opentext™ | Cybersecurity

Position Paper

Artificial Intelligence and Machine Learning 101

Artificial intelligence (AI) is transforming the way that we interact with machines and the way that machines interact with us. This guide breaks down how AI functions, the strengths and limitations of various types of machine learning, and the evolution of this ever-changing field of study. It also explores the role of AI-enabled security analytics or user and entity behavioral analytics (UEBA) to better protect enterprises from today's complex cybersecurity threats.

Table of Contents

Part 1: Machine vs Human Learning

Artificial intelligence (AI) is everywhere—at least, that's how it seems. At OpenText™, the rise of AI is both exciting and challenging. But as we've engaged with our peers, customers, and partners, we have come to realize that the concept of AI is not always easily understood. To start this AI and Machine L, we will unpack the AI puzzle by answering the main question many folks are asking: "What is artificial intelligence, really?"

The easiest way to understand artificial intelligence is to map it to something we already understand—our own intelligence. How does non-artificial, human intelligence work? At the most basic level, our intelligence follows a simple progression: we take in information, we process it, and ultimately the information helps us act.

How does non-artificial, human intelligence work? At the most basic level, our intelligence follows a simple progression: we take in information, we process it, and ultimately the information helps us act.

Ls break this down into a system diagram. In the figure below, the three general steps of human intelligence from left to right: input, processing, and output. In the human brain, input takour eyes, nose, L
tak
process it. On the system's right side is output. This includ s speech and actions, both of which are dependent on how we process the raw input that our brain is receiving. The processing happens in the middle, where knowledge or memories are formed and retrieved, decisions and inferences and made, and l L

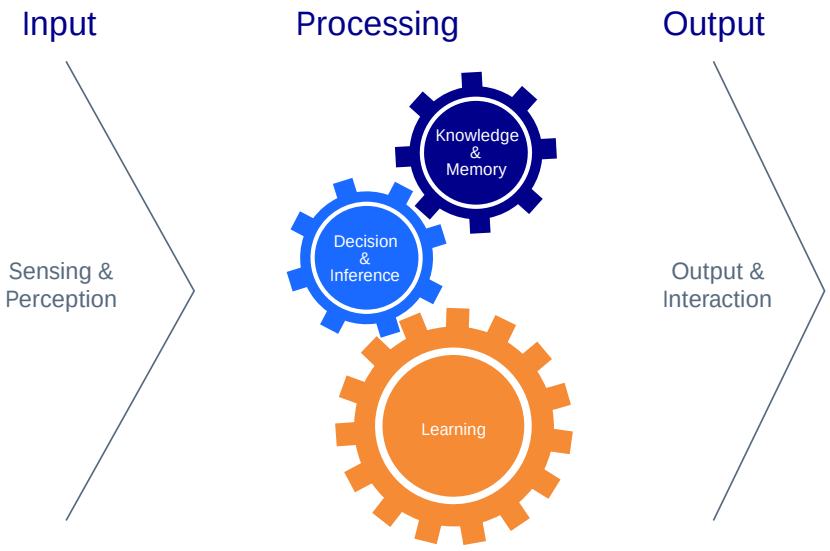


Figure 1. Human intelligence

Picture stopping at a roadway intersection. Your eyes see that the traffic light in front of you has just turned green. Based on what you have learned from experience (and driver's education), you know that a green light indicates that you should drive forward. So, you hit the gas pedal. The green light is the raw input, your acceleration is the output; everything in between is processing.

To intelligently navigate the world around us—answering the phone, baking chocolate chip cookies, or obeying traffic lights—we need to process the input that we receive. This is the core of human intelligence processing, and it is ultimately broken down into three distinct aspects:

1. **Knowledge and memory.** We build up knowledge as we ingest facts (i.e., the Battle of Hastings took place in 1066) and social norms (i.e., saying "Please" and "Thank you" is considered polite). Additionally, memory enables us to recall and apply information from the past to present situations. For example, Edward remembers that Jane did not thank him for her birthday present, so he does not expect her to thank him when he gives her a Christmas present.
2. **Decision and inference.** Decisions and inferences are made based on raw input combined with knowledge and/or memory. For example, Edward ate a jalapeno pepper last year and

In machines, the input part of artificial intelligence is exemplified by natural language processing, speech recognition, visual recognition, and more. You see such technologies and algorithms everywhere, from self-driving cars that need to sense the roadways and obstacles, to Alexa or Siri when it recognizes your speech. The output that follows are ways in which machines interact with the world around us. This might take the form of robotics, navigation systems (to guide those self-driving cars), speech generation (e.g., Siri), etc. In between, we have various forms of processing that takes place.

Similar to our accrual of knowledge and memories, machines can create knowledge representations (e.g., graph databases, ontologies) that help them store information about the world. Just as humans make decisions or draw inferences, machines can make a prediction, optimize for a target or outcome, and determine the best next steps or decisions to meet a specific goal.

Finally, just as we learn by example, observation, or algorithm, machines can be taught using analogous methods. Supervised machine learning is much like learning by example: the computer is given a dataset with “labels” within the data set that act as answers, and eventually learns to tell the di□erence between di□erent labels (e.g., this dataset contains photos label[Frh)toshin is23ladog24 wiois23lacat24 w

Neural networks—a programming paradigm in which we train machines to “learn”—are inspired by neurons, or specialized cells in the human body that form the foundation of our nervous system, and brains in particular. These cells transmit signals throughout our bodies trigger nervous system responses and processes. Neurons are what enable us to see, hear, smell, etc.

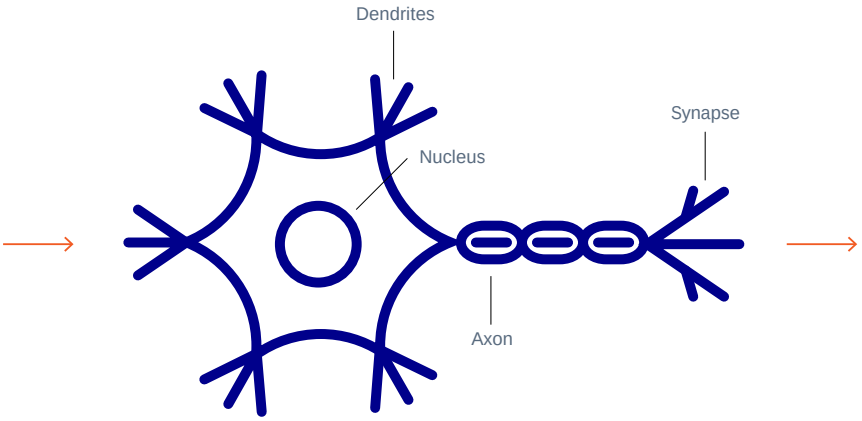


Figure 3. How neurons receive and send messages

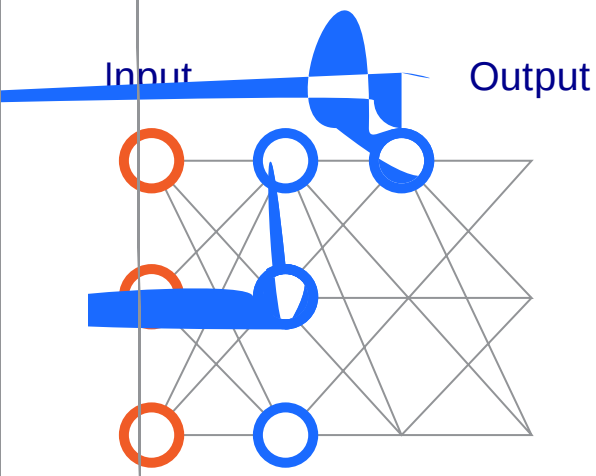
Much of what we think of as human learning can be described by how strong the connection between two neurons in our brain is, along with the strength of the firing of our synapses.

In part one of this guide, we discussed the basic process of human intelligence: input on the left, and output on the right. The neuron (pictured above) plays a critical role in this. On the left side of the neuron, the cell body collects “input.” Once it receives enough input or stimulation, the axon fires, transmitting the information to the right side—the synapse. The “output” is then sent to other neurons.

At any given moment, our neurons are passing messages between each other. These cells are responsible for our ability to perceive our surroundings. And when we learn, our neurons become very active. In fact, much of what we think of as human learning can be described by how strong the connection between two neurons in our brain is, along with the strength of the firing of our synapses.

A neural network is a mathematical simulation of a collection of neuron cells. The image below represents a basic neural network with 3 layers and 12 nodes.

Each circular node represents an artificial, biologically inspired “neuron.” The lines represent a connection from the output of one artificial neuron on the left to the input of another on the right. Signals between these neurons flow along the lines from left to right. In these networks, input—such as pixel data—flows from the input layer, through the middle “hidden” layers, and ultimately to the output layer in a manner described by mathematical equations loosely inspired by the electrical activity in actual biological neurons.


Input
Output

It's worth noting, however, that deep learning is not a silver bullet for machine learning—especially not in cybersecurity, where sometimes there is not the large volume of clean data that is ideal for deep learning methods. It is important to pick the right algorithm, data, and principles for the job. This is the best way for machines to gather evidence, connect the dots, and draw a conclusion.

Neural networks might seem like the stu of the future, but it's been around for a while. In fact, neural networks are based on ideas that started circulating back in the 1940s. In the next section, we will take a short trip back in time to understand how neural networks and machine learning have come to permeate many parts of modern life.

Neural networks might seem like the stu of the future, but it's been around for a while. In fact, neural networks are based on ideas that started circulating back in the 1940s.

Part 3: A Brief History of Artificial Intelligence

For some people, the term artificial intelligence (AI) might evoke images of futuristic cities with flying cars and household robots. But AI isn't a futuristic concept, at least not anymore.

Although not referred to as such, the idea of artificial intelligence can be traced back to antiquity (i.e., Greek god Hephaestus's talking mechanical handmaidens).[1] Since the 1930s, scientists and mathematicians alike have been eager to explore creating true intelligence separate from humans.

AI's defining moment in the mid-20th century was a happy confluence of math and biology, with researchers like Norbert Wiener, Claude Shannon, and Alan Turing having already chipped away at the intersection of electrical signals and computation. By 1943, Warren McCulloch and Walter Pitts had created a model for neural networks. Neural networks paved the way for a brave new world of computing with greater horsepower, and, in 1956, the field of AI research was o cially established as an academic discipline.

The latter half of the century was an exciting age for AI research and progress, interrupted occasionally by "AI winters" in the mid-70s and late 80s where AI failed to meet public expectations, and investment in the field was reduced. But despite setbacks, di erent applications for AI and machine learning were appearing left and right. One particular anecdote of such an application has become a popular parable within the scientific community, speaking quite e ectively to the trials and tribulations of AI research and implementation.

The story goes something like this:

In the 1980s, the Pentagon decided to use a neural network to identify camouflaged tanks. Working with just one mainframe (from the 1980s, keep in mind), the neural net was trained with 200 pictures—100 tanks and 100 trees. Despite the relatively small neural network (due to 1980's limitations on computation and memory), the lab training resulted in 100% accuracy. With such success, the team decides to give it a go out in the field. The results were not great.

With the availability of vast computing resources that were undreamed of back in the 1980's, deep neural networks have quickly become a popular area for research. Deep learning gives a system the ability to automatically "learn" through billions of combinations and observations, reducing the dependency on human resources.

Figure 6. Lab vs field pictures (Source: Neural Network Follies, Neil Fraser, September 1998)

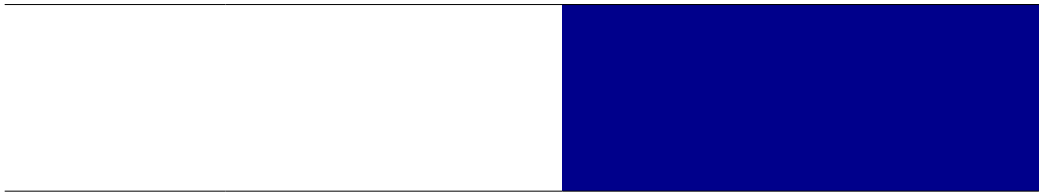
Why did the neural network do so fantastically on the photos in the lab, but fail so completely in the field? It turned out that the non-tank photos were all taken on days where the sky was cloudy; all the pictures of trees were taken on days where the sun was shining. The neural net had been trained to recognize sunniness, not tanks.

Eventually, though, visual recognition via deep learning—facilitated by neural networks that are much more complex than the Pentagon's 1980s mainframe would have been able to handle—became a reality. In 2012, Stanford professor Andrew Ng and Google fellow Je Dean created one of the first deep neural networks using 1000 computers with 16 cores each. The task: analyze 10 million YouTube videos. The result: it found cats.[2] Thanks to its "deep learning" algorithm, the network was able to recognize cats over time, and with very good accuracy.

With the availability of vast computing resources that were undreamed of back in the 1980's, deep neural networks have quickly become a popular area for research. Deep learning gives a system the ability to automatically "learn" through billions of combinations and observations, reducing the dependency on human resources. Within the cybersecurity domain, the method has become particularly promising for detecting malware—scenarios in which we have large datasets with many examples of malware from which the network can learn.

Unfortunately, deep learning methods are currently less e ective when it comes to certain use cases, like insider threat, because we simply don't have the right kind of data on these types of attacks, in the volumes required. Most often, the information we have on insider threats are anecdotal, which cannot be used e ciently by these types of neural networks.

Until we can gather more e□ective datasets (and reduce the cost and complexity of deep



The most common analytics we see in security today involves predictive models, which allow us to identify where risks might be within large amounts of data (this is where anomaly detection fits in). In a nutshell, predictive modeling combines historical data with real-time behavior to understand or predict future behavior. With this, we can answer the question, “What happens next?”

But our vision for security analytics doesn’t stop here. Predictive analytics is just one piece of a much larger puzzle that can give us much more useful insight for security teams. The ideal analytics paradigm combines intelligent sensor and ubiquitous data sources—desktops and servers, mobile, cloud, social networks, open data, etc.—with multiple advanced analytical approaches to behavioral and threat analysis, including forensic analysis, risk modeling, anomaly detection, behavioral and response optimization, and more.

This means that we can do far more than predict or identify a threat. It allows us to go even further to o□er not just advanced detection but insight into how to respond most e□ectively. Security analytics gives us the power to answer other key questions, like “How many threats are there?” and “What is the best possible reaction?”

We haven’t seen other classes of analytics like optimization methods applied to cybersecurity

Connect with Us

www.opentext.com

opentext™ | Cybersecurity

OpenText Cybersecurity provides comprehensive security solutions for companies and partners of all sizes. From prevention, detection and response to recovery, investigation and compliance,