6-14-2024

Socio-Technical AI Design For Public Value

Viviana Bastidas
University of Cambridge, vab44@cam.ac.uk

Jennifer Schooling
Anglia Ruskin University, jenschooling@gmail.com

SOCIO-TECHNICAL AI DESIGN FOR PUBLIC VALUE

TREO Paper

Viviana Bastidas, University of Cambridge, Cambridge, United Kingdom, vab44@cam.ac.uk

Jennifer Schooling, Anglia Ruskin University, Cambridge, United Kingdom, jennifer.schooling@aru.ac.uk

Abstract

The ambition to deploy AI systems in cities that create public value, align with the city's strategic goals, and adhere to ethical standards requires a holistic view of cities as socio-technical systems. However, AI projects in cities can often fail due to a predominant focus on technological aspects, neglecting the complex interdependencies within such socio-technical systems. This limited design perspective has led to AI-enabled solutions that present significant risks and challenges, such as potential bias and discrimination, privacy violations and citizen surveillance. This paper proposes a socio-technical digital architecture design of AI systems, utilising enterprise and domain-specific modelling. This approach can play a crucial role in designing urban systems by providing a structured approach to managing complexity and delivering social benefits while assessing and mitigating potential harm.

Keywords: Socio-technical urban design, artificial intelligence, public value, cities.

1 Introduction

City managers must create public value and achieve socially desirable outcomes for multiple stakeholders and communities (Bastidas, Oti-Sarpong, et al., 2023). The ambition to deploy AI systems in cities (United Nations, 2022) that solve society's problems, align with the city's strategic goals, and adhere to ethical standards requires a holistic view of such socio-technical systems from the perspective of different stakeholders (e.g. city leaders and professionals, citizens and communities, domain experts, data scientists, machine learning engineers, and IT professionals) (Yigitcanlar et al., 2021). Local authorities have started to explore current AI developments such as Large Language Models (LLMs) in various contexts. However, AI projects in cities can often fail because they are predominantly focused on one aspect of the system, commonly technology, and neglect to conceptualise and design the complex interdependencies that exist between people, strategies, policies, regulations, processes, and physical infrastructure. In particular, LLMs are non-deterministic and prone to hallucination, which challenges their adoption in tasks that require robust and reliable responses (e.g. public service provision) (Cabrera, Paleyes and Lawrence, 2024). This has resulted in AI-enabled solutions that present significant risks and challenges such as discriminatory outcomes, lack of transparency, privacy breaches, algorithmic errors, and cybersecurity threats (AIAAIC, 2019). A strategy combining 'social' (urban governance and ethics) and 'technical' (digital technology) can enable designers of AI solutions to plan the development of AI-enabled systems that deliver public value while managing risk and potential social harms. Socio-technical digital architectures can be crucial in designing such urban systems by providing a structured approach to managing complexity and bridging the gap between city strategy and technology deployment (Bastidas et al., 2022).

This TREO paper aims to investigate how to design socio-technical architectures for AI in cities based on enterprise architecture and domain-specific modelling (Bastidas et al., 2022; Bastidas, Bezbradica, et al., 2023). It will explore how to augment existing enterprise modelling languages for cities with

concepts from the AI domain and associated ethics and responsible innovation requirements (e.g. risk management). Thus, this study aims to design socio-technical digital architectures of AI for cities and postulate the research question: "How can socio-technical digital architectures of AI for cities be designed to enable responsible public value creation?".

2 Methodology

This paper follows a design science research approach and research method (Peffers, Tuunanen, Rothenberger and Chatterjee, 2007) due to its relevance to the domain of information systems (IS). This research will use real-world case studies and apply semantic analysis (e.g. using knowledge graphs) to help city managers and planners examine how to align stakeholders' expectations of AI systems with fundamental human and ethical principles. Resulting socio-technical models and tools can help city planners, technology implementers, and citizens share a common understanding of AI solutions and thus plan their development and integration with existing urban systems. Leveraging enterprise modelling can allow public sector stakeholders to move beyond high-level AI ethics guidelines and progress toward an architectural design of the socio-technical aspects of AI (Theodorou and Dignum, 2020).

3 Socio-Technical AI Design for Public Value

This paper is grounded in a socio-technical framework for responsible urban digital innovation (Bastidas, Oti-Sarpong, et al., 2023) that enables public value creation by encompassing three interrelated dimensions: governance and management, ethical and responsible innovation, and digital and technical. Examining the implications of AI systems design for the public sector holds significant relevance and urgency, due to the challenges faced by government agencies regarding the exploitation of AI capabilities and the management of potential risks (Fatima, Desouza, Buck and Fielt, 2022). The primary focus of AI lies in the efficient handling, storage, exposition, and utilisation of data (Cabrera, Paleyes, Thodoroff and Lawrence, 2023). Expanding the perspective beyond the 'technical' dimension of the design of AI systems, we will investigate how the 'social' concept appears and is applied to designing AI systems and related digital architectures.

Building on the work of (Bastidas, Oti-Sarpong, et al., 2023), we introduce a conceptual 'socio-technical AI architecture framework for public value' to provide the first directions towards socio-technical AI design for public value. We define these design components as four distinct socio-technical layers: (i) the 'public value creation layer', (ii) the 'governance and management layer' (iii) the 'ethical and responsible innovation layer' and (iv) the 'digital and technical layer'. The 'public value creation layer' describes the value-oriented features that AI-enabled systems should contribute to society (people and the environment). The 'governance and management layer' presents strategic and operational components needed to create public value enabled by AI technologies. The 'ethical and responsible innovation layer' constitutes the societal risks, human rights, and impacts of implementing and deploying AI-enabled systems. The 'digital and technical layer' describes data, algorithmic models, applications, and technology components of AI systems. However, understanding the components of each layer and the interplay between 'social' (urban governance and ethics) and 'technical' (digital technology) aspects of AI system design that ensures a balanced approach which aligns with public value are still open research challenges. Addressing the challenge of bridging urban planners and AI developers to understand and integrate socio-technical requirements is crucial. Moreover, traditional (enterprise/conceptual) modelling approaches lack the 'social' dimensions of AI systems to meet the specific requirements of this domain. These challenges require further research to enable socio-technical design as a practical tool for guiding the responsible planning and deployment of AI in cities.

4 Discussion and Conclusion

This TREO paper proposes a new research agenda to address the limitations of current architectural approaches for conceptualising and designing AI systems in cities. The research agenda includes developing new socio-technical approaches for specifying and modelling AI-enabled systems. It introduces a novel socio-technical architectural perspective that offers a framework for managing city complexity and designing digital solutions aligned with public value. This alignment must act as a bridge, establishing a robust connection between urban planners and AI developers, and facilitating collaboration and understanding of the relationships between 'social' and 'technical' requirements. Such alignment is challenging and will require exploring in practice essential architectural concepts and tools for designing AI systems and elucidating the complex relationships within and among the proposed layers.

The envisioned research agenda is ambitious and seeks to redefine how city professionals and managers represent and describe their concerns and requirements regarding AI deployment. A socio-technical design approach can foster a shared understanding of AI systems among relevant stakeholders and sectors to ensure the establishment of processes for assessing and mitigating potential harms. This approach not only strengthens leadership capacity within local authorities and the private sector but also prioritises public value creation to ensure the sustained development of responsible AI for cities.

References

AIAAIC. (2019, June). "AI, algorithmic and automation incident and controversy (AIAAIC) Repository." URL: https://www.aiaaic.org/aiaaic-repository (visited on May 14, 2024)

Bastidas, V., Bezbradica, M., Bilauca, M., Healy, M. and Helfert, M. (2023). "Enterprise Architecture in Smart Cities: Developing an Empirical Grounded Research Agenda." *Journal of Urban Technology*, *30*(1), 47–70.

Bastidas, V., Oti-Sarpong, K., Nochta, T., Wan, L., Tang, J. and Schooling, J. (2023). "Leadership for responsible digital innovation in the built environment: A socio-technical review for re-establishing competencies." *Journal of Urban Management*, *12*(1), 57–73.

Bastidas, V., Reychav, I., Ofir, A., Bezbradica, M. and Helfert, M. (2022). "Concepts for Modeling Smart Cities: An ArchiMate Extension." *Business & Information Systems Engineering*, 1–15.

Cabrera, C., Paleyes, A. and Lawrence, N. D. (2024). "Self-sustaining Software Systems (S4): Towards Improved Interpretability and Adaptation." in: *2024 International Workshop New Trends in Software Architecture (SATrends'24)* (p. 5). New York, NY, USA: Association for Computing Machinery.

Cabrera, C., Paleyes, A., Thodoroff, P. and Lawrence, N. D. (2023). "Real-world Machine Learning Systems: A survey from a Data-Oriented Architecture Perspective." *ArXiv Preprint ArXiv:2302.04810*.

Fatima, S., Desouza, K. C., Buck, C. and Fielt, E. (2022). "Public AI canvas for AI-enabled public value: A design science approach." Government Information Quarterly, 39(4), 101722.

Peffers, K., Tuunanen, T., Rothenberger, M. A. and Chatterjee, S. (2007). "A design science research methodology for information systems research." *Journal of Management Information Systems*, *24*(3), 45–77.

Theodorou, A. and Dignum, V. (2020). "Towards ethical and socio-legal governance in AI." *Nature Machine Intelligence*, *2*(1), 10–12.

United Nations. (2022). *AI and Cities Risks, Applications and Governance*. URL: https://unhabitat.org/sites/default/files/2022/10/artificial_intelligence_and_cities_risks_applications and_governance.pdf (visited on May 14, 2024)

Yigitcanlar, T., Corchado, J. M., Mehmood, R., Li, R. Y. M., Mossberger, K. and Desouza, K. (2021). "Responsible urban innovation with local government artificial intelligence (AI): A conceptual framework and research agenda." *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(1), 71.