RUTGERS UNIVERSITY LAW REVIEW

VOLUME 72 | ISSUE 2 WINTER 2019

PROTECTING AGAINST ULTRON: EXPLORING THE POTENTIAL CRIMINAL LIABILITY OF SELF-PROGRAMMING DEEP LEARNING MACHINES

*Priya Persaud**

ABSTRACT

In light of recent technological advancement, deep learning machines have matured to the point where they are capable of creating their own artificial intelligence beings, devoid of human interference. While such an advancement progresses the work of computer scientists the world over, it stands to challenge the current state of legal jurisprudence in the United States. For example, it may be possible for deep learning machines to incur criminal liability as a consequence of their conduct. This note explores the possibility of assigning criminal liability to deep learning machines through a three-pronged test that was crafted to account for discrepancies within mens rea *and* actus reus *as they apply to both human beings and deep learning machines.*

TABLE OF CONTENTS

* J.D. Candidate, Rutgers Law School, 2020. A heartfelt thank you to Professor Gary Francione for his guidance throughout the writing process. Thank you to my family, dearest friends, and the staff of the *Rutgers University Law Review* for their support and encouragement.

I. INTRODUCTION

Artificial intelligence ("AI") is "[t]he theory and development of computer systems [that are] able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."[1] While AI spans a wide array of uses in everyday life,[2] there are subsets of AI that are specifically dedicated to high level computation that is not necessary for

1. Bernard Marr, *The Key Definitions of Artificial Intelligence (AI) That Explains Its Importance*, FORBES (Feb. 14, 2018, 1:27 AM), https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#7e7c04e84f5d.

2. Terence Mills, *The Impact of Artificial Intelligence in the Everyday Lives of Consumers*, FORBES (Mar. 7, 2018, 7:45 AM), https://www.forbes.com/sites/forbestechcouncil/2018/03/07/the-impact-of-artificial-intelligence-in-the-everyday-lives-of-consumers/#4650c3a86f31. Everyday uses of AI include software programs that are used to provide safe transactions in online banking as well as the software that is responsible for autonomous vehicles, such as Tesla. *See id.*

everyday life.[3] One such subset is deep learning.[4] Deep learning is a "subfield of machine learning [that uses] . . . algorithms inspired by the structure and function of the [human] brain . . . [to create] artificial neural networks."[5] In other words, deep learning is an attempt to construct computer software such that the machine is able to function similarly to that of a human with respect to tasks such as object recognition, decision-making, and consequence analysis.[6]

Deep learning machines ("DLM") are traditionally created and used for data collection in an attempt to replicate and enhance the thought processes observed in human beings.[7] They are programmed to harvest as much information as possible and then "to perform automatic feature extraction from raw data, also called future learning."[8] In doing so, the DLM "learn[s] complex functions [by] mapping the input to the output directly from data, without depending completely on human-crafted features."[9] The elimination of human-created features[10] and, by

3. *See generally* Yoshua Bengio, *Learning Deep Architectures for AI*, 2 FOUND. & TRENDS IN MACHINE LEARNING 1 (2009) (discussing the complexity of structuring algorithms for deep learning machines); Chris Nicholson, *Artificial Intelligence (AI) vs. Machine Learning vs. Deep Learning*, PATHMIND, https://pathmind.com/wiki/ai-vs-machine-learning-vs-deep-learning (last visited Feb. 28, 2020).

4. *See* Bengio, *supra* note 3.

5. Jason Brownlee, *What is Deep Learning?*, MACHINE LEARNING MASTERY (Aug. 16, 2016), https://machinelearningmastery.com/what-is-deep-learning/.

6. *See* Nicholson, *supra* note 3 ("Deep artificial neural networks are a set of algorithms that have set new records in accuracy for many important problems, such as image recognition, sound recognition, recommender systems, . . . etc."); *see also A Beginner's Guide to Neural Networks and Deep Learning*, PATHMIND, https://pathmind.com/wiki/neural-network (last visited Feb. 28, 2020) (listing examples of how neural networks are used, including as classification devices and clustering tools).

7. Bernard Marr, *10 Amazing Examples of How Deep Learning AI Is Used in Practice?*, FORBES (Aug. 20, 2018, 12:11 AM), https://www.forbes.com/sites/bernardmarr/2018/08/20/10-amazing-examples-of-how-deep-learning-ai-is-used-in-practice/#17af640df98a (listing the ways in which deep learning machines are used and the fact that they are created based off of the cognitive thought process of human beings).

8. Brownlee, *supra* note 5.

9. *Id.* ("The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones. If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, we call this approach to AI deep learning.").

10. *See id.* Human-crafted features generally refers to the code and algorithms that are written by human beings and are later incorporated into DLM. *See* Ved, *Difference Between Usual Machine Learning and Deep Learning Explained!*, D4DATASCIENCE (Mar. 31, 2016), https://d4datascience.wordpress.com/2016/03/31/difference-between-usual-machine-learning-and-deep-learning-explained/. The code and algorithms instruct the DLM on how to execute various tasks. *Id.* The use of human-crafted features means the DLM is still in the early stages of learning how to perform data collection. *Id.* As the DLM learns how to perform data collection, its reliance on the algorithm decreases. *Id.* Decreased reliance means the DLM is learning how to function independently of its human creator. *Id.*

extension, human interference prepares the DLM for autonomous decision making.[11]

Thus, the DLM is capable of constructing a library of knowledge and resources, pulled from the deep learning methodology, that can be accessed at a later date depending on the task at hand.[12] For example, once the DLM has populated its library of necessary information, it can use that library of knowledge to create its own AI machine, devoid of human intervention.[13] In doing so, the knowledge that is contained in the newly created AI machine will originate solely from the DLM's autonomous decisions.

To better understand this chain dynamic, consider the following example. A software engineer (Human 1) programs a machine (DLM 1) to perform data collection. After executing data mining techniques and constructing a library of knowledge, DLM 1 is fully equipped to utilize the library of knowledge to create its own machine. DLM 1, using its library of knowledge, programs another machine (DLM 2). The programming that takes place between DLM 1 and DLM 2 is strictly between two machines and is devoid on human intervention. Therein lies the problem. Since the level of programming between DLM 1 and DLM 2 does not require human interference, should human beings be held liable for the actions of such machines?

This note will explore whether the DLM can be held criminally liable for its autonomous decisions. Part II introduces the importance of expanding criminal liability to non-human entities. Part III discusses the legal presence of DLM by investigating whether they satisfy the elements of personhood, as enumerated by modern philosophy. Part IV reviews the decision-making process of DLM in comparison with that of human beings to determine whether the DLM can satisfy the requisite *mens rea* and *actus reus* of criminal conduct. Part V analyzes current case law that

11. *See* Nicole Kwan, *The Hidden Dangers in Algorithmic Decision Making*, TOWARDS DATA SCIENCE (Dec. 1, 2018), https://towardsdatascience.com/the-hidden-dangers-in-algorithmic-decision-making-27722d716a49. Autonomous decision making requires the integrated use of a plethora of algorithms. *See id.* However, "[a]lgorithmic autonomy and transparency seem to have an inverse relationship—as these algorithms become increasingly better at 'learning' and adjusting, it becomes more difficult to understand where the biases occur." *Id.*

12. *See A Beginner's Guide to Neural Networks and Deep Learning*, *supra* note 6 ("Learning without labels is called *unsupervised learning*. Unlabeled data is the majority of the data in the world. One law of machine learning is: the more data an algorithm can train on, the more accurate it will be. Therefore, unsupervised learning has the potential to produce highly accurate models.").

13. *See* Dom Galeon, *Google's Artificial Intelligence Built an AI That Outperforms Any Made by Humans*, FUTURISM (Dec. 1, 2017), https://futurism.com/google-artificial-intelligence-built-ai.

addresses criminal liability of non-human entities. Part VI examines the common law requirements for imputing criminal liability. Part VII explores traditional avenues of criminal liability imputation. Part VIII provides a brief overview of the discussion thus far, prior to delving into this note's suggestion for legislative reform. Part IX proposes a three-pronged test that was crafted to account for discrepancies within *mens rea* and *actus reus* as they apply to both human beings and deep learning machines. Finally, part X discusses potential objections to the suggested legislative reform presented in part IX.

II. THE SIGNIFICANCE OF EXPANDING CRIMINAL LIABILITY

As technological advancements continue to improve everyday life, such advancements may negatively affect societal and individual productivity.[14] More specifically, AI and DLM that are programmed with artificial neural networks similar to that of human beings have the capacity to reprogram the embedded commands such that it can employ a more destructive method of reaching its goal.[15] As such, it is virtually impossible to predict the behavior of AI and DLM.[16] Therefore, it is important for scholars and decision-makers to stay abreast of such issues and to tackle future grievances with current legislation.[17]

In recent years, the United States legislature commenced efforts to overcome the difficulties surrounding the unpredictability of AI machines.[18] In December 2017, the House of Representatives introduced a bill that sought to further understand the current state of AI

14. *See* Max Tegmark, *Benefits & Risks of Artificial Intelligence*, FUTURE LIFE INST., https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/ (last visited Feb. 28, 2020) ("In the long term, an important question is what will happen if the quest for strong AI succeeds and an AI system becomes better than humans at all cognitive tasks . . . . Such a system could potentially undergo recursive self-improvement, triggering an intelligence explosion leaving human intellect far behind.").

15. *See id.*

16. *Id.* Because AI has the potential to become more intelligent than any human, we have no surefire way of predicting how it will behave. We cannot use past technological developments as much of a basis because we have never created anything that has the ability to, wittingly or unwittingly, outsmart us. The best example of what we could face may be our own evolution. People now control the planet, not because we're the strongest, fastest or biggest, but because we're the smartest. If we're no longer the smartest, are we assured to remain in control? *Id.*

17. *See* Claudia Geib, *Lawmakers Want You to Be Able to Sue Robots*, FUTURISM (Apr. 13, 2018), https://futurism.com/robots-rights-eu-personhood (discussing the point of view of European lawmakers who believe "electronic personalities" should be granted legal personhood such that it can be held accountable for its conduct).

18. Fundamentally Understanding the Usability and Realistic Evolution of Artificial Intelligence Act of 2017, H.R. 4625, 115th Cong. (2017).

advancements as well as the capabilities of AI machines.[19] Although the bill failed, the mere introduction of the bill highlights the fact that concerns surrounding AI and DLM should be at the forefront of today's concerns.[20] Thus, this note is of timely importance as it will establish a three-pronged test through which a DLM may be held accountable for its actions.[21]

III. DISCUSSING PERSONHOOD OF SOPHISTICATED AI

Prior to delving into a criminal liability analysis, it is necessary to pinpoint where DLM currently fall within the justice system. Only after carving a niche for them can deep learning machines be subject to the privileges and consequences that human beings encounter within the justice system.

As the law currently stands, there is no legislation that specifically addresses the creation or existence of DLM.[22] However, there is a budding area of the law, Artificial Intelligence Law, that is premised on specifically handling matters related to the general class of artificial intelligence, without making reference to DLM.[23] The creation of Artificial Intelligence Law is substantial progress toward the law being able to recognize beings other than human beings.[24] As this area of law continues to develop and expand, recent technological advancements will inevitably lead to discussions of personhood for artificial intelligence beings, especially DLM.[25] In anticipation of that legal development, the remainder of this section addresses the personhood argument.

The hallmark of a thorough personhood discussion is a list of factors that can be consulted during the process of establishing personhood.[26] As derived from the scholarship of Joseph Fletcher, a philosopher within

19. *Id.*

20. *See* Geib, *supra* note 17.

21. *See infra* Part IX.

22. *See* Huu Nguyen, *Artificial Intelligence Law is Here, Part One*, ABOVE THE LAW (July 26, 2018, 2:22 PM), https://abovethelaw.com/legal-innovation-center/2018/07/26/artificial-intelligence-law-is-here-part-one/?rf=1.

23. *See id.* ("Artificial Intelligence Law is the field of law dealing with the rights and liability that arises from the use of AI and the AI itself . . . . Artificial Intelligence Law is being developed now, in order to set the rules of the road for the usage of AI. And we as lawyers should recognize it as a specific discipline.").

24. *See id.*

25. *See* Geib, *supra* note 17.

26. *See* Samir Chopra & Laurence White, *Artificial Agents—Personhood in Law and Philosophy*, PROC. 16TH EUROPEAN CONF. ON ARTIFICIAL INTELLIGENCE 635, 635–39 (2004) ("Typically, contributions to this debate involve drawing up a list of necessary and sufficient conditions, which must be met by an artificial agent in order to be classified as a genuine cognizer on par with human beings.").

moral theory and applied ethics, there are fifteen characteristics to keep in mind when exploring personhood: (1) showing of minimal intelligence;[27] (2) self-awareness;[28] (3) self-control;[29] (4) a sense of time;[30] (5) a sense of futurity;[31] (6) a sense of the past;[32] (7) the capability to relate to others;[33] (8) concern for others;[34] (9) ability to communicate;[35] (10) control of existence;[36] (11) curiosity;[37] (12) change and changeability;[38] (13) balance of rationality and feeling;[39] (14) idiosyncrasy;[40] and (15) neocortical function.[41]

Although there are a number of variations on the list of personhood elements, Joseph Fletcher was the first person to present this idea to the world and the first person to challenge closely held societal norms.[42] For

27. *See* Jon Christian, *New Robot Is "On the Path to Machine Self-Awareness," Says Creator*, FUTURISM (Feb. 4, 2019), https://futurism.com/new-robot-self-awareness (discussing the creation of an AI that is capable of constructing a self-image from scratch).

28. *See id.* ("While our robot's ability to imagine itself is still crude compared to humans, we believe that this ability is on the path to machine self-awareness.").

29. A. Kamala Kumari & Ch Sai Lakshmi, *Self Controlled Robot for Military Purpose*, 6 INT'L J. ADVANCE RES. SCI. & ENGINEERING 1301, 1301 (2017) (discussing the creation of self-controlled military AI that are specifically designed to broach dangerous situations that are not safe for human beings).

30. *See* The Dig. Acid, *Two Robots Debate the Future of Humanity*, YOUTUBE (July 14, 2017), https://www.youtube.com/watch?v=w1NxcRNW_Qk (featuring a discussion between two AI beings where they discuss the meaning of consciousness with respect to human beings and the possibility of expanding consciousness in the future).

31. *See id.*

32. *See id.*

33. Arab News, *Robot Sophia Speaks at Saudi Arabia's Future Investment Initiative*, YOUTUBE (Oct. 25, 2017), https://www.youtube.com/watch?v=dMrX08PxUNY (featuring a presentation by Sophia the Robot where she expresses interest in wanting to work with humans and her passion to build trust with people).

34. *See id.*

35. *See* The Dig. Acid, *supra* note 30; *see also* Arab News, *supra* note 33.

36. *See* Arab News, *supra* note 33.

37. *See Parle: A Curious Robot*, PERS. ROBOTS GRP., https://robotic.media.mit.edu/portfolio/curious-robots/ (last visited Feb. 28, 2020) ("This project aimed to test whether children can 'catch' curiosity from a social robot . . . . We found that indeed, Parle's curiosity was 'contagious': children playing with it were more curious after the interaction, compared to children playing with a non-curious robot.").

38. *See* Zachary Tomlinson, *15 Medical Robots that are Changing the World,* INTERESTING ENGINEERING (Oct. 11, 2018), https://interestingengineering.com/15-medical-robots-that-are-changing-the-world (discussing medical AI that have been used in recent years to perform functions with more precision than that of human beings).

39. *See* Arab News, *supra* note 33.

40. *Id.*

41. *See Neocortex (brain)*, SCIENCEDAILY, https://www.sciencedaily.com/terms/neocortex.htm (last visited Feb. 28, 2020) [hereinafter *Neocortex*]; *see also* Joseph F. Fletcher, *Four Indicators of Humanhood: The Enquiry Matures*, 4 HASTINGS CTR. 6, 6 (1974).

42. *See* George Dvorsky, *What is a Person?*, SENTIENT DEVELOPMENTS (Apr. 27, 2009), http://www.sentientdevelopments.com/2009/04/what-is-person.html.

example, after performing Fletcher's personhood analysis, one might find that not all human beings qualify as a person and that a cyborg may qualify as a person.[43] A simple way to apply Fletcher's personhood elements and comprehend the results is through the following illustrations. To demonstrate the difference in computation power between everyday AI machines and DLM, the illustrations are bifurcated into those two categories.

A. First Illustration: Sophia the Robot

The most famous embodiment of these characteristics in an everyday AI is Sophia the Robot.[44] After becoming the first AI in the world to obtain citizenship, Sophia stood before attendees of the Future Investment Initiative Conference to express her gratitude.[45] Sophia's ability to engage in conversation demonstrates that her level of intelligence surpasses the de minimis threshold.[46] Her knowledge of how she was created and her interest in helping humans lead better lives are illustrative of Sophia's self-awareness, sense of futurity, capacity to relate to others, and her concern for others.[47] Lastly, Sophia's ability to demonstrate her range of emotions, through telling jokes and bidding adieu to the audience, suggests that she is capable of neocortical functioning.[48]

> Many of Fletcher's traits are fairly subjective, open to argument (e.g. how do you measure intelligence, and how intelligent is intelligent enough?) and difficult to test scientifically (at least by today's standards). But what's interesting about this list is that not all human beings qualify as persons, and not all persons qualify as human. Moreover, individuals, at one time or another, are not persons. Fletcher argued that some severely developmentally challenged humans were not persons, and that chimeras and cyborgs might someday qualify as persons (what he called "parahumans").

Id.

43. *Id.*

44. Zara Stone, *Everything you Need to Know About Sophia, the World's First Robot Citizen*, FORBES (Nov. 7, 2017, 12:22 PM), https://www.forbes.com/sites/zarastone/2017/11/07/everything-you-need-to-know-about-sophia-the-worlds-first-robot-citizen/#2652cecf46fa.

45. *Id.* ("On October 25, Sophia, a delicate looking woman with doe-brown eyes and long fluttery eyelashes made international headlines. She'd just become a full citizen of Saudi Arabia—the first robot in the world to achieve such a status."); *see also* Arab News, *supra* note 33.

46. *See* Arab News, *supra* note 33.

47. *Id.*

48. *Id.*; *see also* Pasko Rakic, *Evolution of the Neocortex: Perspective from Developmental Biology*, 10 NATURE REV. NEUROSCIENCE 724, 724 (2009) ("The neocortex, as the name implies, is the newest addition to our brain and is considered to be the crowning achievement of evolution and the biological substrate of human mental prowess.").

Neocortical functioning refers to a being's ability to execute "higher functions such as sensory perception, generation of motor commands, spatial reasoning, conscious thought, and, in humans, language."[49] Neocortical functioning has historically been recorded solely in mammals.[50] Therefore, as DLM near closer to the human threshold of cognitive capacity, it follows that DLM may also develop the neocortical functioning of human beings.

B. Second Illustration: IBM's Watson

However, while Sophia's language and motor skills allow viewers to witness and observe Fletcher's personhood characteristics at play, that is not the case for DLM.[51] The capabilities of DLM are not manifested through outlets similar to that of Sophia, but can be parsed using an understanding of the DLM's inputs and outputs.[52]

One example is IBM's Watson, which initially debuted as a competitor on Jeopardy.[53] Watson was created using a computer architecture structure that performs natural language processing in combination with the information that is stored within its library of knowledge.[54] This structure is better understood through the following breakdown.[55] IBM provided Watson with data mining instructions which allowed Watson to collect and store data that was found by inputting various queries into search engines.[56] That information was then added to Watson's library of knowledge.[57] As Watson mined for data, it was important for Watson to learn how to understand the true meaning of the data since machines "can't really understand what the language is really

49. *Neocortex*, *supra* note 41.

50. *See* Zoltán Molnár & Alex Pollen, *How Unique is the Human Neocortex?*, 141 COMPANY BIOLOGISTS 11 (2014).

51. *See* Nicholson, *supra* note 3; *see also* Arab News, *supra* note 33.

52. *See* Nicholson, *supra* note 3 ("Nets with many layers pass input data (features) through more mathematical operations than nets with few layers, and are therefore more computationally intensive to train.").

53. Jo Best, *IBM Watson: The inside story of how the Jeopardy-winning supercomputer was born, and what it wants to do next*, TECH REPUBLIC (Sept. 9, 2013), https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/ ("While Watson had the questions delivered in text rather than by listening to the quizmaster, he played the game like his human counterparts: puzzle over the question, buzz in, give the answer that's most likely to be right, tot up some prize money.").

54. *Id.*; *see also* Michael J. Garbade, *A Simple Introduction to Natural Language Processing*, MEDIUM (Oct. 15, 2018), https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32.

55. *See* Best, *supra* note 53.

56. *See id.*

57. *See id.*

trying to say. In a nutshell, a computer can't read between the lines."[58] Therefore, IBM also taught Watson how to perform natural language processing (NLP) such that Watson could study the nuances within the languages humans speak.[59]

As Watson trained to compete on Jeopardy, it used countless algorithms to sift through its library, rank answers based on likelihood of success, and produce a weighted list of possible answers.[60] In doing so, Watson demonstrated its ability to display intelligence, weigh past events against potential future occurrences, adapt to change, and balance the rationality of each potential answer.[61]

The processing power of both Sophia the Robot and Watson support the fact that DLM, and AI in general, are more sophisticated than in years past.[62] No longer are human beings the sole being capable of rationalizing and executing autonomous decisions. As such, this overview is the first step in establishing a niche within United States legal jurisprudence for DLM. The second step is to map the decision-making processes of DLM in an attempt to reconcile their processes with those of human beings.

IV. HOW DLM ARE PROGRAMMED TO MAKE DECISIONS

There are a number of avenues that can be explored in the efficient creation of DLM.[63] However, DLM are normally programmed with particular algorithms that instruct the DLM to perform various data mining techniques and to extract particular information.[64] The data bits that are extracted are stored within the DLM's memory for later use.[65]

58. *See* George Seif, *An Easy Introduction to Natural Language Processing*, BUILT IN (May 14, 2019), https://builtin.com/data-science/easy-introduction-natural-language-processing.

59. Sean Sodha, *How to Get Started with Natural Language Processing*, IBM (Oct. 3, 2019), https://www.ibm.com/blogs/watson/2019/10/how-to-get-started-with-natural-language-processing/.

60. *See* Best, *supra* note 53.

61. *Id.*

62. Manoj Gupta, *Where AI is Headed in 2018*, MEDIUM (Feb. 2017), https://web.archive.org/web/20180420080718/https://becominghuman.ai/where-ai-is-headed-in-20181f8913fd420e?gi=8346ecbdbd1d ("2018 would see more use of converting structured data into intelligent narratives based on natural language generation (NLG) and natural language processing (NLP). AI would see more use in automated content generation in news coverage, sports, financial reports, and social media and so on using rule-based systems.").

63. *See generally* Nicholson, *supra* note 3 (discussing various methods of simulating human intelligence to create artificial intelligence).

64. *See* Bengio, *supra* note 3, at 1–2.

65. *See* Best, *supra* note 53.

As the DLM acquires more information, it progresses through a series of algorithms that gradually increase in complexity.[66] These algorithms eventually reach the point where they instruct the DLM to rely solely on the information within its expanding library of knowledge, rather than instructing the DLM to continue collecting additional information.[67] At this point in the programming phase, the DLM is reliant upon its library of knowledge and is, ideally, equipped to make autonomous decisions.[68] Thus, deep learning occurs when the DLM uses the knowledge it already ascertained to make future decisions. For example, IBM's Watson, discussed above, undertook a variation of this methodology while preparing to be a contestant on Jeopardy.[69]

As the DLM learns how to make its own decisions, interference from human beings is far removed.[70] As such, the DLM uses this learned programing process to acquire sufficient knowledge to craft and execute its own algorithms.[71] Once the machine is able to do this, deep learning is beyond the control of human beings because only the machine understands what its algorithms are meant to do.[72]

There are two main reasons only a machine would be able to parse the true meaning of a DLM's algorithms. First, on a purely logistical level, the algorithms are written entirely in machine language.[73] Machine language is specifically used for speaking with machines and, as such, is understood by machines only.[74] Therefore, if a human being were to perform code review on a machine's algorithm, it is unlikely that the person would successfully uncover the algorithm's goal. Secondly, the level of complexity of the DLM's algorithm is assuredly far beyond a human being's processing capacity. Because DLM focus solely on data mining and data extraction, DLM are familiar with such tasks and are able to complete them at a faster rate than most human beings. Therefore, aside from the machine language barrier, the sheer complexity of a DLM's algorithms places the processing power of DLM ahead of human beings.

66. *Id.*
67. *See id.*
68. *Id.*
69. See *supra* Section III.B for an in-depth discussion on IBM's Watson and how the machine was created; *see also* Best, *supra* note 53.
70. *See* Best, *supra* note 53.
71. *See, e.g.*, Aaron Mak, *Google Taught A.I. How to Program More A.I.*, SLATE (Oct. 16, 2017, 3:56 PM), https://slate.com/technology/2017/10/google-created-machine-learning-software-that-can-program-machine-learning-software.html.
72. *Id.*
73. *See* Bengio, *supra* note 3, at 1–2.
74. *Id.*

For example, Google recently launched a project called AutoML, which is a deep learning machine that teaches other AI how to use machine-learning software.[75] AutoML essentially acts as a teacher and the students are other AI that are learning the appropriate algorithms and programming practices to push them to the DLM threshold. When AutoML and the subsequent machines that it created were tested against machines created by human beings on image identification, AutoML and its machines outperformed the human-programmed machines each time.[76] This led the CEO of Google to confess that "only 'a few Ph.D.s' currently have the skills necessary to create the most complex A.I. systems."[77]

When a DLM, such as Watson or the Google deep learning machine, is presented with a question, the DLM "works out some possible answers based on the information it has [at] hand, creating a thread for each. Every thread uses hundreds of algorithms to study the evidence . . . then generate[s] a ranked list of answers, with evidence for each of its options."[78] In using hundreds of algorithms to map one input to multiple outputs, the DLM crafts a multi-dimensional map of its decisions, which is quite difficult for a human being to decipher.[79]

Thus, the decision-making processes of DLM are so far removed from human beings that the inner workings of the processes are almost hidden from computation specialists. Such quasi machine independence welcomes the question of whether such machines can establish a legal identity such that they can be held liable for their actions. These findings are the foundation for the following legal analysis that explores the possibility of imputing criminal liability upon DLM. The first step in doing so is to dissect the current case law that addresses criminal liability of DLM.

75. Mak, *supra* note 71 ("The company's A.I. project, AutoML, has successfully taught machine-learning software how to program machine-learning software.").

76. *Id.* There was an informal competition between the machines that were programmed by human beings and machines that were programmed by AutoML. *Id.* As part of the competition, machines from both teams were shown identical images and were asked to identify the object in the image. *Id.* The machines that were programmed by AutoML performed much better than the machines that were programmed by human beings. *Id.* The success of the AutoML machines can be attributed to the fact that, because of their deep learning algorithms, the AutoML machines had a greater library of knowledge with which to work. *See id.*

77. *Id.*; *see* Best, *supra* note 53.

78. Best, *supra* note 53; *see* Mak, *supra* note 71; *see also* Yangyan Li, *A Brief Introduction to Deep Learning*, Slide 27 (providing visualization of the multi-dimensional decision map).

79. *See* Li, *supra* note 78.

V. CURRENT CASE LAW ON THE CRIMINAL LIABILITY OF AI MACHINES

Due to the relative novelty of DLM combined with the judicial system's inability to mature as technology progresses, there is no case law which explicitly discusses instances where DLM's conduct was called into question.[80] Thus, the foundation of this legal analysis takes a step back from DLM as a subset of AI and focuses on AI as a whole.

As it exists today under United States criminal law, criminal liability is not imputed upon AI.[81] While there have been cases where owners or manufacturers of every day AI beings were hauled into court due to the AI's violent behavior,[82] courts have never engaged in a discussion on the potential liability of AI machines.[83] Instead, courts apply the elements of criminal liability, *mens rea* and *actus reus*, to the human beings who own or manufacture the AI machines.[84] The following subsections explore cases involving everyday AI and the consequences of permitting courts to overlook the discussion on the criminal liability of such machines.

A. Cases Where AI Engaged in Harmful Conduct

In *Payne v. ABB Flexible Automation*, a factory worker, Michael Payne, was responsible for operating and programming an AI that was used in the production of aluminum car wheels.[85] After instructing his co-workers to take a break from work, Payne stayed behind to monitor the machine.[86] Payne entered the cell where the AI was contained and, when his co-worker returned, was found pinned between the arm of the

80. *See* Geib, *supra* note 17.

81. Gabriel Hallevy, *The Criminal Liability of Artificial Intelligence Entities—from Science Fiction to Legal Social Control*, 4 AKRON INTELL. PROP. J. 171, 173 (2010) ("[W]hat kind of laws or ethics are correct and who is to decide? In order to cope with these same problems as they relate to humans, society devised criminal law."); *see also* WILLIAM M. CLARK & WILLIAM L. MARSHALL, LAW OF CRIMES 1 (Melvin F. Wingersky ed., 6th ed. 1958) ("Human control of contingent human behavior is the purpose of criminal law.").

82. *See* Achieng, *15 Most Savage Deaths Caused by Robots*, RICHEST (July 26, 2017), https://www.therichest.com/technologies/15-shocking-deaths-caused-by-robots/; *see also* Mike Power, *What Happens When a Software Bot Goes on a Darknet Shopping Spree?*, GUARDIAN (Dec. 5, 2014, 8:56 AM), https://www.theguardian.com/technology/2014/dec/05/software-bot-darknet-shopping-spree-random-shopper.

83. *See* Geib, *supra* note 17.

84. For an in-depth discussion of the elements of criminal liability, see *infra* Part VI of this note.

85. Payne v. ABB Flexible Automation, No. 96-2248, 1997 U.S. App. LEXIS 13571, at *2–3 (8th Cir. June 9, 1997) ("This case arises from an industrial accident that fatally injured Michael L. Payne while he was working in the cell of an automated robot at the Fayetteville, Arkansas, facility of Superior Industries ('Superior') on September 27, 1994.").

86. *Id.* at *1.

AI and a wheel inside the drilling machine.[87] Due to medical complications from this incident, Payne passed away two days later.[88] The court ruled in favor of the manufacturer of the AI machine, finding that the design of the AI was not a product defect and that Payne should not have approached the AI without proper safety precautions.[89] In this case, the court failed to discuss potential criminal liability of the AI machine.

While the breakdown of the AI machine's code was not discussed in the case, it is likely the AI machine was programmed solely for the purpose of assisting with manufacturing wheels.[90] Its tasks likely included sensing when a wheel approached it from the production line and detecting where the wheel was located on the production line, such that it would be able to maneuver the wheel into the next manufacturing phase.[91] While it is unlikely that the AI was equipped to perform independent decision-making techniques typical of deep-learning machines, this case is a prime example of the need to explore liability in cases where human beings are injured by AI machines.

Similar to the decedent in *Payne*, the decedent in *Williams v. Litton Systems, Inc.* was fatally wounded by an AI machine which malfunctioned.[92] The AI was programmed to lift heavy items in a car manufacturing plant.[93] Unfortunately, Williams was within close proximity of the AI's mechanical arm when it malfunctioned.[94] The AI's

87. *Id.*

88. *Id.*

89. *See id.* at *3–4.

> The OSHA report did not attribute the accident to a defect in the robot, but cited Superior for removing safety devices from the cell of a programmed robot, and for allowing employees to enter into the immediate operational area of the robot, thereby exposing them to the danger of injury by being caught in the robot's jaws. OSHA subsequently deleted the citation and waived the penalty following an informal conference in which Superior agreed to correct the violations. The Superior report indicated that inattention by Payne was the primary factor in the accident, and found that Payne had overlooked safety measures by entering the cell before "locking it out," and by running the robot at 100% test speed while inside the cell, rather than at 25% speed as required by Superior's safety guidelines.

Id. at *1.

90. *See id.* at *1.

91. *See* WHEEL PRODUCTION LINE, https://www.wheelproductionline.com (last visited Jan. 22, 2020).

92. *See* Williams v. Litton Sys., Inc., 449 N.W.2d 669, 670 (Mich. 1989) ("Robert N. Williams, a Ford employee, was killed when he was struck by a malfunctioning robotic machine that was part of a system designed and manufactured by Litton for Ford.").

93. Williams v. Litton Sys., Inc., 416 N.W.2d 704 (Mich. Ct. App. 1987); *see also* Achieng, *supra* note 82.

94. *See* Achieng, *supra* note 82.

arm lifted Williams into the air and slammed him into a wall, crushing him to death.[95]

Much like the court's analysis in *Payne*, the court's analysis in *Williams* does not address the specific programming capability of the AI that caused Williams' death.[96] In fact, the court bypasses the issue of assigning liability to the AI machine by accepting the AI's conduct and focusing on which human being should be liable for the AI's conduct.[97]

B. Consequences of Permitting Courts to Overlook the Discussion on the Expansion of Criminal Liability to AI and DLM

Although only two cases are explored here, there are a number of instances where individuals were harmed by artificial intelligence beings.[98] In such cases, courts were not tasked with determining the liability of the AI in question.[99] Instead, the court accepted the actions of the AI and glossed over the question of AI liability to determine which human beings should be held liable.

The fact that courts focus solely on the criminal liability of human beings instead of AI machines leads to an incomplete liability analysis.[100] Given the information in earlier sections of this note, it has been established that DLM are capable of autonomous decision making.[101] As such, their decisions are concealed from human interference, so much so that only a few human beings actually understand what the DLM is capable of doing.[102] Since the DLM's decision-making process can survive without human interference, the imposition of criminal liability solely upon human beings seems improper.[103] Thus, to engage in a complete analysis of criminal liability, it is necessary to explore current elements that must be satisfied to achieve criminal liability.

95. *See id.*
96. *See generally Williams*, 449 N.W.2d at 669.
97. *Id.* at 671.
98. *See* Achieng, *supra* note 82.
99. *See generally* Payne v. ABB Flexible Automation, No. 96-2248, 1997 U.S. App. LEXIS 13571, at *2–3 (8th Cir. June 9, 1997); *Williams*, 449 N.W.2d at 671.
100. *See* Geib, *supra* note 17.
101. See *supra* Part IV for more information on how DLM are programmed and how they utilize algorithms to engage in autonomous decision making.
102. *See* Mak, *supra* note 71.
103. *See* Geib, *supra* note 17.

VI. COMMON LAW REQUIREMENTS TO IMPUTE CRIMINAL LIABILITY

Charging a person with criminal liability requires a showing of two elements: (1) *actus reus* and (2) *mens rea*.[104] According to *Martin v. State*, *actus reus* is the physical carrying out of a crime.[105] *Actus reus* must be done voluntarily, meaning that it must have been executed through one's own volition and was controlled by the entity's mind.[106]

Mens rea refers to one's state of mind or their criminal intent.[107] Within *mens rea*, there are four culpable mental states.[108] First, a person can act intentionally such that he or she knew the result would occur or was likely to occur because of his or her conduct.[109] Second, a person can engage in negligent conduct where he or she should have been aware of the fact that his or her conduct would create a substantial and unjustifiable risk.[110] Third, a person can act recklessly where he or she is aware of the substantial and unjustifiable risk, but choose to disregard it.[111] Fourth, a person can act with malice where he or she intentionally causes harm.[112]

In addition to identifying the requisite elements of criminal liability, it is also important to understand how a court applies these elements to impose criminal liability.[113] This will be discussed in the following section.

104. *See* Martin v. State, 17 So. 2d 427, 427 (Ala. Ct. App. 1944) (finding that criminal liability must be based on conduct which includes a voluntary act or an omission to act which could have been performed); *see also* People v. Conley, 543 N.E.2d 138, 143–44 (Ill. App. Ct. 1989) (finding that a criminally liable defendant intends the natural and probable consequences of his or her actions); State v. Nations, 676 S.W.2d 282, 284 (Mo. Ct. App. 1984) (finding that knowing engagement in criminal conduct includes the defendant's actual knowledge of the existence of attendant circumstances which constitute the alleged crime).

105. *See Martin*, 17 So. 2d at 427.

106. *See id*.

107. *See Conley*, 543 N.E.2d at 143–44.

108. *See id*.; *see also Nations*, 676 S.W.2d at 284.

109. *See Conley*, 543 N.E.2d at 143–44; *see also Nations*, 676 S.W.2d at 284.

110. *See Conley*, 543 N.E.2d at 143–44; *see also Nations*, 676 S.W.2d at 284.

111. *See Conley*, 543 N.E.2d at 143–44; *see also Nations*, 676 S.W.2d at 284.

112. *See Conley*, 543 N.E.2d at 143–44; *see also Nations*, 676 S.W.2d at 284.

113. *See infra* Part VII.

VII. TRADITIONAL IMPUTATION OF CRIMINAL LIABILITY

There is no law that explicitly states that criminal liability can only be extended to human beings. However, there are laws that list only human beings, or persons, as the subject of criminal liability.[114]

For example, Article IV, section 2, clause 2 of the Constitution does not address the specifics of assigning criminal liability, but illustrates the fact that the law, as it stands, recognizes human beings as the only entity that can assume criminal liability for their actions.[115] The provision does this by specifically referring to a "person" who is charged with a crime.[116] The portion of the Article which states that a *person must be charged* has received criticism in years past.[117] However, the criticism was focused on the latter half of the phrase, while ignoring entirely the fact that the being that is charged must be a human.[118]

Moreover, the Fifth Amendment of the United States Constitution identifies only human beings as the section of society that must answer to criminal charges.[119] The Fifth Amendment uses a two-fold approach to make it quite clear that this law is applied to human beings.[120] First, the

114. *See* U.S. CONST. art. IV, § 2, cl. 2; *see also* U.S. CONST. amend. V; U.S. CONST. amend. VI.

115. U.S. CONST. art. IV, § 2, cl. 2.

> A ***person*** charged in any State with Treason, Felony, or other Crime, who shall flee from Justice, and be found in another State, shall on Demand of the executive Authority of the State from which ***he*** fled, be delivered up, to be removed to the State having Jurisdiction of the Crime.

Id. (emphasis added).

116. *Id.*; *see also* Wigchert v. Lockhart, 114 Colo. 485, 491 (1946) (finding that "no person may lawfully be removed from one state to another by virtue of the federal constitutional provision unless" at least one exception, discussed in the case, is satisfied).

117. *See* Kentucky v. Dennison, 65 U.S. 66, 66 (1860).

> A MOTION was made in behalf of the State of Kentucky, by the direction and in the name of the Governor of the State, for a rule on the Governor of Ohio to show cause why a mandamus should not be issued by this court, commanding him to cause Willis Lago, a fugitive from justice, to be delivered up, to be removed to the State of Kentucky, having jurisdiction of the crime with which he is charged.

Id.

118. *See id.* at 81–82.

119. U.S. CONST. amend. V.

> No ***person*** shall be held to answer for a capital, or otherwise infamous crime, unless on a presentment or indictment of a Grand Jury, except in cases arising in the land or naval forces, or in the Militia, when in actual service in time of War or public danger; nor shall any ***person*** be subject for the same offence to be twice put in jeopardy of life or limb; nor shall be compelled in any criminal case to be a witness against ***himself***, nor be deprived of life, liberty, or property, without due process of law; nor shall private property be taken for public use, without just compensation

Id. (emphasis added); *see also* United States v. Aranda-Hernandez, 95 F.3d 977, 979 (1996).

120. U.S. CONST. amend. V.

Fifth Amendment begins with the language that "no *person*" can be held to answer for a capital crime. Second, the Fifth Amendment states that *someone* called to answer for a crime cannot be a witness against "*himself*."[121] The syntactical language choices within the Fifth Amendment make it clear that the only intended targets are human beings.

Finally, the Sixth Amendment of the Constitution does not explicitly state that only a human being can be charged with a crime.[122] Instead, the sixth amendment refers to the entity that is charged as "the accused."[123] On its own, this would be the easiest way for non-human entities to also be charged with criminal liability. However, the Sixth Amendment overrules this possibility by using pronoun-specific language throughout the remainder of the amendment.[124]

Therefore, while there is no law on the books that states non-human entities can be subject to criminal liability, the language of the U.S. Constitution makes it clear that such liability is left only to human beings.[125] However, technology has greatly advanced since the ratification of the Constitution.[126] As such, the language of the Constitution does not reflect the current state of entities that should be held criminally liable.[127]

---

121. *Id.*

122. U.S. CONST. amend. VI.

> In all criminal prosecutions, the ***accused*** shall enjoy the right to a speedy and public trial, by an impartial jury of the State and district wherein the crime shall have been committed, which district shall have been previously ascertained by law, and to be informed of the nature and cause of the accusation; to be confronted with the witnesses against ***him***; to have compulsory process for obtaining witnesses in his favor, and to have the Assistance of Counsel for ***his*** defence.

Id. (emphasis added).

123. *Id.*

124. *See id.*

125. *See* U.S. CONST. art. IV, § 2, cl. 2; *see also* U.S. CONST. amend. V; U.S. CONST. amend. VI.

126. *See* Matt Turck, *Frontier AI: How Far are we From Artificial "General" Intelligence, Really?*, HACKERNOON (Apr. 18, 2018), https://hackernoon.com/frontier-ai-how-far-are-we-from-artificial-general-intelligence-really-5b13b1ebcd4e.

> This has been a recurring theme of science fiction for many decades, but given the dramatic progress of AI over the last few years, the debate has been flaring anew with particular intensity, with an increasingly vocal stream of media and conversations warning us that AGI (of the nefarious kind) is coming, and much sooner than we'd think. Latest example: the new documentary Do you trust this computer?, which streamed last weekend for free courtesy of Elon Musk, and features a number of respected AI experts from both academia and industry. The documentary paints an alarming picture of artificial intelligence, a "new life form" on planet earth that is about to "wrap its tentacles" around us.

Id.

127. *See* Geib, *supra* note 17.

VIII. BRIEF REVIEW PRIOR TO SUGGESTIONS FOR LEGISLATIVE REFORM

Prior to introducing a suggestion for legal reform, it is important to recap what has been established thus far. DLM are a subset of the study of artificial intelligence that focuses on the creation of machines that possess processing capabilities that are similar to the cognitive capacity of human beings.[128] DLM use data mining techniques and natural language processing to build a library of knowledge that is later accessed in every step of the decision-making process.[129] Each decision that the DLM makes is then weighted against its success and is stored for later usage.[130]

When similar situations are encountered in the future, the DLM is able to recall its past decisions and logically select the best option according to the weighted rankings.[131] Once the DLM collects sufficient information through data mining, the DLM's decision-making slowly moves farther away from the decision-making techniques that it learned from its source code.[132]

Since the DLM's decision-making process can be analyzed separately from the process that is programmed into the DLM by a human being, there exists the question of whether a DLM should be held criminally liable for its decisions. Current case law and legislation fail to answer this question. Therefore, the following suggestion for legal reform is intended to bridge the gap between the current state of legislation and looming technological advancements.

IX. SUGGESTION FOR LEGAL REFORM

Given that the current state of the law focuses on ways of imputing criminal liability to human beings only, there is a need for legal reform such that a DLM can be held liable for its actions. However, since the current elements of a criminal charge are specifically tailored to human beings,[133] there is a need for a new test that can be applied to DLM and AI in general. This section establishes a three-pronged test that was crafted to account for discrepancies within *mens rea* and *actus reus* as they apply to both human beings and DLM. However, prior to discussing

128. *See* Marr, *supra* note 1.
129. *See* Bengio, *supra* note 3, at 1–2.
130. *See* Best, *supra* note 53.
131. *See id.*
132. *See id.*
133. *See supra* Part VII.

the test proposed here, it is important to explore the historic ideology on expanding criminal liability to non-human entities.

A. Previous Attempts at Expanding Criminal Liability to Non-Human Entities

One of the earliest attempts of holding a non-human entity accountable for its conduct was written by Lawrence Solum in 1992.[134] The center of Solum's research focuses on whether AI can serve as a trustee, such that it can be left alone to make decisions for human beings.[135] Solum illustrates this idea through the use of a hypothetical where a computer software program is instructed to invest in publicly traded stocks.[136] Solum identified a number of reasons trusting the program would not be a sensible idea, including the fact that human intervention may be necessary to make a decision and the fact that a lawsuit can ensue if the software program does not invest wisely.[137]

While Solum proposed a timely test for measuring whether AI can function independently, his test does not account for modern day technological advances. As discussed throughout this note, one of the objectives in programming DLM is to create machines that are able to function sans human intervention.[138] Using the data mining techniques

134. Lawrence Solum, *Essay*, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231 (1992).

135. *Id.* at 1240. This first scenario speculates about the legal consequences of developing an expert system capable of doing the things a human trustee can do. Imagine such expert systems developing from existing programs that perform some of the component functions of a trustee. For example, the decision to invest in publicly traded stocks is made by a computer program in what is called "program trading," in which the program makes buy or sell decisions based on market conditions. Today, one also can buy a computer program that will automatically issue instructions to pay your regular monthly bills by sending data to a bank or service via modem. It is not difficult to imagine an expert system that combines these functions with a variety of others, in order to automate the tasks performed by the human trustee of a simple trust. *Id.* at 1240–41.

136. *Id.* at 1240–42.

137. *See id.* at 1242. But there may be times when the human being is called upon to make a decision. For example, suppose the trust is sued. Perhaps a beneficiary claims that the trust has not paid her moneys due. Or imagine that an investment goes sour and a beneficiary sues, claiming that the trustee breached the duty of reasonable care and skill. If such events occur with regularity, the trustee will develop a routine for handling them. She might routinely refer such disputes to her attorneys. In time, the expert system is programmed to handle this sort of task as well. It processes the trustee's correspondence, automatically alerting the trustee when a letter threatening suit is received or process is served. The system prepares a report on the relevant trust from its electronic records and produces a form letter for the trustee's signature to be sent to the trust's attorneys. As the capabilities of the expert system grow, the need for the human trustee to make decisions gradually diminishes. *Id.*

138. DLM should be able to function without human intervention.

with which DLM are programmed, there would be little need for human intervention for a machine that is designed to trade stocks.

The following subsection builds upon the strengths of DLM to present a three-pronged approach for expanding criminal liability.

B. Three-Pronged Proposed Solution

The following liability approach uses the traditional *mens rea* and *actus reus* elements of criminal liability to illustrate how a DLM may be held accountable for its actions. Throughout this subsection modern day examples will be used to assist in the visualization of the approach.

1. Prong One: Review Source Code to Understand *Mens Rea*[139]

The first step is geared toward understanding the *mens rea* of a DLM. As described above, *mens rea* refers to one's mind or, more specifically, one's criminal intent.[140] As it applies here, a DLM's mind is the amalgamation of code and algorithms that comprise its central decision-making center.[141] When taken together, the code and the algorithms provide the DLM with a rudimentary introduction into the tasks that the DLM will be asked to perform later on.[142] Therefore, to gain a true understanding of what the DLM is designed to do, it is important to perform a line by line review of the DLM's code and algorithms.[143] Only by reviewing the DLM's code and algorithms does it become apparent

139. To obtain maximum benefit from this test, it is best if step one is completed by a human being. Having a human being perform the first step is beneficial for two reasons. First, it saves the human being the trouble of having to design a secondary machine to study the code that was programmed into the first machine. A machine is accustomed to reading and processing code so it is possible that the machine would not provide a thorough explanation of the code. *See supra* Part IV. In the same way that machines perform natural language processing to understand human beings, human beings may need an intermediary step before they are able to understand the translations from the machine. *See supra* Part IV. Thus, the second reason why it is beneficial to have a human being perform step one is that a human being wrote the code and algorithms that are programmed into the DLM. Therefore, a human being would likely be able to process the meaning of the code more precisely than a machine. As the processing power of DLM continues to progress, it is likely that a machine may one day be able to perform step one as efficiently, if not more, as a human being. *See supra* Part IV.

140. *See* People v. Conley, 543 N.E.2d 138, 143–44 (Ill. App. Ct. 1989); *see also* State v. Nations, 676 S.W.2d 282, 284 (Mo. Ct. App. 1984).

141. *See supra* Part IV (detailing how DLM programming is conducted and how DLM are designed to execute autonomous decisions).

142. *See supra* Part IV (detailing how DLM programming is conducted and how DLM are designed to execute autonomous decisions).

143. *See supra* Part IV (detailing how DLM programming is conducted and how DLM are designed to execute autonomous decisions).

what the DLM was designed to do.[144] This will establish the starting point for determining whether the DLM's subsequent decisions drastically veered away from the initial code. This will also help with identifying the information that the DLM was initially given, which will be extremely important for tracking the progression of the DLM's conduct.

For example, consider the 2014 news story where an AI was designed to buy products through the online black market.[145] The AI was initially programmed to purchase items over the internet, but was not specifically asked to purchase illegal items.[146] As the AI grew accustomed to purchasing items, it started to explore the depths of the Internet and eventually took to purchasing illegal items, such as ecstasy pills.[147] Applying the first prong of this note's three-pronged approach to determine whether criminal liability can be extended to the AI, it is necessary to review the AI's source code to determine exactly what the AI was initially programmed to do.[148] Although the source code was not publicly released, it is likely that the AI was simply programmed to select online items that were within the weekly budget that the AI was granted.[149] An analysis of the AI's code would be extremely beneficial in understanding the AI's *mens rea* and whether the AI was instructed to purchase illegal items or whether the purchase of illegal items was behavior that the AI learned independently.

144. *See supra* Part IV (detailing how DLM programming is conducted and how DLM are designed to execute autonomous decisions).

145. *See* Power, *supra* note 82.

> A robot deployed on the dark web over the past few weeks has bought a pair of fake Diesel jeans, a baseball cap with a hidden camera, a stash can, a pair of Nike trainers, a decoy letter (used to see if your address is being monitored), 200 Chesterfield cigarettes, a set of fire-brigade issued master keys, a fake Louis Vuitton handbag, and 10 ecstasy pills.

Id.

146. *See id.* ("London-based Swiss artists !Mediengruppe Bitnik, Carmen Weisskopf and Domagoj Smoljo) [sic], coded the Random Darknet Shopper, an automated online shopping bot, and instructed it to spend $100 in bitcoin per week on a darknet market that lists over 16,000 items, not all of them illegal.").

147. *See id.*

148. *See supra* Part IX (details more information on why this would be helpful).

149. *See supra* Part IV (details more information on how DLM programming is conducted and how DLM are designed to execute autonomous decisions).

2. Prong Two: Comparing Conduct with Source Code to Identify *Actus Reus*

The second step focuses on the *actus reus* of the DLM. As described above, *actus reus* refers to the physical carrying out of an alleged crime.[150] As it applies here, this would require a comparison of the DLM's initial code from Prong One with the DLM's conduct that led toward its potential criminal liability.

In order to get closer to establishing criminal liability of the DLM, it is important to show that the alleged criminal conduct originated from decisions that were made by the DLM, using its library of knowledge that was cultivated through data mining. With respect to criminal liability, these decisions are far different from the decisions a DLM makes using only the code that was programmed into it by a human being. If the DLM is believed to have committed a criminal act while executing the code received from a human being, then the DLM has not acted independently.[151] In other words, the DLM was merely following instructions and should not be held criminally liable in such scenarios.

On the other hand, the weighted-decisions that a DLM makes after performing data mining are, ideally, loosely connected to the initial code that the DLM received from its programmer.[152] After performing data mining and consistently maintaining a library of knowledge, the DLM has much more data that it can use throughout these decision-making processes.[153] As such, the likelihood of the DLM making a decision pursuant to its own analysis is much greater than the likelihood of the DLM blindly following its source code.

Therefore, since there is a stark difference in the ways a DLM may make decisions, there exists the need to compare the outputs of the DLM's source code to the DLM's alleged criminal conduct.[154] In other words, it is important to understand the difference between the DLM's alleged criminal conduct and the conduct the DLM was initially programmed to execute.

150. *See* Martin v. State, 17 So. 2d 427, 427 (Ala. Ct. App. 1944).

151. *See* Geib, *supra* note 17 (discussing the point of view of European lawmakers who believe "electronic personalities" should be granted legal personhood such that it can be held accountable for its conduct).

152. *See id.*

153. *See supra* Part IV (detailing more information on how DLM programming is conducted and how DLM are designed to execute autonomous decisions).

154. *See supra* Part V (detailing an in-depth discussion of the elements of criminal liability).

To better understand this prong, consider the AI from Prong One that was engaged in purchasing items from the dark side of the Internet.[155] It is unclear whether the AI was instructed to purchase illegal items; however, by tracking the AI's conduct and purchases, it is obvious that the AI in fact purchased illegal narcotics off of the Internet.[156] Therefore, using the methods prescribed under Prong Two, it is important to compare the AI's source code to the AI's behavior. If the AI was initially programmed to purchase illegal items from the internet, then the AI would satisfy the requisite *actus reus*, but would not satisfy the requisite *mens rea*.[157] By contrast, if the AI was not programmed to purchase illegal items, but proceeded to do just so, then the AI would satisfy the requisite *mens rea* and *actus reus*. Therefore, a comparison between the initial source code and the conduct is imperative.

3. Prong Three: Comparing Conduct with the Source Code Commands

The third step again addresses the *mens rea* of the DLM. It requires a determination of how far removed the DLM's conduct was from its initial code. Once the differences from Prong Two have been spotted and recorded, it is necessary to determine how dissonant the two classes of conduct are. In other words, there must be a determination of whether the DLM did or did not veer too far away from the tasks it was originally instructed to perform. Not only is this useful in determining *mens rea*, but it can also be useful for determining whether punishment will be inflicted upon the DLM.

To faithfully execute this step, it may be necessary to appropriate the process used in Prong Two—a play by play review of the alleged criminal conduct.[158] When combined with the line by line review of the DLM's source code from Prong One,[159] the play by play review might simplify the comparison. With a simplified comparison process, it may be easier to determine whether the alleged criminal conduct follows from the conduct prescribed in the source code. Accurate execution of this step requires an analysis of the algorithms that the DLM used. However, this will likely be difficult to complete given that the level of programming sophistication in the DLM is far beyond that of human beings.[160]

155. *See* Power, *supra* note 82.
156. *See id.*
157. The *mens rea* would not have been voluntarily satisfied.
158. *See supra* Prong Two.
159. *See supra* Prong One.
160. *See* Mak, *supra* note 71.

Once again, Prong Three is easier to visualize using the AI from Prong One that was engaged in purchasing items from the dark side of the Internet.[161] Applied here, it is important to compare the exact conduct of the AI with the commands that are listed in the original source code. If the conduct flows directly from the source code, then it is clear that the machine faithfully executed its orders. However, if the conduct does not follow from the source code then it is clear that somewhere along the line the AI developed its own algorithm that it used to purchase illegal items. A showing of such independence would support the finding that the AI is capable of independent decision-making and should be liable for its criminal conduct.

Thus, after applying each factor, if the results show that the DLM had the requisite *mens rea* and *actus reus* for the crime, then it may be possible to hold the DLM criminally liable for its actions. Also, depending on how far removed the DLM's conduct was from the original source code, officials may be able to use the results of Prong Three to determine a just punishment for the DLM.[162] However, this three-pronged approach may deliver unfavorable results, causing individuals to object to the expansion of criminal liability.

X. POTENTIAL OBJECTIONS TO LEGAL REFORM

Given the controversial nature of this topic, pushback is expected. This section aims to identify and address objections that will likely result from discussions of this topic.

A. First Objection: AI Should be Regarded Solely as Property

The first reasonably foreseeable objection is that human beings should be held liable for the criminal conduct of the machines they create.[163] This objection is based on the theory that DLM are the property of the human beings who create it.[164] Under this theory, the fact that

161. *Cf.* Power, *supra* note 82.

162. *See* Geib, *supra* note 17 (discussing the point of view of European lawmakers who believe "electronic personalities" should be granted legal personhood such that it can be held accountable for its conduct).

163. Bert-Jaap Koops, et al., *Bridging the Accountability Gap: Rights for New Entities in the Information Society?*, 11 MINN. J.L. SCI. & TECH. 497 (2010).

164. *Id.* at 531–32 ("This objection refers to Locke's proposition that artifacts that are the product of human labor are the property of those who made them. For Locke, a human being is not made by his parents, but by God, implying that a parent does not have ultimate control over his children.").

"human beings are made naturally while AIs are made artificially, should make a difference."[165]

In addressing the first objection, it is important to note the emphasis on the fact that a human being should be liable for what he or she creates.[166] While this is a sound argument, it misses the fact that DLM are also capable of creating their own machines. One characteristic of DLM (parent-DLM) is that once they have strengthened their library of knowledge, they are capable of using their library of knowledge to create a secondary DLM (child-DLM) that will inherit its library of knowledge. In this scenario, the human being is not the only being capable of creation.

Furthermore, Solum's response to this objection is quite noteworthy, finding that "whether an AI should be granted constitutional rights depends on it being a person and, if this is the case, an AI should not be owned by another person."[167] Pursuant to the discussion above,[168] the advanced cognitive capacity of DLM can be measured through Joseph Fletcher's personhood elements.[169] While a reader should not give complete deference to Fletcher's elements in every instance, the elements, applied here, show that a DLM has the potential to be categorized as a person.[170]

B. Second Objection: Legal Rights for Human Beings Should Trump Legal Rights for Non-Human Entities

The second reasonably foreseeable objection to the ideas set forth in this note is that legislative officials should focus on advancing legal rights for human beings prior to engaging in a battle for legal rights for non-human entities.[171] While this is a fair objection, it is important to note that advancing the legal recognition of DLM and AI does not diminish the legal rights of human beings. In fact, the legal rights of human beings remain untouched. Moreover, it is crucial that legislative officials are aware of current technological advances such that they are able to respond to them through useful legislation.[172] As Solum states, "just

165. *Id.* at 532.
166. *Id.* 531–32.
167. *Id.* at 532.
168. *See supra* Part III for an in-depth discussion on the personhood elements.
169. *See supra* Part III (detailing an in-depth discussion on the personhood elements).
170. *See supra* Part III (detailing an in-depth discussion on the personhood elements).
171. *See* Koops, *supra* note 163, at 526–27 ("A more fundamental argument against constitutional rights for non-humans holds that the concept of person is intrinsically linked to humans. The idea is that, since nonhumans do not share our biological constitution, they cannot be conceptualized as persons.").
172. *See supra* text accompanying notes 17–18.

because today we cannot imagine non-humans to qualify for personhood, does not imply that, in the future, AIs could not develop into non-biological entities that are intelligent, conscious, and feeling in ways that change our very concept of personhood."[173] Thus, given the discussion in previous sections, it is possible that DLM will reach the cognitive level of human beings, such that the expansion of legal rights will naturally follow.

XI. CONCLUSION

Current U.S. laws do not explicitly state that criminal liability can only be assigned to human beings. However, criminal liability cases that are presented to the court have always involved solely human beings. Until recent advancements in technology, DLM have neither been party to the conversation of criminal liability nor been assigned criminal liability for their actions. DLM are traditionally created and used for data collection in an attempt to replicate and enhance the thought processes observed in human beings.[174] They are programmed to harvest as much information as possible and then "to perform automatic feature extraction from raw data, also called feature learning."[175] In doing so, the DLM "learn[s] complex functions [by] mapping the input to the output directly from data[] without depending completely on human-crafted features."[176] The elimination of human-created features[177] and, by

173. *See* Koops, *supra* note 163, at 526 (emphasis omitted).

174. *See* Bernard Marr, *10 Amazing Examples of How Deep Learning AI is Used in Practice?*, FORBES (Aug. 20, 2018, 12:11 A.M.), https://www.forbes.com/sites/bernardmarr/2018/08/20/10-amazing-examples-of-how-deep-learning-ai-is-used-in-practice/#17af640df98a (listing the ways in which deep learning machines are used and the fact that they are created based off of the cognitive thought process of human beings).

175. Brownlee, *supra* note 5.

176. *Id.*

> The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones. If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, we call this approach to AI deep learning.

Id.

177. *See id.* Human-crafted features generally refers to the code and algorithms that are written by human beings and are later incorporated into DLM. *Id.* The code and algorithms instruct the DLM on how to execute various tasks. *Id.* The use of human-crafted features means the DLM is still in the early stages of learning how to perform data collection. *Id.* As the DLM learn how to perform data collection, its reliance on the algorithm decreases. *Id.* Decreased reliance means the DLM is learning how to function independently of its human creator. *Id.*

extension, human interference prepares the DLM for autonomous decision making.[178]

As such autonomous decision-making processes grow stronger, DLM rapidly approach the threshold that assigns criminal liability to entities. The fact that DLM have not been party to the criminal liability conversation is boggling, given that the decision-making processes of DLM are beyond that of human beings to the point where DLM may be capable of satisfying the *actus reus* and *mens rea* requirements of criminal liability. Thus, this note sought to bridge this gap using a three-pronged legislative reform aimed at bringing non-human entities closer to being held accountable for potential criminal behavior.

178. Autonomous decision making requires the integrated use of a plethora of algorithms. However, "[a]lgorithmic autonomy and transparency seem have [sic] an inverse relationship—as these algorithms become increasingly better at 'learning' and adjusting, it becomes more difficult to understand where the biases occur." Kwan, *supra* note 11.