JeDEM Issue13(1): 203-219, 2021
ISSN 2075-9517
http://www.jedem.org
Date of submission: 06.12.2020
Date of acceptance: 30.05.2021
Article DOI: 10.29379/jedem.v13i1.625

Machine learning in Governments: Benefits, Challenges and Future Directions

Yulu Pi

Author ORCID Nr: 0000-0001-8961-2270
Centre for Interdisciplinary Methodologies, University of Warwick, piyulunju@163.com

Abstract: The unprecedented increase in computing power and data availability has significantly altered the way and the scope that organizations make decisions relying on technologies. There is a conspicuous trend that organizations are seeking the use of frontier technologies with the purpose of helping the delivery of services and making day-to-day operational decisions. Machine learning (ML) is the fastest growing and at the same time, the most debated and controversial of these technologies. Although there is a great deal of research in the literature related to machine learning applications, most of them focus on the technical aspects or private sector use. The governmental machine learning applications suffer the lack of theoretical and empirical studies and unclear governance framework. This paper reviews the literature on the use of machine learning by government, aiming to identify the benefits and challenges of wider adoption of machine learning applications in the public sector and to propose the directions for future research.

Keywords: Machine learning, artificial intelligence, public sector, benefits, challenges

Acknowledgment: I would like to take this chance to thank JeDEM for the opportunity to publish this paper and JeDEM's anonymous reviewers for their feedback that helped to improve this paper. I would also like to thank Professor Johan Elkink, Professor Kalpana Shankar, and Professor Cagatay Turkay for their valuable comments. Particularly, I want to thank Romain Malandre for proofreading this paper.

1. Introduction

Machine learning (ML) algorithms learn from the data and concentrate rules from the data, then apply the learned insight to assess new data, allowing their gained insights to solve many distinct problems ranging from classification and regression to clustering. Nowadays, ML algorithms are the basis for granting loans, online product recommendations, and social media friend suggestions (Adadi and Berrada, 2018). With its unparalleled computational capability to extract information from high-dimensional data as well as unstructured data, ML could help unlock the value of the data, free up high-value work, improve services to citizen queries, enhance predictive capability for decision-making, and thus fuel innovative public services (Eggers et al., 2017). Many scholars have observed the nascent status of governmental ML applications (Sun and Medaglia, 2019) and the scarcity of related research (Gomes de Sousa et al., 2019; Zuiderwijk et al., 2021). Most AI research has focused on technical issues or private sector use, leaving only 3.5% of the nearly 1700 investigated literature focused on the use of AI in the public sector (Gomes de Sousa et al., 2019). Zuidermijk et al.(2019) suggested three reasons to explain such a knowledge gap, namely, less AI expertise within the public sector compared to the private sector, a lower rate of leadership development in AI governance compared to the penetration of AI applications into global governments, and a lack of focus on the specific problems of AI use for public governance compared to technological problems.

Recognizing this knowledge gap, this paper focuses on the discussion of opportunities, challenges, and future directions in the area of algorithmic engagement in the public sector. It strives to serve as a brief primer based on a literature review to answer why ML should be used in public policy and what challenges it may pose to policy makers as well as initiating discussion on further improving the use of ML by government. This paper strives to offer an exhaustive examination in leading ML&AI as well as Public Administration journals (see Table 1), using the keywords: 'machine learning public policy', 'machine learning government', 'machine learning public sector', 'machine learning social good', 'machine learning review'. It synthesizes research and examples in action as its theoretical foundation to illustrate current and potential ML utilization in the public sector, allowing an examination of the central topics as well as a proposal for future research directions. It enables public officials to take the first steps in exploring and becoming familiar with ML. It also helps ML experts to understand the specific needs of the public sector to re-examine and improve existing ML applications. It is structured as follows: it first identifies the sector-specific benefits for public policy engagement on ML, and the second section focuses on the examination of the potential harm and challenges brought by the adoption of ML. Concluding policy advice and research recommendations to maximize its potential and overcome its dangers are given in the last part.

Table 1: List of all the examined journals. The list is ordered alphabetically.

List of examined journals
Academy of Management Review
American Economic Review
Artificial Intelligence and Law
Communications of the ACM
Government Information Quarterly
IEEE Access

IEEE Transactions on Visualization and Computer Graphic
International Journal of Advanced Computer Science and Applications
International Journal of Public Administration
Journal of Advanced Computer Science and Applications
Journal of Economic Behavior & Organization
Journal of Organizational Computing and Electronic Commerce
Journal of Technology in Human Services
Science Robotics
SSRN Electronic Journal
Telecommunications Policy
The American Review of Public Administration
The Annals of Applied Statistics

2. Why the use of machine learning in public policy?

The potential of machine learning in the public sphere is grounded in the tremendous data availability and policy prediction needs of this field. Machine learning has already demonstrated considerable potential to enhance the effectiveness and accuracy of many decisions-making scenarios ranging from medical diagnosis, granting mortgages, tax evasion, and terrorist activities identification (Kononenko, 2001; Nowshath et al., 2019; Rodríguez et al., 2019; Mantari et al., 2020). Machine Learning is an umbrella term encompassing a wide range of algorithms for fields such as natural language processing, data mining, image processing, and predictive analytics. It is also often referred to simply as algorithms and used interchangeably with Artificial Intelligence and automated decision-making (European Commission, 2018). This paper adopts this definition and uses the terms machine learning (ML), artificial intelligence (AI), algorithmic systems, and automated decision-making, interchangeably, ignoring their many other non-overlapping subfields. As a study that enables computers to automatically learn from experience instead of relying on manually, explicitly programmed rules, and generalize the acquired knowledge to new settings (UNECE Machine Learning Team, 2018), ML can automate repetitive tasks, handle the analysis of large datasets, and provide predictive information in the public sector.

First of all, ML algorithms have the unparalleled computational capability to extract information from high-dimensional data as well as unstructured data, such as texts, photos, videos, and blogs. In the digital age, governments have the access to a vast quantity of data collected by public bodies for registration, transaction, and record-keeping. These raw data, however, often do not occur in an understandable form for traditional statistics models because of the huge computation magnitude and unstructured format (Ubaldi, 2013). Therefore, the advancement of ML on data processing and analysis could help unlock the value of the data, automate the repetitive tasks, understand the citizens' needs, and thus fuel the innovative services. With the increasing development of communication, a tremendous amount of information is stored as digital text which can be used as an input to policy, social and economic study. The powerful combination of ML algorithms and digital text is now revolutionizing the way governments seek help from data. A natural language processing algorithm is developed by The United Nations Big Data program, Global Pulse in Indonesia. It can identify tweets that mention the prices of basic food items (beef, chicken, onions, and peppers), allowing the government to track food prices in real-time and have an early warning for unexpected price spikes (UN Global Pulse, 2014). ML algorithms can extract meaningful information from data that is more complex than text. A team at Stanford (Gebru et al, 2017) trained an ML model to extract features from the image of Google Street View in 200 US cities to estimate socioeconomic characteristics. This model is proved to have higher accuracy in predicting household income and thus can not only reduce the cost of labor-intensive door-to-door censure surveys but also help solve the lag of demographic changes between two surveys. Thanks to its powerful data processing capabilities, ML can help the public sector automate many highly labor-intensive data processing and analysis activities, thereby increasing the efficiency and speed of government services and actions.

Another merit of ML is its predictive power. A key challenge when developing policy is that the effects of a new policy are unknown until it is implemented. To address this problem, policy-makers resort to comparing similar policies abroad, performing policy trials, or developing statistical models that assist in predicting likely outcomes. Therefore, ML's predictive power allows the policy makers to anticipate a policy's impact before implementing, supporting the decisions of policy adoption. For example, after the outbreak of the COVID-19 pandemic, an algorithmic system was applied in Qatar to predict the impact of lockdown policy on COVID-19 cases assisting the formulation of social restriction policy (Said et al,2020). Additionally, ML predictions on social service demands can help optimize the allocation of limited social resources. An ML model of homeless family shelter entry and length of stay was developed at New York University to study the likelihood of re-entry and the probability of a homeless family becoming a long-term stayer. The results of the model help shelters understand the demand for the number of beds at any given time, so they can plan more efficient resource allocation based on predicted demand (Hong et al.,2018). Governance by such algorithmic prediction is increasingly interwoven into many scenarios of decision-making, such as predictive policing, smart city planning, and court adjudication (Abdul et al., 2018). Meijer et al.'s study (2019) demonstrates how the application of ML enables the computer to find the spatial pattern of prior criminal activities and automatically forecasts where crime might occur in the future in a considerably accurate manner. There also has been a surge of interest in predicting judicial decisions with ML algorithms. With natural language processing tools and predictive algorithms, Medvedeva's team (2020) learned from the proceeding court reports and automatically predicted future decisions with an average accuracy of 75%, which offers a reliable reference for the verdict.

Other illustrative examples include: (i) tailoring the tax rebate program for the households that are most likely to be consumption constrained (Andini et al., 2018); (ii) improving social welfare by hiring the police who will not use excessive force and promoting teachers that will bring the largest added value (Chalfin et al., 2016); (iii) foreseeing unemployment spell length to assist laborers in savings rates and job search strategies (Kleinberg et al., 2015).

3. Challenges posed by the use of machine learning

Because of the exponential increase in algorithm-assisted decisions within the public policy that can widely affect individuals' rights, interests, and expectations, ML is no longer just about the technical domain, but also has tremendous potential impacts on every aspect of our society. ML can generate numerous benefits in the public sphere. Unfortunately, based on the literature review, there is ample evidence that many concerns around it, such as explainability, fairness, and institutional challenges (Gunning, 2019; Yu et al., 2018; Guenduez et al., 2020), are still unresolved issues perpetuating, and even exacerbating the existing problems. Despite this, the technology community has launched a discussion focusing on the seriousness of the above concerns, the main thrust of research has been either focused on a technical perspective or the applications in the private sector. In light of differences in ownership, administrative culture, and relative reliance on political control versus market forces between the two sectors (Perry and Rainey, 1988), more involvement of public policy perspective to understand and address these issues has become a pressing research agenda.

3.1. Fairness, Bias, and Discrimination

Although ML can generate numerous benefits in the public sphere, there is ample evidence that intentional or unintentional discrimination against certain groups and individuals is still an unresolved issue. For instance, Amazon created an AI recruiting tool to rank the candidates and make hiring decisions. However, for technical jobs such as software engineers, the tool learnt to automatically ignore women's CVs, eliminating women's chance to get such jobs (Kodiyan, 2019). However, bias does not arise as a result of algorithm design in most cases; rather, algorithms inherit existing bias from historical data that contains remnants of bias from human decision-making and culture. In this case, the root of this kind of prejudice lies in the way traditional gender ideology and latent discrimination are captured in the data from which the algorithm learns (Leavy et al., 2020). The significant component of data that was utilized for training the AI recruiting system was the resumes of employees in the company, mostly males. This gender inequality embedded as data imbalance is the natural reflection of existing male dominance in the workplace. This trend was studied by the algorithm, and thus continually sustained by it, meaning that algorithms tend to maintain the status quo instead of making progress without human intervention.

Predictive Policing is another area that has raised concerns regarding racial bias. Although the use of ML technology to predict future crime participants and crime locations has doubled the accuracy of crime prediction over its current practice (Zach, 2013), it has been instrumental in leading to discrimination against a particular racial group. For example, ML predictive policing systems may inappropriately associate darker skin with greater criminal suspicion or lead to more arrests for minor crimes in communities of color (Selbst, 2017). As shown in these real-world incidents, data-driven decision-making is not free from existing, real-world biases. Since the public decisions made by AI will have a large-scale and profound impact on many aspects of society and individuals, they must be as fair as possible. If not developed with awareness, algorithms can inherit or even exacerbate historical inequities underlying the input data. Many discussions of the benefits of ML use in the public sector draw on an argument that

> *'the more data governments and public institutions manage to integrate into their systems, the higher the capabilities of machine learning to make decisions based on this data will be'(Cary &David, 2017).*

Thus putting algorithm design and the quality of input data under scrutiny, to prevent unfair algorithmic decisions is an essential step for integrating AI as a part of service support in the public sphere.

3.2. Explainability, Transparency, and Accountability

Another relevant aspect of fairness is to provide reasoning for the decisions made by algorithms. In the public sector, where decision-makers are hierarchically and democratically accountable, and where transparency is of crucial importance, explanations of the decisions leading to a certain policy are particularly important. It is the right of citizens to access information about the procedures and data which lead to certain decisions affecting them (Scantamburlo, 2019). Transparency of the decision-making process is the mechanism that facilitates accountability (Diakopoulos, 2016), allowing the tracking of the entire procedure, as well as the detection of responsibilities when some failures occur. However, because of the lack of simplicity and observable results, it is challenging for decision-makers to ensure explainability, transparency, and accountability when it comes to machine decisions. For instance, teachers from Houston prevailed in a lawsuit with their schools over an algorithmic system that evaluated their performance (Webb, 2017). Those who received great evaluation won praise, while those who received a bad rating, risked termination. Some teachers believed that the system penalized them unfairly, but they had no way of knowing for sure, since the firm that developed the software, the SAS Institute, considered its algorithm a trade secret and refused to reveal its workings. A federal judge determined that the program had violated the teachers' civil rights when the teachers brought their case to court. A teacher has the right to know the reason behind the decisions that affect his/her career, even if they are made by a computer. This example is a good illustration of the problems faced by using ML in the public sector, where users have no way of knowing for what reason the algorithm made a certain decision and who to blame when the machine makes undesirable decisions. There is a requirement for explainability and transparency of the algorithms to ensure their accountability.

As ML is applied in many sensitive areas in policy making, the need for explainability and transparency of the decision-making process and the clear accountability of the decision in the public sector, becomes even more important. As Rudin believes (2019) explainability and interpretability are necessarily defined in a domain-specific way, these concepts in the context of the public sector have not reached consensus, due to the current nascent status of investigations. Since policy makers rarely have a background in ML and they require explanations for different purposes than model developers do, this paper defines explainability as the ability of algorithmic systems to present understandable post-hoc explanations in various ways such as natural language explanations, visual explanations, and explanations by example, for the causes of their decisions. Once the reasoning behind the machine decisions is revealed to policy makers, it is up to them to decide whether or not to adopt these decisions leaving them to hold accountability for the behavior and the potential impacts of machine decisions.

However, it is not an easy task to make ML models explainable and transparent. There tends to be a trade-off between explainability and predictive accuracy meaning a trade-off between models that are easier to interpret and complex models that provide more accurate predictions. The users are often unable to describe in detail how the decision comes about, or on what aspects of the data the decision is based (Adadi and Berrada, 2018), because the algorithms with higher prediction accuracy, usually take the unexplainable form to understand and discover the subtle correlation between the input variables. The opaque nature of algorithms makes communication between machine learning experts, policy makers, and other domain specialists difficult (Letham et al., 2015). Improving the explainability of machine learning will provide more transparency into how decisions are reached, enabling decision-makers and citizens to track and understand the whole process of decision making, and eventually upgrade the accountability of the algorithmic decision-making system.

3.3. Institutional Challenge

Researchers have also begun to investigate the institutional reasons behind the challenge of incorporating technology-centric solutions into decision-making in the public sector. A decision-making process through which a data-driven approach becomes fully integrated into the organizational culture is crucial to the success of algorithmic support provision. Although most high-level national AI strategies show a welcoming attitude towards their participation in this cutting-edge technology, the lived experience of the bureaucrats at the micro-level are usually in an odd direction because the use of AI challenges the traditionally bureaucratic form of the public sector (Bullock, 2019). The case studies show that innovation adoption in government often takes place within existing systems and utilizes existing tools, leading to bottlenecks and undermining the organization's ability to effectively manage innovation engagement (UNITED NATIONS, 2020), such as ML. Exacerbating this, given the limited infrastructures and the employee resistance to using technology innovation, the public sector is also heavily resource-constrained and path-dependent (Veale et al., 2018). Guenduez's study (2020), also revealed a widespread skepticism of technology innovation among public managers, applying big data as an example.

The use of big data is increasingly studied in the context of policy process or "policy cycle" (Höchtl et al., 2016). In contrast, only a few studies have investigated the opportunities that ML, the working horse in the era of Big Data, can bring to the policy-making process. Most studies in the field of digital transformation in public policy do not pay enough attention to the specific effects of ML on public services. However, this is of particular importance in the context of the current crisis of COVID-19, where algorithmic tools have been developed to put into use in the frontline to keep the economy working and provide a health service. As new situations and a wider range of AI applications emerge, AI and the underlying regulatory mechanism for data ecosystems have become crucial policy concerns (Jordan, 2020). A need to identify the gap between the ML literature and understanding, expectations, and needs of policy makers is justified by the wide-ranging impact caused by the application of ML.

4. Moving towards better use of machine learning

The challenges described above are interlinked with each other and eventually lead to the unreliable and undesirable application of algorithms in the public sector. For instance, researchers found that the PredPol algorithm, a widely used predictive policing platform with the capability to help the policing resource allocation with its future crime risk prediction, improperly link dark skin to greater suspicion of having committed crimes or lead to more arrests for nuisance crime in the colored neighborhoods (Selbst, 2017). Due to the complexity of PredPol's algorithm, it is not an easy task to understand how the decision is made and detect its bias. If police officers cannot comprehend the prediction from the software, they may not take any response to it, hampering the effectiveness of the predictive analytic tools. This might explain the outcome of a recent survey of 70 police agencies where, while 70 percent have predictive tools, only 22 percent were actually using them to help make decisions (Perry et al., 2013). To encourage AI to take a larger role in the policy-making process, coming up with solutions to address these issues is fundamental. Here I present three possible directions that future research can explore to ensure an understandable, accountable and prevalent use of ML in the field of public policy.

4.1. User-centered Explainable AI

Many benefits around ML have yet to be realized within public sectors, due to its opaque nature and the requirements of governments' accountability. To address the opacity, explainable artificial intelligence (XAI) was initiated to "enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners" (Gunning, 2017). XAI is ultimately a human-computer interaction (HCI) problem, a multidisciplinary field on the intersection of social science and artificial intelligence, shifting towards more interpretable AI. Interpreting the reasons behind the predictions can transform an untrustworthy model or prediction into a trustworthy one (Ribeiro, 2016). Friedler proposed two forms of interpretability: global interpretability means understanding the whole of a trained model; local interpretability means understanding the effects of a trained model on a particular input and its corresponding outcome (Friedler et al. 2019). A more relevant way to classify explanations is to divide the users of explanations based on their level of knowledge in AI: technical and non-technical users (Wanner et al., 2020).

However, existing XAI focuses primarily on providing technical interpretations for technical users to debug or improve the performance of algorithms, while non-technical end users, the largest group of XAI users who will be affected by machine decisions or entitled to grant system adoption, are largely ignored. Although the booming research on XAI, the current research tends to take an algorithm-centric view, disregarding the specific needs of real-world users (Liao et al. 2020). There is an increasing awareness that different users will have different requirements in terms of explanation of the prediction. An analyst might be particularly interested in the internal workings of the estimator, while a decision-maker might be more interested in the key facts or data points that lead to the prediction (Sørmo et al. 2005). As a result, based on the distinct level of expertise, need and expectation, appropriate and differentiated explanations should be provided to different user groups.

Although recent research has shifted its attention slightly to the end users of the model explanations (Ehsan et al., 2021), most XAI researchers do not pay enough attention to different audiences so that the explanations are targeted (Ribera, 2019). Furthermore, a large part of this literature is about explaining the inner workings of the algorithms, instead of the justification of the results (Sørmo et al., 2005, Adadi and Berrada, 2018). In the public sector, where decision- makers are hierarchically and democratically accountable, the justification for the policy with its underlying evidence is particularly important. Therefore, the need is urgent, and the time is ripe for the interdisciplinary collaborations of machine learning and social science community to answer the question of what type of algorithmic interpretation is most needed by policy makers and how this need can be met from a technical and practical use perspective. I believe that a usercentered XAI research agenda is becoming more necessary today when algorithms are involved in many high-stake decision-making processes.

Offering tailored explanations based on the characteristics and expectations of the targeted group, user-centered XAI could provide a certain level of transparency to ease the detection of model bias, unveiling hidden correlations between the input data amenable to lead to discriminatory solutions (Ahn, 2019). Besides, it promotes trust and social acceptance for not only the policy design made by the algorithms but also the policy itself, because it enhances the citizens' involvement in the process of service design, helps the communication to the groups influenced by the decisions, and thus raises citizen's awareness and trust of the policy decisions. In the context of AI engagement in the public sector, the development of XAI needs to be supported by multidisciplinary collaboration to ensure that it meets the needs and expectations of public officials. Therefore, future research should shift its focus to a user-centered XAI approach that has the ability to meet different stakeholders' distinct needs and expectations to provide people-centered services with an algorithm-involved approach, in the domain of public policy.

4.2. Algorithmic Impact Assessment

Due to the nature of public sector responsibility and the unresolved concerns around the use of ML, the poorly designed, unregulated algorithms have potentially far-reaching, adverse impacts, often involving the most vulnerable members of society. Public authorities urgently need a practical framework to assess algorithmic systems (European Parliament, 2019). The idea of implementing impact assessments of algorithm to ensure its accountability is gaining momentum (Moss et al, 2020). For instance, The Government of Canada requires a questionnaire-style impact analysis to ensure that its agencies are using ML algorithms in a manner that is compatible with core administrative law principles such as transparency, accountability, legality, and procedural fairness (Kuziemski, 2020). The US Congress proposed The Algorithmic Accountability Act of 2019, requiring companies to perform impact assessments of automated decision systems and evaluate their impact on accuracy, fairness, bias, discrimination, privacy, and security (Booker and Wyden, 2019).

To anticipate, avoid, and mitigate the negative consequences of algorithmic decision-making, many researchers have proposed a framework named Algorithmic Impact Assessment (AIA) as a practical practice for algorithmic accountability (Reisman et al., 2018). The AI Now Institute at New York University outlined four initial goals of AIA, namely providing the public with information about the systems that decide their fate, granting the meaningful access for external researchers to review and audit systems, developing the expertise of public agencies to assess the performance and impact of automated decision systems, and strengthening due process by offering the public the opportunity to engage with the AIA process before, during, and after the assessment (Reisman et al., 2018). Although the research and practice of AIA is still in its infancy, and issues such as the scope and structure of the assessment, when to conduct the assessment, what impacts count as impacts, and who should conduct the assessment are still open to discussion, some basic consensus has been reached on the form of AIA. First, the AIA requires a self-assessment of algorithmic systems within public agencies to assess potential impacts on fairness, justice, bias, and other harms and to make mitigation plans. To complement the insufficient internal self-assessment, AIA also has an emphasis on the establishment of regularly conducted external researcher review processes. Secondly, AIA emphasizes a combined participation of public authorities, domain experts, and the public. Algorithmic decision-making processes can be extremely intricate, and challenges such as bias and systematic error cannot be easily identified by evaluating systems on a case-by-case basis. Therefore, external expert analysis is recommended to be at both a group-level and interdisciplinary on an ongoing basis. The framework also recommends a public disclosure of the purpose, scope, intended use, self-assessment process of the algorithmic system, along with the associated policies at the start of the assessment process, to collect early external feedback and adjust the assessment for the most pertinent public concerns (Koene et al, 2019).

Through agency self-assessment, public disclosure of system adoption, and plans for meaningful access for external researchers and experts, AIA offers agencies a framework to evaluate the automated decision systems they adopt and to provide the public with greater insights into the workings of the systems, making governments ready to face the risks presented by them. Further research and practice is needed to formulate a standard format and application of AIA to ensure that AI moves in a direction beneficial to society.

4.3. Transformation Towards AI-powered culture

The pandemic has shown the critical role that AI plays in facilitating rapid policy decisions based on real-time data analysis and prediction. However, as explained above, the public sector still shows hesitance or lacks the capacity to embrace the arrived AI era. The United Nations has suggested nine key pillars for the AI transformation (UNITED NATIONS, 2020), prioritizing the changes in the government organizational culture, the implementation of the new regulatory framework, and the development of new individual capacities. The easiest start will be the training of the public servants, from the top policy-makers down, to facilitate the implementation of AI. Developing such training projects can improve public officials' knowledge, skills, and attitudes related to ML. As for the specific format of the training, Fountaine suggests the introduction of internal AI academies, which usually include classroom work (online or inperson), workshops, on-the-job training, or on-site visits to experienced commercial companies (Fountaine, 2019).

5. Conclusion

In terms of AI application, public sectors hold a dual role: as AI regulators, governments are obligated to shield citizens from the detrimental impact of the algorithms, while as AI users, governments are facing the pressure to respond to the demands of a rapidly-evolving society, boosting its efficiency with the help of ML (Kuziemski et al, 2021). Therefore, governments should take a distinct approach to respond to the new development of AI technology and social requirements. In this article, I first identified the benefits of ML use in government including automating complex data analysis, innovating public services, supporting policy making with impact prediction, and optimizing resources allocation, due to ML's data processing capabilities and predictive power. This section provided a justification for the pressing need for policy makers to understand, examine and embed an ML-powered approach to delivering services and making decisions, especially in the light of the COVID-19 pandemic. Then I raised my concern about the technical, legal, ethical, and organizational barriers and risks that impede a wider adoption of ML in the public sector. In view of the discussion in this paper, it is suggested that future research directions include:

- User-centered Explainable AI;
- Algorithmic Impact Assessment; and
- Transformation Towards AI-powered culture.

This article comes with some limitations because, as a new field that needs more discussion and research, the application of ML in the public domain is still in its infancy. While this paper intersperses many real-world examples, most ML projects in the public domain are still in the pilot or early implementation phase, and there is a lack of valid results to evaluate the use of ML for public policy. The scope of the reviewed papers may not show the full dimension of current government ML use and discussion, as some papers may have been excluded from the examined range. The analysis and synthesis of the literature discussed in this paper may run the risk of providing only hypothetical ideas about the benefits, challenges, and solutions of using ML in the public sector, as the literature analyzed is often normative and exploratory, lacking empirical support and only focusing on assumptions and expectations (Studinka et al., 2018). For example, the task of addressing explainability challenges in the context of ML engagement in the public sector, may not be an easy task from a technical perspective. It may even conflict with privacy protection in some cases, as people involved in training ML models, refuse their data or inferences about their data to be exposed.

Many important questions remain unanswered, such as the specific definitions and needs of the public sector regarding ML explainability, transparency, and accountability, and the framework for protecting government and citizens from the possible adverse consequences of ML. ML use in the public sector is by nature an interdisciplinary issue that requires the cooperation from AI community, public policy community, and other related communities. This paper calls for more practices and research through a multidisciplinary approach, to further understand the specific benefits, challenges, and solutions of ML applications technically, socially, legally, and politically. More specific case studies should be expected in different policy areas, in different regions, and at specific government levels, to enable the analysis and comparison of the specific benefits, challenges, and solutions. Future research should also aim to develop a practical framework to guide and govern the incorporation of ML innovations and mitigate possible risks through case studies, pilot implementations, or large-scale implementations. Further domain-specific, interdisciplinary research is necessary for ML algorithms to accomplish their full potential in the public policy field. This paper is an introduction to ML in a public policy context that calls for a more specific, systematic, and interdisciplinary investigation of the complex government AI use and regulation.

References

"*Algorithmic Accountability Act of 2019''*, OLL19293, 116th Congress(2019), Retrieved July,7 2021, from https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202019 %20Bill%20Text.pdf

Abdul, Ashraf, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. "*Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda*." In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, 1–18. Montreal QC, Canada: ACM Press, 2018. https://doi.org/10.1145/3173574.3174156.

Adadi, Amina, and Mohammed Berrada. "*Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*." IEEE Access 6 (2018): 52138–60. https://doi.org/10.1109/ACCESS.2018.2870052.

Ahn, Yongsu, and Yu-Ru Lin. "*FairSight: Visual Analytics for Fairness in Decision Making*." IEEE Transactions on Visualization and Computer Graphics, 2019, 1–1. https://doi.org/10.1109/TVCG.2019.2934262.

Aslam, Uzair & Aziz, Hafiz Ilyas Tariq & Sohail, Asim & Batcha, Nowshath. (2019). *An Empirical Study on Loan Default Prediction Models. Journal of Computational and Theoretical Nanoscience*. 16. 3483-3488.https://doi.org/10.1166/jctn.2019.8312.

Brenda Leong, Dr. Sara Jordan (2020), '*Artificial Intelligence and the COVID-19 Pandemic*', Retrieved July,7 2021, from https://fpf.org/2020/05/07/artificial-intelligence-and-the-covid-19-pandemic/

Bullock, J. B. (2019). *Artificial intelligence, discretion, and bureaucracy*. The American Review of Public Administration, 49(7), 751–761. https://doi.org/10.1177/ 0275074019856123.

Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan. 2016. "*Productivity and Selection of Human Capital with Machine Learning.*" American Economic Review, 106 (5): 124-27. https://doi.org/10.1257/aer.p20161029.

Coglianese, Cary and Lehr, David, "*Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*" (2017). Faculty Scholarship at Penn Law. 1734. https://scholarship.law.upenn.edu/faculty_scholarship/1734

Eggers, W. D., Schatsky, D., & Viechnicki, P. (2017). *AI-augmented government. Using cognitive technologies to redesign public sector work*. Retrieved July,7 2021, from https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/artificial-intelligencegovernment.html.

Diakopoulos, Nicholas. "*Accountability in Algorithmic Decision Making.*" Communications of the ACM 59, no. 2 (January 25, 2016): 56–62. https://doi.org/10.1145/2844110.

Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker; *Algorithmic impact assessments: a practical framework for public agency accountability*; AI Now Institute, 2018, https://ainowinstitute.org/aiareport2018.pdf.

Ehsan, Upol, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. "*Expanding Explainability: Towards Social Transparency in AI Systems.*" Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, May 6, 2021, 1–19. https://doi.org/10.1145/3411764.3445188.

European Commission (2018) *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Artificial Intelligence for Europe*, Retrieved July,7 2021, from https://ec.europa.eu/transparency/documentsregister/detail?ref=COM(2018)237&lang=en

European Parliament. Directorate General for Parliamentary Research Services. *A Governance Framework for Algorithmic Accountability and Transparency.* LU: Publications Office, 2019. https://data.europa.eu/doi/10.2861/59990.

Gebru, Timnit, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. "*Using Deep Learning and Google Street View to Estimate the Demographic Makeup of Neighborhoods across the United States.*" Proceedings of the National Academy of Sciences 114, no. 50 (December 12, 2017): 13108–13. https://doi.org/10.1073/pnas.1700035114.

Gomes de Sousa, W., Pereira de Melo, E. R., De Souza Bermejo, P. H., Sousa Farias, R. A., & Oliveira Gomes, A. (2019). *How and where is artificial intelligence in the public sector going? A literature review and research agenda.* Government Information Quarterly, 36(4), 101392. https://doi.org/10.1016/j.giq.2019.07.004.

Guenduez, Ali A., Tobias Mettler, and Kuno Schedler. "*Technological Frames in Public Administration: What Do Public Managers Think of Big Data?*" Government Information Quarterly 37, no. 1 (January 2020): 101406. https://doi.org/10.1016/j.giq.2019.101406.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S. ORCID: 0000-0001-6482- 1973 and Yang, G-Z. (2019). *XAI-Explainable artificial intelligence.* Science Robotics, 4(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Höchtl, Johann, Peter Parycek, and Ralph Schöllhammer. "*Big Data in the Policy Cycle: Policy Decision Making in the Digital Era.*" Journal of Organizational Computing and Electronic Commerce 26, no. 1–2 (April 2, 2016): 147–69. https://doi.org/10.1080/10919392.2015.1125187.

Hong, Boyeong, Awais Malik, Jack Lundquist, Ira Bellach, and Constantine E. Kontokosta. "*Applications of Machine Learning Methods to Predict Readmission and Length-of-Stay for Homeless Families: The Case of Win Shelters in New York City.*" Journal of Technology in Human Services 36, no. 1 (January 2, 2018): 89–104. https://doi.org/10.1080/15228835.2017.1418703.

Huamaní, Enrique Lee, Alva Mantari, and Avid Roman-Gonzalez. "*Machine Learning Techniques to Visualize and Predict Terrorist Attacks Worldwide Using the Global Terrorism Database*." International Journal of Advanced Computer Science and Applications 11, no. 4 (2020).https://doi.org/10.14569/IJACSA.2020.0110474.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. "*Prediction Policy Problems*." American Economic Review 105, no. 5 (May 2015): 491–95. https://doi.org/10.1257/aer.p20151023.

Kodiyan, Akhil Alfons. "*An Overview of Ethical Issues in Using AI Systems in Hiring with a Case Study of Amazon's AI Based Hiring Tool*," Retrieved July,7 2021, from https://www.academia.edu/42965903/An_overview_of_ethical_issues_in_using_AI_systems_in_hiring _with_a_case_study_of_Amazons_AI_based_hiring_tool

Koene, Ansgar, Christopher Wade Clifton, Yohko Hatada, Helena Webb, Menisha Patel, Caio Machado, Jack LaViolette, et al. *A Governance Framework for Algorithmic Accountability and Transparency: Study*, 2019. http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_ EN.pdf.

Kononenko I. *Machine learning for medical diagnosis: history, state of the art and perspective.* Artificial Intelligence in Medicine. 2001 Aug;23(1):89-109. https://doi.org/10.1016/s0933-3657(01)00077-x.

Kuziemski, Maciej, and Gianluca Misuraca. "*AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings.*" Telecommunications Policy 44, no. 6 (July 2020): 101976. https://doi.org/10.1016/j.telpol.2020.101976.

Leavy, Susan, Gerardine Meaney, Karen Wade, and Derek Greene. "*Mitigating Gender Bias in Machine Learning Data Sets*." ArXiv:2005.06898 [Cs], May 18, 2020. http://arxiv.org/abs/2005.06898.

Letham, Benjamin, Cynthia Rudin, Tyler H. McCormick, and David Madigan. "*Interpretable Classifiers Using Rules and Bayesian Analysis: Building a Better Stroke Prediction Model.*" The Annals of Applied Statistics 9, no. 3 (September 2015): 1350–71. https://doi.org/10.1214/15-AOAS848.

Liao, Q. Vera, Daniel Gruen, and Sarah Miller. "*Questioning the AI: Informing Design Practices for Explainable AI User Experiences.*" Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, April 21, 2020, 1–15. https://doi.org/10.1145/3313831.3376590.

Medvedeva, Masha, Michel Vols, and Martijn Wieling. "*Using Machine Learning to Predict Decisions of the European Court of Human Rights.*" Artificial Intelligence and Law 28, no. 2 (June 2020): 237–66. https://doi.org/10.1007/s10506-019-09255-y.

Meijer, Albert, and Martijn Wessels. "*Predictive Policing: Review of Benefits and Drawbacks.*" International Journal of Public Administration 42, no. 12 (September 10, 2019): 1031–39. https://doi.org/10.1080/01900692.2019.1575664.

Monica Andini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio, Viola Salvestrini,*Targeting with machine learning: An application to a tax rebate program in Italy*, Journal of Economic Behavior & Organization,Volume 156, 2018, Pages 86-102. https://doi.org/10.1016/j.jebo.2018.09.010.

Moss, Emanuel, Elizabeth Watkins, Jacob Metcalf, and Madeleine Clare Elish. "*Governing with Algorithmic Impact Assessments: Six Observations.*" SSRN Electronic Journal, 2020. https://doi.org/10.2139/ssrn.3584818.

Pérez López, César, María Delgado Rodríguez, and Sonia de Lucas Santos. "*Tax Fraud Detection through Neural Networks: An Application Using a Sample of Personal Income Taxpayers*." Future Internet 11, no. 4 (March 30, 2019): 86. https://doi.org/10.3390/fi11040086.

Perry, James L., and Hal G. Rainey. 1988. *The public-private distinction in organization theory: A critique and research agenda.* Academy of Management Review 13 (2): 182–201.

Perry, Walter L., Brian McInnis, Carter C. Price, Susan C. Smith, and John S. Hollywood. "Findings for Practitioners, Developers, and Policymakers." *In Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, 115-38. RAND Corporation, 2013. Accessed July 8, 2021. http://www.jstor.org/stable/10.7249/j.ctt4cgdcz.13.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "*'Why Should I Trust You?': Explaining the Predictions of Any Classifier*." In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–44. San Francisco California USA: ACM, 2016. https://doi.org/10.1145/2939672.2939778.

Ribera, M. & Lapedriza, A. (2019). *Can we do better explanations? A proposal of user-centered explainable* AI. CEUR Workshop Proceedings, 2327, http://ceur-ws.org/Vol-2327/IUI19WS-ExSS2019-12.pdf

Rudin, Cynthia. "*Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*." ArXiv:1811.10154 [Cs, Stat], September 21, 2019. http://arxiv.org/abs/1811.10154.

Said, Ahmed Ben, Abdelkarim Erradi, Hussein Aly, and Abdelmonem Mohamed. "*A Deep-Learning Model for Evaluating and Predicting the Impact of Lockdown Policies on COVID-19 Cases*." ArXiv:2009.05481 [Physics], September 11, 2020. http://arxiv.org/abs/2009.05481.

Scantamburlo, Teresa, Andrew Charlesworth, and Nello Cristianini. "*Machine Decisions and Human Consequences*." ArXiv:1811.06747 [Cs], April 30, 2019. http://arxiv.org/abs/1811.06747.

Selbst, Andrew D. "*Disparate Impact in Big Data Policing*." SSRN Electronic Journal, 2017. https://doi.org/10.2139/ssrn.2819182.

Shelby Webb (2017) *Houston teachers to pursue lawsuit over secret evaluation system*, Retrieved July,7 2021, from https://www.houstonchronicle.com/news/houston-texas/houston/article/Houstonteachers-to-pursue-lawsuit-over-secret-11139692.php

Slack, Dylan, Sorelle A. Friedler, Carlos Scheidegger, and Chitradeep Dutta Roy. "*Assessing the Local Interpretability of Machine Learning Models.*" ArXiv:1902.03501 [Cs, Stat], August 2, 2019. http://arxiv.org/abs/1902.03501.

Sørmo, F., Cassens, J. & Aamodt, *A. Explanation in Case-Based Reasoning–Perspectives and Goals.* Artif Intell Rev 24, 109–143 (2005). https://doi.org/10.1007/s10462-005-4607-7

Studinka, Julia & Guenduez, Ali A.: *The Use of Big Data in the Public Policy Process - Paving the Way for Evidence-Based Governance.* 2018. - EGPA Conference 2018. - Lausanne.

Sun, Tara Qian, and Rony Medaglia. "*Mapping the Challenges of Artificial Intelligence in the Public Sector: Evidence from Public Healthcare*." Government Information Quarterly 36, no. 2 (April 2019): 368–83. https://doi.org/10.1016/j.giq.2018.09.008.

Tim Fountaine , Brian McCarthy and Tamim Saleh (2019) "*Building the AI-Powered Organization*", Retrieved July,7 2021, from https://hbr.org/2019/07/building-the-ai-powered-organization

Ubaldi, B. (2013), "*Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*", OECD Working Papers on Public Governance, No. 22, OECD Publishing, Paris, https://doi.org/10.1787/5k46bj4f03s7-en.

UN Global Pulse (2014) *Mining Indonesian Tweets to Understand Food Price Crises*, Retrieved July,7 2021, from https://www.unglobalpulse.org/wp-content/uploads/old_site/UNGP_ProjectSeries_Nowcasting_Food_Prices_2014.pdf

UNECE Machine Learning Team (2018),Yung, Wesley, Jukka Karkimaa, Monica Scannapieco, Giulio Barcarolli, Diego Zardetto, José Alejandro Ruiz Sanchez, Barteld Braaksma, and Joep Burger. "*The Use of Machine Learning in Official Statistics*," n.d., 1

UNITED NATIONS DEPARTMENT FOR ECONOMIC AND SOCIAL AFFAIRS. DESA. UNITED NATIONS *E-GOVERNMENT SURVEY 2020: Digital Government in the Decade of Action for Sustainable Development.* S.l.: UNITED NATIONS, 2020.

Veale, Michael and Brass, Irina, *Administration by Algorithm? Public Management Meets Public Sector Machine Learning* (2019). In: Algorithmic Regulation (Karen Yeung and Martin Lodge eds., Oxford University Press, 2019), Available at SSRN: https://ssrn.com/abstract=3375391

Veale, Michael, Max Van Kleek, and Reuben Binns. "*Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making*." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, April 21, 2018, 1–14. https://doi.org/10.1145/3173574.3174014.

Wanner, Jonas, Herm, Lukas-Valentin and Janiesch, Christian. "*How Much is the Black Box? The Value of Explainability in Machine Learning Models.*." Paper presented at the meeting of the ECIS, 2020.

Yu, Han, Zhiqi Shen, Chunyan Miao, Cyril Leung, Victor R. Lesser, and Qiang Yang. "*Building Ethics into Artificial Intelligence.*" In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 5527–33. Stockholm, Sweden: International Joint Conferences on Artificial Intelligence Organization, 2018. https://doi.org/10.24963/ijcai.2018/779.

Zach Friend (2013) *Predictive Policing: Using Technology to Reduce Crime*, Retrieved July,7 2021, from https://leb.fbi.gov/articles/featured-articles/predictive-policing-using-technology-to-reducecrime?__cf_chl_jschl_tk__=a2e182d81df0d519526af0dba281e6554b7481db-1625708840-0-AVh2B0ML-3USts27YOM86p7FsnEdDey2XIbNAi7HbdOM5uPncDJXIkXNCP6gBLg-PltXH8dpv3mHP2zJlePgZyfTOn fwQAWow0wiI4_VD6oheWxQljKi2CVlJGf2bCFk0uljB09yhgWlmkZzFKTSKB-mEd20O5_cjsAVxsWeVk570Z-sLA6Q6utbXVwVXpW_ETIc-NbOD78oLZUya2TWhxavb9yOEZX4a6YOobbxki5uKlSdvuEkOFEpI0fGZSi Vhkhe42UKrI8vZjBAf8MexHHukHeMHaHNs3gbRli2X8fhUnZriATnx-AGApzhVp3Dyp6l4Pcwpg1BZ7oHpucgXSWSfvshY-yiRD_L4BU_ZtaTtv7m4EV3Q_AYG79l1APOM6Z1f6ZiI0qrMmJMb_ouh74DlU0bFAKAllwaoUShgju-tbSv3bzJdosBpvliPenGPqI6Vgiw7vSRlyv2YdKnqT4UHV03KZk1AIkIg2DxOEhGEgiIMwf1zvvrel-jtsV5HTTw1Eke4Thv1cHAZ2ZqztYWbcLof1p2QZv8PpmsozKdKAByn WyfDXVrnedWjRjHEV57Rw

Zuidermijk, Anneke, Yu-Che Chen, and Fadi Salem. "*Implications of the Use of Artificial Intelligence in Public Governance: A Systematic Literature Review and a Research Agenda*." Government Information Quarterly, March 2021, 101577. https://doi.org/10.1016/j.giq.2021.101577.

About the Author

Yulu Pi

Yulu Pi has a background in data science and economics and is working towards her Ph.D. in Centre for Interdisciplinary Methodologies, University of Warwick. She also worked in several governmental and international organizations before. Her research is focused on the development of interactive algorithmic systems to encourage wider adoption of AI in the public sector.