THE CENTRE FOR internet & society

Artificial Intelligence – Literature Review

By **Shruthi Anand**
Edited by **Amber Sinha** and **Udbhav Tiwari**
Research Assistance by **Sidharth Ray**

The Centre for Internet and Society, India

Designed by **Saumyaa Naidu**

Contents

Introduction

With origins dating back to the 1950s Artificial Intelligence (AI) is not necessarily new. With an increasing number of real-world implications over the last few years, however, interest in AI has been reignited over the last few years. The rapid and dynamic pace of development of AI have made it difficult to predict its future path and is enabling it to alter our world in ways we have yet to comprehend. This has resulted in law and policy having stayed one step behind the development of the technology.

Understanding and analyzing existing literature on AI is a necessary precursor to subsequently recommending policy on the matter. By examining academic articles, policy papers, news articles, and position papers from across the globe, this literature review aims to provide an overview of AI from multiple perspectives.

The structure taken by the literature review is as follows:

1. Overview of historical development;
2. Definitional and compositional analysis;
3. Ethical & Social, Legal, Economic and Political impact and sector-specific solutions;
4. The regulatory way forward.

This literature review is a first step in understanding the existing paradigms and debates around AI before narrowing the focus to more specific applications and subsequently, policy-recommendations.

1. Historical Evolution of AI

The history of the development of AI has been fairly recent, with its origins being traced to the mid-20th century. Despite its seemingly recent origins, however, there exist some influences that have contributed greatly, albeit indirectly, to the envisagement of AI. The genesis of AI can be credited to the contributions of various academic fields, not limited to art, history, philosophy logic and mathematics. This section seeks to identify some of those factors, in addition to providing a brief historical account of the notable breakthroughs in the evolution of AI.

1.1 Contributions to the Genesis of AI

a. Philosophy

The contribution of philosophy to AI is undisputed. For George Luger[1], the natural starting point when examining the philosophical foundations of AI is to begin with Aristotle, as his philosophical work formed the basis for modern science. The great philosopher-scientist Galileo, whose observations and writings contradicted the 'obvious truths of the age' and used mathematics as a tool to test them, challenged our understanding that the world always worked as it appeared.[2] Epistemological work such as Rene Descartes' on the theory of the mind was also very influential to AI, specifically in two ways:

- It established the separation of the body from the mind. This forms the basis of the methodology of AI – mental processes have an independent existence and follow their own laws; and

1 Luger, G. F. (1993). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 5/e.* Pearson Education India.

2 Ibid., 6.

- Once it became established that the mind and body were separate, it became necessary to find innovative ways to connect the two.

Luger considers the empiricist and rationalist traditions of philosophy to be the most prominent pressures for the evolution of AI.[3] For a rationalist, the external world can be clearly reconstructed through the rules of mathematics. Empiricists, on the other hand, do not believe in a world of clear and distinct ideas, but in knowledge being explained through an introspective but empirical psychology. This knowledge, according to Luger, plays a significant role in the development of AI structures and programs.[4] Therefore, the philosophical foundations of AI regarded thinking as a form of computation.

Russell and Norvig[5] premise their philosophical analysis of intelligent agents on the notion that intelligence is a manifestation of rational action; an intelligent agent takes the best action in a given situation. Philosophy conceptualised this idea, which later formed the basis of AI, by equating the behaviour of the mind to that of a machine – *it operates on knowledge encoded in some internal language, and that thought can be used to choose what actions to take*.[6]

b. Mathematics & Logic

Once thinking came to be seen as a form of computation, the next steps were to formalize and mechanise it. Luger[7] defines this as the phase involving the "development of formal logic".[8] Both Charles Babbage and Ada Lovelace's works focussed on this, wherein the patterns of algebraic relationships were treated as entities that could be studied, resulting in the creation of a formal language for thought.[9] The author also credits George Boole for his contribution to the field – Boole's operations of "AND", "OR" and "NOT" have remained the basis for all operations in formal logic.[10] Whitehead and Russell's work has also been acknowledged by Luger, with their treatment of mathematical reasoning in purely formal terms acting as a basis for computer automation.[11]

Russell and Norvig[12] opine that mathematics was used to manipulate statements of logical certainty as well as probabilistic statements, in addition to laying the groundwork for computation and algorithms. Subsequently, the field of economics, by formalizing the problem of decision-making to maximize outcome, furthered the contribution of mathematics.

c. Biology

Apart from philosophy and logic, Nils Nilson[13] believes that aspects of biology and "life" in general have provided important clues about intelligence.[14] This includes principles relating

---

3 Ibid., 8.

4 Ibid., 9.

5 Russell, S., Norvig, P., & Intelligence, A. (1995). A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs, 25*, 27.

6 Ibid., 30.

7 Supra, note 1.

8 Ibid.

9 Ibid., 11.

10 Ibid.

11 Ibid., 12.

12 Supra, note 5.

13 Norvig, P. (2011). The Quest for Artificial Intelligence, Nils J. Nilsson. Cambridge (2010).

14 Ibid., 34.

to neurons & the workings of the human brain, psychology and cognitive science, evolution, development & maturation and bionics.

Russell and Norvig[15] are more specific, pointing out that neuroscience, in discovering that the human brain can be said to be similar to computers in some ways, provided the intuitive basis for AI. This was then supplemented by psychology through the idea of humans and animals as nothing but machines that process information.

d. Engineering

Nilson[16] and Russell & Norvig[17] note that engineering has made a more direct contribution to AI by being the tool used to create machines on which AI application are allowed to run. Particular facets of the field that have made this possible include:

- Automata, Sensing, and Feedback;
- Statistics and probability; and
- The computer – whether through computation theory, the digital computer or the new age "thinking computer".

1.2 Historical Account of AI

The White House's National Science and Technology Council[18] traces the roots of AI to the 1940s, in McCulloch and Pitts', "A Logical Calculus of the Ideas Immanent in Nervous Activity". The idea of artificial intelligence was crystallized by Alan Turing, in his famous 1950s paper "Computing Machinery and Intelligence". The fundamental question posed in that paper was Can machines think?, which Turing sought to answer using what came to be known as the Turing Test. He also believed that a machine could be programmed to learn from experience, much like a child. However, the term 'Artificial Intelligence' itself was not coined until 1956.[19]

The Turing Test became the gold standard for AI-development. Luger[20] identifies its defining features[21]:

- It provides an objective notion of intelligence;
- It enables unidimensional focus by containing a single standard of measurement. This avoids side-tracking with questions such as whether the machine really knows that its thinking;
- It eliminates bias by centering the focus of a neutral third-party on the output.

At the same time, he notes the significant flaws:

- It tests only for problem-solving skills, and not for other forms of human intelligence;
- By using a human standard to measure machine intelligence, the Test straight jackets it into a human mold. This completely avoids a consideration of the possibility that machine and human intelligence are simply different and cannot be compared and contrasted.

---

15 Supra, note 5.

16 Supra, note 13.

17 Supra, note 5.

18 House, W. (2016). Preparing for the future of Artificial Intelligence. *Executive Office of the President, National Science and Technology Council, Committee on Technology*.

19 Ibid., 5.

20 Supra, note 1.

21 Ibid., 14.

According to Nilson[22], the emergence of AI as an independent field of research strengthened and was further strengthened by three important meetings – a 1955 session on Learning Machines held in conjunction with the 1955 Western Joint Computer Conference in Los Angeles, a 1956 summer research project on Artificial Intelligence convened at Dartmouth College and a 1958 symposium on the "Mechanization of Thought Processes" sponsored by the National Physical Laboratory.[23]

Initially, development of AI was primarily to solve mathematical problems, puzzles or games by relying on simple symbol structures. In the 1960s, however, programs were required to perform more intellectual tasks such as solving geometric analogy problems, storing information, answering questions and creating semantic networks, thereby requiring more complex symbol structures termed semantic representations.[24]

The next big breakthrough was in the creation of the General Problem Solver (GPS).[25] The GPS pioneered the first approaches of 'thinking humanly' – it was designed to imitate human problem – solving protocols, solving puzzles using the same approach as humans would.[26]

In 1958, the computer scientist John McCarthy made three crucial contributions to AI[27]:

- He defined Lisp, the language that would later become the dominant programming language for AI;
- He invented time sharing; and
- In a 1958 paper, he described the Advice Taker, which was seen as first end-to-end AI system.

Natural Language Processing (NLP), a fundamental prerequisite for AI-development, got a shot in the arm in the 1950s and 60s due to increased government funding.[28] Natural languages such as English were able to be understood by the machine and translated into a language that was understandable to computers, in turn resulting in the re-conversion into natural language as the output.[29]

The 1960s also saw computer chess programs improve gradually from beginner-level play to mid-level play. However, a fundamental distinction was noticed between how humans and computer played chess- computers would scan through all the permutations and combinations of maneuvers, finally making the one maneuver that would yield maximum benefit. Humans, on the other hand, utilize accumulated knowledge along with reasoning to verify that the proposed maneuver is good in the present instance.[30] According to Nilson, "*specific knowledge about the problem being solved, as opposed to the use of massive search in solving the problem, came to be a major theme of artificial intelligence research during this period*".[31]

While AI research dealt with "toy" problems until the early 1970s – whether it was in solving puzzles or games – the focus gradually shifted to real-world problems, leading to the

22 Supra, note 13.

23 Ibid., 73.

24 Ibid., 131.

25 Supra, note 5.

26 Ibid., 19.

27 Ibid.

28 Ibid., 160.

29 Ibid., 141.

30 Supra, note 13, at 253.

31 Ibid.

creation of sub-categories such as NLP, expert systems, and computer vision.[32] There were two primary reasons for this shift:

- The power of AI had developed to a point where focussing on real-world applications seemed possible; and
- Sponsors of AI research were required by the US government to support research relevant to real time military needs.

While NLP and AI were earlier largely restricted to text-based systems, the 1970s and 80s saw its foray into speech recognition and comprehension.[33] Advances in this field were made in pursuit of specific applications such as computer vision, aerial reconnaissance, cartography, robotics, medicine, document analysis, and surveillance.[34] Overall, the funding for and enthusiasm in AI research was sustained by the promise of its applications, especially those in expert systems.[35] Described as the AI Boom, it was bolstered by Japan's "Fifth Generation Computer Systems" project and Europe's ESPRIT programme.[36] The 1980s also saw an increasing amount of attention being paid to machine learning, which has come to be one of the most prominent branches of AI.[37]

During this time, AI cemented itself as a separate industry, so much so that most major corporations in the US had separate groups working on AI.[38] This can also be seen by the fact that the AI industry had grown from being valued at a few million dollars in 1980 to a billion dollar industry in 1988. This was almost immediately followed by the AI Winter, where a multitude of AI companies, unable to deliver on their earlier grand promises, fell by the wayside.[39]

Research on end-to-end intelligence agents began in 1995, and continues to this day. The availability of computers, large databases and the growth of the internet have allowed AI to expand rapidly and contribute to real-world problems. AI has become both autonomous and ubiquitous, existing in everything from home appliances, driver assistance systems to route-finding in maps. Applications involving full-fledged intelligence agents can be seen in technologies such as internet bots and tools such as search engines, recommender systems and website aggregators.

1.3 AI and Big Data

The AI of today is heavily reliant on the collection, usage and processing of big data. Bernard Marr[40] notes that data is invaluable in AI devices understanding how humans think and feel, thereby speeding up their learning process. It is cyclical – *the more information there is to process, the more data the system is given, the more it learns and ultimately the more accurate it becomes*. In Marr's opinion, AI's growth was earlier restricted due to:

- the limited availability of data sets; and
- Their nature as sample data sets instead of real-time, real-life data.

---

32 Ibid., 265.

33 Ibid., 267.

34 Ibid., 327.

35 Ibid., 343.

36 Ibid., 345.

37 Ibid., 495.

38 Supra, note 5.

39 Ibid.

40 Marr, B. (2017, July 15). Why AI Would Be Nothing Without Big Data. Retrieved November 25, 2017, from **https://www.forbes.com/sites/bernardmarr/2017/06/09/why-ai-would-be-nothing-without-big-data/#6f3b02994f6d.**

With greater availability of real-time data and the increasing ability to process large amounts of it in seconds, AI has transitioned into a data-first approach.

Randy Bean[41] agrees with Marr's argument, noting that the availability of greater volumes and sources of data is enabling capabilities in AI and machine learning that remained dormant for decades due to lack of data availability, limited sample sizes, and an inability to analyze massive amounts of data in milliseconds. There are three critical ways in which big data is now empowering AI:

- Big data technology – Huge quantities of data that previously required expensive hardware and software can now be easily processed; also referred to as "commodity parallelism."
- Availability of large data sets – New forms of data such as ICR, transcription, voice and image files, weather data, and logistics data are now increasingly available.
- Machine learning at scale – "Scaled up" algorithms such as recurrent neural networks and deep learning are powering the breakthrough of AI.

The author concludes with the observation that while the first wave of big data was about speed and flexibility, the next will be all about leveraging the power of AI and machine learning to deliver business value at scale.

The report of the White House's Committee on Technology[42] credits three factors for the current wave of progress in AI, all related to data:

- the availability of *big data* from sources including e-commerce, businesses, social media, science, and government;
- which provided raw material for dramatically *improved machine learning approaches and algorithms*;
- which in turn relied on the capabilities of *more powerful computers*.

Russell and Norvig[43] also point out that while the algorithm was earlier the fulcrum around which computer science revolved, recent work in AI has changed that focus, with data becoming the new fulcrum. In fact, Banko and Brill concluded, using an experiment, that a *mediocre algorithm with 100 million words of unlabeled training data outperforms the best known algorithm with 1 million words.*[44] Similarly, Hays and Efros were able to demonstrate the same principle using photos, concluding that the increase in accuracy of the algorithm was directly proportional to the amount of data fed into it.

On the other hand, Najafabadi et. al.[45] identify areas where Deep Learning would require further exploration to deal with some problems observed in big data analytics. These include determining what volume of input information is necessary for useful representation from deep learning algorithms and defining the criteria for obtaining good data abstractions and representations.

The AI of today, therefore, looks very different to its predecessors. With the number of alterations to the technology and its corresponding capabilities over the years, there might be some uncertainty on the meaning and composition of AI, which the next section seeks to examine.

---

41 Bean, R. (2017, May 08). How Big Data Is Empowering AI and Machine Learning at Scale. Retrieved November 25, 2017, from **http://sloanreview.mit.edu/article/how-big-data-is-empowering-ai-and-machine-learning-at-scale/**.

42 Supra, note 18.

43 Supra, note 5.

44 Ibid., 28.

45 Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1.oo

2. Definitional and Compositional Aspects of AI

As mentioned before, AI has applications on multiple aspects of our live – however, there continues to be ambiguity on both the definition of AI as well as on its constituent elements. This section seeks to understand the literature on these questions.

2.1 What is AI?

Pei Wang[46] does not believe that there is one fixed definition for AI. He lists five ways in which AI can be defined – by structure, by behaviour, by capability by function and by principle – these are in order of increasing generality and decreasing specificity.[47]

Nils J. Nilsson[48] breaks up AI into its components: Artificial (machine, as opposed to human) + intelligence. In order to gauge the intelligence of an entity, Nilsson falls back on the Turing Test. Moreover, he believes that with increased complexity of the machine comes increased intelligence.[49]

James Vincent[50] notes that one of the difficulties in using the term artificial intelligence is that it is tricky to define. In fact, as soon as machines have conquered a task that previously only humans could do — whether that's playing chess or recognizing faces — then it's no longer considered to be (a mark of) intelligence (known as the "AI Effect").

Ben Coppin[51] begins with a simple definition of AI as the *study of systems that act in a way that, to any observer, would be intelligent.*[52] However, this is an incomplete definition, since AI may be used to solve (simple and complex) problems which form part of the internal structure of complex systems. He then redefines the term as *AI involves using methods based on the intelligent behaviour of humans and other animals to solve complex problems.*[53] According to Coppin, the second definition is more useful when considering the difference between strong AI and weak AI.

Russell & Norwig[54] have organized AI into four categories based on their capabilities – Thinking Humanly, Thinking Rationally, Acting Humanly and Acting Rationally.

- Acting Humanly – The relevant test for this category is the Turing Test. These would possess the capabilities of *natural language processing to enable it to communicate successfully in English, knowledge representation to store what it knows or hears, automated reasoning to use the stored information to answer questions and to draw new conclusions, machine learning to adapt to new circumstances and to detect and extrapolate patterns, computer vision to perceive objects and robotics to manipulate objects and move about.*[55]

---

46 Pei, W. A. N. G. (2008). What Do You Mean by "AI"?. *Artificial General Intelligence*, 362-373.

47 Ibid., 6.

48 Nilsson, N. J. (1998). *Artificial intelligence: a new synthesis*. Elsevier.

49 Ibid., 5.

50 Vincent, J. (2016, February 29). What counts as artificially intelligent? AI and deep learning, explained. Retrieved November 25, 2017, from **https://www.theverge.com/2016/2/29/11133682/deep-learning-ai-explained-machine-learning**

51 Coppin, B. (2004). *Artificial intelligence illuminated*. Jones & Bartlett Learning.

52 Ibid., 4.

53 Ibid.

54 Supra, note 5.

55 Ibid., 2-3.

- Thinking Humanly – To create machines that mimic human thinking, we must understand how humans think. This is possible through introspection, psychological experimentation and brain imaging. Computer science and human psychology are brought together via cognitive science, which ultimately allows for the creation of what we now term as AI.[56]
- Thinking Rationally – This "laws of thought" approach stresses the importance of logic in the computational process.[57]
- Acting Rationally – AI is responsible for the creation of computer agents, who are expected to be rational agents and achieve the best possible outcome. This is not the same as "Thinking Rationally" because not all best outcomes are the result of logical inferences. This approach has two advantages – it is more generic in nature and more amenable to scientific development.[58]

2.2 What are its Constituent Elements?

Pedro Domingos[59] details the rival schools of thought within machine learning in his quest to determine the master algorithm – the Symbolists, the Connectionists, the Evolutionaries, the Bayesians, and the Analogizers.[60]

- Symbolists – They believe that all intelligence can be reduced to the manipulation of symbols, similar to the workings of mathematics. They use inverse deduction to find out the missing piece of knowledge and come up with a generic conclusion.
- Connectionists – They believe that the brain performs the action of learning and attempt to reverse-engineer it. The crucial focus revolves around the connections (like neurons) that are responsible for an error and its fix/es. They use backpropagation, which compares the desired output with what the system produces, adjusting the latter accordingly.
- Evolutionaries – They emphasize the phenomenon of natural selection, believing that it is sufficient to depict this onto a computer. They use genetic programming which, according to the author, *mates and evolves computer programs in the same way that nature mates and evolves organisms*.
- Beyesians – They concern themselves with uncertainty and determine how to deal with uncertain information. They use Bayes theorem to *incorporate new evidence into their beliefs* and apply probabilistic inference to *do that as efficiently as possible*.
- Analogizers – They use the process of inference to transpose similarities from one context to the next. They use the vector machine, which determines which contexts to remember and how to combine them to make new predictions.

Another way to understand the compositional mix of AI is to examine its technological applications[61]:

- Internet of Things – AI enables the conversion of unstructured data into knowledge;
- Cybersecurity – AI enables a more proactive cybersecurity system;
- Data analytics – Similar to IoT, AI turns otherwise random masses of data into actionable information.

---

56 Ibid., 3.

57 Ibid., 4.

58 Ibid., 5.

59 Domingos, P. (2016). *Master Algorithm*. Penguin Books.

60 Ibid., 94

61 Artificial intelligence (AI) and cognitive computing: what, why and where. (n.d.). Retrieved November 26, 2017, from **https://www.i-scoop.eu/artificial-intelligence-cognitive-computing/**.

According to Tinholt et. al.[62], AI – like the human brain – is composed of various processes and is multilayered in nature. Just like different brain components have different functions, AI consists of distinct levels of consciousness to fulfil distinct functions.[63] In the authors' opinion, AI systems consist of nine different levels of consciousness and is defined as "*technology which allows digital systems to monitor, analyse, act, interact, remember, anticipate, feel, moralise and create*":

- Monitor – technology that gathers information and records data;
- Analyse – technology that processes information, detects patterns and recognizes trends;
- Act – technology that can carry out tasks and processes;
- Interact – technology that is able to listen and respond with a solution;
- Remember – technology that is capable of finding information;
- Anticipate – technology that can recognize and predict patterns preemptively;
- Feel – technology that is able to understand and act on human emotions;
- Moralise – technology that can integrate morality into its decision-making process.

There is, thus, not too much agreement on the definition and composition of AI. This could be explained by the fact that AI may not be just one 'entity' – it is made up of multiple aspects and applications, each with a different definition, composition and use case. This understanding helps in the following section, which analyzes the impact of AI on various sectors and industries. While neither the magnitude nor the type of impact is uniform across sectors, it is clear that there is more than sufficient impact for stakeholders to sit up and notice.

3. AI – Sectoral Impact

This section broadly seeks to examine the ethical & social, legal, economic and political impact of AI. Under each sub-head, literature on the positive and negative implications are detailed, along with existing literature as regards potential solutions to the negative impact.

3.1 Ethical and Social Impact

The ethical and social impact of AI can be divided into two distinct areas of study – the human perspective and the AI perspective. The first part of the analysis involves looking at the ethical and social aspects of AI's impact on humans. Subsequently, we examine the implications of such progress on the way the technology itself might be perceived.

3.1.1 Human Perspective

The broad questions that are considered are:

- Can moral agency and control be ceded to AI? If so, in what circumstances?
- How will the expansion of AI transform society?
- What are the ethical issues as regards algorithms which form the basis of all AI?

Finally, some mitigating solutions are proposed as regards these questions.

62 Unleashing the Potential of Artificial Intelligence in the Public Sector, Capgemini Consulting Retrieved November 26, 2017, from **https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/10/ai-in-public-sector.pdf**

63 Ibid., 2.

a. Questions of Control

John P. Sullins[64] evaluates whether robots can be accorded the status of a moral agent by posing three questions[65]:

- Is the robot significantly autonomous? – Whether it is under the direct control of another agent or user.
- Is the robot's behaviour intentional? – *As long as the behaviour is complex enough that one is forced to rely on standard folk psychological notions of predisposition or 'intention' to do good or harm, then this is enough to answer in the affirmative to this question.*
- Is the robot in a position of responsibility? – If the robot fulfils a social role, it must possess a duty of care, which is only possible if it is a moral agent.

If all three are answered in the affirmative, he believes that a state of moral agency can be ascribed.

Sean O'Heigeartaigh[66] warns of the dangers of moral outsourcing. According to him, while human biases are many, they are also predictable and are relatively bounded, making it possible to correct for them. However, a machine error could lead to catastrophic and unpredictable consequences.

Pavaloiu and Kose[67] seem to agree, stating that morality cannot be outsourced to AI even if there is algorithmic accountability. Moreover, Paula Boddington[68] contributes to the debate by pointing out that virtue ethics and Kantian morality do not allow for the outsourcing of moral judgments, since another machine, however, sophisticated, would be unable to do the right things for the right reasons and in the right manner.[69]

Jos de Mul[70] believes that the delegation of morality to computer systems, contrary to undermining it, can cause it to strengthen. He debunks the two assumptions that are necessary to believe that moral outsourcing will weaken human morality, which are:

- Computers and humans are viewed as strictly different entities; and
- The formulation of moral goals is exclusively reserved for human beings.

While Jeffrey K. Gurney[71] does not specifically address the issue of moral outsourcing, he examines the use of a crash-optimization algorithm, the method by which an algorithm writer allows an autonomous vehicle to determine who or what to hit. He examines this algorithm through classic moral dilemmas such as the Shopping Cart Problem, Motorcycle Problem, The Car Problem, The Tunnel problem, The Bridge Problem and The Trolley Problem, pointing out the ethical and legal issues that might arise in each case.

---

64 Sullins, J. P. (2006). When is a robot a moral agent. *Machine Ethics*, 151-160.

65 Ibid., 28.

66 Would you hand over a moral decision to a machine? Why not? Moral outsourcing and Artificial Intelligence. (n.d.). Retrieved November 26, 2017, from **http://blog.practicalethics.ox.ac.uk/2013/08/would-you-hand-over-a-moral-decision-to-a-machine-why-not-moral-outsourcing-and-artificial-intelligence/**

67 Pavaloiu, A., & Kose, U. (2017). Ethical Artificial Intelligence-An Open Question. *arXiv preprint arXiv:1706.03021*.

68 Boddington, P. (2017). Towards a Code of Ethics for Artificial Intelligence.

69 Ibid., 90.

70 De Mul, J. (2010). Moral Machines: ICTs as Mediators of Human Agencies. *Techné: Research in Philosophy and Technology, 14(3)*, 226-236.

71 Gurney, J. K. (2015). Crashing into the unknown: An examination of crash-optimization algorithms through the two lanes of ethics and law. *Alb. L. Rev., 79*, 183.

One area in which there is recurring debate regarding the moral control exercised by AI is its application in Autonomous Weapons Systems (AWS). There seems to be overwhelming opinion against the creation of AWS. Purves et. al.[72] state that the dependence of the autonomous system on a variety of abstract factors that cannot be captured by a specific set of rules makes it incapable of replicating the moral judgment of humans, however sophisticated it may be. Even if such systems can make moral decisions similar to human beings, they cannot possibly be made for the right reasons, and will always be morally deficient in at least one respect.

Human Rights Watch (HRW) takes an emphatic stand against autonomisation, providing three reasons for the same:

- AWS lack human judgment and compassion, which humans are uniquely qualified to possess. They also lack prudential judgment – the ability of humans to apply broad principles to situations – and will blindly apply algorithms.
- AWS threaten human dignity, since they cannot comprehend the value of life nor the significance of its loss. Delegating moral life or death decisions in situations of armed conflict dehumanizes the process.
- AWS lack moral agency, which cannot be solved by giving them artificial moral judgment.

HRW lists out a number of others who have raised ethical and moral concerns regarding outsourcing moral decisions to AWS; these include nations such as Chile[73], UN Special Rapporteurs[74] and Nobel Peace Prize laureates[75].

b. Questions of Human Dignity

Nick Bostrom[76] examines whether the intersection of AI and body-mind augmentations is a threat to human dignity. Bostrom ultimately sides with the transhumanists, who believe in the widest possible technological choices for the individual, and addresses the concerns of the bioconservatives, who call for a ban on human augmentation. His underlying argument is that dignity is not restricted to the current state of humanity alone – post-human dignity is a definite possibility.

Jason Borenstein and Yvette Pearson[77] discuss the application of AI (specifically, robots) in the field of caregiving. Utilizing a capabilities approach analysis, the authors believe that the use of robots can maximize care and freedom for recipients of such care.

However, authors such as Noel Sharkey[78] are not in favour of utilizing AI for care-giving, whether it be the care of children or geriatrics. As regards the former, he notes that severe dysfunction occurs in infants (although the tests have been conducted only on animals so far) that develop attachments to inanimate entities. As regards the latter, he notes that leaving the elderly in the exclusive care of machines would deprive them of the human contact that is provided currently by caregivers.

---

72 Purves, D., Jenkins, R., & Strawser, B. J. (2015). Autonomous machines, moral judgment, and acting for the right reasons. *Ethical Theory and Moral Practice, 18*(4), 851-872.

73 Statement of Chile, CCW Meeting of States Parties, Geneva, November 13-14, 2014.

74 Heys, C. (2013). *Report of the special rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns*. UN; Kiai, M. (2013). Report of the Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association. *UN Doc. A/HRC/23/39, 24*.

75 Nobel peace laureates call for preemptive ban on. (2014, May 11). Retrieved November 26, 2017, from **https://nobelwomensinitiative.org/nobel-peace-laureates-call-for-preemptive-ban-on-killer-robots/**

76 Bostrom, N. (2005). In defense of posthuman dignity. *Bioethics*, 19(3), 202-214.

77 Borenstein, J., & Pearson, Y. (2010). Robot caregivers: harbingers of expanded freedom for all?. *Ethics and Information Technology*, 12(3), 277-288.

78 Sharkey, N. (2008). The ethical frontiers of robotics. *Science, 322*(5909), 1800-1801.

At a broader level, Jon Kofas[79] examines the impact of AI on the cybergeneration- the class of people for whom video games, cellphones and computers are the new reality. According to the author, AI will create an impact on the sense of identity and community in society, by undermining community culture and creating a world where transhumanism will be the norm. He paints a picture of an AI household – *the wealthier families will have androids in their homes, most likely helping to raise and educate their children, conditioning them about the existential nature of robots as an integral part of the family like the loveable dog or cat. The less affluent middle class would be able to rent-a-robot for the ephemeral experience of it. The lower classes will feel even more marginalized because AI robotics will be out of reach for them; in fact they will be lesser beings than the robots whose intelligence and functions will be another privilege for the wealthy to enjoy*.

c. Ethics in Algorithms/Machine Learning

Algorithms form one of the pillars on which AI-based applications are created. Understanding the ethical and social shortcomings of algorithms themselves is thus important.

Mike Ananny[80] provides three ethical dimensions through which to access a networked information algorithm (NIA) – the Kantian ("The study of what we ought to do"), Utilitarian ("Maximum benefit for maximum number") and Virtue ("Duty and consequences") models. First, Ananny looks into the manner in which algorithms create associations, whether it be political affiliation, sexuality or even a medical condition, and points out that it is doubtful whether these associations reflect real-life patterns. Second, he states that algorithmic decision making is based on recognizing patterns and similarity. This creates ethical issues such as a false sense of certainty, the discouragement of alternative explorations and the creation of apparent coherence among disparate objects.[81] Finally, the author points out that the disparate focus on time-bound action of an algorithm leads to a situation in which accuracy may be compromised.

Friedler et. al.[82] examine the meaning of a 'fair algorithm', borrowing from the philosophical as well as the computer science community. According to them, bias in algorithmic output stems from the choosing of the 'feature space'. They provide a mathematical definition of 'fairness', and demonstrate that fairness in output depends on the interactions between the construct space, observed space and the decision space of the algorithm.

Noting that algorithms and social actors are inherently different, Anderson & Sharrock[83] state that while the former is bound by mathematical instructions, the latter can exercise discretion.[84] However, they do not believe that fact and ethics are irreconcilable – despite being products of rationality, algorithms can be relied on to make satisfactory ethical decisions.

d. Potential Solutions

Machine Ethics

Some of the ethical issues can be resolved by aligning the objectives of machines with those of humans, ensuring both work toward the same goals. These values can either be

---

79 Kofas, J. (2017). *Artificial Intelligence: Socioeconomic, Political And Ethical Dimensions. Counter Currents*. Retrieved 5 December 2017, from **http://www.countercurrents.org/2017/04/22/artificial-intelligence-socioeconomic-political-and-ethical-dimensions/**

80 Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology, & Human Values*, 41(1), 93-117.

81 Ibid., 104.

82 Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2016). On the (im) possibility of fairness. *arXiv preprint arXiv:1609.07236*.

83 Anderson, R. J., & Sherrock, W. W. (2013). Ethical Algorithms: A brief comment on an extensive muddle. Retrieved November 26, 2017, from **http://www.sharrockandanderson.co.uk/wp-content/uploads/2017/04/Ethical-Algorithms.pdf**

84 Ibid., 5.

imparted during the programming stage, or by the AI itself observing in and learning from its environment. The top-down approach reflects the former, where the AI would be trained to compute the consequences of all its actions before narrowing on the one it decides to undertake. The bottom-up approach depicts the latter, where the AI derives answers from its experiences, making it more spontaneous.

Pavaloiu and Kose[85] note that while the top-down approach is appealing at first glance, it possesses inherent biases, the most prominent one being the data interpretation bias. For example, the AI might indirectly infer if an individual is depressed based on her social media feed, which could affect prospective employment.

Allen et. al[86] point out that the top-down approach allows the designer to tailor ability. At the same time, however, there would be conflict arising from the rules encoded and constant pressure to predict and compute outcomes for every action. It is because of the latter that this approach is untenable—very few computers possess such computational capacity and there would have to be a universal minimum standard for AI.

They note that the bottom-up approach, on the other hand, can be designed either via a reward and punishment system or a system based on levels of pain. However, the manner in which an AI learns from its environment is largely dependent on its design – poorly designed AI learns similar to a child brought up in a rough neighbourhood.

The authors propose a hybrid model which, while framing broad governing rules, would provide scope for learning through experience.

Nick Bostrom[87] compares superintelligent AI to general AI, but opines that it will surpass humans by much more. According to him, the only precaution against such kind of intelligence is to program empathy as one of its core objectives. Once this is done, exponential improvement will lead to an enhancement of this quality, thereby diluting AI's potential threat to mankind. Bostrom also addresses arguments that call for a halt to AI development due to its dangers. He states that AI is an inevitability; thus, utilizing precautionary measures before destructive AI is built would be a better solution.

Accountability Mechanisms

A popular method of ensuring accountability in the algorithm is through openness and transparency. This would enable the examination of the algorithm, its source and its implications on those who would ultimately be the recipients of the decisions made by such algorithms.[88] Academics have argued for disclosure of source code to restore 'a feeling of fairness',[89] to eliminate opacity[90] and also to enable reverse engineering.[91] However, others have pointed out that sheer complexity in machine learning systems means that being able to understand the algorithm in action, during the course of learning, is unlikely.[92]

---

85 Supra., note 67.

86 Allen, C., Smit, I., & Wallach, W. (2005). Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and Information Technology*, 7(3), 149-155.

87 Bostrom, N. (2003). Ethical issues in advanced artificial intelligence. *Science Fiction and Philosophy: From Time Travel to Superintelligence*, 277-284.

88 James, K. 2013. Open Data? The challenges of algorithmic accountability in big data; Diakopoulos, Nick. 2013. Algorithmic Accountability Reporting: On the investigation of black boxes

89 O'Neil, C. (2016). Weapons of Math Destruction.

90 Pasquale, F. (2015). The Black Box Society, 106 Harvard University Press.

91 Diakopolous, N. (2015). Algorithmic Accountability Reporting: On the Investigation of Black Boxes, Tow Centre for Digital Journalism.

92 Burrell, J. (2016). How the Machine 'Thinks' : Understanding Opacity in Machine Learning Algorithms. Big Data and Society, 1-12.

However, some authors[93] dispute the effectiveness of accountability, stating that *is made difficult by the apparent inaccessibility, complexity, obscurity, and intellectual property challenges posed by algorithms and the organisational settings within which they operate*.

Diakopoulos[94] notes that transparency as a solution is limited by:

- The fact that algorithms, more often than not, amount to trade secrets; making them transparent flies in the face of this concept;
- The high overhead costs that are incurred when algorithms are subject to transparency rules, unlike data transparency.

He suggests reverse engineering as an alternative to transparency to act as a check on algorithmic power – *the process of articulating the specifications of a system through a rigorous examination drawing on domain knowledge, observation, and deduction to unearth a model of how that system works*.[95]

Data Mining

Ruggieri et. al.[96] suggest using discriminatory classification rules to identify and analyze discrimination within an algorithm. They utilize data mining techniques to determine the existence of intended and unintended biases on the part of the designer[97] There have also been calls to revisit the works of Fathi et. al.[98] where they have spoken about the use of historical information to favor the choice of elements that have not been selected in the past.

Others

Yampolskiy[99] criticizes the use of machine ethics as a solution, arguing that research in this field is mostly jurisprudential in nature without practical relevance. He then provides recommendations:

- An isolation mechanism needs to be created, which would be capable of containing the interaction of the AI with the outside world;
- In order to truly keep AI in check, safety must be in-built. But super-intelligent AI, due to their self-learning capabilities, can bypass even these. Therefore, all research into Artificial general Intelligence must be banned, since it could lead to the obsoletion and eventual extinction of the human race;
- AI must also not be accorded any legal rights whatsoever.

Pavaloiu and Kose[100] provide some solutions as well:

- Erasing hidden bias;

---

93 Neyland, D. (2016). Bearing accountable witness to the ethical algorithmic system. *Science, Technology, & Human Values*, 41(1), 50-76.

94 Supra, note 91.

95 Ibid., 13; Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 398-415.

96 Ruggieri, S., Pedreschi, D., & Turini, F. (2010). Data mining for discrimination discovery. *ACM Transactions on Knowledge Discovery from Data (TKDD),* 4(2), 9.

97 Ibid., 14.

98 Fathi, Y. and Tovey, C. (1984). Affirmative action algorithms. *Mathematical Programming*, 34(3), 292-301.

99 Yampolskiy, R. V. (2013). Artificial intelligence safety engineering: Why machine ethics is a wrong approach. *Philosophy and theory of artificial intelligence*, 389-396.

100 Supra., note 67.

- Audits and algorithm scrutiny to prevent or correct black-boxes algorithms;
- Real-time simulations in a controlled environment;
- Developing human-friendly AGI algorithms without the power of applying decisions;
- AI Safety Engineering.

3.1.2 AI Perspective

As AI develops, it becomes more autonomous and capable of performing more and greater functions. As an increasingly independent non-human entity, a relevant ethical and social issue involves questioning the existence of a moral status for the technology itself. Such moral status could be a potential precursor to the conferment of more elaborate rights or even legal personhood for AI.

a. Moral Status

Richard Kemp[101] cautions against anthropomorphising AI, stating that three fallacies must be avoided – the 'I Robot fallacy', the 'agency fallacy' and the 'entity fallacy'. According to him, AI, AI systems and AI platforms must be seen as tools for humans beings and must not be assumed to have separate legal/natural personality. Additionally, he advocates for going back to first principles in order to regulate AI, whether it be in contract, tort or copyright law.

When discussing the moral status of machines, Nick Bostrom and Eliezer Yudkowsky[102] point to some exotic properties that AI might possess which humans do not, raising the question as to whether they must be given moral status:

- Non-sentient sapience – AI could be sapient, in that it might possess behavioural traits similar to those of humans, but not be sentient, in that it would not have conscious experiences.
- Principle of Subjective Rate of Time – To explain this, they consider a hypothetical scenario involving the conversion of a human brain into its digital form. They posit the subjective duration of an experience may differ depending on whether the conscious brain is in the human or in digital form.

The authors draw parallels to the animal rights and the right to choice movements, identifying 'sapience' (higher intelligence) and 'sentience' (ability to feel pain) as a popular basis for granting an entity rights.[103] However, this classification creates problems in cases such as those of infants and the mentally challenged. Bostrom & Yudkowsky propose the idea of 'Substrate Non-Discrimination' – *If two beings have the same functionality and the same conscious experience, and differ only in the substrate of their implementation, then they have the same moral status*[104]– and argue that this simplifies ethical considerations and their assessments.

Steve Torrance limits his discussion to the moral status that can be assigned to artificial humanoids that do not possess instantiate phenomenal consciousness (sentience). He notes that the Organic View of Ethical Status holds morality to be the sole domain of organic human persons, since they possess natural, biological and personhood characteristics.[105]

---

101 Legal Aspects of Artificial Intelligence - Kemp IT Law. (n.d.). Retrieved November 26, 2017, **http://www.kempitlaw.com/legal-aspects-of-artificial-intelligence-2/**

102 Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence*, 316-334.

103 Ibid., 7.

104 Ibid., 8.

105 Ibid.

However, non-sentient humanoids may be said to have some moral rights if these rights are closely linked to legal rights – for example, if a property-owning robot is the victim of an investment scam.[106]

Freier et. al.[107], undertook an empirical, rather than normative, analysis of this question. They used AIBO discussion forums to discern human interactions with AI. As regards moral status, they discovered that since the owners of AIBO were aware that it was a technical artefact, they would ignore its 'feelings' whenever convenient or desirable, clearly demonstrating that human do not yet accord any sort of moral status to AI.

Matthias Scheutz[108] is quite critical of authors arguing for emotions in AI. He notes that attempts to create emotions in agents import high-level emotional processes from humans to machines.[109] However, this does not necessarily mean that the processes are captured at lower levels, leading to the exhibition of, *in a very crude way–a relationship between higher level states and some lower level states or processes*. However, the shortcomings of this model may make all the difference between genuine human emotions and counterfeit ones. If the nature of the emotional labels are not more precisely defined, the author warns that there will be no criterion to distinguish alleged emotional agents from real ones.

Mark Coeckelbergh[110] analyzes the issue of robot rights from a philosophical perspective. He relies on the concepts of deontology, utilitarianism and virtue ethics to grant robots moral status, but dismisses them since they rely on ontological, non-relational features of the robot. He offers an alternative, social-relational argument for moral consideration.111 He details some features of his alternative model:

- Moral consideration is extrinsic to an entity, not intrinsic to it;
- Features of the entity as morally significant, but they are apparent, as seen by humans
- The experience is context-dependent instead of being context-independent.

In short, the model believes that *moral significance resides neither in the object nor in the subject, but in the relation between the two*.[112] He argues that emphasis must be placed on the relations forged between humans and robots and that that must act as the basis to our moral considerations.

b. Right to Personhood

A logical extension of the question of whether AI possesses intrinsic morality is whether this can be the basis for more concretized rights such as complete personhood.

Marshal Willick[113] notes that the reluctance to grant AI personhood comes from treating computers as 'the other', which hinders their evaluation on the same terms as humans.[114]

---

106 Ibid., 15

107 Kahn, P. H., Freier, N. G., Friedman, B., Severson, R. L., & Feldman, E. N. (2004, September). Social and moral relationships with robotic others?. *In Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on* (pp. 545-550). IEEE.

108 Scheutz, M. (2002, May). Agents with or without Emotions?. In *FLAIRS Conference* (pp. 89-93).

109 Ibid., 4.

110 Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology, 12*(3), 209-221.

111 Ibid., 214

112 Ibid.

113 Willick, M. S. (1983). Artificial intelligence: Some legal approaches and implications. *AI Magazine*, 4(2), 5.

114 Ibid., 13.

In his opinion, as computers increasingly begin to behave like humans, it will become more and more reasonable for law to treat them as persons unto themselves. Attributing property rather than personhood characteristics, in his opinion, weakens moral foundations of society.[115]

David Levy[116], in comparing robots to children, supports the view that conscious robots ought to have rights. He points out that until an entity is accorded rights, we continue to think of it as being for our own 'use'.[117]

Chopra and White[118] note that personhood will be granted to AI only if[119]:

- There exists internal doctrinal pressure; or
- In the cost-benefit analysis, legal convenience is in favour of granting such personhood.

Such legal convenience is furthered by the potential practical benefits to granting personhood, as the European Parliament[120] discusses as a part of its proposal for the establishment of a Charter on Robotics. The recognition of legal personality also brings with it the potential attribution of legal liability. In other words, a robot is only worth suing for compensation if it is covered by insurance; the Parliament recommends obligatory insurance for intelligent robots. Personhood is also beneficial as regards contract law – if robots can act for themselves under contract, they will also be able to be personally held liable. Finally, granting personhood could enable robots to pay taxes on any earnings, which could secure social welfare systems.

Other authors propose the grant of personhood status on the fulfillment of certain pre-conditions. Lawrence Solum[121] is in favour of granting AI personhood if it behaves the 'right' way and if it is confirmed that the processes that produce such behaviour are similar to those of humans.[122] He even calls for a redefinition personhood, stating that our existing notions of it are inadequate even without accounting for AI (for example, foetuses and people in vegetative states are still considered persons but do not seem to fit the current theory of personhood).

Chopra and White[123] believe that legal personality is an important stepping stone toward being accorded constitutional rights. The factors to be weighed when the legal system debates personhood include:

- *practical capacity to perform cognitive tasks; and*
- *the ability to control money: being able to receive, hold and pay money and other property such as securities, and to remain financially solvent;*

---

115 Ibid., 14.

116 Levy, D. (2009). The ethical treatment of artificially conscious robots. *International Journal of Social Robotics, 1*(3), 209-216.

117 Ibid., 212-213.

118 Chopra, S., & White, L. (2004, August). Artificial agents-personhood in law and philosophy. In *Proceedings of the 16th European Conference on Artificial Intelligence* (pp. 635-639). IOS Press.

119 Ibid., 4.

120 Do robots have rights? The European Parliament addresses artificial intelligence and robotics. (n.d.). Retrieved November 26, 2017, from **http://www.cms-lawnow.com/ealerts/2017/04/do-robots-have-rights-the-european-parliament-addresses-artificial-intelligence-and-robotics**

121 Solum, L. B. (1991). Legal personhood for artificial intelligences. *NCL Rev.*, 70, 1231.

122 Ibid., 57.

123 Supra, 118.

Patrick Hubbard[124] analyzes the liberal theory of personhood and argues for a legal right to personhood if an intelligent artefact:

- Has the ability to interact with its environment;
- Can engage in complex thought and communication;
- Is capable of possessing a sense of self; and
- Can live in a community based on mutual self-interest.

Then, there is the 'partial personhood' theory. Chopra and White[125] put forth the argument that AI may be conferred personhood for some legal purposes and not for others. The rights as regards each type of personhood would then accrue to the AI.[126] Rejecting the notion that a precondition for personhood is the need to be human[127], they divide legal personality into two – dependent and independent. Granting dependent legal personality to AI (like that currently granted to children) would be far easier[128] than independent legal personality, which would require the AI to reach a much higher level of technological sophistication[129]. Marshal S. Willick[130] is in agreement with this proposition, and believes that legal rights for AI can borrow from the current regime of 'partial personality' of corporations.

3.2 Legal Impact

There is wide agreement that the law will struggle to keep pace with the rapid changes in AI. This part of the paper considers the legal implications of AI in the areas of legal liability, privacy , cybersecurity and intellectual property (IP). It will analyze the lens through which various authors have looked at these issues and attempt to provide some solutions as well.

3.2.1 Liability – Civil and Criminal

Andreas Matthias[131] points out that liability with regard to machines is normally contingent upon control- whichever entity exercises control over the machine accepts responsibility for its failures.[132] He says that a "responsibility gap" arises when traditional modes of attribution cannot be transposed to a new class of machine actions, since nobody has enough 'control' to assume responsibility.[133] Matthias details the shift from technology over which the coder exercised control, to the types of technology where the control function gradually erodes. At the same time, the influence of the environment in which the technology operates increases. The extent and type of control dissipation differs with the technology employed[134]:

124 Hubbard, F. P. (2010). Do Androids Dream: Personhood and Intelligent Artifacts. *Temp. L. Rev.*, 83, 405.

125 Chopra, S., & White, L. F. (2011). *A legal theory for autonomous artificial agents*. University of Michigan Press.

126 Ibid., 156.

127 Ibid., 172.

128 Ibid., 160.

129 Ibid., 162.

130 Willick, M. S. (1985, August). Constitutional Law and Artificial Intelligence: The Potential Legal Recognition of Computers as" Persons". In *IJCAI* (pp. 1271-1273).

131 Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology*, 6(3), 175-183.

132 Ibid., 3.

133 Ibid., 7.

134 Ibid., 13-14.

- Logic-oriented programming and symbolic expert systems lead to the developers losing control over the execution-flow of the program;
- Artificial neural-networks (an AI technique) result in complete loss of control with respect to symbolic representation of information and flow.
- Reinforcement learning presents all the same problems as neural networks, in addition to creating another one by blurring the distinction between training and production.
- Genetic programming methods create an additional layer of machine-generated code that comes in between the programmer and his product.
- Finally, autonomous agents create a further spatial gap, quite literally, by engaging in acts outside the observable perimeter of their creator.

He calls for addressing this responsibility gap in both moral practice and legislation.[135]

This element of control is further reduced in the case of reinforcement learning – a training method that allows AI models to learn from their own past experiences – as pointed out by Elman and Castilla[136]. This method of learning was used by the new AlphaGo Zero that beat its earlier version AlphaGo at the board game Go, by learning on its own and with absolutely no human help.[137] When AI embedded with reinforcement learning capabilities is used in more real-world learning applications such as traffic-signals or drones, the traditional liability regime will be ineffective since there would be no human 'fault' at play.

a. Existing Liability Paradigms

A study commissioned by the European Parliament's Legal Affairs Committee[138] (**"EP Study"**) dismisses the idea that a robot may be held – partially or entirely – civilly liable on the grounds that this would require one to assume that it has a legal personality, which is fraught with dangers.[139] While accepting that a strict liability regime is possible, the Study notes that identifying the respondent would be a difficult task. At the same time, it also notes that compensation or liability cannot be reduced on the pretext that a robot (and not a human being) was directly behind the damage.[140]

Other authors like Matthew U. Scherer[141] assess the suitability of existing liability regimes to AI by examining its procedural merits. Substantive and procedural rules in the tort law system lead to focussing attention on the harm that has arisen in a particular case. Because of this, any debate on potential social and economic harms is limited. This leads to courts focussing more on the harms of emerging technology as opposed to its benefits, making adjudication one-sided and stunting technological growth.[142]

---

135 Ibid., 16.

136 Elman, J., & Castilla, A. (2017, January 28). Artificial intelligence and the law. Retrieved November 26, 2017, from **https://techcrunch.com/2017/01/28/artificial-intelligence-and-the-law/**; THOMAS E JONES V W M AUTOMATION INC (Wayne Circuit Court March 8, 2002).

137 DeepMind's New AI Taught Itself to Be the World's Greatest Go Player. (2017, November 08). Retrieved November 26, 2017, from **https://singularityhub.com/2017/10/23/deepminds-new-ai-taught-itself-to-be-the-worlds-greatest-go-player/**

138 Nevejans, N. (2016). European civil law rules in robotics, Directorate General for Internal Policies. *Policy Department C: Citizens' Rights and constitutional Affairs, Study PE, 571.*

139 Ibid.,16.

140 Ibid.,18.

141 Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies.

142 Ibid., 388.

Product Liability

Vladeck[143] notes that on the assumption that liability is the result of human (design or manufacturing) error, the rules of liability applied should be the same as those used when humans create any other machine. So long as the failure of autonomous systems can be linked or imputed to human actions, the existing product liability regime – utilizing tests such as the "consumer expectations" test and the risk-utility test – are sufficient.[144]

George S. Cole[145] discusses four policy principles when considering whether product liability ought to be applied to AI and Expert Systems (ES):

- Stream of Commerce principle – This principle assumes that since the developer voluntarily undertakes to make available the AI/ES product for economic gain, it is acceptable for her to be held liable. The author opines that this principle does not support the imposition of product liability as it restricts innovation by distorting true market costs. Moreover, court interference in what should be decided by private negotiations leads to unnecessary transaction costs.[146]
- Control of risks/environment – This principle justifies supplier liability on the ground that she is in a better position to anticipate and control the risks of harm. According to the author, the application of product liability on this basis would justify a principle of limited liability – if the developer is able to state the limitations of the AI, whether in terms of the range of knowledge or applications, her liability should be able to be constrained.[147]
- Risk cost spreading/Preventive effort allocation – The economic principle is based on the premise that the party better able to absorb and spread the cost of injuries ought to bear it. Basing product liability on this principle would create a market incentive to maintain and improve product quality.[148]
- Deep Pocket/Budding Industry – The inherent motivation in suing for product liability is economic gain. On the other hand, the endeavour to restrict the scope of such suits is to ensure that development in the industry is not stifled. In light of these inherently contrasting priorities, the author notes that the focus of such suits must not be on the harm – if the underlying issue is one of human capacity, both the plaintiff and defendant must share responsibility. On the other hand, if the problem is outside the scope of human capacity, a balance must be arrived at between the humanitarian advancement and human error.[149]

In his conclusion, Cole states that the resolution of the case based on product liability is contingent on the facts and circumstances of the case at hand. What it does, however, is to encourage both the plaintiff and the defendant to act in a more conscientious manner. [150]

---

143 Vladeck, D. C. (2014). Machines without principles: liability rules and artificial intelligence. Wash. *L. Rev., 89*, 117.

144 Ibid., 141.

145 Cole, G. S. (1990). Tort Liability for Artificial Intelligence and Expert Systems, 10 Computer LJ 127 (1990). *The John Marshall Journal of Information Technology & Privacy Law, 10*(2), 1.

146 Ibid., 175.

147 Ibid., 178-179.

148 Ibid., 181.

149 Ibid., 182.

150 Ibid.

Curtis Karnow[151] identifies two concepts of causation that apply to the liability question as popularly understood – causation in fact and proximate cause.[152] However, these concepts cannot be applied to more complex processing environments, which his article restricts itself to.[153] For the purposes of the article, Karnow creates a hypothetical intelligent programming environment called 'Alef', which handles air traffic control. He provides a description of its features, with the objective of emphasizing its networked distribution of agents, their unpredictable variety and complexity, and the polymorphic ambiance of the intelligent environment as a whole.[154]

He points to two drawbacks of fixing liability on intelligent agents themselves[155] seeing as the data and programs that make up a network are scattered, it is not possible to identify all the various points of fault. Moreover, it would be near-impossible for courts to identify what the proximate cause for the failure is, when the various causes (due to the interconnected nature of the network) cannot be sorted out amongst themselves. This leads to a breakdown of the classic cause and effect analysis of liability.

As regards warranties, Gerstner[156] opines that warranties such as warranties of merchantability and fitness of use for a particular purpose are relevant from a liability points of view, especially in the case of expert systems where the buyer relies on the expertise of the seller.[157]

Negligence

Gerstner[158] examines negligence and notes that courts have been reluctant to use it as a standard of liability in cases where software products have malfunctioned.[159] Applying this standard to the software industry is problematic as the duty of care owed to consumers is unclear and tracing the source of the defect to prove causation if difficult.[160]

Peter M. Asaro[161] seems to agree with this point of view (although he equates negligence and product liability). Cole leaves some questions which courts must confront:

- Courts must delimit the nature of duty of care owed in the case of AI/ES applications.[162]
- Given that customers might expect the AI/ES system to be more of an expert than them, how should negligence arising from imbalanced expectations be dealt with?[163]

151 Karnow, C. E. (1996). Liability for distributed artificial intelligences. *Berkeley Technology Law Journal*, 147-204.

152 Ibid., 176-177.

153 Ibid., 182.

154 Ibid., 183.

155 Ibid., 191-192.

156 Gerstner, M. E. (1993). Liability issues with artificial intelligence software. *Santa Clara L. Rev., 33*, 239.

157 Ibid., 253.

158 Ibid.

159 Ibid., 247.

160 Ibid., 258.

161 Asaro, P. M. (2007). Robots and responsibility from a legal perspective. *Proceedings of the IEEE*, 20-24.

162 Ibid., 215.

163 Ibid., 221.

For Asaro[164], the applicability of product liability arises when robots are treated as commercial objects.[165] In order to demonstrate negligence, it is necessary to demonstrate failure to take proper care or avoid foreseeable risks, a task which Asaro acknowledges is less than simple.[166] Moreover, the imposition of liability could slow down the uptake of this technology.

Service Liability

According to George S. Cole[167], service liability is relevant in cases where the distinction between product, service and sale is blurred. He considers the following policy parameters when examining the application of this form of liability to intelligence systems.

- Mass effect v. special application – Only applications that are marketed and sold to a large number of customers and are identical in all cases are considered to be subject to service liability.[168]
- Nature of the service – If the nature of the service is inherently uncertain, the imperfections of the underlying field (like legal or medical services) permeate into the decision-making ability of the AI or ES. However, if the domain is a well-defined one and it can be shown that the AI is operating in such a domain, the author sees a case for the application of service liability.[169]
- Nature of interpretation between representation and real-life events – Courts, when examining the application of service liability, must consider human factors associated with the AI/ES systems. This includes situations wherein the programmer/coder give the AI imperfect reasoning capabilities and does not inform the customer that this is being done, and situations where the cause of action may arise due to human inaction/incorrect action on the basis of the AI system's output.[170]
- Nature of demand for the service – If service liability is imposed in situations where the service is compulsorily consumed by all, it would prevent improvements in the quality of such services.[171]

Apart from policy reasons, Cole also examines practical considerations that would affect the imposition of service liability. These include:

- The class of defect must be correctable or preventable; and
- Intervention of humans, which breaks the chain of causation.

Malpractice Liability

George S. Cole[172] examines the applicability of malpractice liability to AI and ES systems. The standard of liability imposed is higher than that of a reasonable man – it is that of a professional.[173] According to Cole, bringing AI and ES under current malpractice law is difficult. This is due to the fact that there is no legislative standard applicable to a programmer

164 Ibid.

165 Ibid., 1.

166 Ibid., 2.

167 Supra, note 145.

168 Ibid., 193.

169 Ibid., 195.

170 Ibid., 196.

171 Ibid., 199.

172 Ibid.

173 Ibid., 207.

or computer scientist, and indeed that the field is too young and too fast-moving to have such codified standards. Finally, it is difficult to identify the exact profession within which malpractice must be evaluated, since there are many professions involved in the creation of the AI/ES system.[174] Despite the current limitations, however, Cole sees a future in which malpractice liability can be applied to this field.[175]

b. Alternative Frameworks

Strict Liability Regime

Vladeck[176] discusses liability in the context of self-driving vehicles. He advocates for a true strict liability regime, which would be de-linked from notions of fault. The regime would not be based on the premise that the autonomous machines are "ultra-hazardous" or "unreasonably risky", but on the basis that they are so technologically advanced that they are expected not to fail.[177] There are policy reasons as to why such a regime ought to be implemented[178]:

- There is value in providing redress to those who have been injured for no fault of their own;
- The creators/manufacturers are in a better position to absorb and distribute the costs among themselves;
- It will avoid high transaction costs by resolving issues that would otherwise be litigated in courts; and
- It might spur innovation.

Assuming that this strict liability regime is adopted, the next question to be discussed is who bears the cost. The author proposes two solutions[179]:

- Common enterprise liability – each entity within a set of interrelated companies may be held jointly and severally liable for the actions of other entities that are part of the group. This would allow the injured party to obtain redressal without assigning blame to others.
- Vehicle Liability – instead of suing the manufacturer, the autonomous vehicle itself can be sued. Autonomous machines can be reconceptualized as "people" under the law, thereby becoming principles from agents. Solutions such as self-insurance and the creation of an insurance pool would then follow.

Maruerite E. Gerstner[180] also supports strict liability as a potential solution. She believes that the application of strict liability to expert system software is relevant as[181]:

- Software can be brought within the purview of 'product' to which strict liability applies;
- Its application serves public policy considerations since the burden is placed on the party most able to bear it, viz., the manufacturer and/or the vendor.

---

174 Ibid., 210.

175 Ibid., 211.

176 Supra, note 143.

177 Ibid.,146.

178 Ibid.,146- 147.

179 Ibid.,149- 150.

180 Supra, note 156.

181 Ibid., 251.

Gerstner's final liability model requires an examination of both the function of the software program and the method of selling it. If the function is potentially hazardous, strict liability must be applied. In the event that the function is non-hazardous, the method of marketing the software determines the liability standard. Mass-marketed software would attract strict liability whereas custom programs should warrant a negligence standard.[182]

Other Solutions

Ryan Calo[183] proposes a two step process:

- A narrow and selective manufacturer immunity in situations where liability arises out of user changes;
- Users of AI technology must be encouraged to obtain insurance depending on factors such the use to which the AI would be put.

Karnow[184] proposes what he calls the Turing Registry. Noting that the behaviour of intelligent agents is stochastic, he proposes that the risks posed by the use of an intelligent agent is insured against, much like an insurance agency underwriting risk.[185] Under this model, developers would seek risk coverage for the potential risks posed by the agent, with the thumb- rule being the higher the intelligence, the higher the risk and hence the higher the premium. Karnow envisages a network-effects like scenario - users of the agent would begin to trust only those agents that are registered in the Turing Registry, which would in turn lead to more such registrations, and so on.[186] If an untoward consequence occurs due to an intelligent agent on the Registry, compensation is paid irrespective of fault or causal harm.[187] The major drawback of the Registry is its limited scope and the inability to pinpoint agent-reliability or the source of damage caused by the agent.[188]

Matthew U. Scherer[189] proposes a legislative solution, which would create an AI-safety agency, similar to today's FDA. This agency would possess the power to set out rights and liabilities. He advocates for a model wherein Agency- certified AI programs would attract limited tort liability, while uncertified ones would incur joint and several liability.[190]

Elman and Castilla[191] point out that non-human behaviour – such as that of plants or bees – is normally not held liable for their actions. They suggest that definitional and quality standards be adopted – either through treaty or international regulation – which manufacturers and developers would have to adhere to. For them, the benefits of creating these standards outweigh potential harms such as stifling innovation in the field.

c. Criminal Liability

Gabriel Hallevy[192] proposes three models of criminal liability – The Perpetration-via-Another

182 Ibid., 266.

183 Calo, R. (2010). Open robotics.

184 Supra, note 151.

185 Ibid., 193.

186 Ibid., 194.

187 Ibid., 195.

188 Ibid., 197- 198.

189 Supra, note 141.

190 Ibid., 393.

191 Supra, note 136.

192 Hallevy, G. (2010). The Criminal Liability of Artificial Intelligence Entities-From Science Fiction to Legal Social Control. Akron Intell. *Prop. J., 4,* 171.

Liability Model; The Natural-Probable-Consequence Liability Model and The Direct Liability Model. He advocates that while they may be applied separately, a combination of them would come in better use.[193]

Perpetration-via-Another Liability Model

Hallevy compares AI to a child, stating that it does not possess a criminal state of mind. Such innocent agents are deemed to be mere instruments/agents used in committing the crime, and liable as a perpetrator-via-another.[194]

The next question then becomes, who is the "other" who is primarily liable? There are two possibilities- the AI developer and the AI user. The former might design a program that would commit offences, and the latter might use it to do so, both indicating criminal intent or *mens rea*.[195] The AI itself would not be held liable for its actions under this model. However, an obvious limitation of this model is its limited scope – it cannot apply to situations where the AI entity commits offences of its own accord and without being programmed/used to do so.[196]

Natural-Probable-Consequence Liability Model

Hallevy assumes that the programmers and users of AI are involved in its activities, but do not intend, plan or participate in the commission of an offence via the AI. The basis of liability is then based on the ability of the programmers and users to foresee the commission of a potential offence.[197] For them to be held liable, they are required to know that the offence was a natural, probable consequence of their actions. The author borrows this from criminal liability that is imposed on accomplices to a crime.[198] The liability of the AI would be contingent on whether it acted as an innocent agent or not – liability would not accrue in the former case but would in the latter.[199]

Direct Liability Model

Under this, the AI entity is independent of its programmer or user. According to Hallevy, the concept of *actus reus* is quite simple to prove in the case of AI systems – as long as the act/omission is controlled by the AI itself, the element of *actus reus* can be said to be fulfilled.[200] Noting that specific intent is the most important *mens rea* requirement, the author sees no reason why AI systems cannot possess such intent to accomplish a task, for example, to commit murder.[201] Under this model, the criminal liability of an AI system is akin to that of a human being.

Hallevy sees no reason why a combination of the three models cannot be applied; he does not intend for them to be mutually exclusive.[202] In fact, he believes that the three models together create a strong liability net, which would be hard to evade.

As regards punishment for liability, Hallevy examines the theoretical foundations of punishment to humans and attempts to find a corresponding equivalent in the AI world.[203]

193 Ibid., 174.

194 Ibid., 179.

195 Ibid., 180.

196 Ibid., 181.

197 Ibid., 182.

198 Ibid., 183.

199 Ibid., 185.

200 Ibid., 187.

201 Ibid., 188-189.

202 Ibid., 193.

203 Ibid., 195.

He looks at the fundamental significance of different kinds of punishment for humans and proposes forms of punishment that may produce the same effect in AI. For example, he proposes deletion of the Al software controlling the Al entity instead of capital punishment, on the basis that both seek to incapacitate the doer of the crime[204], and he proposes putting the Al entity out of use for a determinate period instead of incarceration, since both seek to take away civil liberty in response to the commission of an offence[205].

Peter M. Asaro[206] is of the opinion that criminal liability cannot be applied to robots directly as criminal actions can only be performed by moral agents and deciding punishment for robots is no easy task. However, he proposes an alternative – criminal liability for robots can be applied akin to such liability for corporations, who are also non-human but separate legal persons. He leaves open the question of how punishment can be meted out to robots, though, since their motivations and reasons for existence are quite different from those of corporations (to make money).

3.2.2 Privacy Concerns

Most privacy concerns related to AI are those that stem from the use of big data – to that extent, the impact of big data on privacy can be said to be relevant for AI as well.[207] This section focuses on privacy concerns above and beyond these, which are likely to be caused by AI as a technology, in addition to providing suggestions as to their resolution.

a. Rethinking Privacy

Some authors like Erik P.M. Vermeulen[208] believe that our over-dependence on concepts like machine learning and AI mean that the old notions of privacy protection (purpose limitation, etc) would no longer be relevant in an AI-centric world. According to him, privacy no longer remains a well-defined concept and needs re-thinking.

Bohn et. al.[209] seem to agree, noting that the very nature of AI technologies have the potential to create surveillance frameworks, thereby invading our privacy.[210] The EP Study[211] points to the fact that robots will not only collect data from their surroundings, but may also exchange it between themselves or to another entity without the knowledge of humans.[212]

These invasions are ignored by users since the short term gains, in the form of increased productivity and efficiency, seem more definite than the vague future threats to privacy. This opinion is shared by the IEEE[213], the world's largest technical professional organization, which

204 Ibid., 196.

205 Ibid., 197.

206 Supra, note 161.

207 Fiander, S., & Blackwood, N. (2016). House of Commons Science and Technology Committee: Robotics and artificial intelligence: Fifth Report of Session 2016–17.

208 Vermeulen, E. P. (2017, April 27). What You Should Do in a World of Artificial Intelligence (and No Privacy). Retrieved November 26, 2017, from **https://medium.com/startup-grind/artificial-intelligence-is-taking-over-privacy-is-gone-d9eb131d6eca**

209 Bohn, J., Coroamă, V., Langheinrich, M., Mattern, F., & Rohs, M. (2005). Social, economic, and ethical implications of ambient intelligence and ubiquitous computing. In *Ambient intelligence* (pp. 5-29). Springer Berlin Heidelberg.

210 Ibid., 9.

211 Supra, note 138.

212 Ibid., 22.

213 Shahriari, K., & Shahriari, M. (2017, July). IEEE standard review—Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In *Humanitarian*

points to data asymmetry as the biggest threat to personal data protection. In other words, the organization collecting the data is more benefited by data collection than the user is by giving it up.[214]

From an analysis of the literature, privacy risks seem to vary depending on the type of AI technology used.

Bohn et. al.[215] identify some privacy problems caused due to ambient intelligence[216]:

- Ephemeral and transitory effects – Such technology can contain all information within itself forever, even the most minute details of the user;
- Spatial and temporal borders – Threats include accidental informational leaks and withholding of personal information.

In the context of Ambient Intelligence, Cook, Augusto and Jakkula[217] note that applications based on this technology have intentional and unintentional privacy and security risks.[218] They, along with multiple other authors[219] are of the opinion that privacy should be customizable according to the personal preferences of users and depending on the context.

Ackerman, Darrell & Weitzner[220] warn of the privacy risks of context-aware applications – those that adapt their behaviour according to a given physical environment. Regulating such applications is complex as[221]:

- One person's contextual awareness is another's lack of privacy; and
- The notions of user-participation in the control and dissemination of data are not straightforward – in many circumstances, users may not want to/may not be effectively allowed to do so.

b. Methods of Resolution

Existing Methods

According to some[222], existing laws like the General Data Protection Regulation (GDPR) of Europe are posing considerable hurdles to AI. Article 22(1) of the GDPR states that *the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her*. While there exist some exceptions to the above rule (including consent of the user), these must be resorted to subsequent to putting in place measures that

---

Technology Conference (IHTC), 2017 IEEE Canada International (pp. 197-201). IEEE.

214 Ibid., 56.

215 Supra, note 209.

216 Ibid., 11-12

217 Cook, D. J., Augusto, J. C., & Jakkula, V. R. (2009). Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and Mobile Computing*, 5(4), 277-298.

218 Ibid., 15

219 Monteleone, S. (2011). Ambient Intelligence and the Right to Privacy: The challenge of detection technologies.

220 Ackerman, M., Darrell, T., & Weitzner, D. J. (2001). Privacy in context. *Human–Computer Interaction*, 16(2-4), 167-176.

221 Ibid., 6

222 Artificial intelligence: what privacy issues with the GDPR? (2017, November 05). Retrieved November 26, 2017, from **http://www.gamingtechlaw.com/2016/10/privacy-gdpr-artificial-intelligence.html**

safeguard the rights, freedoms and legitimate interests of the data user.[223] Especially in cases of profiling, the law grants individuals the right to demand a justification for the decision made by the automated system.

Moreover, the GDPR requires automated decision systems to provide *meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.*[224] Finally, the Regulation requires data controllers and processors to undertake a data protection impact assessment, as long as there exists an evaluation of personal aspects based on automated processing (however limited) of data.[225]

Policy-led Solutions

The EP Study calls for a balance between the benefits of autonomous robots and the threat to personal privacy. Importantly, the Study places user consent at the top, pointing to the necessity of establishing protocols/rules which would be based on primacy of user-consent. The user must be allowed to control the points at and conditions under which third-party access to her private sphere is provided.[226]

The Science and Technology Committee of the UK House of Commons, in its report on Robotics and Artificial Intelligence[227], recommends the formation of a 'Council of Data Ethics' to address the challenges in the area of Big Data, which is now in the implementation phase due to the efforts of the UK government.[228] Such government-led solutions would make it easier to bridge the gap between technology and its regulation.

The IEEE[229] envisages a policy that[230]:

- Allows a minimum threshold of user control over her personal data, both in terms of the kind of data and how it is collected/shared;
- Simplifies the framework governing the usage of personal data;
- Encourages citizen training regarding data-management

Shara Monteleone[231] is in favour of differentiated data protection and privacy solutions depending on the type of Ambient Intelligence technology or application used, arguing that such micro-policies (sector-based rules) instead of domain policies would be more effective.[232] apart from such contextual regulation, privacy rights could also be considered as 'packages' that could be acquired, just like any other service.[233] She notes, however, that this second suggestion has been heavily criticized as it fails to ensure decisional autonomy.[234]

---

223 Articles 22(3) and 22(4), General Data Protection Regulation: Call for Views. (n.d.). Retrieved November 26, 2017, from **https://www.gov.uk/government/consultations/general-data-protection-regulation-call-for-views**

224 Ibid, Articles 13(2)(f), 14(2)(g) and 15(1)(h).

225 Supra, note 223, at Article 35.

226 Supra, note 138.

227 Supra, note 207.

228 Ibid., 20.

229 Supra, note 213.

230 Ibid., 57.

231 Supra, note 219.

232 Ibid., 32.

233 Ibid., 34.

234 Ibid.

Privacy and security training forming a part of AI curriculum must also find place in the solution-matrix, according to the White House report on AI ("**WH Report**")[235]. The WH Report remains critical of transparency as a potential solution. Implementing transparency will be easier said than done as:

- Private entities, afraid that it would compromise their competitiveness, would be reluctant to be more transparent, making it more difficult to monitor AI development; and
- The 'black box' nature of more sophisticated AI technology would be difficult for regulators to understand and deal with.

Techno-legal Solutions

Ackerman, Darrell & Weitzner[236] support the usage of context-aware tools to counter the very privacy risks they create. Examples include the use of context-aware computing in privacy agreement negotiation or in early-warning systems relating to the collection/use of user data.[237] They have also advocated for a security requirement in automated systems that would encrypt/secure data before it is stored or processed, with access being granted only to the user.[238] Alyssa Provazza[239] agrees, stating that AI and machine learning can themselves be used to counter privacy risks – For example, privacy assistant apps could be used to predict the type and nature of user privacy over a period of time.

Carnegie Mellon University[240] is working on one such personalized privacy assistant – the project envisions a system that would learn the privacy preferences of users and make privacy decisions on their behalf. The completed program would offer assistance in:

- Detailing the implications of different types of data processing;
- Alerting users to unusual privacy settings/practices; and
- Nudging users to consider and select certain kinds of privacy settings.

Wallace and Freuder[241] deal with privacy in the context of multi-agent systems, which assist in problem-solving. Using constraint-based reasoning, they explore, verify and propose a method in which such problem-solving agents operate under conditions of partial ignorance – while all information may not be available to them, they will have access to a range of possibilities from which to make decisions.[242]

Monteleone recognizes the potential of Privacy Enhancing Technologies (PET), wherein regulation is inbuilt in the design of the technology.[243] However, she points to three shortfalls

---

235 Privacy Perspectives | Why artificial intelligence may be the next big privacy trend Related reading: Uber concealed breach of 57M users for more than a year. (n.d.). Retrieved November 26, 2017, from **https://iapp.org/news/a/why-artificial-intelligence-may-be-the-next-big-privacy-trend/#**

236 Supra, note 220.

237 Ibid., 8.

238 Ibid.

239 Artificial intelligence data privacy issues on the rise. (n.d.). Retrieved November 26, 2017, from **http://searchmobilecomputing.techtarget.com/news/450419686/Artificial-intelligence-data-privacy-issues-on-the-rise**

240 Privacy Assistant Index, (n.d.). Retrieved November 26, 2017, from **https://www.privacyassistant.org/index/**

241 Wallace, R. J., & Freuder, E. C. (2005). Constraint-based reasoning and privacy/efficiency tradeoffs in multi-agent problem solving. *Artificial Intelligence*, 161(1-2), 209-227.

242 Ibid., 210.

243 Supra, note 219.

of the method which must be taken into account[244]:

- The difficulty in transforming legal rules into computer codes;
- The confusion on whether such technical rulemaking will supplement or replace traditional law;
- Automatic data mining and analysis could lead to profiling.

3.2.3 Cybersecurity Impact

This section focuses on two aspects concerning the relationship between AI and cybersecurity- the threats posed by AI to the field of cybersecurity, as well as using cybersecurity itself as a means to combat those threats.

a. Cybersecurity Risks

The threat to cybersecurity caused by AI will differ depending on the type and nature of AI application. Yampolskiy[245] classifies and surveys the potential kinds of dangerous AI. He suggests that cybersecurity threats from AI will be less the robot science-fiction kind, and more arising out of deliberate human action, side effects of poor design or factors attributable to the surrounding environment. He concludes that that most dangerous types of AI would be those that are created to harm.[246] He opines that deciding what constitutes malevolent AI would be an important problem in AI safety research, and advocates for the intentional creation of malevolent AI to be recognized as a crime.

Unlike other cybersecurity research and publication which focuses on the design of malicious machines (which would then prompt cybersecurity solutions from other scholars), research on AI seems to focus primarily on the creation of safe machines themselves. Pistono and Yampolskiy[247] argue that there is great value to the former kind of work, which would encourage collaboration between hackers and security experts, something that is currently not present in the AI sphere. They mention two points that make the creation and spread of malevolent AI easier[248]:

- Lack of oversight with regard to AI implementation in areas such as human cloning would make it easier to develop potentially dangerous Artificial General Intelligence in those areas.
- Existence of a closed source code is a key attribute of malevolent AI.

Risks caused by AI systems can range from the more simple phishing to the catastrophic risks of Artificial General Intelligence (AGI). By combining phishing with automated and intelligent technology, AI can and is being used to influence the next-generation of phishing methods, which would be more sophisticated than before.[249]

Yampolskiy[250] believes that a failure by existing cybersecurity systems, while unpleasant, is, for the most part, curable, since it would only cause monetary or privacy risks. However, the

244 Ibid., 35.

245 Yampolskiy, R. V. (2015). Taxonomy of Pathways to Dangerous AI. *arXiv preprint arXiv:1511.03246*.

246 Ibid., 4.

247 Pistono, F., & Yampolskiy, R. V. (2016). Unethical research: How to create a malevolent artificial intelligence. *arXiv preprint arXiv:1605.02817*.

248 Ibid., 4-5.

249 AI will supercharge spear-phishing. (n.d.). Retrieved November 26, 2017, from **https://www.darktrace.com/blog/ai-will-supercharge-spear-phishing/**

250 Yampolskiy, R. V. (2017, July 12). AI Is the Future of Cybersecurity, for Better and for Worse. Retrieved November 26, 2017, from **https://hbr.org/2017/05/ai-is-the-future-of-cybersecurity-for-better-and-for-worse**

failure of a superintelligent AI (SAI) or an AGI system has the potential to cause an existential risk event, causing large-scale damage to human well-being. While attempting to analyze the failures of existing AI systems, Yampolskiy and Spellchecker[251] paint a dismal picture by indicating that AI (especially AGI) will never be 100% safe – we can only hope to end up with a probabilistically safe system.[252] They note that while solutions such as AI Boxing or AI Safety Engineering can be developed, this would only delay, and not prevent, the problems from occurring.

With specific reference to Brain Computer Interfaces (BCIs), Yampolskiy[253] notes that the rise of BCIs – devices that are controlled by brain signals, as of now being used in the medical devices and gaming sectors – would present an attractive target for hackers. On gaining access to such systems, hackers would be able to tamper with not only critical personal information, but also more abstract albeit important important personality traits such as preferences and thoughts.

While examining privacy and security issues specifically in the context of BCIs, Bonaci et. al.[254] note that while these neural engineering systems can be, and indeed are, trained to follow ethical guidelines, there is no protection currently against third-party exploitation. In fact, researchers were even able to create the first 'brain spyware', a malicious software designed to detect private information through a BCI.[255] The authors argue that it is not hard to imagine applications that could extract much more private and intimate information such as memories, prejudices and beliefs.[256]

Governments can also cause greater threats to cybersecurity via AI by way of increased and more efficient surveillance. Jon Kofas[257] is skeptical of the use AI will be put to by governments; according to him, governments will use AI for all the wrong reasons. The pervasive dependence on robots by humans will allow intelligence agencies to use AI for surveillance, jeopardizing civil and political rights.

b. AI as a Cybersecurity Tool

The advantage of using AI over earlier cybersecurity methods, Morel[258] notes, is that earlier methods approached the problem in a manner akin to "fixing the plumbing". AI would play a bigger role in making the paradigm shift from this method to a more comprehensive cybersecurity framework. Moreover, AI technology can also help software improve its own cybersecurity abilities, thereby increasing effectiveness and supplementing a shortfall in trained personnel.[259]

---

251 Yampolskiy, R. V., & Spellchecker, M. S. (2016). Artificial Intelligence Safety and Cybersecurity: a Timeline of AI Failures. *arXiv preprint arXiv:1610.07997*.

252 Ibid., 8.

253 Supra, note 250.

254 Bonaci, T., Calo, R., & Chizeck, H. J. (2014, May). App stores for the brain: Privacy & security in Brain-Computer Interfaces. In *Ethics in Science, Technology and Engineering, 2014 IEEE International Symposium* on (pp. 1-7). IEEE.

255 Ibid., 3.

256 Ibid., 3.

257 Artificial Intelligence: Socioeconomic, Political And Ethical Dimensions. (2017, April 22). Retrieved November 26, 2017, from **http://www.countercurrents.org/2017/04/22/artificial-intelligence-socioeconomic-political-and-ethical-dimensions/**

258 Morel, B. (2011, October). Artificial intelligence and the future of cybersecurity. In *Proceedings of the 4th ACM workshop on Security and artificial intelligence* (pp. 93-98). ACM.

259 Hoffman, K. (2017, March 23). For black and white hats, AI is shaking up infosec. Retrieved November 26, 2017, from **https://www.the-parallax.com/2017/03/27/hackers-ai-artificial-intelligence-infosec/**

These are underscored by Dilek et. al.[260] as well, stating that AI techniques offer unique advantages in combating cybercrime[261], including the ability to learn by example, the resilience to noise and incomplete data, intuitiveness, adaptability to the environment and user preferences and the ability to collaborate. At the same time, however, they also suffer from limitations including the inability to create a model of what constitutes an attack, resulting in a number of false positives.[262] The authors examine AI applications used to combat cyber-crimes, including technology in the areas of Artificial Neural Network Applications[263], Intelligent Agent Applications[264], Artificial Immune System Applications[265] and Genetic Algorithm and Fuzzy Set Applications[266].

Landwehr[267] points to more specific benefits, pointing out that AI techniques could assist in the explanation of complex cybersecurity policies to users, in addition to detecting (lower-level) anomalies in the system which might otherwise not be noticeable. AI-led solutions could also be useful in countering more high-level attacks such as spoofing that exploit social engineering approaches.

The benefits of AI have been recognised by the United States National Science and Technology Council[268], which has published a report stating that AI could assist the government in *planning, coordinating, integrating, synchronizing, and directing activities to operate and defend U.S. government networks and systems effectively, provide assistance in support of secure operation of private-sector networks and systems, and enable action in accordance with all applicable laws, regulations and treaties.*[269]

The use of AI in cybersecurity ranges from the more humble CAPTCHA systems to Intrusion Detection. Zelkowitz270 details the history, concept and utility of the former, one of the more prominent applications of AI to cybersecurity. Morel[271], on the other hand, examines the role of AI in Intrusion Detection, which is used to detect and alert networks to ongoing attacks or abuses.

Other specific sectors of cybersecurity in which AI is and would play a role are detailed by Golan[272]:

- Hacking in IoT devices – AI-based prediction models can be used to detect and block suspicious activity;

---

260 Dilek, S., Çakır, H., & Aydın, M. (2015). Applications of artificial intelligence techniques to combating cyber crimes: A review. *arXiv preprint arXiv:1502.03552*.

261 Ibid., 32-33.

262 Ibid., 33.

263 Ibid., 25.

264 Ibid., 26.

265 Ibid., 28.

266 Ibid., 29.

267 Landwehr, C. E. (2008). Cybersecurity and artificial intelligence: From fixing the plumbing to smart water. *ieee seCurity & PrivaCy*, 6(5), 3-4.

268 Supra, note 18.

269 Ibid., 36.

270 Zelkowitz, M. (Ed.). (2011). *Security on the Web* (Vol. 83). Academic Press.

271 Morel, B. (2011). Anomaly based intrusion detection and artificial intelligence. In *Intrusion Detection Systems.* InTech.

272 Golan, M. Y. (2016, November 23). How AI will transform cybersecurity. Retrieved November 26, 2017, from **https://venturebeat.com/2016/11/22/how-ai-will-transform-cybersecurity/**

- Preventing malware – AI is harnessed to examine the millions of features in suspicious-looking files and detect mutations in the code;
- Operating efficiency of cyber security departments – AI can increase efficiency by going through the many security alerts received by these departments, flagging the ones which might contain actual threats;
- Cyber-risk estimation for organizations;
- Network traffic detection – AI can be used to detect anomalies in network traffic, which might indicate malicious activity, by analyzing metadata created from traffic; and
- Malicious mobile application detection.

There is quite a bit of focus on BCI systems here as well, with authors differing on how they can be used to offer cybersecurity solutions. Bonaci et. al.[273] suggest an engineering solution in the BCI Anonymizer. The anonymizer would pre-process neural signals before they are stored or transmitted, acting like a filter before the information reaches the processor.[274]

Victoria Turk[275], on the other hand, states that policy rather than technology should be relied on to control data collection and usage by BCI systems. She argues that standards as regards data usage must be developed jointly by lawyers, ethicists and engineers. This could be supplemented by a system similar to an app-store certification, which would certify those apps that adhere to these standards, thereby incentivizing developers and programmers.

However, Bonaci et. al.[276] do not seem to be using technology to the exclusion of policy. They examine the vulnerable components of a BCI system, the types of attackers and the possible methods used to extricate personal information, and come up with a 'threat model'. They advocate a combined approach using both technology and policy. They advocate for the former through privacy and security by design. For the latter, they envisage a 'triangle' approach- the development solutions must be a three-tier collaborative effort between neuroscientists, neural engineers, ethicists, as well as legal, security and privacy experts, with systems manufacturers and application developers developing the tools meeting the criteria laid down by the first two.[277] The issues required to be addressed include amount and nature of access to individuals' neural signals, purpose for which these signals can be used and risks of misuse.

3.2.4 Intellectual Property Issues

Schafer[278] believes that AI impacts IP laws in two ways:

- AI is being used to design creative works, either along with humans or entirely on their own. Whether the traditional notions of 'creator', 'inventiveness' and 'original' will be relevant with regard to AI is yet to be seen.
- AI's dependence on others' creative works. Being primarily data-driven, AI will require great amounts of input which can all be subject to different IP regimes, potentially hindering economic access.

---

273 Supra, note 254.

274 Ibid., 6.

275 Turk, V. (2016, August 03). How Hackers Could Get Inside Your Head With 'Brain Malware'. Retrieved November 26, 2017, from **https://motherboard.vice.com/en_us/article/ezp54e/how-hackers-could-get-inside-your-head-with-brain-malware**

276 Supra, note 254.

277 Ibid., 5.

278 Schafer, B. (2016). The Future of IP Law in an Age of Artificial Intelligence.

Apart from authors of creative works, Schafer[279] also notes AI's impact on IP as regards the legal profession. Lawyers will be forced to provide value in the IP sector, either instead of or in conjunction with AI.

This section examines whether IP rights can be said to exist in AI- driven work in the first place, and if so, the attributability of the same. A brief overview is also provided of some copyright applications as well as AI's contribution to intellectual property management.

a. Existence of IP rights

Can IP rights be said to exist at all, when they arise from entities that are not human? Tuomas Sorjamaa[280] argues that if copyright's primary role is to incentivize the production and dissemination of creative works, it would not be advisable to leave AI-produced work out of its realm.[281] If the premise is correct, copyright law must then be able to develop to respond to technological challenges such as this one.

On analyzing existing case-law and scholarship on the matter, Annemarie Bridy[282] indicates that copyright protection is presently granted – however, she restricts her analysis to psychographic works and procedurally generated artworks.[283]

Erica Fraser[284] describes AI techniques such as genetic programming, artificial neural networks and robot scientists that are used to generate inventions. She notes that patents have earlier been granted for inventions using AI and that the method of creation of the invention does not seem (so far) to factor into the patent granting process.[285] However, she sees a need to redefine inventiveness and patentability in light of the increased role played by computer programs in the inventive process.

To identify the existence of an inventive step in patent, it becomes important to identify the notion of "person of ordinary skill in the art". Since AI will effectively raise the skill level of ordinary inventors, this notion must be rethought in light of the contemporary inventor and the technology she typically might use.[286] Similarly, the (vast) knowledge that AI technologies possess must be taken into account when assessing obviousness, failing which there will be a flood of patent filings and grants. Both Shamnad Basheer[287] and Ryan Abbot[288] seem to agree with the idea that this test would need rethinking in light of AI. Samuelson[289] resorts to this only in the event that[290]:

---

279 Ibid.

280 Sorjamaa, T. (2016). *I, Author–Authorship and Copyright in the Age of Artificial Intelligence* (Available on Internet) (Master's thesis, Svenska handelshögskolan).

281 Ibid., 57.

282 Bridy, A. (2012). Coding creativity: copyright and the artificially intelligent author. *Stan. Tech. L. Rev.*, 1.

283 Ibid., 20.

284 Fraser, E. (2016). Computers as Inventors-Legal and Policy Implications of Artificial Intelligence on Patent Law. *SCRIPTed*, 13, 305.

285 Ibid., 319.

286 Ibid., 320.

287 Basheer, S. (2016). Artificial Invention: Mind the Machine. *SCRIPTed*, 13, 334.

288 Abbott, R. (2016). I Think, Therefore I Invent: Creative Computers and the Future of Patent Law. *BCL Rev.*, 57, 1079.

289 Samuelson, P. (1985). Allocating ownership rights in computer-generated works. *U. pitt. L. rev.*, 47, 1185.

290 Ibid., 1224.

- The ownership dilemma cannot be resolved satisfactorily through the application of traditional authorship tests; and
- Joint authorship as a concept proves to be unworkable.

In the context of completely autonomous AI systems, Fraser is of the opinion that patentability should not be denied.[291] She calls for the evolution of the law toward wider patentability, except in situation where there is a sound policy reason not to. Examining whether AI-inventions fit within the incentive justification of the patent system, she notes in the affirmative, stating that there are economic and social benefits to innovation that will arise as a result of patenting AI-led innovations.[292]

Vertinsky and Rice[293] call for an increase in the 'utility' threshold in a world where there are AI- led inventions.[294] This would ensure that 'useful' ideas are granted patents, but the mere generation of 'new' ideas – which will become easier to do with machines – will not.

Lasse Øverlier[295] obtains industry perspective on this issue through his thesis. While the predominant view with regard to patents is that it is not possible under current laws, respondents seemed more positive as regards copyright protection in the era of machine learning technologies.[296]

b. Attribution/Ownership of Rights

The United Kingdom (UK) is probably the only nation whose copyright legislation deals with computer-generated work. Section 9(3) of the Copyright, Designs and Patents Act (CDPA) states:

"*In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken*."

Section 178 defines a computer-generated work as one that "*is generated by computer in circumstances such that there is no human author of the work*", making the law quite clear in this regard.

However, Andres Guadamuz[297] notes that despite a seemingly clear wording of the law, there is ambiguity as to the actual author. Drawing an analogy to Microsoft Word (programmed by Microsoft, but the company does not have copyright over the works creates using it), Guadamuz states that there could potentially be authorship attribution to either the programmer or the user, and under the law, it is unclear which.[298]

Guadamuz analyses the laws of jurisdictions such as the EU, the US and Australia, concluding that there are wide gaps in the interpretation of originality for copyright protection, more so in the case of computer-generated works.[299] He details two areas in which failing to provide copyright protection would lead to negative commercial implications – Computer code and

291 Ibid., 324.

292 Ibid. 328.

293 Vertinsky, L., & Rice, T. M. (2002). Thinking About Thinking Machines: Implications of Machine Inventors for Patent Law. *BUJ Sci. & Tech. L.*, 8, 574.

294 Ibid., 35.

295 Øverlier, L. (2017). *Intellectual Property and Machine Learning: An exploratory study* (Master's thesis).

296 Ibid., 42-46.

297 Guadamuz, A. (2017). Do androids dream of electric copyright? Comparative analysis of originality in artificial intelligence generated works.

298 Ibid., 8.

299 Ibid., 18.

databases.[300] In his final analysis, he recommends that the model adopted in the UK, despite its limitations, be followed more widely around the world.

Lasse Øverlier[301] point out that appropriability – the term used to denote companies securing the future value of their inventions – is of two types – primary and generative. Generative appropriability will become increasingly important for companies using Machine Learning Systems (MLS), as everything created by such a system can, in turn, be used to create new IP. But in order for companies to maximize this potential, the rights to the creations of the MLS have to vest in them.[302]

Presenting his views in the form of a trial dialogue, Shamnad Basheer[303] grapples with the issue of who possesses rightful IP ownership to an invention – the person who coded/created the software for the AI system that then generates the invention, or the AI system itself. In his final analysis, Basheer (through the judge in the trial) finds that, under current law, the patent cannot rest with either, since machines cannot yet be considered inventors or authors. The IP rights fall to the public domain or the commons, free to be used by all. Mark Perry and Thomas Margoni[304] seem to agree with this view, arguing that it is a much more efficient allocation of resources compared to its alternatives.[305]

Erica Fraser examines the issue of both inventorship as well as ownership. As regards inventorship, she considers three possibilities – granting inventorship to the AI algorithm designer, to the computer itself or doing away with the requirement for it – and notes that the current climate is in favour of identifying human inventors where they can be reasonably identified.[306] As regards ownership, she considers two possibilities – granting ownership rights to the computer vis-a-vis the first owner of the computer – and debates their merits.[307]

Ryan Abbot[308] analyses US patent and copyright law, including the history of the Copyright Office's Human Authorship Requirement and case law interpreting the Patent and Copyright Clause. He concludes that on the basis of the above analysis and with the assistance of dynamic statutory interpretation, computers must qualify as legal inventors. He dismisses the counter arguments, being that inventors be individuals and that there must exist a subjective mental process by which the invention is made.

Abbott also states that the default assignee for the invention must be the owner of the computer responsible for it, as this would most incentivize innovation. In the event that the owner, developer and user are distinct parties, alternative ownership arrangements could be reached through contract.[309]

Pamela Samuelson[310] is of the opinion that the user of a computer-generated program ought to be considered the author, unless *the work generated by a computer incorporates a substantial block of recognizable expression from the copyrighted program*, in which case

300 Ibid., 18.

301 Supra, note 295.

302 Ibid., 17.

303 Supra, note 287.

304 Perry, M., & Margoni, T. (2010). From music tracks to Google maps: Who owns computer-generated works?. *Computer Law & Security Review*, 26(6), 621-629.

305 Ibid.

306 Ibid., 331.

307 Ibid., 331.

308 Supra,note 288.

309 Ibid., 1114.

310 Supra, note 289.

it should be considered derivative work.[311] This is supported by both doctrinal and policy considerations – legally, the user is responsible for 'fixing' the work and hence responsible for bringing it out into the world[312], and such an arrangement would also not fall afoul larger policy goals[313].

Some authors, while not offering specific solutions, offer some assistance in terms of how to look at the problem of attributability. Tuomas Sorjamaa[314] argues that copyright as a legal concept and as a cultural concept have developed simultaneously. Acknowledging that there is no ready answer, he calls for a semiotic study into the concept of authorship, which will then make copyright ownership more clear.[315] Annemarie Bridy[316] suggest two lenses through which vesting of copyright can be considered – the derivative work doctrine and the work for hire doctrine.[317] In her opinion, while neither doctrine is perfectly placed to solve this problem, the work for hire doctrine is more easily modifiable without requiring an expansion in the notion of copyrightable subject matter, while at the same time avoiding the difficult discussion of whether such rights can be vested in machines. Other authors such as Kalin Hristov[318] also endorse the work for hire approach, calling for a reinvention of the terms employee and employer.

Others, such as James Grimmelmann[319] are dismissive of the very existence of computer authorship. He notes that the underlying problems of assigning authorship to computer-generated are more apparent than real – they are no different from human-generated works. In any case, by the time future authorship is attributed to future computer programmes, law would have progressed enough such that they will already have been assigned personality/personhood rights, and copyright law will then simply fall in line.[320]

An interesting issue to be considered is liability for IP infringement by AI. Eran Kahana[321] argues that the default strict liability standard is misguided, and proposes an iterative liability standard. Under this form of enquiry, if it is show that the AI behaved independent of its human deployer/developer, the individual must not be held liable.

c. Specific Copyright Applications

Christian Geib[322] examines the intersection of copyright and data mining, which is *an automatic or semi-automatic way of manipulating large quantities of data to discover patterns and rules*. The author notes that the EU lags behind the US and much of Asia when it comes to data mining and he attributes this to strong copyright protection laws in the former, which limits the progress of data mining. The relationship between copyright and database

311 Ibid., 1192.

312 Ibid., 1202.

313 Ibid., 1203.

314 Supra, note 280.

315 Ibid., 60.

316 Supra, note 282.

317 Ibid., 25.

318 Hristov, K. (2016). Artificial Intelligence and the Copyright Dilemma.

319 Grimmelmann, J. (2015). There's No Such Thing as a Computer-Authored Work–And It's a Good Thing, Too.

320 Ibid., 414.

321 Kahana, E. Intellectual Property Infringement by Artificial Intelligence Applications.

322 *From infringement to exception: why the rules on data mining in Europe need to change*. (2016). CREATe. Retrieved 5 December 2017, from **http://www.create.ac.uk/blog/2016/06/28/from-infringement-to-exception-why-the-rules-on-data-mining-in-europe-need-to-change/**

law and data mining is studied in three impact sectors are considered – pharmaceuticals, law enforcement and marketing. It is concluded that data mining is *prima facie* copyright infringing, with the exceptions and defences being insufficient, at best and inapplicable, at worst.

Geib considers potential solutions, evaluating their applicability and risks:

- Copyright owners can grant licenses to researchers to mine data;
- A US-style fair use exception to copyright can be introduced, or in the alternative, a hybrid (fair dealing-fair use) approach with US law forming a part of it.

Noting that a *closed, mandatory, data mining-specific exception rather than a technology-neutral, future-proof solution seems to be the most likely outcome of the latest round of EU copyright reform*, he urges European policymakers to err on the side of openness and legal certainty.

d. IP Management

Stading[323] is of the opinion that AI could revolutionize the way IP data is managed and analyzed. At the moment, administration of various IP rights, while extremely important for companies, is a mismanaged and inefficient process, not to mention cost and time-consuming. A switch to AI-enabled technology from a manual one would increase the efficiency and accuracy in processing and analyzing large datasets. AI would be useful not only in automating the database search process, but also provide insights into an IP market, which would help rights-holders strategically plan their filings.

Vertinsky and Rice[324] also point to practical problems that will arise in terms of the increase in number and complexity of patent applications and licensing strategies.[325] They advocate for the implementation of smart technologies into the patent examination process itself, thereby arming the examiners with the same technological capabilities as the potential patentee.[326]

Lasse Øverlier[327] examines the impediments that IP rights pose for a company within the field of machine learning. By way of a literature survey as well as personal interviews, the author concludes that ambiguity regarding the use and rights to input data used in machine learning technologies were essential in controlling the freedom-to-operate of companies.[328] According to Overlier, this is an IP management problem, especially in cases where the owner of the input also owns the right to any output from the MLS. If there are multiple owners of the input data, it acts as a restriction on later MLSs, which use billions of inputs which may all possess corresponding IP rights.[329] Even if input data is used without permission by a MLS, the copyright owner will have a hard time proving the same due to lack of understanding of what and how the data was used by the MLS.[330]

---

323 *The Role of Artificial Intelligence in Intellectual Property - IPWatchdog.com | Patents & Patent Law. (2017). IPWatchdog.com | Patents & Patent Law.* Retrieved 5 December 2017, from **http://www.ipwatchdog.com/2017/07/27/role-artificial-intelligence-intellectual-property/id=86085/**

324 Supra, note 293.

325 Ibid., 22.

326 Ibid., 34.

327 Supra, note 295.

328 Ibid., 33.

329 Ibid., 53.

330 Ibid., 54.

3.3 Economic Impact

AI affects the economy both at a micro (jobs) and macro (economic development) level. From an analysis of the literature, authors seem to have mixed opinions as to how much of an impact AI will have and whether it be, on balance, negative or positive.

3.3.1 Employment/Labour

Frey & Osborne[331] note that historically, computerization was restricted to tasks involving *explicitly rule-based activities*; now, however, they are being used in domains that involve *non-routine cognitive tasks*.[332] Using economic and statistical analysis techniques, they predict that developments in machine learning and computerization will reduce demand for those forms of labour that rely on pattern recognition, but will increase demand for tasks that cannot be computerized.[333] They divide occupations into low, medium and high-risk depending on the likelihood of likelihood of computerization, and conclude that 47% of total US employment falls under the high-risk bracket.[334]

Analyzing data, Brynjolfsson & McAfee[335] discover a paradox they call the "great decoupling" – despite the growth in productivity due to technology, growth in jobs & median income becomes weaker and inequality increases. They attribute this to the fact that the advancement in technology too fast for human skills to keep up with. Interestingly, they find that while the demand for high-skill and low-skill jobs are growing, the mid-level jobs are losing out due to automation and AI. in what they term to be the race against the machine, progress will not be equal – *some will win while many others will lose*.

However, many authors are confident that the economic impact would not be as bad as is feared. They point out that instead of replacing humans, AI will lead to a change in how 'work' is looked at. People will no longer work for money but for pleasure, which will, in turn, result in greater societal and market contributions.[336] A study conducted by Pricewaterhouse Coopers ("**PwC Study**") supports this,[337] emphasizing on the fact that increase in productivity and automation increases real wage, which allows for greater efficiency.[338]

Pavaloiu & Kose[339] argue that managers and executives would be able to prioritize the larger and more important issues, leaving the simpler and more tedious ones for AI. This would encourage collaboration in the short term and inclusive growth in the longer term.[340] IBM CEO Ginni Rometty refers to this as the creation of – not blue collar or white collar - but "new collar" jobs, where humans will have the time to do what they do best and leave the rest to AI-enabled machines.[341] The PwC Study also points out that despite a potential redundancy in

---

331 Frey, C. B., & Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation?. *Technological Forecasting and Social Change*, 114, 254-280.

332 Ibid.

333 Ibid.

334 Ibid.

335 Rotman, D. (2013). How technology is destroying jobs. *Technology Review*, 16(4), 28-35.

336 Ibid., 21.

337 The *impact of artificial intelligence on the UK economy*. (2017). PwC. Retrieved 5 December 2017, from **https://www.pwc.co.uk/economic-services/assets/ai-uk-report-v2.pdf**

338 Ibid., 11.

339 Supra., note 67.

340 Ibid 21.

341 Balakrishnan, A., & Jr., B. (2017). IBM CEO: *Jobs of the future won't be blue or white collar, they'll be 'new collar'*. CNBC. Retrieved 5 December 2017, from **https://www.cnbc.com/2017/01/17/ibm-ceo-says-ai-will-be-a-partnership-between-man-and-machine.html**

existing types of employment, new types will be created – *along with jobs in the development and application of AI, the technologies will need to be built, maintained, operated and regulated*.[342] The increase in demand due to the adoption of AI will also indirectly create jobs in other sectors.

Kolbjørnsrud, Amico and Thomas[343] term this synergy between man and machine 'organizational intelligence'. In their opinion, organization based on principles of collaboration between humans and AI (intelligent enterprises) improve decision-making than if they are solely human or machine driven.[344]

Purdy & Daugherty[345] point to three avenues where AI can drive growth, ultimately resulting in a broader structural transformation[346]:

- Intelligent automation – AI has the ability to automate complex physical tasks, solve problems across different sectors and self-learn.
- Labour and capital augmentation – AI will not so much replace labour as enable its effective usage by complementing existing workforce and improving capital efficiency. As a result of innovation allowing more efficient use of workman hours, AI can increase labour productivity by upto 40% in 2035.[347]
- Driver of innovation – AI has the ability to generate innovations as it diffuses through different sectors of the economy.

The Mckinsey Global Institute[348] has undertaken a detailed study ("**McKinsey Study**") on the impact of automation on employment and economic productivity. The Study notes that the threat posed by AI as regards automation is different from earlier technologies – while earlier technology could only perform routine physical tasks, AI can also perform cognitive tasks that were considered difficult to automate. As per the study, *about half of all the activities people are paid to do in the world's workforce could potentially be automated by adapting currently demonstrated technologies*.

However, the net impact will be positive – as the study notes, "*At a microeconomic level, businesses everywhere will have an opportunity to capture benefits and achieve competitive advantage from automation technologies, not just from labor cost reductions, but also from performance benefits such as increased throughput, higher quality, and decreased downtime*. Acknowledging that existing opinion is skewed toward the negative effects of AI on jobs, the Study states that the nature of work will undergo a transformation – comparing the present scenario with that of the shift from agriculture to industry, the Study points out that humans will perform tasks complementary to machine-labour. In order for this to happen, however, policy-makers must evolve regulation in areas such as education, income support and safety nets that allows workers to take advantage of the new economic shift.

According to Darrell West[349], the real challenge is in understanding how to navigate the age of abundance that technology brings. He calls for a reinterpretation of the social contract

342 Supra, note 337.

343 Kolbjørnsrud, V., Amico, R., & Thomas, R. J. The promise of artificial intelligence.

344 Ibid., 22.

345 Purdy, M., & Daugherty, P. (2016). Why Artificial Intelligence is the future of growth. Remarks at *AI Now: The Social and Economic Implications of Artificial Intelligence Technologies in the Near Term*, 1-72.

346 Ibid., 12.

347 Ibid., 17.

348 Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). A future that works: Automation, employment, and productivity. *McKinsey Global Institute, New York, NY.*

349 West, D. M. (2015). What happens if robots take the jobs? The impact of emerging technologies on employment and public policy. *Centre for Technology Innovation at Brookings*, Washington DC.

in light of the shift in employment and leisure time. Concepts such as income generation, employment and public policy need to be re-thought in this new light. The resulting inequality creates both economic (job loss) and social (social disruptions and unrest) consequences. Some solutions West proposes include creating multiple avenues for learning new skills, including those in arts and culture, supplementing incomes and benefits and the encouragement of volunteering opportunities.

3.3.2 Wealth Gap and Economic Disparity

Despite the potential increase in productivity and employment, the benefits of AI on the economy may be skewed towards a few. A report by the Executive Office of the President[350] points out that identifying the exact nature and type of jobs that would be affected by AI is difficult, since AI comprises a collection of technologies. More importantly, the threatened sectors are those that comprise low-paid, low-skilled workers, and AI-led progress will decrease their demand and increase inequality. The benefits of AI may accrue to a select few, leading to concentration of power, reduced competition and increased inequality.[351]

The report notes that the extent of disparateness in impact depends largely on how policy measures handle AI's impact on the labour market. Some possible solutions suggested include[352]:

- The development of pro-competition policies;
- Altering education and training to fit jobs of the future;
- Empowering workers by modernizing the social safety net, ensuring wage insurance and critical social safeguards like health and retirement.

Other authors[353] provide similar solutions, calling for a greater investment in diverse types of more useful education and training. Another feasible solution in Universal Basic Income – which provides income to workers who have lost out due to automation and generates consumption, which keeps the economy going. However, the problems with this solution are that it may be unaffordable for governments, destroy incentives or need to be pegged too low to be effective.[354]

Purdy & Daugherty[355] caution of the risk of economic disparity and ask regulators to take these concerns seriously. The response, according to them, must be twofold[356]:

- Policy-makers must highlight the tangible benefits of AI – both at a micro level (assist workers) and a macro level (can alleviate climate change) – in order to ensure a more positive framework for its development.
- In order to control the negative externalities of AI, policy-makers must identify and address those groups that would be disproportionately affected by its uptake.

---

350 House, W. (2016). Artificial Intelligence, Automation, and the Economy. *Executive office of the President.* **https://obamawhitehouse.archives.gov/sites/whitehouse. gov/files/documents/Artificial-Intelligence-Automation-Economy. PDF.**

351 Ibid., 2.

352 Ibid., 3-4.

353 PricewaterhouseCoopers, U. K. (2009). *UK Economic Outlook March 2017*. URL: **https://www.pwc.co.uk/economic-services/ukeo/pwcukeo-section-4-automation-march-2017-v2.pdf**

354 Ibid., 45.

355 Supra, note 345.

356 Ibid., 23.

3.3.3 Economic Progress

Apart from job-creation and job-supplementing, there are also macro-positives that AI will bring to the economy. Robert D. Atkinson[357] argues that those who claim that AI undermines the labour market consider only the first order effects whereby the machine replaces the worker. There exist second order effects as well which have gone unnoticed – there is increased productivity, leading to increased savings which is ploughed back into the economy in the form of lower prices, higher wages for the remaining workers, or higher profits. He relies on an OECD analysis between productivity and employment, noting that:

"*Historically, the income-generating effects of new technologies have proved more powerful than the labor-displacing effects: technological progress has been accompanied not only by higher output and productivity, but also by higher overall employment.*"

Analyzing the impact of AI on 12 developed economies, Purdy & Daugherty[358] conclude that AI has the potential to double annual economic growth in them.[359]

Chen et. al.[360] focus on the broad economic impact of AI. They primarily utilize two approaches for their analysis[361]:

- The bottom-up approach – using the premise that investment in technology is an indicator of its future potential, the authors examine private sector and venture capital investments in AI. Accounting for overlap, they estimate that the total economic impact of investments by these sectors would imply $359.6 billion to $773.2 billion in economic growth over the next ten years.
- The top-down approach – the authors examine the impacts of prior technologies as benchmarks. Relying on the impacts of technologies such as IT investment, broadband internet, mobile phones and industrial robotics, the authors conclude the economic impact of AI to be between $1.49 trillion and $2.95 trillion.

The Mckinsey Study points out that automation can help in closing the GDP gap. The declining birth rates will lead to an increase in the average age and consequent decrease in the working capacity of the population. This will create an economic growth gap as labour, the factor of growth, evaporates. Automation, through AI, can compensate for some of this. The macroeconomic factor of an ageing demographic would require that all remaining humans and robots engage in productive labour to ensure sustained economic growth.[362] The Study calls for policy-makers to encourage *investment and market incentives to encourage continued progress and innovation*.

3.3.4 Greater Consumer Choice

According to a study conducted by PwC[363], the bulk of the impact of AI on the GDP of the United Kingdom – slated to increase by 10% – will arise due to *consumption-side product enhancements*. They detail four ways in which this will happen – better quality products, wider consumer choice, saving consumer time and lowering of prices.

---

357 Atkinson, R. D. (2013). Stop Saying Robots Are Destroying Jobs–They Aren't.

358 Supra, note 345.

359 Ibid., 15-16.

360 Chen, N., Christensen, L., Gallagher, K., Mate, R., & Rafert, G. (2016). Global Economic Impacts Associated with Artificial Intelligence. *Study, Analysis Group, Boston, MA, February*, 25.

361 Ibid., 4.

362 Ibid., 15.

363 Supra, note 337.

3.4 Impact Global Geopolitics and Democracy

AI has the potential to alter relations not only at an interpersonal level, but also at an international one. This section discusses the implications specifically for global inequality and democratic elections.

3.4.1 Global Inequality

David Gosset[364] notes that AI will create a digital divide in the world – the benefits of AI will reach only a small section of the global population, while a large majority still waits to gain access to the internet as tool. He warns of a duopoly in AI development, with the West and China taking the lead. He recommends the creation of a United Nations International Artificial Intelligence Agency, involving participation from academia, industry civil society and government, which will have the following objectives:

- The discussion of the implications of AI for humanity as a whole;
- Prevent the enlargement of the digital divide;
- Ensure transparency in AI research, both with regard to the government and the private sector;
- Encourage knowledge sharing and international cooperation to prevent concentration of AI-harnessed power in the hands of a few.

Kyle Evanoff[365] terms this an alteration of power dynamics on the international stage. He points out that the power harnessed by AI will be concentrated in a few countries, non-state actors and corporations. This raises the potential for international conflict, which can only be addressed by a multilateral, multi-stakeholder approach diverse ethical, cultural and spiritual values.

3.4.2 Public Opinion and Elections

Vyacheslav W Polonski[366] examines the use of AI in elections campaigns. Noting that machine learning already uses algorithms to predict which US congressional bills will be passed, the author points out that it is now being used to keep voters informed and engaged regarding political issues.

Polonski recounts the allegations that AI-powered technologies were used in the election campaign of Donald Trump and in the Brexit vote, wherein they were used to manipulate individual voters by delivering personalized advertising. Such targeting was effective, since voters' opinions were influenced depending on their susceptibility to different arguments, eventually leading to an unexpected Trump victory.

In addition to targeted advertising, Polonski notes that AI has also been used in the form of bots to manipulate public opinion, notably in the 2017 UK general election, the 2017 French presidential election and the 2016 US presidential election. By spreading biased political messages and manufacturing the illusion of public support, these bots are able to shape public opinion and hence discourse.

---

364 *Artificial Intelligence (AI) And Global Geopolitics.* (2016). *HuffPost.* Retrieved 5 December 2017, from **https://www.huffingtonpost.com/david-gosset/artificial-intelligence-a_2_b_10710612.html**

365 *A Sputnik Moment for Artificial Intelligence Geopolitics*. (2017). *Council on Foreign Relations*. Retrieved 5 December 2017, from **https://www.cfr.org/blog/sputnik-moment-artificial-intelligence-geopolitics**

366 *How artificial intelligence conquered democracy*. (2017). *The Conversation*. Retrieved 5 December 2017, from **https://theconversation.com/how-artificial-intelligence-conquered-democracy-77675**

However, Polinski concludes that the same technology that is used to distort public opinion and interfere with free and fair elections can equally be used to support democracy. Political bots can be used to check misinformation and warn citizens, as much as targeted advertising can be used to educate users on real issues.

Data Analytics companies such as Avantgarde[367] do exactly this – they utilize machine learning techniques to assist social movements and political campaigns. They claim to also using AI to prevent computational propaganda that distorts political sentiment.

4. Proposed Solutions for the Regulation of AI

Having examined the impact of AI on diverse sectors, it is pertinent to look at regulatory solutions. While the previous sections have provided some suggestions for resolution of sector- specific issues, this section aims to analyze the methods and tools that can be used to in policy- making and regulation as regards artificial intelligence unto itself. This part begins with questioning the need for regulation in the space, moving on to examining substantive bases of regulation as well as methods used for the same.

4.1 Should we regulate?

Comparing the AI of today to the early days of the internet, Jeremy Straub[368] points out that the lack of regulation in the latter case is what allowed the internet to develop to its full potential. Similarly, with the potential to change nearly everything humans do, AI should be regulation-free to encourage as much innovation as possible.

Amitai Etzioni & Oren Etzioni[369] are also not in favour of generic AI regulation. According to them:

- AI does not possess a motivation of its own, unlike humans. The chances of intelligence being turned to motivation that leads creates trouble is relevant only for the purposes of science fiction.
- The regulation of AI at this stage will be challenging, since AI is already being used and developed by many government and private entities around the world.
- Regulation might lead to restrictions, which are likely to impose high human and economic costs

A report by Stanford's "One Hundred Year Study on Artificial Intelligence"[370] acknowledges that regulation in AI will be an inevitability due to its ability to effect profound change. The report warns that knee-jerk regulation would be detrimental to innovation and counterproductive in the long run.[371] The regulators recommend regulation in AI to borrow from aspects of privacy regulation, which ultimately creates *a virtuous cycle of activity involving internal and external accountability, transparency, and professionalization, rather than narrow compliance*.

---

367 *Avantgarde Analytics | Data, Networks, Behaviour.* (2017). *Avantgarde Analytics | Data, Networks, Behaviour*. Retrieved 5 December 2017, from **http://www.avntgrd.com/**

368 *Does regulating artificial intelligence save humanity or just stifle innovation?*. (2017). *The Conversation.* Retrieved 5 December 2017, from **https://theconversation.com/does-regulating-artificial-intelligence-save-humanity-or-just-stifle-innovation-85718**

369 Etzioni, A., & Etzioni, O. (2017). Should Artificial Intelligence Be Regulated?. *ISSUES IN SCIENCE AND TECHNOLOGY,* 33(4), 32-36.

370 Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., ... & Leyton-Brown, K. (2016). Artificial Intelligence and Life in 2030." One Hundred Year Study on Artificial Intelligence, Report of the 2015-2016 study panel.

371 Ibid., 10.

However, some authors find the current regulatory and policy environment wanting as regards AI regulation. Nicolas Petit[372] examines the need for regulation of AI in a context of technological emergence. He details the trade-offs that will need to be made if emerging technology such as AI were to be regulated:

- Regulation can be consciously and unconsciously disabling for the development of the technology;
- The lack of knowledge and inherent suspicions regarding AI could lead to knee-jerk regulation;
- Powerful private interest groups will have the ability to steer regulation to their own benefit;
- The protracted time taken to enact and enforce reactive regulation will lead to its obsolescence by the time it is adopted.

According to Guihot, Matthew and Suzor[373], the problems with regulating AI include:

- The continuing development of the technology ensures that it stays one step ahead of regulation;
- Information asymmetry between regulators and private corporations in understanding AI and its applications;
- Coordination between regulatory bodies is a necessity;
- Regulatory failure could occur due to agency capture;
- There are limited enforcement mechanisms.

4.2 Basis of Regulation

Because AI is composed of many different sub- parts, its regulation will not be straightforward. Substantive law and policy may regulate AI based on a number of different factors, which are detailed below.

4.2.1 Application-Based Regulation

One way to regulate AI is to regulate its applications – each industry or sector in which AI may/does have an application can be regulated individually.

This is an approach that the United States has followed[374]– AI is regulated by a number of government agencies:

- AI that is responsible for medical diagnostics and treatment is regulated by the Food and Drug Administration (FDA).
- Drones are regulated by the Federal Aviation Administration (FAA).
- Consumer-facing systems are regulated by the Federal Trade Commission (FTC).
- Financial market applications are subject to the Securities Exchange Commission (SEC).

The IEEE-USA[375] points out that the distribution of oversight responsibilities requires that

372 Petit, N. (2017). Law and Regulation of Artificial Intelligence and Robots-Conceptual Framework and Normative Implications.

373 Guihot, M., Matthew, A. F., & Suzor, N. P. (2017). Nudging robots: Innovative solutions to regulate artificial intelligence. *Vanderbilt Journal of Entertainment & Technology Law*.

374 Supra, note 370.

375 IEEE-USA POSITION STATEMENT ARTIFICIAL INTELLIGENCE RESEARCH, DEVELOPMENT & REGULATION

policies be consistent across the board. A first step would be for academia, industry and government to collaborate in examining the status quo as regards governance of AI.

The report of the Committee on Technology of the National Science and Technology Council[376], while broadly supporting sector specific AI regulation, provides a few specific guidelines[377]:

- Those AI-products that are tasked with the protection of public safety must be examined from the prism of a risk-analysis – the risks that the AI-products will increase against those that it will decrease.
- The existing regulatory regimes must be analyzed to check whether they can adequately address AI-risk.
- The cost of compliance of any future policy must also be thoroughly examined.

The European Parliament, in its resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics[378] also follows a sectoral regulatory approach, with separate recommendations for distinct industries such as autonomous travel, care robots and medical robots, among others.

AI may also be subject to the 'critical infrastructure' rule, which is composed of "*the assets, systems, and networks, whether physical or virtual, so vital to the United States that their incapacitation or destruction would have a debilitating effect on security, national economic security, national public health or safety, or any combination thereof.*" AI-technologies that have their application in critical sectors may be subject to higher regulatory thresholds.[379]

One such critical sector is national defence. Amitai Etzioni & Oren Etzioni[380] point to the fact that while generic AI regulation is not called for, applicational-regulation is warranted. In fact, over 20,000 AI researchers, public intellectuals and activists signed an open letter calling for a ban on weaponized AI that operates beyond meaningful human control.[381] Others such as the UN Special Rapporteur on extrajudicial, summary, or arbitrary executions, have also weighed in, calling for a moratorium on armed robots.[382]

4.2.2 Principle/Rule-Based Regulation

Principle-based regulation is another possibility. Fairness and transparency seem to be two often repeated principles that would find place in AI regulation. Although Andrew Fogg[383] advocates for an application-based regulatory model, he makes a reference to these

(2017). *Globalpolicy.ieee.org*. Retrieved 5 December 2017. Retrieved on 5 December 2017, from **https://ieeeusa.org/wp-content/uploads/2017/07/FINALformattedIEEEUSAAIPS.pdf**

376 Supra, note 18.

377 Ibid., 1.

378 Delvaux, M. (2016). Draft report with recommendations to the Commission on Civil Law rules on robotics. *European Parliament Committee on Legal Affairs* **http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0005+0+DOC+XML+V0//E**N

379 Supra, note 370.

380 Supra, note 369.

381 *Open Letter on Autonomous Weapons -Future of Life Institute.* (2017). *Future of Life Institute.* Retrieved 5 December 2017, from **https://futureoflife.org/open-letter-autonomous-weapons/**

382 Cumming-Bruce, N. (2014). *U.N. Expert Calls for Halt on Military Robots. Nytimes.com.* Retrieved 5 December 2017, from **http://www.nytimes.com/2013/05/31/world/europe/united-nations-armed-robots.html**

383 Fogg, A. (2016). *Artificial Intelligence Regulation: Let's not regulate mathematics! - Import.io. Import.*

principles. In the context of fairness, he notes that it is essential that technology contain the widest and most diverse the data-set possible to avoid any subsequent bias. Fogg sees transparency as another essential principle, given the black-box nature of AI. Fogg also refers to what not to regulate – the inner working of Deep Learning, or, as he calls it, the mathematics of AI.

Through the enactment of the GDPR, the European Union also takes into account these two principles. As mentioned above, the GDPR[384]:

- Restricts complete automated decision making without any human intervention; and
- Grants users a 'right to explanation' for algorithmic decisions involving them.

Jia Kai & Tao Tong[385] stress on two principles:

- Data regulation – The authors point to this as a precondition for AI regulation. Data fed into machines forms the basis of all AI activity; there are grave risks of the data is biased or incomplete. Regulation of data sharing and application would better guide development in AI.
- Machine optimization rules – Transparency and open-source must be included as part of AI regulation, to ensure that the methodology of the conversion of input into output is better understood and can be checked.

Some authors have proposed extremely specific principled regulatory rules.[386] Oren Etzioni proposes three rules of AI that are inspired by Isaac Asimov's "three laws of robotics", which are:

- *An AI system must be subject to the full gamut of laws that apply to its human operator.* This would prevent illegal behaviour being excused on the ground that it was committed by the AI system.
- *AI systems must clearly disclose that it is not human* to prevent a potential fraud.
- *AI systems cannot retain or disclose confidential information without explicit approval from the source of that information.* This is a necessary precaution since AI is able to absorb, analyze and act on information much faster than humans can.

4.2.3 Risk- Based Regulation

A third way is to regulate for the existing and potential risks of the technology. Guihot, Matthew and Suzor[387] point out that despite the existence of distinct identifiable classes of AI, regulating each class separately is not a feasible option.[388] AI poses a range of risks which does not necessarily correlate to a specified class. Moreover, one class of AI has the potential to become stronger and become an entirely different class with different risks. This is enhanced by the fact that AI also presents systemic risk, which is the embedded risk

io. Retrieved 5 December 2017, from **https://www.import.io/post/artificial-intelligence-regulation-lets-not-regulate-mathematics/**

384 Goodman, B., & Flaxman, S. (2016, June). EU regulations on algorithmic decision-making and a "right to explanation". In *ICML workshop on human interpretability in machine learning (WHI 2016), New York, NY.* **http://arxiv. org/abs/1606.08813 v1.**

385 *AI Regulation: understanding the real challenges - Paris Innovation Review.* (2014). Retrieved 5 December 2017, from **http://parisinnovationreview.com/articles-en/ai-regulation-understanding-the-real-challenges**

386 Etzioni, O. (2017). *How to Regulate Artificial Intelligence. Nytimes.com.* Retrieved 5 December 2017, from **https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html**

387 Supra, note 373.

388 Ibid., 26.

'to human health and the environment in a larger context of social, financial and economic risks and opportunities'[389], which requires the coordination of multiple stakeholders in the regulatory process.

The authors support risk-based regulation, especially in light of the fact that regulators possess limited resources.[390] They advocate for a staggered approach, with the most serious risks being tackled first. The risk-based framework would operate in the following manner, in consultation with the industry[391]:

- The regulator will assess the levels and types of risks;
- Risk assessment will lead to the likelihood of the occurrence of each (class of) risk;
- The regulated entities will be evaluated on the basis of risk and ranked accordingly; and
- Resources will be allocated according to the above evaluation.

4.3 Regulatory Tools

Apart from substantive regulatory methods, there are a variety of regulatory tools that can be employed.

4.3.1 Self-Regulation

Some methods of self-regulation have already been resorted to/attempted by the AI community. The Asilomar AI Principles are a byproduct of such self-regulation. The Principles are aspirational in nature, to ensure the beneficial development of AI. The principles cover issues ranging from AI safety, transparency and value alignment to longer-term issues such as potential risks and development for the common good.[392]

Similarly, the Association for Advancement of Artificial Intelligence appointed a panel in 2009 to examine *the value of formulating guidelines for guiding research and of creating policies that might constrain or bias the behaviors of autonomous and semi-autonomous systems so as to address concerns*.[393] However, the resultant deliberation decided against the need for concern or to halt research.

4.3.2 Legislative/Agency Regulation

Matthew U. Scherer[394] examines the competencies of three regulatory systems – national legislatures, administrative agencies and the common law tort system. He notes that in all three, the greater the financial resources, the greater the chances of 'working' the system and influencing policy.[395] He proposes a regulatory regime for AI to manage its public risks without stifling innovation.[396] Under his Artificial Intelligence Development Act (AIDA), an agency tasked with certifying AI-safety would be created. AIDA would create a joint but

389 Ibid., 27.

390 Ibid., 51.

391 Ibid., 55.

392 AI Principles - Future of Life Institute. (2017). Future of Life Institute. Retrieved 5 December 2017, from **https://futureoflife.org/ai-principles/**

393 Supra, note 369.

394 Supra, note 141.

395 Ibid., 377.

396 Ibid., 393.

limited liability system between the designers, manufacturers, and sellers of agency-certified AI programs, whereas uncertified ones would be subject to joint and several liability. Under his proposed Act, the respective institutional strengths of legislatures, agencies, and courts would be leveraged.

Arguing that the development of AI and robotics would require an understanding of the technology and the economic incentives of the stakeholders as much as the law, Ryan Calo[397] proposes the creation of a Federal Robotics Commission (FRC), which would advise as opposed to regulate. The FRC would consist of experts in engineering and computer science, along with those in law and policy. The tasks of the FRC would include[398]:

- Allocation of money for research;
- Advising federal agencies such as the DOT, the SEC and the FAA on specific AI applications they oversee;
- Advise lawmakers as regards policy;
- Engage in soft-diplomacy by bringing together stakeholders for discussions.

Building on Calo's work, Aaron Mannes[399] provides a list of potential federal and state agencies that could assist the US government with the development of AI policy, along with an analyses of their advantages and drawbacks.

4.3.3 Regulatory structures

Nicolas Petit[400] proposes a regulatory framework from the point of view of a public-interest minded social planner and examine its normative applications.[401] Beginning with the proposition that law and regulation seek to address externalities – both positive and negative, Petit creates a distinction between three types:

- Discrete externalities – harms or benefits that are personal, random, rare or endurable;
- Systemic externalities. third party harms or benefits that are local, predictable, frequent or unsustainable;
- Existernalities – the group of externalities that comprises existential threats and opportunities created by AI.

Petit's model has the following normative implications:

- Since their impact on society is not of a high magnitude, discrete externalities must be resolved by the legal infrastructure, viz., court the system. They must be solved *ex post* through the application of property, contract and liability and other such specific generic rules.
- Systemic externalities, which are more severe, must be handled with an *ex ante* approach by the social planner.
- Given that externalities create much larger and more abstract concerns, they must be tentatively regulated by International Organizations.

397 Calo, R. (2014). The Case for a Federal Robotics Commission.

398 Ibid., 12.

399 Mannes, A. (2016). Anticipating Autonomy: Institutions & Politics of Robot Governance.

400 Supra, note 372.

401 Ibid., 25

4.3.4 Cross-Industry Partnerships

Pavaloiu & Kose[402] suggest cross-industry partnerships as a method of regulation- engineers would design robots according to rules and ethics, and governments and other organizations would understand the nuances of the technology, which, in turn, will allow for more accurate and effective rule-setting.[403]

Amitai Etzioni & Oren Etzioni[404] recommend a tiered decision making system. The lower levels of the system will consist of the AI-workers, above whom will exist an oversight system that provides parameters within which the work ought to be done.

Conclusion

The term “Artificial Intelligence” include within its scope a wide range of technological processes, making it tricky to understand and hence create policy for. This literature synthesis attempts to provide a broad overview of the key technologies that compose the umbrella term referred to as AI and the key common factors/issues to its different disciplines. As is evident from this literature synthesis, the field of AI offers tremendous promises as solutions and optimisation for a variety of problem statements we face. However, equally importantly, AI also throws up key normative and practical questions of ethics and governance that will play a central role with increased adoption of these technologies. While the some of the tensions between efficiencies promised by AI, and the criticisms pointed to by those advocating greater caution in its adoption may appear irreconcilable, it is important to delve into these points of conflict, so that we are able to rethink some the existing legal and regulatory paradigms, and create new ones if required.

402 Supra., note 67, at 15-27.

403 Ibid., 22.

404 Supra, note 369.