

The State of Ethical AI in Practice: A Multiple Case Study of Estonian Public Service Organizations

Charlene Hinton, Erasmus Mundus PIONEER, Belgium*

ABSTRACT

Despite the prolific introduction of ethical frameworks, empirical research on AI ethics in the public sector is limited. This empirical research investigates how the ethics of AI is translated into practice and the challenges of its implementation by public service organizations. Using the Value Sensitive Design as a framework of inquiry, semi-structured interviews are conducted with eight public service organizations across the Estonian government that have piloted or developed an AI solution for delivering a public service. Results show that the practical application of AI ethical principles is indirectly considered and demonstrated in different ways in the design and development of the AI. However, translation of these principles varies according to the maturity of the AI and the public servant's level of awareness, knowledge, and competences in AI. Data-related challenges persist as public service organizations work on fine-tuning their AI applications.

KEYWORDS

AI Ethical Principles, Artificial Intelligence, Estonia, Ethical AI Design and Development, Ethics, Practical Application of AI Ethics, Public Sector, Value Sensitive Design

1. INTRODUCTION

Artificial intelligence (AI) has a deep potential to change various aspects of citizens' daily lives and of society as a whole. A systematic review of academic literature has shown growth in the uptake of artificial intelligence in the public sector (Gomes de Sousa et al., 2019; Berryhill et al., 2019; van Noordt & Misuraca, 2020). In Europe alone, the use of AI in public services is increasing, with over 230 empirical use cases identified (van Noordt & Misuraca, 2020). Researchers have noted that AI applications bring significant benefits to institutions that deploy them, from improving public services to reducing the costs and administrative burden (Mehr, 2017; Misuraca et al., 2020). However, these benefits are countered with sobering risks. Concerns for citizens' privacy and security, loss of decision-making autonomy, and unintentional harm that arise from AI systems may reinforce existing discriminatory practices (Sun & Medaglia, 2019).

DOI: 10.4018/IJT.322017

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

As a response to the risks, international organizations and institutions have increasingly advocated for the ethical design and development of AI. The results of their endeavors are realized through the introduction of ethical guidelines, standards, and governance frameworks, or soft law (Bartneck et al., 2021). More recently concrete actions toward operationalizing ethics have emerged in the form of legislative proposals for AI (EU Proposal AI Regulation, 2021). As technical developments in AI flourish, the ethics of AI persists as a contentious yet important discussion for communities, putting into question the human values that are deemed important by society.

Against the background of the multidisciplinary field of AI, empirical research on AI in the public sector has been inadequate (Sun & Medaglia, 2019; Zuiderwijk et al., 2021). Even less has been published about the practical implementation of the ethics of AI in this sector. Only a handful of empirical studies address the state of AI ethics in practice, and they have either focused on companies in the private sector (Vakkuri et al., 2020) or on a broad mix of both (Desouza et al., 2020; Ryan et al., 2021). Researchers note that in practice, most governments have a limited understanding of the implications of the use of AI. They hypothesize that insufficient research on empirical, context-based AI usage in governments can induce systemic failures that may negatively impact not only governments but also societies as a whole (Zuiderwijk et al., 2021). Therefore, this research aims to address this knowledge gap in the rapidly-evolving field of AI by addressing the following questions:

1. How do public service organizations ensure ethically-aligned AI public services in practice?
 - a. What are the key issues that public service organizations face in the design and development of AI?
 - b. In what ways are AI ethical principles considered in practice by public service organizations in the design and development of AI for public service delivery?

By answering these questions, this empirically-grounded research contributes to a broader academic discussion about the practical implementation of AI ethics and concurrently maintains focus on the under-researched public sector within the AI discipline. Furthermore, Estonia is chosen as the country context of study given its highly digitalized public services, its aggressive AI strategy, and the extensive collection of use cases of AI in the public sector. The rest of this research is organized as follows: Section 2 offers research background on AI in the public sector and the debates concerning AI ethics in practice. Section 3 presents the Value Sensitive Design framework used as the theoretical lens through which the research questions are addressed. Section 4 details the methodology used to prime the research analysis. Section 5 presents the empirical results that emerged from this analysis, the implications of which are critically discussed in Section 6. Finally, Section 7 concludes with a summary of the findings and future avenues of research.

2. BACKGROUND

2.1 Defining AI in the Public Sector

The ambiguity surrounding the definition of artificial intelligence continues to challenge researchers, practitioners, and policy-makers alike as there is still no universally accepted definition available for it (Grosz et al., 2016). A number of international organizations have offered definitions to address the ambiguity regarding the lack of a standard definition for what is meant by artificial intelligence when developing policy in the field. In particular, the European Commission, as of April 2021, presented a proposal for regulating AI. Because this paper inquires into the state of AI ethics in practice within the European context, it adopts the definition established by the European Commission in its proposal for regulating AI. Hereto, AI can be any “software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/the-state-of-ethical-ai-in-practice/322017

Related Content

Cyber-Bullies as Cyborg-Bullies

Tommaso Bertolotti and Lorenzo Magnani (2015). *International Journal of Technoethics* (pp. 35-44).

www.irma-international.org/article/cyber-bullies-as-cyborg-bullies/124866

The Context of IT Ethical Issues

Robert A. Schultz (2006). *Contemporary Issues in Ethics and Information Technology* (pp. 33-43).

www.irma-international.org/chapter/context-ethical-issues/7044

Belmont 2.0: Ethical Research Expectations for PII in AI, ML, and Data Mining/Scraping

Robin Throne and Michalina Hendon (2024). *Methodologies and Ethics for Social Sciences Research* (pp. 1-19).

www.irma-international.org/chapter/belmont-20/337047

A Moral Analysis of Effective Prediction Markets on Terrorism

Dan Weijers and Jennifer Richardson (2014). *International Journal of Technoethics* (pp. 28-43).

www.irma-international.org/article/a-moral-analysis-of-effective-prediction-markets-on-terrorism/108853

Digital Literacy: A Review of Literature

Pritika Reddy, Bibhya Sharma and Kaylash Chaudhary (2020). *International Journal of Technoethics* (pp. 65-94).

www.irma-international.org/article/digital-literacy/258971