BRIEFING
STOA Options Brief

European Parliament

Artificial intelligence: From ethics to policy

There is little doubt that artificial intelligence (AI) and machine learning (ML) will revolutionise public services. While AI holds great power in solving some of the world's most dangerous and complicated problems, the power for positive change that AI provides simultaneously has a potential for negative impacts on society.

Based on a framing of 'AI as a social experiment,' this study arrives at regulatory options for public administrations and governmental organisations who are looking to deploy AI/ML solutions, as well as the private companies who are creating AI/ML solutions for use in the public arena. The reasons for targeting this application sector concern: the need for a high standard of transparency, respect for democratic values, and legitimacy. The policy options presented in the study demand targeted procedural solutions. Together, these chart a path towards accountability in AI; procedures and decisions of an ethical nature are systematically logged prior to the deployment of an AI system. This logging is the first step in allowing ethics to play a formidable role in the implementation of AI for the public good.

Policy options	It is proposed that all AI/ML system developers are required to have a data hygiene certificate (DHC) to be eligible to sell their solutions to government institutions and public administrative bodies.
	It is proposed that all public administrative bodies must purchase AL/ML solutions and/or systems from developers who can produce a DHC.
	It is proposed that all public and government organisations using AI systems should conduct an ethical technology assessment (eTA) prior to deployment of an AI system.
	It is proposed that all public administration institutions and government bodies are required to show clear goals for the AI/ML application as part of the eTA.
	It is proposed that all organisations deploying AI systems should produce an 'accountability report' in response to the eTA.
Procedural options	Develop criteria for auditing data provenance.
	Establish a systematic process for granting a DHC.
	Develop standards for the procedure and criteria of an eTA.
	Develop standards for the creation and criteria of an 'Accountability report'.
	Develop governance mechanisms to ensure the necessary competency to fulfill the eTA and 'accountability report'.

Background

The way in which AI/ML progress on a global, national, or international scale is dependent upon the vision put in place by policy-makers, academics, industry leaders, public administration organisations, consumer rights organisations and the like. This study is intended to show a vision of a future world that conceptualises AI as a real world experiment and thus requires that it meet the conditions of an experiment, i.e. that it only be conducted when: 1) appropriate ethical constraints are in place to protect citizens; 2) that the experiment is aimed at assessing a predicted amount of good to be achieved by the AI/ML system; and 3) that any (acceptable) risks are appropriately balanced against the assured benefits for users/society.

Although there has been an increase in attention to, and focus on, ethics in the AI debate, there is still little done to show what ethics means for the creation of policy and regulation of AI beyond the development of guidelines and/or principles. Thus, the question at the centre of this study was to ask: **how can we move from AI ethics to specific policy and legislation for governing AI.**

Ethics provides a variety of conceptual tools for understanding how to evaluate actions and people. Some of these ethical tools may be partially translated into technical, procedural or governance solutions whereas other ethical concepts (e.g. dealing with moral overload) cannot. What ethics as a study provides us with is the capacity and tools for deliberation about the kinds of people we want to be, the kinds of communities we want to build, and the kinds of technologies we want to create and use.

In recent decades, academics have uncovered a range of ethical issues pertaining to AI. Some of these issues relate to how AI/ML algorithms are made, e.g. how the data is acquired, sourced, and labelled; the computing power required to train an algorithm; the asymmetry in power, and the lack of transparency between the private companies who have both the data and the computing power, and the consumer, who is reliant on private companies for their services. Relatedly, ethical issues result from how the AI/ML algorithm is applied in society, for example: differential impact in society seen through an unequal distribution of risks and benefits between groups; the potential for consumers to be unknowingly nudged to act in certain ways; lack of opportunity for meaningful, explicit informed consent; and the threat to constitutional democracy if AI/ML applications influence political power and the decision-making of citizens.

Although AI engenders unique ethical discussions, the literature on the ethics of technology provides helpful conceptual tools to think about the ethics of AI. The ethics of technology shows us that AI should be understood as a complex confluence of both society **and** technology, rather than society and technology isolated from one another until the moment AI/ML is introduced into the real world. The consequence of this is that AI/ML should be evaluated with reference to the society in which it has been created. Furthermore, the fact that AI/ML is a complex technology demands that the variety of actors involved in AI/ML development, implementation, use, and regulation decide together with users about the accountability-responsibility relations they wish to enforce.

The fact that AI/ML is a complex technology demands that the variety of actors involved in AI/ML development, implementation, use, and regulation decide together with users about the accountability-responsibility relations they wish to enforce.

Given the lack of operational experience we have with AI and the level of uncertainty and risk, it is wise to frame AI as a social experiment and to usher in experimental conditions for the real world applications of AI, namely ethical constraints and learning goals. Ethical technology assessments (eTAs) can be a powerful tool to uncover the qualitative ethical issues of AI at an early stage.

In short, when there is much that is unknown about a technology, it should only be allowed under constrained circumstances, with careful logging of the possible ethical risks prior to deployment, which will give us the information necessary for making more concrete policies and legislation around the technology at hand.

Policy options

The policy options presented in this study are aimed at the public administrations and governmental organisations who are looking to deploy AI/ML solutions, as well as the private companies who are creating AI/ML solutions for use in the public arena. These options center around the practice of logging and its relationship to ensuring accountability. The logging discussed here places ethical considerations relevant to the technology at the centre of the decision-making process. With such logging in place, the creation of systematic procedures for ethical evaluations that is thoroughly documented, becomes achievable. The policy options chart a path towards accountability insofar as procedures and decisions of an ethical nature are logged, transparent, and accessible to the public.

The reasons for targeting this application sector have to do with both the desire to use AI/ML in these spaces, along with the need for this sector to maintain a high standard of transparency, respect for democratic values, and legitimacy. The reasons to legislate are multiple: the criticality of ethical and human rights issues raised by AI/ML development and deployment; the need to protect people (i.e. the principle of proportionality); the interest of the state (given that AI/ML will be used in state governed areas such as prisons, taxes, education, child welfare etc); the need to create a level playing field (e.g. self-regulation is not enough); and the need to develop a common set of rules for all government and public administration stakeholders to uphold.

Based on a framing of 'AI as a social experiment,' this study arrives at four policy options for European Parliamentary policy-makers:

It is proposed that all AI/ML system developers are required to have a **data hygiene certificate (DHC) to be eligible to sell their solutions to government institutions and public administration bodies**. It is well known that the quality (or hygiene) of the data plays a key role in the efficacy and accuracy of an algorithm. Without accurate algorithms, the autonomously developed rules (of ML) will also be skewed. Consequently, a first ethical constraint is to ensure the quality of the data being used to train the algorithm, where quality is measured according to its sourcing, acquisition, diversity, and labelling. Such a certificate does not require insight into the proprietary aspects of the AI system (i.e. companies do not have to divulge their algorithm) and, of equal importance, such a certificate does not require organisations to share their data sets (which may be their source of income) with competing organisations.

It is proposed that all public and government organisations using AI systems are required to conduct **an ethical technology assessment (eTA) prior to deployment of the AI system**. The eTA is a written document intended to capture and log the dialogue that occurred between ethicist and technologist and/or ethicist and officials of the public administration about to implement the AI/ML solution. The eTA is a list of ethical issues related to the AI/ML application, made by an expert trained to engage in ethical reflection (or at the very least one who is able to envision possible moral risks related to the implementation of the AI/ML). The eTA is the moment at which all the possible ethical risks that could result from the AI/ML application in question must be considered.

It is proposed that all public administration institutions and government bodies are required to show **clear goals for the AI/ML application**. With this policy option, it is not possible to deploy AI/ML in society in the hopes of learning an unknown 'something'. Instead, there must be a specific and explicit 'something' to be learned. The specific aim and scope of the AI/ML experiment must also be stated as part of the eTA.

It is proposed that all organisations deploying AI systems should produce an **'accountability report' in response to the eTA**. The accountability report is the third step in logging the AI/ML usage in public administration and/or in government. Whereas the eTA is meant to draw out the possible negative consequences of implementing an AI system (completed by an external third party), the accountability report is a response to the eTA, completed by the organisation implementing the AI/ML system. It is meant as a response to the ethical and human rights issues that were identified in the eTA. Thus, in the accountability report, it is proposed that institutions will be required to account for how they have mitigated or corrected the concerns raised in the eTA.

Conclusion

Although there has been an increase in attention to ethics in the AI debate, little has been done to show what ethics means for the creation of policy and regulation of AI. What ethics provides as a field of study is a vision of the future: a normative perspective, rather than descriptive, with an eye to 'the good life'. Given our lack of operational experience with AI – the level of uncertainty and risk – it is wise to introduce experimental conditions for the real world applications of AI, especially when it comes to ethical constraints and requirements for demonstrating clear benefit. The regulatory options provided in this study are the ethical constraints under which AI may be introduced into society, and are the first step in allowing ethics to play a considerable role in the implementation of AI for the public good.

MAIN REFERENCES

Bryson J., A smart bureaucrat's guide to AI regulation [Internet]. 2019. Available from: https://joanna-bryson.blogspot.com/2019/01/a-smart-bureaucrats-guide-to-ai.html

Floridi L., Cowls J., Beltrametti M., et al, AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, Minds Mach, December 2018, 28(4):689–707.

Johnson D. G., Technology with No Human Responsibility?, J Bus Ethics., April 2015, 127(4):707–15.

Van de Poel I., Why New Technologies Should be Conceived as Social Experiments, Ethics Policy & Environment, October 2013, 16(3):352–5.

Van de Poel I., Nuclear energy as a social experiment, Ethics, Policy & Environment, 2011, 14(3), 285-290.

This document is based on the STOA study on 'Artificial intelligence: From ethics to policy' (PE 641.507) published in June 2020. The study was written by Dr Aimee van Wynsberghe of Delft University of Technology and co-director of the Foundation for Responsible Robotics, at the request of the Panel for the Future of Science and Technology (STOA) and managed by the Scientific Foresight Unit within the Directorate-General for Parliamentary Research Services (DG EPRS) of the European Parliament. STOA administrator responsible: Mihalis Kritikos.

DISCLAIMER AND COPYRIGHT

stoa@ep.europa.eu (contact)

http://www.europarl.europa.eu/stoa/ (STOA website)

www.europarl.europa.eu/thinktank (internet)

http://epthinktank.eu (blog)