DOI: 10.2478/ijasitels-2022-0003

Data, governance and tackling the "tragedy of the commons"

Valentin Bogdan Crețu
[1]*Projectix Software SRL, Bistrița, Romania,*
cretu.bogdan.valentin@gmail.com

Abstract

The purpose of this paper is to highlight the importance of active and continued public engagement in the debate regarding the regulation of artificial intelligence (AI). The results of several studies are presented, in reference to the benefits of this technology, its risks and limitations, and the shortcomings of traditional and hybrid approaches. Certain conceptual and practical approaches are presented, that aim to facilitate the participation of citizens and other stakeholders in the decision-making process of AI governance.

Keywords: governance, artificial intelligence, human rights

1. Statement of the problem

1.1. Introduction

The concept of tragedy of the commons was originally coined by the British economist William Forster Lloyd [1], and it refers to a worst-case scenario, regarding the dysfunctional exploitation of a of commons. The following elements usually define such situation:

- individual actors;
- with access to a resource;
- where each act independently, according to their narrow self-interest;
- without shared social structures or formal rules for access and use;
- leading to depletion of the resource through their uncoordinated action, contrary to the common good of all users.

The classic example of this concept is the overgrazing of a pasture (seen as a common), leading to its demise. Increasingly, climate change is seen as a tragedy of the commons as well [2].
In the following sections, we will explore the applicability of this concept to the topic of artificial intelligence (AI) and its impact on society, technology and economy.

1.2. Benefits of AI

The following benefits of AI have been identified [3]: clinical decision support systems that make medical diagnoses; filtering algorithms that categorise and provide personalised content for users; policing systems that predict the likelihood of criminal

activities; increasing cost savings and quality of life; meeting labour shortages; tackling ageing populations; strengthening national defence.

1.3.AI risks and limitations

Several risks regarding the deployment of AI have been identified [4]:

- unexpected behaviours that pose safety hazards for users. Some examples: an algorithm for landing aircraft exploited overflow errors in the physics simulator by creating large forces that were estimated to be zero, resulting in a perfect score; a robotic arm trained using hindsight experience replay to slide a block to a target position on a table achieves the goal by moving the table itself; a neural network evolved to classify edible and poisonous mushrooms took advantage of the data being presented in alternating order, and didn't actually learn any features of the input images. The use case with the highest impact, currently, regards the use of autonomous driving technology.
- biases in the data and algorithm can lead to discrimination in credit scoring and criminal sentencing. The risk is that AI algorithms might identify as indicators of bad credit/higher risk of criminality certain factors that are discriminatory such as: gender, ethnicity, sexual or political orientation, among others. The use of AI for decisions regarding bail, incarceration, parole or sentencing has already led to several erroneous results (for example, unclear details about the face of someone in the vicinity of a crime has led to them being registered in a database that relies upon an algorithm to match it to a mugshot or driver's license).
- loss of human autonomy and control in caregiving and military combat. The International Committee of the Red Cross has noted in a position paper ("Artificial intelligence and machine learning in armed conflict: A human-centred approach") that "autonomous weapon systems – weapon systems with autonomy in their "critical functions" of selecting and attacking targets – are an immediate concern from a humanitarian, legal and ethical perspective, given the risk of loss of human control over weapons and the use of force".
- displacement of millions of jobs, unevenly distributed within and across countries;
- privatization of benefits and socialization of costs;
- increased risk regarding: the formation of surveillance states; accelerating losses of democratic freedoms and personal autonomy.

Certain limitations are also of particular relevance: lack of generalizability to different contexts; the difficulty in understanding why and how a decision was made [5].

1.4.The tragedy of the commons, in the context of AI

Taking into consideration the wide impact of this technology, on virtually all sectors of society, economy and technology, the initial formulation of the collaboration problem can be restated in these terms:

Table 1. Reformulation of the problem

Initial formulation	**Reformulation in the context of AI**
individual actors;	*All* members of economy and society
with access to a resource;	*All* natural resources; social trust, social cohesion, civic liberties
where each act independently, according to their narrow self-interest;	where each act independently, according to their narrow self-interest
without shared social structures or formal rules for access and use;	without shared social structures or formal rules for access and use;
leading to depletion of the resource through their uncoordinated action, contrary to the common good of all users.	leading to depletion of the **resources** through their uncoordinated action, contrary to the common good of all users.

2. Current approaches and their issues

2.1. Traditional governance approaches

The following issues can be highlighted regarding the traditional governance approaches (regulation, taxes or subsidies): the need to need to reassess their efficacy; the lagging behind due to rapid developments in the technology; the reluctance to define specific rules and duties for algorithm programmers to allow for future innovation, that can in turn allow programmers to evade responsibility and accountability for the system's resulting behaviour in society [6].

2.2. Hybrid/adaptive governance approaches – definition, examples

Hybrid/adaptive governance approaches can be described in several ways: diminished role of government; nonbinding norms and techniques; various forms of legal experimentation regarding the testing or implementation of new technologies.
Certain issues regarding such governance approaches can be identified: their voluntary nature cannot assure that the outlined principles will always be adhered to; ensuring consistent application of these guidelines in designing the same AI technology across different sectors; the challenge of ensuring inclusivity and representation of diverse stakeholders.

An example in this regard is the ISO 26000 voluntary standard for "Guidance in Social Responsibility". While it recognizes distinct categories of responsible organizational behaviour which are based on identifiable corporate practices and subsequent measurable performance, it also states that what social responsibility "means" ultimately varies from one company to another. Unfortunately, this leaves ambiguity in place regarding, for example, what a "socially responsible" firm is within the same industry.
Another matter of concern regarding these technologies pertains to their opacity: algorithms are often kept intentionally opaque by their developers to prevent cyber-attacks and to safeguard trade secrets; most individuals lack sufficient technical literacy or the willingness to pay for accessing such expertise to help them to interpret these explanations, therefore they are unlikely to benefit from the provisions of requirements by the likes of GDPR in making informed decisions; data fragmentation and lack of interoperability between systems.
On the other hand, a counter-criticism can also be formulated: these arguments obfuscate that algorithms are fundamentally understandable; "rather than discounting systems which cause bad outcomes as fundamentally inscrutable and therefore uncontrollable, we should simply label the application of inadequate technology what it is: malpractice, committed by a system's controller".
In addition to opacity, the problem of undue corporate influence has been identified: major technology companies and AI developers such as Google, Facebook, Microsoft, and Apple possess huge informational and resource advantages over governments in regulating AI; such information asymmetries increase the difficulty for regulators in understanding and applying new or existing legislation to AI applications.
With AI in general, the concern is that if we continue blindly forward, we should expect to see increased inequality alongside economic disruption, social unrest, and in some cases, political instability, with the technologically disadvantaged and underrepresented faring the worst [7].

3. Proposed solutions

3.1. General formulation

An element that has been conspicuously missing from all the previously mentioned approaches is *wide, active and persistent* public engagement in defining:

- social values, applicable to the design, implementation and regulation of AI;
- acceptable risks regarding the use of this technologies.
- distribution of benefits and costs across all the actors in society.

The absence of this element makes the implementation and regulation of AI prone to risk amplification and the "dread risk" that has plagued the regulation of other advanced technologies, such as nanotechnology or genetically modified crops.

3.2. The efficacy of the ethical approach

An ethical approach to the implementation of AI (that is, one that promotes the common good and shared cultural values above economic indicators) ca be said to have the following benefits: the ability to identify and leverage new opportunities that

are socially acceptable or preferable; the ability to anticipate and avoid or at least minimize costly mistakes.
Certain prerequisites are required in order to achieve this: there must be an environment of public trust and clear responsibilities, and benefits must be publicly perceived as meaningful and risks as potential, yet preventable, minimizable, or at least something against which one can be protected, through risk management (e.g. insurance) or redressing. These attitudes will depend in turn on public engagement with the development of AI technologies and openness about how such technologies operate [8].

3.3.Society-in-the-loop

The original concept of human-in-the-loop refers to a situation where one or more human operators are intermittently programming and continually receiving information from a computer that itself closes an autonomous control loop through artificial effectors to the controlled process or task environment; can be used to both improve and also regulate the AI.
The society-in-the-loop concept is extension of this approach: society is first responsible for finding consensus on the values that should shape AI and the distribution of benefits and costs among different stakeholders [9].

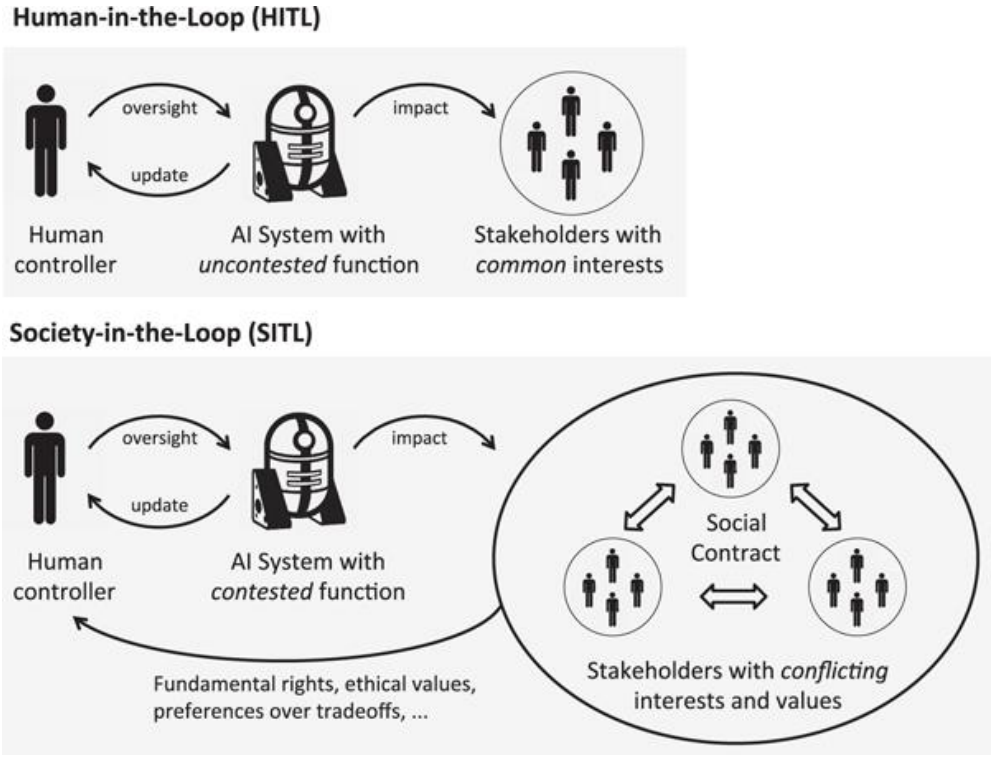

Figure 1. Society-in-the-Loop (SITL) = Human-in-the-Loop (HITL) + Social Contract

3.4. Asilomar principles

The Asilomar Conference on Beneficial AI was a conference organized by the Future of Life Institute, held in 2017, at the Asilomar Conference Grounds in California. Its outcome was the creation of a set of guidelines for AI research – the 23 Asilomar AI Principles, such as:

- Support for the creation of educational curricula and public awareness activities around the societal, legal, and ethical impact of Artificial Intelligence;
- Initiatives to educate employees in companies and public authorities on the societal, legal, and ethical impact of working alongside AI;
- Implementation of structured public consultation mechanisms to design policies and rules related to AI;
- Assessment of which tasks and decision-making functionalities should not be delegated to AI systems;
- Development of agreed-upon metrics for the trustworthiness of AI products and services.

3.5. Activism for digital rights

Last, but definitely not least, we should mention areas of digital rights initiatives, from various NGO's and grassroots organizations, that have been successful in securing the public interest:

- Privacy International: has influenced the EU to classify some digital surveillance technologies as dual-use for the purpose of exportation;
- Stop the Cyborgs: has contributed to Google's eventual decision not to fully commercialize the explorer version Google Glass;
- Access to Information Program (AIP), Bulgaria: has prevented 'passive access through a computer terminal' by the Ministry of Interior, as well as access without court permission by security services and other law enforcement bodies, to all retained data by Internet and mobile communication providers;
- Germany: over 30,000 German citizens brought a class action suit, leading to the suspension of a Directive that violated citizens' rights to privacy;
- Digital Rights Ireland: has improved EU approach to data retention [7].

4. Conclusions

On the one hand, artificial intelligence holds great promise for improving the quality of life of citizens and for boosting economic activity. On the other hand, there are also very significant risks regarding this technology, from the proper application of law, to secure driving, to lawful use of force, or to the safeguard of fundamental human rights.

The asymmetries in information and understanding of this technology, between the private sector and the public authorities, has created a tremendous pressure toward softer/hybrid approaches for regulation, if not outright "laissez-faire". However, this carries a grave risk of the un-democratizing of the regulatory process and of alienation of public opinion.

In contrast, approaches that put front and center the public interest, and dialogue with all the interested shareholders, hold the greatest potential in ensuring that the public is adequately aware of the benefits and risks of this technology and that it is invested in its implementation, in a manner that harmonious with current social values and norms.

References

[1] Thompson, N, *"Thompson, William (1775–1833), socialist and economist"*, Oxford University Press, Oxford Dictionary of National Biography (online ed.), doi:10.1093/ref:odnb/27284.
[2] Caggiano, H., & Landau, L. F., *A new framework for imagining the climate commons? The case of a Green New Deal in the US*, Planning Theory, 0(0), https://doi.org/10.1177/14730952211042765, 2021.
[3] Helbing, D, *Societal, economic, ethical and legal challenges of the digital revolution: From big data to deep learning, artificial intelligence, and manipulative technologies*, Towards Digital Enlightenment, Springer Cham, City, ISBN 978-3-319-90868-7, 2019.
[4] He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. *The practical implementation of artificial intelligence technologies in medicine*, Nature Medicine, 25(1), 30–36, https://doi.org/10.1038/s41591-018-0307-0, 2019.
[5] Osoba, O. A., & Welser, W. *An intelligence in our image: The risks of bias and errors in artificial intelligence*, Rand Corporation, https://www.rand.org/pubs/research_reports/RR1744.html, 2017.
[6] Taeihagh, A., Ramesh, M., & Howlett, M. *Assessing the regulatory challenges of emerging disruptive technologies*, Regulation & Governance 15, 1009-1019, https://doi.org/10.1111/rego.12392, 2021.
[7] Leiser, M., & Murray, *The role of non-state actors and institutions in the governance of new and emerging digital technologies*, The oxford handbook of law, regulation and technology, A. Eds. R. Brownsword, E. Scotford, K. Yeung, & O. U. Press, Oxford, https://doi.org/10.1093/oxfordhb/9780199680832.013.28, 2016.
[8] Floridi, Luciano & Cowls, Josh & Beltrametti, Monica & Chatila, Raja & Chazerand, Patrice & Dignum, Virginia & Lütge, Christoph & Madelin, Robert & Pagallo, Ugo & Rossi, Francesca & Schafer, Burkhard & Valcke, Peggy & Vayena, Effy. *I4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, Minds and Machines 28, 10.1007/s11023-018-9482-5, 2018.
[9] Rahwan, I. *Society-in-the-Loop: Programming the Algorithmic Social Contract*, Ethics and Information Technology 20, 5-14, https://doi.org/10.1007/s10676-017-9430-8, 2018.