# PERSONA EX MACHINA: THE LEGAL PERSONHOOD OF ARTIFICIALLY-INTELLIGENT MACHINES*

*Noemi M. Mejia***

**ABSTRACT**

As artificial intelligence becomes more developed and advanced, the issue of their legal personhood arises. Machines in recent history have become more person-like. The more independent and autonomous they are, the less their actions can be imputed to their owners and users. Because they are not considered "persons" in the eyes of the law, ordinary rules on liability cannot be made applicable to them. With the influx of technological advancements, it is not entirely impossible for AI to eventually be perfected for widespread use. As such, modern legal measures need to be put in place to protect society from possible dangers related to their use.

* *Cite as* Noemi M. Mejia, *Persona Ex Machina: The Legal Personhood of Artificially-Intelligent Machines*, 98 PHIL. L.J. 539, [page cited] (2025).

** J.D., Dean's Medal for Academic Excellence, University of the Philippines (2023). B.S. in Applied Mathematics, *magna cum laude,* University of Asia and the Pacific (2018). Concurrent with her legal studies, she worked as a machine learning engineer for an insurance company. Through the UP Law Center - Technology Law and Policy Program, she assisted in the drafting of the Implementing Rules and Regulations of Republic Act No. 11967 or the Internet Transactions Act.

Noemi is currently an associate at the Disini Law Office, a consultant for The Asia Foundation - Cybersecurity-Artificial Intelligence, and a fellow at Data and AI Ethics PH. She extends her sincere appreciation to Atty. JJ Disini for his invaluable advice and guidance as supervised legal research adviser.

## I. INTRODUCTION

"*And now I see with eye serene*
*The very pulse of the machine*;"
— William Wordsworth,
*She Was a Phantom of Delight*

From beating the best chess grandmaster in the world[1] to predicting the global spread of COVID-19 a year early,[2] artificial intelligence ("AI") has

[1] Jack Watson, *Artificial Intelligence*, 22 LAWNOW 36, 36 (1997).

[2] Paul W. Grimm, Maura R. Grossman & Gordon V. Cormack, *Artificial Intelligence as Evidence*, 19 N.W. J. TECH. & INTELL. PROP. 9, 32 (2021).

greatly contributed to many fields.[3] It has made many things in life easier, quicker, and more convenient. Although some may argue that AI can never replace human beings, industries that used to only employ manual work are now infused with AI in one way or another. In China, for example, a company called Squirrel developed an AI tutor to help students prepare for annual standardized tests.[4] Doctors now use AI technology not just to diagnose patients, but also to predict the effectivity of treatments.[5] Racter and Hal, both computer programs, have written books.[6] DABUS, itself a patented device, has invented and obtained a patent over a food container.[7] Out of 800 executives surveyed, almost half believe that an AI machine can sit on the board of directors by 2025.[8] AI has even penetrated the legal profession—a survey showed that 26% of attorneys use AI-based technology[9] for drafting documents, developing strategy, and predicting trial outcomes, among others. Among surveyed law firm leaders, 35% can envision AI associates, while 47% can imagine AI paralegals.[10]

Given these rapid advances in AI technology, it is not impossible for AI to eventually be perfected for widespread use. Should that happen, they will be thinking machines—completely autonomous, able to act and think with little to no human supervision. It will become more and more difficult to distinguish between human and machine. In response, legal measures need to be put in place to protect society from possible dangers related to AI. When the AI makes a mistake, such as a self-driving car getting into an

---

[3] *See* Victor M. Palace, *What If Artificial Intelligence Wrote This: Artificial Intelligence and Copyright Law*, 71 FLA. L. REV. 217 (2019). *See also* Alicia Solow-Niederman, *Administering Artificial Intelligence*, 93 S. CAL. L. REV. 633 (2020).

[4] Karen Hao, *China has started a grand experiment in AI education. It could reshape how the world learns*, MIT TECH. REV. WEBSITE (2019), *at* https://www.technologyreview.com/2019/08/02/131198/china-squirrel-has-started-a-grand-experiment-in-ai-education-it-could-reshape-how-the/.

[5] Lucie Rutherford, *Medicine Meets Big Data: Clinicians Look to AI for Disease Prediction and Prevention*, UVATODAY, Feb. 18, 2022, *at* https://news.virginia.edu/content/medicine-meets-big-data-clinicians-look-ai-disease-prediction-and-prevention.

[6] Palace, *supra* note 3, at 221.

[7] *DABUS Gets Its First Patent in South Africa Under Formalities Examination*, IPWATCHDOG, July 29, 2021, *at* https://ipwatchdog.com/2021/07/29/dabus-gets-first-patent-south-africa-formalities-examination/id=136116/.

[8] Palace, *supra* note 3, at 221.

[9] Adam N. Eckart, *Transactional Artificial Intelligence*, 26 LEG. WRITING 273, 275 (2022).

[10] Palace, *supra* note 3, at 221.

accident,[11] or an algorithm making medical misdiagnoses, resulting in further errors in treatment, who becomes responsible? Traditionally, liability would fall upon the owners, manufacturers, and developers of the AI technology employed. There may have been errors in the code or a defect in the logic.[12] But what if the AI is already capable of independent decision-making? To whom will the blame shift then?

Much of what has been written on the subject stops at the philosophical—the question of whether the law ought to recognize AI as legal persons, and the implications of such recognition on natural persons. Others delve immediately into the liability framework—if AI were indeed legal persons, how can they be held liable for their actions? Some focus on a specific area of law, such as intellectual property, tort, or criminal law. This research builds upon these existing literature, and explores the possibility of recognizing AI as legal persons and determining the extent of liability that may be allocated on them.

Part II of this research dives into the concept of AI and how it came about to be the form we know of today. While experts still cannot agree on a single definition for AI, it is defined based on its capabilities, as well as the advantages and disadvantages associated with its use. This also examines recent jurisprudence and current laws and regulations involving AI.

Part III focuses on legal personhood and whether or not such a status is possible to be accorded to AI. AI is compared to other entities that are considered legal persons, such as corporations. There is also an exploration of existing objections and counter-objections to such personhood. The consequences of recognizing AI as legal persons are also discussed.

In Part IV, traditional liability frameworks are discussed in relation to their applicability to AI. Possible liability models are also proposed in the hopes of filling in these gaps.

---

[11] Aaron J. Snoswell, Henry Fraser, & Rhyle Simcock, *When self-driving cars crash, who's responsible? Courts and insurers need to know what's inside the 'black box'*, THE CONVERSATION, May 24, 2022, *at* https://theconversation.com/when-self-driving-cars-crash-whos-responsible-courts-and-insurers-need-to-know-whats-inside-the-black-box-180334.

[12] *See* Megan Sword, *To Err Is Both Human and Non-Human*, 88 UMKC L. REV. 211 (2019).

## II. "THINKING" MACHINES

"Can machines think?"[13] Alan Turing, considered as the father of modern computer science, had asked. Suppose a person and a machine are placed in two separate rooms, one labeled "X" and one labeled "Y." There is an interrogator between them, who does not know which room contains the person and which one contains the machine. This interrogator communicates with them only through typewritten questions which, in turn, those in the rooms can respond to through typewritten answers. Will the interrogator be able to tell which is the person and which is the machine?

This is the essence of the Turing Test, also known as the Imitation Game. The Turing Test aims to determine whether a machine can be so powerful as to be able to fool the interrogator that it can think. Under the Turing Test, if there were ten interrogators, the machine only needs to fool at least seven of them. But this begs the question: is a machine really thinking, or is it just *simulating* the process of thinking?

AI was borne out of the desire to replicate human cognitive processing, if not human intelligence itself. AI takes Thomas Hobbes' words: "[b]y ratiocination, I mean computation"[14] and applies them almost literally. With these in mind, algorithms called *artificial neural networks* mimic the neurons in the human brain. Each artificial neuron is a mathematical model, such that when grouped together, these neurons form a network capable of cognition.[15] It is this knowledge of the basic physiology and function of neurons in the brain, along with propositional logic and Turing's theory of computation, that gave rise to AI.[16]

Turing is often considered the most influential person in the field of AI. As early as 1947, he lectured on the topic and introduced the aforementioned Turing Test, machine learning, genetic algorithms, and reinforcement learning in his 1950 article *Computing Machinery and Intelligence*. Meanwhile, the term "artificial intelligence" was coined in 1956 by John McCarthy. He, along with other US researchers, organized a two-month

---

[13] A.M. Turing, *Computing machinery and intelligence*, 59 MIND 433, 433 (1950).

[14] Thomas Hobbes, *Of Philosophy*, in THE ENGLISH WORKS OF THOMAS HOBBES 3 (I Sir William Molesworth, Bart. ed., 1839).

[15] Yavar Bathaee, *Artificial Intelligence Opinion Liability*, 35 BERK. TECH. L.J. 113, 139 (2020).

[16] STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 16 (3rd ed. 2010).

workshop in Dartmouth with the goal of having a machine simulate any aspect of learning or feature of human intelligence.[17] It was here that two researchers, Allen Newell and Herbert Simon, were able to prove most of the theorems on the foundations of mathematics.[18] Then, from 1952 to 1969, various other researchers developed the first AI programs. These programs were greatly limited, however, by the available computing resources and programming tools of the time.[19]

At that point, it seemed as though AI research was on a road to continuing success. But from 1966 to 1973, Newell and Simon encountered three main difficulties. First, most programs relied heavily on syntax and lacked the required background knowledge on the subject matter involved. For example, attempts to translate Russian to English proved futile because mere syntactic translation did not preserve the intended meaning of the sentences. "The spirit is willing but the flesh is weak" was famously retranslated as "the vodka is good but the meat is rotten."[20] Second, early AI models lacked generalizability. Researchers were disappointed to find out that although their AI programs worked for certain problems, they were unable to scale the programs to solve larger queries.[21] Third, the basic structures being used for machine learning were very limited. The two-input perceptron, a simple form of neural network, performed poorly when given two different inputs. As a result, confidence in neural network research waned and research funding decreased. From 1969 onwards, however, the demand for knowledge-based systems grew due to their increase in applications to real-world problems.[22] And since 1980, the AI industry has boomed from massive investment, with companies pouring in billions of dollars on expert systems and robots.[23]

### A. Defining AI

There is no single definition for AI. This issue comes not from what "artificial" means, but from what "intelligence" encompasses.[24] An AI can

---

[17] *Id.* at 17.
[18] *Id.* at 18
[19] *Id.*
[20] *Id.* at 21.
[21] *Id.*
[22] *Id.* at 24.
[23] *Id.*
[24] Matthew Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J. L. & TECH 353, 359 (2015).

be viewed using four approaches, as a system that: (1) thinks as human, (2) acts as human, (3) thinks rationally, or (4) acts rationally. An AI thinks humanly when it behaves as a human being does. In this approach, it does not matter whether the system decides correctly, or if it even solves the problem at all. What experts are more concerned with is whether the system's decision-making process resembles that of a human solving the same problem.[25] An AI acts as human when it is capable of natural language processing, knowledge representation, automated reasoning, and machine learning.[26] Lastly, an AI thinks rationally when it follows logic, and acts rationally when it is able to achieve the best possible outcome.[27]

Nevertheless, the term AI has been used to refer to "a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."[28] In short, there is AI when computer programs can solve problems that usually require human intelligence.[29] An intelligent entity may have the five features of: (1) communication, or the ability to communicate, (2) internal knowledge or the knowledge about oneself, (3) world knowledge or awareness of the outside world, (4) intentionality, which refers to having a goal-driven behavior, and (5) creativity.[30] Although an entity need not possess all five features to be considered intelligent, most AI entities have all of them.[31]

An AI may be an entity (hardware) or a tool (software). As an entity, AI has a physical manifestation capable of acting upon its environment, while as a tool, it is a mere computer program.[32]

As to its capabilities, AI can be weak or strong. Weak AI only acts as if it were intelligent, whereas strong AI is one that actually thinks. The AI

---

[25] RUSSELL & NORVIG, *supra* note 16, at 3.

[26] *Id.* at 2.

[27] *Id.* at 4.

[28] Organisation for Economic Co-operation and Development, *Explanatory Memorandum on the Updated OECD Definition of an AI System* (OECD Artificial Intelligence Papers No. 8, 2024), *available at* https://doi.org/10.1787/623da898-en.

[29] Damodar Singh Rajpurohit & Rishika Seal, *Legal Definition of Artificial Intelligence*, 10 SUPREMO AMICUS 87, 89 (2019).

[30] Roger C. Schank, *What is AI, Anyway?*, 8 AI MAG. 59, 60 (1987).

[31] Gabriel Hallevy, *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, 4 AKRON INTELL. PROP. J. 171, 176 (2010).

[32] Rajpurohit, *supra* note 29, at 89.

that exist today are only of the weak type as these systems still rely on humans for functioning. As of the time of writing, no machine or software has yet to achieve strong AI.[33]

AI can also be narrow or general, depending on its use. It is narrow when it is developed for a specific purpose or task. Examples include virtual assistants, and speech and text recognition bots. It is general, on the other hand, when it is developed to function like a human being.[34]

Another classification can be made according to the degree of human intervention involved. AI in this case may be automated or autonomous. When the AI is manually configured and pre-programmed with human instructions, it is said to be automated. If the AI is heavily reliant on machine learning, such that it learns and makes decisions by itself given its inputs, the AI is autonomous. This does not mean that human intervention is totally absent; The human input is limited to assigning the tasks and how the AI accomplishes said task is left to the system.[35] These classifications of AI become significant later on when discussing liability allocation.

## B. The Benefits of AI

Quality of life has improved immensely because of AI. We discuss here the various dimensions of human action improved by it.

### *1. Speed*

One of AI's biggest advantages is the speed at which tasks are accomplished and completed. The use of AI in medical diagnoses, for example, results in faster decision-making, allowing healthcare providers to deliver treatment to patients quickly and accurately.[36] AI can also be used in securities trading, where the difference between losses and huge financial

---

[33] Rafael Dean Brown, *Property ownership and the legal personhood of artificial intelligence*, 30 INFO. & COMM'N TECH. L. 2, 211 (2021).

[34] B Darshan Bhora & Kuldeep Shravan, *Demystifying the Role of Artificial Intelligence in Legal Practice*, 8 NIRMA UNIV. L J. 1, 3 (2019).

[35] Sahara Shrestha, *Nature, Nurture, or Neither?: Liability for Automated and Autonomous Artificial Intelligence Torts Based on Human Design and Influences*, 29 GEO. MASON L. REV. 375, 391–92 (2021).

[36] Sarah Kamensky, *Artificial Intelligence and Technology in Health Care: Overview and Possible Legal Implications*, 21 DEPAUL J. HEALTH CARE L. 1, 6 (2020); *See also* Sword, *supra* note 12, at 215, *and* Michael Hatfield, *Professionally Responsible Artificial Intelligence*, 51 ARIZ. ST. L.J. 1060 (2019).

gains is just a few milliseconds.[37] Lawyers, meanwhile, tediously scour through huge amounts of data particularly during the discovery phase. With the help of Technology Assisted Review, though, data are sorted into well-organized files to make discovery easier and faster for lawyers.[38] AI tools also assist lawyers in creating and closing documents in just a matter of minutes.[39]

*2. Adaptability*

Unlike other machines or software that need frequent configuration to perform a new or additional function, AI has the advantage of adaptability. In cybersecurity infrastructure, for example, AI detects vulnerabilities not previously found by traditional security measures or that are similar to others previously found. Rather than depending on rules, the AI relies on observation to find associations and patterns.[40] There is no need for developers to exhaustively specify the possible threats to trigger the defending mechanism of the AI; the AI does this by itself.

*3. Accuracy*

In the medical field, AI is commonly used to make diagnoses based on patient data like scans and biopsies. The AI developed at the Auckland University of Technology's Knowledge Engineering Discovery Research Institute has a 95% chance of accurately predicting a stroke occurring one day ahead, and a 70% chance for those occurring 7 and 11 days ahead. This accuracy is revolutionary, given that stroke is the third most common cause of death in the United States.[41]

The accuracy of an AI also contributes to e-commerce. Instead of using the traditional methods of advertising, such as making personalized pitches and displaying relevant content, online retailers can rely on customer data to make accurate product recommendations. This increases sales and retention rates.[42]

---

[37] Grimm, *supra* note 2, at 34.

[38] Bhora & Shravan, *supra* note 34, at 5.

[39] Eckart, *supra* note 9, at 282.

[40] Cristian-Vlad Oancea, *Artificial Intelligence Role in Cybersecurity Infrastructures*, 4 INT'L. J. INFO, SEC. & CYBERCRIME 59, 61 (2015).

[41] Callaghan Innovation, *Thinking Ahead: Innovation through Artificial Intelligence*, *at* 25, *available at* https://www.callaghaninnovation.govt.nz/sites/all/files/ai-whitepaper.pdf.

[42] *Id.* at 18.

But AI is a double-edged sword. With the advantages come the risks associated with its use.

## C. The Risks of AI

### *1. The "Black Box Problem"*

One of the biggest challenges when it comes to assessing AI is the "black box problem," or that there is no readily transparent way to understand how a neural network comes up with a result, hence, the term "black box." It is largely a consequence of how algorithms, such as neural networks and support vector machines, usually make decisions using high-dimensional patterns. If there are two parameters, the algorithm searches for a pattern in two or more dimensions. Humans can visualize a two- or even three-dimensional space, but not when there are a hundred or more parameters, resulting in a hundred or more dimensions.

The black box problem can also be caused by nonintuitive patterns. Algorithms can make decisions that do not follow the usual cause-and-effect reasoning that drives the scientific method.[43] This aspect of the problem can best be explained by the "beer and diapers story."[44] It is said that beers and diapers are placed next or close to each other in shelves in supermarkets and retail stores. While the connection is not so obvious nor logical, statistics have shown that middle-aged men who purchase diapers are likely to purchase beer at the same time. The final insight, therefore, is that items that are likely to be bought together are placed next or close to each other in stores. The reasoning behind such insight, on the other hand, is not as scientific. The same is true for the proximity of data points in the dimensional vector. Some data points may be closer in the vector than others for no apparent, intuitive reason, and may, therefore, result in seemingly confusing patterns.

The black box problem is a challenge precisely because understanding the process of coming up with the result may sometimes be as important as the result itself, no matter how optimized or efficient the result may be. While a patient may not need to understand how an AI was able to come up with its diagnosis, a bank depositor might perhaps want to

---

[43] Solow-Niederman, *supra* note 3, at 657, *citing* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines,* 87 FORDHAM L. REV. 1085 (2018).

[44] Christopher Jermaine, *Finding the most interesting correlations in a database: How hard can it be?*, 30 INFO. SYS. 21, 21 (2005).

understand why, for example, his loan was not granted. The bank cannot simply say that such a decision was arrived at using code.[45] An AI "can only be trustworthy if it can be explained to humans."[46]

Recent US jurisprudence is instructive on this point. In *K.W. v. Armstrong*,[47] the plaintiffs were adults suffering from disabling and chronic conditions, and whose respective Medicaid assistances were substantially reduced by the Idaho Department of Health and Welfare ("IDHW") allegedly without explanation.[48] The calculation for the budget was done using Idaho's automated decision-making system algorithm, which received as input factors such as the type of disability, the need for nursing and living services, and the level of hearing, vision, and mobility, and outputted the amount that Medicaid needed to pay.[49] When asked to explain the reduction in budget, IDHW claimed that the formula is a trade secret. The American Civil Liberties Union ("ACLU"), however, was able to demand disclosure of the formula. The court determined the same to be unconstitutionally arbitrary and ordered Medicaid to reconfigure its system.[50]

*Berliner v. Nassau County*,[51] meanwhile, involved the use of a predictive model for reassessing real property tax for the period 2020-2021. When the residents were informed of their preliminary market values, they claimed that the reassessment was conducted arbitrarily because the algorithms and software used were undisclosed.[52] The county eventually disclosed the algorithm, causing residents to challenge its validity because it "had missing files that would not let the code run effectively." A New York lower court ruled in favor of the plaintiffs, holding that the use of the reassessment algorithm must be discontinued.

Both *K.W.* and *Berliner* involved property rights, and the respective courts ruled for the disclosure of the algorithm. The court, however, thought

---

[45] William Magnuson, *Artificial Financial Intelligence*, 10 HARV. BUS. L. REV. 337, 353 (2020).

[46] Grimm, *supra* note 2, at 61.

[47] D.W. *ex rel.* K.W. v. Armstrong, No. 1:12-cv-22-BLW (D. Idaho Mar. 28, 2016) (mem. decision and order).

[48] *Id.* at 2.

[49] *Id.* at 6.

[50] *Id.* at 20–25.

[51] Berliner v. Nassau Cnty., 605904/2019, 2020 NYLJ LEXIS 311 (N.Y. Sup. Ct. Nassau Cnty. 2019).

[52] *Id.* at *10–12.

differently in *Houston Federation of Teachers v. Houston Independent School District*,[53] which also involves a property right: employment. In this case, the Houston Independent School District ("HISD") developed a data-driven teacher appraisal system to evaluate teachers based on instructional practice, professional expectations, and student performance, with this last criterion measured based on student growth on standardized tests.[54] Scores were compared to the statewide average for students in the same grade or course using the Educational Value-Added Assessment System ("EVAAS").[55] The plaintiffs claimed that their employment contracts, retirement plans, and teaching plans were impacted by this scoring system without any transparency. The parties eventually agreed to settle, but the court nevertheless ruled that "[w]hile teachers may not be able to verify the accuracy of their EVAAS scores, a suitably definite rule or regulation is not rendered unconstitutionally vague simply because it may be unfair or prone to error."[56]

Even more surprising is the court's ruling against disclosure when it comes to the fundamental right to liberty. The subsequent cases involve the Correctional Offender Management Profiling for Alternative Sanctions ("COMPAS"), a risk assessment tool that determines a person's likelihood for recidivism. In *State v. Loomis*,[57] respondent Eric Loomis pled guilty to two crimes. Before he was sentenced, he was assessed using COMPAS. Since he scored high, he was sentenced to prison rather than being placed on probation.[58] He argued that he was denied due process because of COMPAS,[59] but the Supreme Court disagreed. It held that there is no violation of the right to be sentenced based upon accurate information even if the defendant cannot assess the tool's accuracy.[60] It was sufficient that judges using COMPAS risk scores are aware of the following cautions:

> [(1)] the proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are to be determined; (2) risk assessment compares defendants to a national sample, but no

---

[53] Houston Fed'n of Teachers v. Houston Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1183 (S.D. Tex. 2017).

[54] *Id.* at 1171.

[55] *Id.* at 1172–73.

[56] *Id.* at 1183.

[57] State v. Loomis, 371 Wis. 2d 235 (Wis. 2016).

[58] *Id.* at 246–47.

[59] *Id.* at 277.

[60] *Id.* at 277–78.

> cross-validation study for a Wisconsin population has yet been completed; (3) some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism; and (4) risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations.[61]

Additionally, the algorithm was not deemed discriminatory, even though it takes gender into account, "[b]ecause men have higher recidivism rates."[62]

As shown in these cases, a certain degree of explainability is crucial for purposes of liability allocation. Should the algorithm be a total "black box" such that the developer cannot explain how the result came to be, an injury or harm caused by the algorithm cannot be attributed to the developer or, at the very least, the liability is mitigated. The opacity of the model equals the developer's degree of knowledge.

*2. Bias*

A machine learning algorithm is only as good as its data. If the input data was skewed or biased from the beginning, the algorithm cannot be expected to perform for a general population. As such, algorithms may tend to "perpetuate the very biases they are often intended to prevent."[63] This could happen when the algorithm is either purposely fed bad data, or seemingly neutral data that has unrecognized bias in it.[64]

In insurance, for example, an algorithm may be developed to speed up the approval of policy applications. Risky applications, such as those where the insured is into extreme sports like skydiving or paragliding, require more time for review, but those considered less risky may be approved immediately. To determine the degree of risk involved for each application, the algorithm may take as input basic information like age, gender, income, and educational background. An algorithm must be trained on huge volumes of data for it to be accurate and non-discriminatory. If it was solely trained on applications by those aged 30 years and below, an application by one aged 50 is highly likely to be rejected. The same can be said for an algorithm

---

[61] *Id.* at 270.

[62] *Id.*

[63] Grimm, *supra* note 2, at 42.

[64] Magnuson, *supra* note 45, at 356.

trained solely on applications by males, those with low annual income, or those with no college degrees.

The bias may at times not even be this direct. An algorithm using proxy variables, like a type of job typically held by either men or women, can be indirectly discriminatory. It is even feared that the use of deep learning systems, or those highly complex and dynamic algorithms, may blur the line between direct and indirect discrimination. A user may claim that their decisions "are justified by high-accuracy and blind to any kind of bias (because the prohibited variable is not in the training data.)"[65]

Discrimination by algorithms has already been considered by courts, with a recent case decided in favor of the alleged victim of discrimination. In Canada, the Corrections and Conditional Release Act requires the Correctional Services of Canada ("CSC") to ensure that all information about an inmate is accurate and complete.[66] To this end, the CSC uses five different algorithmic tools— the Hare Psychopathy Checklist-Revised ("PCL-R"), the Violence Risk Appraisal Guide ("VRAG"), the Sex Offender Risk Appraisal Guide ("SORAG"), the Static-99, and the Violence Risk Scale – Sex Offender ("VRS-SO")—to assess the likelihood of recidivism. Ewart, a Métis[67] serving two sentences for murder and attempted murder, challenged these tools in *Ewart v. Canada*.[68] Ewart argued that these tools were modeled on non-indigenous people and therefore less accurate for him. The Supreme Court sided with Ewart and held that the CSC breached their statutory duty when it failed to study the cross-cultural validity of the tools. As a result, indigenous inmates were "less likely to be granted early release" and were classified as "high-risk" since the tools had been developed using information from non-indigenous populations.[69]

While courts have sided with victims of discrimination in some cases, there are also cases that demonstrate how courts may refuse to pass

---

[65] Mireille Hildebrandt, *Discrimination, Data-driven AI Systems and Practical Reason*, 7 EUR. DATA PROT. L. REV. 358, 361–62 (2021).

[66] The Corrections and Conditional Release Act, S.C. 1992 c. 20, § 24(1) (1992). (Can.).

[67] *See* Larry Chartrand, *Métis Identity and Citizenship*, 12 WINDSOR REV. LEGAL & SOC. ISSUES 1, 34–35 (2001), where it is explained that the French word Métis means "mixed", referring to the indigenous peoples who are of mixed European and Indian blood in Canada.

[68] Ewert v. Canada, [2018] 2 R.C.S. 165 (2018) (Can.).

[69] *Id.* at 198.

upon the accuracy of algorithms. In *Lynch v. Florida*,[70] undercover police officers were able to take a photograph of a man called "Midnight" who sold them cocaine. They sent the said photograph to an analyst, who used a facial recognition program called Face Analysis Comparison Examination System to compare the photograph with others from the law enforcement database.[71] Plaintiff Lynch's photo was a weak match, but was the first listed photo from the results. The analyst then told the officers that Lynch and Midnight were probably the same person. Although Lynch questioned the accuracy of the system, the officers were able to identify him as the one who sold them the cocaine, which satisfied the court enough for his conviction.[72]

*3. Lack of Robust Testing for Validity and Reliability*

Owing to their complicated nature, many AI systems in use today lack proper evaluation. While algorithms may be tested, these tests are "rarely independent, peer-reviewed, or sufficiently transparent."[73] Part of the reason may be due to the lack of central authority for algorithms, unlike, for example, the Food and Drug Administration, which undertakes a rigorous testing and approval process before authorizing the distribution of a drug for public use.[74]

Courts have had several encounters with defective algorithms. *Zynda v. Arwood*[75] involved a fraud detection system used by Michigan's Unemployment Insurance Agency ("UIA") for its claimants. Upon identifying discrepancies in the claimants' information, the UIA would send a questionnaire to determine if the claims were fraudulent and the benefits unlawfully received, before automatically adjudicating the matter.[76] Should the system find the claim to be fraudulent, the benefits would be terminated and the claimant would be asked to return all the benefits previously received. The plaintiffs in this case, among them unemployment benefits claimants, contended that even if the UIA failed to send the fraud questionnaires, the system nevertheless issued automatic fraud determinations.[77] They also claimed that returned questionnaires were

---

[70] Lynch v. State of Florida, 260 So. 3d 1166 (Fla. Dist. Ct. App. 2018).
[71] *Id.* at 1169.
[72] *Id.*
[73] Grimm, *supra* note 2, at 48.
[74] *Id.*
[75] Zynda v. Arwood, 175 F. Supp. 3d 791 (E.D. Mich. 2016).
[76] *Id.* at 797–98.
[77] *Id.*

sometimes ignored or erroneously tabulated, resulting in incorrect fraud determinations and even when such errors occurred, the plaintiffs claimed that the UIA does not attempt to ensure that the fraud determinations are accurate. Plaintiffs thus argued that the UIA's fraud determination system violates various constitutional and federal statutory rights. Defendants moved to dismiss mainly on procedural grounds, and on the ground that plaintiffs failed to state a claim. All claims except for the plaintiffs' Eighth Amendment excessive fine claim survived,[78] although the parties eventually settled.[79]

The focus of *Barry v. Lyon*[80] is the federal Supplemental Nutrition Assistance Program ("SNAP"), the food assistance program administered by the US Department of Agriculture to low-income families and individuals, which disqualified convicted felons who fled custody or confinement to avoid prosecution, among others.[81] The State of Michigan developed the system which automatically terminated benefits upon a database match, without verifying whether the individual was actively feeling or has an outstanding warrant. The plaintiffs, SNAP recipients whose benefits were terminated because of the interface, sought an injunction to refrain the state "from automatic disqualifications based solely on the existence of a felony warrant and to provide adequate notices in the event of valid disqualification."[82] The district court ruled for the plaintiffs, holding that under the law, the disqualification must be based on three criteria: (1) that the felon is "actively fleeing", (2) to avoid prosecution for a crime that is a felony "under the law of the place from which the individual is fleeing," and (3) the authorities must be "actively seeking" to prosecute him or her for the offense.[83] Michigan's interface disregarded the third criterion, thereby unjustly depriving the plaintiffs of their right to assistance under SNAP. In addition, the disqualification notices sent by the state were also inadequate under the due process clause.[84]

---

[78] *Id.*

[79] *See* Kreps v. Michigan Unemployment Ins. Agency, No. 22-12020, 2023 WL 4494339 (E.D. Mich. July 12, 2023).

[80] Barry v. Lyon, 834 F.3d 706 (6th Cir. Mich. 2016).

[81] *Id. at 711.*

[82] *Id.* at 714.

[83] *Id.* at 717–18.

[84] *Id.* at 718–20.

*4. Uncertainty*

Given that the aim of AI developers is to teach a computer to think as a human does, developers do not provide the AI an algorithm on how to solve the problem. Rather, it lets the AI recognize a solution by itself.[85]

In 2002, scientists at Chalmers University of Technology in Sweden developed a winged robot without a flight algorithm. It only had wings that could move up, down, forwards, and backwards, and it was only programmed to achieve maximum lift. Although the robot struggled at first, performing only random movements, it eventually recorded the successful combinations of movements that achieved the flapping technique. In three hours, the robot learned to flap its wings and fly.[86]

Also in the same year, the Magna Science Centre in England initiated a project called Living Robots to confirm that the principle of survival of the fittest applies even as to robots. Robots were assigned as either predators or prey. The prey robots, while looking for food, were hunted by the predator robots which drain their energy. After being left unsupervised for 15 minutes, one of the prey robots called Gask has managed to escape out into the streets where it was eventually hit by a car.[87]

These examples show that despite their accuracy and the speed at which they arrive, AI decisions still tend to be uncertain. An AI system may react differently than what was intended by the developer[88] and it is not always possible to predict what an AI might do. Bound by the cognitive limitations of the brain, humans tend to settle for satisfactory solutions instead of optimal answers. AI, on the other hand, with their massive computational power, are not similarly limited. They can come up with numerous possibilities in a short amount of time, until they arrive at the most optimal solution. Despite this degree of uncertainty, Scherer argues, the AI

---

[85] Paulius Cerka, Jurgita Grigiene & Gintare Sirbikyte, *Liability for damages caused by artificial intelligence*, 31 COMPUT. L. & SEC. REV. 376, 379 (2015).

[86] Peter Augustsson, Krister Wolff, & Peter Nordin, *Creation of a learning, flying robot by means of evolution*, *in* GENETIC AND EVOLUTIONARY COMPUTATION CONF. 1279–85 (2002). *See also id.* at 381.

[87] *Robot wars for real*, BBC NEWS, Feb. 5, 2022, *at* http://news.bbc.co.uk/2/hi/sci-ence/nature/1801985.stm.

[88] Solow-Niederman, *supra* note 3, at 663.

is not doing anything other than executing code. Still, one cannot ignore the possibility that learning AI systems act so unpredictably that it would be unfair to hold the developers liable for harm.[89]

*5. Data Privacy Concerns*

Machine learning algorithms heavily rely on large volumes of data. This dependence urges companies to gather more of data, at times at the expense of privacy.[90] In marketing, for example, the more sensitive the data, the more useful for purposes of customer profiling. Once companies obtain a holistic view of their customers—from basic information like age, nationality, and gender, to more sensitive information like spending patterns, existing medical conditions, and internet browsing history—advertisements become more targeted and personal.

In *Patel v. Facebook*,[91] users from Illinois assailed how Facebook collected, used, and stored their face geometry from photos they uploaded without notice and consent, violating their privacy.[92] The case was eventually settled for $550 million.[93] A similar incident happened in *In re Google Assistant Privacy Litigation*,[94] where plaintiffs alleged that Google violated their right to privacy by secretly recording conversations—often in users' homes—through Google Assistant, sharing them with human subcontractors for transcription accuracy checks without the users' knowledge and consent, and storing the recordings for further analysis rather than destroying them after use.[95]

---

[89] Scherer, *supra* note 24, at 364–66.

[90] Magnuson, *supra* note 45, at 357–58.

[91] Patel v. Facebook, Inc., 932 F.3d 1264, (9th Cir. 2019); *See also* Calderon v. Clearview AI, Inc., 20 civ. 1296 (CM), (S.D.N.Y. 2020); Mutnick v. Clearview AI, Inc., Case No. 20 C 0512, (N.D. Ill. 2020), Vance et al. v. Amazon.com, Inc. (W.D. Wa. 2024); Vance et al. v. Facefirst, Inc. (C.D. Cal. 2023); Vance et al. v. Google LLC (N.D. Cal. 2024); Vance et al. v. Microsoft Corp. (W.D. Wa. 2022); Janecyk v. IBM Corp. (Cook Cty. Cir. Ct. Ill. 2020); *and* Stein v. Clarifai, Inc. (Cook County Cir. Ct. Ill. 2020).

[92] *Id.* at 1273.

[93] Rachel Pester, *Patel v. Facebook: Facebook Settles Illinois Biometric Information Privacy Act ("BIPA") Violation Suit*, HARV. J.L. & TECH. DIG., Feb. 14, 2020, *at* https://jolt.law.harvard.edu/digest/patel-v-facebook-facebook-settles-illinois-biometric-information-privacy-act-bipa-violation-suit.

[94] *In re* Google Assistant Privacy Litigation, 546 F. Supp. 3d 945 (N.D. Cal. 2021). *See also* Kumandan v. Google LLC, 19-cv-04286-BLF (N.D. Cal., 2022).

[95] *Id.* at 953–54.

In *Carpenter v. McDonald's Corp.,*[96] which also involved the use of audio recordings, the plaintiff Shannon Carpenter alleged that McDonald's collected customers' voiceprint biometrics from drive-through audio recordings using an AI voice assistant and stored biometric information without their customers' consent.[97] Likewise, in *Dinerstein v. Google, LLC,*[98] Dinerstein sued Google for using his personal health information—obtained from the University of Chicago Medical Center and de-identified—to develop a predictive health model without his express consent, arguing that Google could re-identify the data using information it already has from users' phones and mobile applications.[99] The court, however, dismissed the case, finding that Dinerstein failed to sufficiently prove damages.[100]

## D. Governing legislation

It is within the inherent police power of the state to regulate persons and property in order to promote the public welfare.[101] Given how much AI has proliferated people's daily lives, it is only right that it become the subject of regulation. Although not formal legislation, several rules and laws have been formulated governing robots and AI in other countries.

Many ethical guidelines surrounding the regulation of Artificial Intelligence have been based off of science fiction author Isaac Asimov's Three Laws of Robotics, provided in his novel *Runaround*:[102]

a. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

b. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.

c. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.[103]

---

[96] Carpenter v. McDonald's Corp., 580 F. Supp. 3d 512 (N.D. Ill. 2022).
[97] *Id.* at 514.
[98] Dinerstein v. Google, LLC, 484 F. Supp. 3d 561 (N.D. Ill. 2020).
[99] *Id.* at 570.
[100] *Id.* at 591–92.
[101] *See* United States v. Toribio, 15 Phil. 85 (1910).
[102] ISAAC ASIMOV, *Runaround*, *in* ASTOUNDING SCIENCE FICTION 94-103 (1942).
[103] Robin R. Murphy & David D. Woods, *Beyond Asimov: The Three Laws of Responsible Robotics*, 24 IEEE INTEL. SYS. 14, 15–16 (2009).

Due to the fictional nature of the novel, the three laws have no binding effect. But Asimov's principles have heavily influenced scholars such as Robin Murphy and David Woods, who have presented their own alternative laws using the former as a jumping-off point:

> a. A human may not deploy a robot without the human-robot work system meeting the highest legal and professional standards of safety and ethics.
>
> b. A robot must respond to humans as appropriate for their roles.
>
> c. A robot must be endowed with sufficient situated autonomy to protect its own existence as long as such protection provides smooth transfer of control to other agents consistent the first and second laws.[104]

Others, like Keith Miller, have also proposed their own, completely independent of Asimov's:

> a. The people who design, develop or deploy a computing artefact are morally responsible for that artefact, and for the foreseeable effects of that artefact. This responsibility is shared with other people who design, develop, deploy or knowingly use the artefact as part of a sociotechnical system.
>
> b. The shared responsibility of computing artefacts is not a zero-sum game. The responsibility of an individual is not reduced simply because more people become involved in designing, developing, deploying or using the artefact. Instead, a person's responsibility includes being answerable for the behaviors of the artefact and for the artefact's effects after deployment, to the degree to which these effects are reasonable foreseeable by that person.
>
> c. People who knowingly use a particular computing artefact are morally responsible for that use.
>
> d. People who knowingly design, develop, deploy or use a computing artefact can do so responsibly only when they make a reasonable effort to take into account the sociotechnical systems in which the artefact is embedded.

---

[104] *Id.* at 17–18.

> e. People who design, develop, deploy, promote or evaluate a computing artefact should not explicitly or implicitly deceive users about the artefact or its foreseeable effects, or about the sociotechnical systems in which the artefact is embedded.[105]

With the novelty of the technology, the development of new, universal ethical standards for AI demonstrates how there remains much to be explored in our social understanding of AI. Similarly, the policy and legal frameworks governing AI remain to be in their early stages globally.

In 2018, the Public Voice Coalition issued the Universal Guidelines for AI ("UGAI"),[106] which paved the way for policy discourse on AI worldwide. The UGAI combined universally-recognized human rights principles and data protection law in order to establish ethical guidelines on the use of AI. It established all individuals' right to transparency, or the right to be informed of the basis of AI decisions, including the factors considered, the logic, and the techniques that produced the outcome.[107] It also reaffirmed the individual's right to a human determination, or the right to be made subject to a final determination by a human.[108]

In 2019, the Organisation for Economic Co-operation and Development (OECD) issued their AI Principles, which highlighted the five principles for responsible stewardship of trustworthy AI: inclusive growth, sustainable development, and well-being; human-centered values and fairness; transparency and explainability; robustness, security, and safety; and accountability.[109] These same principles were adopted by the G20 in the same year.[110]

In 2021, the United Nations Educational, Scientific and Cultural Organization (UNESCO) issued its Recommendation on the Ethics of AI,[111]

---

[105] Keith Miller, *Moral Responsibility for Computing Artifacts: "The Rules"*, 13 IT PROF'L 57, 58–59 (2011).

[106] The Public Voice, *Universal Guidelines for Artificial Intelligence*, Oct. 23, 2018 *at* https://archive.epic.org/international/AIGuidleinesDRAFT20180910.pdf.

[107] *Id.*

[108] *Id.*

[109] OECD, *OECD AI Principles Overview,* OECD WEBSITE, *at* https://oecd.ai/en/ai-principles.

[110] OECD, *G20 Ministerial Statement on Trade and Digital Economy*, June 8–9, 2019, *available at* https://wp.oecd.ai/app/uploads/2021/06/G20-AI-Principles.pdf.

[111] *Recommendation on the Ethics of AI*, UNESCO WEBSITE, Nov. 23, 2021, *at* https://unesdoc.unesco.org/ark:/48223/pf0000381137.

aiming to establish a universal framework for AI governance that guides states, individuals, and organizations in embedding ethics throughout the AI lifecycle while ensuring alignment with international law. It sought to uphold human rights, dignity, equality, and environmental sustainability, fostering cultural diversity and protecting future generations. Additionally, it promoted inclusive, multi-stakeholder dialogue on AI ethics and equitable access to AI advancements, particularly benefiting low- and middle-income countries, least developed countries, landlocked developing countries, and small island developing states.

Up to this point, governments had only formally considered and adopted national strategies on AI without taking any concrete steps. No law had come to fruition yet, and previously-existing ones were either merely conjectural or focused on the economic implications of AI.[112]

This all changed in June 2024 when the European Union (EU) passed the AI Act,[113] , the world's first-ever comprehensive legal framework on AI. It classified AI according to its risk for purposes of regulation.[114] AI systems with unacceptable risk, such as those involving social scoring, were completely prohibited.[115] Meanwhile, high-risk[116] and limited-risk AI systems,[117] like chatbots and deepfakes, were regulated, while minimal-risk AI systems,[118] or those comprising the majority of AI applications currently available on the EU market, were unregulated.

Under the AI Act, providers of regulated AI systems are required to comply with certain obligations depending on the risk involved. For high-risk AI systems, providers must establish a risk management system, conduct data governance, draw up technical documentation, establish record-

---

[112] Avinash Dadhich, *A Critical View of Laws and Regulations of Artificial Intelligence in India and China*, 6 KATHMANDU SCH. L. REV. 1, 9 (2018).

[113] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 [hereinafter, "EU AI Act"] (2024), O.J. L. 2024/1689, *available at* https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (EU).

[114] EU AI Act, recital 26.

[115] Art. 5.

[116] Art. 6.

[117] Art. 52.

[118] Recital 165.

keeping, provide instructions for use, implement human oversight, and establish a quality management system.[119]

Upon its enactment, the EU AI Act faced various criticisms. According to the French IT consulting group Capgemini, the EU AI Act made it harder for global technologies to deploy AI systems in the region.[120] It also appears that the law left out many AI systems from its application, such as biometric recognition systems used for mass surveillance, and AI systems used for "predictive policing" to assess the security risk of alleged illegal immigrants.[121] In terms of scope, Article 6(3) of the AI Act impliedly allowed developers to exempt themselves from obligations for high-risk AI systems, while Article 49 exempted law enforcement and migration authorities from some of the transparency obligations attached to the use of high-risk AI systems.[122]

Most recently, the Council of Europe (COE) opened the Framework Convention on AI or the AI Treaty for signatures in September 2024.[123] It is the first ever international legally binding treaty on AI with 41 signatories from all over the world as of writing. Unlike the EU AI Act which uses a risk-based approach, the AI Treaty adopted a rights-based approach, meaning that it ensures the responsible use of AI with respect to "human rights, democracy, and the rule of law."[124] To ensure its effective implementation, the AI Treaty established a follow-up mechanism in the form of a Conference of the Parties.[125]

---

[119] Arts. 8–21; For a summary of the act, *see also High Level Summary of the AI Act*, EU ARTIFICIAL INTELL. ACT WEBSITE, Feb. 2, 2024, *at* https://artificialintelligenceact.eu/high-level-summary/.

[120] Florence Loeve & Supantha Mukherjee, *Capgemini CEO Says EU Went 'Too Far' with AI Rules*, REUTERS.COM, Feb. 10, 2025, *at* https://www.reuters.com/technology/artificial-intelligence/capgemini-ceo-says-eu-went-too-far-with-ai-rules-2025-02-10.

[121] *Statement on Proposed AI Regulation*, CTR. FOR AI & DIGITAL POL'Y, July 28, 2021, *at* https://www.caidp.org/app/download/8334787563/CAIDP-Comments-on-EU-AI-Act-28072021.pdf.

[122] *The EU AI Act: A Failure for Human Rights, a Victory for Industry and Law Enforcement*, ACCESS NOW, Mar. 13, 2024, *at* https://www.accessnow.org/press-release/ai-act-failure-for-human-rights-victory-for-industry-and-law-enforcement.

[123] Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, C.E.T.S. No. 225, Sept. 5, 2024, *available at* https://rm.coe.int/1680afae3c.

[124] Art. 1.

[125] Art. 23.

Like the EU AI Act, the AI Treaty was met with several critiques from various stakeholders. Under Article 3, states are allowed to exclude all AI systems designed, developed or deployed for the protection of national security interests from its application.[126] This exclusion was criticized since it leaves important areas of AI application—such as autonomous weapons or surveillance—unregulated, posing risks to privacy and civil liberties. The AI Treaty also had limited application, such that only the public sector and the private sector acting on their behalf are covered. Each state therefore had a wide discretion as to how the treaty will apply to private actors, who are ultimately involved in the AI lifecycle, which may lead to discrepancies in implementation.[127]

In the Philippines, there are essentially no laws governing AI. Republic Act No. 9262 or the Electronics Engineering Law of 2004, the only existing legislation in the Philippines that includes the word "robotics", only provides for the regulation of the practice of electronics engineering and electronics technician professions in the country.[128]

On the other hand, while it may be argued that the terms "labor-saving devices" under the Labor Code[129] and "automated processes" under the Data Privacy Act[130] include AI, these laws still fail to capture the unique nature of AI. "Labor-saving devices" under the Labor Code, as one of the authorized causes for termination of employment, refers to mechanical or technological innovations like machinery or systems that directly replace manual labor in industrial or physical tasks. In contrast, AI encompasses a broader scope, including cognitive and decision-making functions, which were likely not contemplated when the Labor Code was enacted.

"Automated processes" or "automated decision-making" under the Data Privacy Act, on the other hand, refers to "a wholly or partially automated processing operation that can make decisions using technological means totally independent of human intervention,"[131] often involving

---

[126] Art. 3.

[127] Karine Caunes, et al., *Open Letter to Council of Europe AI Convention Negotiators: No to the Abdication of Our Rights!*, CTR. FOR AI & DIGITAL POL'Y, Jan. 24, 2024, *available at* https://www.caidp.org/app/download/8496552763/COE-AI-Treaty-ExpertsLetter24012024.pdf.

[128] Rep. Act No. 9292 (2004), § 5.

[129] LABOR CODE, art. 283.

[130] Rep. Act No. 10173 (2012), § 16(c)(6).

[131] Nat'l Privacy Comm'n (NPC) Circ. No. 1701 (2017), § 3(b). Registration of Data Processing and Notifications regarding Automated Decision-Making.

profiling. While AI may fall under this in certain applications, it is not exclusively covered, as AI extends beyond data processing to creative, predictive, and adaptive systems. Stretching the interpretation of existing laws to cover AI is therefore a legislative overreach that may result in inconsistent application and further ambiguities.

Meanwhile, Philippine jurisprudence is also just as barren. The word "robot" was mentioned solely for the purpose of describing someone as acting in a mechanical, automatic, or "mindless" way.[132]

As of this writing, however, there are several bills pending in the19th Congress of the Philippines which attempt to address the country's most pressing concerns about AI.

In terms of education, House Bill No. 10751 seeks to establish the Generative AI in Education Council ("GAIEC"), which will oversee the integration of AI in educational practices to ensure ethical and responsible use, and to address potential impact on students. The GAIEC will formulate policies on curriculum development, teaching methods, and student support, while also monitoring AI implementation and conducting capacity-building activities for educators.[133]

Similar pending bills seek to create governing bodies, albeit with overlapping functions, to regulate AI. House Bill No. 7396 establishes the Artificial Intelligence Development Authority,[134] House Bill No. 7913 the Philippine Council on Artificial Intelligence,[135] House Bill No. 7983 the National Center for Artificial Intelligence Research,[136] House Bill No. 10385 the AI Bureau,[137] House Bill No. 10845 the AI Integration Council,[138] and

---

[132] *See* Feliciano Aureus v. Sec'y of Agric. and Commerce, 85 Phil. 1, 5 (1949). *See also* Ople v. Torres 354 Phil. 948 (1998); St. Luke's Hospital, Inc. v. Minister of Lab., 201 Phil. 706 (1982); People v. Dela Cruz, 203 Phil. 36 (1982); Republic v. De Los Angeles, 242 Phil. 590 (1988); Cortes v. Catral, 344 Phil. 415 (1997); *and* Jabon v. Usman, 510 Phil. 513 (2005).

[133] H. No. 10751, 19th Cong., 3rd Sess., § 5 (2024). Philippine Generative Artificial Intelligence (GenAI) in Education Bill.

[134] H. No. 7396, 19th Cong., 1st Sess., § 5 (2023). Artificial Intelligence Development and Regulation Bill.

[135] H. No. 7913, 19th Cong., 1st Sess., § 6 (2023).

[136] H. No. 7983, 19th Cong., 1st Sess., § 4 (2023).

[137] H. No. 10385, 19th Cong., 2nd Sess., § 3 (2024). AI Regulation Bill.

[138] H. No. 10845, 19th Cong., 3rd Sess., § 6 (2024). Artificial Intelligence Integration in Government Bill.

House Bill No. 10944 the Philippine Artificial Intelligence Board.[139] More notably, however, aside from creating the governing bodies, House Bill No. 7913 proposes the creation of an AI Bill of Rights.[140] House Bill No. 10944, on the other hand, penalizes certain prohibited uses of AI, such as the creation and spread of deepfake videos, the use of lethal autonomous weapon systems (LAWS), and the use of any AI applications that manipulate or control individuals in ways that could harm them physically or psychologically. AI-based social scoring and practices that violate privacy rights are also prohibited.[141] Additionally, the Act mandates the creation of a central database of AI companies.[142]

Several bills covering economic implications of AI have also been filed. Senate Bill No. 2762 considers cybersecurity, AI, and data center facilities as Tier III activities for purposes of income tax-based incentives.[143] The bill proposes that such facilities be allowed to avail of income tax holiday for a set amount of time, as well as special corporate income tax or enhanced deductions.

When it comes to labor, which is a matter of national interest, the pending bills take on a more conservative approach. For instance, House Bill No. 9448 prohibits employers and recruitment entities from using AI or automated systems as their sole or primary basis in the hiring and termination of employees.[144] The bill also prohibits the use of AI and automation technologies to replace human workers, unless an equivalent alternative employment opportunity for affected human workers is made available.[145] House Bill No. 10460 seeks to further amend the Labor Code to protect employees from layoffs caused by automation through AI.[146] It mandates fair compensation, re-skilling programs, and transition plans to new roles or opportunities for affected employees.[147] Unlike authorized causes, which require notification to employees and the Department of Labor and Employment (DOLE) at least thirty (30) days prior to the intended

---

[139] H. No. 10944, 19th Cong., 3rd Sess., § 5 (2024). Artificial Intelligence Bill.

[140] H. No. 7913, § 5.

[141] H. No. 10944, at § 9.

[142] § 6.

[143] S. No. 2762, 19th Cong., 3rd Sess., § 19 (2024).

[144] H. No. 9448, 19th Cong., 2nd Sess., § 5(a) (2023). Protection of Labor Against Artificial Intelligence (AI) Automation Bill.

[145] § 5(b).

[146] H. No. 10460, 19th Cong., 2nd Sess., § 1 (2024).

[147] § 2.

termination,[148] the bill requires employers to notify and consult with employees and the DOLE at least six (6) months before implementing AI-related layoffs.[149]

Clearly, regulating AI is not an easy endeavor. Owing to its nature, AI requires special, technical expertise, which makes it a difficult subject to regulate. Policymakers are generally unequipped with the technical knowledge to draft informed policies on code.[150] AI itself is already a complex topic even for specialists or experts, more so for those without the general knowledge at all. This is compounded by the dearth of experts trained to tackle AI or machine learning.[151]

Regulation proves to be difficult as well when considering the pace at which technology develops. Law tends to be reactionary; it merely responds to a past or an existing need. With how rapid AI technology advances, the law may not be able to keep up. Although the language of laws can be made as broad as possible to account for changes in the technology, this is not always possible.[152] Take for example regulatory interventions which can be implemented either before or after the fact. *Ex ante* legislation can incentivize manufacturers to prioritize safety, or require intensive testing and oversight before products hit the market. *Ex post* regulation, on the other hand, may include liabilities under tort and/or criminal law. However, this approach does not consider the fundamental constraints that exist when it comes to computer programming.[153]

There also exists the fear that subjecting AI to regulation will stifle invention, also known as the Collingridge dilemma. When technology was just beginning to emerge, their potential harmful effects were "not well enough known to support effective regulation or control."[154] Now that the harmful risks are known, two things have been deemed necessary for AI regulation: sufficient knowledge of what the risks are and their causes, and the capability to make interventions to reduce or completely eliminate these

---

[148] LABOR CODE, art. 298.

[149] H. No. 10460, 19th Cong., 2nd Sess. § 3.

[150] Solow-Niederman, *supra* note 3, at 638.

[151] *Id.* at 659.

[152] Harrison Fawcett, *Artificial Intelligence, Robots and the Law*, 39 U. TAS. L. REV. 168, 169 (2020).

[153] Solow-Niederman, *supra* note 3, at 667.

[154] Edward A. Parson, *Social Control of Technological Risks: The Dilemma of Knowledge and Control in Practice, and Ways to Surmount It*, 64 UCLA L. REV. DISCOURSE 464, 467 (2017).

risks. Often, however, these two elements are in tension with each other.[155] In the medical field, for instance, an algorithm developed by researchers at Stanford has an accuracy rate of less than 75%. Each time the algorithm makes a misdiagnosis, which is 25% likely, the researchers may be held liable under a strict liability theory. Production of the technology will then slow down, if not cease completely, out of fear of liability. If this is to happen, it would amount to a huge setback for the medical community, since the full potential of the use of AI technology can be hampered.[156] Additionally, there is the fear that the burden of responsibility is disproportionate to the person eventually held liable—whether it is the manufacturer or the developer. This could then result in fear on the part of the manufacturer or developer of publicly revealing their identity, potentially moving technological development from official to unofficial markets.

## III. LEGAL PERSONHOOD

The goal of the Turing Test is for the interrogator to be able to tell which entity between X and Y is the machine and which is the human. This endeavor certainly has become more difficult since Turing's time. The court has already recognized the difficulty in "[p]roving that an actual person is behind something like a social-networking account […] in an era when Twitter bots and other artificial intelligence troll the internet pretending to be people."[157] While no machine or software has yet to achieve strong AI, the distinction between a machine and a human being gets more confusing as technology advances. Google employee Blake Lemoine was recently put on leave after claiming that LaMDA, one of Google's AI models, is sentient. This is reportedly after, when asked what the word "soul" meant to it, LaMDA responded with, "To me, the soul is a concept of the animating force behind consciousness and life itself."[158]

Obviously, AI cannot be considered human, although it may possess human-like attributes. Humans are able to perceive color, smell, sound, taste, and touch using their senses, while AI can only do so in the form of data. Once it does, it processes this data through code, unlike humans who

---

[155] *Id.*

[156] Kamensky, *supra* note 36, at 13–15.

[157] *In re* C.W., Appeal No. C-180677, 7-8 (Ohio Ct. App. 2019).

[158] Brian Christian, *How a Google Employee Fell for the Eliza Effect*, THE ATLANTIC, June 21, 2022, *at* https://www.theatlantic.com/ideas/archive/2022/06/google-lamda-chatbot-sentient-ai/661322.

understand their environment by intuition. The unpredictability of human discourse comes from the differing use of language which, in turn, is caused by many different factors. AI unpredictability, however, is caused by none other than code.[159]

AI is technically a machine, just as a calculator and a microwave oven are machines. But calculators and microwave ovens are not capable of holding conversations like AI chatbots or of suggesting products one may similarly be interested in like an AI-powered recommender. The law assumes that the distinction between individuals and tools is night and day,[160] but the reality presents a gray area—AI is neither purely human nor purely machine. How must it be treated, then, in the eyes of the law?

### A. Legal persons

Diana Mădălina Mocanu simplifies subjecthood in Figure 1. Under the all-encompassing meta-category of subjects, there are legal subjects and things. Legal subjects can be personal and non-personal, where non-personal subjects are "not quite legal persons, but not things either," although they are holders of limited rights. This category includes animals. Personal legal subjects, on the other hand, can be legal and natural persons.[161]

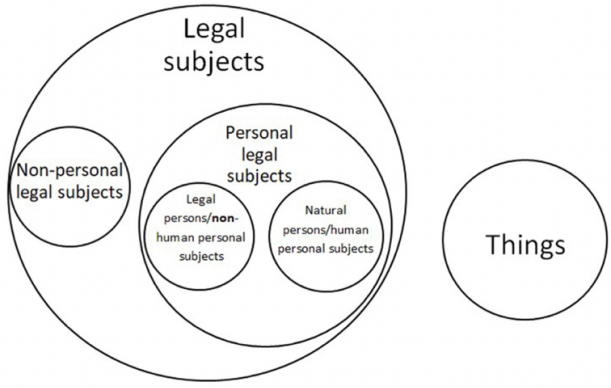


**Figure 1.**[162]

[159] Mireille Hildebrandt, *The Artificial Intelligence of European Union Law*, 21 GERMAN L.J. 74, 76–77 (2020).

[160] Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 535, 546 (2015).

[161] Diana Mădălina Mocanu, *Gradient Legal Personhood for AI Systems—Painting Continental Legal Shapes Made to Fit Analytical Molds*, 8 FRONT ROBOT AI 1, 5–6 (2022).

[162] *Id.* at 8.

Human beings are natural persons, but they did not always have the status of legal subjects. In the past, slaves and women are not considered such, leaving them unable to own property, vote, or express themselves freely.[163] They have only been conferred the status of legal subjects during the French Revolution, which resulted in the recognition of the dignity of a human person. Before 1990, or before the United Nations Convention on the Rights of the Child came into force, children were seen as objects of protection, rather than active participants or agents capable of legal duties in their own right.[164] What constitutes a legal subject, therefore, is more a political decision than anything else. The concept of a legal person, on the other hand, traces its origins back to when there was a need to confer rights upon an individual or a group of individuals.[165] Its purpose is mere convenience[166]—to "facilitate the regulation of human conduct by an organi[z]ed society."[167]

An artificial person[168] is "[a]n entity, such as a corporation, created by law and given certain legal rights and duties of a human being; a being, real or imaginary, who for the purpose of legal reasoning is treated more or less as a human being." It is also known as a fictitious person, juristic person, juridical person, legal person, or moral person.[169] It is a "being of legal existence, susceptible of rights and obligations, or of being subject of juridical relations"[170]; an "abstract being, formed for the realization of collective purposes to which the law has granted capacity for rights and obligations."[171]

---

[163] MIREILLE HILDEBRANDT, LAW FOR COMPUTER SCIENTISTS AND OTHER FOLK 5–6 (2020).

[164] Visa Kurki, *Active but not independent: the legal personhood of children*, 30 GRIFFITH L. REV. 395, 395 (2021).

[165] Brown, *supra* note 33, at 215–16.

[166] *Id.*

[167] S. M. Solaiman, *Legal Personality of Robots, Corporations, Idols and Chimpanzees: A Quest for Legitimacy*, 25 A.I. & L. 155, 159 (2017), *citing* Bryant Smith, *Legal Personality*, 37 YALE L.J. 283 (1928).

[168] For purposes of this research, the term "artificial person" is used interchangeably with "legal person" and "juridical person."

[169] BLACK'S LAW DICTIONARY 3618 (8th ed. 2004).

[170] Roldan v. Phil. Veterans Board 105 Phil. 1081, 1085 (1959), *citing* 2 Sanchez Roman, p. 119, *quoted in* Padilla's Civil Code Annotated, Vol. 1, 94 (1956 ed).

[171] ARTURO M. TOLENTINO, I COMMENTARIES AND JURISPRUDENCE ON THE CIVIL CODE OF THE PHILIPPINES 179 (2004 ed.).

The *Código Civil* or the Spanish Civil Code of 1889, the basis of the Philippines' New Civil Code and still the governing code in Spain at present, enumerates the following as legal entities:

> Article 35. The following are juridical persons:
>
> a. Corporations, associations and foundations of public interest recognized by the law. Their personality shall begin from the very moment in which they should have been validly incorporated in accordance with the law.
>
> b. Associations of private interest, whether civil, commercial or industrial, to which the law grants legal personality independent of that of each member.[172]

The New Civil Code of 1950, an evolution of the Spanish Civil Code, expands the definition and recognizes the following as juridical persons in Article 44:

> a. The State and its political subdivisions;
>
> b. Other corporations, institutions and entities for public interest or purpose, created by law; their personality begins as soon as they have been constituted according to law;
>
> c. Corporations, partnerships and associations for private interest or purpose to which the law grants a juridical personality, separate and distinct from that of each shareholder, partner or member.[173]

Juridical persons under the Spanish Civil Code may "acquire and possess property of all kinds, and contract obligations and exercise civil and criminal actions, in accordance with the laws and internal regulations."[174] They are still capable of the same under the New Civil Code.[175] These capabilities of juridical persons however, are "not completely at par" with those of natural persons. Juridical persons cannot exercise family rights nor make wills, but they can have a nationality, a domicile, and even a name. They

---

[172] SPANISH CIVIL CODE (1889), art. 35.
[173] CIVIL CODE, art. 44.
[174] SPANISH CIVIL CODE, art. 38.
[175] CIVIL CODE, art. 46.

can also enter into relations that do not require ties of blood, such as guardianship.[176]

Attempting to secure legal personhood is often seen as a potential pathway to get certain rights.[177] Animals, fetuses, trees, and rivers have been considered legal persons for their protection and preservation.[178] Inanimate objects like temples in Rome and church buildings in the Middle Ages have been made the subject of legal rights. A ship can be found guilty under admiralty law, and a family idol in India has become the subject of a custody dispute.[179] Non-human entities such as corporations and governments are likewise treated as juridical persons under Philippine law and other jurisdictions. This personhood serves as a shield or a veil to separate the corporation or government from the persons composing it or directing its affairs.

A corporation's separate personality is based on three prominent theories, which all provide interesting frameworks for the legal personhood of AI. Under the artificial and dependent person theory, the corporation is a mere creation of law for purposes of convenience. It is dependent on agents to act on its behalf. It is also at the mercy of the law – it does not come into being until after it has complied with all the requisites for incorporation.[180] Under Roman law, a legislative act is enough to bring a *persona ficta* into existence. During the Middle Ages, however, there needs to be some individual member upon whose rights the existence of a *persona ficta* depends. Without such individual member, the *persona ficta* cannot be granted rights, such as property ownership, and cannot sue or be sued. A will is necessary for the *persona ficta* to confer rights, but not to impose duties. In jurisdictions that adhere to this approach, including the United States and England, corporations are considered fictitious entities.[181]

Meanwhile, under the aggregate person theory, a corporation is an organization or association of people, without which it does not come into

---

[176] TOLENTINO, *supra* note 176, at 184.

[177] Roman Yampolskiy, *AI Personhood: Rights and Laws, in* STEVEN JOHN THOMPSON, MACHINE LAW, ETHICS, AND MORALITY IN THE AGE OF ARTIFICIAL INTELLIGENCE 1 (2021).

[178] *Id.*

[179] Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1239 (1992).

[180] Solaiman, *supra* note 172, at 163–65, *citing* John Niman, *In support of creating a legal definition of personhood*, 3 J. L. & SOC. DEVIANCE 142 (2012).

[181] Brown, *supra* note 33, at 216.

being. Its rights only stem from the rights of those composing it. As such, it exists as an aggregate of individuals, not a separate entity independent of its components.[182] The juristic person is composed of a person or a group of persons who, individually, is capable of holding a set of definite legal rights. Unlike the *persona ficta*, the juristic person is not required to have will power. Corporations in jurisdictions that adhere to this approach, like Germany, Spain, and France, are considered real persons.[183]

Lastly, there is the real and independent person theory, which states that corporations are real entities independent of the law creating them. They have a real presence in society, such that they can incur liabilities and be held liable for them.[184]

The main purpose of creating a specific legal status for AI, according to proponents, is to make it responsible for any damage it may cause because of its autonomous decisions and interactions with third parties.[185] In medical malpractice claims, for example, it is easier to apportion responsibility when AIs are involved in decision-making and are considered legal persons.[186]

Should AI be brought under the scheme of liability of Philippine law, the New Civil Code must be amended to further expand the definition of juridical persons. The provision, as amended, can then be read as follows:

> Art. 44. The following are juridical persons:
>
> (1) The State and its political subdivisions;
>
> (2) Other corporations, institutions and entities for public interest or purpose, created by law; their personality begins as soon as they have been constituted according to law;
>
> (3) Corporations, partnerships and associations for private interest or purpose to which the law grants a juridical personality, separate and distinct from that of each shareholder, partner or member;

---

[182] Solaiman, *supra* note 172, at 163–65.

[183] Brown, *supra* note 33, at 216.

[184] Solaiman, *supra* note 172, at 163–65.

[185] Mady Delvaux, *Draft Report with recommendations to the Commission on Civil Law Rules on Robotics* (2015/2103(INL)), Introduction art. AD.

[186] Sword, *supra* note 12, at 230–31.

> *(4) Artificially intelligent entities, whether physical or in the form of software, which have successfully passed the Turing Test.*[187]

This amendment should be deemed comprehensive enough to include all AI entities capable of legal personhood without affecting the existing system of liability for other legal persons.

Admittedly, however, granting legal personhood to AI entities at this early stage may open a Pandora's box of new issues, rather than simply addressing the problem of liability. A more measured approach, consistent with the bills pending in Congress, would be to first enact separate legislation to govern AI entities.[188]

### B. Objections to Legal Personhood for AI

Proponents of conferring legal personhood upon AI based on corporate personhood fail to consider that unlike an AI, a corporation is composed of a group of people. Compared to idols, whose rights or duties depend on their human managers and custodians, who are human beings, AI entities have more independence from human intervention.

Personhood comes with a "bundle of rights,"[189] and whether AI should be bestowed with rights at all is still a topic for debate among experts. It has been argued that AI cannot own copyright since, under copyright law, "author" and "authorship" are not clearly defined.[190] Granting them copyright ownership creates even more uncertainties when it comes to legal standing. It also does not serve the purpose of the copyright law, which is "to incentivize people to create works by offering them exclusive rights," since AI needs no incentive.[191] In one author's words, "[c]omputer authorship is a law of the horse."[192] The court is more definite when it comes to patents, saying that Congress intended to limit the definition of an "inventor" to natural persons. The time may come when AI can satisfy the meaning of inventorship, "[b]ut that time has not yet arrived, and, if it does,

---

[187] Priyam Jhudele, *On Robot Crimes and Punishments*, 6 NLIU L. REV. 1, 22 (2016). (Emphasis supplied.)

[188] *Id.* at 23.

[189] Kurki, *supra* note 169, at 398.

[190] David W. Opderbeck, *Artificial Intelligence, Rights and the Virtues*, 60 WASHBURN L.J. 445, 459 (2021).

[191] Palace, *supra* note 3, at 234.

[192] Opderbeck, *supra* note 196 at 459.

it will be up to Congress to decide how, if at all, it wants to expand the scope of patent law."[193] Free speech rights may be conferred to AI, but only if the AI is of the strong type and if such rights "would be valuable to human listeners."[194] Weak AI can be conferred property rights, since the legal duties incident to property ownership can be performed by human agents. Property rights cannot be conferred to strong AI, however, because it is uncertain how it will exercise these rights as against other legal persons.[195]

Legal personhood does not stop at the conferral of rights, though. There must necessarily be a corresponding imposition of obligations. This, according to some authors, is where AI fails. They can never become more than property as they lack the requisite intentionality, desire, and feelings, and therefore, cannot be subject to duties.[196] Brown argues, however, that the existence of a will, intent, desires, and interests should not be the main determinants when deciding whether or not to impose a legal duty.[197] Just because an entity is capable of rights but incapable of duties does not mean it cannot be conferred the status of a legal person. Take children and infants, for example. Kurki argues that they are passive legal persons since they do not bear legal duties,[198] but they are the subject of rights under the UN Convention on the Rights of the Child.[199] Comatose patients and persons with dementia likewise lack the ability to bear duties, but they have the same rights as others. Nonetheless, intention and desire can easily be programmed into the AI depending on the purpose it is intended to serve. Such intent can be attributed to the AI itself, in the case of strong AI, or to its developer, in the case of weak AI.[200]

Other arguments against granting AI legal personality are more philosophical than legal. Yampolskiy says it is tantamount to an "assault on

---

[193] Thaler v. Hirshfeld, 558 F. Supp. 3d 238, 249 (E.D. Va. 2021). *See also* Thaler v. Comptr. Gen. of Patents Trade Marks and Designs (UKSC 2021), where the England and Wales Court of Appeal likewise ruled that only a person can be an inventor. *But see* Thaler v. Comm'r of Patents [2021] FCA 879 (Fed. Ct. of Australia 2021), where the Full Federal Court of Australia ruled that an AI system or device can be an inventor, but it cannot be an applicant for or a grantee of a patent.

[194] Opderbeck, *supra* note 196, at 459.

[195] Brown, *supra* note 33, at 230.

[196] Solum, *supra* note 185, at 1258. *See also* Solaiman, *supra* note 172, at 173, although referring to industrial robots, and Brown, *supra* note 33, at 220.

[197] Brown, *supra* note 33, at 220.

[198] Kurki, *supra* note 169, at 403.

[199] UN Convention on the Rights of the Child, pmbl., Nov. 20, 1989, 1577 U.N.T.S. 3.

[200] Brown, *supra* note 33, at 220.

human dignity," as humans could be relegated to an inferior position with AIs becoming more intelligent.[201] To illustrate, in 2017, a humanoid robot named Sophia was granted citizenship by Saudi Arabia, making her (or it?) the first robot citizen in the world.[202] Although the legal basis for the grant of citizenship to a robot was not specified, this granting of citizenship to a female robot caused outrage among Saudi Arabian women, who were still required to undergo the male guardianship system.[203] The outrage is understandable. It was seemingly so easy for the Kingdom to grant citizenship to a female robot, when it took over 60 years for Saudi Arabian women to be even allowed to drive.[204] However, while these indignities can be understood, it can also be counterargued that giving an entity, such as AI, legal personality does not necessarily bring with it all the rights and privileges enjoyed by natural persons. Corporations, for example, can invoke the right against unreasonable searches and seizures,[205] but they cannot invoke the right against self-incrimination.[206]

Mark Coeckelbergh further argues that the act of giving rights to an entity requires moral consideration. It implies "that the entity in question has inherent worth and that therefore the entity needs to be treated as such irrespective of all other (human or non-human) considerations."[207] The robots of today are not moral subjects, adds Gunkel. Being mere tools and instruments, they are considered "neutral" beings "without valuative of [their] own." They are only means to an end. They are neither "true" nor "just," but are simply efficient.[208] Solaiman further states that legal personhood must only be conferred upon an entity that is able to enjoy the rights and perform the duties associated with such status.[209] It cannot be a mere object. However, it can also be argued that the robots they consider as

---

[201] Yampolskiy, *supra* note 183, at 3.

[202] Olivia Cuthbert, *Saudi Arabia becomes first country to grant citizenship to a robot,* ARAB NEWS, Oct. 26, 2017, *at* https://www.arabnews.com/node/1183166/saudi-arabia.

[203] Kristine Beckerle, *Boxed In: Women and Saudi Arabia's Male Guardianship System*, HUMAN RIGHTS WATCH, July 16, 2016, *at* https://www.hrw.org/report/2016/07/16/boxed/women-and-saudi-arabias-male-guardianship-system.

[204] *Saudi Arabia's ban on women driving ended at midnight*, June 24, 2018, BBC NEWS, *at* https://www.bbc.co.uk/newsround/44586038.

[205] *See* Stonehill v. Diokno, 126 Phil. 738 (1967).

[206] *See* Bataan Shipyard & Eng'g Co., Inc. v. Pres. Comm'n on Good Gov't, 234 Phil. 180 (1987).

[207] Mark Coeckelbergh, *Robot rights? Towards a social-relational justification of moral consideration,* 12 ETHICS INFO. TECH. 209, 210 (2010).

[208] David J. Gunkel, *The other question: can and should robots have rights?*, 20 ETHICS INFO. TECH. 87, 90 (2017).

[209] Solaiman, *supra* note 172, at 160.

mere tools and instruments are not as advanced as the ones existing now. Marx argues that there must be a distinction between the tools used by the worker and the machines that eventually replace the worker. The former are definitely not moral subjects, but the latter have the possibility to be. To quote Darling, "[w]hile toasters are designed to make toast, social robots are designed to act as our companions."[210]

## IV. FRAMEWORK FOR LIABILITY

As part of the broader approach of enacting separate legislation to govern AI entities, one critical aspect that requires careful consideration is the establishment of a liability framework. While some authors argue that such a framework is premature, this view reflects the understanding that the law is often reactionary—that legislation should follow actual, rather than merely theoretical, needs. Hildebrandt asks whether the level of autonomy assumed in, say, driverless cars, will ever be achieved.[211] Solum agrees by saying that "[n]o existing computer program currently possesses the sort of capacities that would justify serious judicial inquiry into the question of legal personhood."[212] According to William Magnuson, adopting a precautionary approach this early in the game will end up "stifling innovation," and if the fears never even come about, regulation then becomes a waste of resources.[213] The mere possibility of harm is not enough, he adds, and that there must be actual proof that current laws are not working, in order to justify the need for new ones.[214]

This research believes the contrary. That AI is being urged to be taught in law schools underscores the necessity of this discussion. Given how AI has already impacted a range of practice specialties, Johnson emphasizes that law students should be equipped to handle AI-related questions.[215] Otherwise, they will be left behind as the legal practice quickly shifts.[216] They

---

[210] *Id.*

[211] Hildebrandt, *supra* note 168, at 5.

[212] Solum, *supra* note 185, at 1231.

[213] Magnuson, *supra* note 45, at 379.

[214] *Id.* at 365.

[215] Brendan Johnson & Francis X. Shen, *Teaching Law and Artificial Intelligence*, 22 MINN. J.L. SCI. & TECH. 23, 35 (2021).

[216] *Id.* at 36.

must also be introduced to transactional AI tools that are increasingly being used in legal practice.[217]

According to Moore's law, the capacity of computer processors doubles every 18 months.[218] As knowledge level increases, people also become unable to control and process the volume of information, leading to increasing reliance on powerful computer systems. The divide between human capabilities and the computing power of computer systems widens as the latter surpass their human counterparts. It is not impossible, therefore, to anticipate computers that are autonomous and independent of human will. Google engineer Ray Kurzweil goes as far as to predict that AI will become superior to humans, so-called "technological singularity," by 2045.[219] And the more autonomous AI become, the less they can be controlled, which calls for new guidelines as to the legal liability of the various actors (i.e., the manufacturer, the operator, the owner, the user, etc.) involved. This is especially significant if the acts and omissions of the AI cannot be pinpointed to a specific cause or actor, or if the acts and omissions which caused harm could have been foreseen and otherwise avoided.[220]

Although the thinking machines of today are not the "killer robots"[221] that Tesla and SpaceX CEO Elon Musk had feared, precautions are still necessary because they can—and already have—the capacity to cause injury and harm. In an open letter to the United Nations in 2017, Musk, along with more than a hundred other signatories, called for a ban on the development and use of AI weaponry. These weapons are feared to bring armed conflict to a whole new scale and give rise to the Third World War. "Once this Pandora's box is opened, it will be hard to close," the letter says.[222] As the advancement of AI systems increases, the likelihood of damage also increases.[223]

---

[217] Eckart, *supra* note 10, at 285.

[218] *See* Gordon E. Moore, *Cramming more components onto integrated circuits*, *in* 38 ELECTRONICS 114 (1965).

[219] Cerka, et al., *supra* note 89, at 381–82.

[220] Delvaux, supra note 191, at 242.

[221] Cade Metz, *Mark Zuckerberg, Elon Musk and the Feud Over Killer Robots,* N.Y. TIMES, June 9, 2018, *at* https://www.nytimes.com/2018/06/09/technology/elon-musk-mark-zuckerberg-artificial-intelligence.html.

[222] Catherine Clifford, *Hundreds of A.I. experts echo Elon Musk, Stephen Hawking in call for a ban on killer robots*, CNBC, Nov. 8, 2017, *at* https://www.cnbc.com/2017/11/08/ai-experts-join-elon-musk-stephen-hawking-call-for-killer-robot-ban.html.

[223] Cerka, *supra* note 89, at 382.

The purpose of this discussion is not to cause panic. Rather, it is to introduce the possibility or impossibility of classifying AI as legal persons if only for purposes of allocating liability, which is the most pressing concern given recent advancements in technology. That AI has the ability to cause damage or injury is already acknowledged. What this aims to achieve is to "encourage those inclined to study robotics and the law to think systematically about emergence."[224]

## A. Why the Traditional Frameworks are not Enough

Existing frameworks hold the actor either directly or indirectly liable. Liability is direct when the person who committed the act or omission himself bears the sanction or punishment, and indirect when the person suffers the punishment because of the act or omission of another. As applied to AI, however, it is never held directly liable. An AI's act or omission is always traced back to a specific human agent, such as the manufacturer, operator, owner, or the user.[225] The imposition of liability further depends on how the AI is treated—whether as a tool or product, or as an agent, but without passing upon its legal personhood.

### *1. AI as a Tool or Product*

Those against granting legal personhood to AI, even if only for purposes of liability, argue that traditional frameworks and existing legal doctrines are sufficient. One alternative involves treating the AI as a mere tool or product, thereby holding their manufacturers, developers, and owners responsible for any potential liability.[226] In other words, the laws on product liability must be applied.

Under a typical products liability case, a consumer purchases a product that may be either fungible, such as food, or non-fungible, such as an AI. If he consumes it or uses it in accordance with the instructions by the manufacturer and yet suffers injury because of a defect, the consumer may bring an action against the manufacturer of the product.[227] The plaintiff only needs to prove that damage has occurred, and that such damage is caused by a defect or a failure to warn. A defect occurs when the product has not been

---

[224] Calo, *supra* note 165, at 545.

[225] Delvaux, *supra* note 191, at 244.

[226] Brown, *supra* note 33, at 213.

[227] JORGE COQUIA & ELIZABETH AGUILING-PANGALANGAN, CONFLICT OF LAWS: CASES, MATERIALS AND COMMENTS 425 (2000 ed.).

manufactured according to standards, or when a foreseeable risk of harm exists and the same could have been avoided or reduced by resorting to an alternative design.

Failure to warn also triggers a products liability case because the manufacturer failed to provide instructions on how to safely use or consume the product.[228] Manufacturers are not obliged to warn the user of every danger, but only those that they are aware of or could have reasonably foreseen.[229]

In this framework, since the AI is treated as a product, the remedy for injury caused by it—whether due to a defect or a failure to warn—rests upon its manufacturers, developers, and owners. According to Alejandro Zornoza et al., manufacturers can be:

1. The producer of a raw material and the manufacturer of a finished product or of a component part

2. The importer of the product

3. Any person putting their name, trademark, or other distinguishing features on the product

4. Any person supplying a product whose producer or importer cannot be identified[230]

As Zornoza et al. note, "a robot is composed of two things: software and hardware."[231] The producer (or the manufacturer) is responsible for the hardware, such as sensors and actuators, while the programmer (or developer) is liable for the software, including the AI's decision-making processes and learning capabilities.[232] When the harm results from a software failure or error, liability rests with the programmer.[233]

---

[228] Sword, *supra* note 12, at 224. *See also* Shrestha, *supra* note 35, at 381.

[229] Shrestha, *supra* note 35, at 381–82.

[230] Alejandro Zornoza, et al., *Robots Liability: A Use Case and a Potential Solution*, *in* ROBOTICS – LEGAL, ETHICAL AND SOCIOECONOMIC IMPACTS 57, 64 (George Dekoulis ed., 2017), *at* http://dx.doi.org/10.5772/intechopen.69888.

[231] *Id.* at 64.

[232] *Id.*

[233] *Id.* at 65.

Users fall under the category of developers. They are the ones who use the AI. When an AI has a closed architecture, meaning that it does not provide for any customizability or personalization, the developers become solely liable. Further, "When the robot allows a degree of personalization, there is an effect of displacement of liability and voluntary assumptions of risk by the user."[234] However, liability on the part of the user will not be absolute if the user's expertise on robots is not at par with the developer's and the risk is limited to the risk that he or she knows. The developers in this case share the burden with the users.[235]

For instance, the Supreme Court did not allow the application of strict liability in *Pascual v. Ford Motor Company Philippines, Inc.*,[236] because the car owner altered the vehicle's rear axle to carry more weight beyond the vehicle's capacity as represented by the manufacturers.[237] Since the manufacturers have already warned users of the weight limit, they cannot be held liable for damage sustained as a result of going beyond the vehicle's capacity.[238]

Zornoza et al., distinguish between the concepts of *owner* and *user* of a product: the owner is the person who purchased the product (and is not necessarily the user, but one who maintains or preserves the product), and the user is the one who uses. But if the owner also uses the product he or she has bought, they become one and the same person. If this is the case, the owner-user's liability is the same as discussed above.[239] Article 12 of the UN Convention Use of Electronic Communications in International Contracts places the responsibility upon the owners. Thus, when a computer is programmed to generate electronic communications, and in the process causes damage to another, the person on whose behalf the computer is programmed becomes liable for any message generated by the machine. Such message is deemed to come, not from the computer, but from the owner, who may be a natural person or a legal entity.[240]

---

[234] *Id.* at 66.
[235] *Id.*
[236] G.R. No. 220667 (Res.), slip op., Jan. 27, 2016.
[237] Pascual v. Ford Motor Company Phil., G.R. No. 220667, slip op. at 1.
[238] *Id.* at 4.
[239] *Id.*
[240] Bhora & Shravan, *supra* note 34, at 6. *See also* Cerka, *supra* note 89, at 383.

Existing jurisprudence on the matter follows this liability model. In *Umeda v. Tesla, Inc.*,[241] Yoshihiro Umeda was standing on the expressway near Tokyo, Japan when he was struck and killed by a Tesla Model X vehicle that was on autopilot. Umeda's spouse and child filed the case against Tesla, Inc. to seek damages under the theory of strict products liability and negligence.[242] The court dismissed the case, however, on the ground of *forum non conveniens*.[243]

Two other cases similar to *Umeda* involve self-driving vehicles and are based on strict liability grounds. In *Hudson v. Tesla, Inc.,*[244] the plaintiff sustained permanent injuries while operating a Tesla Model S vehicle and sued for damages, alleginig that, among others, the vehicle's autopilot system is defective and that Tesla failed to provide adequate warnings and instructions.[245] The case remains pending and has not yet been resolved by the court. In *Nilsson v. General Motors,*[246] the plaintiff, Oscar Nilsson, was riding his motorcycle alongside a self-driving 2016 Chevrolet Bolt vehicle when the vehicle suddenly veered into his lane and knocked him to the ground. He sued General Motors for negligence, although the parties eventually settled.

The problem with applying the strict product liability concept to AI is that the law generally only recognizes tangible products.[247] When the AI is embodied as a computer program without any physical manifestation, there may be a difficulty in establishing a product liability action.

Unlike products that are attributable solely to their manufacturer, AI technologies and systems are usually created and developed by multiple entities. Designers, engineers, and developers work together with varying

---

[241] Umeda v. Tesla Inc., No. 20-cv-02926-SVK, 2020 U.S. Dist. LEXIS 175286 (N.D. Cal. Sept. 23, 2020)

[242] *Id.* at *4.

[243] *Id.* at *22–23.

[244] Hudson v. Tesla Inc., 2018-CA-011812-O, Complaint (Fla. Cir. Ct. 2018), *available at* https://media.forthepeople.com/wp-content/uploads/2018/10/Tesla-autopilot-lawsuit-complaint.pdf

[245] *Id.* at 8–9.

[246] Nilsson v. Gen. Motors LLC, No. 3:18-cv-00471- KAW, Complaint (N.D. Cal. 2018), *available at* https://www.courthousenews.com/wp-content/uploads/2018/01/Self-driving.pdf.

[247] Calo, *supra* note 165, at 535–36.

degrees of involvement and contribution. This makes it more difficult to impose the blame on one specific individual.[248]

Further, despite the existence of warning labels or instruction manuals on AI, this may not be enough to constitute the informed consent requirement for the assumption of risk. In the first place, instruction manuals on AI will be extensive. Most people do not even bother to read through the terms and conditions of digital products and services. It will therefore be difficult on the part of manufacturers and developers to invoke assumption of risk when users do not find it necessary to heed warning labels that already exist.[249]

Additionally, this framework may be possible for weak AI, or AI that is incapable of independent volition, but not for strong AI. The element of foreseeability, necessary to establish causation, is not always present, especially when it comes to truly emergent AI behavior. An act is said to be negligent if an ordinary prudent person would be able to foresee an appreciable risk of harm to others, such that he would act in a more careful manner or avoid the act altogether.[250] For AI, however, it is unavoidable that the acts of otherwise useful technology "will legitimately surprise all involved."[251] Plaintiffs will have difficulty proving causation because autonomous AI is not just following code but making decisions on its own,[252] for which they cannot be solely faulted. The software they released initially may not be the same software at the time of the injury. This is especially true for AI that learn by reinforcement, or through its own experiences without human intervention. When this happens, control over the AI is beyond the developers' reach, as developers normally lack the capability to predict how the AI learns, adapts, and makes decisions.[253]

*2. AI as an Innocent Agent sans Legal Personhood*

Another alternative involves treating the AI as an agent. Under this framework, the AI entity is an "innocent agent" that does not possess any

---

[248] Kamensky, *supra* note 36, at 13–15. *See also* Delvaux, *supra* note 191, at 242.

[249] Amy L. Stein, *Assuming the Risks of Artificial Intelligence*, 102 B.U. L. REV. 979, 1032 (2022).

[250] Capili v. Sps. Cardaña, 537 Phil. 60 (2006).

[251] Calo, *supra* note 165 at 665.

[252] Shrestha, *supra* note 35, at 398.

[253] Sword, *supra* note 12, at 226. *See also* Brown, *supra* note 33, at 213–14, and Delvaux, *supra* note 191, at 243.

human attributes. Also called the perpetrator-via-another model, this approach is similar to a situation when an offense is committed by a child or a mentally incompetent person.[254] These types of persons, from the Revised Penal Code,[255] are exempt from criminal liability. AI systems under this liability model therefore do not incur any criminal liability, but their developer and/or user do, even if the robot actually perpetrates the crime. The AI is a mere instrumentality that does not and cannot possess the requisite criminal intent. AI systems of the present cannot be made subject to this model, as they possess advanced capabilities beyond their developer or user.[256]

The liability ultimately falls upon the principal or the person orchestrating the offense.[257] In the context of Philippine law, this means that the law on agency applies.[258] The principal becomes liable for the acts of the agent—in this case, the AI—as long as the acts are within the latter's authority. If the AI acts beyond the principal's authority, the latter is not bound except when the act was ratified expressly or impliedly. Furthermore, if the third person—in this case, the user of the AI—knows that the agent was acting beyond its power or authority, the principal cannot be bound by the AI's acts.[259]

This liability framework is at play in Commission on Elections (COMELEC) Resolution No. 11064, or the "Guidelines on the Use of Social Media, Artificial Intelligence, and Internet Technology for Digital Election Campaign and the Prohibition and Punishment of its Misuse for Disinformation and Misinformation in connection with the 2025 National and Local Elections and the BARMM Parliamentary Elections." Under this Resolution, an individual that uses AI as "false amplifiers" (e.g. fake accounts, bots, and astroturf groups) to propagate disinformation and misinformation in endorsing or campaigning against candidates or organizations, or to propagate disinformation and misinformation targeting the Philippine election system, the COMELEC, and electoral processes

---

[254] Hallevy, *supra* note 31, at 174.

[255] REV. PEN. CODE, art. 12.

[256] Hallevy, *supra* note 31, at 174.

[257] *Id.* at 179.

[258] CIVIL CODE, art. 1898. If the agent contracts in the name of the principal, extending the scope of his authority and the principal does not ratify the contract, it shall be void if the party with whom the agent contracted is aware of the limits of the powers granted by the principal. In this case, however, the agent is liable if he undertook to secure the principal's ratification.

[259] *See* Safic Alcan & Cie v. Imperial Vegetable Oil Co., Inc., 407 Phil. 884 (2006).

during the election and campaign period may be held guilty of an election offense.[260] The individual in this case acts as the principal who uses AI in the form of social media accounts and bots as agents to propagate disinformation and misinformation online. Although AI is the one committing the disinformation and misinformation, the act being within the authority of the principal means the criminal liability ultimately falls upon the individual who used the AI.

Some authors argue that the agent has to be a legal person.[261] When it comes to imposing criminal liability or recognizing constitutional rights, the agents must be entities that can be called to account for their actions.[262] Others say, however, that it is "immaterial whether or not the agent is legally competent to commit the act." What is important is that the two parties are in an agent-principal relationship.[263]

### B. Exploring Possible Frameworks for AI

AIs "have no soul to damn and no body to kick."[264] The problem with the aforementioned frameworks is that they only consider human acts as a ground for punishment. Even if the AI committed the crime, the *actus reus* is still ascribed to a human. These traditional frameworks fail to consider the situation where the human actor has no control over the inner workings of the AI.[265]

If traditional heuristics such as intent may not work with Al, then a new set of factual heuristics is needed for liability: tailored to machine-learning models, not to human beings.[266]

The first factor to consider is whether or not the AI caused the harm or injury. According to Courtney K. Meyer, this serves as a liability shield for AI, or emerging technologies in general, although not intended to completely

---

[260] COMELEC Res. No. 11064, art. V, §1(1). Guidelines on the Use of Social Media, Artificial Intelligence, and Internet Technology for Digital Election Campaign and the Prohibition and Punishment of its Misuse for Disinformation and Misinformation in connection with the 2025 National and Local Elections and the BARMM Parliamentary Elections.

[261] Brown, *supra* note 33, at 214.

[262] Hildebrandt, *supra* note 163, at 10.

[263] Jhudele, *supra* note 193, at 8.

[264] Calo, *supra* note 165, at 554.

[265] Rajpurohit & Seal, *supra* note 29, at 94.

[266] Bathaee, *supra* note 15, at 153.

excuse manufacturers from liability. AI caused the harm or injury if a car's self-driving feature malfunctions, but not if the airbag malfunctions, as the latter has no implication with AI technology.[267] Regardless, the car manufacturer remains liable.

Once it is determined that AI caused the harm, the next issue to be determined is whether the same was foreseeable by the manufacturer or user. If it were, the courts expect that the defendant has taken the necessary precautions. Another issue is as to the comparative level of safety and precautionary actions taken by the manufacturer or user to prevent these foreseeable dangers. Manufacturers are expected to impose basic restrictions into the AI, and users are also expected to take adequate precautions, similar to how, in Philippine law, a plaintiff cannot recover damages if the proximate cause of the injury was their own negligence; and can only recover mitigated damages if the plaintiff contributed any negligence.[268]

In case the risk was highly unforeseeable, the other factors must be weighed more heavily than whether or not safety precautions were undertaken. The next factor to consider is the length of time spent by the manufacturer and user controlling the AI. While the time factor varies depending on the type of AI technology involved—since some AIs might require months of training, whereas others may only need hours—the court must account for situations where the user has exerted more influence over the AI than the manufacturer, or where the harm occurred only after the user purchased the machine. Lastly, courts must also take into consideration the various influences over the AI, which resulted in the AI behaving in different manners, and the context where the AI was used. Each industry has its own custom and standard of care, so this must be carefully compared with the harm caused by the AI.[269] This so-called balancing test is simplified in Figure 2.

---

[267] Courtney K. Meyer, *Exculpatory Clauses and Artificial Intelligence*, 51 STETSON L. REV. 259, 280 (2022).

[268] *See* CIVIL CODE, art. 2179

[269] Shrestha, *supra* note 35, at 403-405.

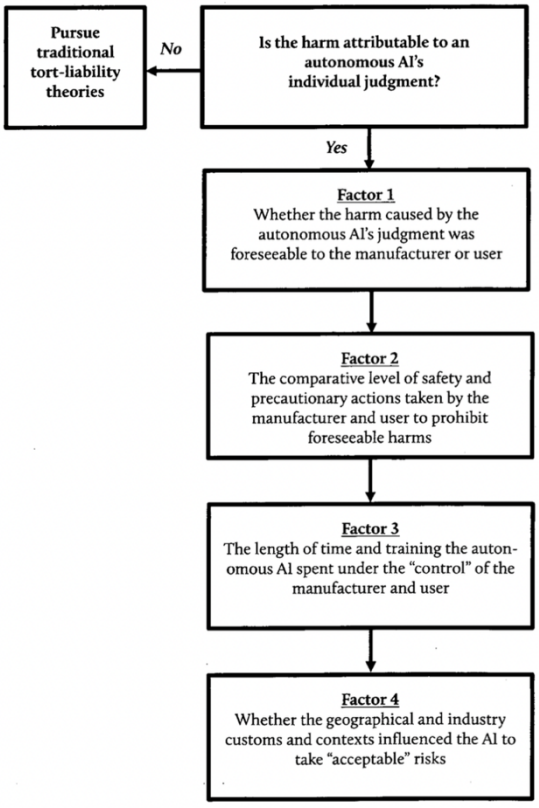


**Figure 2.**[270]

In principle, the liability of the defendant must be directly proportional to the level of instruction given to the AI[271] and the defendant's degree of control over the same, such that the closer the defendant is, the easier it is to argue for his responsibility for the harm caused.[272] This is similar in nature to the rule in Philippine law on quasi-delicts: that obligations resulting from damage caused to another are not only demandable to the person's act, but for the acts of those for whom they are responsible.[273]

Conversely, for autonomous AI, its liability can be imposed directly as a legal person.

---

[270] Shrestha, *supra* note 35, at 402.
[271] Delvaux, *supra* note 191, at 249.
[272] Rajpurohit & Seal, *supra* note 29, at 94.
[273] CIVIL CODE, art. 2180.

*1. AI as a Legal Person – Directly Liable*

In the same way that the law punishes individuals with imprisonment or the death sentence, or private enterprises such as corporations with the revocation of its certificate of registration, AI persons may be punished by the deletion of its software,[274] which is what gives AI persons their "life." An AI may also be put out of use for a certain period of time,[275] which is tantamount to its incarceration. It can also be made to suffer community service,[276] contributing labor to the community. Similar to probation, AI systems can be kept under close supervision. The courts may order a tweak in the code or a certain standard upon the AI's autonomy.[277]

Opponents of this approach argue that the main purpose of punishment is to educate, and punishing a computer program does not serve this function. Any kind of punishment on AI does not have any educative effect, whereas punishment on erring members of society communicates the message that being part of society carries with it certain rights and obligations that cannot be shirked.[278] The separate personality of the AIs must also not be used as a shield by erring manufacturers, developers, and users for their own wrongdoing.[279]

In addition to *ex post* legislation, regulation can also be done *ex ante*. For the following models of liability, recognition of the legal personality of AIs is not even necessary, unlike in the previous framework.

*2. Regulation of Inputs*

The main inputs of AI research and development can be broadly categorized into three: computing power, human expertise, and data. It seems more prudent to regulate AI through what goes into the algorithm than to regulate it as a whole.[280] The regulation of data, in particular, is most significant. The same algorithm can perform differently depending on the

---

[274] Hallevy, *supra* note 31, at 196. *See also* Ankit Kumar Padhy & Amit Kumar Padhy, *Criminal Liability of the Artificial Intelligence Entities*, 8 NIRMA U. L.J. 15, 20 (2019).

[275] *Id.* at 197. *See also* Padhy & Padhy, *supra* note 284, at 20.

[276] *Id.* at 198. *See also* Padhy & Padhy, *supra* note 284, at 20.

[277] Jhudele, *supra* note 193, at 21.

[278] Solum, *supra* note 185, at 1247.

[279] Jhudele, *supra* note 193, at 16. *See also* Sergio M. C. Avila Negri, *Robot as Legal Person: Electronic Personhood in Robotics and Artificial Intelligence*. 8 FRONT. ROBOT. AI 1, 7 (2021).

[280] Solow-Niederman, *supra* note 3, at 673.

input data, and a model is only as good as its data. Inserting poisoned data into the system can cause the same to malfunction.[281] Garbage in, garbage out.

To this end, a legislator has proposed the creation of a Big Data Center, among the functions of which is making available the best possible data for analysis.[282] It aims to collaborate with mobile and internet companies to establish data sharing agreements.[283] Said data, once anonymized, becomes part of the public dominion,[284] and must be published and made available for download[285] by the Big Data Center. At the same time, the collection, storage, and processing of the data must also be regulated to avoid privacy concerns. Pursuant to the Data Privacy Act of 2012, collection must be for a specified and legitimate purpose, processing must be fair and lawful, and retention must only be for as long as necessary for the purpose for which the data was acquired.[286] Consent of the data subject is generally required, unless the processing falls under certain exceptions according to law.[287]

*3. Comprehensive Insurance Scheme*

Another way to hold AI liable is by developing a comprehensive insurance scheme. According to the deep pocket theory, "a person engaged in dangerous activities that are profitable and useful to society should compensate for damage caused to the society from the profit gained." The one with the "deep pocket" is required to have insurance in exchange for his engagement in hazardous activities.[288] The comprehensive insurance scheme for AIs is similar to that required of cars. Motor vehicle insurance is limited, however, in that it only covers human acts and failures. The insurance scheme for AIs must cover all potential responsibilities in the chain.[289] And since not all legal liabilities can be met,[290] there must also be an established

---

[281] Magnuson, *supra* note 45, at 365.

[282] H. No. 269, 19th Cong., 1st Sess., Explanatory Note (2022). *See also* S. No. 2214, 16th Cong., 1st Sess. (2014).

[283] H. No. 269, 19th Cong., 1st Sess., § 11(b) (2022).

[284] § 18.

[285] § 10(c).

[286] § 11.

[287] §§ 12–13.

[288] Jomon P. Jose, *Legal liability issues and regulation of Artificial Intelligence (AI)* 74–75 (Unpublished Dissertation Work for Post Graduate Diploma in Cyber Laws and Cyber Forensics, Natl. Law School of Indi Univ. Bengaluru, 2018).

[289] Delvaux, *supra* note 191, at 249–50.

[290] Solum, *supra* note 185, at 1245.

compensation fund that would ensure reparation for damage not covered by insurance.[291] To inform anyone interacting with the AI about its fund, the limits of its liability, and other relevant details, AIs may be required to always have a visible registration number.[292]

Overall, these frameworks are not mutually exclusive and can concurrently exist. Ultimately, the purpose of these models of liability is to hold responsible according to their level of participation the ones guilty of the acts or omissions that caused the harm,[293] regardless of their legal personality.

## V. CONCLUSION

As Vladeck argues, had robots remained in their state of complete dependence on their human users and programmers, there would be no need to reexamine present laws on liability.[294] There would be no need to establish legal personhood for AI, as well. But with the influx of innovations such as driverless cars, unmanned aircrafts, and humanoid robots, legislators and decision-makers now have to take into consideration whether existing laws provide adequate safeguards for the liability of these thinking machines. The law "tends to assume a dichotomy between individuals and tools," [295] but this dichotomy is upset as it becomes more difficult to distinguish one from the other. "Where people have difficulty categorizing something as being more object- or person-like, the law may similarly struggle."[296]

It is admitted that this perspective proposes a radical change. Recognition of AI systems as legal persons is not going to happen overnight. Courts are slow to embrace change, and usually do so if only to accommodate more inclusive interpretations of the law.[297] Acceptance will develop as a process and in phases. As AIs proliferate, jurisprudence will be replete with them. They may first be recognized as mere tools, which is the

---

[291] Delvaux, *supra* note 191, at 250.

[292] *Id.*

[293] Jhudele, *supra* note 193, at 16.

[294] David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 120 (2014).

[295] Calo, *supra* note 165, at 546.

[296] *Id.* at 553.

[297] Matter of Nonhuman Rights Project Inc. v Stanley, 2015 N.Y. Misc. LEXIS 2785 (N.Y. Sup. Ct. July 29, 2015).

common conception today as seen in existing jurisprudence. Eventually, they will be seen as agents, and thereby recognized under agency law.[298] The question of legal personhood will surely arise as AIs become more advanced and autonomous. But first, we must ask: Is there injury? Who should be held liable and up to what extent? What is the appropriate penalty? These are the questions we must face eventually.

Then again, our cars are now self-driving. Robots serve us food in restaurants, clean our houses, and write the books we read. They give us medical and legal advice. Soon, they will teach our children in schools, defend us before courts of law, and perform our surgeries. That future is not something remote or far away—that future is here.

**- o0o -**

[298] Brown, *supra* note 33, at 233.