

McKinsey on Payments Special Edition on Advanced Analytics in Banking



The analytics-enabled collections model



How machine learning can improve pricing performance



Combating payments fraud and enhancing customer experience



Using data to unlock the potential of an SME and mid-corporate franchise



Hidden figures: The quiet discipline of managing people using data



Using analytics to increase satisfaction, efficiency, and revenue in customer service



Designing a data transformation that delivers value right from the start



Building an effective analytics organization



"All in the mind": Harnessing psychology and analytics to counter bias and reduce risk



Mapping AI techniques to problem types



Data sheet: Advanced analytics



Mapping AI techniques to problem types

As artificial intelligence technologies advance, so does the definition of which techniques constitute AI (see sidebar, “Deep learning’s origins and pioneers”).¹ For the purposes of this paper, we use AI as shorthand specifically to refer to deep learning techniques that use artificial neural networks. In this section, we define a range of AI and advanced analytics techniques as well as key problem types to which these techniques can be applied.

Michael Chui
James Manyika
Mehdi Miremadi
Nicolaus Henke
Rita Chung
Pieter Nel
Sankalp Malhotra

Neural networks and other machine learning techniques

We looked at the value potential of a range of analytics techniques. The focus of our research was on methods using artificial neural networks for deep learning, which we collectively refer to as AI in this paper, understanding that in different times and contexts, other techniques can and have been included in AI. We also examined other machine learning techniques and traditional analytics techniques (Exhibit 1, page 81). We focused on specific potential applications of AI in business and the public sector (sometimes described as “artificial narrow AI”) rather than the longer-term possibility of an “artificial general intelligence” that could potentially perform any intellectual task a human being is capable of.

Neural networks are a subset of machine learning techniques. Essentially, they are AI systems based on simulating connected “neural units,” loosely modeling the way that neurons interact in the brain. Computational models inspired by neural connections

have been studied since the 1940s and have returned to prominence as computer processing power has increased and large training data sets have been used to successfully analyze input data such as images, video, and speech. AI practitioners refer to these techniques as “deep learning,” since neural networks have many (“deep”) layers of simulated interconnected neurons. Before deep learning, neural networks often had only three to five layers and dozens of neurons; deep learning networks can have seven to ten or more layers, with simulated neurons numbering into the millions.

In this paper, we analyzed the applications and value of three neural network techniques:

- Feed forward neural networks. One of the most common types of artificial neural network. In this architecture, information moves in only one direction, forward, from the input layer, through the “hidden” layers, to the output layer. There are no loops in the network. The first single-neuron network was proposed in 1958 by AI pioneer Frank Rosenblatt. While the idea

Editor’s note: This article is a reprint of a chapter from the April 2018 McKinsey Global Institute discussion paper, “Notes from the AI frontier: Insights from hundreds of use cases.” The full paper can be downloaded here: <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>.

¹ For a detailed look at AI techniques, see *An executive’s guide to AI*, McKinsey Analytics, January 2018. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/an-executives-guide-to-ai>.

Exhibit 1

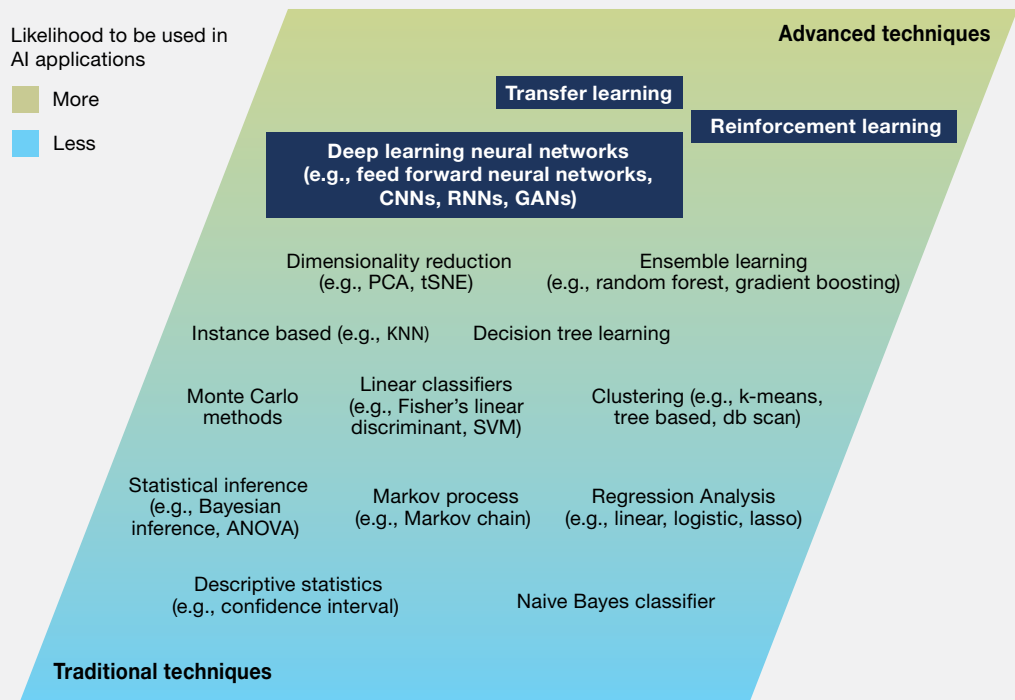
Artificial intelligence, machine learning, and other analytics techniques examined for this research.

Techniques

■ Considered AI for this research

Likelihood to be used in AI applications

■ More
■ Less



Source: McKinsey Global Institute analysis

is not new, advances in computing power, training algorithms, and available data led to higher levels of performance than previously possible.

- Recurrent neural networks (RNNs). Artificial neural networks whose connections between neurons include loops, well-suited for processing sequences of inputs, which makes them highly effective in a wide range of applications, from handwriting, to texts, to speech recognition. In November 2016, Oxford University

researchers reported that a system based on recurrent neural networks (and convolutional neural networks) had achieved 95 percent accuracy in reading lips, outperforming experienced human lip readers, who tested at 52 percent accuracy.

- Convolutional neural networks (CNNs). Artificial neural networks in which the connections between neural layers are inspired by the organization of the animal visual cortex, the portion of the brain that processes images, well suited for visual

Deep learning's origins and pioneers

It is too early to write a full history of deep learning—and some of the details are contested—but we can already trace an admittedly incomplete outline of its origins and identify some of the pioneers. They include Warren McCulloch and Walter Pitts, who as early as 1943 proposed an artificial neuron, a computational model of the “nerve net” in the brain.² Bernard Widrow and Ted Hoff at Stanford University developed a neural network application by reducing noise in phone lines in the late 1950s.³ Around the same time, Frank Rosenblatt, an American psychologist, introduced the idea of a device called the Perceptron, which mimicked the neural structure of the brain and showed an ability to learn.⁴ MIT’s Marvin Minsky and Seymour Papert then put a damper on this research in their 1969 book “Perceptrons,” by showing mathematically that the Perceptron could only perform very basic tasks.⁵ Their book also discussed the difficulty of training multi-layer neural networks. In 1986, Geoffrey Hinton at the University of Toronto, along with colleagues David Rumelhart and Ronald Williams, solved this training problem with the publication of a now famous back propagation training algorithm—although some practitioners point to a Finnish mathematician, Seppo Linnainmaa, as having invented back propagation already in the 1960s.⁶ Yann LeCun at NYU pioneered the use of neural networks on image recognition tasks and his 1998 paper defined the concept of convolutional neural networks, which mimic the human visual cortex.⁷ In parallel, John Hopfield popularized the “Hopfield” network which was the first recurrent neural network.⁸ This was subsequently expanded upon by Jurgen Schmidhuber and Sepp Hochreiter in 1997 with the introduction of the long short-term memory (LSTM), greatly improving the efficiency and practicality of recurrent neural networks.⁹ Hinton and two of his students in 2012 highlighted the power of deep learning when they obtained significant results in the well-known ImageNet competition, based on a dataset collated by Fei-Fei Li and others.¹⁰ At the same time, Jeffrey Dean and Andrew Ng were doing breakthrough work on large scale image recognition at Google Brain.¹¹ Deep learning also enhanced the existing field of reinforcement learning, led by researchers such as Richard Sutton, leading to the game-playing successes of systems developed by DeepMind.¹² In 2014, Ian Goodfellow published his paper on generative adversarial networks, which along with reinforcement learning has become the focus of much of the recent research in the field.¹³ Continuing advances in AI capabilities have led to Stanford University’s One Hundred Year Study on Artificial Intelligence, founded by Eric Horvitz, building on the long-standing research he and his colleagues have led at Microsoft Research. We have benefited from the input and guidance of many of these pioneers in our research over the past few years.

² Warren McCulloch and Walter Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bulletin of Mathematical Biophysics*, volume 5, 1943.

³ Andrew Goldstein, “Bernard Widrow oral history,” *IEEE Global History Network*, 1997.

⁴ Frank Rosenblatt, “The Perceptron: A probabilistic model for information storage and organization in the brain,” *Psychological review*, volume 65, number 6, 1958.

⁵ Marvin Minsky and Seymour A. Papert, *Perceptrons: An introduction to computational geometry*, MIT Press, January 1969.

⁶ David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, “Learning representations by back-propagating errors,” *Nature*, volume 323, October 1986; for a discussion of Linnainmaa’s role see Juergen Schmidhuber, Who invented backpropagation?, Blog post <http://people.idsia.ch/~juergen/who-invented-backpropagation.html>, 2014.

⁷ Yann LeCun, Patrick Haffner, Leon Botton, and Yoshua Bengio, Object recognition with gradient-based learning, *Proceedings of the IEEE*, November 1998.

⁸ John Hopfield, Neural networks and physical systems with emergent collective computational abilities, *PNAS*, April 1982.

⁹ Sepp Hochreiter and Juergen Schmidhuber, “Long short-term memory,” *Neural Computation*, volume 9, number 8, December 1997.

¹⁰ Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, ImageNet classification with deep convolutional neural networks, *NIPS 12 proceedings of the 25th International Conference on Neural Information Processing Systems*, December 2012.

¹¹ Jeffrey Dean et al., Large scale distributed deep networks, *NIPS 2012*.

¹² Richard S. Sutton and Andrew G. Barto, *Reinforcement learning: An introduction*, MIT Press, 1998.

¹³ Ian J. Goodfellow, Generative adversarial networks, *ArXiv*, June 2014.

perception tasks.

We estimated the potential of those three deep neural network techniques to create value, as well as other machine learning techniques such as tree-based ensemble learning, classifiers, and clustering, and traditional analytics such as dimensionality reduction and regression.

For our use cases, we also considered two other techniques—generative adversarial networks (GANs) and reinforcement learning—but did not include them in our potential value assessment of AI, since they remain nascent techniques that are not yet widely applied in business contexts. However, as we note in the concluding section of this paper, they may have considerable relevance in the future.

- **Generative adversarial networks (GANs).** These usually use two neural networks contesting each other in a zero-sum game framework (thus “adversarial”). GANs can learn to mimic various distributions of data (for example text, speech, and images) and are therefore valuable in generating test datasets when these are not readily available.
- **Reinforcement learning.** This is a subfield of machine learning in which systems are trained by receiving virtual “rewards” or “punishments,” essentially learning by trial and error. Google DeepMind has used reinforcement learning to develop systems that can play games, including video games and board games such as Go, better than human champions.

Problem types and the analytic techniques that can be applied to solve them

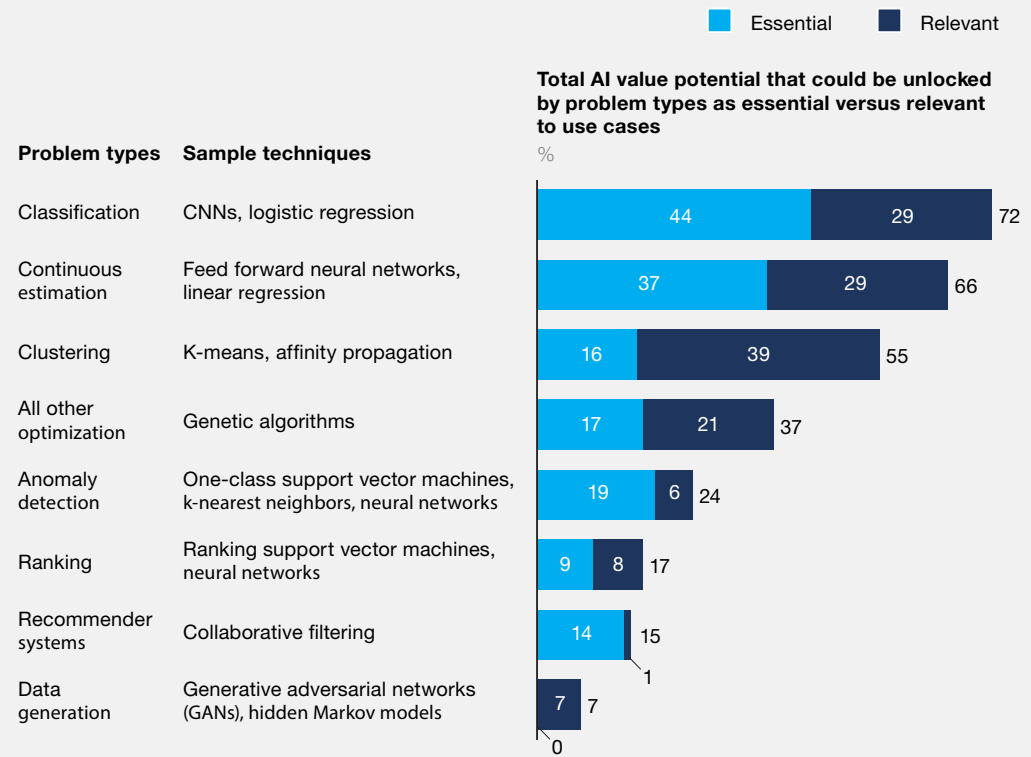
In a business setting, those analytic techniques can be applied to solve real-life

problems. For this research, we created a taxonomy of high-level problem types, characterized by the inputs, outputs, and purpose of each. A corresponding set of analytic techniques can be applied to solve these problems. These problem types include:

- **Classification.** Based on a set of training data, categorize new inputs as belonging to one of a set of categories. An example of classification is identifying whether an image contains a specific type of object, such as a truck or a car, or a product of acceptable quality coming from a manufacturing line.
- **Continuous estimation.** Based on a set of training data, estimate the next numeric value in a sequence. This type of problem is sometimes described as “prediction,” particularly when it is applied to time series data. One example of continuous estimation is forecasting the sales demand for a product, based on a set of input data such as previous sales figures, consumer sentiment, and weather. Another example is predicting the price of real estate, such as a building, using data describing the property combined with photos of it.
- **Clustering.** These problems require a system to create a set of categories, for which individual data instances have a set of common or similar characteristics. An example of clustering is creating a set of consumer segments based on data about individual consumers, including demographics, preferences, and buyer behavior.
- **All other optimization.** These problems require a system to generate a set of outputs that optimize outcomes for a specific objective function (some of the other problem types can be considered types of

Exhibit 2

Problem types and sample techniques.



NOTE: Sample techniques include traditional analytical techniques, machine learning, and the deep learning techniques we describe in this paper as AI. Numbers may not sum due to rounding.
Source: McKinsey Global Institute analysis

optimization, so we describe these as “all other” optimization). Generating a route for a vehicle that creates the optimum combination of time and fuel use is an example of optimization.

- **Anomaly detection.** Given a training set of data, determine whether specific inputs are out of the ordinary. For instance, a system could be trained on a set of historical vibration data associated with the performance of an operating piece of machinery,

and then determine whether a new vibration reading suggests that the machine is not operating normally. Note that anomaly detection can be considered a subcategory of classification.

- **Ranking.** Ranking algorithms are used most often in information retrieval problems in which the results of a query or request needs to be ordered by some criterion. Recommendation systems suggesting next product to buy use these types of algo-

rhythms as a final step, sorting suggestions by relevance, before presenting the results to the user.

- **Recommendations.** These systems provide recommendations, based on a set of training data. A common example of recommendations are systems that suggest the “next product to buy” for a customer, based on the buying patterns of similar individuals, and the observed behavior of the specific person.
- **Data generation.** These problems require a system to generate appropriately novel data based on training data. For instance, a music composition system might be used to generate new pieces of music in a particular style, after having been trained on pieces of music in that style.

Exhibit 2 (page 84) illustrates the relative total value of these problem types across our database of use cases, along with some of the sample analytics techniques that can be used to solve each problem type. The most prevalent problem types are classification, continuous estimation, and clustering, suggesting that meeting the requirements and developing the capabilities in associated techniques could have the widest benefit. Some of the problem types that rank lower can be viewed as subcategories of other problem types—for example, anomaly detection is a special case of classification, while recommendations can be considered a type of optimization problem—and thus their associated capabilities could be even more relevant.

Michael Chui is a partner in McKinsey's San Francisco office, where **James Manyika** is a senior partner. **Mehdi Miremadi** is a partner in the Chicago office, **Nicolaus Henke** is a senior partner in the London office, and **Rita Chung** is a consultant in the Silicon Valley office. **Pieter Nel** is a digital expert in the New York office, where **Sankalp Malhotra** is a consultant.