December 2020

Human-Machine Teaming and Its Legal and Ethical Implications

Jim Q. Chen

National Defense University, College of Information and Cyberspace, Jim.chen@ndu.edu

Thomas Wingfield

National Defense University, College of Information and Cyberspace, thomas.wingfield@ndu.edu

Follow this and additional works at: <https://digitalcommons.usf.edu/mca>

Part of the [Computational Engineering Commons](#), [Computer Law Commons](#), [Legal Ethics and Professional Responsibility Commons](#), and the [Science and Technology Law Commons](#)

Recommended Citation

Chen, Jim Q. and Wingfield, Thomas (2020) "Human-Machine Teaming and Its Legal and Ethical Implications," *Military Cyber Affairs*: Vol. 4 : Iss. 2 , Article 2.

<https://doi.org/10.5038/2378-0789.4.2.1074>

Available at: <https://digitalcommons.usf.edu/mca/vol4/iss2/2>

This Article is brought to you for free and open access by the Open Access Journals at Digital Commons @ University of South Florida. It has been accepted for inclusion in Military Cyber Affairs by an authorized editor of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

Human-Machine Teaming and Its Legal and Ethical Implications

Erratum

Format changes (various)

Human-Machine Teaming and Its Legal and Ethical Implications

Jim Q. Chen, Ph.D. and Thomas Wingfield, Esq.

Abstract

Humans rely on machines in accomplishing missions while machines need humans to make them more intelligent and more powerful. Neither side can go without the other, especially in complex environments when autonomous mode is initiated. Things are becoming more complicated when law and ethical principles should be applied in these complex environments. One of the solutions is human-machine teaming, as it takes advantage of both the best humans can offer and the best that machines can provide. This article intends to explore ways of implementing law and ethical principles in artificial intelligence (AI) systems using human-machine teaming. It examines the existing approaches, reveals their limitations, and calls for the establishment of accountability and the use of a checks-and-balances framework in AI systems. It also discusses the legal and ethical implications of this solution.

Human-Machine Teaming

The recent development of AI led by humans has made machines more intelligent and powerful. Machines are capable of performing more tasks now than ever before, either in automated mode or in autonomous mode. They are widely utilized in various fields.

By default, machines, or AI-systems, are not equipped with the capability of applying law and ethical principles. How to enable such a capability within AI-systems is obviously a challenge. Without such a capability, AI systems can hardly be trusted. Therefore, this challenge has to be tackled. To this end, various approaches have been proposed and/or utilized in dealing with this challenge. However, the majority of the approaches available lack a checking system that verifies the legal and ethical decision-making.

To establish such a checking system calls for human-machine teaming, which takes advantage of the best that both humans and machines can offer. Machines are good at speedy calculation, fast data analysis, quick pattern recognition, rapid processing, and repetitive work. They can learn much more quickly than humans. They can process a large amount of information simultaneously. They can “calculate, analyze and perform tasks tireless and round the clock”, while not being “affected or influenced by emotions, feelings, wants, needs and other factors that often cloud the judgment and intelligence of us mere mortals” (Whitney, 2017). These capabilities ensure speed and precision. Humans are good at imagination and creativity in addition to resolving ambiguity and exercising judgment. They are superior to machines with respect to the use of instincts, common sense, life experience in performing tasks, making decisions, and solving problems (Whitney, 2017). When humans and machines are teamed together, humans can complement machines with training, explanation, and sustainment while machines can complement humans with amplification, interaction, and embodiment. Amplification refers to extraordinary data-driven insights. Interaction refers to the employment of advanced communication interfaces such as voice-driven natural-language processing devices. Embodiment refers to the integration of AI-systems with sensors, motors, and actuators. Such an integration allows “robots to share workspace with humans and engage in physically collaborative work” (Daugherty and Wilson, 2018).

Taking advantage of what human-machine teaming can offer, this article explores a new way in which law and ethics can be implemented, checked, and applied in decision-making within AI systems. Before this exploration, existing approaches are examined.

Current Approaches in Implementing Law and Ethics in AI Systems

There are various approaches in implementing law and ethical principles in AI systems. Generally speaking, they fall into the categories of the top-down approach, the bottom-up approach, the hybrid approach, and the no-use approach.

The no-use approach argues for the regulation and prohibition of decision-making by autonomous AI systems. This avoids potential legal and ethical issues, reducing the complication of introducing legal and ethical mechanisms within AI systems. This is an approach at the policy level. The representatives of this approach are the Data Protection Act 2018 and the European Union's General Data Protection Act 2018. Under these data protection frameworks, solely automated decision-making is prohibited (Edward and Veale, 2017). Another representative is the European Civil Law Rules in Robotics, which clearly states that a robot's presence should "not dissolve social ties". While acting as an agent helping people, "the robot may not substitute humans entirely" (Nevejans, 2016). Still another representative is Underwood (2017), who argues for a ban on the development, manufacture, and deployment of lethal autonomous weapon systems (LAWS) or such type of machines/systems because these systems lack responsibility and predictability. The lack of responsibility refers to the difficulty in apportioning blame when something goes wrong. The lack of predictability refers to the unpredictability of autonomous systems (Underwood, 2017). Underwood (2017) is correct in pointing out the issue of responsibility. However, she does not look into ways of creating legal and ethical mechanisms and establishing responsibility and accountability within AI systems. Instead, like other representatives in this approach, she counts on a solution at the policy level, i.e. prohibiting the use of autonomous AI systems in decision-making, especially in conflict environment.

Other approaches look into ways of implementing law and ethical principles into AI systems. In general, they consist of the following three approaches, i.e. the top-down approach, the bottom-up approach, and the hybrid approach.

The top-down approach attempts to formulate legal guidance and ethical principles first and apply them to specific cases. This application helps to make legal and ethical decisions and determine legal and ethical courses of action. For instance, it formulates "moral principles like Kant's categorical imperative, the utilitarian principle of maximizing utility, or Asimov's laws of robotics as rules in a software program which is then supposed to derive what has to be morally done in a specific situation" (Misselhorn, 2018). Besides, a maximally-just autonomy using artificial intelligence (MaxAI) is a good example. It requires an ethical governor should provide an assessment on proposed courses of action based on preloaded law and ethical principles such as the laws of war and the rules of engagement (Arkin et al., 2009). It "involves the codification of normative values into rule sets and the interpretation of a wide range of inputs through the application of complex and potentially imperfect machine logic" (Galliot and Scholz, 2018). A minimally-just autonomy using artificial intelligence (MinAI) is another good

example. Having realized the challenges in MaxAI, the designers of MinAI do not use an ethical governor. Instead, they only use a preloaded, hard-coded and narrow set of constraints in identifying and avoiding what is ethically impermissible, such as “lawfully protected symbols and locations, signs of surrender (including beacons), and sites that are hors de combat” (Galliot and Scholz, 2018). Within the top-down approach, the precise interpretation and accurate application of legal guidance and ethical principles in particular cases are always challenges. So is the implementation of the interpretation and application in software development. In other words, within these environments, there are a lot of dynamics. It is hard to capture all the possibilities and then select the most appropriate ones from them. It is even harder to code all these in software and make them work as required and as designed. At present, the approach is mainly at the conceptual level. There is not much implementation of this approach. Besides, this approach faces the challenge of the frame problem (Horgan and Timmons, 2009; McCarthy and Hayes, 1969). Since “potentially every new piece of information may have an impact on the whole cognitive system of an agent”, there are “reasons to doubt that moral normativity is fully systematizable by exceptionless general principles” (Misselhorn, 2018).

The bottom-up approach intends to judge each particular case individually and then derive certain legal and ethical rules of thumb. In software design, it is suggested that artificial neural networks be used to find relationships or patterns that eventually help to formulate legal guidance and ethical principles (Misselhorn, 2018; Dancy, 2013). Besides, there are recommendations for using modeling for the inputs received or the data captured. The modeling ranges from human moral learning process (Froese and Di Paolo, 2010) to human socialization (Breazeal and Scassellati, 2002; Fong et al., 2002). However, in the bottom-up approach, it is difficult to question, interpret, explain, supervise, and control AI systems “because deep-learning systems cannot easily track their own ‘reasoning’” (Ciupa, 2017). Also, as a solution, the bottom-up approach poses “problems of operationalization, safety and acceptance”. Therefore, it is “of limited suitability for implementing moral capacities in autonomous artificial systems” (Misselhorn, 2018). The four reasons for this claim are as follows: First, it is “difficult to evaluate when precisely a system possesses the capacity for moral learning and how it will, in effect, evolve”. Second, there is “no component or mechanism to point to which embodies the system’s moral capacities”. Third, it is hard to make prediction. Fourth, it is difficult to reconstruct the way in which a moral decision is arrived. Therefore, this approach is “hardly suitable for practical purposes” (Misselhorn, 2018).

The hybrid approach seeks to combine the strengths of both the top-down approach and the bottom-up approach. It has a pre-defined framework of legal and moral values. It also uses a specific context to generate the individual legal and moral value profile of a specific user. The values from the two different sources are mapped together for decision-making. The hybrid approach is used in designing an elder care system with moral capacities. The conceptual design consists of four steps. “The first step is to find out the moral values that are relevant in care from the perspective of the people concerned.” Interviews are held “with the aim to gain a *prima facie* list of values that are considered relevant by the potential users” and to find out different weights that people assign to each moral value. The second step is to have an artificial system recognize the moral values and to “weigh them according to the moral value profile of the user”. A user’s preference is thus captured. The third step is to translate the scenarios into algorithms. The fourth step is to update the individual moral value profile of a user through continuous interaction with the user. The system, therefore, “is able to learn what is morally good or bad” via constantly interacting with users. It then “acts for the corresponding moral reasons”. “It even treats persons

according to moral standards that these people endorse” (Misselhorn, 2018).

It is obvious that the hybrid approach has a lot more advantages than other existing approaches discussed above. The unique capability of customizing moral values by using weight does make a difference. However, this approach still faces the same challenges for both the top-down approach and the bottom-up approach.

In general, there are several issues that all approaches discussed above share. First, all approaches lack a checking and balancing mechanism for decision-making. Second, in all these approaches, it is not clear if humans should be completely out of the loop in legal and ethical decision-making. Third, no roles and responsibilities are assigned in the process of legal and ethical decision-making. Consequently, it is not clear whether humans or machines should be held liable if bad decisions are made. Fourth, it is not clear what should be done if there are conflicting views on one matter. It is also not clear whether humans or machines should make the final decision. Fifth, it is not clear how the weight of a legal and ethical value is determined. Sixth, it is not clear how errors can be detected and corrected as there is no checking mechanism.

To summarize, there are different types of limitations in all the approaches discussed above. One type of limitations is associated with ways of applying abstract law and ethical principles to specific cases or ways of deriving legal guidance and ethical principles from specific cases. This type of limitations is technically related. The solution depends upon the further development of artificial general intelligence (AGI). Another type of limitations is associated with incomplete architecture of solutions. In other words, some important components are missing in all these approaches, such as the lack of roles and responsibilities as well as the lack of checks-and-balances mechanism. Without those, it is hard to know when humans should be responsible for decision-making and when machines should be responsible. Likewise, it is also difficult to determine who should be held accountable for decision-making. This article is only focused on this particular type of limitations. The solution provided below only addresses the limitations in this respect. The technical solution that requires the support of AGI will be explored in a separate article.

Establishment of Accountability and Checks-and-Balances

To hold relevant parties accountable, roles and responsibilities have to be assigned to them ahead of time and their performance needs to be constantly evaluated. Meanwhile, to prevent biases or unexpected consequences and to balance power in legal and ethical decision-making, checks-and-balances should be established.

The establishment of accountability and the establishment of checks-and-balances call for human-machine teaming. The reasons are as follows:

First, it is hard or almost impossible to hold machines liable for making illegal and unethical decisions. Humans are normally held liable instead. In this case, the people who may be held accountable consist of the person who designed the algorithm, the person who coded the program, the person who gave the initial instruction to the relevant AI system, and the person who had control over the relevant AI system. Without human involvement, there is no accountability per se in AI systems, including autonomous AI systems.

Second, the establishment of checks-and-balances requires involvement of more than one party. It is possible to have two AI systems for the sake of checks-and-balances. One system makes decisions while the other system performs checks-and-balances. However, as all AI

systems are made to help humans to achieve humans' goals, humans' judgment should have high priority. Besides, humans' instincts, common sense, and life experience, imagination, and creativity can provide tremendous guidance for AI systems. Therefore, humans' involvement can significantly enhance this process. It is for these reasons that human-machine teaming is the best solution in dealing with the challenges.

A checks-and-balances framework empowered by human-machine teaming is thus called for. It should be capable of establishing accountability and providing checks-and-balances. This framework consists of three components: the law/ethics/rule-making component, the judicial component, and the execution component. Both actors, i.e. humans and machines, are included in these three components in the form of human-machine teams. Their responsibilities are assigned so that accountability in each component is established. It also needs to be pointed out that the leadership role in these components varies.

In the law/ethics/rule-making component, humans are chosen as leaders and they have been assigned with the responsibility of making or selecting relevant law and relevant ethical values as well as creating rules. Machines are assigned with the responsibility of assistants in improving efficiency within this component. Humans are held liable if something goes wrong.

In the judicial component, humans are also chosen as leaders. Humans in this component are legal experts, moral experts, and subject-matter-experts in relevant fields. They verify and interpret the law made, the ethical values selected, and the rules created for AI systems. Some machines are employed as assistants in quickly locating or finding relevant legal and ethical documents, references, and cases. In addition, the humans in this component render a judgment if there are any conflicting views on one matter in any component. Needless to say, humans in this component are held accountable for decisions that they make.

In the execution component, machines are chosen as leaders while humans are assigned with the responsibility of assistants. Machines are in charge of the execution on most occasions. Humans only get involved whenever needed. Even machines are in charge, human managers, coders, implementers, and operators are still held accountable if something goes wrong. This is because machines are doing all these jobs in helping humans to accomplish humans' missions, not machines' missions.

This framework allows multiple levels of checks-and-balances to be conducted. In the execution component, either humans or machines can verify then approve or block the decisions made by the other party. If there are conflicting views on one matter, they are sent to the judicial component, where the experts from different fields are called in to make a judgment. Final decisions are then made based on the judgment. Moreover, the law cited, the moral values selected, and the rules created within the law/ethics/rule-making component can be further checked and verified by the experts from different fields in the judicial component. This checking and verification mechanism also makes it possible to detect then correct errors should there be any. All these guarantee accuracy and reliability in legal and ethical decision-making.

Obviously, this checks-and-balances framework is able to address the issue of lacking accountability and the issue of lacking checks-and-balances. It is also able to support speed, completeness, accuracy, reliability, and dynamics simultaneously. As a result, it is able to ensure legal and ethical decisions should be made.

This new approach successfully addresses the issues and challenges that other approaches are facing. It has several strengths.

With the employment of the checks-and-balances mechanism, this approach successfully establishes accountability by assigning responsibility ahead of time. This makes it certain that legal and ethical decisions be made since the people who do not follow legal guidance and ethical principles are held liable. Besides, this approach makes sure that all decisions made should be further checked and verified within the same component and/or in another component. This can avoid making some obvious mistakes and help to reduce some potential bias should there be any. This ensures the legal guidance and ethics principle be faithfully followed. The mechanism for balances is extremely useful under some circumstances. If a human accidentally or intentionally makes a mistake, the machine within the law/ethics/rule-making component, upon detection, can instantly block the decision from being undertaken and call for immediate intervention and investigation from the judicial component. Likewise, if a machine makes a mistake, a human within the same component can also block the decision and call for immediate intervention and investigation from the judicial component. As a consequence, this mechanism can effectively prevent the worst scenario from happening in some cases. It is also able to handle conflicting views on one matter by calling in the experts in the judicial component to solve the dispute. In addition, this mechanism effectively handles error detection and error correction.

This approach is built on the belief that AI systems are made by people and for people. In other words, AI systems are built and used as a means for helping humans to reach their ends or enhance their quality of life. Therefore, humans should be involved in critical decision-making; even the level of involvement may be varied depending upon contexts. As a result, human-machine teaming should be employed in AI systems to ensure fast and accurate decisions-making following legal guidance and ethical principles.

Conclusion

Human-machine teaming takes advantage of both the best humans can offer and the best that machines can provide. It has a great potential in AI systems. This article clearly shows its importance in AI systems, especially in implementing legal guidance and ethical principles in AI systems. Most approaches available do not employ human-machine teaming. As a consequence, they fail to establish accountability and fail to provide checks-and-balances in legal and ethical decision-making. The approach recommended in this article promotes human-machine teaming. It uses human-machine teaming as a foundation for a checks-and-balances framework, which assigns responsibilities to both humans and machines, establishes accountability for humans, and provides checks-and-balances for legal and ethical decision-making within AI systems, including autonomous AI systems. There are three components within the framework: the law/ethics/rule-making component, the judicial component, and the execution component. Each component plays the role of checks-and-balances, thus making this function resilient. Ultimately, this framework, empowered by human-machine teaming, can successfully ensure lawful and ethical decision-making in AI systems.

References

- Arkin, R., Ulam, P., and Duncan, B. (2009). An ethical governor for constraining lethal action in an autonomous system, technical report. Fort Belvoir, VA: Defense Technical Information Center. DOI:10.21236/ADA493563.
- Breazeal, C. and Scassellati, B. (2002). Robots that imitate humans, *Trends in Cognitive Sciences*, 6, pp.481–487.
- Ciupa, M. (2017). Is AI in jeopardy? The need to under promise and over deliver - the case for really useful machine learning, *Computer Science and Information Technology (CS & IT)*, Fourth International Conference on Computer Science and Information Technology, Academy & Industry Research Collaboration Center (AIRCC), pp.59-70. DOI:10.5121/csit.2017.70407.
- Dancy, J. (2013). Moral particularism. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2013 Edition. Retrieved from <http://plato.stanford.edu/archives/fall2013/entries/moral-particularism/>.
- Daugherty, P. and Wilson, H. (2018). *Human + Machine: Reimagining Work in the Age of AI*. Boston, Massachusetts, U.S.A.: Harvard Business Review Press.
- Edward, L. And Veale, M. (2017). Slave to the algorithm? Why a ‘right to an explanation’ is probably not the remedy you are looking for, *Duke Law & Technology Review*, 16(1), pp.18-84.
- Froese, T. and Di Paolo, E. (2010). Modelling social interaction as perceptual crossing: an investigation into the dynamics of the interaction process, *Connection Science*, 22, pp.43– 68.
- Fong, T., Illah, N., and Dautenhahn, K. (2002). A survey of socially interactive robots: concepts, design, and applications, *Technical Report CMU-RI-TR*, pp.2–29.
- Horgan, T. and Timmons, M. (2009). What does the frame problem tell us about moral normativity? *Ethical Theory and Moral Practice*, 12, pp.25–51.
- Galliot, J. and Scholz, J. (2018). Artificial intelligence in weapons: the moral imperative for minimally-just autonomy, *The Air Force Journal of Indo-Pacific Affairs*, 1(2), pp.57-67.
- McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer, & D. Michie (eds.), *Machine Intelligence*, pp.463–502. Edinburgh: Edinburgh University Press.
- Misselhorn, C. (2018). Artificial morality: concepts, issues and challenges, *Society*, 55(2), pp.161-169.

Nevejans, N. (2016). *European Union Policy Department C: Citizens' Rights and Constitutional Affairs: European Civil Law Rules in Robotics*.

Underwood, S. (2017). Potential and peril: the outlook for artificial intelligence-based autonomous weapons, *Communications of the ACM*, 60(6), pp.17-19.
DOI:10.1145/3077231.

Whitney, L. (2017). Are computers already smarter than humans, *Time*, 29 September 2017.