Research Article

On the utilization of Disruptive Technologies for Municipality-wide Policy Making

Maria Tsourma*[iD], Anna Kougioumtzidou, Anastasios Drosou and Dimitrios Tzovaras

Information Technologies Institute, Centre for Research and Technology Hellas (CERTH), Thermi-Thessaloniki, 57001, Greece, Greece.
*Corresponding author. E-mail: mtsourma@iti.gr

Abstract

Disruptive technologies (DTs) aim to introduce innovative elements with the immediate goal of changing citizens' behavior and the optimization of certain processes in either the private or public sector. The sector, especially "e-governance" has adopted several DT-based systems and is open to exploring new technologies that could be potentially useful for the sector's administration. This paper proposes a DT-based framework for the early detection of misinformation, the identification and prioritization of emerging and significant issues that refer to the municipality, and the recommendation of counter-actions that could be applied for each identified issue. The proposed framework aims to be applied initially at the municipal level as this is a more limited community for the framework's study and evaluation. News feeds and information posted on social media platforms will be the key information included in the proposed framework, aiming at sensing the public opinion and extracting issues and topics that concern a municipality's citizens. Through social media platforms, users tend to post, among other, concerns or issues they face within the local community, and thus useful content can be extracted and used for the identification of emerging and significant issues. The main advantage of this framework is the utilization of information from different data sources and their processing in order to immediately inform the relevant authority about the most important issues, allowing them to act in a proactive way. The expected outcome of the proposed framework is to raise the citizens' awareness of the issues of the local society and restrict criminality and violence in the municipality.

Policy Significance Statement

The Disruptive technologies (DTs)-based framework proposed in this paper will leverage DTs, aiming to propose a toolset for municipalities, in order to protect themselves and their citizens against misinformation. The proposed framework has some significant advantages, as it helps to inform the competent authorities in a timely and valid manner on issues concerning the local community. The time-consuming process of data collection and evaluation is automated through the proposed framework, enabling the competent authorities to have more time to make decisions and plan their counter-actions. In addition, this framework does not have any restrictions, as it can collect information concerning a municipality from multiple sources, and extract and prioritize any issues referring to the local community, including issues concerning violence and crime or social issues such as immigration.

1. Introduction

The so-called Disruptive Technologies (DTs), including technologies such as Artificial intelligence (AI), Internet of things (IoT), Natural Language Processing (NLP), Virtual and Augmented reality (VR, AR), big data and block chain (Ronzhyn et al. (2017)), show nowadays the possibility to increase dramatically the efficiency of the business processes typically operated by Public Administrations (PAs) and public service providers. These technologies allow the enhancement of data and information used for decisions making, as well as the automation of processes. At the same time, DTs are often interaction-based solutions developed by making extensive use of Learning Functions[1], and this makes them particularly prone to ethics-related issues.

DTs have the potential to transform public sector, providing numerous benefits to society. McKinsey in World Bank Group (2017) estimates that big data can reduce the EU's administrative costs by 15 to 20 percent, while Deloitte's simulations in Viechnicki et al. (2017) shows that the adoption of AI has the potential to save government workforces by 30 percent of all their labour hours. New technologies have the potential to improve efficacy, speed and cost of public services but they also introduce new challenges and security concerns related with the accuracy of decisions of DTs driven systems. The transforming capabilities of DTs can unlock a range of benefits across the entire public sector spectrum, achieving a dual objective; improve quality and speed of frontline service delivery, meeting citizen's needs and expectations, while at the same time optimize work processes, increase productivity and cost efficiency.

The public sector, in general, is somewhat reluctant to adopt DTs and related innovations as compared to the private sector and the entrepreneurial community. This can be attributed to the fact that the public sector plays an important role in the society and the economy, having the function of service provider, employer, and regulator, and therefore, the room for experimentation and testing of new technologies is limited, as is the room for error. The public sector requests the adoption of technological solutions with a clear scope, proof of concept, and a successful track record of application, as well as with clear and documented impacts and implications in ethical, political, socio-economic, legal, cultural, and technical terms.

On the other hand, local authorities across Europe are striving to address the modern societal, economic, and practical challenges of their local communities. The most common critical factors in municipalities are management and organization, technology, economy, built infrastructure, environment, communities, policy framework, and governance. Among the aforementioned factors "e-governance" is considered as the most critical challenge as through the public network different stakeholders are responsible for policies' definition and services' provision.

In the context of "e-governance", most of the municipalities are equipped with online communication infrastructures for their citizens. In the era of information explosion, a large amount of public discussion/deliberation and news broadcasting has now moved on digital news and social media platforms. Thus it becomes apparent that massive amounts of incoming information, sensing the public pulse, concerns, and emerging issues can nowadays be collected and processed in automatic ways using big data and AI technologies. DTs can be leveraged at this point concerning the address of information collection issues, such as the safe-marking of trusted sources, avoidance of false information dispersion, detection of fraudulent accounts identification and prioritization of emerging and significant issues, application of counter-actions with minimum cost, effort and time, etc. with the use of the appropriate information handling and diffusion through online social media platforms.

This paper presents a DT-based framework that can be adopted by the public sector for the early detection of misinformation, the identification and prioritization of emerging and significant issues, and the application of counter-actions with minimum cost.The identification of trusted information,

---

[1]Class of algorithms and models that allows a machine to infer how to perform a task from a given set of examples, thus without executing explicit instructions. Examples of Learning Functions are Decision Trees, Neural Networks, Support Vectors machines, etc.

analysis and extraction of emerging issues is a time-consuming process, that takes place in daily basis on a municipality and in any public services, in order to collect all the available information concerning significant issues existing within a municipality. The collected data may include false information that may lead to distortion of public opinion, use of resources in a wrong manner and lack of immediate application of counter-measurements to demanding circumstances.

The retrieved data from the aforementioned sources are at a next step delivered to the corresponding authority within the municipality in order to be processed and used in decision-making. The corresponding personnel of the municipality is responsible for the decision of which pieces of the collected information is important and for which authority. Due to the nature of this process, and according to the fact that the personnel of the municipality is responsible to decide which topics are significant, which should be distributed or not, this leads to the creation of human-induced bias and the elimination of transparency. Therefore, and due to the human-induced bias, the corresponding public authority does not have an overall picture of the issues concerning the citizens, because it receives only the information passed from the corresponding personnel.

By leveraging DTs and the features that each DT has available, the proposed DT-based platform aims to automate these processes by collecting data from news feeds, social media platforms, and also receive feedback from citizens, analyze these data and identify, extract, and prioritize emerging and significant issues mentioned in these data sources. The automation of emerging and significant issues extraction is not only important because it will reduce the execution time of this process, but it is also important in preventing serious issues related to violence and crime. Emerging topics/issues concerning violence or crimes will be detected in a pro-active way by analyzing the views and local authorities will have the time required for trying to mitigate them with the appropriate social actions. Furthermore, this framework tends to be used also for improving migration reporting, through the analysis of public opinion and by forming privacy preserving groups of opinions and views. Alongside, the automation of these processes aims to achieve the minimization of the prejudices associated with the introduction of the human factor at all stages of the processes. Thus, the information collection processes become easier and faster, they are evaluated and the false information is discarded and only the correct data are used for decision-making. Through the adoption of the proposed framework by a municipality, we aim at achieving transparency in the decision-making process, the direct communication of the public with the respective authority and the optimization of the elements used in decision-making in all procedures of a public service.

The main objective of the proposed framework is to sense the public pulse and the needs of its citizens both in important problems, such as criminality and violence, or in smaller issues that may exist, from news feeds and social media platforms, aiming at helping the municipality to apply counter-actions with minimum cost. In addition, the improvement of migration reporting is an important issue that could be handled by the proposed DT-based framework. One of the major challenges that this framework has to tackle concerns the detection and elimination of false information, i.e., misinformation or disinformation, the detection of fraudulent accounts and the investigation of the propagation of misinformation among networks. The DT-based framework presented will be able also to detect news that contain biases or promote propaganda. Misinformation might disorient the public sentiment, might upset locals and also disorient the municipality and therefore hinder decision-making.

This framework can be applied in public sector, without requiring any specific skills or knowledge for its utilization, since all extracted information will be provided to the end users in order to be used by the corresponding authorities for decision-making and counter measurements planning. At this point it should be mentioned that the proposed framework is not responsible for the decision-making processes, however its main scope is to provide the competent authority with the appropriate and correct information available on any issue raised in order to be informed and take timely measures to resolve them, without wasting valuable time and resources.

The proposed DT-based framework has several advantages. Firstly, with the use of big data and the identification of trusted-sources functionality, the proposed framework is a new initiative in policy-making and regulation. It facilitates digital interaction between administrations and citizens/businesses

for high-quality public services. Furthermore, through sensing of public opinion, emerging and significant issues can be raised directly, thus giving the municipality the flexibility to act directly. In this context, the time-consuming task of information retrieval and analysis in order to extract emerging issues from the local community is withdrawn from the municipality's responsibilities, giving in this way more time to the municipalities to plan and execute the counter-actions required for each emerging task. Alongside, through the use of this framework, citizens' awareness of general issues and problems could be raised, fostering positive behavioral changes. Environmental sustainability could be favored by boosting citizen's trust and confidence in public authorities and services and enhancing public involvement, engagement, and participation.

Concerning the environment, the introduction and deployment of DTAs in the public sector through the stimulation and uptake of the proposed framework will have a significant impact. More specifically, the improved management of incoming information reported through online social media platforms results in a more reliable information about public services (e.g. waste management, transport, water and energy systems), translated into efficient urban planning and design as well as natural resources management, also avoiding unnecessary vehicle usage by public servants and reducing carbon footprint of the public sector.

The rest of the paper is organized as follows: Next section presents related work. In section "Proposed framework", the overall concept and the objectives of the proposed framework are presented. The third section presents the ethical, social, and legal aspects that will be investigated within the proposed framework. Lastly, the final section summarizes the main conclusions and includes topics for future work.

2. Related work

Disruptive Technologies are cutting-edge technologies that allow gauging and satisfying some latent needs of a community of users. Their adoption typically follows Christensen et al. (2013) "Disruptive Innovation" theory. According to such theory, disrupters start by appealing to low-end or unserved markets and then move to the mainstream one as soon as the quality of their products or services catches up to their standards.

The opportunities and advantages of introducing DTs in the public sector have been extensively studied and presented in several research papers Wirtz et al. (2018). Petriv et al. (2019) have presented a study aiming to determine the concerns of stakeholders as well as the major limitations existing concerning the chatbot development in public sectors. The outcome of this study was based on a comprehensive analysis of the state-of-the-art DT methodologies and twelve interviews with experts in Ukraine who were involved in the LvivCityHelper [2] bot project. A similar study concerning the use of DTs in German municipalities aiming to examine any challenges existing of its adoption from employees' perspective is described by Schaefer et al. (2021).

The public sector can leverage DTs in several fields of application. DTs can be applied and assist municipalities in different sectors as presented in Wirtz et al. (2019). Herrera et al. (2018) in their paper present a framework for fraud intention recognition of public transport bus operators based on DTs. In particular, this framework permits the recognition of the fraud intention in the operators delivers account time, based on date, route, count of passengers on board, among other features. Another application proposed in the literature for public sectors is the one presented by Hengstler et al. (2018) concerning a new field, the use of autonomous vehicles in traditionally human-executed tasks.

Knowledge Management systems aiming at the generation and knowledge sharing are presented by Lin et al. (2018). The authors of this research present a system based on DTs for the automation of clinical records documentation. Furthermore, process automation systems that interact with the users and software that can perform tasks for humans have been adopted by the public sector. In this context,

---

[2]This project concerns the implementation of a Lviv City Helper - a chatbot of the City Council of Lviv, which quickly and easily gives access to city public information in 24/7 mode.

Zheng et al. (2018) have presented a system for tasks allocation according to the respective area of responsibility that could be applied in the public and private sector. Alongside, Power (2016) provide a software combined with big data, advanced analytics, and identity to automate risk identity checks aiming at restricting violence in a region.

Available studies, produced both by academic and research organizations and by institutional studies, on the responsible adoption of DTs are still mostly theoretical (Patil et al. (2018); Princeton (2017); Siau et al. (2018); EC (2019)). However, scientific advancements and practical solutions in this domain are fundamental to unlock the use of DTs in line with the principles recommended at European level (which are often shared at national level), including, among others, the attention to accountability, privacy, sustainability, trustworthiness, security, transparency and interpretability, openness, fairness, safety, or the respect of human agency. Looking upon these aspects reduces significantly the exposition to unnecessary risks of malicious exploitation and manipulation and unforeseen consequences thus, taking full stock of the DTs potential.

3. Proposed DT-based framework

3.1. Concept and objectives

In general the isolation of trusted information, analysis and extraction of emerging issues is a time consuming process in a municipality and in the public sector. The proposed DT-based framework is oriented towards municipalities and local authorities, since it is a toolset to protect themselves and their citizens against misinformation. This framework aims at the automation of these processes by collecting data from news feeds and social media platforms, analyses them in order to identify, extract, and prioritize emerging and significant issues mentioned in these data sources.

Another source of information will also be the citizens themselves, who will be encouraged to be involved in this process as their point of view and vision should reach the respective public authority. Citizens will be able to interact with this platform and have the ability to post any issues related with the municipality or the public service and provide any information required without having to spread them in social media.By allowing all citizens, independent of status and group, to express their opinion and report any issues they face in the municipality, the reduction of discrimination between different groups of citizens is achieved, vulnerable populations, if their voices are not heard from the online sources used in the system, are thus allowed to introduce their issues and make suggestions via the proposed platform. The information provided by the users will be posted anonymously and it will be also processed with the rest of the data collected from social media and news articles.

The automation of emerging and significant topics extraction is not only important because it will reduce the execution time of this process, but it is also important in preventing serious issues related to violence and crime. Emerging topics/issues concerning violence or crimes will be detected in a pro-active way by analyzing the views and local authorities will have the time required for trying to mitigate them with the appropriate social actions. Furthermore, this framework is aimed to be used also for improving migration reporting, through the analysis of public opinion and by forming privacy preserving groups of opinions and views.

The main concept of the framework proposed is displayed in Figure 1. It is based on four main steps, each of which contributes to a different level to achieve the framework's objectives. A detailed description of the information flow and the interconnection and data exchange among the steps is presented in detail in subsection 3.2.

The recipients of this framework will learn how to leverage AI and information retrieved for multiple data sources to improve migration reporting and restrict criminality and violence, in compliance with all legal and ethical standards. This will increase confidence and trust in AI-related breakthroughs for the public good, paving the way to unleash the whole potential of the application of DTs in the wider society.
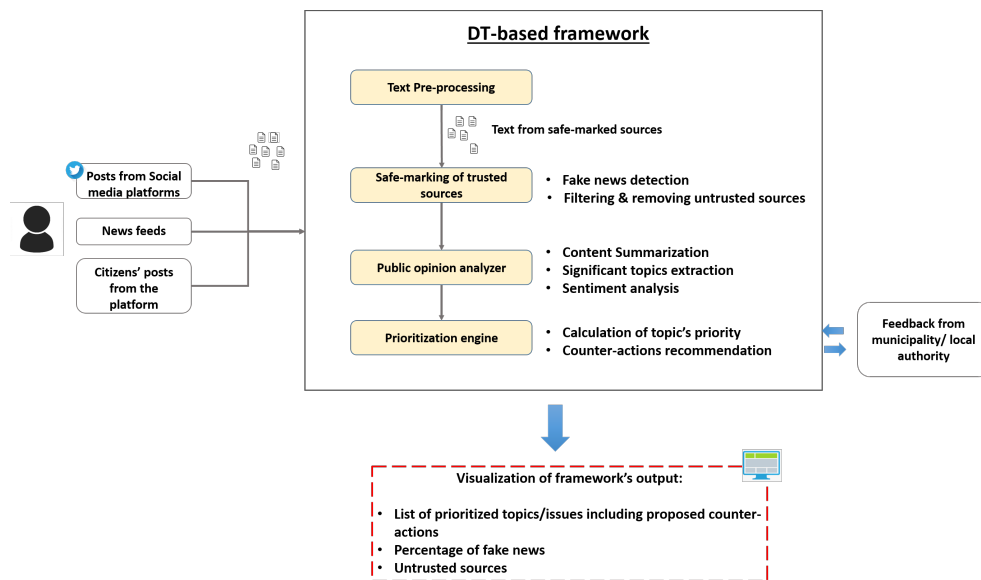
Figure 1. *Information flow according to the concept of the proposed framework..*

For the implementation of the proposed framework to be used as a real-life system, specific algorithms and methods are required, based on the characteristics that will be used. The DT-based framework will be accompanied by a user-friendly User Interface (UI) allowing data input, settings configuration, and visualization of each component's output. Furthermore, through its interface the citizens will also be able to provide their feedback and send information concerning topics and issues related with the municipality to the corresponding public authority. A central database is also required for data storage and data exchange among different steps. Furthermore, a system for the data retrieval from multiple data sources, e.g. different social media platforms and news feed sources needs to be available.

3.2. Conceptual architecture

The proposed DT-based framework is based on four main steps, as depicted in Figure 1, which exchange information with each other aiming at the unobtrusive execution of the main tasks in order to achieve the objectives of the framework. The proposed DT-based framework will allow the continuous exploitation of the information received from social media platforms, news feeds and information posted from citizens through the presented platform. Before applying any method to the collected information, data retrieved from the sources defined will be processed in order to be used in the next steps of the procedure. The text preprocessing step of the DT-based framework pipeline will employ Natural Language Processing (NLP) techniques such as the removal of unnecessary textual elements, lemmatization, tokenization and use of word embeddings, always in accordance with the specifications and demands of the model used in the next steps.(Mikolov et al. (2013);Nayak et al. (2016))

The preprocessed text will then be inserted as input in the safe-marking of trusted sources step. This steps concerns the identification of mis-information and of untrusted sources, aiming at eliminating them. Through this process, web pages that include mis-informative content will be detected early and will be removed from the list of valid sources. Textual content-based approaches that rely on linguistic lexical and syntactical features and thus capture deceptive cues or writing styles, propagation-based approaches that are founded on the premise that fake news propagate differently from true news

forming spreading patterns similar to those of infectious diseases, and potentially social context approaches that leverage user demographics, reactions and connections between users will be evaluated to safe-mark trusted sources (Zhang et al. (2019)). This process is related with ethical aspects, and in particular with fairness, accountability and transparency. The automated detection and/or removal of false or dubious content coming from news feeds and social media is prone to bias risks arising from multiple sources (e.g. incomplete or unrepresentative training data, bias in the engineering and training processes) that might cause the incorrectly flagging of a truthful statement as false, embedding and amplifying discrimination and unfairness in public sector practice.

False information detection is a challenging problem in deception detection with real-world political and social impacts. Challenges include that the detection often involves mostly short statements, as well as the fact that the interpretation of the news requires the knowledge of political or social context or "common sense", which current NLP algorithms are still missing. To tackle these challenges three main categories of detection methods have emerged; methods based on content, on social context, and propagation-based methods (Zhang et al. (2019); Wang (2017)). Within this scope, AI graph–based algorithms will be used in order to detect any false information available in the retrieved data.

The filtered list of text provided as output from the safe-marking of trusted sources step is merged and summarized in the so called "public opinion analyser" step. The summarized text is used in the next step for the extraction of topics included in the retrieved text. Automatic text summarization techniques based on transformer neural networks will be employed in cases when content needs to be summarized (condensed) and significant topic extraction will be performed using automated keyword extraction methods and topics clustering. Similarly to automatic text summarization techniques, state-of-the-art sentiment analysis techniques will be used, also primarily based on transformer neural networks, making thus this type of neural network the main applicable methodology in the "public opinion analyser" step. (Campos et al. (2020); Devlin et al. (2018); Zhang et al. (2019))

The advantage of the "public opinion analyser" step concerns the immediate and direct information of the municipality or local authority about the current topics discussed locally, by the analysis and extraction of emerging topics, and give to the authorities the time needed to consider taking counter measures if required. For example, if an emerging topic discussed in social media concerns a criminal case or a violent assault occurred, the system will be able to inform the municipality, providing all the required information, in order to take instantly the required counter-measures and help in the restriction of criminality. Alongside, when the topics concern important public issues, e.g. immigration or air pollution, local authorities can raise campaigns or start discussions according to the topics in order to engage citizen awareness.

From each topic extracted, a sentiment analysis will be performed. The purpose of this procedure is to form privacy preserving groups of opinions and views for the current topic. The sentiment of each topic will be also one of the parameters taken into consideration for the prioritization of the extracted topics. Sensing public opinion is an important asset not only for the prioritization of topics or emerging issues, but also for the municipality.

A list of topics, including the each topic's sentiment and text along with the emergency score, are provided as input in the Prioritization Engine. The purpose of this system is initially to calculate a priority score, using AI methods, in order to depict how emergent and severe a topic is according to the data collected. The more emergent the issue, the more direct and targeted should be the counter-actions required. Concerning the significance of the extracted topics, it will be calculated based on information previously provided by the municipality, also using weights per topic.

After the calculation of the priority score, the initial list of topics will be re-ordered according to this score. The final list of topics can be used as input on the counter-actions recommender supported, aiming at providing recommendations of counter-actions that should be taken on each topic. Historical counter-actions used in the past for similar topics will be also provided as input in the decision-making tool, in order to allow the selection of the most suitable counter-action. In this context, machine learning algorithms will be explored and utilized, in order to assess the data sources provided. The final output of the Prioritization Engine will be a list of prioritized topics/issues including also proposed

counter-actions. Feedback from the municipality's personnel will be provided in the Prioritization Engine, in order to be adapted and produce an output of higher quality, resulting in the development of an evolving intelligent system. At this point, it needs to be highlighted that the corresponding authorities will conduct the final decision-making of the counter-actions that will be performed and they will also have the ability to change the significance level of an extracted issue based on their experience.

Aiming to provide an easily accessible system that could be used by all users without any restrictions, a user interface (UI) will be available as mentioned previously. This UI will be accessible from the municipality's personnel, aiming to allow them receive information about the eliminated sources, the most significant and frequently discussed topics about the municipality, in order to have available all the data required in order to take the appropriate actions and make any decisions. The list of the prioritized emergency topics will be depicted including also a scale displaying how urgent a topic is. Counter-actions recommended for each topic will be also available. Through this UI, users can provide feedback to the Prioritization Engine and perform any correction actions on the results provided by each step, such as the indication of a trusted source as untrusted, the manual update of the prioritized topics, etc.

4. Expected impact at the policy level

The adoption of DTs leads to clear problems in terms of accountability and transparency of the policy making process. Especially in the case of artificial intelligence, it is important that decisions are properly explained and communicated, and that policy makers can be considered responsible for the decision taken.

The introduction and deployment of DT-enabled solutions in the public sector through the stimulation and uptake of the proposed framework's implementation is aimed to have a great policy impact. First, the framework will define a governance model of disruptive technologies aimed to gain a comprehensive understanding of how different types of disruptive technologies transform public administrations on the inter-/intra-organizational and individual levels with respect to governance, engagement and participation, as well as other emergent transformational impacts. In this regard, the proposed solution will promote transparency and explainability of DTs in the public sector, therefore increasing the accountability of policy makers. This will aim to restore trust into policy making in general, and to public sector work processes, services, politicians and public servants in particular. The new solutions will lead to increased compliance with legal and regulatory obligations and standards. In general, the proposed framework will enact simpler, faster, more secure and trustable processes of the public services. Overall, the proposed framework will lead to an improvement in the quality of policy making - which becomes more evidence based - an increase in the collaboration across government bodies, as well as between government and citizens, increase the awareness of citizens on specific issues, as well as their information about government policies. The proposed framework fits in this context contributing to the achievement of the following policy objectives:

1. Contributing to the responsible adoption and use of DTs in public administrations;
2. Enhancing trust in digital public services by fostering ethical consideration to the adoption of DTs;
3. Advancing testing on the responsible digital transformation of public administrations;
4. Facilitating digital interaction between administrations and citizens/businesses for high-quality public services;
5. Contributing to new initiatives in policy-making and regulation.

As already mentioned in the introduction concerning the description of the process currently followed by the personnel of the municipality and due to the nature of the process and the fact that the estimation of the significance of the topics is currently assigned to humans, it is possible that human-induced biases are introduced and the transparency of the process is minimized. DTAs can automate tasks such as the estimation of the topic's significance, aiming at making this process more transparent, straightforward, and robust (Danks et al. (2017)). Yet, although algorithmic decision-making employed

by DTAs can simplify and enhance the relevant processes performed by humans, it may also contribute to the creation or the propagation of biases. The presence of biases in algorithmic decision-making, or in decision-making assisted by algorithms, may perpetuate discrimination based on sensitive attributes such as gender, race, religion, disability etc., leading thus to potential unfavourable treatment or even exclusion of persons and/or groups from services and benefits. In general, biases that appear in algorithmic decision-making systems can originate from hidden or neglected biases in the data (e.g., zip code might be indicative of race) or be introduced by the algorithms themselves (e.g., the order in which an algorithm processes data) or by the application of a system designed for one context in another context (e.g., with a different population). (Bender et al. (2018); Mehrabi et al. (2019); Srinivasan et al. (2021); Winfield et al. (2018))

Moreover, since nowadays algorithmic decision-making often involves machine learning, becoming this way "intelligent", some additional risks arise. In particular, neural networks - a Learning Function that has been used to achieve state-of-the-art results in many fields, such as computer vision and NLP – have been characterized as black box models, i.e., models that are not offering acceptable explanations regarding the rationale behind the output they produce. This inadequacy in explainability leads also to issues in establishing the transparency of the (decision-making) process and defining the accountability of the persons involved (Bellamy et al. (2019)).

Both biases and the deficiency of explainability in the automated/assisted decision-making lead to lack of reliability and trust in the quality and fairness of the produced results. This potential unreliability concerning machine learning algorithms (Learning Functions) renders human control and human capability to intervene in such automated processes an absolute necessity. Privacy preservation and protection of personal, potentially sensitive, data, is another relevant serious concern that is being addressed in the EU by General Data Protection Regulations (GDPR). And lastly, as in the case of conventional systems and humans, DT-based systems are also prone to manipulation, use for malignant purposes and being hacked. (Lepri et al. (2018); Shin et al. (2019); Winfield et al. (2018)). Even if the automation of process by DT is successful from a functional point of view, this still may entail the risk of increasing the unemployment, since people might lose their jobs if all tasks are automated. These people may lack the qualifications to take on a different job and/or experience stress by the process of being forced to do so.

To tackle the above risks, emerges the need to design human-centered algorithms and by extension human centered systems, along with applying end-to-end ethical design principles regarding both the process applied and the outcome. Essential elements of this type of design are the development of observable or explainable algorithms and adherence to the principles of fairness, accountability, sustainability, and transparency, as well as treating user needs, beliefs and preferences as an important component in shaping the overall system (Leidner et al. (2017); Shin et al. (2019)).

5. Conclusion

In this paper, a DT-based framework for safe-marking of trusted sources, identification and prioritization of emerging and significant issues, and recommendation of counter-actions that could be applied, by leveraging news feeds, information posted on social media platforms, and from the feedback posted from citizens has been proposed. The framework exploits dynamic information available from different sources and uses AI technologies to accomplish its purpose. It leverages this information for sensing the public opinion and improve the social impact in local communities. The framework offers some advantages by design, such as the use of big data and the identification of trusted-sources functionality, increasing citizens' awareness on general issues and problems, and foster positive behavioral changes. As a future work, we plan to implement the proposed framework and evaluate it within a municipality under long-term use. During the evaluation, the framework's usefulness, validity, the ethical and legal aspects of its adoption will be studied, along with the risk and mitigation indicators that need to be

taken into consideration with the adoption of such technologies. Furthermore, the possibility of using the framework in real-time can be explored.

The transforming capabilities of DTs demonstrated by the proposed DT-based framework will unlock a range of benefits across the entire public sector spectrum, achieving a dual objective; improve quality and speed of front line service delivery meeting citizen's needs and expectations, while at the same time optimize work processes, increase productivity and cost efficiency. The AI-based tools proposed are expected to improve the decision-making capabilities of the public authorities in the management of health-related issues, efficient urban planning and resource allocation management as well as management of migration flows by leveraging valid information and taking into consideration the public opinion. Furthermore, through the presented processes, citizens will be engaged and actively involved in community issues as their opinion is heard.

DTs are still in their infancy stage of development in the public service landscape and the emerging associated risks and challenges are relatively unknown, thus remaining widely unaddressed, potentially threatening the successful implementation of DTs in the public sector. Legal-regulatory, ethical, social, technological implementation-related barriers and risks should be considered in the future during the implementation of the proposed DT-based system.

Competing Interests. All authors declare none.

Data Availability Statement. Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Ethical Standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Author Contributions. M.T. and A.K. designed the framework, and wrote the first draft of the manuscript. A.D., D.T. revised the manuscript and approved the final version.

Supplementary Material. No supplementary material has been provided with the submission.

References

Bellamy R K E, **Mojsilovic A**, **Nagar S**, **Ramamurthy K N** , **Richards J**, **Saha D**, **Sattigeri P**, **Singh M**, **Varshney K R**, **Zhang Y**, **Dey K**, **Hind M**, **Hoffman S C**, **Houde S**, **Kannan K**, **Lohia P**, **Martino J and Mehta S** (2019) AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias *IBM Journal of Research and Development 63*(4–5).

Bender E M and Friedman B (2018) Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science *Transactions of the Association for Computational Linguistics 6*, pages 587-–604.

Campos R, **Mangaravite V**, **Pasquali A**, **Jorge A**, **Nunes C and Jatowt A** (2020) YAKE! Keyword extraction from single documents using multiple local features. *Information Sciences 509*, pages 257—289.

Christensen CM, **Raynor ME and McDonald R** (2013) Disruptive innovation

Danks D and London A J (2017) Algorithmic bias in autonomous systems *IJCAI International Joint Conference on Artificial Intelligence 0*(January), pages 4691-–4697.

Devlin J, **Chang M. W**, **Lee K and Toutanova K** (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding *arXiv preprint arXiv:1810.04805*

European Commission (2019) A definition of AI: Main capabilities and scientific disciplines *The European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG)*

Hengstler M, **Enkel E and Duelli S** (2018) Applied artificial intelligence and trust – The case of autonomous vehicles and medical assistance devices *Technological Forecasting and Social Change 105*, pages 105–120.

Herrera JLL, **Figueroa HVR and Ramírez EJR** (2018) Deep fraud. A fraud intention recognition framework in public transport context using a deep–learning approach *2018 International Conference on Electronics, Communications and Computers (CONIELECOMP)*. IEEE, pages 118–125.

Leidner J L and Plachouras V (2017) Ethical by Design: Ethics Best Practices for Natural Language Processing *In Proceedings of the First ACL Workshop on Ethics in Natural Language Processing* pages 30-–40.

Lepri B, **Oliver N**, **Letouzé E**, **Pentland A and Vinck P** (2018) Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges. *Philosophy and Technology 31*(4), pages 611-–627.

Lin SY, **Shanafelt TD and Asch SM** (2018) Reimagining clinical documentation with artificial intelligence. *Mayo Clinic Proceedings 93*(5), pages 563–565.

Mehrabi N, **Morstatter F**, **Saxena N**, **Lerman K and Galstyan A** (2019) A Survey on Bias and Fairness in Machine Learning

Mikolov T, **Sutskever I**, **Chen K**, **Corrado G and Dean J** (2013) Distributed representations of words and phrases and their compositionality *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 1389-–1399.

Monti F, **Frasca F**, **Eynard D**, **Mannion D and Bronstein M M** (2019) Fake news detection on social media using geometric deep learning *arXiv preprint arXiv:1902.06673*, pages 1-–15.

Nayak A. S and Kanive A. P (2016) Survey on Pre-Processing Techniques for Text Mining *International Journal Of Engineering And Computer Science 5*(16875), pages 16875-—16879.

Patil DJ, **Mason H and Loukides M** (2018) *Ethics and Data Science*. O'Reilly Media Inc.

Petriv Y, **Erlenheim R**, **Tsap V**, **Pappel I and Draheim D** (2019) Designing effective chatbot solutions for the public sector: A case study from Ukraine *In International Conference on Electronic Governance and Open Society: Challenges in Eurasia*. Springer, pages 320–335.

Power D J (2016) "Big Brother" can watch us *Journal of Decision systems 25*(sup1), pages 578–588.

Princeton (2017) *Dialogue on AI and Ethics*. Workshop Series on AI and Ethics.

Ronzhyn A, **Wimmer M A**, **Spitzer V**, **Pereira G V and Alexopoulos** (2019, September) Using disruptive technologies in government: Identification of research and training needs. *In International Conference on Electronic Government*. Springer, Cham, pages 276–287.

Schaefer C, **Lemmer K**, **Samy Kret K**, **Ylinen M**, **Mikalef P and Niehaves B** (2021) Truth or Dare? – How can we Influence the Adoption of Artificial Intelligence in Municipalities? *In Proceedings of the 54th Hawaii International Conference on System Sciences*.

Shin D and Park Y J (2019) Role of fairness, accountability, and transparency in algorithmic affordance *Computers in Human Behavior 98*(November 2018), pages 277-–284.

Siau K. and Wang W. (2018) Building Trust in Artificial Intelligence, Machine Learning, and Robotics *Cutter Business Technology Journal 31*(2), pages 47–53.

Srinivasan R and Chander A (2021) Biases in AI Systems *Queue 19*(2), pages 45–64.

Viechnicki P., **Eggers WD and Deloitte** (2017) How much time and money can AI save government?

Wang W Y (2017) "Liar, liar pants on fire": A new benchmark dataset for fake news detection *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics*, Proceedings of the Conference (Long Papers), *2*, pages 422–426.

Winfield A F and Jirotka M (2018) Ethical governance is essential to building trust in robotics and artificial intelligence systems *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376*(2133)

Wirtz B W and Müller W M (2018) An integrated artificial intelligence framework for public management *Public Management Review*

Wirtz BW, **Weyerer JC and Geyer C** (2019) Artificial intelligence and the public sector – Applications and challenges *International Journal of Public Administration 42*(7), pages 596–615.

World Bank Group (2017) Big Data in Action for Government : Big Data Innovation in Public Services, Policy, and Engagement

Zhang Y, **Zhao Y**, **Saleh M and Liu P J** (2019) PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. *International Conference on Machine Learning*, pages 11328–11339. PMLR.

Zheng Y, **Han Y**, **Cui L**, **Miao C**, **Leung C and Yang Q** (2018) SmartHS: An AI platform for improving government service provision *The Thirtieth AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI18)*, pages 7704–7711.