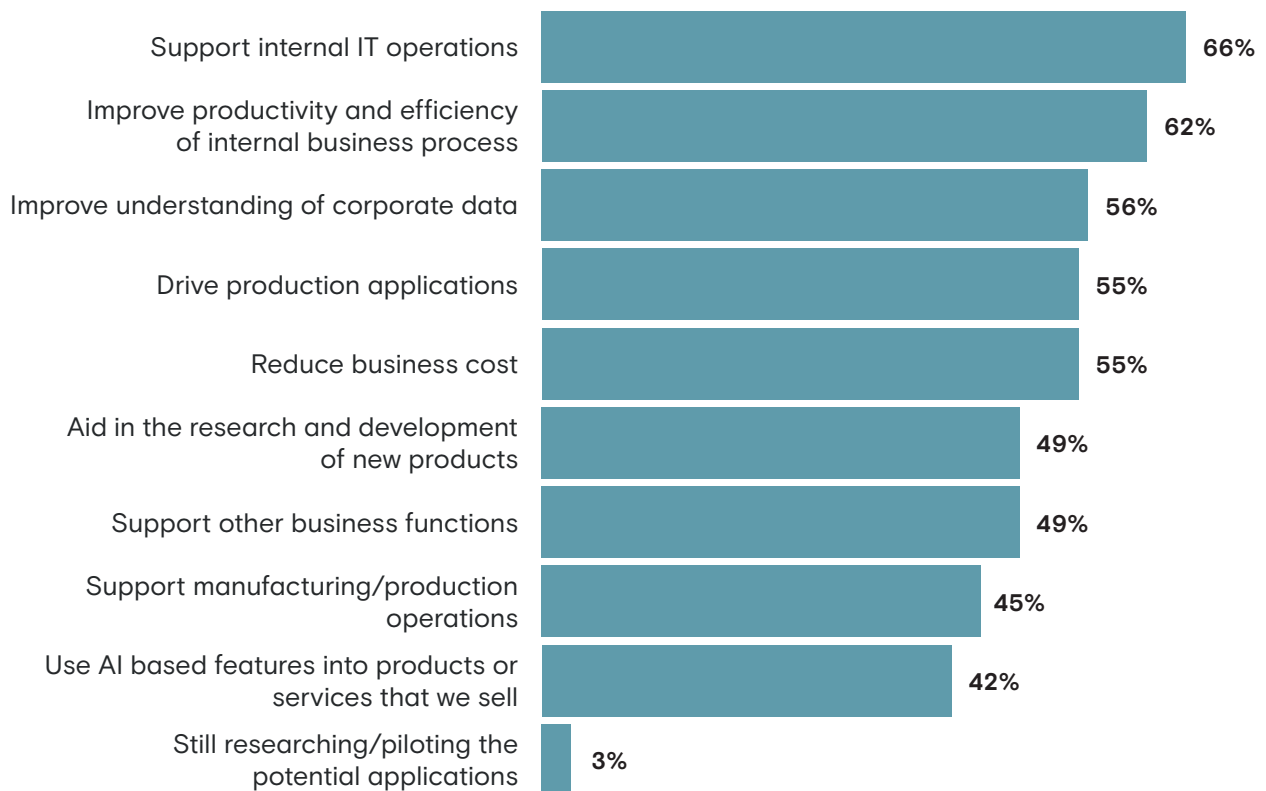


To what extent do you foresee COVID-19 impacting your AI strategy in the current year?



While AI use continues to expand to the consumer sector, we see continual growth in internal applications. Companies are seeking the benefits of artificial intelligence in supporting internal tech processes, improving productivity, and managing data. More than half of those surveyed leverage AI to reduce the cost of business and increase production.

Which of the following best describes the use of AI in your organization?



Ethics

One of the challenges in our industry is the perception that artificial intelligence poses ethical risks.

We are pleased to report that the large majority (93%) agree that responsible AI is a foundation for all AI projects within their organization. As diversity and inclusion become more prominent parts of mainstream AI and ML conversation, ethics at all stages of the AI lifecycle is more important than ever—especially regarding reducing bias and ethical data sourcing.

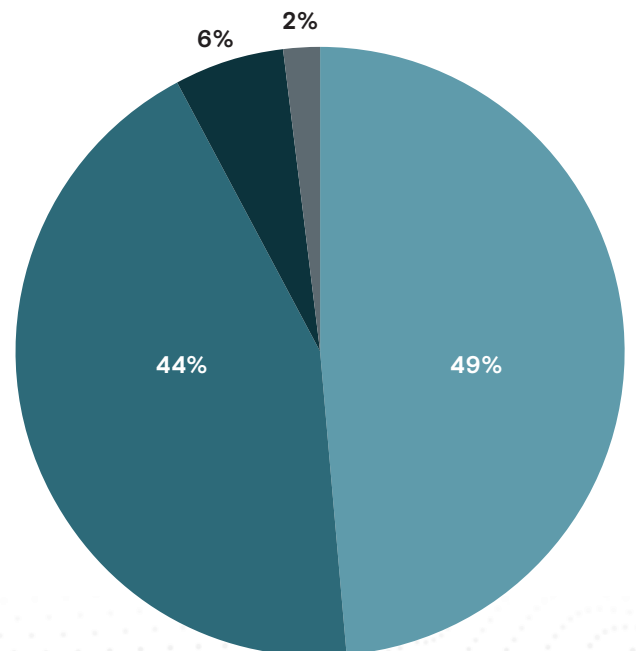


“Data ethics isn’t just about doing the right thing, it’s about maintaining the trust and safety of everyone along the value chain from contributor to consumer.”

Erik Vogt
VP of Enterprise Solutions - Appen

To what extent do you agree or disagree with the following statement? At my organization, responsible AI is a foundation of all AI projects.

- Strongly agree
- Somewhat agree
- Somewhat disagree
- Strongly disagree



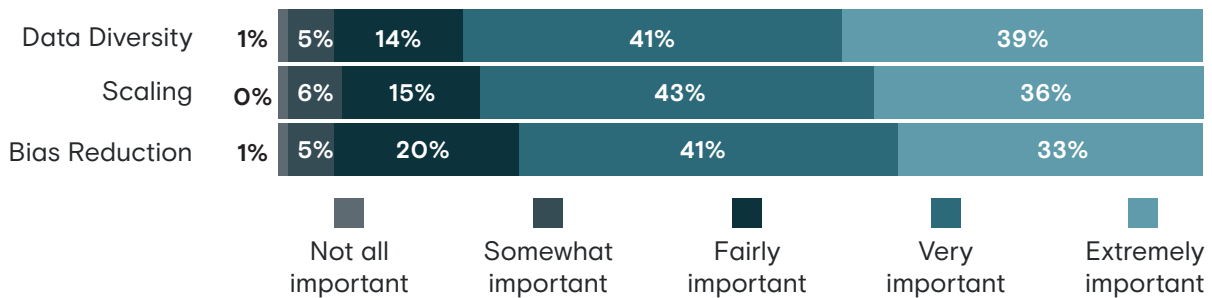


“There is a strong stance in Europe prioritising trust and ethical AI to ensure fundamental rights are protected. Europe is striving to lead the way in setting the global gold standard in these areas as well as become a powerhouse fostering excellence in AI to compete globally. We hear this time and again from business and government leaders in the UK and Europe.”

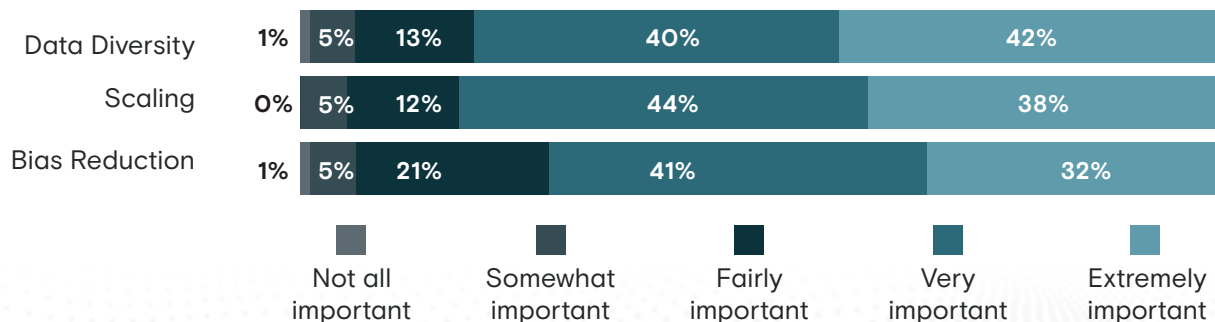
Sarah Lowe
 VP of Business Development
 in EMEA and APAC - Appen

With a few exceptions, data needs have largely held year over year. Last year, while bias reduction and scaling were strong AI priorities (75% and 76% extremely/very important in the US, respectively), data diversity (82%) topped the list. This year (in the US), there was a large jump in the importance of scaling, putting it on par with data diversity (82% each).

Importance of AI Features (n=504)



Importance of AI Features for the US (n=404)



“As a data optimist, I believe the data revolution has the potential to bring immeasurable benefits to people in ways that are only beginning to become apparent. But with this emerging power, comes the potential for harm from abuse or misuse of data, often carelessly or unintentionally. At its core I feel that data ethics is fundamental to our core sense of trust and integrity in, and for, ourselves, as well as in the technology we interact with.”

Erik Vogt
VP of Enterprise Solutions - Appen



While custom-sourced datasets are the preferred format across use cases, we’re seeing significant use of synthetic and pre-labeled datasets as well. In the US, more than 90% of participants indicated the value of pre-labeled datasets in the model training stage and more than 97% stating the importance of synthetic data in developing inclusive training datasets. With the importance of data diversity and desire to reduce bias, we predict this level of interest in new data formats will be mirrored globally as each market reaches maturity in their AI initiatives.

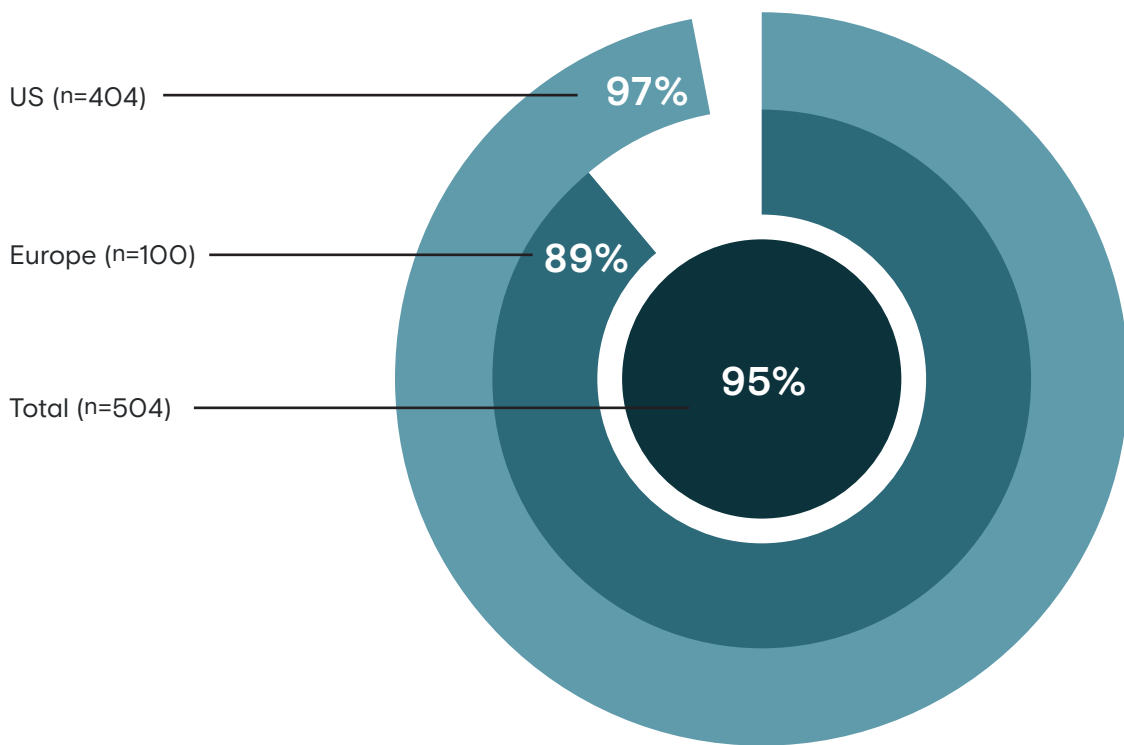
Most Commonly Used Dataset By Use Case

Use Case (from most to least common use)	Custom-collected	Synthetic	Pre-labeled
Data Classification (n=286)	42%	29%	27%
Search Relevance (n=189)	47%	24%	29%
Content Relevance (n=172)	45%	26%	28%
Computer Vision (n=161)	41%	27%	31%
Augmented and Virtual Reality (AR/VR) (n=161)	45%	29%	25%
Point-of-Interest (n=154)	39%	28%	31%
Chatbot (n=154)	38%	31%	30%
Translation and Localization (n=152)	39%	30%	28%
Natural Language Processing (NLP) and Speech (n=137)	39%	32%	26%
Autonomous Mobility (n=134)	38%	38%	23%

97% OF RESPONDENTS AGREE THAT SYNTHETIC DATA IS IMPORTANT TO CREATING INCLUSIVE DATASETS.



Synthetic Data Is Important To Provide Inclusive Training Datasets (% Agree)



"As more and more companies adopt the concept to supplement human-collected datasets, we can expect much more inclusive and representative datasets that will lead to safer and more equitable applications for all genders and races."

Excerpt from **Fast Company**

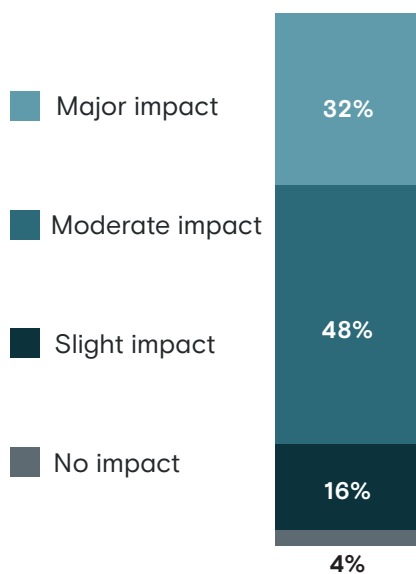
"How 'fake' data can make a real difference for people of color"

80% BELIEVE AI WILL HAVE A MODERATE TO MAJOR IMPACT ON HELPING TO CUT EMISSIONS AND FIGHT CLIMATE CHANGE.



Responsible AI is not only applied to the model build and deployment, but also what the models are being used for. 80% believe AI will have a moderate to major impact on helping to cut emissions and fight climate change. AI technology has the power to make the world a better place.

Impact Of AI In Helping To Cut Emissions And Slow Down Climate Change



In your opinion, how much of an impact do you think AI will have in helping to cut emissions and slow down climate change?

The World Economic Forum shares how AI can support diversity, equity and inclusion, quoting Appen CEO Mark Brayan who said “creating AI that’s inclusive requires a full shift in mindset throughout the entirety of the development process.” Outlined in the article were four ways to ensure AI is inclusive; diversity, transparency, education and advocacy. These are required throughout the entire data for the AI lifecycle from sourcing to deployment to evaluation. Excerpt from World Economic Forum Agenda "How can AI support diversity, equity and inclusion?"

Conclusion

Change is good, especially when it means businesses and leaders can put more time and energy into planning for their future with AI. There's been strong alignment, this year, in priorities that resulted in some positive changes. Among those surveyed, there's an agreement that the data used for AI models should be sourced responsibly, diverse, and remain high-quality.

Our findings show that consensually business leaders and practitioners agree that high-quality data is critical in creating successful AI models. Some responses indicate that more time is being spent in the AI adoption and focus stage. Instead of racing to stay ahead of the AI competition, companies feel they are in a secure place and can afford to spend more time planning their next steps. Spending more time making decisions can correlate to less money actively spent. However, we are still seeing budget increases for larger sized companies as they are looking to partner with external data partners to meet their AI needs. With these budget increases we also anticipate growth in the AI adoption stage in the coming years.

In conjunction with the changes and requirements companies have for their data sourcing; they have been looking to find a single partner to help manage all stages of data for the AI lifecycle. Thus, ensuring high-quality data is obtained. These focal shifts aren't just for larger sized companies to stay ahead, but for companies of any size to ensure they keep up with consumer demands and expectations. AI may not be a critical offering for all organizations yet, but growing interest within and outside of the US indicates that strategic decision makers see AI initiatives as a means to establish market leadership.

As is true of 2021, COVID-19 played an important part in the continued growth that's been witnessed in the industry. Companies are no longer playing catch-up but are now planning ahead for their next AI advances, albeit carefully, to ensure there won't be a repeat in the future. Our goal in developing this report is to provide the current state of AI and machine learning from the perspective of top decision-makers across industries and companies. If you have any questions on the information presented here, please reach out to our team.



Methodology

The objective of the State of AI 2022 survey was to evaluate AI and data for the AI lifecycle implementation across organizations by gathering responses from business leaders and practitioners.

The Harris Poll, our research partner, surveyed 504 respondents through an online survey between June 2, 2022, to June 14, 2022*. The random sample consisted of 404 decision-makers in the United States and 100 decision-makers in the United Kingdom, Ireland, and Germany. All respondents worked for companies with 100+ employees and our qualification questions ensured that 33% of respondents were from companies with 1,000 or fewer employees, 44% were from companies with between 1,001 and 10,000 employees, and 23% were from companies with more than 10,000 employees.

Raw data were not weighted and are therefore only representative of the individuals who completed the survey.

Respondents for this survey were selected from among those who have agreed to participate in our surveys. The sampling precision of Harris online polls is measured by using a Bayesian credible interval. For this study, the sample data is accurate to within +/- 4.4 percentage points using a 95% confidence level. This credible interval will be wider among subsets of the surveyed population of interest.

All sample surveys and polls, whether or not they use probability sampling, are subject to other multiple sources of error which are most often not possible to quantify or estimate, including, but not limited to coverage error, error associated with nonresponse, error associated with question wording and response options, and post-survey weighting and adjustments (not applicable in this case).



Industry

- Advertising & Marketing
- Automotive & Transportation
- Business Support & Logistics
- Construction, Machinery & Homes
- Education
- Entertainment & Leisure
- Finance & Financial Services
- Food & Beverages
- Healthcare & Pharmaceuticals
- Insurance
- Manufacturing
- Real Estate
- Retail, Consumer Durables & E-Commerce
- Telecommunications, Technology, Internet & Electronics
- Utilities, Energy, and Extraction

Role

- Software/App Developer
- Technical Manager Data
- Engineering
- DevOps/Developer
- Data Analyst
- Data Architect/Applications
- Architect/Enterprise Architect
- Software/SaaS Engineer
- Data Engineer
- Data Scientist
- AIOps Engineer
- Machine Learning Engineer
- VP/C-Level Executive
- Program Manager/Director
- Business Process/Dept Owner
- Data and Analytics Manager

Company Size

- 101 – 500
- 501 – 1,000
- 1,001 – 5,000
- 5,001 – 10,000
- 10,001 – 25,000
- 25,000+

* The 2021 study was conducted by The Harris Poll in March 2021 among 501 US ITDMs; the 2020 study was conducted using a different method among 290 ITDMs



About Appen

Appen is the global leader in data for the AI lifecycle. With more than 26 years of experience in data sourcing, data annotation, and model evaluation, we enable organizations to launch the world's most innovative artificial intelligence systems. Our expertise includes a global crowd of more than 1 million skilled contractors who speak over 235 languages, in over 70,000 locations and 170 countries, and the industry's most advanced AI-assisted data platform. Our products and services give leaders in technology, automotive, financial services, retail, healthcare, and governments the confidence to launch world-class AI products. Founded in 1996, Appen has customers and offices globally.

- ❖ Experience working in 170+ countries
- ❖ Expertise in 235+ languages
- ❖ Access to a curated crowd of over 1 million flexible contractors worldwide
- ❖ Over 1,125 employees located in offices around the globe
- ❖ Nearly 1 billion judgements made and 3 million images and videos collected in 2020.
- ❖ 25+ years working with leading global technology companies