

Ethics and Trustworthiness of Artificial Intelligence Enabled Systems in The Public Sector

Michael M. Farayola, Regina Connolly, Malika Bendeche, Irina Tal

1 RESEARCH CONTEXT:

Artificial Intelligence-Enabled Systems (AIES) are increasingly pervasive and revolutionizing services across different industries.

The challenges and issues associated with AIES deployment in high stake contexts are increasingly discussed in scholarly outlets to increase efficiency, reduce potential harm, as well as promote fairness, equity, accountability, and transparency in decision-making.

Particularly, in the judiciary system, AIES may entrench, amplify, and even obscure human bias and discrimination in the decision process something that underpins public concern.

However, to date, neither theoretical models nor fit-for-purpose AIES process models exist to guide the development of ethical and trustworthy AIES.

2 OBJECTIVE:

The goal of the thesis is to examine how the **development, test, and validation** of fit-for-purpose AIES in the public sector, particularly the judicial system, could be evaluated from a multidisciplinary perspective to provide insights into their ethics and trustworthiness for different AIES stakeholders (e.g., the adopting organizations, end-users/the public, regulators, auditor).

3 REQUIREMENTS OF TRUSTWORTHY AI:

- Human Agency and Oversight
- Technical Robustness and Safety
- Privacy and Data Governance
- Transparency
- Non-discrimination, Fairness
- Society and Environmental Wellbeing
- Accountability

4 FUTURE PLANS:

- Creation of a framework for the application of AI in the judiciary system
- Application of technical methods to ensure trustworthy AI such as:
 - Architectures for Trustworthy AI
 - Ethics and rule of law by design
 - Explanation methods
 - Testing and validating
 - Quality of service indicators

References:

- HLEG AI. 2019. High-level expert group on artificial intelligence.
- Yerlikaya, S. and Erzurumlu, Y.Ö., 2021. Artificial Intelligence in Public Sector: A Framework to Address Opportunities and Challenges. The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success, pp.201-216.
- Castelluccia, C. and Le Métayer, D., 2019. Understanding algorithmic decision-making: Opportunities and challenges. European Parliament.

HOST INSTITUTION



PARTNER INSTITUTIONS



FUNDED BY:

