# ARTIFICIAL INTELLIGENCE TERMINOLOGY CAN BE MISLEADING: A FRAMEWORK FOR RATIONALIZING THE DISCOURSE

Kenneth R. Walsh, University of New Orleans

Sathiadev Mahesh, University of New Orleans

## ABSTRACT

Steven Hawkins and Elon Musk have both commented on the potential terror that could occur as machines develop true artificial intelligence powers. However, Wasserman shows that much machine learning is little different from age old statistical analysis, although supercharged by the latest computer technology. Terms such as artificial intelligence, machine learning, and deep learning can evoke emotions in the general public and in the political arena, inconsistent with the true state of the art. In this paper we debate whether a machine can truly learn and conclude that the more useful question is about the manner in which business practice and the legal environment permit a machine to operate autonomously within the decision context. In answering the latter question we develop a stage model of machine learning systems based on the decision level of the system governed autonomously by machine algorithms. The model provides a useful framework for discussion, understanding, and governance of machine learning systems and reduces the hyperbole that can follow loaded terms such as learning and intelligence.

*Keywords*: artificial intelligence, machine learning, scientific framework

## INTRODUCTION

Machine learning (ML) systems play a role in many facets of our daily lives. Not only do they direct us to preferred sites through search engines and match faces in our photos, but they also playing a role in health care diagnostics, insurance pricing, law enforcement, and employment decisions. Unlike the in-your-face approach of social news feeds which have aroused serious discussion at many levels of the community and government, most business applications of ML have replaced human decision makers in making many decisions that have serious consequences, with little press coverage and even less public attention. The consequences of machine learning algorithms can be lifesaving or life threatening, depending on the situation. Carr (2014) in his book, "The Glass Cage", describes the problems of highly automated environments which strip the human decision maker of any real work, reducing human learning and human alertness, creating cascading problems when the human is suddenly charged with making a decision. Diakopoulos (2016) notes that many news articles are written by machine learning algorithms and are often well written, but sometimes have single word errors that completely change the meaning of the article.

At the same time, the methods used in particular decisions may be opaque, giving little understanding to users about how a decision or classification was made (Burrell, 2016). Yarkoni and Westfall (2017) make the argument that in the field of experimental psychology, the emphasis on explanation, has reduced the predictive power of some machine learning techniques. A "black box" could fit data perfectly, while a well-explained model would only partially explain the data. While the causes of those predictions may lead to better explanation in the future, since knowledge is almost always built on top of current knowledge, at the present time, an explained model may provide a less useful, and less commercially valuable, predictive technique. Businesses focus on the commercial value of the ML model, and a good black box predictor which provides the best stock trade or optimal inventory policy is more suitable to the business than a well explained model with less short-term predictive power. Burrell (2016) identified three types of opacity stemming from the secrecy of the organization, the complexity of the algorithm, and interpretability by human decision makers. The secrecy of the organization is the desire to protect proprietary corporate methods, either for profit or political advantage. Statistics is a challenging field and most managers have only a limited knowledge of basic, often century old-methods. Only a few experts even understand complex models used in ML and even fewer understand the computer code used to process the data with the algorithm. Many ML tools yield black box models which are opaque to even the sophisticated user. The lack of interpretability by human decision makers renders the human decision makers useless and irrelevant in the process, leading to a surrender to the decision of the system. With such levels of opacity, it is more difficult for those affected by such systems to understand or dispute decision made by the systems. Garfinkel, Mathews, Shapiro, and Smith (2017) and Knight (2017) argue for the need for algorithmic transparency and accountability. The Defense Advanced Research Projects Agency has recognized the problem and initiated a program to study the discipline of explainable artificial intelligence (Gunning, 2016).

If machines are learning, in the manner of humans, then to what extent can people understand what they have learned? Terms such as artificial intelligence and machine learning may give the impression that these creatures are autonomous learners, however, many such systems fall short of this hyperbole. This paper opens with a debate as to whether or not machines can learn followed by a classification of machine learning levels, differentiated by the extent to which the computer systems are operating autonomously. Diakopoulos (2016) identified five broad categories of information that should be disclosed about machine learning systems. These areas will be used to inform our categorization. By classifying machine learning systems that have different components and levels of learning, more specificity can be given as to what aspects are technically opaque and what information could be provided to reduce their opacity when accountability is needed. The implication of the categories follows.

**Machine Learning is not Learning**

"Statistics is the science of learning from data. Machine Learning (ML) is the science of learning from data" (Wasserman, 2013, p. 2). Wasserman clearly pointed out the similarities between statistical analysis and machine learning and concluded they were largely the same. However, few would describe statistical analysis as learning. It would be more accurate to describe statistical analysis as a mathematical technique that humans use to identify patterns in data. A data-driven decision maker uses these patterns in decision making. We would therefore argue that machine learning is not about a machine learning either, but rather merely a different implementation of

statistical analysis. This leads to two important implications. First, many predictions of what machine learning can achieve are overblown and ill-informed and much that has been written about the implications of machine learning in various industries may not be realistic. Second, public policy is often shaped by word choice, and since learning represents one of the higher achievements of human beings, the use of the word "learning" has caused an over-reaction among policy makers on the ability of machines to completely replace human workers. An interesting corollary to this argument is whether we could take steps to actually get machines to learn as humans do.

An argument can be made that the user interface of machine learning gives the impression of learning because of how the output is used in practice. If a user of Google types a search term and receives a useful result, or verbally states a question to a home automation device, and receives a useful answer, they tend to be amazed by how well the system works. They may over-estimate the "smartness" of the machine to have learned so much and come up with the answer. Consider a different scenario where human statisticians work with scientists or business leaders to use data collected by either a targeted survey instrument or by surveillance equipment, follow the rules of accepted statistical analysis, and interpret and present the results. Often, a satisfied client will likely attribute the good work to the statistician rather than the computer software and hardware used to obtain the results. This attribution of skill to the human statistician happens primarily because the underlying computer system is not well understood by the user and the human is the interface through which the results are made available. Many user-focused applications using ML have a user-friendly interface, often designed with much thought given to user-stories, the script describing user interaction with the system. Often, this user interface (UI) is designed to be user-friendly to even an unsophisticated user. The major difference between the first and second scenarios is that in the first the machine directly communicates with the user, while in the second a human interprets and delivers the result to the users.

To take the Google search example further, the consumer may go to google the next day and re-enter the same query. If the results are better than the day before, the consumer may observe that the system is getting better. This improvement may have happened without the intervention of any human and so it may give the appearance that the machine has learned. However, the machine has just run the same algorithms or statistical analysis on a new dataset with more data. Updating a dataset and rerunning an analysis is a useful task for computers, but we would argue that automated data gathering is not enough to be called learning. Consider the scenario of the human statistician who returns the next day with one more day's data in the data set and provides even better results than the previous day. One could argue that the biggest leap in the new ML systems is that they have finessed the toolset to continually add new data and re-analyze larger data sets in a convenient manner, eliminating the delays and manual effort to add and re-analyze data. From this perspective, ML is just an automated tool within a normal statistical update processes.

A simple example of when statistical analysis and machine learning yield the same result would be a simple classification problem. Whereas some may call it supervised learning and others may call it classification, the results are mathematically equivalent. Wasserman (2013) describes how researchers in both statistics and machine learning are working on an extension of supervised learning call semi-supervised inference. When using ML to classify data, for example, in a classification of photos by background location, data needs to be initially classified, often by

human classifiers. This manually classified data provides the training set for the ML tool. In supervised learning, a large training set is used to train the ML tool, and the creation of this training set is a major expense. In semi-supervised inference, the set of manually classified data is analyzed along with a much larger unclassified dataset, using techniques of likelihood maximization to obtain better results than with only the manually classified data set. The tools are part of the repertoire of a sophisticated statistician; the difference is that ML automates the process of data gathering and analysis, and also provides the results in a decision impelling format, and may even complete the decision process automatically. An interesting extension, but just a more in-depth example of how analysis performed on a static data set, can be updated with a new data in a dynamic dataset, presenting the appearance of learning over time.

Much machine learning research is done on improving the algorithm and tackling certain problems that can arise. Often machine learning approaches use a large number of independent variables. Since so many input variables are used, the approach may uncover independent variables that are surprising. This is again the consequence of larger dataset and not real learning. A downside to using a large number of independent variables can be over fitting. Over fitting occurs when a model very accurately represents the training set but does not predict new datasets well because it is too highly influenced by idiosyncrasies of the training data set. Regularization techniques are used to prevent such over fitting. For example, the dropout regularization technique is a method of dropping random nodes when training a neural network to prevent large scale neural networks from over fitting (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014).

The term deep learning may conjure up the idea that something is being learned. Many stunning improvements in natural language recognition and translation have been attributed to deep learning. However deep learning simply refers to a more complex version of machine learning or neural networks, often implemented as several hidden layers in a neural network design. The benefits have accrued from better hardware, massive datasets, and algorithmic improvements in the weights and processing within the neural network (Monroe, 2017). "Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction" (LeCun, Bengio, & Hinton, 2015, p. 436). By this definition, deep learning is an extension of the concept of the neural network that has been afforded business and science through increases in computing power. Deep learning neural networks can have many layers and many nodes. A principle advantage of handling many nodes and layers is for machine learning systems to be developed without the often time consuming and largely manual stage of feature engineering. With many forms of ML, real world data is not in the right format for the input to the system as features. Feature engineering creates features out of real world data. Deep learning neural networks can use a large number of input nodes to access real world data directly while using many layers to effectively extract features on its own. As such, deep learning may not strictly be considered learning, but it does automate a time consuming human tasks in many ML projects. Chainer is an open source framework for deep learning models (Tokui, Oono, Hido, & Clayton, 2015). Blocks and Fuel are software applications that help deal with complexity of deep neural networks having many layers and nodes as well as large data sets (Van Merriënboer, 2015).

## Machine Learning is Learning

What is learning? Consider the case of a hunter-gatherer tribe twelve millennia ago with no knowledge of farming. They find that a particular variety of grass provides seeds which are edible. The learning process in this case involves trial and error, where the tribe tries out many different seeds and finds one that is edible and available for easy foraging. They then determine visible patterns in this type of grass and use them to identify the grass which yields this type of seed. Later, maybe in order to avoid having to forage for this plant, or by accident because they carry the seed home frequently, they find that the plant grows in their vicinity and can be cultivated to enhance yields. There are many learning steps in the process of moving from a hunter-gatherer group to a farming community. A trial and error process leads to understanding how to detect one type of grass which yields wheat from another that does not provide a healthy yield. Communication between members of the tribe speeds up the learning process.

One major aspect of learning is the learning of a society rather than the learning of an individual. When viewing the historical arc of research, we often jump from one discovery to another, often highlighting the "winners" in the intellectual arms race. In reality however, the words of Isaac Newton (1675, p. 1) hold true, "If I have seen further, it is by standing on the shoulders of giants." Learning typically develops in a community where the successes and more often the failures of others are used to work toward a new concept. Machines can learn almost instantaneously from other machines, since the most successful neural network's weights, identified clusters, or detected rules from a decision forest can be immediately replicated across multiple machines. Just as viral concepts rise and flow rapidly through internetworked social media, multiple networked machines can quickly learn from one another. Hence, ML can be learning if the machines learn from one another rather than working in isolation.

Consider adaptive exponential smoothing (McClelland, 1971), a computer-based time series forecasting approach developed in the early 1970s. Exponential smoothing developed in the 1950s used a smoothing parameter and the value of this parameter was selected to provide the best fit to the data. However, if the parameter selected at one point in time was no longer optimal, the forecast error grew and the predictions became unreliable. A correction mechanism using computer code to re-calculate the optimal value of a smoothing parameter when errors exceeded a threshold made the forecasts much better. Can this be called a learning system? It is trivial by the standards of today's machine learning tools, but it clearly encapsulates the learning process. The approach uses a simple model, determines the error, and if the cumulative error exceeds a threshold, it launches a corrective action. We may not term this intelligent in a machine, and of course the knowledge is limited to one type of time series forecasting, but we have the essence of a learning system built into the program. Machine learning is a form of learning if it has a built-in feedback mechanism which monitors the outcome of its actions and takes appropriate corrective actions to its module when the outcomes are deemed undesirable.

Intelligence is not merely in computational complexity but also lies in the interface. If we make an entry in a cell formula in a spreadsheet and the intelligent formula advisor signals an error, it is clearly more intelligent than a formula entry mechanism which blindly accepts the bad formula and merely fails to calculate a result. As the formula advisor in a workbook becomes more capable and provides a wider range of advice, it is perceived as smarter. However, if it repeats the same

error statement repeatedly, users term it a "dumb machine", and seek human help. How do humans exhibit intelligence in this scenario? An intelligent human advisor, will not monotonously repeat the same error statement multiple times. After repeating the error statement, maybe a couple of times for emphasis, the error statement will be rephrased. Failure at this step in getting the user to enter the correct formula will open up a more detailed analysis of the error and better targeted advice to the user to help fix the problem. This is the adaptability expected of a human instructor who learns about the problem faced by the user after interaction. Machine learning is a form of learning when the system does not repeat itself interminably, and monitors a log of its actions to ensure that alternate approaches are tried out. If the entire logic of the action sequence is pre-programmed, then it does not constitute learning. The system should have basic rules of behavior about what constitutes rational behavior and what constitutes stupidity.

An aspect of learning is to get better over time, by the repeated exercise of cognition. One part of this learning is increased hand to eye coordination gained by repetitive action. While the motor abilities of robots have been continually improved, they still face challenges opening a door, climbing a flight of stairs, or grasping a soft item without crushing it. Another part of this learning is the development of mental maps of situations in the brain of the decision maker. These mental maps help the decision maker make sense of data that deviates from the norm, such as a scene which does not fit the known profile of a route. While human drivers often use landmarks to assist in navigation, automated navigation tools often use a strictly algorithmic, turn by turn signaling approach which requires constant attention to, and complete dependence on the system's instructions. Human drivers may either surrender to the system or undertake the intense cognitive effort to maintain a mental map of their location, in addition to following the turn-by-turn instructions. An approach to make the system more intelligent incorporates landmarks in the directions and creates and maintains maps of these landmarks in a manner similar to human drivers (Zhu & Karimi, 2015). Machines can learn if they can create and maintain multiple higher-level maps of the data. Regression software merely calculates the vector values of the data and does not understand the variables, and this applies to the most sophisticated statistical tools in ML toolsets. However, if the variable names are mapped to real-word objects in a map of images, their physical locations, and assigned names, and other ML tools connect the data to the objects, as well as to other prior data analysis, we can have real learning.

The human body is highly adaptable and the muscles become more adept at physical activities by training. The body and brain often re-allocate resources to meet the needs of the environment. Repeated mathematical training makes the student better at math while physical training on a wood-working tool makes the human a better carpenter. The way robots are designed today, they are incapable of transferring their capabilities to different parts to get better at what they do. A CPU which may be as good at math on the day it is fabricated in a chip plant as it will be years later, may also deteriorate or fail completely, but it will not improve. Hardcoded software has the same problem; it does not learn. The only part of the system getting better is that of pattern recognition systems which improve as they get more data, and are corrected when they make errors. However, they too do not learn anything new unless they place their results in a mental map that displays the connections between the data.

Rather than recognizing patterns by mere trial and error training of pixelated images, multi-modal neural networks generate pattern similarity scores and assign text to describe the image and its

context. The combination these two modes, text descriptors of the object and its context and image patterns yield better recognition of objects than mere pattern similarity. In essence, this process not only recognizes the pattern in an image but also assigns text descriptors to the objects in the image and uses both to make a recognition. Stores of these mental maps have enhanced the performance of robots in many areas (Di Nuovo, De la Cruz, & Marocco, 2013).

When ML is applied to large data sets, the approach is not very different from traditional statistical analysis. Data is collected, cleaned, and then processed through dimensionality reduction, clustering, and regression to find patterns in the data which can be used to explain and predict human behavior. Of course, the availability of ML tools makes it easy to set up data acquisition and cleaning processes to repeatedly access and clean the data, and in fact to set up the entire statistical sequence to generate classifications of targeted customers or predictions of human behavior. For example, Microsoft's Azure ML toolset has models for Regression, Logistic Regression, Boosted Decision Trees, Random Decision Forest Algorithm, Support Vector Models, and Neural Networks, all readily available in drag and drop format. The availability of larger datasets ensures more frequent use of split data to test the model, and the availability of automated data processing tools supports the testing of data on multiple models to select and use the model providing the most effective classification. In addition, tools to monitor performance, can be linked back to the model to modify it when the outcomes are unsatisfactory.

What about the new tools available through the cloud which have learned to detect objects from videos, speech, or even emotions? These tools provide knowledge gained from a vast pool of training data that can be applied to the data available to a researcher. This is different from the statistical tools described earlier, which encoded knowledge developed to recognize patters in data, but did not include the knowledge gained from using these tools. To summarize, when we use factor analysis, or logistic regression, or even an ANN tool, we do not use any of the knowledge gained from prior use of the tool. SPSS's regression has been used millions of times, and it has not become any better because of prior use.

However, when a cloud-based voice recognition or object recognition tool is used, it becomes better and the dominant tools in the marketplace become superior to human classifiers. This improved ability of cloud based apps such as Google's Tensor Flows, Apple's Core ML or Amazon's Polly, or Microsoft's emotion detection API offer continually improving ability to identify objects in images, detects specific types of video and add labels to identify content as well as to detect changes in scenes and content, to detect speech and provide either a transcript from audio or generate speech from text, and to detect emotions from images and videos. This is where the real machine learning is taking place, i.e. where machines are becoming more capable than humans at certain tasks and available at next to no marginal cost. When individuals and businesses use this learning, and connect it to their in-house systems, we have machines with greater capability than many humans.

## LEVELS OF MACHINE LEARNING

One way to better understand the extent to which machines are learning is to classify them by what steps in the process are automatically generated by the computer versus what part of the process involves decisions made by the machine learning analyst (MLA). For example, in a simple system,

an MLA may choose the design of the neural network and define the number of layers and nodes at each layer as well as the aggregation system. In a more automated and hands-off system, the ML system will split the data, run different models and decided on whether or not to use a network and select the parameters of the network. Knowing what the computer is choosing autonomously helps understand how the term learning is being used and what strengths and weaknesses might occur in practice.

We propose a set of levels of increasing machine autonomy to classify machine learning systems. In general the levels are increasing orders of autonomy and would indicate a system with greater ability to learn.

A Level 1 ML system would be the most basic type of ML system. The MLA would choose a training dataset and type of model and run an algorithm to determine the model parameters. The algorithm may have been coded by the MLA or may have been provided in an ML library. At this level, ML is being used in much the same ways as it would be in traditional statistical analysis. The predictions based on new data may be presented to the end user either by the MLA or directly embedded in the computer system. In traditional statistical analysis, often the analyst presents the finding personally or in a written report to decision maker. In the ML community, the results may be embedded in a computer application.

A Level 2 ML system is a Level 1 system where the training dataset is automatically updated and the model parameters are recalculated. This may be one of the most common forms of systems colloquially referred to as learning. In such a system, the computer collects new observations which improve the training dataset. The model parameters are recalculated with each update to the training dataset or at a periodic interval. The methodology can allow a system to be deployed in practice even without adequate training data if sufficient data is expected to be forthcoming. This level can appear to an end user as learning because the system may initially make weak decisions and improve those decisions over time.

A Level 3 system is one in which the computer autonomously chooses the type of model. Many ML models exist ranging from simple regression to decision trees and multilayer neural networks. In a Level 3 system, the ML compares all models at its disposal and choose the model best suited to the data. In such a system, the MLA is responsible for providing the range of available models. A Level 3 system may or may not have a training dataset that is updated over time such as in Level 2. If such a dataset were updated, then a Level 3 system could choose different models over time as well as the parameters for such models.

A Level 4 system chooses its own input regularization method. Regularization is used to keep model from overfitting the training data and a regularization method is often selected by the MLA. In a Level 4 system, the computer chooses the regularization method and parameters from those at its disposal.

A Level 5 system conducts its own feature engineering based on raw real-world data. Deep learning systems usually fall into this category whereby raw data can be fed directly to the systems inputs.

A Level 6 system chooses its own feature set. Typically ML systems are given training data by the MLA. A Level 6 system may or may not have an initial dataset. The Level 6 system will search the data at its disposal on internal networks to the organization or the public Internet to find feature sets that are useful to the desired outcome. Over time, features may be added or dropped by the system.

In a Level 7 system, results of machine learning prediction are implemented without human oversight. For example, many ML systems provide recommendations to humans for decision making. However, one distinguishing characteristic of ML systems from previous statistical techniques is technical suitability for making embedded systems that can automatically act upon their results. In simple information retrieval tasks, no human stands between the computer and end user in displaying search results, however, in medical diagnosis, this is still an important step. If an artificial intelligence system were connected to an intervening drug delivery system administering pain medication, it may in fact do a better job than medical personnel while it may also have consequences for errors. Note that this characterization of a Level 7 system could be combined with other levels and does not strictly follow Level 6.

A Level 8 system chooses the outcome objective. For example, in most systems, the MLA choses and objective such find the modest relevant document of the shortest path. In a Level 8 system, the computer automatically choose an objective. In an organization, it might have the power to hire or fire people in a certain job description. A Level 8 system might choose not only how a self-driving car will get to its destination, but also what the destination should be. Table 1 summarizes the levels identified.

Table 1. Levels of Machine Learning

| **Level** | **Computer Learning** | **MLA Oversight** |
|---|---|---|
| 1 | Choose model parameters | Choose Model, Choose training dataset |
| 2 | Update training dataset and update parameters | Choose Model, Choose training dataset |
| 3 | Choose Model | Provide Available Models, Choose training dataset |
| 4 | Choose input regularization | |
| 5 | Automated or no feature engineering | Provide raw data source |
| 6 | Find new training data | |
| 7 | Results acted upon without human intervention | |
| 8 | Set Objective of ML System | |

## AUDITING MACHINE LEARNING

Auditing of machine learning is critical when important decisions are made by a relatively autonomous computer system. Machine learning systems can be seen as both socially consequential and opaque (Burrell, 2016). The systems are certainly consequential as they can make decisions on employment, credit worthiness, and health, all vital to individual well-being. They are also consequential even in less threatening systems such as web advertising where it is

still unclear to what extent a society can be influenced by machine learning controlled social media. The system are also opaque in that the user of the system and the target of such systems may have little insight into the black box that created the decision. They must simply accept or reject the results, but being at a loss for the argument behind the results, they may be in a position where they must accept the results. Further, if such results are challenged on a legal or ethical basis, it may be difficult to find or prove the flaw in the system. Content management systems may be one area where the consumer may demand algorithm transparency (Mittelstadt, 2016). In such a system, the consumer relies on system to filter the wide range of information sources available for news and opinion. The system is intentionally biased for the purpose of providing the person with the type of information they are interested in, but at the same time the consumer may be looking for an unbiased view of those issues of interest.

This combination of social consequence with the significance of those consequences expected to increase over time and the opacity of the decision models lead to a great need for a systematic audit procedure. In some contexts, legal issues may demand such audits, while in other contexts consumers may demand it. Such a systematic audit procedure can analyze a system and shed light on operations of different stages of processing. An audit check can be developed based on the level of autonomy of the ML system.

Auditing using such a classification scheme, however may have limited use if the audit is sought by a third party seeking to audit the organization. Since the business organization using the ML will not reveal the details of the model used for the protection of proprietary algorithms, business processes, and proprietary data, the organization may not share enough internal information for systems to be categorized and for the audit procedure appropriate for the category to be applied. As seen earlier, opacity in machine learning system can be divided into three levels, organizational opacity, technical opacity, and mismatches between human decision making and machine (Burrell, 2016). The classification most directly addresses technical opacity and gives a start at considering differences in human decision making and machine. Although, it does not give a solution to organizational opacity, it does provide a framework for what questions can and should be asked as well as a possible framework for what questions organization may be compelled to answer.

## IMPLICATIONS FOR PUBLIC POLICY

If public policy experts are not aware of the similarities between machine learning and statistical analysis, they are likely to misunderstand the problem and propose solution that can be circumvented. Policy makers should be more focused on the improvements in the science of making inference from datasets. If the public or a company has access to an array of datasets, modern analysis techniques can be used to conduct a fine grained classification of individuals including buying power, purchase likelihood, emotional triggers, heath status, financial status, political affiliation or criminal participation.

Governmental decision making based on machine learning is likely to expand and can be regarded as having a positive potential if care is taken in the implementation (Coglianese and Lehr, 2017). Three aspects of machine learning can be seen as factors for worry about implementation, the complexity of the algorithm, the black box nature of the decision, and the automation of the decision process (Coglianese & Lehr, 2017). First, because of the self-learning property of ML,

the computer derives algorithms that are not prescribed by computer analysts. Second, because the derived algorithm is defined with such a wide variety of variables and combinations it acts like a black box whereby the human decision maker would find it hard to tell why a decision was made. Finally, due to the fact that ML systems are often design to be embedded in other computer systems, their results can be acted upon with no human intervention. Debate on these proposed challenges of ML systems can enhanced by using the proposed framework. In such a framework we can separate sources of data available to the ML system and types of algorithms applied. For a given algorithm we could debate what parameters may be useful as audit points. And in the case of Level 7 systems where the output of the ML systems is applied autonomously, the benefits and risks can be weighed and mitigating systems could be proposed.

For example, if advanced machine learning techniques were applied in a health care institution then much may be predicted about its patients' future health conditions. Once such future health conditions are found, is it ethical to disclose or not disclose the findings? Further, it is the nature of machine learning analysis that many inputs and many outputs are simultaneously considered in what could be described as exploratory approach. If a machine learning study were conducted for the purpose of capacity planning for a hospital, it may be approved will little concern for ethical questions. However, the study may simultaneously predict which patients are likely to return because of new health concerns.

Insurance companies in many fields from health care to auto insurance use machine learning to classify customers into profitability categories. Such classification could lead to fine grained insurance pricing. The internal dataset of hospitals and insurance companies may be able to make significant predictions, possibly more accurate predictions are combined with publicly available data such as Facebook or public government records.

## POSSIBILITIES FOR INCREASED LEARNING

The more the techniques of ML are classified and described by what they do autonomously, the less they look like the general public's idea of learning. However, future developments and the use of multiple layers of ML systems could change this.

### Choosing Relevant Problems

Today, ML analysts choose the problem for which to apply ML techniques, but if ML systems begin to make this choice autonomously, new systems with goals not explicitly enumerated by humans could emerge. The assumption behind all research papers we have found to date is that a human decided upon the problem to be solved and selected or designed ML systems to solve this problem. Even if the system appeared to be somewhat of a black box with hidden complexity, ultimately, the human picked a goal and assessed the output. Particularly in the area of non-labeled classification, one could imagine a computer system that autonomously creates categories and possibly had a way to act upon such results.

### Choosing Relevant Data

When humans craft problems for machine learning systems to solve, the humans choose relevant data sets. The machine learning algorithm is applied and a mathematical prediction model is calculated. One area that could bring machine learning closer to human learning would be an architecture where the machine chooses the relevant data sets. Zhang et al. (2016) show how their transformation, hybrid orthogonal projection and estimation tool, improves the performance of neural network training. In their article they describe the machine learning approach as having two stages, extraction and data modeling. Extraction, sometimes called feature engineering, is the process largely done by the human analyst which select which data, interactions, and transformation should be made on the raw data to best prepare it for machine learning. The importance of this process its labor intensity show why much human guidance is given to the machine learning algorithm. Zhang et al. point out that better neural network approaches, such as their own, allow the neural network to function effectively will little or no feature engineering. Raw data set can be fed to the neural network. As this science progresses, it does remove the human decision maker from an important step in the analysis process and may be yet another small step to independent learning.

## CONCLUSION

The eight level ML scale was designed to focus debate on where and how ML systems derive their effectiveness and deliver either benefit or risk. ML system are so diverse that it not useful to say in general what they do and this classification can inform ethical and legal debates over what constitute and issue. It also highlights that while computers may be considered by some as running the world because of the vast number of decisions they make, in all cases there are humans with goals involved at some point in the process and a computer that can control humans is not likely to be developed in our life time.

The eight level categorization scale for ML systems was designed based empirical evidence of ML design. It is robust in that the introduction of newer deep learning methods do not change the scales but just fit within the framework. However, one can imagine that new systems may introduce new technique that could be consider as new categories. More work should be done investigating the latest techniques to document whether they are improved methods of techniques without altering the classification of the technique or in fact change the classification. Deep learning networks are a good example of where the neural network techniques have evolved into new areas. Further, one should be on the lookout where new techniques offer truly new categories. In any of these cases, though, the framework is useful tool for having a consistent debate about the benefits and implications of such new techniques.

The model is also useful at classification of new ML systems. It would be useful to know as new systems are implemented if they are refinements on an existing class of ML system or if they warrant a new class.

## REFERENCES

Burrell, J., (2016, Januay - June). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 1–12.

Carr, N. (2014). *The glass cage: automation and us.* New York, NY, US: W W Norton & Co.

Coglianese, C., & Lehr, D. (2017). Regulating by robot: Administrative decision making in the machine-learning era. *Georgetown Law Journal, 105*, 1147.

Di Nuovo, A., De la Cruz, V., & Marocco, D., (2013). Special issue on artificial mental imagery in cognitive systems and robotics. *Adaptive Behavior*, *21*(4) 217 – 221.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of ACM*, *59*(2), 56-62.

Garfinkel, S., Matthews, J., Shapiro, S. S., & Smith, J. M. (2017). Toward algorithmic transparency and accountability. *Communications of The ACM*, *60*(9), 5.

Gunning, D. (2016). Explainable artificial intelligence (XAI). Defense Advanced Research Projects Agency, Retrieved from https://www.darpa.mil/program/explainable-artificial-intelligence.

Knight, W. (2017). The dark secret at the heart of AI. *MIT Technology Review, 120*(3), 54-65.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

McClelland, C. A. (1971). The management and analysis of international event data: A computerized system for monitoring and projecting event flows. WEIS Technical Report, USC, Los Angeles, CA.

Mittelstadt, B. (2016). Automation, algorithms, and politics: Auditing for transparency in content personalization systems. *International Journal of Communication, 10*, 12.

Monroe, D. (2017). Deep learning takes on translation. *Communications of the ACM*, *60*(6), 12-14.

Newton, I. (1675). Isaac Newton letter to Robert Hooke. Feb. 5, Retrieved from https://discover.hsp.org/Record/dc-9792/.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*, 1929-1958.

Tokui, S., Oono, K., Hido, S., & Clayton, J. (2015). Chainer: a next-generation open source framework for deep learning. Proceedings of workshop on machine learning systems in the twenty-ninth annual conference on neural information processing systems, 5.

Van Merriënboer, B., Bahdanau, D., Dumoulin, V., Serdyuk, D., Warde-Farley, D., Chorowski, J., & Bengio, Y. (2015). Blocks and fuel: Frameworks for deep learning. Cornell University Library, Retrieved from https://arxiv.org/abs/1506.00619.

Wasserman, L. (2013). Rise of the machines. Retrieved from http://www.stat.cmu.edu/~larry/Wasserman.pdf

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives On Psychological Science, 12*(6), 1100-1122.

Zhang, S., Jiang, H., & Dai, L. (2016). Hybrid orthogonal projection and estimation (HOPE): A new framework to learn neural networks. *Journal of Machine Learning Research*, *17*, 1-33.

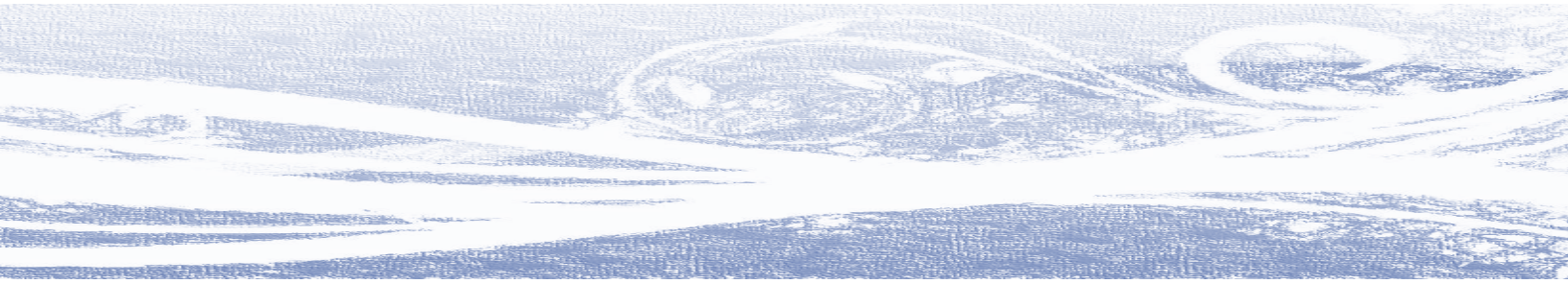
Zhu, R., & Karimi, H. A. (2015). Automatic selection of landmarks for navigation guidance. *Transactions In GIS, 19*(2), 247-261.

# QRBD

# QUARTERLY REVIEW OF BUSINESS DISCIPLINES

May 2018

Volume 5
Number 1



A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES
SPONSORED BY UNIVERSITY OF NORTH FLORIDA
ISSN 2334-0169 (print)
ISSN 2329-5163 (online)