NOTE FOR NATIONAL DEFENCE:

Challenges of Public Policy in Autonomous Systems

Authors: Shahram Shahkar[1] and Kash Khorasani [2]

[1] Graduate student, Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada

[2] Professor, Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada

SUMMARY

- In every autonomous system one needs to consider and incorporate a balance between accuracy and safety, performance, ethics, and cost.
- The most important challenge in successful application of autonomous systems is their public acceptance. This requires involvement of public and military stakeholder's opinion at the development stage.
- Public policy directives and guidelines are needed to decide and determine the extent to which ethics and cost have to be compromised in order to achieve safety, reliability, assurance, and performance.
- As human involvement and command diminishes in autonomous systems there needs to be specific form of ethics to replace human conscience in the decision-making process of computer-based systems.
- The most important challenge in successful application of autonomous systems is explainability of decisions and how decisions have been made.
- Explainability of decisions generally contradicts autonomy given that in the first place the main motivation is to have computers undertake decisions that are complex and unexplainable by humans.
- Given that ethics are difficult to formulate in autonomous systems is it possible to replace ethics and morality by social norms and values. If this is indeed a desirable option and approach, then how does this help the public acceptance?

CONTEXT

- Autonomy is strategically important in the future development of our civilization, for optimizing and equally sharing our existing resources and infrastructure [2].
- Accuracy and Safety in Autonomy are generally in competition with optimal use of resources and infrastructure. In other words, increasing accuracy or safety inevitably compromises efficient consumption of energy and available resources and infrastructure.

- Ethics in Autonomy is also important, since human oversight is removed or minimized. In its compensation, an artificial form of ethics has to be substituted in order to maintain moral standards in our civilization [3].
- As human oversight gradually diminishes, machine decision making processes have to be understood and accepted among the users. Therefore, public opinion has to be taken into consideration throughout the development stage and process of these emerging technologies [1].
- It is challenging to formulate ethics in computer systems since computer operations are based on numeric quantities rather than qualitative parameters, and in order to follow ethical standards all alternative options in a decision have to be quantized and then compared in order to deduce the most preferable choice.
- Social norms are more feasible to be implemented in computer systems since computers can learn repeated behaving patterns and replicate these patterns under similar and given circumstances.
- Social norms are also more readily and well explainable and favour public acceptance. Therefore, autonomous systems that follow norms are expected to be more readily acceptable and approved in societies.
- Public policy needs to address these challenges in the form of civil law and/or technical codes and standards.

CONSIDERATIONS

- Development of autonomous systems should have a clear and specific justification and goal in every area, rather than a tendency based on herd mentality or global trends. Application of autonomy simply because it has been successfully implemented in a given country does not necessarily guarantee that it would also result satisfactorily in Canada.
- Autonomy inherently involves probabilistic methods and models. Hence, it is impossible to assure safety and accuracy in a definite manner due to the probabilistic nature of autonomy. It is possible to increase safety in an autonomous system, but it invariably entails certain compromises. The required level of safety and ethics in autonomous systems have to be decided in view of this fact.
- Responsibility in autonomous systems is another important challenge that needs special consideration by the public policy decision makers. It would be unwise, at the current time, to let robots assume full responsibility of their decisions. On the other hand, system designers should be held accountable and responsible despite the fact that decision making in autonomous systems is a very complicated process and often beyond apprehension of human operators. Therefore, it is crucial to determine to what extent one should hold designers responsible for acts that are beyond their control and knowledge. This remains a consideration for the public policy and decision makers.
- Complete validation and verification of autonomous systems is also quite complex and very challenging. Simulation of every possible scenario under which autonomous systems are expected to function is a challenging endeavor. Consequently, one has to prevent the possibility of encountering an output from a robot that has not been foreseen and programmed before.

ARTIFICIAL ETHICS

Diminishing human oversight resulting from future autonomous systems requires some form of artificial guard against manifestation of immoral behaviors. Researchers are beginning to consider implementing artificial ethics in autonomous systems. Philosophical and historical approaches pertaining to human ethics addressed the following main doctrines [3], [4], [5].

- **Consequentialism:** the normative properties of an act depend only on the con- sequences of that act. Thus, whether an act is considered morally right can be determined by examining the consequences of that act.
- **Deontological Ethics:** the moral worth of an action is judged by its conformity to a set of rules, irrespective of its consequences. Specifically, "the concept that one must act only according to that precept which he or she would will to become a universal law, so that the rules themselves are grounded in reason alone" [5].
- **Virtue Ethics:** morally good actions will exemplify virtues and morally bad actions will exemplify vices. Traditional virtue ethics are based on the notion of Eudaimonia, translated as happiness or flourishing.

The main idea here is to translate the ethical concepts into computer codes and algorithms. Consequently, possible decisions of an autonomous system could be ranked based on numerical grades, and the option entitling the highest rank would then be taken as the final decision [3].

- It should further be investigated to determine if social norms are equivalent to ethics. Only then one could substitute ethics by norms in autonomous systems.
- Social norms are behaviors that are culturally approved and/or widely accepted. Morals on the other hand can be understood as distinguishing criteria for a behavior to be classified into as right or wrong regardless of whether or not said behavior is culturally approved or widely accepted. Ethics then is the systematic approach one uses in deciding on a behavior [6], [7].
- Values on other hand comprise of ideas that are preferred. In other words, what is good, right, wise or beneficial. According to [10] "Values are implanted early in a person's life and once they are fixed, serve as a guide in choosing behaviour and in forming attitudes". "Values account for the stability of social order, and they provide the general guidelines for social conduct".
- Therefore, values can be considered as standards of social behaviour derived from social interactions and accepted as constituent facts of social structure. In this sense, it could be possible to consider them equivalent to ethics in certain context [8], [9], [10].

RELATIONSHIP BETWEEN NORMS AND VALUES:

- Norms and values have salient relationships. Norms are specific, values are not. There may be, in a particular situation, delusion of norms, but values are commanding [10].
- Norms are rules for behaving. They identify more or less specifically what should or should not be done by particular types of actors under given circumstances.
- Values are standard of desirability that are more nearly independent of specific situations. The same value may be a point of reference for a great many specific norms; a particular norm may represent the simultaneous application of several separable values.

RECOMMENDATIONS

- Implementation of artificial ethics is a major challenge that is presently at its very early stages and needs significantly more development. Without artificial ethics there would be few replacements for the diminishing human oversight and one would be constrained in maintaining human moral values in the future societies.

- It is important to consider and formulate responsibility of autonomous systems behavior and artificial decisions.
- Involving public opinion in development of autonomous systems is a matter that needs to be considered in early stages. This is only possible through an open dialogue among designers, the regulatory body, and the public.
- Validation and verification of products are challenging issues and require very urgent considerations by the public policy decision makers.

References:

[1] F. Alaieri, A. Vellino, Ethical Decision Making in Robots: Autonomy, Trust and Responsibility Auton- omy Trust and Responsibility, University of Ottawa.
[2] Royal Academy of Engineering, Safety and Ethics of Autonomous Systems, National Engineering Policy, Center, June 2020.
[3] T.J.M. Bench-Capon, Ethical Approaches and Autonomous Systems, Artificial Intelligence 281 103239, 2020.
[4] Colin Allen,Iva Smit, Wendell Wallach, Artificial Morality: Top-Down, Bottom-Up, and Hybrid Approaches. Ethics and Information Technology, 7(3):149-155, 2005.
[5] I. Kant, The Moral Law: Groundwork of the Metaphysics of Morals, first published 1785, Routledge, 2013.
[6] Stanford Encyclopedia of Philosophy, March 2011.
[7] Thomas Schelling, T. C. (1966), The strategy of conflict, New York: Oxford University Press.
[8] Thibaut, J.W. and Kelley, H.H. (1959), The social psychology of groups. John Wiley & Sons, New York.
[9] Ragnar Rommetveit (1955), Social Norms and Roles, Minneapolis: University of Minnesota Press.
[10] Puja Mondal, Essay on Values, Norms and Beliefs, https://www.coursehero.com/file/24476489/Meaning-and-Nature-of-Societydoc/.