Ethical and legal responsibility for Artificial Intelligence

Patrick Henz¹

Received: 24 April 2021 / Accepted: 22 July 2021

Published online: 22 September 2021

© The Author(s) 2021 [OPEN](#)

Abstract

"In civilized life, law floats in a sea of ethics", a quote by the former Chief Justice of the United States, Earl Warren. In a democratic society, the constitution defines the country's values, and the laws define the preferred or at least still tolerated behavior, making deviations sanctionable. As the society is in a continuous flow, also based on scientific and technical developments, law always lags behind. Until regulations can catch up, ethics has to lead society. In less democratic societies, the water gets polluted up to poisoned, ethical behavior may be against ruling law. As for the latter, Robin Hood was a thief, but for most parts of the population, a hero. The less transparent the water, the more difficult to adapt law to new developments. This includes direct corruption, but also unfaithful lobbying. This article discusses the "nature" of Artificial Intelligence, including the risks its posing, and who is responsible for systematic errors, from a moral, but also legal point.

1 Legal responsibility

Most presentations and discussions about Artificial Intelligence ask the listeners to not imagine the technology as the pop cultural well-known Terminator. For our cause, let us make an exception and imagine the famous "T-800", as interpreted by Arnold Schwarzenegger. Originally created to kill all humans, but as the resistance could kidnap one of these machines (in the second movie), they programmed the new task to protect a particular human. Doing so, the machine could learn and adapt to different situations, but nevertheless stayed true to its general programming [1].

A circus tiger can learn tricks from the tamer, but nevertheless stays a predator. If such a creature would hurt or kill the tamer, it is not to blame, as it is part of its nature. Accordingly, if a T-800 hurts or kills a human, the machine is not to blame, as it is following its basic programming, similar to animal instincts. We can replace the tiger with an elephant, no predator and furthermore, based on today's science, one of the few animals with self-awareness. Bad treatment can lead to aggressive behavior, up to that elephants would kill humans. Again, the animal is not to blame, and self-awareness is no gamechanger to determine responsibility.

So far, law confirms this understanding. The British photographer David Slater placed a camera installed on a tripod deep in Indonesia's North Sulawesi national park. A curious macaque monkey had been attracted by the light reflecting lens and instinctively pressed the camera's shutter button, shooting a series of selfies. The U.S. Copyright Office defines in article 306: *"The U.S. Copyright Office will register an original work of authorship, provided that the was created by a human being"* [2]. In this particular case, the regulation stays grey, as a court would have to decide if the photo was really just done by the monkey, or the photographer's experience in nature and according set-up of the camera was relevant to create the image, especially as Slater was nearby. The animal rights organization PETA (People for the Ethical Treatments of Animals) sued Slater for commercially using this photo, claiming that the monkey should be the copyright holder. Similar arguments had been used by Wikipedia, who concluded that based on the regulation, the photo would be without

✉ Patrick Henz, cerpheus27@gmail.com | ¹Governance, Risk and Compliance, Atlanta, USA.

copyright holder (public domain). PETA and Slater came to a settlement, where the photographer agreed to donate 25% of the future revenues of the photo usage [3]. Due to this, it never came to a court decision, which would have been an example to be followed by potential future cases, including the creation of art by an Artificial Intelligence. Art is the result of the artist's skills and decision-making. The ownership depends on the skills, which animals and AI possess, but also the responsibility for own decisions, what is defined as not given for animals and AI.

Wikipedia still displays the photo as part of the article "Monkey selfie copyright dispute" [4], citing various legal experts as University of Michigan law professor Jessica Litman: "*No human author has right to a photograph taken by a monkey. ... The original monkey selfie is in the public domain*" [5]. Due to missing funds, Slater was unable to legally challenge Wikipedia [6].

Another lawsuit causally related to AI is the fatal accident including Uber's test of autonomous automobiles. The car's sensors failed to correctly identify and classify a pedestrian, which was crossing the street pushing a bike, and as consequence, the algorithm not initiated to brake. The human co-driver, who had the task to supervise the car, also not reacted in a timely manner, so could not prevent the accident. The court indicated the co-driver for criminal negligence, as she was distracted looking at her phone as the accident occurred. In opposite to this, Uber did not face any charges. This based on the understanding that today no autonomous self-driving technology ("Level 5 Full Autonomy" as defined by the Society of Automotive Engineers [7]) exists, so always the human must be ready to take over at every given moment [8].

2 System-thinking

Leaving Law, let us see what a system-thinker like W. Edwards Deming can teach us. He understood an organization (a company, but we can extend this also to society) as one interlinked system. Connecting humans with other humans, but also with values, information and knowledge [9]. As described in his model of the "System of Profound Knowledge", an effective and efficient system requires four pillars: Appreciation for a system, Knowledge about variation, Theory of Knowledge, and Psychology [10].

If the system is out of its equilibrium, like one pillar not adequately addressed, Deming concluded that "*a bad system will beat a good person every time*" [11]. An example for this is the effect "overtrust in robots". An experiment at the Georgia Institute of Technology confirmed that depending the situation, here an emergency evacuation, humans may prefer to follow a robot, even if it is recognizable going in circles, instead to use their own knowledge of the building and leave as fast as possible [12]. The participants perceived the machine as knowable and officially authorized to lead in the emergency situation (comparable to the famous Milgram experiment [13]). Furthermore, potential higher stress levels blocked the individuals' ability to understand that the machine was lacking capability. Cognitive biases lead to human vulnerabilities in the cooperation with AI. Companies have a moral responsibility when using such technologies with employees or customers. Future court decisions must be observed, as judges may include the appliance of system-thinking and behavioral science into their decisions.

3 Behavioral science

Humans perceive Artificial Intelligence on the rise, as they have to witness a continuous series of defeats:

- 1977: Chess
- 2016: Go
- 2017: Poker
- 2018: StarCraft II
- 2019: Quake III

While the first two are boardgames with a relative limited number of possibilities, is Poker a game of luck, but also strategy and psychology. The benefit of the victory does not go to the algorithm, but its owners, as IBM could market the victory of its Big Blue against the reigning Chess world-champion Garry Kasparov. Or some casinos in the Chinese territory of Macau implemented digitally enabled poker chips and baccarat tables, in addition with hidden cameras and facial recognition technology to observe the human players and predict where they may lose their money [14]. This with

the goal to lead them consciously or subconsciously to these places. A potential trap, this time with higher costs than just a selfie.

As we perceive AI being on a higher level, similar to the effect of learned helplessness [15], humans tend to accept the recommendations of the algorithm, without properly questioning them. Monotonous situations like being co-driver for hours, amplifying such effects. But we cannot reduce this to perceived helplessness and giving up. A study conducted in 2019 by Oracle and Future Workplace concluded that “64% of people trust a robot more than their manager.” In particular, the algorithms are perceived as superior in providing unbiased information, maintaining work schedules, problem solving and managing a budget [16].

Increased numbers of legislators understand the importance of behavioral science and require organizational systems fostering the human employee [17]. Accordingly, in future, companies could be held legally responsible, if the required human supervisor is not part of a “good system”, in other words, the company’s internal processes (including the defined interaction with intelligent algorithms) dehumanize the employees, so that they are not able developing responsibility or accountability for their tasks. The risk is far from being new, the author Franz Kafka described disorienting and illogically complex bureaucracy in his various novels and short-stories, such like “The Metamorphosis” (1915) and “The Process” (1915). Internal training and efficient processes have to avoid “Kafkaesque” scenarios; less limiting the employees, but defining a safe space for them to act, and so keep them human. This does not mean to exclude technology, in opposite, it should include intelligent solutions, for example an internal knowledge database, a chatbot or nudging solutions. This to reduce the automation bias, as described by Linda J. Skitka:

- Omission Error: The human fails to notice a software failure.
- Commission Error: The human fails to identify an incorrect automated error-message, which wants to replace correct information with incorrect one [18].

4 Ethics

Artificial Intelligence started its conquest to automate further white- and blue-collar jobs. As conclusion, intelligent algorithms must accept internal responsibility (this in addition, as it cannot take-away responsibility from the human). AI can take the place of one set of eyes inside the four-eyes-principle. Human and AI will work as a team. The four-eyes-principle is widely used to reduce internal risks. To ensure this, each part must be competent and independent from each other. Intelligent algorithms are able to fulfill these two requirements.

As law has not defined its boundaries yet, ethics has to support us. Even if highly unlikely, the trolley dilemma, adapted for autonomous vehicles, is an ideal model to discuss the limitation of ethics. In the example, a self-driving car is not capable to avoid a crash, only can decide between two different groups to sacrifice. This could be two externals or also include its own passengers. As it is such a drastic example, it is ideal for discussion, including that the Massachusetts Institute of Technology set up a “Moral Machine” in the web, where users can analyze their own decisions for such scenarios [19].

In the healthcare-sector, an Artificial Intelligence could cause suffering of humans, also on lower levels. What about the decision between five minutes of suffering of four humans against 15 min of suffering of only one person? What about to decide strong suffering of two against light suffering of 50? The Stanford Encyclopedia of Philosophy defines Utilitarianism: “*Though there are many varieties of the view discussed, utilitarianism is generally held to be the view that the morally right action is the action that produces the most good*” [20]. A clear mathematical reason, one the first view perfect for algorithms to solve as a first step towards a potential “theory of everything”. Philosophies as the “effective altruism” combine evidence and reasoning to exactly answer the question, which action produces the most good. Base of such a concept is that everything has a value. The problem already starts with simple questions like, which building has a higher value, the one including a library, a place of faith, or a server room ensuring efficiency of the Cloud? What about the value of human life? Each society will find other answers, including that each subgroup up to individual members will have different opinions. Actual movements as “Black Live Matter” show us that people perceive that their lives have lower values than the ones from others, which is not acceptable. A sophisticated philosophical discussion what produces the most good (or in the trolley example the less evil) is more than a pure counting of individuals. Universal answers may never be found, and surely, we cannot leave such topics to companies and organizations to define, but need bold discussions inside societies. Artificial Intelligence can support with data and information, but only human knowledge and wisdom can decide, including to decide not to decide. Results would be laws and regulations, also known as “rule utilitarianism” (or “rule consequentialism”), where

not the potential action producing the most good gets executed, but the defined rule or process, which predicts to most positive outcome [21].

Immanuel Kant's Categorical Imperative [22] stands in opposite to Utilitarianism, as he defines that a moral decision has to be acceptable by everybody who is involved. Accordingly, a decision, like sacrificing someone in the trolley example is not acceptable. Based on the society's values and philosophers, laws can be interpreted differently. Due to Germany's guidelines for self-driving cars, the machine has the top priority to protect human life. Doing so, for example in the dilemma situation, the system must not consider attributes such as age, gender, physical, or mental constitution. Furthermore, *"parties involved in the generation of mobility risks must not sacrifice non-involved parties"* [23]. This can mean a decision in contradiction of Utilitarianism, as if the autonomous car has to decide to sacrifice one pedestrian or the four passengers inside it, the choice has to be the second. In opposite to a pedestrian, the driver and passengers of a car, independent if autonomous or not, always pose a risk, accordingly, have a responsibility, at least if the manufacturer of the technology or the mentioned pedestrian not acted, unknown to the driver, negligent deviating existing laws.

5 Conclusions

Humans are morally responsible for the AI they create, deploy and/or use. Artificial Intelligence acts based on sensors, data, algorithms and computing power. This underlines the need of adequate maintenance. In the case of the autonomous car, not only cameras and other sensors have to be working on the required levels, but also traditional parts, such as headlights and tires. If we speak about Artificial Intelligence today, we mostly mean Machine Learning. AI acts on information to predict the outcomes of its potential decisions. The person or organization who is using the AI, is responsible that the algorithms get feed by adequate datasets to reduce biased decisions. Not only datasets can be biased, also the algorithm itself. Best protection is to have diverse and inclusive programmer groups, and, of course, regular auditing of the algorithm, including its integration into product or system.

From a legal point, regulations still lack behind, and relevant will be the first court decisions to understand how the judicative sector interprets existing and coming law. Nevertheless, it can be assumed that governments understand AI as tool, which includes the responsibility of ownership, as a request for information by the US banking regulators, including Board of Governors of the Federal Reserve System, Bureau of Consumer Financial Protection, Federal Deposit Insurance Corporation, National Credit Union Administration, and Office of the Comptroller of the Currency, shows: *"With appropriate governance, risk management, and compliance management, financial institutions' use of innovative technologies and techniques, such as those involving AI, has the potential to augment business decision-making, and enhance services available to consumers and businesses"* [24]. Similar had been expressed by the European Parliament in its procedure 2020/2013: *"Autonomous decision-making should not absolve humans from responsibility, and that people must always have ultimate responsibility for decision-making processes so that the human responsible for the decision can be identified"* [25]. The document serves as inspiration to implement adequate legal requirements, as concluded in 2020, active legislation lacks *"of clear requirements and the characterizes of AI technologies ... make it difficult ... for persons having suffered harm to obtain compensation under the current EU and national liability legislations."* Irrespective of the involvement of AI, product liability laws apply, nevertheless without more detailed regulations it is difficult for potential victims of systematic errors to make their case at court [26]. In countries as in the US, companies settled cases related the responsibility of AI before a court decision, making them not suitable yet to function as base for future cases.

In opposite to the human, all AI decisions, even the fastest, are systematic as algorithms are involved. As conclusion, actions lead always to responsibility, even if not to legal consequences (as efficient legal frameworks are not established yet), to moral ones surely. Not by the machine itself, but its creators and users. The question is not, if they had been aware of the potential risks and flawed decision-making, but had they the possibility to understand the risk and detect such.

Authors' contributions This article was written by me, reflecting my perspective. The author read and approved the final manuscript.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Cameron J. Terminator 2: judgment day; 2014.
2. U.S. Copyright Office. Compendium of U.S. Copyright Office Practices, Article 306; 2021. <https://www.copyright.gov/comp3/chap300/ch300-copyrightable-authorship.pdf>. Accessed 31 Aug 2021.
3. Slotkin J. 'Monkey Selfie' Lawsuit ends with settlement between PETA, photographer; 2017. <https://www.npr.org/sections/thetwo-way/2017/09/12/550417823/-animal-rights-advocates-photographer-compromise-over-ownership-of-monkey-selfie>. Accessed 31 Aug 2021.
4. Wikipedia. Monkey selfie copyright dispute; 2021. https://en.wikipedia.org/wiki/Monkey_selfie_copyright_dispute. Accessed 31 Aug 2021.
5. Axelrad J. US government: Monkey selfies ineligible for copyright; 2014. <https://www.csmonitor.com/Technology/Tech-Culture/2014/0822/US-government-Monkey-selfies-ineligible-for-copyright>. Accessed 31 Aug 2021.
6. Wong JC. Monkey selfie photographer says he's broke: 'I'm thinking of dog walking'; 2017. <https://www.theguardian.com/environment/2017/jul/12/monkey-selfie-macaque-copyright-court-david-slater>. Accessed 31 Aug 2021.
7. Society of Automotive Engineers. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles; 2018. https://www.sae.org/standards/content/j3016_201806/. Accessed 31 Aug 2021.
8. Marshall A. Why Wasn't Uber Charged in a Fatal Self-Driving Car Crash?; 2020. <https://www.wired.com/story/why-not-uber-charged-fatal-self-driving-car-crash/>. Accessed 31 Aug 2021.
9. Henz P. The good system—ethics by design; 2020.
10. Deming WE. A system of profound knowledge. In: McConnel J, editor. Safer than a known way. The Deming approach to management; 1991. p. 150–9.
11. Deming WE. at Deming Four Day seminar in Phoenix, Arizona; 1993.
12. Robinette P, Li W, Allen R, Howard A, Wagner AR. Overtrust of robots in evacuation scenarios. 2016. <https://doi.org/10.1109/HRI.2016.7451740>.
13. Milgram S. Behavioral study of obedience. *J Abnorm Soc Psychol*. 1963;67(4):371–8. <https://doi.org/10.1037/h0040525>.
14. Hong J. China's biggest casinos use hidden cameras, facial recognition to spot losers; 2019. <https://www.hindustantimes.com/world-news/china-s-biggest-casinos-use-hidden-cameras-facial-recognition-to-spot-losers/story-7U7CnNkZpU3TMk38XFOJGO.html>. Accessed 31 Aug 2021.
15. Seligman M, Maier S. Failure to escape traumatic shock. *J Exp Psychol*. 1967;74(1):1–9. <https://doi.org/10.1037/h0024514>.
16. Oracle. New study: 64% of people trust a robot more than their manager; 2019. <https://www.oracle.com/corporate/pressrelease/robots-at-work-101519.html>. Accessed 31 Aug 2021.
17. HM Government. United Kingdom anti-corruption strategy 2017–2022; 2017. p. 65. http://data.parliament.uk/DepositedPapers/Files/DEP2018-1255/Anti-Corruption_Strategy_Year1_Update.pdf.
18. Skitka L. Automation Bias; 2021. <https://lskitka.people.uic.edu/styled-7/styled-14/#:~:text=Automation%20bias%20refers%20to%20a,to%20monitor%20on%2Dgoing%20tasks>. Accessed 31 Aug 2021.
19. Moral Machine (fetched 8.4.2021): www.moralmachine.net.
20. Driver J. The history of utilitarianism. *Stanford encyclopedia of philosophy*; 2014. <https://plato.stanford.edu/entries/utilitarianism-history/>. Accessed 31 Aug 2021.
21. Hooker B. Rule consequentialism. In: *Stanford encyclopedia of philosophy*; 2015. <https://plato.stanford.edu/entries/consequentialism-rule/>. Accessed 31 Aug 2021.
22. Kant I. *Grundlegung zur Metaphysik der Sitten*. In: Kant I (1903): "Gesammelte Schriften; 1785. p. 421.
23. German Federal Ministry of Transport and Digital Infrastructure. Ethics Commission Report: automated and connected driving; 2017. p. 7. https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission-automated-and-connected-driving.pdf?__blob=publicationFile. Accessed 9 Sept 2021.
24. Federal Register—The Daily Journal of the United States Government. Request for information and comment on financial institutions' use of artificial intelligence, including machine learning"; vol. 86, No. 60, p. 16839. <https://www.federalregister.gov/documents/2021/03/31/2021-06607/request-for-information-and-comment-on-financial-institutions-use-of-artificial-intelligence>. Accessed 06 Jan 2021.
25. European Parliament Procedure 2020/2013. Artificial intelligence: questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice"; 2021. <https://oeil.secure.europarl.europa.eu/oeil/popups/printficheglobal.pdf?id=710011&l=en>. Accessed 31 Aug 2021.
26. European Commission. White paper—on artificial intelligence—a European approach to excellence and trust; 2020. p. 12. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en. Accessed 31 Aug 2021.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.