# LIMITATIONS OF ARTIFICIAL INTELLIGENCE WITH BIG DATA

# ETHICS AND RISKS OF AI: THE NEED FOR REGULATION

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Henry Todd

December 1, 2022

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Catherine D. Baritaud, Department of Engineering and Society

Brianna Morrison, Department of Computer Science

Artificial intelligence has developed greatly since its early inceptions. What grew out of the idea of modeling human neurons with the Perceptron (Rosenblatt 1958) has since become a sophisticated and widely used technology. Li, X. and Zhang, T. believe this technology "will change our economy and society significantly in the near future" due to its ability to aid with "dangerous and tedious tasks," however it comes with "potential risks and threats" (2017, p. 1). As the applications of artificial intelligence become more widespread, its potential for misuse or result in other damages due to its limitations must be assessed. To address the potential risks of AI, its limitations and how they may be improved with future innovations will be analyzed, as well as the need for regulations (Fournier-Tombs, 2021) to ensure the ethical use of artificial intelligence.

While artificial intelligence has developed greatly since its origin of the Perceptron, it is still young and has a number of weaknesses. Author Kefei Zhang (2020) describes, "artificial intelligence algorithms rely on big data," and as such, without "integrity of original data…. the conclusion will be less objective even if the calculation process is impeccable" (p. 1). As a technology that is becoming increasingly prevalent in modern systems, the study of shortcomings of artificial intelligence is imperative to ensure safe and proper use. Potential future developments must also be considered. Most artificial intelligence today is known as "weak AI," which as Lu et al. (2017) describe "is designed to perform a special task" (p. 1). Recent research has focused on developing an improved "strong AI" or "general AI" which would "outperform humans at nearly every cognitive task" (Lu et al., 2017, p. 1).

To gain an understanding of the limitations of artificial intelligence, the state-of-the-art report will consider the development path from the inception of the Perceptron (Rosenblatt 1958) to the current state-of-the-art technologies such as deep neural networks and machine learning.

Specifically, it will introduce issues in modern artificial intelligence that arise from its dependence on big data (Zhang, K. 2020). Issues from big data, which itself is inherently flawed (Gudivada et al., 2015), are propagated to the outputs of machine learning models. Figure 1 represents a modern artificial intelligence system, specifically a deep neural net machine learning model. Each point in the model represents a “neuron” which takes in input data and performs some actions to produce output data. Each layer of neurons is connected by lines in the diagram such that the output of one layer acts as the input to the next until a final output value is reached. If the initial data set has issues such as statistical bias, these issues will be propagated from one layer to the next. Specifically, when biased data is used to train the model the weight values ($\mathbf{W}^{i}$) will be trained with the same bias. When such a network is used on testing data, output predictions will be made using biased weighting.

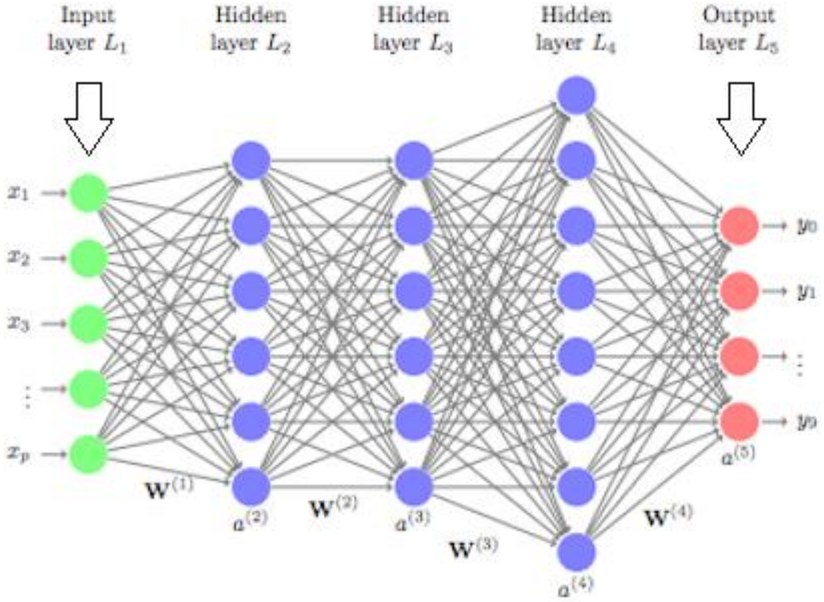

Figure 1: Diagram of a Deep Neural Network. The diagram represents a deep neural network consisting of nodes called “neurons.” Each neuron receives a set of input and produces output through its connections to other nodes. The weights, shown as $\mathbf{W}^{i}$ in the diagram, are “learned” by the model for each layer through training data. The output from one layer of the network is used as input in the next layer. (Boehmke, 2018)

The paper will provide an overview of proposed solutions to these issues stemming from big data, mainly the introduction of statistical methods to mitigate issues within the original data set and to provide context for interpreting the output (Friedrich et al., 2021).

The tightly coupled STS paper considers another facet of the increasing prevalence of artificial intelligence algorithms, namely the potential for misuse as algorithms become more sophisticated as well as the complex ethical dilemmas surrounding the data which algorithms collect and use. The risks for purposeful unethical use of AI are numerous, as Li, X. and Zhang, T. (2017) describe, “attackers can launch a large-scale attack with [few resources] when using intelligent methods” to, for example, “access private information illegally” (p. 2). According to Li, G. et al. (2019), “the current policy system is not perfect” and insufficient for the threats of “personal privacy leakage, widening the gap between the rich and poor, and environmental pollution” (p. 1). McStay and Rosner (2021) introduce one example of the ethical issues concerning the data collected and used by artificial intelligence, in particular in the case of children’s toys. In contrast to Li, X. and Zhang, T., McStay and Rosner believe that new legislation is not necessary for regulating the use of AI, however current legislation must be better enforced (p. 8). The STS paper will review the current policies on data protection and compare these with the proposed means of regulating artificial intelligence algorithms to determine whether regulation is required to regulate this emerging technology.

All research of articles and opinions on the topics described above will be conducted by myself, with preliminary research having been conducted already and continuing into the coming Spring 2023 semester. Both the state-of-the-art report and the STS paper will be written by myself, with guidance from advisors Catherine D. Baritaud of the Department of Engineering and Society and Brianna Morrison of the Department of Computer Science.

# LIMITATIONS OF ARTIFICIAL INTELLIGENCE WITH BIG DATA

The technical state-of-the-art report will consider the issues with modern artificial intelligence systems as they pertain to reliance on ‘big data.’ As Zhang, K. (2020) describes, “artificial intelligence algorithms rely on big data… [which] is incomplete in some levels” (p. 1). He goes on to describe how if these issues are not addressed, or are even “strengthened and amplified by artificial intelligence algorithms,” then “a series of political and ethical issues will occur” (Zhang K., 2020, p. 1). Some specific challenges introduced by big data are “how to capture, transfer, store, clean, analyze, filter, search, share, secure, and visualize data” (Gudivada et al., 2015, p. 3). Dealing with these issues, as Gudivada et al. describe, “require making several tradeoffs among desired scalability, availability, performance, and security” (p. 3). Other issues introduced include interpretability, uncertainty quantification, applications with limited training data, and selection bias (Dunson, 2018, p. 1).

The use of artificial intelligence has grown tremendously in recent years, in large part due to a number of “dramatic developments” which led to “transformative performance [increases]” and the ability “to apply deep learning to a wide variety of data sets and settings” (Dunson, 2018, p. 2). As a result, any issues with artificial intelligence technology are of great importance. In a mortal example provided by Li, X. and Zhang, T. (2017), technical issues with an AI system resulted in the death of a worker by an industrial robot (p. 2). While not always life-threatening, issues in artificial intelligence can have a number of real social and political implications. For example, if some algorithm is trusted to make decisions, “the needs and wishes of [some] disadvantaged groups [may not be] reflected in the final calculation results” if “data shielding” occurs during data collection, i.e. when certain groups are underrepresented in a given data set (Zhang, K., 2020, p. 3).

The research paper will attempt to analyze the problems introduced above and discuss the potential improvements proposed in current state-of-the-art research, especially the integration of statistical methodology towards designing better artificial intelligence algorithms and improved handling of big data (Bülmann & Geer, 2018; Dunson, 2018; Friedrich et al., 2021). Specifically, the paper will discuss how statistics can be used to (1) mitigate issues that arise from big data, (2) improve data collection, (3) provide context and degrees of certainty to results, and (4) minimize bias. Additionally, the paper will briefly consider the limitations of current artificial intelligence as "weak AI" which is constrained to solve specific problems, and evaluate current research towards designing "strong AI," which can extrapolate learning to create "new ideas" (Lu et al., 2018, p. 1).

The paper will be conducted by first collecting and analyzing current state-of-the-art research concerning the reliance of artificial intelligence on its training and testing data sets and resulting limitations, then analyze what methods have been proposed for improving upon these limitations, namely through the use of statistical methodology, and finally considering the likelihood of artificial intelligence undergoing a paradigm shift towards strong AI. Conclusions will be drawn about best practices for current systems, and the best means of improving algorithms in the future. No additional resources or funding will be necessary to complete this paper. It is hoped that the research will take the isolated studies of the topics described above, limitations of current AI, the potential to improve methodology with statistical analysis, and other future improvements, and put them in context of each other, thus giving an overview of the development of artificial intelligence systems up to this point, and how they might be advanced going forward. The style of this paper will be that of a scholarly article.

## ETHICS AND RISKS OF AI: THE NEED FOR REGULATION

The STS research paper will discuss the problem of regulation and policymaking concerning artificial intelligence. Specifically, it will consider the current policy that governs data protection and privacy rights compared to the proposed legislation specific to artificial intelligence. The paper will discuss whether current policy is sufficient to ensure ethical use of AI, or if new legislation is required due to the unique cases introduced by AI algorithms. As discussed in the technical report prospectus, artificial intelligence has developed greatly in recent years (Dunson, 2018, p. 1). As awareness of artificial intelligence methods has grown, so has the desire for ethical usage of the technology (Kieslich et al., 2022, p. 11). However, opinions on the need for new regulation vary. Li, G. et al. (2019) argue that “the current policy system is not perfect” and changes are necessary to protect against “personal privacy leakage, widening the gap between the rich and poor, and environmental pollution” (p. 1). Others such as McStay and Rosner (2021) argue that current policy is sufficient, but must be better enforced (p. 8). These opinions come from different contexts, however. Li, G. et al. discuss the dangers of artificial intelligence methods when used by an adversary who intentionally seeks to do harm, whereas McStay and Rosner discuss the ethics of using artificial intelligence for emotion recognition and emulation, and the privacy of data collected by AI.

This illustrates a complication in discussing the regulation of artificial intelligence. As the number of domains in which artificial intelligence is used continues to increase, the ability to regulate its use under the same policies becomes increasingly difficult. Figure 2 shows the wide variety of algorithms that fall under the label of artificial intelligence. Each of these algorithms are applicable to a number of different problems, for example classification and clustering are often used for classifying text, called "natural language processing."

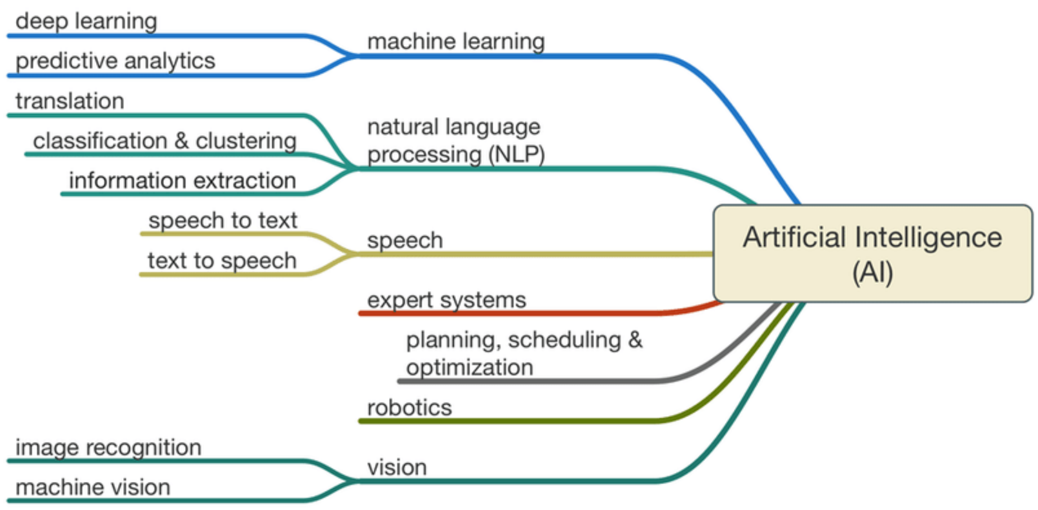

Figure 2: Branches of Artificial Intelligence. The diagram displays a wide variety of groups and subgroups of algorithms which are considered to be artificial intelligence (Villanueva & Salenga 2018).

The following two examples show how two instances of artificial intelligence algorithms can vary greatly, and how regulating AI should vary accordingly. The case study on emotional AI in children's toys discussed by McStay and Rosner states the datification of childhood can have potential negative effects such as (1) generational unfairness: children have little control over the datafication of their childhood with potential "insidious outcomes" such as "a generation of people who have a very confined set of personalities that were heavily influenced by algorithms," (2) manipulation of emotions for economic gain of companies, (3) the longevity of data removing the right to forget the past, or let children "experiment with adult topics," and (4) interaction with "synthetic personalities" that result in an "imbalance of care" (McStay, A., Rosner, G., 2021, p. 4-5). Despite the numerous ethical issues with emotional AI, the study concludes that new regulation is not necessary as emotion data is already protected under legislation regarding sensitive data and/or biometric data (McStay, A., Rosner, G., 2021, p. 8). In contrast, Fournier-Tombs (2021) discusses the need for new regulations by the United Nations as "uses of AI come under the European Commission regulations list of high-risk activities" (p. 5). In addition to protection against high-risk AI, she also discusses the benefits of regulation in forcing companies "to be much more considerate of human rights and privacy" (p. 5).

The variety of scenarios in which artificial intelligence is applied gives rise to a complex web of potential risks and ethical considerations, for which additional research is necessary to design regulatory methods and/or policies that can handle such a dynamic technology. The objective of this paper will be to analyze the SCOT framework for the development of artificial intelligence in order to understand the social influences that affect, and are affected by, these algorithms. Through analyzing this framework, the paper will discuss the proposed methods of regulation and come to a conclusion on how to ensure the ethical use of artificial intelligence. In the SCOT diagram shown in Figure 3, the social groups which influence the development of AI and are influenced by its use are listed. The engineers develop the artificial intelligence algorithm, and choose the data used to train the model. Designers develop the context within which the AI is used. Users interact with the algorithm, providing real-world data as input. For example, in the case of emotional AI toys (McStay, A., Rosner, G. 2021), the engineers train the model, the developers build the toy and implement how the AI is used, and the users provide interactive data which the AI algorithm uses to influence the toy's responses. Legislators and statisticians are the social groups more pertinent to my technical and STS work. The regulators have the ability to affect who is able to be a user, e.g. to protect users such as those children, define who can provide input data, both for training of the model as well as in deployed

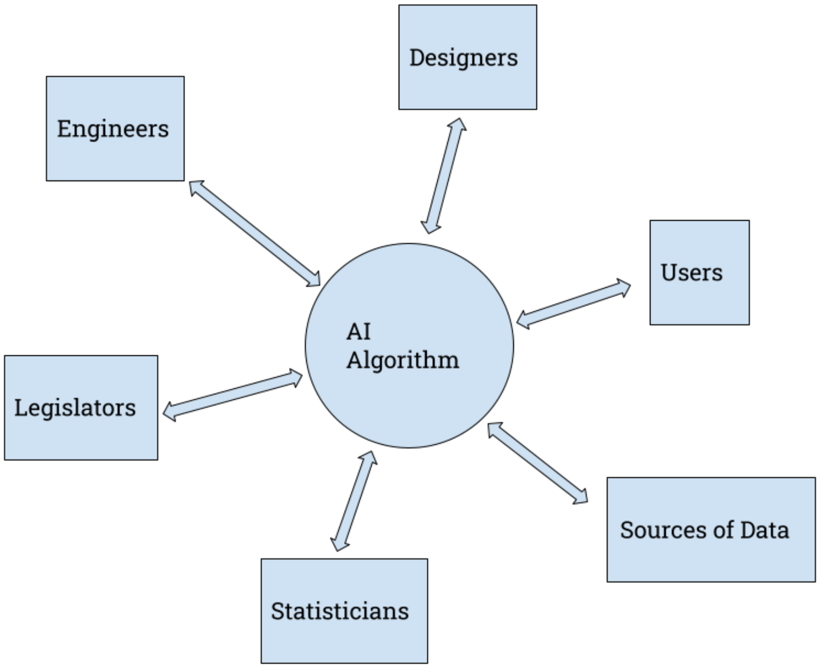

Figure 3: SCOT Diagram of AI. Diagram displays the social groups which influence the development of artificial intelligence algorithms, and which are affected by its development and use. (Adapted by Henry Todd from Pinch T., Bijker, W. 1984)

algorithms. Various methods for accountability and regulation include the study of algorithm transparency (Ananny, M., Crawford, K., 2016; Felzmann et al., 2019), the introduction of new policy (Li, G. et al., 2019) or improved enforcement of existing policy (McStay, A., Rosner, G., 2021), and human-in-the-loop (Waldman, A., Martin, K., 2022). This paper is anticipated to compare the proposed legislation and other methods of accountability with arguments that current data protection laws are sufficient and make a decision on what course of action is necessary to ensure ethical use of AI in the future.

# IMPROVING AI AND DEFINING ITS PLACE IN SOCIETY

Due to the presence of artificial intelligence in widespread applications from economics, healthcare, education, etc, concerns about the misuse or failure of intelligent computing have indicated the need to analyze any technical limitations of current AI and the current policies that govern its ethical use. The technical state-of-the-art report seeks to understand these limitations of the current technology. It will take special care to discuss the dependence of AI on big data and the potential for error propagation from the data sets to outputted results. The technical report will furthermore consider how these issues may be improved upon in the future, mainly by analyzing the proposed use of statistical methods. The tightly coupled STS research paper will study the place of policy and regulatory bodies in ensuring the proper use and development of artificial intelligence systems. It will analyze the SCOT framework of current artificial intelligence systems to discuss proposed regulatory mechanisms and come to a conclusion on how artificial intelligence should be regulated. The conclusions of my technical report and STS research will together provide insight into how artificial intelligence should be treated and improved upon in the coming years, both from a technical standpoint and in terms of its place in society.

## REFERENCES

Ananny, M. Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its applications to algorithmic accountability. *New Media and Society, 20*(3). https://doi.org/10.1177/1461444816676645

Boehmke, B. C. (2018). Feedforward deep learning models. *UC Business Analytics R Programming Guide.* http://uc-r.github.io/2018/04/09/feedforward-deep-models/

Bühlmann, P., van de Geer, S. (2018). Statistics for big data: A perspective. *Statistics and Probability Letters, 136*(1). https://doi.org/10.1016/j.spl.2018.02.016

Cochran, W. G., Rubin, D. B. (1973). Controlling bias in observational studies: A review. *The Indian Journal of Statistics, 35*(4), 417-446. https://www.jstor.org/stable/25049893

Dunson, D. B. (2018). Statistics in the big data era: Failures of the machine. *Statistics and Probability Letters, 136*(1), 4-9. https://doi.org/10.1016/j.spl.2018.02.028

Felzmann, H., Villaronga, E. F., Lutz, C., Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data and Society, 6*(1). https://doi.org/10.1177/2053951719860542

Fournier-Tombs, E. (2021). Towards a United Nations internal regulation for artificial intelligence. *Big Data and Society, 8*(2). https://doi.org/10.1177/20539517211039493

Friedrich, S., Antes, G., Behr, S. et al. Is there a role for statistics in artificial intelligence?. Adv Data Anal Classif (2021). https://doi.org/10.1007/s11634-021-00455-6

Gudivada, V. N., Baeza-Yates, R., Raghavan, V. V. (2015). Big data: Promises and problems. *Computer, 48*(3). http://doi.org/10.1109/MC.2015.62

Kieslich, K., Keller, B., & Starke, C. (2022). Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence. *Big Data and Society*, *9*(1). https://doi.org/10.1177/20539517221092956

König, P. D., Wurster, S., Siewart, M. B. (2022). Consumers are willing to pay a price for explainable, but not for green AI. Evidence from a choice-based conjoint analysis. *Big Data and Society, 9*(1). https://doi.org/10.1177/20539517211069632

Li, G., Deng, X., Gao, Z., Chen, F. (2019). Analysis on ethical problems of artificial intelligence technology. *ICMET 2019,* 101-105. https://doi.org/10.1145/3341042

Li, X., Zhang, T. (2017). An exploration on artificial intelligence application: From security, privacy and ethic perspective. *IEEE*. https://doi.org/10.1109/ICCCBDA.2017.7951949

Lu, H., Li, Y., Chen, M, Kim, H., Serikawa, S. (2018). Brain intelligence: Go beyond artificial intelligence. *Mobile Networks and Applications, 23*, 368–375. https://doi.org/10.1007/s11036-017-0932-8

Mahapatra, S. (2016). Why deep learning over traditional machine learning. *Towards Data Science*. https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063

McStay, A., Rosner, G. (2021). Emotional artificial intelligence in children’s toys and devices: Ethics, governance, and practical remedies. *Big Data and Society, 8*(1). https://doi.org/10.1177/2053951721994877

Pinch, T. J., Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Sciences of Science, 14*(3), 399-441. https://www.jstor.org/stable/285355

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386-408. https://psycnet.apa.org/doi/10.1037/h0042519

Villanueva, M. B., Salenga, L. M. (2018). Bitter melon crop yield prediction using machine learning algorithm. *International Journal of Advanced Computer Science and Applications, 9*(3). https://doi.org/10.14569/IJACSA.2018.090301

Waldman, A., Martin, K. (2022). Governing algorithmic decisions: The role of decision importance and governance on perceived legitimacy of algorithmic decisions. *Big Data and Society*, *9*(1). https://doi.org/10.1177/20539517221100449

Zhang, K. (2020, December 28-29). *The data limitations of artificial intelligence algorithms and the political ethics problems caused by it*. Big Data Analytics for Cyber-Physical System in Smart City, Singapore. https://doi.org/10.1007/978-981-33-4572-0_64