CyberRes

White 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 behavioural analytics (UEBA) to better protect enterprises from today's complex cybersecurity threats.

Table of Contents

Part 1: Machine vs Human Learning 1

Part 2: The Neural Network and Deep Learning 3

Part 3: A Brief History of Artificial Intelligence 6

Part 4: A New Vision for Security Analytics. 8

Part 1: Machine vs Human Learning

Artificial intelligence (AI) is everywhere—at least, that is how it seems. At Micro Focus, the rise of AI is both exciting and challenging. But as we have engaged with our peers, customers, and partners, we have come to realise that the concept of AI is not always easily understood. To start this AI and Machine Learning 101 guide, 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.

Let us 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 takes place in the form of sensing and perceiving things. Your eyes, nose, ears, etc., take in raw input on the left, such as photons of light or the smell of pine trees, and then process it. On the system’s right side is output. This includes 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 learning occurs.

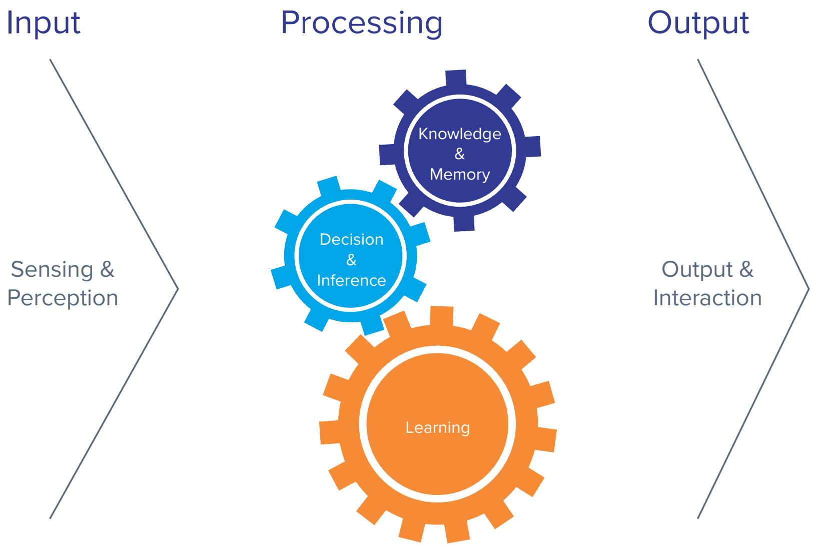

Figure 1. Human intelligence

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.

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 did not like it. When Johnny offers a pepper to Edward, he decides not to eat it.
3. **Learning.** Humans can learn by example, observation, or algorithm. In learning by example, we are told that one animal is a dog, the other is a cat. In learning by observation, we figure out on our own that dogs bark and that cats meow. The third learning method—algorithm—enables us to complete a task by following a series of steps or a specific algorithm (e.g., performing long division).

These aspects of human intelligence parallel artificial intelligence. Just as we take in information, process it, and share output, so can machines. Let us take a look at the figure below to see how this maps out.

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.

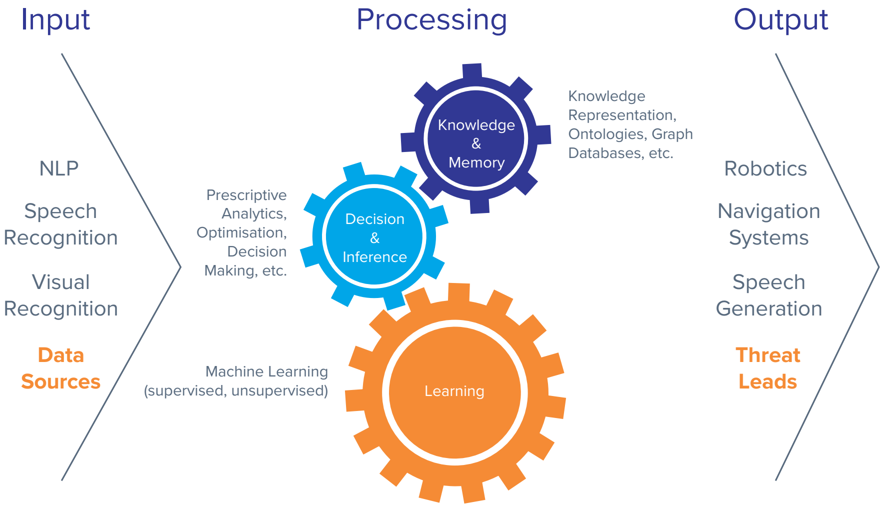

Figure 2. Artificial intelligence

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 recognises 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, optimise 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 difference between different labels (e.g., this dataset contains photos labelled as either “dog” or “cat”, and with enough examples, the computer will notice that dogs generally have longer tails and less pointy ears than cats).

Unsupervised machine learning, on the other hand, is like learning by observation. The computer observes patterns (dogs bark and cats meow) and, through this, learns to distinguish groups and patterns on its own (e.g., there are two groups of animals that can be separated by the sound they make; one group barks—dogs—and the other group meows—cats). Unsupervised learning does not require labels and can therefore be preferable when data sets are limited and do not have labels. Finally, learning by algorithm is what happens when a programmer instructs a computer exactly what to do, step-by-step, in a software programme.

Ideally, the most accurate and efficient artificial intelligence results require a combination of learning methods. Both supervised and unsupervised machine learning are useful methods—it is all about applying the right approach or approaches to the right use case.

Ideally, the most accurate and efficient artificial intelligence results require a combination of learning methods. Both supervised and unsupervised machine learning are useful methods—it is all about applying the right approach or approaches to the right use case.

Next, we will put machine learning under the microscope to understand how this part of AI mirrors the neurons in our brain to turn input into to optimal output.

Part 2: The Neural Network and Deep Learning

Machine learning is just one part of AI, although it has a massive subset of algorithms within it. One method that you hear frequently today is “deep learning,” an algorithm that has received a fair share of attention in the news in recent years. To understand its popularity and success, it is helpful to understand how it works. Deep learning is an evolution of a machine learning algorithm that was popular in the 1980s that you may recognise: neural networks.

Neural networks—a programming paradigm in which we train machines to "learn"—are inspired by neurons, or specialised 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.

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.

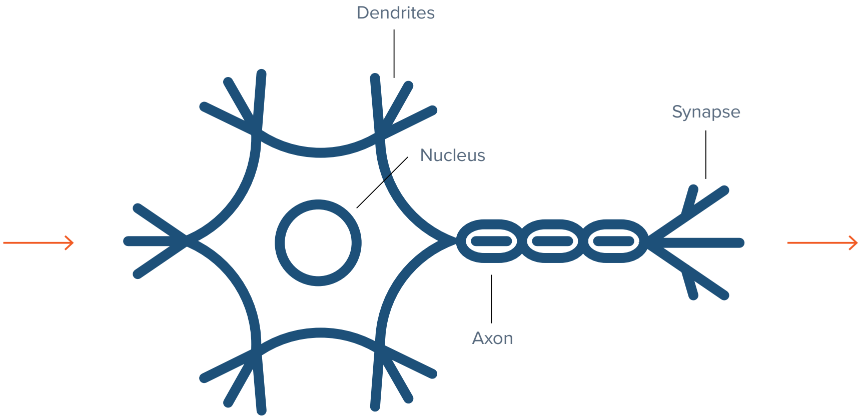

Figure 3. How neurons receive and send messages

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.

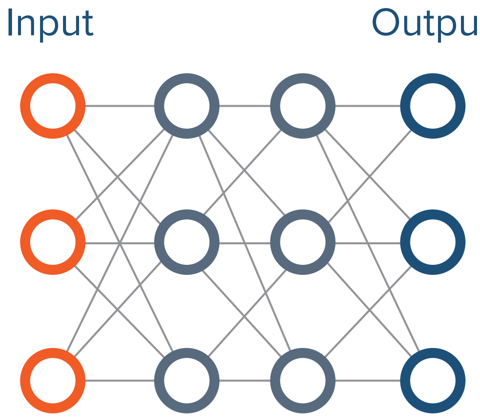

Figure 4. Simple neural network

Neural networks learn by trying to match data sets presented to the input layer to desired outcomes in the output layer. The mathematical equations compute the outputs, compare the simulated output to the desired outcome, and the resulting differences then produce tweaks to the strength of the connections. These tweaks are iteratively modified until the computed output is close enough to the desired outcome, at which point we say the neural network has "learned."

Neural networks learn by trying to match data sets presented to the input layer to desired outcomes in the output layer. The mathematical equations compute the outputs, compare the simulated output to the desired outcome, and the resulting differences then produce tweaks to the strength of the connections.

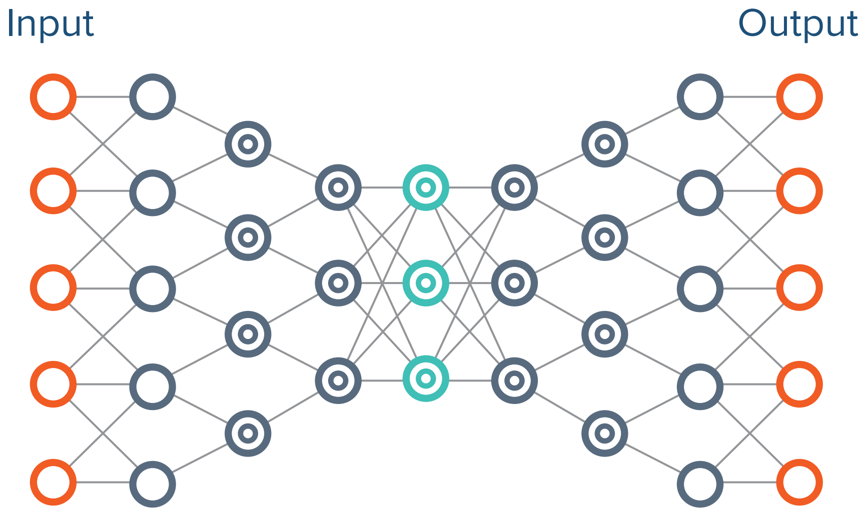

Figure 5. Complex neural network

These "deeper" neural networks can do much more complex predictions. There can be thousands of nodes and hundreds of layers, which means thousands of different calculations. Deep learning models have become very good at specific problems, such as speech or image recognition.

It is 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 stuff of the future, but it is 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 stuff of the future, but it is 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 is not 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 officially 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, different 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 effectively 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.

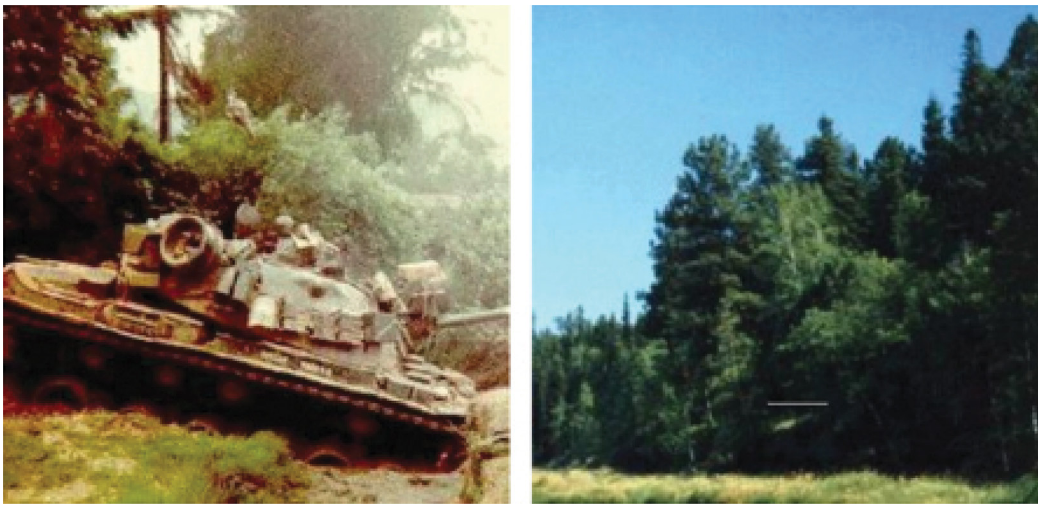

Figure 6. Lab vs field pictures (Source: Neural Network Follies, Neil Fraser, September 1998)

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 recognise 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 Jeff Dean created one of the first deep neural networks using 1000 computers with 16 cores each. The task: analyse 10 million YouTube videos. The result: it found cats.[2] Thanks to its "deep learning" algorithm, the network was able to recognise cats over time, and with very good accuracy.

With the availability of vast computing resources that were undreamed of back in the 1980s, 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.

With the availability of vast computing resources that were undreamed of back in the 1980s, 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.

Unfortunately, deep learning methods are currently less effective when it comes to certain use cases, like insider threat, because we simply do not 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 efficiently by these types of neural networks.

Until we can gather more effective datasets (and reduce the cost and complexity of deep learning systems), deep learning is not the right choice for all use cases. And that is okay. Deep learning is just one of many machine learning algorithms, and these approaches can be just as if not more valuable—it all depends on the job at hand.

We have seen immense potential of AI technologies in the six decades since its official "birth," and we have only just scratched the surface, especially in security. Next, we will take a deeper dive into the potential applications for AI and analytics to change the way that we identify and respond to security threats.

> We have seen immense potential of AI technologies in the six decades since its official "birth," and we have only just scratched the surface, especially in security.

Predictive analytics is just one piece of a much larger puzzle that can give us much more useful insight for security teams.

Part 4: A New Vision for Security Analytics

So far, this guide has taken a close look at machine learning, understanding its limitations and strengths. There is enormous potential for machine learning to facilitate AI, but it is worth noting that the broader game of threat detection is not just about deep learning or machine learning as we know it today. New analytical methods combined with new data types can give us entirely new frameworks in which to analyse and act upon security threats.

New Methods	Adaptive Analysis Continual Analysis Optimisation under Uncertainty	Responding to context Responding to local change/feedback Quantifying or mitigating risk	
Traditional	Optimisation Predictive Modelling Simulation Forecasting Alerts Query/Drill Down Ad hoc Reporting Standard Reporting	Decision complexity, solution speed Casually, probabilistic, confidence levels High fidelity, games, data farming Larger data sets, nonlinear regression Rules/triggers, context sensitive, complex events In memory data, fuzzy search, geo spatial Query by example, user defined reports Real time, visualisations, user interaction	Today
New Data	Entity Resolution Relationship, Feature Extraction Annotation and Tokenisation	People, roles, locations, things Rules, semantic inferencing, matching Automated, crowd sourced	

We have seen what analytics can do for other industries, and there is potential for analytics to have a profound impact on cybersecurity, too. We see this taking shape in a new field that we refer to as security analytics, which essentially takes the battle-tested algorithms and methodologies that we have discussed (and more) and applies them help solve the really difficult problems in security.

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 modelling combines historical data with real-time behaviour to understand or predict future behaviour. With this, we can answer the question, "What happens next?"

But our vision for security analytics does not 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 behavioural and threat analysis, including forensic analysis, risk modelling, anomaly detection, behavioural and response optimisation, and more.

This means that we can do far more than predict or identify a threat. It allows us to go even further to offer not just advanced detection but insight into how to respond most effectively. 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 have not seen other classes of analytics like optimisation methods applied to cybersecurity yet, but they have immense potential. These techniques look at all the possible reactions to a security risk and determining the best response. Yes, there are ways to do this with math.

For example, optimisation methods are used when you place a call to your cell phone service provider with an issue. They are not randomly making a recommendation on whether or not to upgrade your service plan at a discount; they rely on a set of mathematics in the background that looks at your call logs, the number of dropped calls, how your history compares with that of other users, etc. It even calculates the probability that you might switch to another service provider. Then, out of all the possible next steps, it computes the best next step to maximise customer retention.

The same math can be applied to a security team to identify a risk, provide a number of ways in which to react, and determine mathematically the best response to maximise containment of this particular risk.

The rapid rise and evolution of security threats make this type of response efficiency critical. We have more data today than ever before. Thankfully, we also have more compute power, better algorithms, and broader investment in research and technologies to help us make sense of this data through mathematics. By all accounts, we believe security analytics is just getting started.

We have more data today than ever before. Thankfully, we also have more compute power, better algorithms, and broader investment in research and technologies to help us make sense of this data through mathematics. By all accounts, we believe security analytics is just getting started.

Contact us at **CyberRes.com**

Like what you read? Share it.

CyberRes
A Micro Focus Line of Business