# Potential Risks of DTs in public sector

## The development of a Risk Framework for the adoption of Disruptive technologies within public organizations

Sara Mancini[1], Sven-Ove Hansson[2], Carolina Beniamina Rutta[3], Noemi Luna Carmeno[3], Giovanna Galasso[3], Barbro Fröding[4]

[1] PwC, Milan, Italy, [2] Karolinska institutet, University in Solna, Sweden, [3] Intellera Consulting, Milan, Italy, [4] KTH Royal Institute of Technology, University in Stockholm, Sweden

Corresponding author email: sara.mancini@pwc.com

### PROJECT OVERVIEW

ETAPAS, which stands for Ethical Technology Adoption in Public Administration Services, is a project funded by the European Union Horizon 2020 Programme which is carried out by 14 partners from 8 countries. The project aims at supporting public sector organisations to responsibly adopt Disruptive Technologies (DTs) - with a particular focus on AI, Big/Open Data and Robotics, to improve the services they provide in an ethically compliant manner. The aim will be achieved by introducing three main tools: (1) a Responsible Disruptive Technology Framework (RDT Framework), including a Code of Conduct (CoC) setting out social, legal and ethical principles for DTs' adoption in the Public Sector (PS), an extensive mapping of the associated ethical risks and social impacts, a legal review and a RDT indicators framework to practically measure these risks and related mitigation actions; (2) a Governance model which will allow other public organisations to replicate the methodology and adopt innovative solutions while monitoring risks; and (3) a prototypical software platform to automate the risks and mitigation assessment which will be tested within four real-life use cases adopting AI, robotics and Big/Open data technologies.

### 1. INTRODUCTION

This paper aims at introducing the ETAPAS Risk Framework which identifies the legal, ethical and social risks that may be associated with the introduction of AI, Robotics and Big Data in the public sector. In particular, we will explain the methodology which was used for structuring the Risk Framework, followed by the description of the identified risks.

### 2. METHODOLOGY

The Risk Framework is the results of a three steps analysis carried out as follows: (1) a thorough literature review addressing ethical, social and legal aspects of the disruptive technologies, (2) the development of a structured framework that categorizes these risks, and (3) the review of the framework. In the first step, different sources, both academic papers and grey literature, were analyzed to review the literature on the ethical, social and legal issues that the public sector faces

The ETAPAS project has received funding from the European Union's Horizon 2020 Research and innovation programme under Grant Agreement n°101004594

when dealing with disruptive technologies. In particular, we identified commonalities to consolidate the most common and important risks and issues related to Disruptive Technologies, spotting potential contradiction among different sources and tailoring the risks and issues to the specific context of the public sector. In the second step, we structured a draft of the Risk framework based on the analysis of the literature review. Finally, in the last step, the Risk framework was reviewed and validated by the ETAPAS consortium partners - and in particular the public sector organisations - for ensuring that the results were relevant and based on both science and practical experience.

## 3. RESULTS

The findings from the literature review and the validation process result in the final draft of Risk Framework, which identifies eight risk categories including 35 risks, as follows:

1. *Risks concerning direct interaction with humans,* such as the risks of physical and psychological harm*;*
2. *Legal Risk,* including liability risk and law infringement risk;
3. *Governance risks,* which comprises all the risks arising from mismanagement, weak supervision and lack of standard procedures;
4. *Enanced inequality and discrimination,* which covers risks related to possible perpetuation of discrimination with DT-based solution and the unequal access to and benefit from the DTs, due to economic, cultural and social differences;
5. *Errors and misuse,* including risks emerging from the DT lack of robustness, biased performances or security branches, as well as the risk of the DT being used as autonomous weapon, for malicious surveillance purposes or for spreading disinformation while limiting the decision-making power of the citizens;
6. *Security and data protection risks,* such as the risk of the DT being hacked to get information access or to change its behaviour;
7. *Unsustainable use,* stemming from high level of energy consumption or wide use of environmentally unsustainable materials;
8. *Workplace issues,* including job displacement and lack of competences by public sector employees.

## 4. FUTURE STEPS

The Risk Framework, together with the CoC, the literature review and the legal framework, represents the basis for further analyses in the ETAPAS project, in particular for the development of the Indicators Framework. The Indicators Framework will provide a practical methodology to assess the ethical, social and legal risks of a disruptive technology, such as AI and Robotics, within a public sector organisation, as well as the effectiveness of mitigation measures the organisation might put in place. Then, in the following project phase, a prototypical software platform will be designed and developed on the basis of the Indicators framework. The whole RDT framework, as well as the prototypical platform, will be tested and validated through the tailoring

The ETAPAS project has received funding from the European Union's Horizon 2020
Research and innovation programme under Grant Agreement n°101004594

and application to four real-life use cases involving the adoption of AI, Robotics and Big Data applications, namely:

- *Ethically Responsible Big Open Data* - focuses on NoiPA, a platform that offers HR shared services to around 100 Italian central and local public administrations;
- *Robot-mediated rehabilitation* - focuses on humanoid robots used for the assessment of patients' walking abilities;
- *Municipality chatbot Kar*i - focuses on Kari, an AI-based virtual agent adopted in more than 80 municipalities in Norway and Denmark to answer the citizen's questions on the municipality services;
- *Public Organizations Multi-factor Misinformation Handling* - focuses on the deployment of AI for fake news detection and prioritization of emerging issues in the municipality of Katerini (Greece).

Finally, a Governance model will be structured to provide guidelines for applying the ETAPAS RDT Framework and methodology every time a public sector organisation wants to adopt a DT application.

## REFERENCES


Boholm, M. et al., 2016, "The Concepts of Risk, Safety, and Security: Applications in Everyday Language.", Risk Analysis, 36 (2): 320–338.

Hansson, Sven Ove, "Risk", The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.).

Leitner, Christine, et al., "Disruptive technologies and the public sector: the changing dynamics of governance." Public Service Excellence in the 21st Century. Palgrave Macmillan, Singapore, 2019, 237-274.

Renda A., (February 2019), "Artificial Intelligence Ethics, governance and policy challenges - Report of a CEPS Task Force", CEPS, ISBN 978-94-6138-716-5, Brussels. https://www.ceps.eu/wpcontent/uploads/2019/02/AI_TFR.pdf

Ronzhyn, Alexander, Maria A. Wimmer, Vera Spitzer, Gabriela Viale Pereira, and Charalampos Alexopoulos. "Using disruptive technologies in government: Identification of research and training needs." In International Conference on Electronic Government, pp. 276-287. Springer, Cham, 2019.

Ubaldi, B., et al. (2019), "State of the art in the use of emerging technologies in the public sector", OECD Working Papers on Public Governance, No. 31, OECD Publishing, Paris.

Winfield, Alan (2019) "Ethical Standards in Robotics and AI." Nature Electronics 2, 46-48. https://www.nature.com/articles/s41928-019-0213-6



The ETAPAS project has received funding from the European Union's Horizon 2020 Research and innovation programme under Grant Agreement n°101004594