**Assignment 2 – IT Technologies**

**Machine learning**

Although most agree Machine learning is a branch of Artificial Intelligence, definitions and explanations of machine learning vary, including:

*Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed****. Machine learning focuses on the development of computer programs****that can access data and use it learn for themselves*. (Expert system 2019), and

*Machine learning is a method of data analysis that automates analytical model building. It is a branch of* [*artificial intelligence*](https://www.sas.com/en_au/insights/analytics/what-is-artificial-intelligence.html)*based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.* (SAS 2019)

But what does learn mean for computers, is it the same as learning for humans? The Macquarie Dictionary defines learning as:

*Learning is to acquire knowledge of or skill in by study, instruction or experience (Macquarie 2010)*

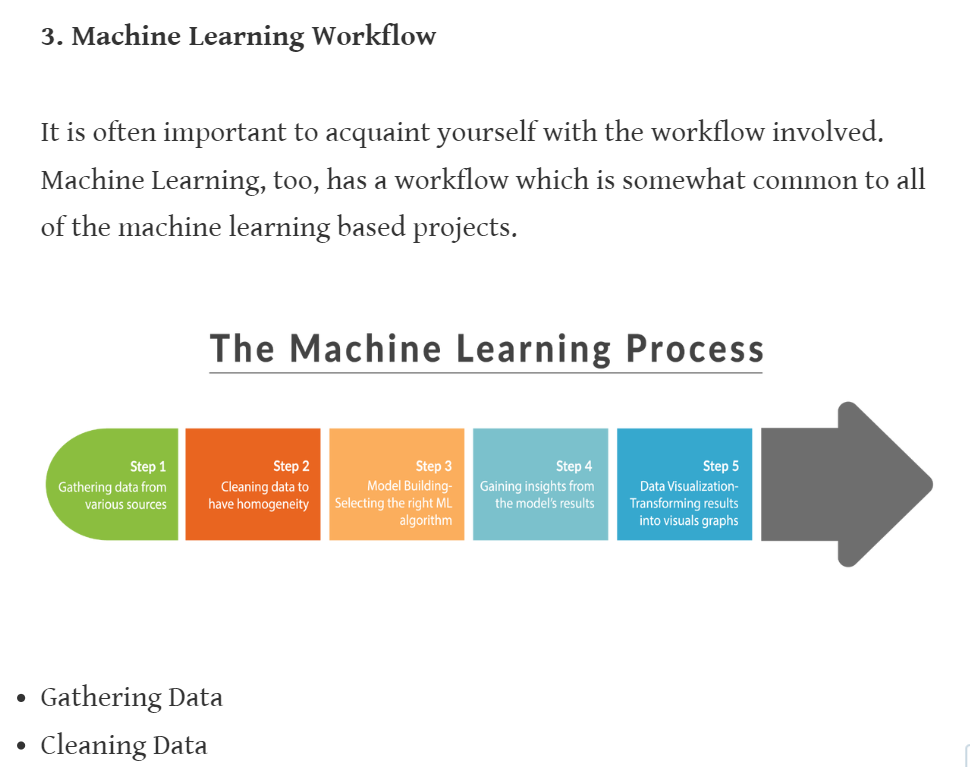
Is this what it means for computers? How would we know when a computer has knowledge or a skill? Afterall, you can receive instruction without benefiting from it at all if it can’t be applied. It seems that learning in this context is more linked to performance and ability to apply learnings rather than building knowledge (Witten 2017).

This learning occurs by using [algorithms](https://en.wikipedia.org/wiki/Algorithm) and [statistical models](https://en.wikipedia.org/wiki/Statistical_model)  in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. (Wikipedia 2019) These algorithms build mathematical models using sample data, known as training data, and they are able to adapt without being specifically programmed to do the task (Wikipedia 2019). They learn from previous computations to produce reliable, repeatable decisions and results. (SAS 2019)

This sample data needs to include robust, inclusive samples, otherwise outputs can contain bias that can lead to misleading information that may advantage or disadvantage. Datasets used to train AI systems must be appropriate for the system they are used to train (Dawson - CSIRO 2019). Data use must be ethical and useful and must not allow unfair bias, breaches to privacy or security and be of sufficient quality and relevance for input into algorithms (Dawson - CSIRO 2019). If the data that is input is not appropriate, then the outputs will not be reliable.

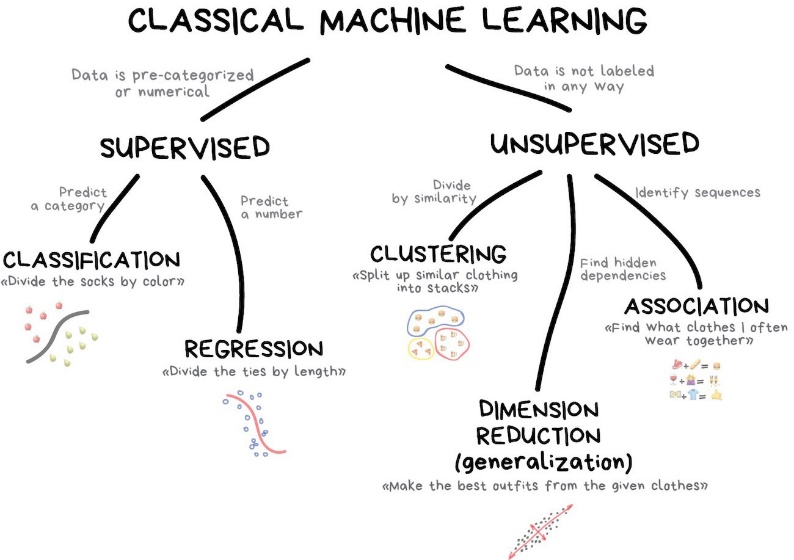
**The Machine Learning Process**

The following are the steps common to most machine learning based projects:



Python tips, *Introduction to Machine Learning and its Usage in Remote Sensing*, <https://pythontips.com/2017/11/11/introduction-to-machine-learning-and-its-usage-in-remote-sensing/> viewed 7/7/2019

As illustrated above, the first two steps are about gathering the data and cleansing it to ensure that it is complete, accurate, consistent and unique. This data is then used to build the model to select the right algorithm in the step 3, that will then be used in step 4 to gain insights which can be done in a number of ways. Two of the most widely adopted machine learning methods are **supervised learning** and **unsupervised learning,** but there are other types such as **semi-supervised** and **reinforcement learning** which I will also examine (SAS 2019).



VAS3K, 2019, *Machine Learning for everyone*, <https://vas3k.com/blog/machine_learning/>

**Supervised learning**algorithms are trained using labelled examples, such as an input where the desired output is known. (SAS 2019) As the example above illustrates using clothes. The clothes are sorted using a decision tree based on attributes known before ie socks labelled by colour - is it green or red, and ties by length – long or short. Often used for diagnostics, medicine and finance where historical data predicts future events (Vas3k 2019). Using methods like classification and regression, supervised learning uses patterns to predict the values of the label, that can then be used on additional unlabelled data, (SAS 2019) also known as data mining.

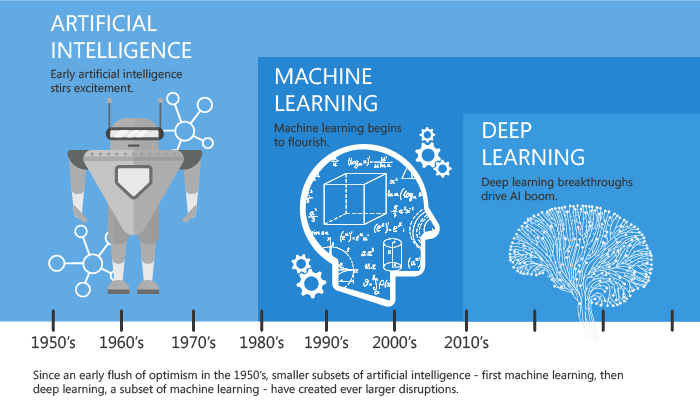
**Unsupervised learning**is used against data that has no historical labels (SAS 2019). Following on with the example above, the clothing is not labelled at all, so you’re dividing up socks when you don’t know what colours you have (VAS3K 2019). The goal is to explore the data and find some groupings or structure within it (SAS 2019). It works well for grouping data with similar attributes or identifying the main attributes that differentiate data, such as in marketing campaigns.

**Semi-supervised learning**is used for the same applications as supervised learning. But it uses both labelled and unlabelled data for training, typically a small amount of labelled data with a large amount of unlabelled data. This type of learning can be used with methods such as classification, regression and prediction. Semi-supervised learning is useful when the cost associated with labelling is too high to allow for a fully labelled training process. Early examples of this include identifying a person's face on a web cam (SAS 2019).

**Reinforcement learning** is not related to assessing a data set or sets, it is carried out by processing the deluge of data from an environment, like those encountered by autonomous cars (VAS3K 2019). Using reinforcement learning, the algorithm discovers through trial and error what actions produce the greatest rewards (SAS 2019). The goal is to minimise the error rate, accidents or injuries in the autonomous car example, not predict everything (VAS3K 2019). This type of learning has three primary components, the agent (the learner or decision maker), the environment (everything the agent interacts with), and actions (what the agent can do). The objective is for the agent to choose actions that maximize the expected reward over a given amount of time (SAS 2019).

**Machine learning over time**

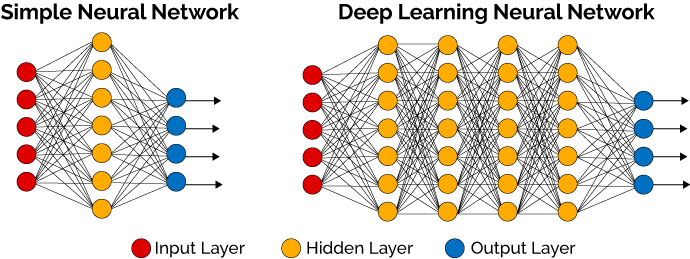
As you can see this has been around since the 1980s so what is behind the lastest surge in popularity? New computing technologies are transforming machine learning so it is not like machine learning of the past.



*The evolution of artificial intelligence, machine learning, and deep learning. (Image source: Tandon 2016)*

Deep learning is a subset of machine learning where artificial neural networks, algorithms inspired by the human brain, learn from large amounts of data. (Marr 2018). Deep learning combines advances in computing power and special types of neural networks to learn complicated patterns in large amounts of data (SAS 2019).

Neural networks are computing systems with interconnected nodes (illustrated by red and blue input and output points below) that work much like neurons in the human brain (SAS 2019). A Neural network is made up of a number of interconnected nodes and hidden layers. A hidden layer is any layer between the input layer and output layer and there can be multiple hidden layers (Zhou 2019). Using algorithms, neural networks can recognise hidden patterns and correlations in raw data, cluster and classify it, and, over time, continuously learn and improve. (SAS 2019)



# *Favio Vazquez, 22 December 2017, Deep Learning made easy with Deep Cognition in*[*Becoming Human: Artificial Intelligence Magazine*](https://becominghuman.ai/) *viewed 10 July 2019*

These are being used for virtual assistants like Alexa or Siri or Cortana, translations, autonomous cars, chatbots, facial recognition, personalised marketing and medical diagnosis, the list goes on (Marr 2019). Amazon used deep learning to train its software to analyse medical records. And according to Taha Kass-Hout, *the software performed as well as or better than similar programs. It was also able to pull out data regarding patients' illnesses, prescriptions, lab orders and procedures, all of which is organized into a spreadsheet-like report.* (Locklear - Engadget 2018). This work is currently done by around 60 people, so this advance has the potential to replace them in pulling data from approximately 500,000 patient records.

The potential applications seem limitless as deep learning algorithms continue to perform better as they gain more experience, however, it is good to keep in mind that ethical principles should be followed when developing and using them. These should include ensuring benefits are greater than costs, they cause no harm or disadvantage, they’re legal, protect privacy, transparent, contestable and creators are accountable for impacts of algorithms.

**REFERENCES**

* Bernard Marr (Forbes 2018), viewed on July 2019, *What Is Deep Learning AI? A Simple Guide With 8 Practical Examples*, <https://www.forbes.com/sites/bernardmarr/2018/10/01/what-is-deep-learning-ai-a-simple-guide-with-8-practical-examples/#4faa26308d4b>
* Expert system, July 2019, *What is Machine Learning? A definition* <https://www.expertsystem.com/machine-learning-definition/> viewed 9 July 2019

# Favio Vazquez, 22 December 2017, viewed 10 July 2019, *Deep Learning made easy with Deep Cognition* in [Becoming Human: Artificial Intelligence Magazine](https://becominghuman.ai/) <https://becominghuman.ai/>

* Locklear M, Engadget 2018 <https://www.engadget.com/2018/11/27/amazon-software-mine-medical-records/>
* The Macquirie Library Pty Ltd, 2010, The Macquarie Dictionary, Third edition
* Python tips, 2017, viewed 10 July 2019, *Introduction to Machine Learning and its Usage in Remote Sensing*, <https://pythontips.com/2017/11/11/introduction-to-machine-learning-and-its-usage-in-remote-sensing/>
* SAS, 2019, viewed on 10 July 2019,*Machine learning – What is it and why it matters*, [https://www.sas.com/en\_au/insights/analytics/machine-learning.html](https://www.sas.com/en_au/insights/analytics/machine-learning.html%20)
* (SAS 2019), 2019, viewed on 10 July 2019, Neural Networks: What they are and why they matter <https://www.sas.com/en_au/insights/analytics/neural-networks.html>
* Tandon, K. (2016) AI & Machine Learning: The evolution, differences and connections <https://www.linkedin.com/pulse/ai-machine-learning-evolution-differences-connections-kapil-tandon/>
* VAS3K, 2019, viewed 7 July 2019, *Machine Learning for everyone*, <https://vas3k.com/blog/machine_learning/> viewed 7 July 2019
* Wikipedia, 2019, viewed on 10 July 2019, <https://en.wikipedia.org/wiki/Machine_learning#cite_note-bishop2006-2>
* Witten H, Frank E, Hall M, Pal CJ, 2017, viewed 10 July 2019, *Data Mining: Practical Machine Learning Tools and Techniques* 4th edition, <https://books.google.com.au/books?hl=en&lr=&id=1SylCgAAQBAJ&oi=fnd&pg=PP1&dq=machine+learning&ots=8IFLpcnyta&sig=d7Yxy-RNub2pon9VtYT_PiaUkAk#v=onepage&q=machine%20learning&f=false>
* Zhou V, March 2019, viewed 14 July 2017, *Machine Learning for Beginners: An Introduction to Neural Networks, viewed 14 July 2017*, <https://towardsdatascience.com/machine-learning-for-beginners-an-introduction-to-neural-networks-d49f22d238f9>

**Natural language processing**

Natural language refers to the way we, communicate with each other through speech and text. (Brownlee 2017). While the processing of natural language, Natural Language Processing (NLP), is not a new science, the technology is rapidly advancing thanks to an increased interest in human-to-machine communications, plus an availability of big data, powerful computing and enhanced algorithms. (SAS 2019)

Moore, an AI and Language Analytics Strategist, SAS describes Natural language processing as:

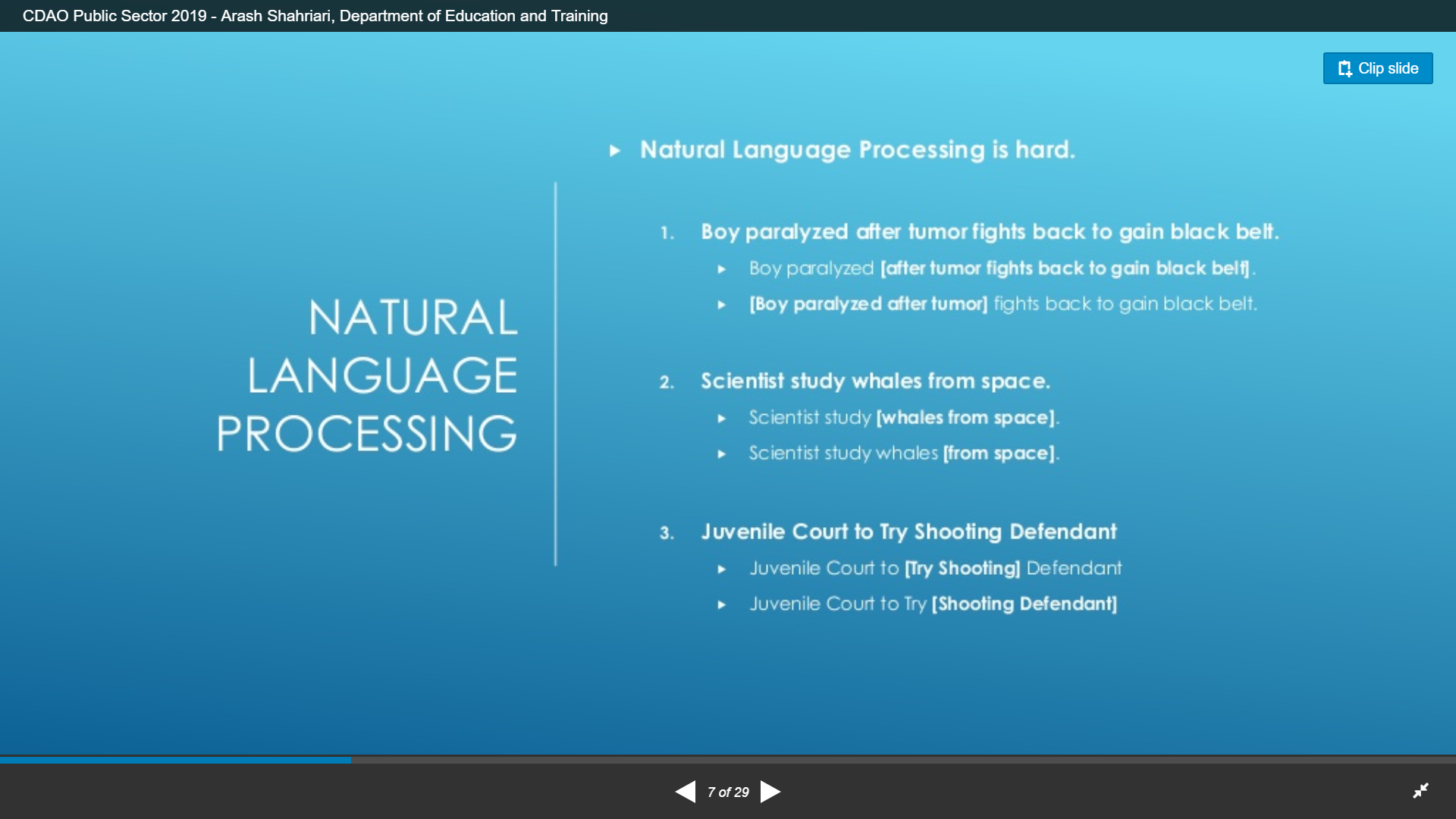
# … *a branch of artificial intelligence (AI) that helps computers understand, interpret and manipulate human language. In general terms, NLP tasks break down language into shorter, elemental pieces, and tries to understand relationships among those pieces to explore how they work together to create meaning. The combination of NLP, machine learning and human subject matter expertise holds the potential to revolutionize how we approach new and existing problems (Moore 2019)*.

Given the volume and importance of this type of data, we must have ways to understand and harness the value of natural language, as we do for other types of data. (Brownlee 2019).

But Natural Language Processing is difficult because of the:

* complexity of representing, learning, and using
* ambiguity in meaning
* interpretation needs real world experience, common sense, and contextual knowledge. (Shahriari 2019)

As the example below illustrates, depending on how the language is broken down into smaller pieces can dramatically affect the processing of its meaning:



Shahriari, A, viewed 11 July 2019, *Natural Language Processing in practice: Facts and fiction*, <https://www.slideshare.net/Chief_Data_Officer_Forum/cdao-public-sector-2019-arash-shahriari-department-of-education-and-training?>

Computational modelling of human language requires an understanding of:

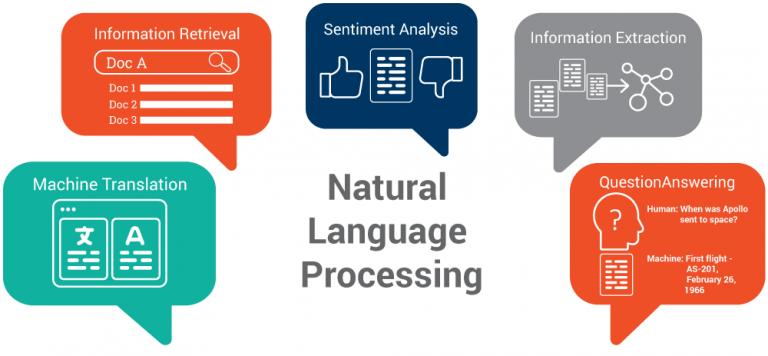
* Morphology: structure of words
* Syntax: the way words are used to form phrases
* Semantics: construction of meaning by syntax
* Discourse: meaning in context (Shahriari 2019), and
* a general knowledge about the world.

Acquiring and encoding this knowledge is one of the fundamental impediments to developing effective and robust language systems (Ritkov 2005).

Moore explains that to be able to make sense of all this information requires a combination of three capabilities:

* *Natural language processing - Performs linguistic analysis that essentially helps a machine read text. It analyzes and converts text into form representations for text processing and understanding. This includes methods such as tokenization, part of speech tagging, stemming, named entity recognition and more.*
* *Machine learning - Once NLP has been applied to text, machine learning uses the output for data mining and machine learning algorithms to automate the generation of key insights and descriptive analytics.*
* *Human input - When it comes to analyzing text, human input is still incredibly important. Subject matter expertise is applied in the form of linguistic rules to help the machine capture slang, detect sarcasm and provide relevant context.*

The technology retrieves information to analyse unstructured text, then actively learns from the data through sentiment analysis and machine translation along with human direction to generate new insights. The purpose of this is to build and deploy text analytics models that enable understanding through topic detection, contextual extraction, document categorisation and sentiment analysis (SAS 2019).



Ontotext, 2017, Top 5 Semantic Technology Trends to Look for in 2017, viewed 13 July 2019, <https://www.ontotext.com/top-5-semantic-technology-trends-2017/>

**NLP and Deep Learning**

As illustrated in Shahriari’s example above, machine learning algorithms traditionally use well defined fixed-length inputs and outputs. But they cannot work with raw text directly; the text must be converted into numbers (specifically, vectors of numbers). This is called feature extraction or feature encoding, and this is one of the key areas where deep learning is challenging statistical methods with singular and simpler models. For this reason new methods are starting to outperform traditional statistical methods (Brownlee 2019).

The following are three examples of what deep learning is capable of in the in the field of natural language processing: (Brownlee 2019)

#### Automatic Image Caption Generation - is the task where, given a photograph, the system must generate a caption that describes the contents of the image.

#### Automatic Translation of Text - is the task where you are given sentences of text in one language and must translate them into text in another language.

#### Automatic Text Classification - is the task of assigning a class label given a text document such as a review, tweet, or email. (Brownlee 2019)

From these examples you can see that developing systems capable of these tasks would be valuable in a broad range of domains and industries. (Brownlee 2019)

## ****Chatbots****

Chatbots use deep learning capabilities to automate user interactions with websites and social media sites. They are designed to interact or converse with users either via text input or verbal conversation (Hubspot 2019). They aim to seamlessly create a fully online interaction, that is available anytime.

Sometimes though, users have unrealistic expectations of chatbots and expect that they will converse in the same way as another human would. While this is a goal, it is not possible to do this yet using current technology (Botpress 2019).

An example of how this technology can still go wrong was Tay the Twitter chatbot, which was developed by Microsoft. Tay learned how to communicate with Twitter users to better understand how AI interacts with users. However, Tay only lasted 24 hours before being taken offline. As the inputs from users contained abusive content, Tay learned and created responses that were offensive, sexist and racist. The data being input was not appropriate, parameters were not sufficiently developed to filter these out, therefore outputs were not appropriate. (Dawson - CSIRO 2019).

**Other applications**

As NLP can be applied to any situation that needs rapid analysis of unstructured data, the applications are incredibly diverse. This has amazing potential to identify and categorise the overwhelming amount of data currently stored in uncontrolled environments, as the volume variety and velocity of growth mean that this is no longer possible for humans to do.

Intelligently analysed data is a valuable resource as it can lead to new insights, better decisions making and in commercial settings, competitive advantages. Conversely, not understanding data holdings could compromise service standards, increase exposure you to reputational damage, loss of data, high discovery, litigation and management costs throughout its lifecycle. Although it is difficult to quantify the full cost of managing this data throughout its lifecycle, we can all relate to the pace of growth qualitatively, and the exponential growing risks.

**Looking ahead**

Future developments in NLP are occurring in the subfield of Natural language understanding (NLU). Its potential in cognitive and AI applications go beyond the structural understanding of language to interpret intent, resolve context and word ambiguity, and even generate well-formed human language on its own. These developments have a number of important implications but will first have to address the complex problem of semantic interpretation, to understand the meaning and subtleties of human language in many contexts (SAS 2019).

Inexpensive storage has made it too easy to postpone decisions about what to do with data. Simply getting more memory and keep everything has been the solution for far too long but lying hidden in all this data is information, potential useful information, that we rarely make explicit or take advantage of (Witten 2017). The need to address the growing volume and variety of unstructured data, more than ever, relies on Natural Language processing and Machine learning in order to manage it and harness its full potential (SAS 2019).

**References**

Brownlee, J, 22 September 2017, viewed *What is Natural Language Processing?* in Deep Learning for Natural Language Processing <https://machinelearningmastery.com/natural-language-processing/>

The Botpress, 2019, viewed on 13 July 2019, *The what’s & why of chatbots* <https://botpress.io/learn/what-and-why/>

Dawson D and Schlieger E, Horton J, McLaughlin J, Robinson C, Quezada G, Scowcroft J, and Hajkowicz, (2019) viewed on 10 July 2017, Artificial Intelligence: Australia’s Ethics Framework, Data 61 CSIRO, Australia. <https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf>

Hubspot, by Munroe, M 2019, viewed 13 July 2019, Understanding the Difference Between a Bot, a Chatbot, and a Robot, <https://blog.hubspot.com/customers/understanding-the-difference-between-a-chat-bot-and-robot>

Mitkov, Ruslan, 2005, Page 377,  [The Oxford Handbook of Computational Linguistics](http://amzn.to/2uHeERE).

Moore, M B, SAS, 2019, viewed 6 July 2019, The untapped potential in unstructured text, SAS <https://www.sas.com/en_au/insights/articles/analytics/the-untapped-potential-in-unstructured-text.html>

Ontotext, 2017, Top 5 Semantic Technology Trends to Look for in 2017, viewed 13 July 2019, <https://www.ontotext.com/top-5-semantic-technology-trends-2017/>

# SAS, 2019, viewed 10 July 2017, *Natural Language Processing (IoT)*, <https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html>

Shahriari, A, Department of Education and Training, viewed 11 July 2019, *Natural Language Processing in practice: Facts and fiction*, <https://www.slideshare.net/Chief_Data_Officer_Forum/cdao-public-sector-2019-arash-shahriari-department-of-education-and-training?>

Witten H, Frank E, Hall M, Pal CJ, 2017, viewed 10 July 2019, *Data Mining: Practical Machine Learning Tools and Techniques* 4th edition, <https://books.google.com.au/books?hl=en&lr=&id=1SylCgAAQBAJ&oi=fnd&pg=PP1&dq=machine+learning&ots=8IFLpcnyta&sig=d7Yxy-RNub2pon9VtYT_PiaUkAk#v=onepage&q=machine%20learning&f=false>