Understanding AI: business uses and strategies

A brief introduction to AI and neural networks: What they are, how they work, and what you need to be aware of

Artificial intelligence (AI) has emerged as one of the most significant drivers of business transformation in history. Yet its extreme complexity can leave business leaders feeling like outsiders in their own organizations, dependent on a technology they may not fully understand and on others who possess the knowledge and the keys to their organization's success–and perhaps survival.

We believe business leaders can benefit from straightforward, jargon-free explanations of the key concepts behind the AI, neural networks and deep learning technologies that have been making headlines and affecting their business.

What's the difference between AI, machine learning, deep learning and neural networks?

Let's start with some definitions.

As its name implies, **artificial intelligence** attempts to be an artificial version of human intelligence—machines thinking and making decisions like humans.

Machine learning (ML) is a type of AI, where machines learn to identify patterns in datasets to make predictions or decisions or perform complex tasks without having been given explicit instructions. ML systems are dynamic; they have the ability to modify themselves based on the data they are exposed to. Not all AI models involve machine learning—for example, expert systems or rules engines that give predefined responses based on certain keywords.

Deep learning is a category of machine learning algorithms that employ increasingly sophisticated mathematics

powering their ability to detect highly complex patterns in data. Conventional machine learning algorithms have a finite capacity to learn patterns no matter how much data they're given, but deep learning algorithms can improve their performance with access to more data and can model more complicated patterns in data.

Neural networks (NN) are one type of machine learning—a specific approach to getting a model to learn things, recognize patterns, and make predictions or decisions in a humanlike way. Neural network models are designed to mimic the structures in human brains: the network of neurons that receive, store and transmit information that enable the brain to learn and make decisions.

The terms "deep learning" and "neural networks" are often used interchangeably because most deep learning algorithms today rely on neural networks, and most neural networks employ deep learning techniques. The term **deep neural networks** is increasingly being used in recognition of their growing inseparability.

What's the difference between AI and generative AI?

Generative AI is a new type of AI that has garnered a lot of attention in recent years. As its name implies, its focus is on generating new content—primarily text or images—designed to be indistinguishable from human-generated content. All currently popular generative AI models are built on deep neural networks.

There's a common misperception that generative AI is "AI 2.0" or that it has replaced more traditional AI models, but they are really designed for two very different purposes. Put simply, "traditional AI excels at pattern recognition, while generative AI excels at pattern creation. Traditional AI can analyze data and tell you what it sees, but generative AI can use that same data to create something entirely new."[1]

1. Source: "The Difference Between Generative AI And Traditional AI: An Easy Explanation For Anyone," Barnard Marr, Forbes, July 24, 2023.

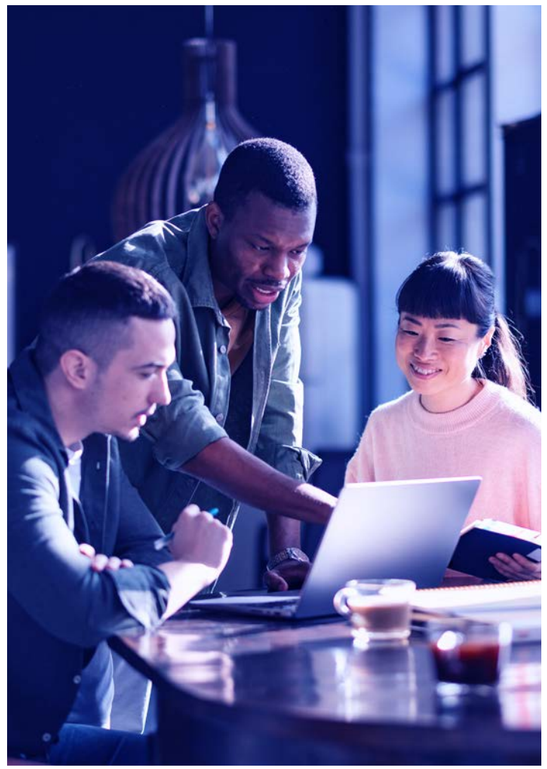
Types of machine learning training

What makes ML unique—and NNs in particular—is that instead of people programming in all the rules that enable them to make accurate predictions or what appear to be intelligent decisions based on novel data (data they have never seen before) these models "learn" the rules by themselves. Therefore, before a model can be used in the real world, it must be **trained** to learn these rules.

There are multiple ways to train machine learning models, depending on the intended use. In the simplest of terms, the biggest difference may be how the model is told it has gotten a "right" answer during training.

Supervised learning

Supervised learning involves giving the model a **training dataset**—a set of examples that reinforce what you want the model to predict (the answer you'd like it to give) when it's ultimately given novel data. The dataset explicitly says what the "ground truth" answer to be predicted is. The supervised machine learning model then learns to recognize patterns in the dataset to make a prediction. The model knows when it has gotten the right answer because the training dataset tells it exactly what the right answer is.



Supervised learning is most often employed for models used for prediction for continuous number values (e.g., regression) and classification (e.g., dividing into specific categories).

With regression, for example, a technology company could use supervised learning to predict how many users it might expect on its web application during a given hour, using a training dataset that included what web traffic looked like previously. With classification, an IT organization could classify ticket requests for open IT issues into categories such as access requests, installation issues, password resets, etc. based on previous tickets IT had received and resolved.

Unsupervised learning

Machine learning models can also employ unsupervised learning: instead of being told what the connection is between the input and expected output, the model recognizes a "right" answer during training by whether or not the data fits a pattern defined by a statistical formula.

There are multiple statistical techniques that can be employed to detect correlations, associations, clusters, and other patterns. In many cases, humans may not know that these patterns exist in the data or can't

easily recognize them on their own. This is one reason this method can be so valuable.

Unsupervised learning models are often used for customer segmentation, such as identifying that customers who share certain characteristics also tend to share similar purchase patterns. For example, the model might identify that older customers who live in a certain geographic region tend to purchase different items than younger customers in another region.

They're also used to make customer recommendations based on complementary purchases. For example, the model might identify that customers who bought certain ski gear were also more likely to buy a winter coat, or that bank customers who opened CDs were also more likely to open a new credit card account.

Semi-supervised learning

Semi-supervised learning, as the name implies, is a mix of supervised and unsupervised learning; some of the training data is labeled and some isn't. This approach can increase the speed and accuracy of unsupervised learning by giving the model some clues about what patterns the model designers think are relevant.

It can also be used for the opposite reason, giving the model designers some clues about what characteristics or variables in a dataset might be relevant and useful for improving supervised learning.

It's often used for anomaly detection such as fraudulent credit card activity (e.g., this transaction doesn't fit within the historical purchase patterns of this user). Some of the indicators of fraud may be well known and can easily be labeled; others may be patterns that are more difficult to recognize or describe by humans.

Reinforcement learning

With reinforcement learning, the model is given a goal and then through trial-and-error finds ways to achieve the goal. The model is "rewarded" when it achieves the goal in the most optimal way. Instead of training data, it uses what it has learned on its previous attempts to improve its decision-making over time.

A recommendation engine, for example, might measure clicks on its recommendations as a measure of its performance—it's rewarded by a click if the user found the recommendation valuable and ignored if the user did not. It can be useful for teaching models how to solve puzzles such as a Rubik's cube or playing games such as chess. The goal is clear and the reward is solving the puzzle or winning the game in the fewest number of steps.

One key difference in this type of learning is that the techniques used are designed to teach even when the feedback the model receives is delayed from its decisions. It won't recognize that the decisions it's making about moves in the chess game were "good" decisions until after the game is complete—and, more realistically, after many games have been played.

Reinforcement learning works well in dynamic environments. Self-driving cars rely on it as do the AI controlled characters that humans play against in video games. It can be used in healthcare to help find treatments that achieve the best results.

Types of neural networks

Feedforward neural networks

The most basic type of neural network is a feedforward neural network (FFN), so called, as you might imagine, because the information is processed linearly, fed in the forward direction only.

One of the key limitations of FNNs is that they don't consider any previous inputs (other than their training data) when they are asked to make a prediction based on the current input. That lack of memory limits their ability to make certain predictions such as what input is likely to come next.

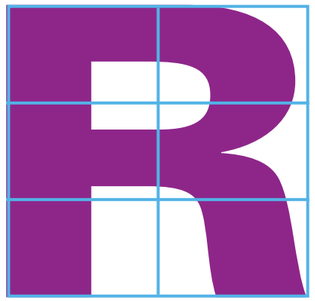
Some example use cases for FNNs:

- Financial institutions may leverage FNNs to help with loan application approvals or determine loan amounts by predicting the likelihood of default based on a set of characteristics.
- The manufacturing sector often leverages FNNs to predict when preventative maintenance should occur to pre-empt failures (e.g., when you see these values, a failure is more likely to occur).
- Retailers use FNNs to help predict foot traffic into brick-and-mortar stores, using a variety of signals from day of week and expected weather to interest and unemployment rates.

In each of these cases, the model will make a decision or prediction based on a given set of conditions and the rules it has learned, not on any set of conditions it had been given previously (other than its training dataset).

Convolutional neural networks

A convolutional neural network (CNN) includes specialized layers designed to detect if a certain feature or pattern is present in the data. That process is called convolution. They specialize in processing data that's organized in a grid—like pixels in an image. Consider this simple grid:



When viewed as a grid, we can clearly make out the letter "R". But if the data is serialized, as it is when image data is fed into a computer like this:

The pattern becomes disjointed and the image is much more difficult to decipher visually.

CNNs are commonly used for applications that involve image analysis:

- Banks leverage image analysis to identify which ATM locations need servicing due to garbage or loitering.
- The manufacturing industry also uses CNNs for images analysis to verify the quality of manufactured items and flag any potential defects; the agricultural industry similarly uses AI in this way to inspect produce.
- CNNs are also used to moderate content by media companies, to help identify and proactively remove harmful content from their platforms.

Recurrent neural networks

A recurrent neural network (RNN) is designed to handle sequential data, such as time sequences or speech.

There are use cases where sequence does matter. The order of words matters, for example, if we want to extract meaning from them. “John ate a fish” has a very different meaning than “A fish ate John” even though they are the very same words.

Unlike an FNN, RNNs include short-term memory. Instead of feeding data forward only, RNNs loop back to consider both the current input and the recent or immediate past input. This helps them recognize sequences, so they’re better suited to handle data where the order matters, and as a result, are better able to predict what comes next. The word suggestions your phone makes when you type out a text message—its prediction of what word you might want to type next—is a good example of an RNN.

Some example use cases of RNNs:

- Finance teams across industries leverage RNNs for time-series forecasting to help predict cash flow, revenue, operating expenses and more.
- Large event locations also use image analysis with RNNs to track movements and foot traffic in crowds to better deploy services to event attendees in real time.
- RNNs are often used for natural language processing tasks such as translation and speech-to-text conversion.

Transformers

Introduced in 2017, a transformer is one of the more recent neural network architectures. It’s arguably one of the more significant evolutions in AI models. Transformers are like extreme versions of RNNs. They are also designed to handle sequential data, but they do it bi-directionally—it’s not just about the current word and the words that came just before it, for example, but everything around it. In short, they learn context and consider the current context when they make predictions. Transformers are the driving force behind generative AI and large language models (LLMs) such as ChatGPT.

The sheer size of these models is also what makes them so impressive. ChatGPT-3 uses 175 billion parameters[2] (the things it remembers) and was trained on the entire crawlable internet—hence the “large” in LLM. This was possible due to several advances in the transformer architecture, and in particular, the concept of attention. Instead of considering everything in its training dataset, a transformer is designed to focus more on certain parts, specifically the parts that appear to be more important based on the commonalities and patterns in the data.

2. Source: “Is Bigger Better? Why The ChatGPT Vs. GPT-3 Vs. GPT-4 ‘Battle’ is just a family chat, “ Aleks Farseev, Forbes, February 17, 2023.

What are the sources of AI risks?

Much has been written about the wide range of risks associated with the use of any AI model, including poor or inaccurate decisions, intellectual property loss, bias and discrimination, privacy concerns, legal and regulatory challenges, and so on. Understanding where and how these risks can be introduced can help to address them.

Training risks

Many risks associated with AI often have their roots in model training.

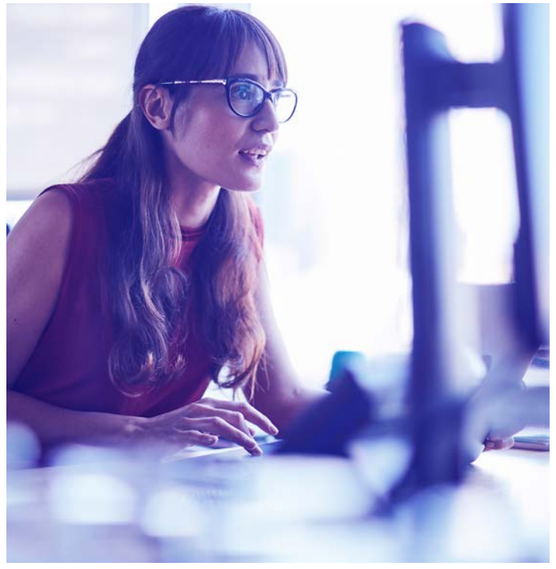
- Non-representative training datasets create the classic "garbage in, garbage out" risk, where the data the model is trained on doesn't accurately reflect or represent the real-world use cases it will encounter after training. The training dataset should be diverse and balanced and include edge cases when possible. Non-representative datasets have been blamed in part for issues with racial or gender bias in some AI systems—for example, facial recognition software that has consistently far greater failure rates in subjects who are female, black, and between 18 and 30 years old.[3]

- Overfitting errors are similar to those caused by non-representative datasets in that the model performs well on the training dataset but fails to generalize when given new or unique data in real-world use. However, the cause is different. Overfitting can occur if the model trains too long or has excess capacity that it uses to memorize the training data (instead of recognizing the underlying patterns in it) or to learn irrelevant patterns in it. For example, an email spam detection model might memorize the specific words or phrases in emails marked as spam during training. When those exact words don't appear in real-world spam emails, it fails to detect them as spam.

- Underfitting errors occur when the model is too simple to recognize the complexity of underlying patterns in the data, or if the training dataset contains too much noise for the model to recognize the relevant patterns.

- Data drift occurs when the distribution of the input data gradually changes over time, so much so that it is significantly different than the data that the model was trained on, causing the machine learning system to no longer perform well.

- Legal, copyright and intellectual property (IP) issues can easily arise during training—and beyond. It's not uncommon, for example, to use data that was "scraped" off the internet as part of training datasets. Unless you have clear ownership of data, IP rights may be violated.



3. Source: "Racial discrimination in face recognition technology," Alex Najibi, Harvard Kenneth C. Griffin Graduate School of Arts and Sciences, October 24, 2020.

- Privacy concerns are a particular challenge with any model designed to learn from or about people. If the data isn't anonymized, you must ensure that consent has been obtained from everyone whose data is used. That can be challenging if the data is coming from public or third- party sources.

Black-box risks

Because NNs are very complex, one of the most significant risks is that they are a black box; even their designers or programmers may not know or understand how they arrive at specific decisions. This lack of transparency can lead to unintended and potentially devastating consequences.

If, for example, an autonomous vehicle strikes a pedestrian when we'd expect it to hit the brakes, the black box nature of the system means we can't trace the system's decision process and see why it made this decision. Was it because it was sunny and a bit foggy, or they've just salted the roads and the asphalt now appears whiter than it usually does? There are an infinite number of permutations so you may never know if the system is robust enough to handle every situation.[4]

People risks

One of the biggest weaknesses of any AI implementation isn't related to the technology or the data it relies on; it's the people who are involved in the process.

At some level, a human must take unstructured data and encode it in such a way as to give meaning to the model and the computer on which it runs. That human element can introduce labeling **errors**—the label for "age" is inadvertently swapped with "height," for example. Errors of this type can also be more subtle and therefore difficult to detect. In some cases, a label used in one dataset might have a different meaning in another—one defines "customer age" by how long they've been a customer, for example, and another defines it as their biological age. It's a relatively common problem. Researchers have found a 3.4 percent average labeling error rate across all datasets.[5]

Even if the data is flawless, a human data scientist must decide which variables or characteristics are important and therefore will be included in the training dataset and ultimately in the production data. For example, a data scientist developing a model to predict the likelihood of loan default for a given applicant—to decide if the lender should offer that person a loan or not—might select age, gender, income, and current address as variables. But what about marital status?

People-related risks aren't limited to actions or decisions by data scientists. Someone decides for what purpose a model will be used, and how much weight its predictions or decisions are given. Someone decides how often models are reviewed for accuracy or to confirm they're delivering their promised value. Someone also must decide how much control and autonomy AI is given.

4. Source: "AI's mysterious 'black box' problem, explained," Samir Rawashdeh, University of Michigan-Dearborn, March 6, 2023.
5. Source: "Major ML datasets have tens of thousands of errors," Adam Conner-Simons. MIT Computer Science and Artificial Intelligence Laboratory, March 29, 2021.

How do you mitigate AI risk?

There may be straightforward ways to overcome some of these risks: increasing training dataset size or adjusting model parameters to help prevent overfitting, or regular model retraining to help prevent data drift, for example.

However, there are no such quick and easy answers for many of the black-box or people-related risks. The magnitude of the challenge and the consequences of a failure require a more holistic approach.

Data modernization

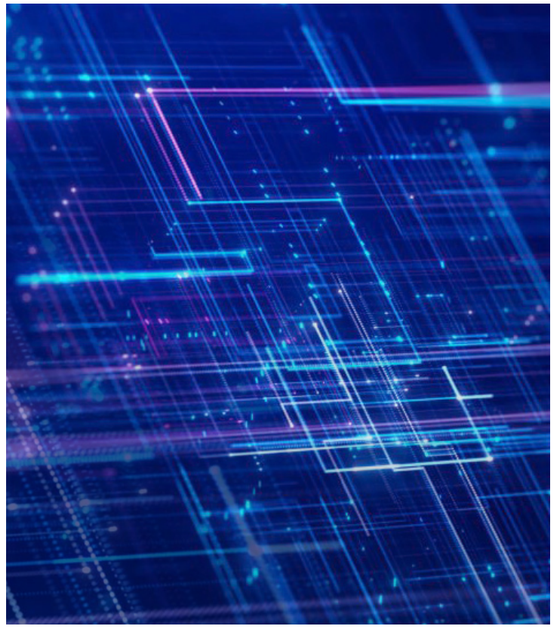
It's not difficult to see why data plays such an important role in AI. As more organizations become more reliant on AI, they also become more data dependent and therefore more data vulnerable. Identifying and addressing risks across the entire data lifecycle becomes an essential part of business. **Data governance**—defining and meeting standards for how data is gathered, stored, protected, used, shared and disposed of—is as vital a capability as any other.

Model validation

Every model and AI solution should be continually and rigorously tested and evaluated for performance to help ensure it delivers results in line with the expectations of both its users and the business. Multiple standard metrics have been developed to help organizations evaluate the performance of AI models depending on their design, training type and purpose. Selecting the appropriate measures is as important as any other AI risk mitigation decision.

Human-in-the-loop

It's essential to identify human-in-the-loop (HITL) points—i.e., places in AI model training, execution, or operation where a human must become involved to help avoid the model making predictions or decisions that could lead to negative impacts on the business. For example, HITL may not be necessary when suggesting complementary purchase items to a customer on a website, but it is critical if your AI system is automating accounting conclusions that would impact financial reporting.



Trusted AI

The practice of responsible or **trusted AI** has been designed to help organizations address the risks and challenges associated with AI and NN models. It provides a framework designed to help organizations implement practices for designing, building, training, deploying and using AI solutions that consider accuracy, trust, fairness, privacy and security at every step.

Using AI to detect and mitigate risk

While AI and ML models are known to introduce risks, perhaps ironically they are also valuable tools for detecting and mitigating them.

For example:

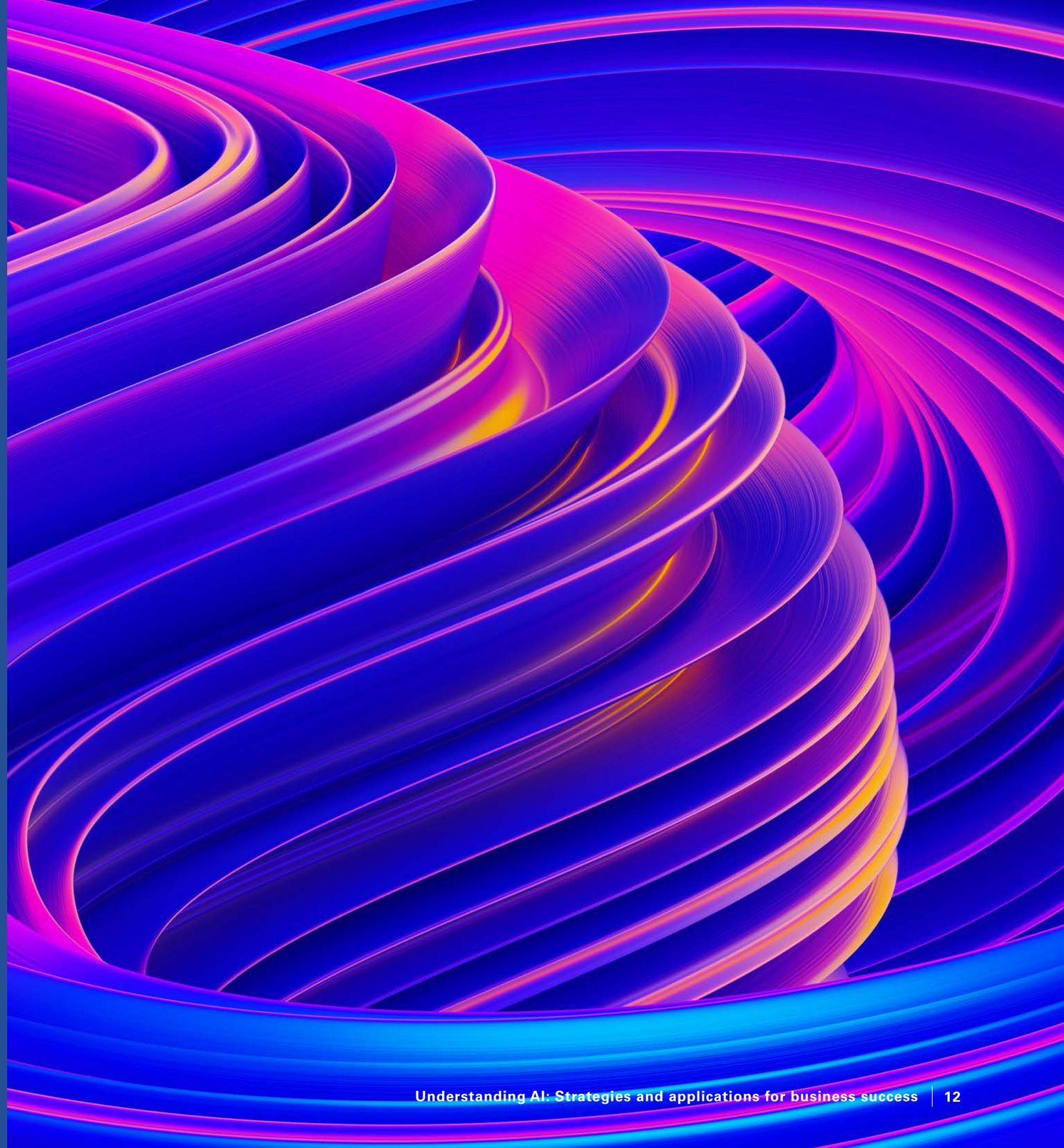
- AI can be used to proactively review invoices and payment processing, enabling companies to automatically examine 100 percent of transactions for all suppliers. This can help detect fraud, prevent rate discrepancies between invoiced amount and master rates, identify unrealized benefits such as early payment or volume discounts, and highlight new preferred suppliers to enhance contract negotiations.
- AI can help automatically analyze contracts to identify outdated, altered, or missing contract clauses that expose the organization to risk. The analysis can identify the issue with a specific contract clause and make suggestions for improved contract language to mitigate the risk.
- ML models for identity and access management (IAM) can automate developer recertifications to prevent unauthorized access to internal resources, or automatically assign suspected vulnerabilities to security teams for remediation, saving both time and costs.

How KPMG can help

Arguably more than any other technology, AI requires an incredibly broad range of skills and experience to implement safely and securely while helping to maximize value.

AI requires a deep understanding of data analytics and technology skills to design, build, and implement. But it also requires a detailed understanding of the business and the use cases for which it's applied, as well as risk, legal, cybersecurity, and organizational change management skills to help ensure models achieve business goals and deliver value without introducing unacceptable risks.

At KPMG, we bring together technology and business professionals with the necessary skills and experience to help organizations harness the power of AI in an effective, secure, and responsible way. We can help identify use cases and design and build AI models designed to deliver business value. We can help increase awareness of AI risks, improve collaboration between AI and cyber teams, improve tooling for visibility and protection, and help build a foundation that can allow your organization to stay ahead of the AI innovation wave.



Contact us

Kanika Saraiya Havelia

Director
KPMG Consulting
KPMG US

khavelia@kpmg.com

Rachel Wagner-Kaiser

Director and KPMG
Data Scientist
KPMG US

rwagnerkaiser@kpmg.com

Sreekar Krishna

Principal
KPMG Advisory
KPMG US

sreekarkrishna@kpmg.com