An Overview of AI, ML and DL*

Tony S. Wirjanto†

August 16, 2019

1 Introduction

Artificial Intelligence (AI), as a field, has existed for many years. However, recent increases in computing power coupled with an increase in the availability and quantity of data have resulted in a renewed interest in applications of (AI). These applications are used to diagnose diseases, translate languages, and drive cars; and they are increasingly being used in the financial industry as well.

Originally, AI was defined by John McCarthy, who is a computer scientist, at the Dartmouth Conference in 1956 as the science and engineering of making intelligent machines, especially intelligent computer programs. More informatively, AI is defined in the Oxford Dictionary as a theory and development of computer systems able to perform tasks that traditionally have required human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. According to this definition, one way to implement AI is to develop an expert system; i.e., to build a database of knowledge from human experts and apply this data to offer advice or make decisions. This technique was popular in the 1980s, but it has attracted relatively few attention as people working on expert systems have come to understand better the complexity of many seemingly simple problems.

AI is a broad field, of which Machine Learning (ML) is one of its branch. In 1959, Arthur Samuel defined ML as a field of study that gives computers the ability to learn without being explicitly programmed. Thus an alternative way of implementing AI is by way of having the machine learn directly from the data. ML may be defined as a method of designing a sequence of actions to solve a problem, known as algorithms, which optimize automatically through experience and with limited or no human intervention. These techniques can be used to find patterns in large datasets from increasingly diverse and innovative sources.

In order to understand how AI, ML, and Deep Learning (DL), a subset of ML, technologies affect the financial industry, it is important to have a big picture of what these technologies are and how they engage in learning.

---

*This note is prepared for ACTSC 974 / STAT 974: Financial Econometrics, under the theme: AI in Finance & Business for Fall 2019. The discussion in this note is drawn from the materials in https://towardsdatascience.com/notes-on-artificial-intelligence-ai-machine-learning-ml-and-deep-learning-dl-for-56e51a2071c2

†Department of Statistics & Actuarial Science, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada, email: twirjanto@uwaterloo.ca.
According to McKinsey, AI is by far the most popular buzzword around. It should be pointed out that AI occupies #1 and #3 spots of Gartner Top 10 Strategic Technology Trends in 2019. Indeed, AI has become a catch-all term that refers to any computer algorithm that automatically does something. However, to fully appreciate how these technologies may influence finance both as a discipline and as an industry, it is important to have a rather high-level understanding of what these technologies are, how they engage in learning, and what are some of their critical strengths and weaknesses.

As mentioned earlier, two important advances have occurred recently: (i) availability of big data, and (ii) affordability of high computing power. AI works best when a large amount of data sets is combined with a set of fast, iterative processing and intelligent algorithms. This makes it possible for the AI software to learn automatically from patterns (or features) in that vast data set. The ubiquitous presence of the AI in our daily life was jolted by an AlphaGo AI program introduced in May 2017 at the board game GO, which used Reinforcement Learning (RL) algorithms. Nowadays, AI is intricately associated with self-driving cars, Alexa/Siri like digital assistants frenzy, real-time face recognition at airports, human genome projects, Amazon/Netflix algorithms, AI composers/artists, hand-writing recognition, Email marketing algorithms and the list can go on and on.

AI technologies will continue to be disruptive in 2019 and beyond and they will become even more prevalent in our daily life due to increasing affordability of cloud computing and increasing accessibility to big data.
2 What is the difference between AI, ML and DL?

What are the conceptual differences between AI, ML and DL? AI is a broadly defined term and this has undoubtedly emboldened many companies to claim freely nowadays that their product has an AI component to it. Machine Learning (ML) is a subset of AI, and consists of more advanced methods/models that enable computers to figure things out from the data and deliver AI applications; i.e., ML is the science of getting computers to act without being explicitly programmed. Lastly, Deep Learning (DL) is an area within ML that uses a multi-layered Artificial Neural Network (ANN) to deliver high accuracy for tasks such as object detection, speech recognition, language translation and other recent breakthroughs. The advantage of DL is that it can automatically learn/extract/translate features from data sets, such as images, video or text, often without using traditionally hand-coded algorithms.

3 Machine Learning

There are three broad categories associated with ML methods/models: (1) Supervised Learning, (2) Unsupervised Learning, and (3) Reinforcement Learning.
3.1 Supervised Learning

In Supervised Learning (SL), algorithms are fed a set of training data that contains labels on some portion of the observations. For instance, a data set of financial transactions may contain labels on some data points identifying those that are fraudulent and those that are not fraudulent. The algorithm will learn a general rule of classification that it will use to predict the labels for the remaining observations in the data set. In particular, SL involves an output label (called a response variable) associated with each instance (called a feature/predictor) in the dataset. In general, this output can be discrete/categorical or continuous. In the SL case, the data has known labels as output. It involves a supervisor that is more knowledgeable than, say, the NN. For instance, the supervisor feeds some example data about which she already knows the answers. She guides the system by tagging the output. For example, a supervised ML system that can learn which emails are ”spam” and which are ”not spam.” The associated learning algorithm would be first trained with available input data set (of zillions of emails) that is already tagged with this classification to help the ML system learn the characteristics or parameters of the ”spam” email and distinguish it from those of ”not spam” emails. Just as a three-year-old child would learn the difference between a ”block” toy and a ”soft” toy, the supervised ML system learns which email is ”spam” and which is ”not spam.” Methods such as linear regressions, logistic regressions, or decision tree classification, fall under the SL category.

ML can be applied to different types of problems, such as regression analysis or classification. Regression
algorithms estimate the outcome of problems that have an infinite number of solutions (continuous set of possible outcomes). This outcome can be accompanied with a confidence interval. Regression algorithms can be used for the pricing of options. Regression algorithms can also be used as one intermediate step of classification algorithm. Classification algorithms, in contrast, which are far more frequently deployed in practice, group observations into a finite number of categories. Classification algorithms are probability-based, meaning that the outcome is the category for which it finds the highest probability that it belongs to. An example might be to automatically read a sell-side report and label it as "bullish" or "bearish" with some probability, or estimate an unrated company's initial credit rating.

**Regression.** We wish to predict/forecast continuous response values. As an example, *stock prices*. In this case, we have a data set and outputs and our learning algorithm predicts the outcome based on a fitting function.

**Classification.** We need to categorize a certain observation into a group. In the graph below, we are given a dot that needs to be classified as either a blue dot or a red dot. Other examples include prediction (i) if a given email is spam or not spam, (ii) if a detected particle is a Higgs Boson or a normal sub-atomic particle, (iii) if it will rain today or not, (iv) if this picture is a cat or not? Still other applications involve (i) detecting fraud in financial transactions or evaluating risk for frauds or insurance underwriting, and (ii) assigning a certain news article into a group, such as sports, weather, or science.
3.1.1 Unsupervised Learning

Unsupervised Learning (UL) refers to a situation where the data provided to the algorithm does not contain labels. The algorithm is asked to detect patterns in the data by identifying clusters of observations that depend on similar underlying characteristics. For example, an unsupervised ML algorithm could be set up to look for financial securities that have characteristics similar to an illiquid security that is hard to price. If it finds an appropriate cluster for the illiquid security, pricing of other securities in the cluster can be used to help price the illiquid security.

In particular, UL is an “unaided” type of learning, where the data has no known output labels or any feedback loop. There is no example data set with known answers and we in effect are searching for a hidden pattern. In this case, clustering i.e. dividing a set of elements into groups according to some unknown pattern is carried out based on the data sets. The system has to understand itself from the data set we provide. In general, UL is a bit more challenging to implement and, as a result, it is not used as widely as SL. There are two popular types of UL: (i) Clustering and (ii) Association.

**Clustering.** This is a type of UL problem where similar things are being grouped together. For example, given news articles or books, we cluster them into different types of themes. Given a set of tweets, we cluster them based on content of tweet. It could also be used in the other areas such as health care, shopping, real estate, etc.

**Association.** The goal is to find exact rules that will describe a large portion of the data. For example, people who buy X are also the ones who tend to buy Y. We also encounter this type of situation when we receive a
book or movie recommendation based upon previous purchases or searches. Moreover, these algorithms are used for market basket analysis using online or offline retailers shopping (point of sales) data. Given many baskets, association helps us to understand which items inside a basket predict another item in the same basket.

As an example, Figure 3.4 depicts associations between selected items using a data set on an actual grocery transaction over 30 days. Larger circles imply higher support, while red circles imply higher lift; i.e., the most popular transaction was of pip and tropical fruits. Relatively, many people buy sausage along with sliced...
cheese.

3.2 Reinforcement Learning

Reinforcement Learning (RL) falls in between SL and UL. In this case, the algorithm is fed an unlabelled set of data, chooses an action for each data point, and receives feedback (perhaps from a human) that helps the algorithm learn. For instance, reinforcement learning can be used in robotics, game theory, and self-driving cars.

Suppose that instead of telling a child which toy to put in which box, we reward the child with a big hug when the child makes the right choice, and make a sad face when the child makes the wrong choice. Very quickly, just after a few iterations, the child would learn which toys need to go into which box. This is (Figure 3.5: Source: Wikipedia

This strategy is built on observation and trial and error in order to maximize a reward. The agent makes a decision by observing its environment. If the observation is negative, the algorithm adjusts its weights to be able to make a different decision the next time. We can view RL as part of DL based on the number of hidden nodes and the complexity of algorithms (more on this later). RL algorithms try to find the best ways to earn the greatest reward. Rewards can be winning a game, earning more money or beating other opponents.

Google DeepMind used RL to develop systems that can play games, including video games and board games,
such as GO. AlphaGo won a game with more board states than chess at $10^{170}$, which is greater than the number of atoms in the universe, against a den 9 GO master. A combination of RL and human SL was used to build value and policy NNs, which also used the search tree to execute its game play strategies. The software learned from 30 million moves played in human-on-human games.

More recently, it was announced that AI agents developed by Google DeepMind had defeated human professional players at Starcraft II?, which is the first in the AI world. **Note:** Starcraft II game is much harder for computers to play than board games such as chess or Go.

Google DeepMinds researchers used RL to train the AlphaStar agents. These agents play the game by trial and error while trying to reach the goal of winning or simply staying alive in the game. They learn first by imitating human players and then play against each other, where strongest agents survive, and weakest agents are left behind. DeepMind estimated that its AlphaStar agents each racked up to about 200 years of game time in this way at an accelerated rate.

Below is a graphical summary to wrap up the overview on ML.

### 4 Deep Learning

Many AI techniques are hardly new. Indeed, [NNs], the basic concept for DL, were first developed in the 1960s. However, after an initial burst of excitement, AI failed to live up to their promises and research
funding dissipated for over a decade, in part because of the lack of sufficient computing power and data. There was renewed research funding and interest in applications in the 1980s, during which many of the research concepts were developed for later breakthroughs. By 2011 and 2012, driven by the vast increase in the computational power of modern computers, ML algorithms, especially DL algorithms, began to consistently
win image, text, and speech recognition contests. **Noticing this trend** major tech companies began to acquire DL start-ups and rapidly accelerate deep learning research. Also new is the scale of collection of big data, for example the ability to capture data on the scale of every single credit card transaction or every word on the web, and even “mouse” hovers over websites. Other advances have also helped, such as increased interconnectedness of information technology resources with cloud computing architecture, with which big data can now be organized and analyzed. Using data sets of this size and complexity and with the increase in computing power, ML algorithms results have improved, some of which are highlighted in the sections that follow. This has also spurred large investments in AI start-ups. **The World Economic Forum** reported that global investment in AI start-ups rose from $2.4 billion in 2011, to $2.4 billion in 2015.

Deep Learning **DL** is a form of ML that uses algorithms that work in “layers” inspired by the structure and function of the brain. DL algorithms, whose structure are called Artificial Neural Networks **ANN**, can be used for SL, UL, or RL. **McKinsey** claims that DL techniques have the potential to create between $3.5 trillion and $5.8 trillion in value annually in 19 industries! Like ML, DL is also a method of statistical learning that extracts features (or attributes) from raw data sets. **The main difference** is that DL does this by utilizing a multi-layer **ANN** with many hidden layers stacked one on top of the other. DL also has somewhat more sophisticated algorithms and requires more powerful computational resources. These are specially designed computers with high performance CPUs or GPUs. They could be on premise or as workloads on Cloud.

Below, I discuss three widely used DL methods/models: (i) Convolutional NN, (ii) Recurrent NN, and (iii) Generative Adversarial Networks. I will also share my view on how Natural Language Processing (NLP) can use DL.

**What is an ANN?** How does a small child learn to recognize the difference between a school bus and a regular transit bus? How do we subconsciously perform complex pattern recognition tasks without even noticing it? The answer is that we have a biological neural network that is connected to our nervous systems. Our brains are very complex networks with about 10 billion neurons each connected to 10 thousand other neurons. Each of these neurons receives electro-chemical signals and passes these messages to other neurons. Actually, we do not even know well how our brain neurons work. We do not know enough about neuroscience and the deeper functions of the brain to be able to correctly model how the brain works. **DL** is only inspired by the functionality of our brain cells called neurons, which lead to the concept of artificial neural networks (**ANN**). **ANN** is modeled by using layers of artificial neurons to receive input and apply an activation function along with a human set threshold. **DL** has already achieved closed to, or better than, human level image classification, speech/hand writing recognition and of course the autonomous driving. Complex ad targeting or news feeds are all over the net these days.

In the most basic feed forward NN (top right), there are five main components to artificial neurons. From
left to right, these are:

1. **Input nodes.** Each input node is associated with a numerical value, which can be any real number. Example could be one pixel value of an image.

2. **Connections.** Similarly, each connection that departs from the input node has a weight \( (w) \) associated with it and this can be any real number. The ANN runs and propagates millions of times to optimize these "w" values. We need a high computational power to make this run in a relatively short time.

3. Next, all of the values of the input nodes and weights of the connections are brought together. They are used as inputs for a **weighted sum**.

4. This result will be the input to a **transfer or activation function.** Just like a biological neuron, which only fires when a certain threshold is exceeded, the artificial neuron will also only fire when the sum of the inputs exceeds a threshold. These are parameters set by us (this procedure inevitably leads to an important discussion of AI vs. Ethics, which we are not covering in this course.)

5. As a result, we have the **output node**, which is associated with the function of the weighted sum of the input nodes.

**What is the Deep in DL?** DL networks are distinguished from the more general single-hidden-layer NN by their depth. Depth is the number of node layers, where there are more than one hidden layers, and thus DL needs more computational power for forward/backward optimization when it trains, tests and eventually runs these ANNs.

Among the layers, we can distinguish an input layer, hidden layers and an output layer. The layers act like biological neurons. The outputs of one layer serve as the inputs for the next layer.
Convolutional Neural Networks. These are one of the most popularly applied DL cases. They are great for image/video processing or computer vision applications. CNNs are deep ANN that are used primarily to classify images (e.g. label what they see), cluster them by similarity (photo search), and perform object recognition within scenes. These are algorithms that can identify faces, individuals, street signs, tumors, flowers and many other aspects of visual data. Self driving cars or drones use CNN capabilities. The most popularly applied corporate cases are probably an optical character recognition (OCR) to digitize text to automate data entry.

In the above example, the CNN algorithm sees the image differently than the human brain does. Each image is a 3-dimensional arrays of numbers, known as pixels with width, height and depth. Width and height depends on the image resolution. The 3rd dimension (depth) is of the Red-Green-Blue (RGB) values for the color code (unless we are using a black and white image as input).
Technically, DL-CNN receives these images to pass through a series of convolution layers with filters (basic depiction below). The way in which the CNN layers work requires a more elaborate discussion. Initially, these filters do not know where to look for image features like edges or curves and the previously mentioned weights are random numbers. We typically have a large training data set with thousands of images with pre-identified labels. The model first makes a forward pass, calculates the initial weights, makes a prediction of the outcome label (i.e. this is a dog) and compares it with the truth that is the existing training set labels. As this is a training set, we already know the outcome labels; thus depending on the success of the prediction, a loss function is calculated and the network makes a back pass while updating its weights. The way the computer is able
to adjust its weights in order to reduce the loss is through a back propagation. Next, the model performs a backward pass through the network to determine which weights have contributed most to the loss and finding ways to fine tune these weights, so that the loss decreases through consecutive passes.

Initially, the calculated loss is expected to be very high and it is expected to decrease to a minimum after many (but fixed) times of forward/backward passes. In the end, it is hoped that the network should be trained well enough, so that the weights of the layers are tuned correctly. Then we run testing to check whether the CNN model works. We should have a different set of images plus its respective labels and pass the testing set of images through the CNN. We compare the outputs to the testing set to see if and how well the network works! Naturally the more data we have at our disposal, the better the model could be tuned through training and testing. That is why big data enables deep learning to be applied. Once we have obtained a good enough model, it is ready to be used for real life scenarios, even though we still continue to tune the model.

Another real-life example of computer vision has recently taken place in China. Alibaba launched City Brain System in its birthplace in Hangzhou, Zhejiang, China, where an AI center optimizes the traffic controls. Much of CNN like algorithms have already dominated our daily life: Facebook - automatic tagging, Google - photo search, Pinterest - home feed personalization.

Recurrent Neural Networks. It is also known interchangeably as recursive neural network. This is just a generalization of a recurrent network while having the same acronym. An RNN simply uses previous input sources within the calculations. Suppose that we are analyzing handwriting; we can predict words and future letters much better if we remember the previous letters. Another way to think about RNNs is that they have a "memory" which captures information about what has been calculated so far. RNN can remember the former inputs, which gives them a big edge over other ANN, when it comes to sequential and context-sensitive tasks such as speech recognition.

RNNs are the most powerful model for NLP. RNNs are also used for language translations, composing music, writing novels, writing Shakespearean poems, or composing non-existent Beatles songs.

Generative Adversarial Networks. GAN was introduced by Ian Goodfellow, who is a research scientist at Google Brain, and his associates from the University of Montreal in 2014. Yann LeCun, the director of Facebook AI said: Generative Adversarial Networks is the most interesting idea in the last ten years in Machine Learning. GAN makes the neural nets more human by allowing it to create instead of just training it with data sets.

A GAN is composed of two NNs: (i) a generative network and (ii) a discriminative network. In the starting phase, a Generator model takes random noise signals as input and generates a random noisy (fake) image as the output. Gradually with the help of the Discriminator, it starts generating images of a particular class that look real. The Discriminator which is the advisory of Generator is fed with both the generated images as well
Neural Networks That Cling to the Past

Figure 4.6: Source: https://towardsdatascience.com/understanding-recurrent-neural-networks-the-preferred-neural-network-for-time-series-data-7d856c21b759

as a certain class of images at the same time, allowing it to tell the generator how the real image looks like.

Figure 4.7: Source: OReilly.

Generator and Discriminator are pitted against each other (thus the word ”adversarial”) and they compete each other during the training phase, where their losses push each other to improve the performance (via backpropagation). The goal of the generator is to pass without being caught, while the goal of the discriminator is to identify the fake images. After reaching a certain point, the Discriminator will be unable to tell if the generate image is a real or a fake image, and that is when we can see images of a certain class (class that the discriminator is trained with) being generated by out Generator that never actually existed before! Experts sometimes describe this as the generative network trying to ”fool” the discriminative network, which has to
be trained to recognize particular sets of patterns and models. GAN could be used to increase the resolution of an image, recreate popular images or paintings, or generate an image from text, produce photo realistic depictions of product prototypes, generate realistic speech audio of real people and produce fashion/merchandise shots. GAN was used to create the 2018 painting Edmond de Belamy, which was sold for $432,500. Lastly, GAN is also very popular in social-media. But watch out for deepfake videos. If you feed it with enough faces data set, it can create completely new fake faces that are very realistic but in reality do not exist!

5 Natural Language Processing

Recently, (DL) has led to remarkable results in diverse fields, such as image recognition and Natural Language Processing (NLP). (DL) algorithms are capable of discovering generalizable concepts, such as encoding the concept of a “car” from a series of images. An investor might deploy an algorithm that recognizes cars to count the number of cars in a retail parking lot from a satellite image in order to infer a likely store sales figure for a particular period. (NLP) allows computers to read and produce written text or, when combined with a voice recognition, to read and produce spoken language. This allows firms to automate financial service functions previously requiring manual intervention.

(NLP) encompasses a broad list of things. It has become increasingly popular in recent time primarily due to the explosion of ML applications. NLP captures the ability of computers to analyze, understand and generate human language, including speech. For example, we can perform a sentiment analysis given any text. NLP can make AI recommendations after parsing through movie/book reviews or web. NLP can also run chatbots/digital assistants for front end tasks using text or audio interactions. Alexa/Siri/Cortana/Google Assistant are the famous digital personas using NLP engines.

![The Artificial Intelligence (AI) behind Chatbots](image-url)
The next stage of NLP is a natural language interaction, which enables people to communicate with computers using everyday language to complete given tasks. Google CEO, Sundar Pichai, showed how the Google Assistant can make a few calls and book a haircut appointment for you. Other known applications include enterprise search or opinion mining (sentiment analysis). There is already a large number of choices of NLP engines that are available to be embedded into an everyday usage whether it is call centers, chat-bots, translators, auto-predictors, spam filters or the new vast domain of digital assistants.

6 Conclusion

How does the overview presented in this note fit in the field, known as Data Science (DS)? Nisarg Dave’s image below illustrates, very well, the inter-disciplinary nature of DS. DS is at the intersection of all of these diverse fields. Data scientists need to have multi-disciplinary skills to be able to create a data set to test, create the code needed for the learning algorithms and, in the end, deliver an innovative business insight into it.

Figure 6.1: [https://www.nisargdave.com/blog/interdisciplinary-data-science](https://www.nisargdave.com/blog/interdisciplinary-data-science)

It is important to stress what kind of tasks AI is unable to perform, such as drawing causal inference. Generally speaking, ML algorithms are used to identify patterns that are correlated with other events or patterns. Thus the patterns that AI identifies are in the form of correlations, some of which are unrecognizable to the human eye. However, AI applications are being used increasingly to help understand complex relationships, along with other tools and domain expertise.

Many applications tend more toward augmented intelligence, or an augmentation of human capabilities, rather than a replacement of humans. Even as advancements in ML continue, including in the area of DL, most
industries are not attempting to fully replicate human intelligence. As noted by one industry observer, "... a human in the loop is essential: we are, unlike machines, able to take into account context and use general knowledge to put AI-drawn conclusions into perspective."

It is also important to stress that ML and DL, just like DS in general, is as much an art as a science. When we start working on an AI project, we tend to focus more on data sets and models/methods. It may be a good idea to choose a favorite ML domain and go deeper into it. The proficiency acquired for an AI project only comes with practice just like everything else in life.

We will see many more news and inventions in AI domain in 2019 and beyond. We also will witness many more advances and applied cases at the "edges" for ML to be available with mobile phones, ear buds, watches and other portable devices beyond just high power computers per se. Satya Nadella, the CEO of Microsoft, said so aptly that I believe in a world that will have an abundance of artificial intelligence, but what will be scarce is real intelligence and human qualities, like empathy. I think great innovation comes from the empathy you have for the problems you want to solve for people.