ETAPAS

Ethical Technology Adoption in Public Administration Services

# Potential Risks of Disruptive Technologies in the Public Sector

## The development of a Risk Framework for the adoption of Disruptive Technologies within public organisations

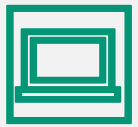
***Sara Mancini*** *presenting the ETAPAS Project*
*Italian Responsible AI lead in PwC Italy, collaborating with Intellera Consulting*

*The Culture of Trustworthy AI. Public debate, education, practical learning*

*2nd of September 2021*
*Venice International University*

Funded by the Horizon 2020
Framework Programme of the European Union

# Different trends pushing Public Bodies towards an ethical-aware adoption of Disruptive Technologies (DT)

| *Why is the Public Sector adopting Disruptive Technologies?* | | *Why should they consider ethics?* | |
|---|---|---|---|
| 1 | **To improve service delivery for citizens** | Citizens' demand for transparency and fairness should be addressed |  |
| 2 | **To be aligned with European and national strategies** | Legal and regulatory framework is evolving towards a risk-based framework |  |
| 3 | **To be resilient to the Covid-19 pandemic** | Accelerated digitalisation of the public sector stressed the importance of compliance ethical requirements |  |

# The ETAPAS Consortium

The ETAPAS Consortium is formed by 14 Partners from 8 different countries including:

- Public Administrations and Public Services Providers: Italian Ministry of the Economy and Finance (MEF), Municipality of Katerini (MUKA), Fondazione Don Carlo Gnocchi Onlus.
- Digital Innovation Hubs: CERTH, SINTEF, CEA List.
- Industry player: Intellera Consulting.
- IT SMEs: Prokom, 2021 AI.
- Universities and Research Organisations: KTH Royal Institute of Technology, Karolinska Institutet, University of Graz, Italian Institute of Technology (IIT).
- Think Thank association: the Lisbon Council.

# The ETAPAS Project at a glance

**Responsible Disruptive Technology Framework including Risk Assessment Methodology**

**Co-designed with Public Bodies, Digital Innovation Hubs and universities**

**Focus on three main Disruptive Technologies**

**Tested on four different use cases**

**Prototypical technological platform to be developed**

1 **Phase 1 – Research & Design**

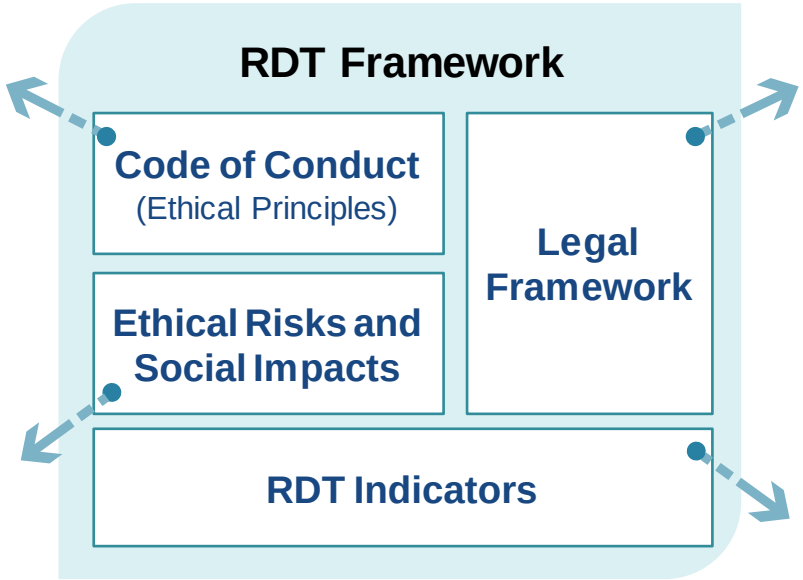
2 **Phase 2 – Test & Validation**

3 **Phase 3 – Synthesys & Impact Pathways**

# The ETAPAS Responsible Disruptive Technologies (RDT) Framework provides both the ethical and legal direction and the indicators to help mitigate the risks of DT-based applications

**ETAPAS Generic Code of Conduct for DTs in the PS provides 10 principles** to guide public sector organisations conduct. It can serve both as affirmations of values and as the basis of an internal accountability mechanism for the organizations.

The **ETAPAS Risk Framework** provides a detailed mapping of all the risks to which public bodies are exposed when adopting DTs.




**European Legal Framework** provides an analysis of Primary, Secondary and Case Law for DTs in the PS.

It describes identified gaps in binding and non-binding sources regarding the regulation of DTs in the PS and consideration of potential regulatory solutions.

The **ETAPAS RDT Indicator Framework** provide a practical methodology to evaluate & assess the impact of the DT application as well as if and how the risks are mitigated

# How we developed the ETAPAS Risk Framework?

## Step 1

1

**Literature review** addressing ethical, social and legal aspects of the disruptive technologies

## Step 2

2

Development of a structured **framework** that categorizes the **risks**

## Step 3

3

**Consortium partner consultation** to ensure the results were relevant and based on both science and practical experience

## Step 4

4

**Final review of the framework** to ensure consistency and put the focus on the undesirable outcomes

# How the Risk Framework fits into the whole ETAPAS RDT Framework?

**ETAPAS Generic Code of Conduct**

10 principles

**ETAPAS Risk Framework**

8 risk categories

34 risks

**ETAPAS RDT Indicators**

25 risk indicators

110 ca mitigation indicators

P1. Environmental sustainability
P2. Justice, equality, and the rule of law
P3. Transparency and explainability
P4. Responsibility and accountability
P5. Safety and security
P6. Privacy
P7. Building an ethical culture involving the employees
P8. Retaining human contacts
P9. Ethical public-private cooperation
P10. Continuous evaluation and improvement

1. Risks concerning direct interaction with humans
2. Legal risks
3. Security and data protection risks
4. Governance risks
5. Enhanced inequality and discrimination
6. Errors and misuse
7. Unsustainable use
8. Workplace issues

- Each indicator addresses **one or more risks**, but it is traced back to **one main risk category**
- Each indicator assesses compliance with **one or more principles**, but **one main principle** represents the **central ethical issue** addressed by each indicator.

# The ETAPAS Risk Framework – Risks concerning direct interaction with humans

**Risks concerning direct interaction with humans**

- Undisclosed technology use
- Replacement of human agency
- Exclusion of individuals
- Social isolation
- Psychological harm
- Physical harm

# The ETAPAS Risk Framework – Legal risks & Security and data protection risks

**Legal Risks**

- Illegal behaviour
- Liability risks
- Lack of certification procedures

**Security and Data Protection Risks**

- Rogue applications
- Hacking to change the behaviour of the DT
- Hacking to get information access
- Loss of service continuity

# The ETAPAS Risk Framework – Governance risks

**Governance Risks**

- Lack of transparency and explainability
- Use for unintended purposes
- Poor human oversight
- Unclear accountability and responsibilities
- Distrust
- Relations with the private sector
- Waste of resources

# The ETAPAS Risk Framework – Enhancing inequality and discrimination & Errors and Misuse

**Enhancing inequality and discrimination**

- Biased outcomes
- Unequal access and benefit
- Concentration of power
- Loss of cultural diversity

**Errors and misuse**

- Unreliable or poor-quality outcomes
- Autonomous weapons proliferation
- Malicious surveillance
- Disinformation
- Democratic deficit
- Citizen storing

# The ETAPAS Risk Framework – Unsustainable use & Workplace issues

**Unsustainable use**

- High level of energy consumption
- Use of environmentally unsustainable materials

**Workplace issues**

- Job displacement
- Lack of competences

# A bird's eye view on the ETAPAS Indicator Framework

## Risk indicator

assessing if the risk is relevant

*If the risk is present than mitigation questions are addressed*

## Mitigation indicator

evaluating the mitigation actions

*We foresee different types:*

- *Text question*
- *Numeric question*
- *Yes/No question*
- *Single choice question*
- *Multiple choice question*
- *Ratio Grid question*
- ***Direct feedback to the users***
- ***Computational***

**Example**

Does DTAinteract with the decision-making process of human end-users (e.g. recommended actions or decisions to be taken, presentation of options)? [Yes and it is fully autonomous; Yes, with a human-over-the-loop approach->; Yes, with a human-out-of-the-loop approach->; Yes, with a human-in-the-loop approach->; No]

- How does the users feel when taking decisions based on the DT application?
- Has your organization considered the task allocation between the DTA and humans for meaningful interactions and appropriate human oversight and control? [Yes; No]
- When the DTAis making a decision for which it is significantly unsure of the answer/prediction, is it able to flag the case and triage them for a human to review? [Yes; No]
- The relevant personnel will be able to assume control where necessary? [Yes; No]
- Does the DT solution provide sufficient information to assist the personnel to make an informed decision and take actions accordingly? [Yes, No]

# Testing the RDT Framework with the use cases: the process explained on the IIT & FDG Robot-mediated rehabilitation Use Case

## 1 Tailoring of the RDT framework to each use case

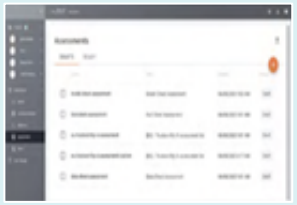
Example:
Let's consider the risk "psychological and mental impacts":

- ✓ **Addresses 1 or more ethical principles** such as Retaining human contacts; Safety & security
- ✓ **Requires 1 or more indicators to measure its impact** e.g. patient evaluation; presence of experts' oversights

## 2 Uploading of the RDT framework on the platform

Prototype Platform

The platform will both measure and compute some of the indicators, as well as provide a dashboard for the use case owner for monitoring purpose



## 3 Synthesis & Impact pathways

The feedback and inputs collected during the test cases, especially those from the PAs, will be used to refine the RDT framework itself and feed into the Governance model

The Governance model will provide PAs with guidelines for applying the ETAPAS RDT Framework and methodology every time a PA wants to adopt a DT application

# ETAPAS Project general information

For further information, contact me or get in touch with the ETAPAS Project team


*Send me email*
*Sara.mancini@pwc.com*

*Follow me*
*www.linkedin.com/in/mancinisara*


ETAPAS
Ethical Technology Adoption in Public Administration Services

***Send us email***
***etapas@etapasproject.eu***

***Follow us***
***ETAPAS Project***

***Follow us***
***@ETAPAS_Project***

***Follow us and subscribe to the newsletter***
***ETAPAS Website***

| ETAPAS | |
|---|---|
| ***Full Title*** | Ethical Technology Adoption in Public Administration Services |
| ***Programme*** | Horizon 2020 |
| ***Grant Agreement ID*** | 101004594 |
| ***Topic*** | DT-TRANSFORMATIONS-02-2018-2019-2020 Transformative impact of disruptive technologies in public services |
| ***Funding scheme*** | Research and Innovation Action (RIA) |
| ***Start date*** | 1st November 2020 |
| ***End date*** | 31st October 2023 |
| ***Cordis page*** | https://cordis.europa.eu/project/id/101004594 |

# Thank you!