Toward developing a Social Impact Assessment for AI in the public sector (SIA*4*AI) framework: key design considerations

SAMANEH BAGHERI[*], Open University, The Netherlands
VANESSA DIRKSEN[*], Open University, The Netherlands

Abstract

The rapid development and use of Artificial Intelligence (AI) systems in the public sector necessitate a structured approach to assess their societal impacts. This research focuses on identifying key design considerations for a Social Impact Assessment for AI (SIA*4*AI) framework tailored to AI's unique challenges and opportunities in the public sector. We address the limitations of existing AI impact assessment methodologies, which often take a top-down approach, focusing primarily on technical and ethical concerns while overlooking broader societal implications. Through an analysis of different assessment methodologies, we identify key design considerations to guide the development of a SIA*4*AI framework. This framework aims to ensure AI systems align with the values and priorities of diverse citizen groups, particularly vulnerable social groups, to promote societal benefits and mitigate potential harms of AI systems in the public sector.

Keywords: Artificial Intelligence, Societal Impact, Social Impact Assessment, AI Impact Assessment, Public Sector, Empirical Ethics

Introduction

The growing integration of AI systems into the operations of public sector organizations has the potential to enhance both extrinsic values, such as optimized resource allocation and improved service delivery, and intrinsic benefits, including the promotion of the public good and collective interests [1, 2]. AI has the potential to promote societal values, reflecting the needs and ambitions of local stakeholders. However, these advancements may also introduce unintended negative consequences and societal challenges, such as threats to inclusivity, equity, social well-being, and public trust [2-4]. To effectively address these societal challenges, a proactive social impact assessment is essential. Conducting impact assessments early in the AI system lifecycle helps identify potential issues and implement mitigation strategies [1, 5]. This approach allows public organizations, such as local governments, to ensure that AI applications align with the societal values and priorities of affected communities in the social domain.

To date, we see a sprawl of Artificial Intelligence impact assessments (AIIAs) being developed [5]. Nevertheless, current AIIA frameworks often fall short in addressing these broader societal concerns in the public sector. The state of AIIAs studies in the public sector reveals several key gaps [1]: (1) Only a few studies address impact assessments, primarily emphasizing normative rather than descriptive analyses. (2) Most studies are conceptual, with limited empirical accounts of AI's real-world effects. (3) Efforts to "measure" the actual impact of AI on public values are

[*] Both authors contributed equally to this research.

Contact Information: Samaneh Bagheri, samaneh.bagheri@ou.nl, Vanessa Dirksen, vanessa.dirksen@ou.nl, Open University, The Netherlands.

EWAF'25, June 30–July 02, 2025, Eindhoven, NL

lagging behind, often remaining at the level of "AI readiness" or serving as “promotional tools” for public value creation rather than conducting comprehensive impact assessments. (4) The perspectives and experiences of citizens and affected communities are underrepresented, and there is a notable lack of critical analysis regarding the broader societal implications of AI in the public sector. Fitzgerald and Taylor [6] emphasize the need for the development of well-defined, theoretically-based frameworks to assess the social impacts of AI, moving beyond simplistic references to social "issues" or "consequences." This research responds to this call by developing a structured Social Impact Assessment for AI (SIA*4*AI) framework that addresses the unique societal challenges and opportunities presented by AI systems in the public sector. It aims to evaluate the broader societal implications of AI in the public sector, ensuring that these systems deliver social benefits and address the needs of diverse local communities, particularly vulnerable and marginalized groups.

In this paper, we critically compare the similarities and distinguishing characteristics of the most prevailing existing impact assessment frameworks to understand what a framework for societal impact assessment of AI in the public sector should abide to. In so doing, we lay the groundwork for our SIA*4*AI framework by identifying key design considerations, making a critical first step toward its development.

Related Work

A growing number of AIIAs have been developed in recent years, varying in scope and focus [5]. Some AIIAs address technical issues, like explainability and bias, while others address broader ethical concerns. Moreover, most of these assessments are designed for the private sector, whereas AI in the public sector faces unique challenges. While businesses prioritize efficiency and cost savings, public organizations must balance these goals with the need to serve the public good, maintain public values, and address the needs of diverse stakeholders [7]. Therefore, AIIAs in the public sector cannot simply adopt tools and methods designed for the private sector [1].

Current AIIAs often focus on ethical and technical issues while ensuring compliance with relevant regulations. For instance, FRIA [8] and FRAIA [9] are designed to protect fundamental rights, while ALTAI focuses on the HLEG guidelines for trustworthy AI [10]. Furthermore, these frameworks typically rely on predefined ethical principles, such as the Z-inspection® process [11] and ALTAI [10], and often overlook broader societal implications. Moreover, they tend to neglect the specific needs of the public sector and lack meaningful community involvement. This lack of engagement limits their ability to capture a comprehensive understanding of its effects on diverse communities and stakeholders in the public sector. As a result, their evaluations may fail to reflect varied community viewpoints or effectively safeguard societal benefits. Instead, we propose a shift inspired by traditional Social Impact Assessment (SIA), which emphasizes community profiling, explores potential adverse impacts on societal value, and uses a large variety of participatory methods [12]. In this study, social impact encompasses a wide range of effects on rights, responsibilities, benefits, harms, justice, fairness, well-being, and the common good. It includes both proactively addressing societal issues and preventing negative outcomes. It refers to something that is experienced or felt in a perceptual or corporeal sense at the level of an individual, social unit (family/household/collectivity), or community/society [12, 13]. However, SIA-based approaches, in turn, require adaptation to capture the socio-technical complexities of AI implementation in the public sector. Our work addresses this gap by developing the SIA*4*AI framework, specifically designed for the public sector. It incorporates key considerations drawn from a comparative analysis of existing frameworks while emphasizing community engagement and a comprehensive and bottom-up understanding of social impacts.

Methodology

To understand what a framework for societal impact assessment of AI in the public sector should entail, we conducted a comparative analysis of existing impact assessment approaches. This includes both AI-specific impact assessment frameworks and broader SIA approaches. The findings from this analysis provided insights into the key design considerations for our SIA*4*AI framework. Our analysis covers some of the key relevant impact assessment frameworks that are implemented in multiple projects, including:

Fundamental Rights and Algorithms Impact Assessment (FRAIA): A risk-based assessment strongly substantiated on the protection of human fundamental rights. Applied for governmental contexts in the Netherlands [9].

Assessment List for Trustworthy AI (ALTAI): A risk-based assessment providing a checklist for self-assessment of Trustworthy AI systems, consisting of four ethical principles and seven requirements of Trustworthy AI [10].

Z-Inspection® Process: A Process- and risk-based approach to assess the trustworthiness of AI, in terms of technical robustness, legal compliance, and ethical compliance [11].

Generic Model for AI-IA: A structured approach to assess diverse impacts of AI from an organizational perspective [5].

Public value analysis (PVA)/Public Value Mapping (PVM): Evaluates public values in science policy and assesses the relationship between research activities and societal impact [14].

Societal Impact Assessment (SIA): "the processes of analyzing, monitoring, and managing the intended and unintended social consequences, both positive and negative, of planned interventions (policies, programs, plans, projects) and any social change processes invoked by those interventions" [13].

To conduct a structured comparative analysis, we focused on the following aspects:

Technology/AI Lifecycle: the assessment's applicability across all stages of an AI system, from development to deployment and post-deployment monitoring, to capture emerging social challenges at each lifecycle stage.

Social Impact Consideration: whether each framework explicitly covers social impact considerations.

Approach: whether the framework uses process-based methodologies (step-by-step assessment process), value-based assessments (underlying values that steer the use of technology), risk-based evaluations (identifying and mitigating technology-related risks), or a combination.

Evaluation type: if assessments are conducted before or after implementation, and follow a cyclical or linear process.

Methods of assessment: the instruments used for assessment (e.g., scenario analysis, stakeholder workshops).

Stakeholder engagement: whether each framework actively involves relevant stakeholders throughout the assessment.

Results

Table 1 provides the results of our comparative analysis of the impact assessment frameworks.

Table 1. Pros and cons of Impact Assessment Frameworks.

Framework	Pros	Cons
FRIAs [8]	Focus on protecting rights and ensuring compliance with fundamental rights. Process-based approach for rights-related risk assessment.	Missing broader societal impacts and public value considerations.
FRAIA [9]	Very comprehensive, combining a large range of assessment methods, amongst which FRIA, DPIA, FACT principles. Specifically developed for governmental contexts. Includes consideration for societal impacts as well, e.g., equality rights. Includes attention for value tradeoffs and mitigation. Ex ante evaluation type. Includes a large variety of stakeholders	A rather linear approach to evaluation. Almost too comprehensive to put into use.
ALTAI [10]	Risk-based assessment list aligned with EU values of Trustworthy AI.	While useful for regulatory compliance and ethical AI, it lacks the depth needed to assess the impact on public and societal values. It also lacks guidance on applying the assessment list, including implementation steps and responsible stakeholders [8].
Z-inspection® [11]	A systematic approach to assessing the trustworthiness of AI. Able to mobilize a large international pool of multidisciplinary expert panels. Cyclical evaluation type. Acknowledges that assessment requires continuous monitoring (dynamic). Includes a large variety of stakeholders (domain experts, vendors, governmental organizations). Applies a large variety of methods.	Better suited for in-depth evaluations of the trustworthiness of AI rather than public and societal value.
Generic Model for AI-IA [5]	A structured process and risk-based approach to assess diverse impacts (e.g., ethical, social) of AI. Takes a large variety of stakeholders into account, including vulnerable communities. Its iterative process allows continuous monitoring.	By aiming to cover all potential impacts, it lacks specificity for certain contexts, like the public sector.
PVA/PVM [14]	Emphasizes societal values while also highlighting public value failures (e.g., inequities, lack of transparency), Practice-based take on values [15].	Lacks detailed methodological tools for assessment, i.e., how to identify and reflect on public values [16]. It does not specifically address AI-related concerns.
SIA [12]	Broadly considers societal implications, uses well-developed notions of social impact to benefit local and marginalized communities; considers stakeholder perceptions/experiences, truly ex ante, context-awareness [17], participatory methods, meaningful participatory engagement [18], draws attention to long-term impact.	Its generality and the absence of detailed practical guidance for technology assessment limit its ability to address the challenges posed by AI in the public sector.

Based on this analysis, we identified several key design considerations for our SIA*4*AI framework:

(1) The assessment should go beyond technical and ethical concerns to incorporate a broad range of potential negative and positive social impacts, with a focus on 'for social good' such as empowerment and social inclusion (considering "many societal risks [and benefits] are not specific to AI, and it may be counterproductive to treat them as such" [19].)

(2) Employ an interpretive explorative approach to identify 'societal values' involved instead of relying on a pre-defined list of values, i.e., societal impact as experienced by relevant community groups.
(3) Context-sensitivity whilst 'delving deep' into a particular social impact domain.
(4) Active involvement of a diverse range of stakeholders, including vulnerable and marginalized groups, to reflect on the diversity of societal perspectives and values.
(5) Focusing on community impact rather than holding a project focus. "Too often, planned interventions are conceived, decided, and designed outside the locality of the intervention" [13].
(6) Forward-looking, anticipatory impact assessment, including both short-term and long-term impact, making use of participatory methods such as community mapping, scenario analysis, and stakeholder workshops.
(7) Continuous monitoring as values are not static [15].
(8) A critical reflexive stance towards the assessment procedure itself and the (unintended) consequences it may have ('assessing the assessment').
(9) Meaningful community engagement whilst maintaining a cautionary stance on citizen participation (as a form of window-dressing, 'educating the public', enforcing the AI system on citizens, and the like) [20].

Moving forward, by integrating these key design considerations, we will develop the SIA*4*AI framework in the next phase of our research, subsequently validating and refining it through public sector use cases.

References

[1] S. Bagheri, and V. Dirksen 2024. *Public Value-Driven Assessment of Trustworthy AI in the Public Sector: A Review*. Lecture Notes in Computer Science-Springer, City.

[2] Y.-C. Chen, et al. 2023. Artificial intelligence and public values: value impacts and governance in the public sector. *Sustainability* 15, 6.

[3] D. S. Schiff, et al. 2022. Assessing public value failure in government adoption of artificial intelligence. *Public Administration* 100, 3.

[4] B. W. Wirtz, et al. 2019. Artificial intelligence and the public sector—applications and challenges. *International Journal of Public Administration* 42, 7.

[5] B. C. Stahl, et al. 2023. A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review* 56, 11, https://doi.org/ 10.1007/s10462-023-10420-8.

[6] G. Fitzgerald, and C. N. Taylor. 2024. AI and SIA: some reflections. *Impact Assessment and Project Appraisal*

[7] K. C. Desouza, et al. 2020. Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. *Business Horizons* 63, 2.

[8] H. Janssen, et al. 2022. Practical fundamental rights impact assessments. *International Journal of Law and Information Technology* 30, 2.

[9] J. Gerards, et al. 2022. Fundamental rights and algorithms impact assessment (FRAIA).

[10] A. HLEG. 2020. Assessment list for trustworthy Artificial Intelligence from https://altai.insight-centre.org/

[11] R. V. Zicari, et al. 2021. Z-Inspection®: a process to assess trustworthy AI. *IEEE Transactions on Technology and Society* 2, 2.

[12] F. Vanclay, et al. 2015. Social Impact Assessment: Guidance for assessing and managing the social impacts of projects.

[13] A. J. Imperiale, and F. Vanclay. 2024. Re-designing social impact assessment to enhance community resilience for disaster risk reduction, climate action and sustainable development. *Sustainable Development* 32, 2.

[14] B. Bozeman, and D. Sarewitz. 2011. Public value mapping and science policy evaluation. *Minerva* 49.

[15] M. Boenink, and O. Kudina. 2020. Values in responsible research and innovation: from entities to practices. *Journal of Responsible Innovation* 7, 3.

[16] R. Huijbregts, et al. 2022. Public values assessment as a practice: integration of evidence and research agenda. *Public management review* 24, 6.

[17] I. Bianchi, and I. Tosoni 2023. *Culture and the City: Towards a Context-Aware Assessment Framework*. Springer, City.

[18] E. Smyth, and F. Vanclay. 2017. The Social Framework for Projects: a conceptual but practical model to assist in assessing, planning and managing the social impacts of projects. *Impact Assessment and Project Appraisal* 35, 1.

[19] N. A. Smuha. 2021. Beyond the individual: governing AI's societal harm. *Internet Policy Review* 10, 3.

[20] F. Lysen, and S. Wyatt. 2024. Refusing participation: hesitations about designing responsible patient engagement with artificial intelligence in healthcare. *Journal of Responsible Innovation* 11, 1.