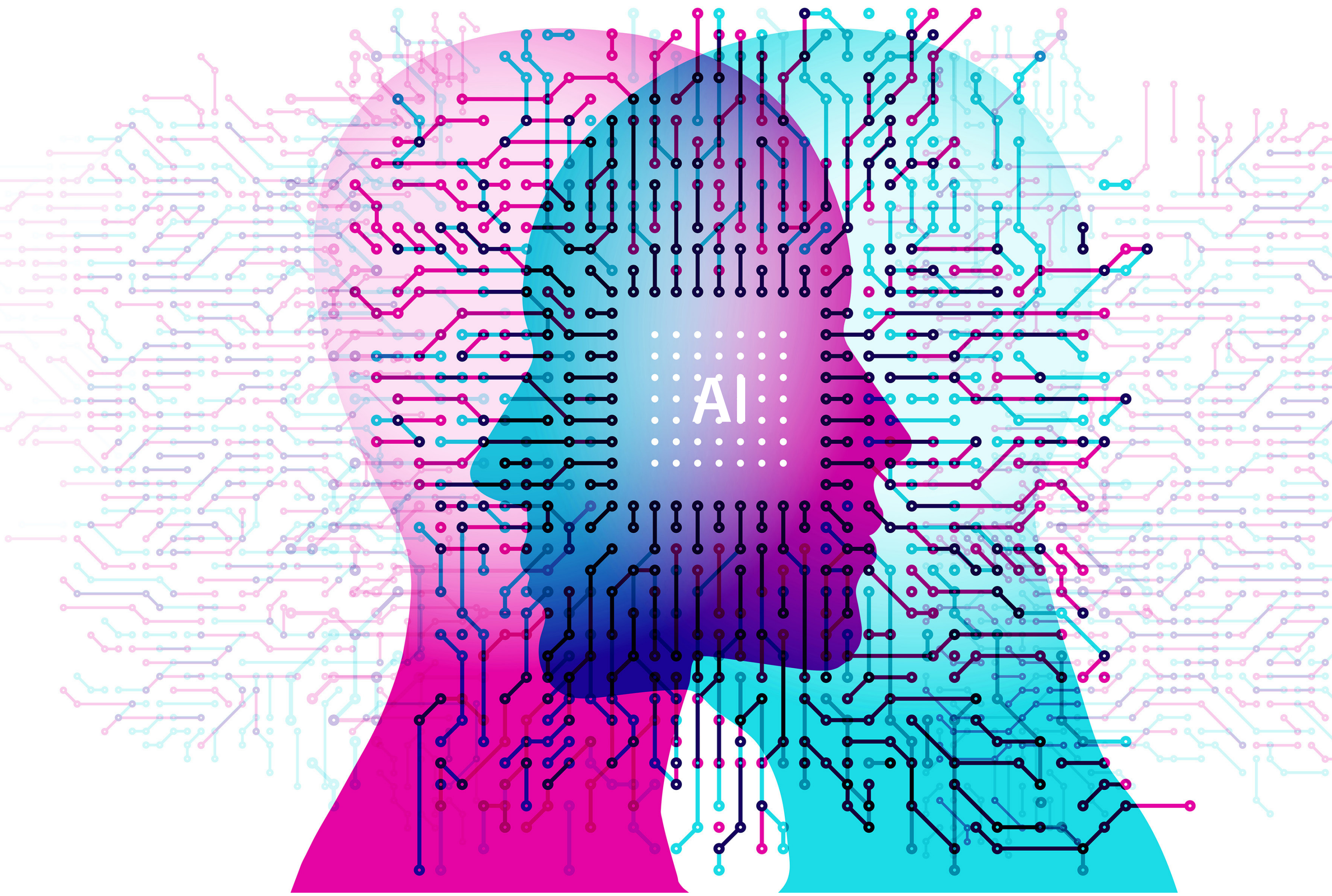
Infosys
Navigate your next

DATA+AI RADAR 2022

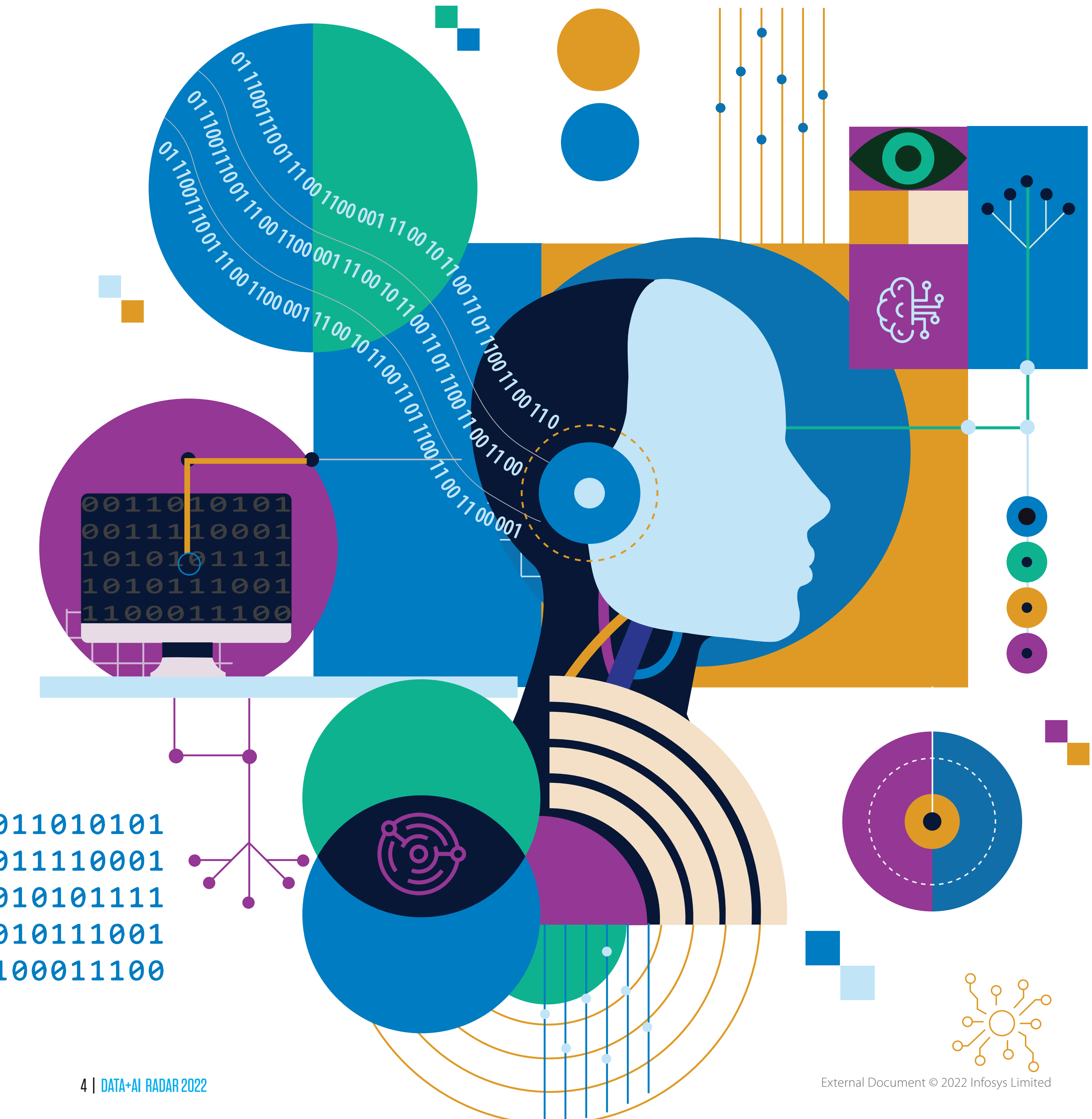
MAKING AI REAL: FROM DATA SCIENCE TO PRACTICAL BUSINESS

Infosys | Knowledge Institute


AI

Contents

Generating real value from data and AI requires companies to share data, build trust in the results delivered by advanced systems, and keep business goals at the fore.



Executive summary

Making AI Real: From data science to practical business

Companies need to think differently about data and artificial intelligence (AI). They have invested heavily in AI systems in the past few years. Global spending on AI-centric systems will approach $118 billion in 2022 and grow to more than $300 billion by 2026.[1] All that work is not paying dividends. Companies have achieved basic AI capabilities. This is not what they want. Three out of four companies in our survey want to operate AI at enterprise scale.

Infosys Knowledge Institute's inaugural Data+AI Radar identifies why AI fails to deliver on heightened expectations, and recommends three areas for improvement: develop data practices that encourage sharing, bind explanations into advanced AI, and focus AI teams on business. If companies improve on these fronts, they can add up to $467 billion in profit growth, collectively, and increase internal satisfaction with data and AI.

Companies are new to advanced AI – and it shows

Our survey of 2,500 AI practitioners found that 81% deployed their first AI system in the past four years. However, most companies (85%) have not achieved advanced capabilities, and most AI models (63%) are still driven by humans. Compounding this, outcomes are middling at best: Users are highly satisfied with their data and AI results only about a quarter of the time.

Data is not the new oil

Businesses can no longer afford to think of their data as oil, extracted with great effort and valuable only when refined. Data today is more like currency: It gains value when it circulates. This data-sharing economy is already up and running. In our study, companies that shared data, in and out of their organization, are more likely to have higher revenue and use AI better. Refreshing data closer to real time also correlates with increased profits and revenue. However, companies can't share data until they trust it. Shortcomings in data verification, data practices, and data strategies continue to hold companies back.

Advanced AI requires trust in all directions

Advanced AI requires a new way of thinking about technology and business. It requires trust: Trust in your own and others' data management, and trust in AI models. Pristine data and perfectly programmed AI models mean nothing if humans do not trust and use what data and AI produce. Companies that are the most satisfied with their AI consistently have strong, responsible AI practices.

Trust, share, and focus on value

Our analysis identifies ways data and AI work better together:

1. **Get your data right, and share it.** Focus on data sharing capabilities and hub-and-spoke data management.
2. **Build trust in advanced AI.** Strengthen ethics, bias management, deep learning, AI cloud, and scaling across the organization.
3. **The AI team needs a bias for value.** Business leaders matter as much as data scientists.

Combined, these actions will scale AI and unlock value – positively impacting the bottom line and user satisfaction.

Companies are new to AI – and it shows

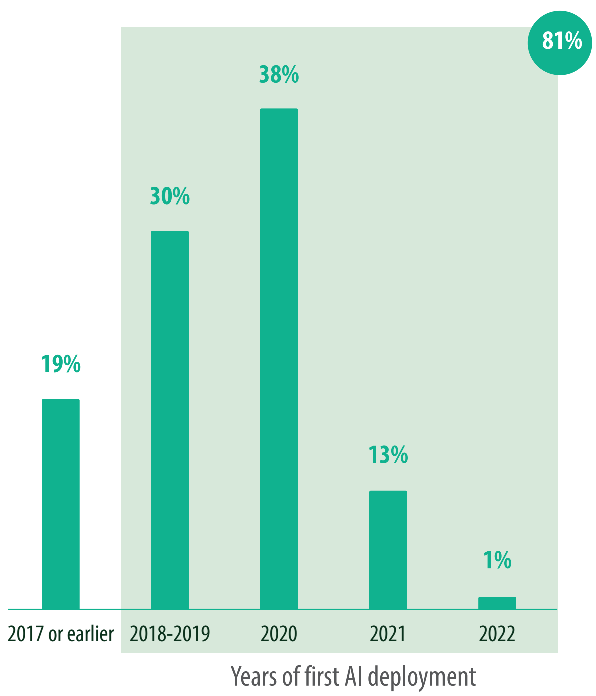
It's not hard to find a business that could be improved by data and AI. Simply look for an exercise with routine, highly manual processes and lots of data.

That notion led many real estate companies to pursue "iBuying" – developing algorithms to speed up and optimize residential home-buying in the US. It hasn't worked.

Zillow, the market leader in public-facing residential real estate data, shut down its algorithm-driven iBuying unit in November 2021 and took a $300 million charge for the quarter.[2] OpenDoor, a tech startup built on iBuying has yet to show a profit, in spite of raising $1.9 billion in eight years in business.[3] Zillow began offering up computer-generated home estimates 15 years ago. After all that time, its algorithm still couldn't deliver reliable results.

These outcomes should serve as cautionary tales, but scores of big businesses are coming new to AI, eager to apply data and AI in big ways.

Figure 1. Most companies deployed their first AI system within the past four years



As the saga of iBuying illustrates, AI is not new. But it's being newly applied in more businesses than ever. A decade ago, digital giants including Amazon, Google, and Microsoft accounted for most data and AI activity.[4]

Around the turn the millennium, these companies had two things going for them that most enterprises lacked: zettabytes of data and petahertz of computing power. Now every company is like Amazon in 2001: Gallons of data flow into corporate reservoirs and computing power can be ramped up with the touch of a button.

The result: AI is spreading like wildfire. But most companies are new to attempting their own advanced AI. We surveyed 2,500 AI practitioners from companies around the globe, and found that four out of five companies put their first AI model in production less than five years ago (Figure 1).

To be sure, many enterprises have been making use of classic rule-based artificial AI and simple automation for years. For instance, a European paper producer has prioritized maintenance work and reduced pump failures in its paper mills using a system developed by Nokia, Infosys and the firm formerly known as Pöyry (now a unit of AFRY).[5]

But today's interest stems from the belief that AI can derive value from massive amounts of data and develop novel consumer products.[6] Data scientists, AI experts, and academics have achieved significant advances in AI.[7] But companies are struggling to answer: "What can AI do for business?"

Most people, when they think of AI, think of super-advanced models designed to predict the future and change the business, says Balakrishna DR, Infosys senior vice president and head of AI and automation.

"Companies want to use AI across their enterprise, to uncover hidden insights and construct new business models. But few are achieving this because it is a multi-faceted challenge involving data processes, AI techniques and interdisciplinary teams," he says.

Most AI is basic, and not autonomous

More often than not, data and AI fail to deliver high satisfaction from users because most companies use only basic AI. A full 85% of AI practitioners have not achieved top-tier capabilities – the sort of capabilities that are closer to AI that can predict the future.

We asked respondents what capabilities AI systems deliver, and scored answers across our Sense, Understand, Respond, Evolve (SURE) taxonomy (Figure 2). The Infosys Knowledge Institute developed SURE with the guidance of AI expert Rajeshwari Ganesan. The framework draws inspiration from the eight layers of an AI stack articulated by Shailesh Kumar, chief data scientist at India telecom firm Jio. (He describes the eight-layer AI stack, or Ashtang-AI, in his keynote address to the virtual 2020 Open Data Science conference.)[8]

The SURE taxonomy includes four tiers, ranging from the basic Sense capabilities (simple signal processing, such as being trained to recognize an object) to the most advanced, Evolve level (a system that senses, finds causes, acts on recommendations, takes feedback, and refines its performance). Survey respondents said only 15% of AI systems reach the evolve tier.

"The digital giants, including the cloud giants and others such as Apple, Facebook, and Netflix, are able to attain top-tier Evolve AI capabilities, but other large enterprises can't," explains Balakrishna. "The Fortune 100 aren't there because of their old systems and ways. They want to do AI, but they don't know how to get something out of it."

More than one-third (36%; Figure 2) of AI delivers basic Sense capabilities. While that may not result in game-changing insights for an enterprise, it's a pretty reliable use for AI, said John Bohannon, director of science at Primer.AI, a California-based AI and natural language processing business.

"From an engineering and data science point of view, monitoring is the most successful AI task," he said. "That's because the AI doesn't have to be as good as a human at it. The AI can provide value by prioritizing the flood of inbound information."

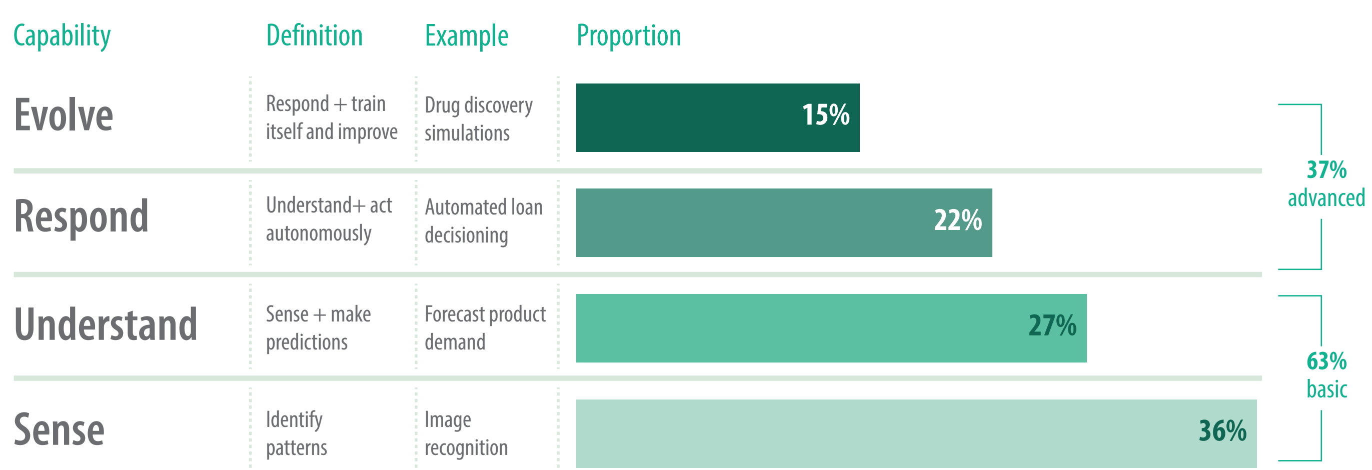
AI systems that rank and order enable workers to look at things more efficiently.

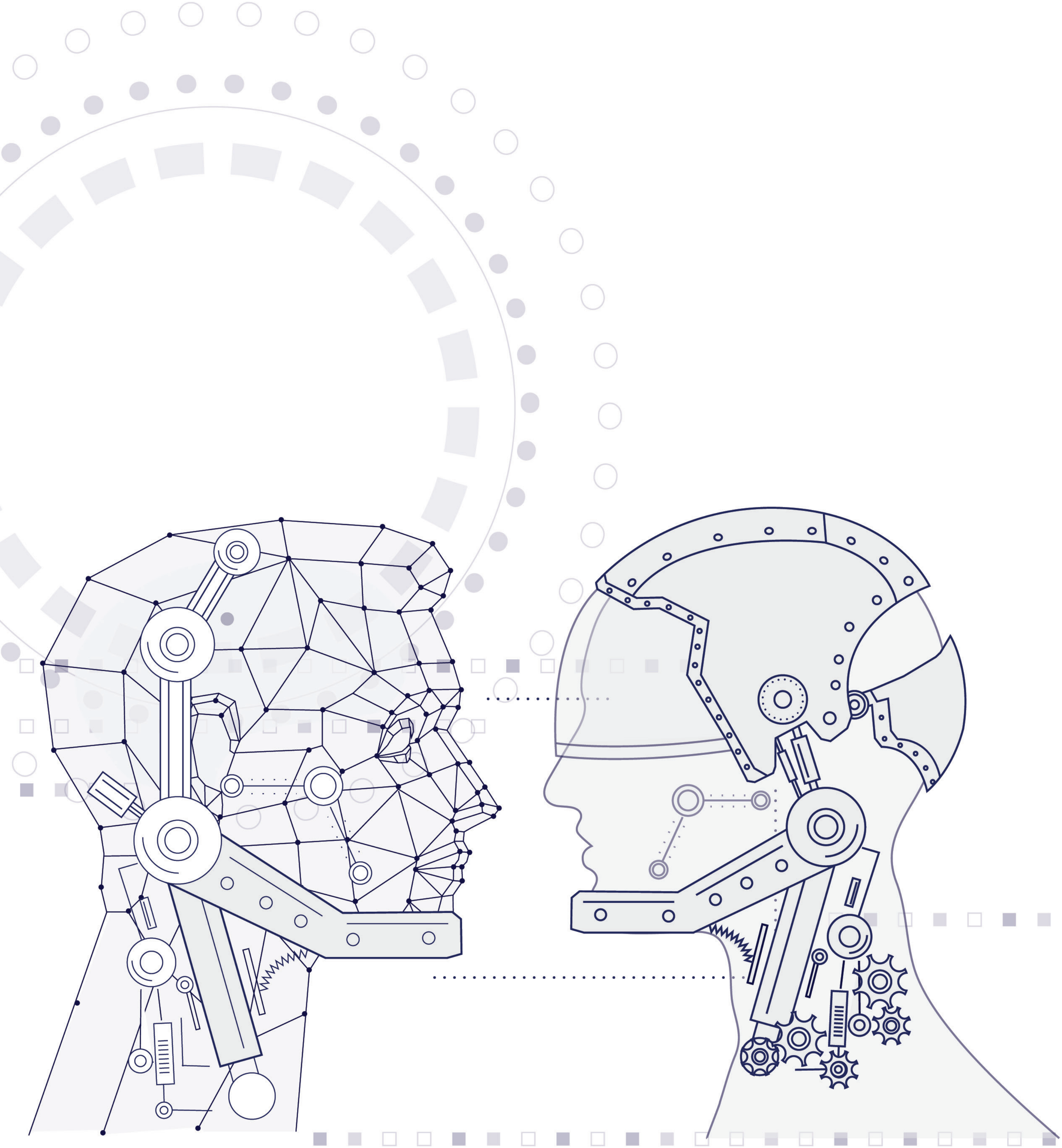
Sensing inputs is also the most forgiving sort of AI when it comes to data quality. "You just have to match format," Bohannon says. Sense-level AI typically comprises only a few different data formats. Of course, matching formats grows more complex as AI models grow more intricate.

From a business perspective, these capabilities are commodities. Most companies have them, and they are not differentiators.

"Companies need advanced AI if they are to achieve the loftiest ambitions of AI and stand out from competitors," says Sunil Senan, Infosys senior vice president head of data and analytics.

Figure 2. Sense, Understand, Respond, Evolve (SURE) taxonomy: Only 15% achieve top AI capabilities





Humans limit the speed of AI

The SURE taxonomy reflects that companies cannot ascend to advanced capabilities without first achieving the basics. Think of AI in the context of an automobile. Collision-avoidance alarms and lane-assist technologies are at the basic Sense and Understand levels. Adaptive cruise control and fully autonomous vehicles are at the advanced Respond and Evolve levels. Simply put, the big break between basic and advanced AI capabilities is who's in the driver's seat.

For example, UK medical insurance company Bupa Global takes in claims from 220 countries around the world, and has worked with automation and AI to extract relevant details from different sorts of claims submissions.

"What you're saying to an AI is, 'Here is an invoice in a format you've never seen before. Can you pull the data out?'" says Bupa Global IT strategy head Steve Williams. "If you don't get to a high enough confidence level, someone's still going to look at that piece of paper."

Williams said it's important for companies to think about what problem they are trying to solve with AI. While using AI to review invoices globally still needs a human touch, he says Bupa is achieving good outputs with that process trained on UK health claims, where a monolithic National Health Service and standardize invoices make the AI learning more manageable.[9]

In our survey some 63% of AI models fall into the Sense and Understand capability levels, which require user interventions. This restricts ultrafast computers to run only at the speed of humans, and limits how much data and AI can do.

This limitation is an artifact of early AI implementations in business, Rajeshwari notes. A decade ago, organizations typically began their AI work with rule-based automation using tools such as robotic process automation (RPA). Once deployed, the rule-based system added intelligence incrementally. For example, RPA tools added process discovery using deep neural networks, or added intelligent document extraction using machine vision algorithms. In these systems, automated AI is a small part embedded in a larger system.

The challenge, though, is that when 63% of AI systems rely on humans, they can't be ported from automated to autonomous. Automated AI can be built incrementally from rule-based systems – autonomous AI cannot.

To operate in dynamic and unpredictable environments, autonomous AI systems must be constructed through reinforcement learning, generative adversarial networks (GANS), and advanced neural networks. All this enables companies to engage in continuous learning and evolve in response to new stimuli. The transition to a live enterprise that can sense, respond and evolve smoothly can be daunting.[10] But this can be made easier by employing a technology architecture built with space and flexibility to incorporate the autonomous flow of information. Harnessing autonomous AI is not simply a matter of adding bandwidth and processing more bits. It requires companies to think differently.

"Autonomy is the way of the future, but I don't think it's going to be a flip-the-switch situation where everything goes autonomous. It's going to be a slow, gradual change." says Saurav Agarwal, founder and CEO of Siera.AI, an Austin, Texas, maker of safety and autonomous driving systems for use in warehouses. "We're at that stage where people are trying to figure out, 'What can I do with the automation? Where can I plug it in? What can I do with it?'"

Sparse satisfaction, but financial services a bright spot

Enterprises were attracted to AI that is intelligent, fast, and automated. So far, most have built AI that is simple, slow, and manual.

With companies struggling with what their AI capabilities actually are (basic) and what they hope they will achieve (advanced and autonomous), business and AI leaders must do more to manage expectations.

More often than not, data and AI fail to deliver. We asked AI practitioners what they used their data and models for, and if it was working. Respondents rated satisfaction with five use cases for their industry. They said data and AI left them highly satisfied one out of four times (Figure 3). We also mapped use cases by satisfaction levels and usage levels (frequency of use case), and found that only 18 of 63 (29%) scored in the high satisfaction zone (top of Figure 4).

Additionally, 12 of the 18 high-satisfaction use cases are more basic detect or automate functions, such as enabling equipment to self-diagnose potential problems in high tech and manufacturing applications or automated compliance and customer services tools in financial services. The financial services industry recorded the strongest satisfaction with its data and AI uses. It is the only industry to rate all five of its use cases in the high-satisfaction category.

As the financial services sector shifted from analog to digital, the quantity and complexity of products increased drastically, notes Mohit Joshi, Infosys president and financial services technology expert. This creates new business opportunities that require robust data governance.

"Companies must be able to securely and instantly share data across platforms and services to enable seamless services to the customer. A strong governance strategy for managing data while ensuring quality and minimizing risk will enable faster development and more sophisticated data-driven decision-making capabilities," he says. "A good mix of secure data controlled with the right privacy controls is a good way for firms to keep in line with regulation but still derive maximum value from the data for the end user," he says.

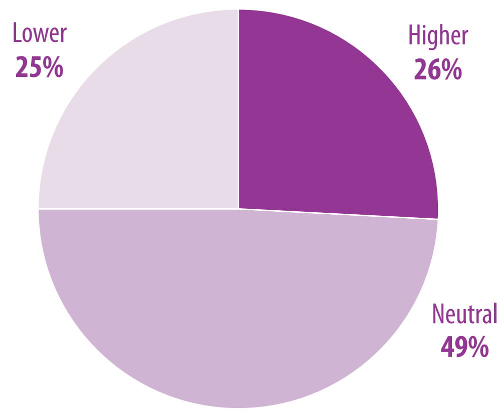
The next-highest rates of satisfaction are, in order, in retail and hospitality, healthcare, and high tech (Figure 5). But general satisfaction doesn't mean companies have optimized AI for business. For example, in retail (Figure 6), checking for inventory and streamlining checkout received high satisfaction ratings. But flashier and popular tech, such as virtual reality and augmented reality, doesn't meet expectations.

On the dissatisfied end of the spectrum, AI tools that aim to deliver personalized recommendations or products received lower satisfaction ratings. Telecom, energy, and CPG report low satisfaction with their data and AI uses.

Financial services firms can unlock additional value for their customers by implementing strong privacy controls coupled with seamless data exchange with partners, amplifying the power of customer's own data."

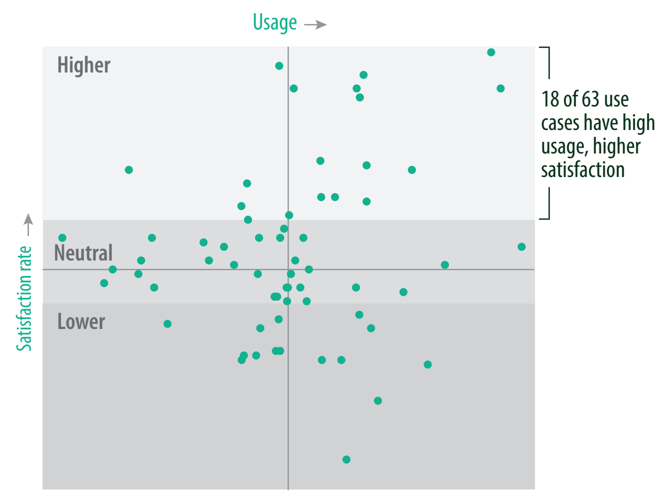
– Mohit Joshi
President, Infosys

Figure 3. Nothing special: AI outputs fail to excite



Overall satisfaction rates for data and AI

Figure 4. Only one in three highly satisfied with industry use cases



Satisfaction and usage rates for the 12 industries

• Industry-specific use cases

Figure 5. Financial services most satisfied, telecom least

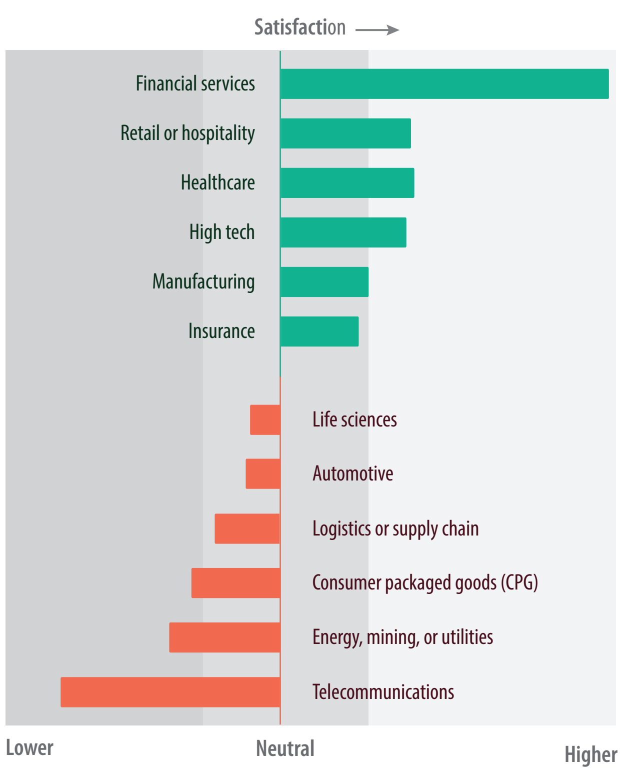
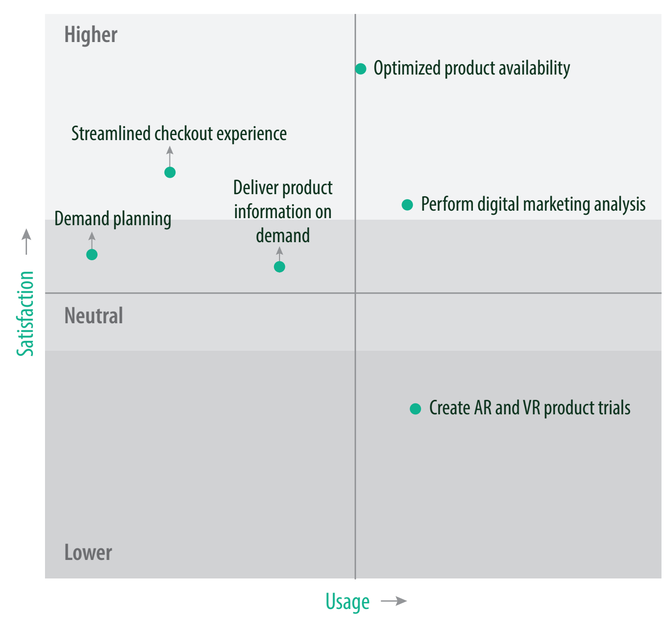


Figure 6. Retail: Low satisfaction for AI in AR and VR



Satisfaction and usage levels for retail

A conversation about AI with AI

In researching data and AI, the Knowledge Institute interviewed more than 100 AI, IT and business experts from Infosys, industry and academia. In the final stages of drafting Data+AI Radar we spoke with Infosys Consulting's Chief AI Architect Harald Gunia and Chief Data Scientist Serge Plata on some follow-up questions. In addition to their own expertise, they enlisted GPT-3, a popular AI language model that can generate its own answers, to weigh in on our questions. Here's how that conversation went.

Knowledge Institute: Does the nature of some industries ask more from AI than others?

GPT-3: There is no one-size-fits-all answer to this question, as the nature of AI and its applications vary greatly from industry to industry. However, some industries may require more AI resources than others, depending on the complexity of the tasks involved and the amount of data available.

Harald Gunia: That's generally true, but further industry-specific criteria are: type of data (life science & insurance have more texts, images & time series and hence can make more use of deep learning than consumer packaged goods and banking), type of regulation (the black-box nature of machine learning makes its use harder in tightly regulated industries like life science & banking), and importance of mobility (industries like agribusiness and logistics require AI-solutions to be integrated into 5G-frameworks).

Knowledge Institute: How can you keep the AI effort focused on the most critical business problems?

GPT-3: There are a few ways to keep the AI effort focused on the most critical business problems:

1. Define the business problem that you want to solve with AI, and then identify the AI technology or approach that is best suited to solving that problem.
2. Work with business stakeholders to identify the most critical business problems that AI can help with, and then prioritize those problems.
3. Conduct a feasibility study to assess whether AI can help solve a particular business problem, and if so, how effective it is likely to be.

Harald: I would also recommend using AI-specific business case templates with successfully calculated example cases. Great candidates of critical business problems in the past dealt with time-critical decisions to be taken based on incomplete information, processing information-rich unstructured data or acceleration/automation of routine tasks, where humans have become unavailable or too expensive

Serge Plata: You need to superimpose a continuous improvement program on the AI roadmap …. the secret is to align the AI resources and data strategy with the overall strategy of the business.

Knowledge Institute: Is there something specific to telco, energy or CPG that makes good AI uses more challenging?

GPT-3: There is no one-size-fits-all answer to this question, as the challenges associated with AI applications in telco, energy and CPG industries vary depending on the specific use case. However, some common challenges include the need for large amounts of data to train AI models, and the need for specialized expertise to develop and deploy AI applications.

Harald: More specifically, telco and energy have a lot of time-series data, which require specialized machine learning models. All industries have data-privacy challenges, while computer vision plays an increasing role in CPG (in the form of visual search and visual question-answering).

Serge: Agree with Harald. Data privacy is a main issue in all industries, but especially in telco. Another aspect that concerns many energy companies is social responsibility, so AI should also be responsible AI.

In line with our survey results, GPT-3 delivers agreeable but not particularly satisfying answers to our questions. Cocktail party-grade insights about emerging technology are not the outputs companies want to achieve from their data and AI.

Data is not the new oil – it needs a new metaphor

Sixteen years ago, business strategists began talking about data as the new oil.[11] Data, like oil, is hard to extract and valuable only after it is refined, they argued. And for many global enterprises in 2006, good, refined data was rare and held a potential akin to jet fuel. Today, acquiring and manipulating data is fairly frictionless. Drilling for oil remains a complex engineering feat. Refining crude oil takes enormous capital expenditures, specialized equipment, and trained professionals. Refining data still comes with challenges, but can be done with on-demand cloud, off-the-shelf tools, and citizen data scientists.

In 2022, data needs a better metaphor.

For organizations trying to extract value from data, we believe there are three more modern metaphors that can help them keep in mind the challenges and best practices they require for success:

Data is more like nuclear power.

Data is enriched with potential, in need of special handling, and dangerous if you lose control. The advance to data in cloud and AI in cloud have flipped the script for data. Companies don't lose data. Twenty-first century data has a long half-life. When to use it, where to use it, and how to control it are as critical as where to put it.

Business leaders see a full spectrum of AI-informed data uses, Suresh Renganathan, chief technology officer at Teachers Federal Credit Union says.

"All leaders would like to see a data-driven business model. Data's a fuel to propel business growth right now," he says. "We want to accomplish personalized experience. We want to enable fast investment decisions. We want to predict and reduce delinquencies, and to track branch performances against objectives. There is a plethora of use cases."

Data is a new currency.

It gains value when it circulates. Companies that import data and share their own data more extensively achieve better financial results and show greater progress toward operating AI at enterprise scale – a critical goal for three out of four companies in our survey.

Companies recognize that the emerging data economy holds great potential. But as recently as last year, only about one-third had taken steps to start collaborating with partners.[12]

Building the capability and the will to share data between customers and suppliers could drive tremendous value in the manufacturing space, says Priya Almelkar, vice president of IT manufacturing operations at North Carolina semiconductor firm Wolfspeed.

"Data becomes the new currency because that's how you're able to add more revenue from those shared insights," she says. "It's a new culture for semiconductor makers to get to the point where they feel comfortable sharing their data across. From a manufacturing background, that has been a big no-no, I think now companies are opening up to be willing to share data."

Data is gold yet to be mined.

Data is a mountain of material. Below the topsoil and above the bedrock is a mix of minerals varying in value. The challenge for the data and AI team is to locate and mine the gold, and leave the lead.

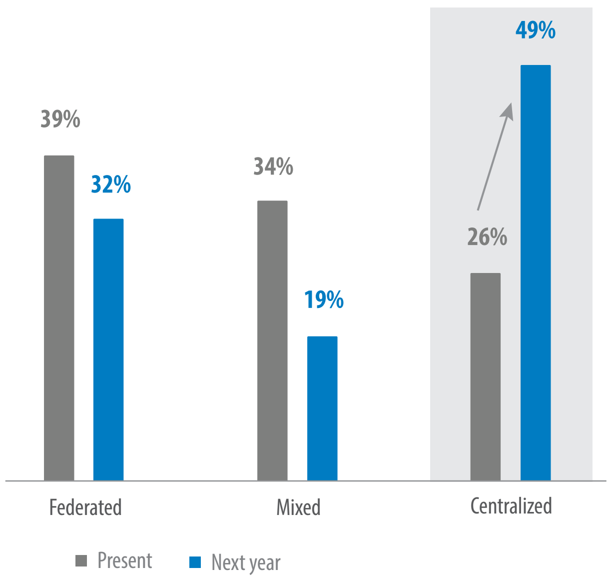
Like the unmined reservoir, data doesn't have any intrinsic value, until you know what's in it, says Sameli Mäenpää, chief data officer of OP Financial in Finland.

"When you know what's in it, you can derive assets out of it," he says. "It only has value if you can connect it to something real."

Ever-evolving methods of data management

Companies need a clear and comprehensive data strategy to manage data properly and ingest new data smoothly. The trouble is, most companies don't have a consistent data management strategy. Respondents tell us they want to manage data centrally, but this is not what most do right now. Our analysis shows that centralized data management links to better profit and revenue growth. However, a shift to fully federated data management also increases profit growth.

Figure 7. Companies moving to centralized data management



If companies (and their technology partners) can execute, centralized data management will be the most common strategy for big businesses in 2023 (Figure 7). But this doesn't necessarily mean it's the best choice. In fact, our analysis found that most AI practitioners who currently have centralized data management intend to change strategies, with the biggest portion shifting to federated approaches. Figure 8 shows this flip-flop from centralized to federated, and vice versa. Most organizations have yet to settle on their preferred strategy, and in this immature phase, the market is yo-yo-ing between the two extremes.

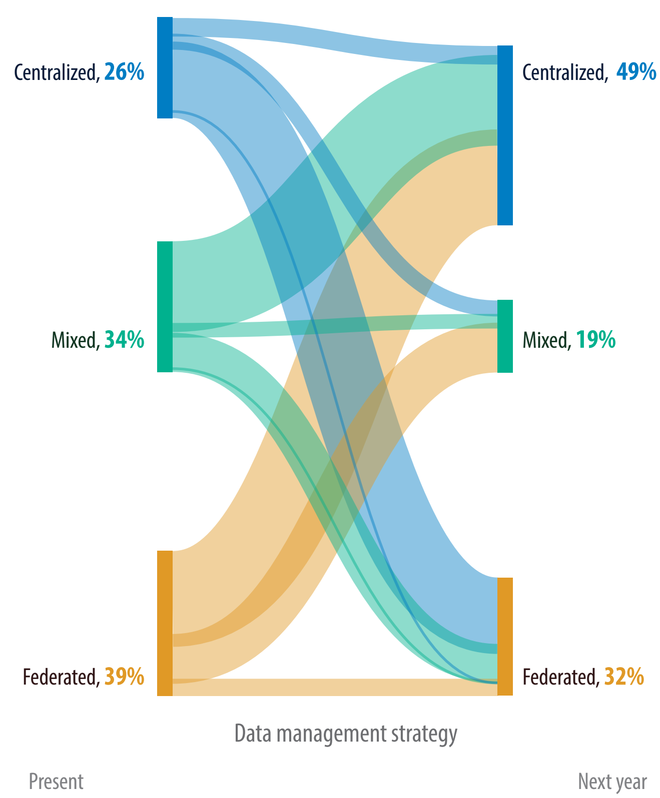
The reality is that these two extremes are too simple to adequately serve as a comprehensive corporate data strategy. Companies deal with so much data in so many sources for so many uses, a one-size-fits-all solution does not work. And yet a data warehouse, data estate, data lake, or data lakehouse without a central organizational authority would leave an organization at risk of a metaphorical data meltdown. Over time we expect more organizations to find a sophisticated balance of centralization and federation – a middle ground that suits their context and needs.

Raj Savoor, AT&T Labs vice president of network analysts and automation, shares the story of how AT&T organized its big data and then made it accessible for AI uses. The US telecommunications giant first invested heavily in big data management capabilities, and its chief data officer put extra effort in establishing a democratized ecosystem where data and AI capabilities can be put to work, he says.

"There's a step function here in complexity as the amount of data increases… we get a kind of finer grain visibility and we have a lot more intelligent controls to then apply decisions," Savoor told Laurel Ruma with MIT Technology Review's Business Lab podcast.[13]

This structure has allowed AT&T to cultivate many AI use cases ranging from internal planning to customer support and threat detection. It also has led the company to better use feedback loops to optimize its models and develop additional use cases, Savoor says.

Figure 8. Strategy in flux: No clear pattern emerges



Data sharing delivers a dual benefit

Data in the 21st century is not a scarce, nonrenewable asset. Savvy businesses know that establishing a data-sharing ecosystem with partners and peers delivers greater benefits than a solitary data lake or warehouse. Siera.AI's Agarwal says his customers have grown to understand that sharing data with Siera benefits the borrower and the lender.

"Any company that needs to build a good AI needs data for the system to train and get better. Customers these days are asking us, 'What can you do with our data? Can you give us custom AI models?'" he says. Siera's best and toughest customers are the ones who push the company to engage in the most vigorous data sharing.

"They give you access to their people, they give you access to their resources, they give you access to data, they give you feedback to give you insights."

"We've had a customer who has come down to our facilities and spent four days with us, testing the equipment and giving us different test cases and scenarios, throwing curveballs at us. The ones that help us make the technology better, they are the early adopters."

Inbound data sharing and outbound data sharing give companies new ways to ensure they have the right information for their data scientists and AI models. All else being equal, importing data from third parties and high levels of data sharing delivered bigger boosts to the corporate bottom line than any other data or AI action. Of the $467 billion in global profit increase available, $105 billion of that links to importing 75% or more of data from third parties, our analysis shows.

Companies that have built a foundation to trust and share their own data can be more agile in their AI work and using AI at scale more readily, says Satish H.C., Infosys executive vice president and co-head delivery. Businesses that lack that can't clearly see the value of their data or trust it. Companies that don't trust their data risk getting caught in a vicious cycle of experimenting and side work preparing and cleaning data for a specific task.

"Then you're working very hard and only able to use data and AI to solve small problems," says Satish.

Effective AI requires data to be fresh and clean

Data and AI work best with fresh data. Companies that refresh the data in their AI models in near real time or at minimum every six months after launching a model achieve better financial outcomes. Our respondents are doing well on this front. Some 79% of respondents refresh data in real time or less than every six months.

But for all that progress on keeping data fresh, companies do not have clean data – Data verification was one of the top challenges facing analytics and AI teams, along with AI infrastructure and compute resources (Figure 9).

In other words, data scientists are needed to do as much data cleaning as they do science. Jacqueline De Rojas, president of techUK, a national trade group said that low-quality data emitted from legacy systems is the most hidden and important challenge for driving AI adoption.

"The power of algorithms has to be driven on clean data," says de Rojas.[14]

Northwestern University Professor and AI expert Mohanbir Sawhney said AI in business is at a critical juncture. In the way e-commerce and commerce were once two separate things that inevitably united into an omnichannel, data analysis has entered AI modeling's field of gravity. And yet, the challenge of data verification persists, regardless of experience (Figure 10) and even though cloud-based AI systems have made AI computing power more available and less costly. Other major challenges, such as bias management and framing AI questions, can be managed with experience. Time and experience also improve companies' abilities to manage biases. AI practitioners at companies that have been implementing AI for more than five years identify bias management as less of a problem than companies that are newer to AI.

Figure 9. Data verification and AI infrastructure are top challenges

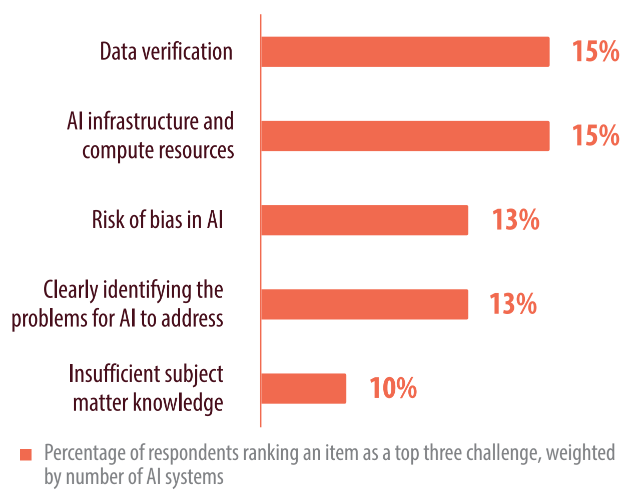
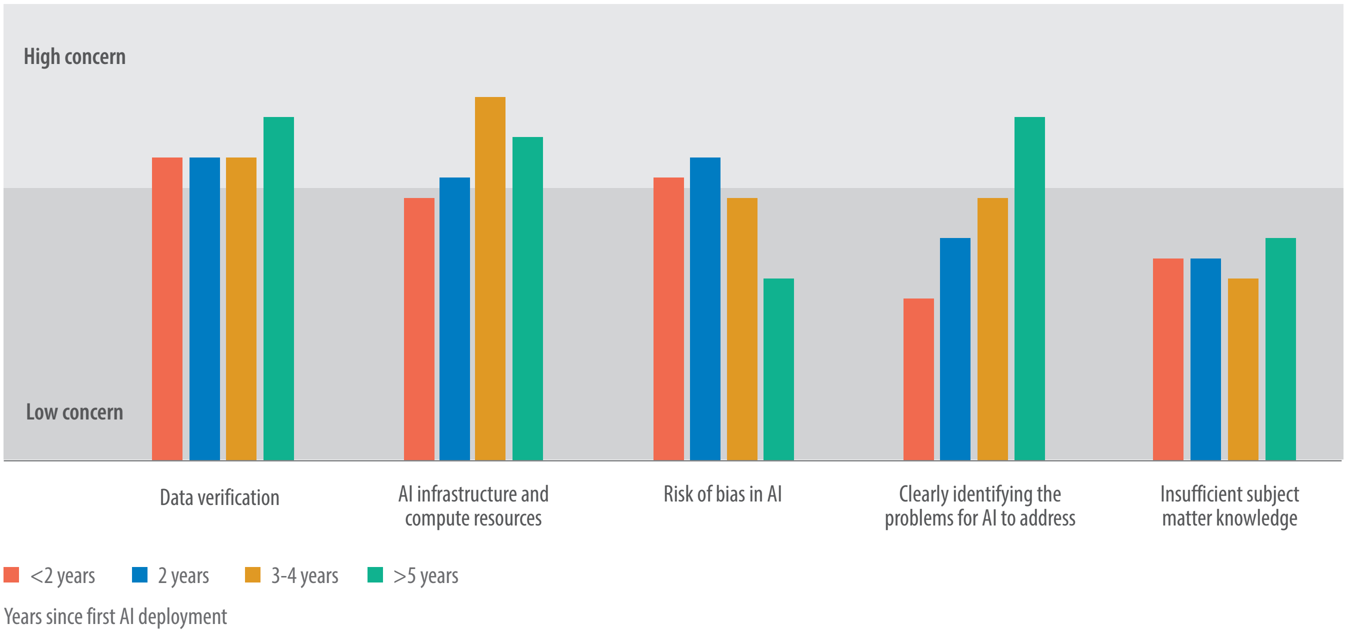


Figure 10. Data verification is a persistently high concern



Exploratory data analysis and data cleaning take time. Companies can't build a model until they know and trust that data is clean and accurate. It's common to spend two months exploring and cleaning data for every one month experimenting with the model, our experts tell us. "Thirty-40% of the time, and 70% of the effort, is in data discovery, preparation, and augmentation," says Karthik Andhiyur Nagarajan, industry principal and data expert at Infosys.

Data and AI return value only once a model is in the field and doing work – but the model must use clean data.

This has companies more focused on data governance, says Andrew Duncan, CEO and managing partner at Infosys Consulting.

"Corporate clients view the ability to collect, organize and analyze clean data as a new differentiator," explains Duncan. "Companies must take a forensic approach to real-time data capture across all forms and formats – both internally and externally – since this will form the backbone of all business units."

"To be more democratic about data and AI decisions, you must be more dictatorial about your data."

– Mohanbir Sawhney

McCormick Foundation Chair of Technology, Kellogg School of Management, Northwestern University

Poor processes invite bias

Even with clean data, AI practitioners face additional pitfalls, starting with missing data – an inevitable problem in most data sets. We used a question about how practitioners handle missing data as a proxy to assess data processes. We found that 34% of our survey respondents say they handle missing data in ways that can allow bias to creep into AI models (Figure 11). That is, they simply impute values for missing data based on other values in the data set or they delete observations. Both of these methods can introduce bias.

Most commonly (42% of respondents), companies deal with missing data by studying the data set to see how much the unknowns may vary from the included data. This requires a high degree of clarity in what the missing data ought to represent, says Rajeev Nayar, Infosys vice president and data strategist.

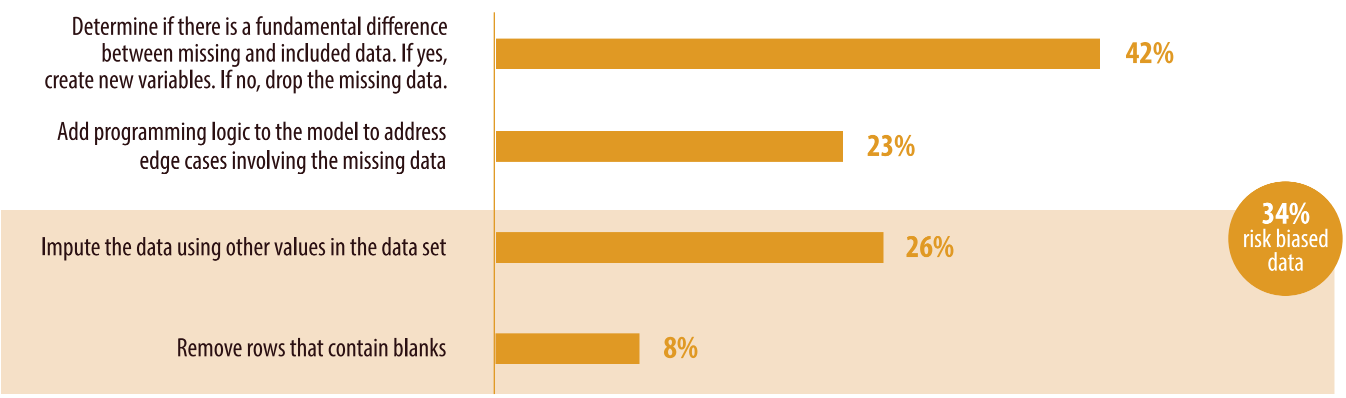
"Anytime you introduce artifacts, there's some kind of problem that you introduce – how do you minimize it?" Nayar says.

Advanced AI brings additional tools, such as generative adversarial networks (GANS) and neural networks, that can shift this work from a slow, manual data science task to an exercise in AI itself. Adding programming logic to address edge cases is less popular, but holds more promise, than heavy data study and scrubbing, says Gary Bhattacharjee, Infosys data strategist.

At times, getting data processes right is something companies cannot control, because of regulations or business considerations.

AI practitioners from the highly regulated healthcare and financial services industries were two to three times more likely to simply delete data from data sets as a way to manage incomplete data. Because they deal in sensitive personal information with heavy regulations, many companies in those industries don't want to leverage sensitive data in AI models or for alternative uses, our experts tell us.

Figure 11. One in three companies use data processes that increase the risk of AI bias



Advanced AI requires trust across all dimensions

Grocery businesses operate on thin margins. So when a US grocer had a chance to save more than $4 million annually by automating work schedules, a well-defined process, it looked like an obvious move. The trouble was, automating this process would have shifted scheduling work shifts from human hands to a machine learning system. The leader of the business unit was against it, and employee union rules didn't immediately align with the software tools that the AI would use to put it in place. In the end, the grocer never used the system, and scheduling remained a 40-hour-a-week job conducted by humans.

The best data and the most elegant AI model amount to nothing if humans do not believe the system is effective, fair, and adaptable. Otherwise, AI systems end up like the grocery store scheduler: developed but unused. The most valuable and most used AI systems instill trust as they operate, Bohannon says. "When you get it right, you'll find that explainability and trust is baked in," he says.

Our survey results support this. Companies that develop strong ethics (Figure 12, top) and bias management capabilities (Figure 12, bottom) report higher satisfaction and trust for their data and AI use cases.

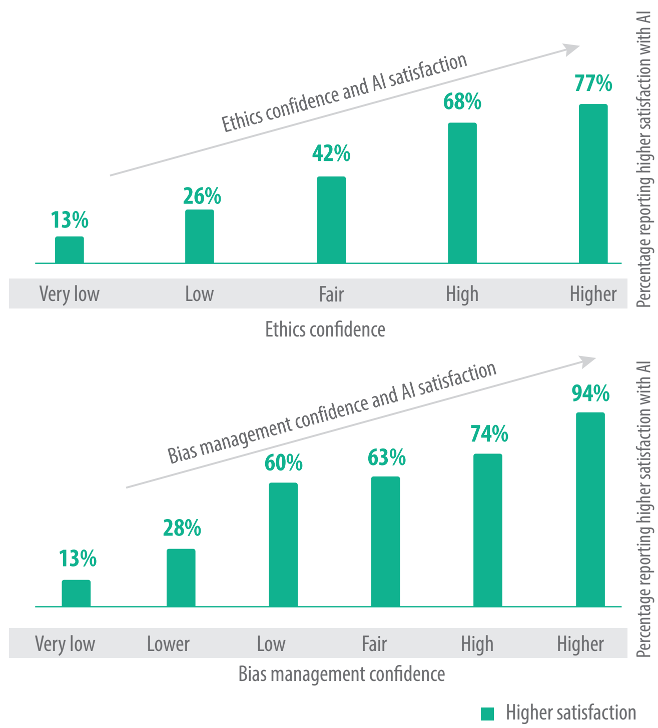
We measured AI ethics and bias management capabilities in 12 ways. Companies with high confidence in these measures are more likely to be satisfied with data and AI. This holds for every measure of ethics and bias control that we studied.

Enterprises that have only recently deployed their first AI models can start by cataloging the usage of AI across the business, said Adriano Koshiyama, a research fellow in computer science at University College London and co-founder of Holistic AI Inc., a London-based AI risk management firm.

"A simple strategy is to peg the cataloging system to other widely used and adopted processes, such as privacy risk assessments or information security reviews," he says. "The next step is to start risk-assessing in a way that a priority list can emerge on what the focus of the enterprise should be."

Over time, robust AI risk management practices will be a substantial component of new enterprise risk management processes, Koshiyama says.

Figure 12. Strong ethics and bias management correlate with greater trust in AI



Five measures of bias management

1. Sampling bias
2. Participation bias
3. Convergence bias
4. Measurement bias
5. Bias derived from overfitting the model

Seven measures of AI ethics

1. Understandable models
2. Clear, useful outputs
3. Explainable algorithms
4. Processes design to systematically detect bias
5. Active efforts for humans to report bias
6. Clear data provenance
7. Data stewardship including routine ethics, compliance and privacy reviews

Trust is a blind spot

Trust is important for simple automation-grade AI, and even more critical for advanced AI, where calculations are too fast or complex for humans to quickly understand.

Primer.AI's Bohannon calls this explainability.

"To get an AI system to the level of performance where you even have a shot at them performing well enough to be trusted, you have systems that help you understand," he says. "There's an explainability all built in around it. That's the most sophisticated version of an AI system. It's very hard to achieve and won't work for all tasks today."

Where will it work? Bohannan says, in cases where AI practitioners have achieved very high data quality that is properly formatted, large in volume, sufficiently diverse, and specifically instructive for the problem AI is aiming to address. This final "instructiveness" dimension of data quality is the most difficult to attain, and requires business knowledge along with AI and data science acumen, he says.

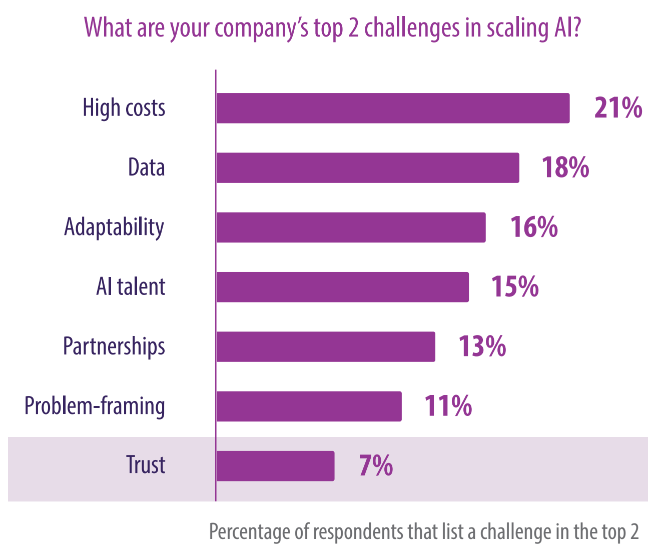
Autonomous AI and self-training AI are at the core of the corporate dream of enterprise-wide AI systems. Nearly three out of every four AI practitioners surveyed want to scale AI across their enterprise. But only 7% say trust is a top challenge in scaling AI. Instead, they most frequently rated the business reality of managing costs (21%) as the top challenge (Figure 13).

With economic uncertainty looming[15], controlling costs are likely to continue to remain atop the list of corporate concerns. The challenge for companies focused on cost is to not neglect trust as they engage more with AI.

Trust is the next big horizon in implementing AI systems, says Bonnie Holub, a data science leader with Infosys Consulting. Teams have been struggling to simply get AI systems implemented at all, let alone efficiently, but best-in-breed companies are industrializing AI systems at scale now, she says.

"At that point, trust and responsible AI systems become a major issue. We see trust and responsible AI as crucial parts of the non-financial governance issues investors are demanding from companies," Holub says.

Figure 13. Despite its importance, trust rated a low concern by executives



The where, what, and who of AI

Before businesses can scale AI, they must make some smart choices about deploying it. Our analysis shows that focusing AI on customer experience and core operations can contribute to profit and revenue growth. In capability terms, the middle tiers (Understand and Respond) link to profit increases, while the most basic tier (Sense) and most advanced tier (Evolve) do not at this point deliver business value.

Adding systems at the Understand and Respond levels has a smaller impact on business outcomes than data-sharing behaviors. But these middle tiers represent a pivotal point in the maturity of an enterprise's AI efforts.

Siera.AI offers two types of autonomous AI systems for use with forklifts and industrial environments. On the basic end are tools that prevent collisions. The higher end includes systems that study how things get moved and looks for ways to optimize warehouse operations.

Agarwal says company managers get excited about the optimization and analysis. But then they hear a different story from the workers closest to the systems.

"The people on the ground level feel like there's a computer watching them and there's a tattletale," he says. "I realized that what they really wanted was something between the two."

Perhaps that system could eventually add more analysis and autonomy down the road, but more immediately warehouse managers and ground-level employees need tools that enhance safety and make their people feel comfortable adopting it.

"I think that's been the biggest, aha, moment. This is not about the sexiest technology that the people at the top want. There's a middle ground, where technology is advanced, but simple enough so that people at the top and the bottom can meet in the middle," Agarwal says.

Holub, editor of Infosys Consulting's AI Journal said top performers are able to differentiate themselves by implementing managed, value-oriented AI project development pipelines.

"They evaluate the potential ROI, rapidly prototype to prove the concepts, and use disciplined data engineering processes," she says.

Disciplined data engineering spans data cleansing, ModelOps for model maintenance and retraining, and system monitoring.

"That complex orchestration is required to industrialize AI at speed and at scale," Holub says.

Deep learning helps achieve AI at scale

Deep learning – an advanced form of AI that imitates how humans learn – holds great promise for advanced AI capabilities and delivering AI at enterprise scale. Infosys Tech Compass Research found that companies can move out of the lowest tier of AI maturity when they integrate deep learning or related advanced algorithms into more than 30% of their AI systems.[16] Data+AI Radar found evidence that this benchmark also links to good business results and drives companies toward AI at scale.

Embracing deep learning draws enterprises closer to AI at enterprise scale, because the two share common technical and data requirements. First, both rely on scalable computing power. Overall, our survey respondents are evenly split on where they prefer to host third-party AI solutions, four out of five companies engaged in deep learning prefer cloud-based AI tools to on-premises AI tools.

Deep learning and AI at scale both also require massive volumes of data. Some 70% of companies heavily engaged in deep learning also extensively share data, our research found (Figure 14).

Most respondents in our survey are on this path – Two-thirds use deep learning in more than 30% of systems (Figure 15); most of that group (73%) share data extensively and most (79%) use cloud for AI (Figure 16). Companies in different industries showed more variation in data and AI practices, capability and satisfaction. On a region-to-region basis, appetite for cloud and levels of advance AI adoption showed some distinction. AI practitioners in Australia, New Zealand, France and Germany showed a slight preference for third-party AI tools in the cloud, and a greater percentage of deep learning in their AI systems. To a large extent, deep learning and other forms of advanced algorithms are what corporate executives mean when they talk about AI today. That's led to a gap between the reality of contemporary AI and the expectation of what companies think they can easily achieve, Agarwal says.

"When you give customers an AI enabled product, they expect it to basically do magic and suddenly solve all their problems," he says. "The AI is there to make humans smarter and safer and better, and that's what it can do practically."

The misinformed notion that AI is practical magic is starting to change, he said, as more managers and IT staffers grow familiar with what AI can practically do.

"There is more excitement from people familiar with AI. They understand that it is not a silver bullet," he says.

Figure 14. Deep learning correlates with extensive data sharing

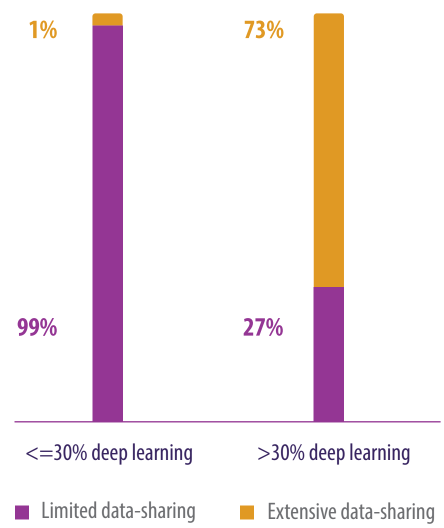
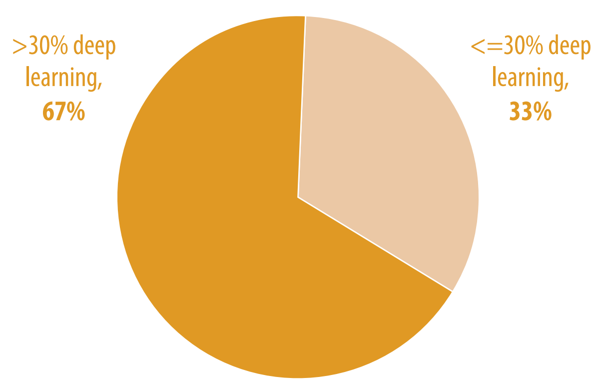


Figure 15. Two-thirds of respondents use deep learning in 30% or more of their AI systems



Proportion using deep learning in more than 30% of systems

What's at stake: Data+AI must prove it's nothing to fear

"Due to their significant influence on society, large enterprises have a distinct responsibility to deliver explainable, responsible AI. By explaining their data and AI outputs with transparency, stakeholders (including customers) are more likely to trust their intent, as well as their products and services."

– Nandan Nilekani
Chairman, Infosys

This is the state of AI in the corporate world: It's everywhere and not often a differentiator. Most AI practitioners are not satisfied with or trusting of AI outcomes – and neither is the public.[17]

Advanced AI requires a new way of thinking about technology and business. It requires trust: Trust in your own and others' data management, and trust in AI models. Clear data processes, strong ethics and bias management build trust. The right choices of where to apply AI and the right mix of advanced AI and AI-cloud systems can reinforce and improve AI results, for both companies and the greater good.

Around the globe, governments are trying to guard against data misuse and autonomous AI running amok. China and California have followed Europe's General Data Protection Regulation with their own data-privacy rules.[18] Business groups and consumer advocates in the US want to see federal data privacy legislation enacted.[19]

Both the US and Europe have regulatory sights set on AI. European leaders have been debating the proposed EU AI Act for more than 12 months.[20] On Oct. 4, 2022, the White House Office of Science and Technology Policy (OSTP) published a Blueprint for an AI Bill of Rights.[21] The blueprint covers five big areas: safe and effective systems; algorithmic discrimination protections; data privacy; notice and explanation; and human alternatives, consideration, and fallback. Large enterprises must demonstrate that they can get value from data and AI while upholding ethical and legal standards. If they don't, they face increased compliance and reputational risks.

"Explaining the outcomes of the AI will continue to require human ingenuity and the inclusion of a greater diversity of perspectives in training algorithms."

– Sunil Senan
Senior vice president, data and analytics, Infosys

Recommendations: Precision, trust, business value

Reduced to their simplest forms, data is quantities and AI is math. Data+AI is a formula written to solve for business value. Advanced AI requires enterprises to be precise in their numbers, calculations, and equations.

Data must be true, accessible, and shareable. Advanced AI requires zettabytes of data and the wisdom to understand which bits and bytes are most instructive for the present business problem. The wrong digits and right computation will not yield an accurate, trustworthy result.

Our study found that AI models must be informed by bias management and ethical AI practices, and they must bring explanation. This is critical to build trust and ensure that AI systems get used. If you can't show your work, you can't persuade skeptics to believe your solution.

AI teams must clearly frame and explain both their definition of business value and choice of variables used in an AI model (or formula). This requires the perspectives of data scientists, industry specialists, and business executives. A model or formula developed without business value can only deliver a solution in search of a problem.

Companies that think differently about data and AI will get the most business value out of AI. Here's how:

1. **Get your data right, and share it.** Focus on data sharing capabilities and hub-and-spoke data management.
2. **Build trust in advanced AI.** Strengthen ethics, bias management, deep learning, AI cloud, and scaling across the organization.
3. **Compose an AI team biased to business value.** Business leaders matter as much as data scientists.

"You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you."

– A Blueprint for an AI Bill of Rights
The White House Office of Science and Technology Policy

1. Get your data right, and share it

Get your data right, and share it. Focus on data-sharing capabilities and hub-and-spoke data management.

Before putting any AI into production, companies must have accurate, organized data. This has long been true, but the method for achieving that has evolved.

The old process – extract, transform, and load data into a private warehouse – faced limits. Followers of that procedure can only apply AI to the data contained in the four walls of their warehouse. Data management strategies that foster data sharing, both importing in and sharing out, expand the universe of available data. Complex data and AI capabilities should leverage the best features of centralized and federated data management strategies, Infosys data strategist Nayar says.

"The best way to get control of your data is to centralize it. But as companies move toward AI-driven initiatives, they're finding centralization is not the answer because the most pertinent data is all over the place," he says.

Rafee Tarafdar, Infosys chief technology officer, recommends "a hub-and-spoke strategy where companies centralize platform and technology but give teams flexibility to operate on their own."

This approach takes the best from centralized and federated and melds data management strategies. The hub is a common repository that defines the data and its location. The spokes, figuratively radiating from the hub, contain and share data as dictated by business specialists and AI teams. Companies are increasingly establishing a hub as a common tech layer that organizes data and connecting spokes that empower business units to manage their own data .

"Business units are saying, Data and AI is my asset," Nayar says.

Sameli Mäenpää, the chief data officer (CDO) at Finland banking and insurance firm OP Financial, embraces this sort of strategy. He describes his CDO role as more of a library architect than a librarian.

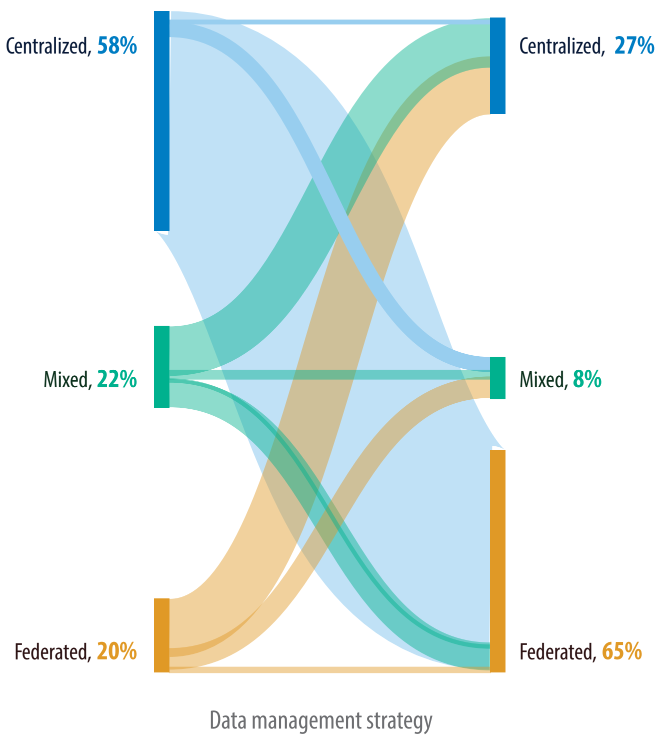
"We provide the library building, electricity, water, and the shelves for books. We do not take responsibility for which books are on the shelves. We do not take responsibility for indexing the books. We provide the indexing service and solution, but we do not do the actual indexing," he says. "If the books are worn out or quality is bad, it's the business owner who needs to fix that. The data owner is responsible for that quality and they also need to know who they loan the books to."

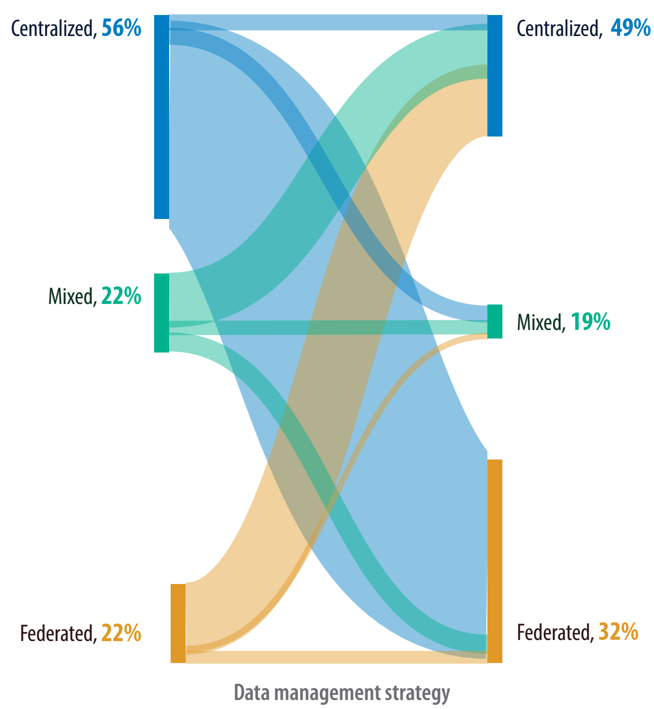
Mäenpää rejects the notion that the CDO should fix and clean all the company's data. Rather, each individual business unit should own their own data, with the data unit providing data warehouse infrastructure and a cloud-based data-science workspace.

Centralized structuring of data followed by distributed data analysis is a hallmark of companies engaged in extensive data sharing, our research shows. More than half of companies in our highest tier of data sharing are pursuing a centralized-to-federated data management trajectory. Technology and insurance companies survey share data the most, and they are also the two industries that showed the strongest preference for advancing from centralized to federated data management (Figure 16).

Think of changes in data management strategies as a maturity curve or a cycle, Nayar says. Data that's not organized must be centralized. Once that's established, a shift to federated data analysis brings business executives into the data+AI equation.

Figure 16. Insurance and technology industries show strong preference to move to federated than do other industries





"A hub-and-spoke strategy where companies centralize platform and technology but give teams flexibility to operate on their own."

– Rafee Tarafdar

Chief technology officer, Infosys

2. Build trust in advanced AI

"Companies have a lot of value in their data that they have not unlocked. This is leading to a realization that they cannot approach data and AI as individual use cases or projects. They must establish an enterprise-wide strategy to discover, democratize and de-risk AI implementation at scale."

– Balakrishna DR
Executive vice president and head of AI and automation, Infosys

Build trust in advanced AI. Strengthen ethics, deep learning, AI cloud, and scaling across the organization.

When employees trust that their AI systems are operated responsibly, they are more likely to work with AI outputs, experts say. This grows increasingly critical as companies put more advanced AI systems to work.

AI that cannot quickly and clearly be explained ends up unused and leaves the people receiving its outputs dissatisfied. Fortunately, strong ethics and bias management practices can increase trust and satisfaction in data and AI.

Specific advanced AI capabilities and practices can work together to create a virtuous cycle that leads to more advanced AI. Deep learning is an example of this.

"The foundational aspect of doing deep learning is to get tons of different kinds of data, and you can only do that in an environment where data is conveniently shared, and that's on the cloud," Bhattacharjee says.

More AI in cloud correlates with greater data sharing. Expanding deep learning and data sharing each have a positive influence on corporate profits, our analysis shows. Sharper data honed by robust data sharing leads to better outputs from deep learning.

Recall that most of the high-satisfaction data and AI use cases deliver signal detection or automation functions – in essence, narrow solutions. Three-quarters of our survey respondents say they want something more broad: advanced AI at scale. Advanced algorithms such as deep learning, AI cloud, and extensive data sharing are the capabilities guide companies to better business outcomes, and a shot at scale.

3. Compose an AI team biased to business value

"The mission of the AI team is to deliver democratized tools to employees, customers and partners to enable them to develop and use data responsibly, accurately and ethically across their entire ecosystem."

– Andrew Duncan
CEO and managing partner, Infosys Consulting

Compose an AI team biased to business value. Business leaders matter as much as data scientists.

Good AI teams typically involve multiple disciplines. Regardless of the question, three groups of people should always be on the AI team: data scientists, experts in the business problem, and senior executives.

It's obvious that data scientists should be involved in AI work, and our study backs this up. But experts with intimate knowledge of the business problem at hand matter just as much, our analysis shows. Business experts are critical to framing out AI models and identifying the most instructive data.

Including a key subject matter expert has the dual benefit of properly training a new AI model and establishing trust in the organization when the AI enters full-scale operation, says Richard Donaldson, vice president of digital transformation at Duke Energy Corp.

"You're going to let a computer make a decision that a human once made. You're only going to be successful if you include that decision maker on the front end of the process," he says.

Duke included these decision makers when its data team built an algorithm to prioritize work at a group of its generation stations. In the testing phase, people reviewed all the results from the algorithm, and delivered feedback when the model made a mistake and when it prioritized correctly.

"You're giving that feedback into your machine learning and it's actually getting smarter and smarter every day," he said.

Duke's AI teams focus on building trust as they scale AI.

"It's not enough to solve it in one business unit and then everybody else will jump on board. You have to build that trust in every single use case," Donaldson says.

Data+AI doesn't belong only to data scientists. Senior executives play critical roles in keeping AI efforts focused on core strategies. Experts in the business problem guide data scientists and AI systems in properly framing the problem at hand. A diverse, business-focused team keeps Data+AI tethered to business priorities.

Senior executives are critical to scaling AI because of their knowledge of strategic operations. And when the boss is absent from the AI team, it can be costly. Companies that never involved senior executives on their AI teams reported significantly lower profit growth and revenue growth. On average, this was more than 10 percentage points lower than those who often use senior executives on their AI projects.

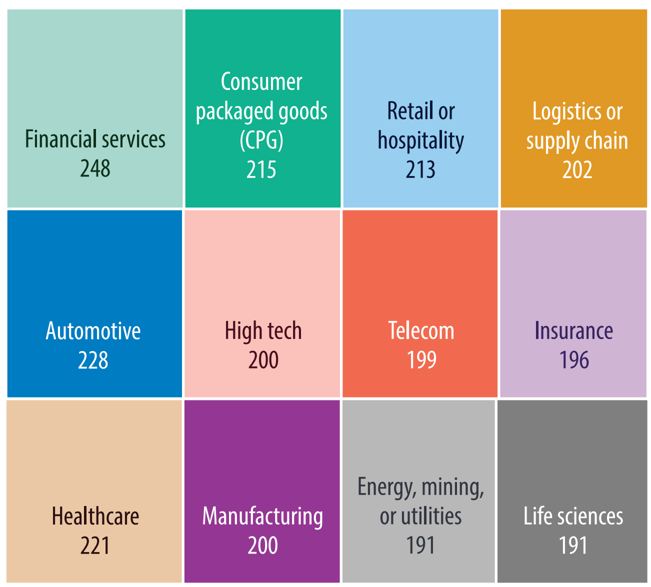
Once again, in their simplest form: Data is numbers. AI is math. Both must be accurate. The AI team must develop formulas to solve two questions simultaneously: "What value do you seek?" and "Who feels the impact of that value?"

Appendix: Research approach

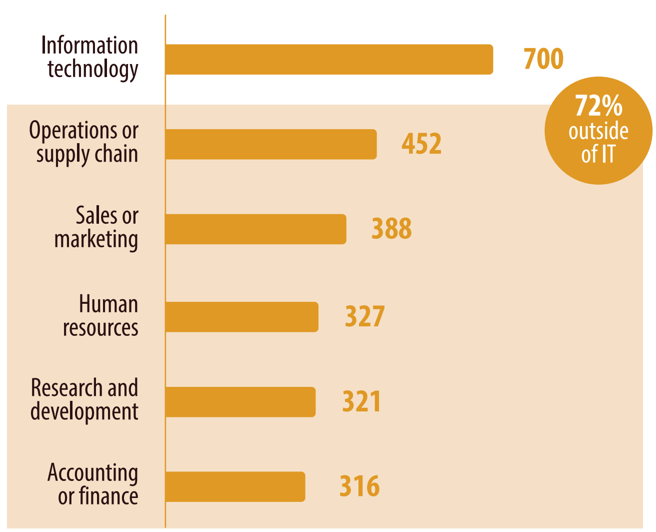
Infosys commissioned an independent third-party survey of 2,504 AI practitioners. In addition to questions about data, AI and technology practices and capabilities, we asked survey respondents for financial details including revenue range and yearly revenue and profit growth rates. The survey was conducted in May to July 2022. It included respondents from companies with more than $500 million in annual revenue in the United States, United Kingdom, Germany, France, Australia and New Zealand.

We identified and analyzed a large set of actions that could affect profit and revenue change related to data and AI. We then set base cases and found via linear regression 23 actions (of 69 analyzed) that showed evidence of a statistically significant impact on profit or revenue growth. The $467 billion in potential profit growth derives from a ~10% increase in profit growth that can be achieved from 13 actions with statistically significant uplifts.

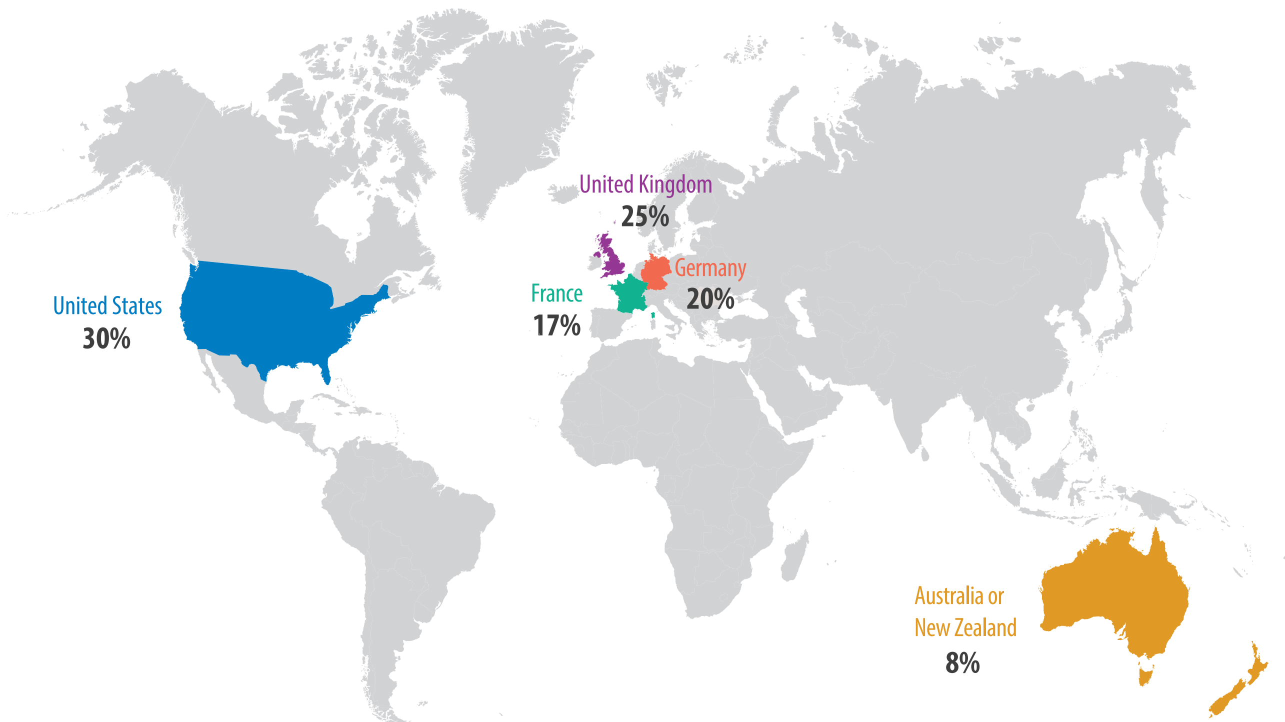
Industry



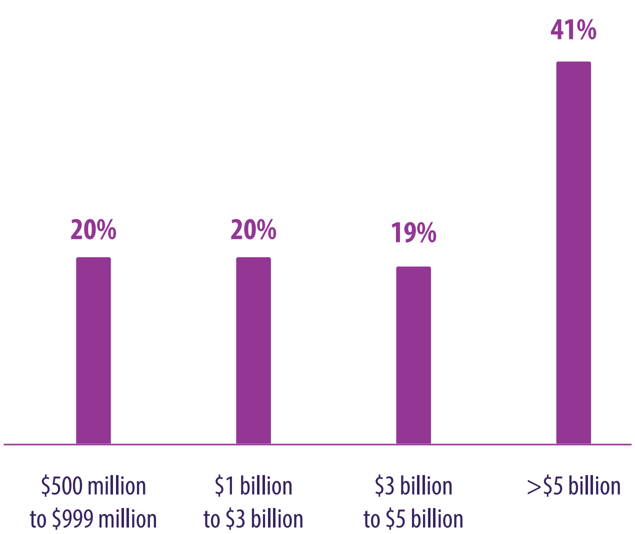
Job function



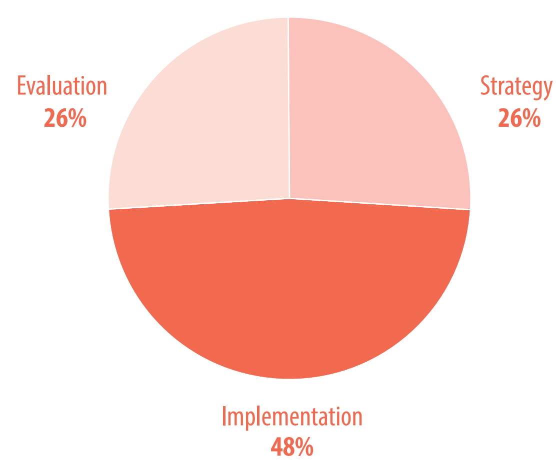
Country



Annual revenue



Role in AI



References

1. Worldwide Spending on AI-Centric Systems Will Pass $300 Billion by 2026, According to IDC, Michael Shirer, Sept. 12, 2022, International Data Corporation.
2. Why the iBuying algorithms failed Zillow, and what it says about the business world's love affair with AI, John Cook, GeekWire, Nov. 3, 2021.
3. Opendoor company financials, Crunchbase.com, accessed Oct. 17, 2022.
4. AI: It's Not Just For the Big FAANG Dogs Anymore, Alex Woodie, Datanami, June 11, 2021.
5. KRTI 4.0 – The Next Generation AI Framework for Operational Excellence, Infosys, May 24, 2019.
6. Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance, Aphra Kerr, Marguerite Barry, and John D. Kelleher, Big Data & Society, April 30, 2020.
7. What are the most important advances in AI?, AI100: A One-Hundred-Year Study on Artificial Intelligence, Stanford University, Sept. 16, 2021.
8. Open Data Science APAC Virtual Conference 2020, Keynote, Shailesh Kumar, Dec. 8, 2020.
9. Automation Journey of Bupa Global, Abhiram Mahajani, Automation Journey of Bupa Global, The Applied AI Podcast, Dec. 22, 2021.
10. The Live Enterprise: Create a Continuously Evolving and Learning Organization, Jeff Kavanaugh and Rafee Tarafdar, Infosys Knowledge Institute, 2021, p. 129.
11. Data Is the New Oil, Clive Humby, ANA Marketing Maestros, November 2006.
12. Capitalizing on the data economy, MIT Technology Review Insights, Nov. 16, 2021.
13. Building tomorrow's telecommunications network today, MIT Technology Review Insights, June 15, 2022.
14. Applied AI: For a diverse, more human enterprise, Mohit Joshi, Infosys Insights.
15. Global Economic Outlook 2023: Global Recession Fears Resurface, Risks Mount, The Conference Board, Oct. 10, 2022.
16. Pervasive Intelligence for a Live Enterprise, Infosys Tech Compass, 2022.
17. Managing expectations of artificial intelligence, Nature, Jeff Hecht, Nov. 28, 2018.
18. Cross-Border Data Transfers: PIPL vs. GDPR vs. CCPA, Allison Kutner, et al., Cooley LLP, April 11, 2022.
19. What Should and Should Not Be Included in a National Privacy Bill, Jordan Crenshaw, U.S. Chamber of Commerce, Sept. 13, 2022.
20. The EU AI Act will have global impact, but a limited Brussels Effect, Alex Engler, Brookings Institute, June 8, 2022,
21. Blueprint for an AI Bill of Rights, White House Office of Science and Technology Policy, Oct. 4, 2022.

Contributors

The Knowledge Institute would like to thank the following people for providing guidance, insight and feedback throughout the Data+AI Radar research project. We are grateful to be surrounded by intelligent professionals in and outside our organization that know the value of sharing data.

Bill Sniffen, Cigna; Richard Donaldson, Duke Energy; Rich Cooper, Fusion Risk Management; Adriano Koshiyama, Holistic AI Inc.; Hajo von Reinhardt, Kao; Dr. Tamara StClaire, Vikas Agrawal, Oracle; Austin Gunter, John Bohannon, Primer AI; Saurav Agarwal, Siera. AI

Aisha Mendez, Akhilesh Mairpady Rao, Alec Boere, Arunkumar Ranganathan, Ashok Panda, Asit Karapurkar, Balakrishna DR, Bonnie Holub, Brent Field, Daniel Farenzena, Deepak P.N., Dr. Ravi Kumar G.V.V., Erik Newlin, Gary Bhattacharjee, Gaurav Bhandari, Gaurav Bansal, Harald Gunia, Jagadamba Krovvidi, Jinu Koshy, Jon Warren, Joseph J. Alenchery, Kamalkumar Rathinasamy, Kamlesh Kumar, Karthik Andhiyur Nagarajan, Madhav Timalsina, Mohammed Rafee Tarafdar, Mythreya Reddy Kotha, Nandha Kumar Muthusrinivasan, Narendar P., Prabhat Kaul, Praveen Gonabal, Pushpak Banerjee, Rajeev Nayar, Rajeshwari Ganesan, Rajnish Sharma, Ramgopal Natarajan, Randy Chambers, Ravishankar Shetty, Ritesh Saluja, Ryan Elmore, Saibal Samaddar, Sandeep Roy, Sankar S.D.V., Satish Doreswamy, Satish H.C., Serge Plata, Seth Siegel, Shashank Mathur, Shatam Bhattacharyya, Soumya Nandy, Srinivas Gopal Sugavanam, Sriram Sundar, Sudarshan Gopalan, Sunil Senan, Syed Ahmed, Ulku Guneysu, Uttam C.N. Ritesh, Vishwanath Taware

Authors

Chad Watt | Infosys Knowledge Institute, Dallas

Jeff Kavanaugh | Infosys Knowledge Institute, Dallas

Analysis and Production

Isaac LaBauve | Infosys Knowledge Institute, Dallas

Jitesh Gera | Infosys Knowledge Institute, Bengaluru

Ryan Mathew | Infosys Consulting, Bengaluru

About Infosys Knowledge Institute

The Infosys Knowledge Institute helps industry leaders develop a deeper understanding of business and technology trends through compelling thought leadership. Our researchers and subject matter experts provide a fact base that aids decision making on critical business and technology issues.

To view our research, visit Infosys Knowledge Institute at infosys.com/IKI or email us at iki@infosys.com.

For more information, contact askus@infosys.com