Openness And Transparency by Design: Crafting an Open Generative AI Platform for the Public Sector

Per Persson
University of Gothenburg, Municipality of Sundsvall
per.persson@ait.gu.se

Yixin Zhang
University of Gothenburg
yixin.zhang@ait.gu.se

Abstract

This paper outlines the foundational importance of openness and transparency in the public sector to foster trust, accountability, and democratic principles. It discusses the potential of generative AI to enhance public sector services through content generation, data analysis, and personalized services, aiming to increase accessibility and efficiency. The study aims to address the research question: What design principles are needed to foster openness and transparency in the design of generative AI-platform solutions within public sector organizations? The open generative AI platform designed and implemented in this innovation project in the Sundsvall municipality, Sweden has already been in use. This action design research study aims to establish design principles for developing trustworthy, scalable generative AI solutions that promote openness, transparency, and collaboration across the public sector.

Keywords: Openness, Transparency, Generative AI, Public Sector, Design Principles

1. Introduction

In today's governance, the principles of openness and transparency within the public sector stand as foundational pillars, essential for fostering trust, strengthening accountability, and enhancing efficiency within democratic societies. Openness and transparency are intricately intertwined concepts. Transparency, defined as the openness and accessibility of information held by governing bodies, not only ensures citizens' rights to information but also plays a pivotal role in reinforcing democratic principles. The significance of openness and transparency transcends access to information it serves as the foundation of democratic principles and good governance (da Cruz et al., 2016; Innerarity, 2017, Lyrio et al., 2018).

Openness and accessibility in government ensure that citizens gain unimpeded access to a comprehensive understanding of the decision-making processes, actions, and financial allocations of their governance (Hood, 1991). This accessibility empowers citizens by providing them with the tools to hold their leaders accountable, make well-informed decisions, and actively participate in civic matters (Welch et al., 2004). Furthermore, transparency fosters a sense of trust, acting as the glue that binds the government and its constituents, ultimately enhancing social cohesion and reinforcing the legitimacy of governance (Megersa & Cassimon, 2015).

Generative AI has the potential to significantly impact the public sector by offering capabilities such as content generation, data analysis, and personalized services, all of which can drive openness and accessibility in the public sector. By leveraging generative AI, public organizations can enhance efficiency, automate tasks, and improve service delivery, potentially leading to cost savings and better resource allocation. This, in turn, enables public organizations to provide efficient services accessible 24/7 to citizens (van Noordt & Misuraca, 2022).

However, the use of generative AI in the public sector also raises significant challenges and risks. These challenges include the potential misuse of generative AI, such as the generation of adversarial examples that could lead to incorrect classifications and decisions. Ethical and social risks associated with generative AI applications, including the potential for misinformation and the need for explainable AI to support decision-making processes, are critical considerations in the public sector (Berryhill et al., 2019). The responsible integration of generative AI in the public sector demands careful consideration to avoid risks of automation bias, overdependence, and long-term staffing problems (Arora, 2020). Addressing these challenges hinges on the public sector exerting control over model selection and data curation processes. To maintain autonomy in modifying models and datasets as

URI: https://hdl.handle.net/10125/109067
978-0-9981331-8-8

needed, it is essential to avoid vendor lock-in. Given the current lack of commitment from AI practitioners in the industry towards transparency and explainability in AI solutions, as highlighted by Denford et al. (2024), the public sector cannot rely solely on the market.

There are also concerns about the responsible use of AI that need to be considered. This includes the need for human-centered and sustainable AI approaches to ensure that AI applications align with societal values and contribute to ecological integrity and social justice (Wynsberghe, 2021). Generative AI not only requires substantial knowledge and understanding of AI solutions within an organization but also demands significant resources for training models and developing solutions. These are crucial to ensure that power and control remain within public sector organizations. In addition, to ensure the development of AI in the public sector is both economically sustainable and resource-efficient, it is essential to adopt open and scalable solutions. These solutions should enable public organizations to collaborate and share the development

Though there has been considerable number of discussions about potential and risks of generative AI use in the public sector, there is a lack of research about the actual implementation of AI tools in the public sector (Mergel et al., 2024). This action design research (ADR) study focuses on how to achieve openness and transparency through the design of the AI-platform solutions, and address the research question*: What design principles are needed to foster openness and transparency in the design of generative AI-platform solutions within public sector organizations?*

2. Background

2.1. Openness and transparency in the public sector

Openness and transparency in the public sector are multifaceted concepts with profound implications for accountability, trust, and governance. In the context of public administration, from information related perspective, openness refers to availability of data and the readiness of public institutions to engage with citizens in a transparent manner (Janssen et al., 2012), and transparency is defined as information about government activities that are made available in a clear, accessible, and timely manner, ensuring that citizens can understand and evaluate the performance of public institutions (Grimmelikhuijsen et al., 2013). Openness and transparency have intersections and are dynamic forces that can either empower or complicate public administration, depending on their context and implementation. It is linked to technological advancements that have the potential to revolutionize the landscape of governance and accountability, help citizens keep an eye on, and build trust in, public services, encourage public employees to better serve the people, and ensure public services are efficient and responsible (Holzner & Holzner, 2006; Günal & Peçe, 2022; Young & Yin, 2020).

Technology is changing how transparency works. Modern technology can facilitate transparency in government and improve the planning and evaluation of public programs and services (Boudreau, 2020). Openness in technology is more than a symbol as it indicates a transformation of the relationship between citizens and authorities. As Maier-Rabler & Huber (2011) promotes, open systems not only suggest a mindset willing to share data and decision-making processes but also encompass actual designs that make such transparency possible. Open initiatives also have the potential to foster collaboration, knowledge sharing, and co-creation of data and software, leading to greater transparency and accessibility of public sector information (Linåker & Runeson, 2020). However, we should be careful when using advanced technology like generative AI, as it can bring transparency challenges that need to be properly managed (Bogner et al., 2021).

Another challenge in achieving desirable transparency in public sector solutions is the tension between protecting private sector interests and the need for shared knowledge in the public sector (Garud et al. 2002). This complicates matters – how can we facilitate open and transparent public sector operations if we rely on proprietary solutions? This places significant demands on public sector information systems procurement processes or forces the public sector to develop solutions by themselves when the market cannot meet the requirements.

2.2. Generative AI adoption in the public sector

The public sector holds significant potential to enhance governance and citizen services through AI adoption. Research suggests that AI can streamline public services, improve digital experiences, and enhance decision-making processes (Kuziemski & Misuraca, 2020). Batarseh & Freeman (2022) highlight the need for transparency in AI adoption to build trust among citizens and argue that being clear about AI's intended use and its potential impacts is crucial for addressing concerns related to accountability, ethics, and privacy.

Generative AI, especially platforms like OpenAI, has recently heightened interest in AI technologies and presents an opportunity for the public sector to leverage AI in ways that were previously impossible. With user-friendly interfaces, generative AI tools can empower

public sector employees, often burdened with bureaucratic tasks, to enhance operational efficiency and effectiveness (Bright et al., 2024).

However, the adoption of AI in the public sector is not without challenges. Technical challenges include integrating AI systems into existing IT infrastructures, ensuring data quality, and addressing security concerns. Legacy systems, common in the public sector, often struggle to support modern AI technologies, making integration complex (Batty et al., 2012). Social challenges revolve around public trust, ethics, and the digital divide. There is a risk that AI technologies could unintentionally worsen existing inequalities or erode public trust if not implemented carefully (Wirtz et al., 2019). Public trust in AI must be informed, with citizens having influence over the conditions under which algorithms operate (Wilson & Van Der Velden, 2022). Political challenges include creating regulatory frameworks that keep pace with technological change and ensuring that AI initiatives align with public policy goals. Regulatory bodies must navigate the rapid evolution of AI technologies, requiring strong political will and leadership to ensure that AI adoption serves the public interest (Yeung & Lodge, 2019).

As highlighted, openness and transparency are critical to foster accountability and public trust in AI. Ethical considerations and participatory governance are essential to ensure that AI technologies benefit the public and address societal concerns. However, there is a notable gap in research on the practical implementation of AI services in the public sector, particularly from a platform perspective.

This study seeks to fill that gap by exploring how openness and transparency can be designed into AI systems. It responds to the call for more research on implementation and public service management related issues in AI, as highlighted by Mergel et al. (2024). By focusing on the intersection of technical and social aspects, this research aims to provide actionable insights for the responsible adoption of AI in the public sector, ensuring that these technologies are implemented in ways that are ethical, transparent, and aligned with public interests.

3. Methodology

The first author is the chief IT architect in the municipality and has been participating in the solution design activities throughout the project. His high level of involvement enabled us to study in situ the design and development of the AI platform. In what follows, we describe the research setting as well as the initiative, and explain our methodological choices around data collection and analysis.

3.1. ADR approach

To investigate how generative AI platforms can be designed and built to facilitate openness and transparency and to cater to collaboration, reuse, and scalability, we conducted an ADR study in the municipality of Sundsvall, Sweden.

ADR is a research approach that integrates action research and design science to address complex organizational problems and develop innovative solutions. ADR aims to not only understand and explain organizational phenomena but also to intervene and bring about change (Sein et al., 2011). This approach allows researchers to actively engage with practitioners and stakeholders, co-create knowledge, and develop practical solutions and can be compared with Design Science Research (DSR) which surfaced as a method for developing artifacts grounded in academic theory.

Action Research (AR) (Coughlan & Coghlan, 2002) is an iterative research method where researchers intervene in the real world to solve practitioners' problems and to gain scientific knowledge (Avison et al., 1999). AR always involves user participation throughout the creation of artifacts, while in DSR users may be assumed by the researcher during the design process (Peffers et al., 2007).

ADR combines AR and DSR to focus on artifact development while taking into consideration user participation and feedback during the experimentation with different versions of the artifact. It considers the artifact development and organizational impact assessment at the same level, with the artifacts being shaped by the organizational context throughout their development and use. The ADR method as described by Sein et al. (2011), is an iterative method divided into steps 1) Problem Formulation, 2) Building, Intervention, and Evaluation (BIE), 3) Reflection and Learning, and 4) Formalization of Learning. It involves close collaboration with practitioners, aligning with the participatory nature of the approach, which emphasizes the involvement of stakeholders in the research process (Becker et al., 2019). This collaborative aspect is essential for developing and evaluating tools and solutions in real-world settings. It results in an artifact, in the context of this study, the artifact is an IT architecture with specific material properties.

The research contributions are delivered as DPs, capture the knowledge "*... about creating other instances of artifacts that belong to the same class*" (Sein et al., 2011) and are statements that guide or constrain actions (Hevner & Chatterjee, 2010). In this study, the DPs have been derived primarily from justificatory knowledge. The scope of justificatory knowledge is broader than that of kernel theory, which refers to "*any descriptive theory that informs artifact*

construction" (Gregor & Jones, 2007). Justificatory knowledge also includes informal knowledge from the field and the experience of practitioners (Gregor & Hevner, 2013). This type of knowledge is important in this ADR research, especially considering that designing generative AI platforms in the public sector is pioneering work in the field.

3.2. Research context

As mentioned, this ADR study focuses on an innovation project in the municipality of Sundsvall, Sweden with the aim to investigate the possibilities of creating an open generic platform for establishing and scaling generative AI solutions. The municipality of Sundsvall has about 100,000 residents and is a relatively big municipality, given that the median size municipality in Sweden has about 20,000 residents. With around 8,000 employees, the municipality organization is the largest employer in the area, and its primary objective is to provide services and welfare to meet the needs of residents.

Sundsvall municipality is an intriguing case for analysis as it is well-known for its digitalization efforts nationally and is recognized as one of the forerunners in digitalization in of Sweden's municipality sector. In December 2023, the municipality's IT director was given a great responsibility as the only representative of the municipal sector in the Sweden government's AI commission (Sveriges Riksdag (the Parliament of Sweden), 2023). Also, the project owner is a representative of an AI council for Swedish municipalities.

Sundsvall municipality is particularly suitable for investigating how to foster openness and transparency in the design of generative AI-platform solutions within public sector organizations, as the municipality puts a great emphasis on openness and transparency in public sector operations. This is evidenced, for example, by Sundsvall's strategy for sustainable digital development, *"Openness and transparency are important both for democracy but also for increased cooperation between business, the public sector, and society"* Translated from Swedish, A strategy for sustainable digital development, Sundsvall Municipality (2022).

3.3. Data collection

Apart from actively participating in the project, we used documents, reports, pre-studies, project descriptions, solution descriptions, and public communication related to the project in the study (several of which the first author contributed to). This information, though scattered throughout the different stages of the project, is important to be included, as it can be part of justificatory knowledge. The archival data consists of documents like; the project descriptions and basis for decisions to invest in the project (phase 1 decided in April 2023 and phase 2 decided in January 2024), the API Strategy in the municipality of Sundsvall (decided in September 2020), the municipality of Sundsvall opens accessible information on several aspects of AI (October 2021 – June 2024), and solution descriptions from the project evolved in iterations during the project's timeline (May 2023 – June 2024).

We also conducted interviews with key personnel involved in the project. The first round of interviews was conducted in October and November 2023 and lasted approximately 60 minutes on average. Follow-up interviews with a selected set of interviewees were conducted in May 2024 and lasted approximately 30 minutes on average. The interviews were recorded with the interviewees' consent and transcribed for analysis. The interviewees were; the project owner at the Sundsvall municipality (also member of the Swedish municipalities AI Council) (2 interviews), a senior IT project manager at Sundsvall municipality (2 interviews), an IT security expert representing one of the municipalities owned companies (1 interview), and the CEO and the Technical Lead from the vendor company (a start-up company sprung from a world-leading academic institution) building the solution (1 interview).

4. The ADR project

The project core team in the municipality comprised a project owner, an IT project manager, and an IT security expert. The project IT architect, participating in the design phases, is also the first author of this article and acts as both the practitioner and the researcher in the ADR project context. The project team is lean while highly competent with experienced expert competencies in all roles, a success factor for maintaining the innovative approach. The project also involves employees at different departments in the municipalities, while they are not part of the project core team and are mainly users of the different use cases for the generative AI platform.

The vendor selected as a partner in the project had a clear vision to make AI tools adaptable to various public sector processes, ensuring that domain experts can influence AI behavior and outcomes, and focused on building an AI framework that emphasizes transparency, flexibility, and domain expertise integration for public sector applications.

Phase 1 of the project (below referred to as the first iteration of the ADR project) started in May 2023 and delivered results in December 2023. Phase 2 of the

project started in January 2024 and will span over several iterations in the whole year of 2024 (where the first iteration in phase 2 of the project maps to the second iteration of the ADR project). As of May 2024, the platform has been put into use with real use cases, such as a public assistant for adult education services, an assistant for domestic services employees, and a support services assistant for the municipal internet services provider.

4.1 Problem formulation

Problem formulation is a crucial step in the design of the IT artifact. Identifying problems and formulating goals will shape the outcome of the project, in this case, both the generative AI platform (the artifact) and the DPs. The overall vision stated by the project owner for how to scale the use of generative AI in the municipality is to establish an open AI platform that supports various future use cases while ensuring thorough consideration of data privacy and ethical concerns. Below, we introduce the themes that emerged from the analysis of our interviewee's sharing about the problems that need to be addressed and the aims of the project.

One of the project's goals, to address the "*big problem with transparency in AI*" (as stated by the project owner) and how to achieve a satisfying level of transparency was one of the most important and complex areas of discussion. We adopted the definition of transparency as a "*human-level understanding of the inner workings of the model*" (Belle & Papantonis, 2021). Every time an AI is asked a question or to perform a step in a process, the AI model makes a form of a decision by providing an answer. "*We have to be able to be 100% transparent with our citizens and say that this answer comes from this set of rules or this description or whatever*" (IT project manager).

Transparency is even more important in use cases that involve decision-making about citizens. *"What if we must analyze a large data set about how an individual is behaving …, could be an indication that the individual needs more help to cope with their everyday business? … transparency will be a much more important factor in those use cases"*. (Project owner)

Control is raised as another important requirement, and it is considered closely linked with transparency. Control is exhibited in various dimensions, including the selection of models to utilize, control over data, and oversight of the operations within AI tools. "*The control is very much based on the principle of transparency. As we all know, AI has a quite large problem, a big problem with transparency. The models, the processes, there's a lot of black boxing in this industry where we just use services, where we send our data and we get the response, but we have no idea what happened in there between."* (Project owner)

Because the municipality offers various services to citizens (e.g., building permit application, welfare decision making), and owns different companies (e.g., water company, energy company), they hope the AI platform can enable them to provide AI solutions to different scenarios. Scalability and versatility are considered important requirements.

"It's more infrastructure or platform where we basically could have hundreds or several hundreds of use cases in the future that we run on the same infrastructure. … So that's why we must think and build for a much broader perspective than a single use case." (Project owner)

Furthermore, the solution shall also be scalable in the sense that "*more municipalities can use it*" (Project owner), and "*makes it possible to share … knowledge that is built up between municipalities*" (Projektbeskrivning Utforskande av AI – Project description Exploring AI).

4.2 First iteration

The main goal of the first iteration was to test if technology in the market was ready to deliver on this vision and whether it was feasible to develop an open digital platform for AI. The platform should be scalable from a performance perspective (that you can add more use cases and users without affecting response times), from a perspective of the amount of information (that you can add the amount of information that is needed), and from a security perspective (that you can configure the solution to be able to handle more sensitive data). It should provide a solution demonstrating that a more open and scalable digital platform enables increased collaboration between municipalities in the field of AI and lowers the thresholds for adopting new technology. The idea was that open and scalable digital platforms that support open models make it possible to share model training and the knowledge that is built up in different municipalities.

Due to the innovative nature of the project in the public sector, there are limited prior examples to draw from. To the best of our knowledge, there is no overarching theory that can guide the project. However, we found literature on public sector information systems and digital transformation that shed light on issues we also aim to address, such as over-reliance on external suppliers and vendor lock-in (Welby & Tan, 2022), lack of access to data, and loss of data ownership (Persson et al., 2024), as well as organizational siloing resulting in a lack of collaboration among departments (Øvrelid, 2023). The project team also engaged in multiple rounds of discussions regarding how to address these problems.

4.2.1 Building, intervention, and evolution

Two real-life use cases were proposed to be implemented in the first iteration. The user cases make it possible to evaluate the solution with real-life data. The first use case was general, it used the municipality website sundsvall.se as a data source, and a chatbot was developed to answer questions about the services provided by the municipality. This use case was tested with a wide set of questions, such as "How do I apply for a building permit?", or "When is the bathhouse open on Sunday?", to verify the solution's functionality. The other use case was more specific and covered advanced support questions for a municipality-owned Internet service provider. If the evaluation proves successful, these use cases would be candidates for the first Minimum Viable Product (MVP) to set into production.

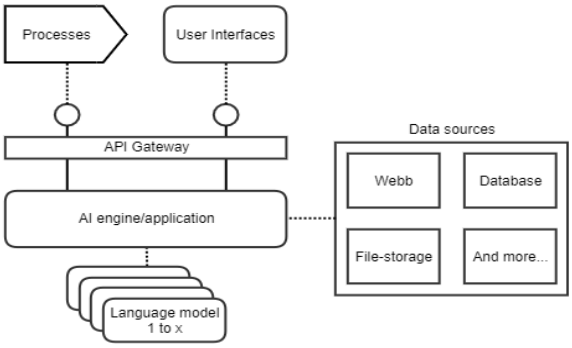

Figure 1. High-level solution architecture.

Guided by these discussions, the solution architecture (the artifact, illustrated in Figure 1) was designed including the following components: user interfaces and processes that integrate via APIs exposed in a central API gateway, the AI engine/application managing and configuring the AI platform (e.g., configuration and function for indexing/joining data sources with information, permission management, logging, configuration of language models, etc.), the data sources that the project team chooses to us and the language models (each application built on the platform can use any language model of choice).

4.2.2 Reflection and learning

Overall, the first iteration of the project was evaluated as successful, delivering a solution that representatives from different parts of the municipality's organization could evaluate in a real-life work environment. It demonstrated the feasibility of introducing an open AI platform that allows a municipality to continuously develop AI services with different requirements. The collaboration with the vendor resulted in a solution that will be released as open source. The level of transparency was also satisfying – "*we know exactly what is happening in the process as a whole ... we know exactly what model is being used*" (project owner).

Highlighted in the project report for the first iteration, are the following as important factors for a long-term sustainable solution for generative AI. First, as a municipality, there is a need to be able to create AI services for different use cases with simple means. Implementing separate solutions for each requirement is impractical, so the municipality needs a platform capable of accommodating a variety of use cases over time. Second, the municipality needs a platform that can keep pace with technological advancements, such as adopting newly developed language models. Otherwise, it will be difficult to fully harness the benefits of AI in the long term. Third, every application of AI must comply with regulatory requirements and offer transparency right down to how the data was processed during a decision to exactly which models were used at the time of the decision. Fourth, the municipality needs secure solutions that allow it to implement certain use cases that do not have personal data via certain language models and other use cases with sensitive data through other language models. Language models can be run in a municipalities own data center, so the municipality has control over data in the AI platform. Furthermore, the municipality must be able to share what it is developing with other municipalities, and the municipality also wants to be able to take part in what other municipalities are developing. This sharing applies both in the form of source code, but also around concrete AI applications.

4.3 Second iteration

4.3.1 Building, intervention, and evolution

The goal of the second iteration of the project is to roll out the use of the solution in the municipality, and in parallel continue to evolve the platform adding features, especially concerning transparency, in iterations. The project found a solution that should enable full transparency in exactly what was the basis for the conclusion the AI service reached. If every interaction with the AI service generates and logs a data package in the background, this package should describe exactly which conditions were the basis for the conclusion reached by the AI service.

At a minimum, the question asked/task given to the AI service, the prompt that was sent in the question or given as input to a task, the settings in the AI service (e.g., setting for temperature), the embedding model used to transform the question/task into a vector, the source that the model identified as containing the answer, the data set that the model identified as

containing the answer, the language model used to translate the response and the final answer/conclusion that the language model translated the answer into should be available for each response or decision.

The first external AI assistant for citizen use, an assistant for adult education services, was released in early May alongside several internal AI assistants used for a variety of purposes. All these assistants were designed to assist the user in specific contexts using specific sets of data with varying efficiency depending on the use case. The next planned assistant will help officials create official letters, as required by laws and regulations, estimated to save 50% of the time for administrative tasks.

4.3.2 Reflection and learning

The second iteration demonstrates that the platform is prepared for scaling and that it is possible to support transparency to a certain extent in the platform. The reflection and learning during the second iteration were integrated with that from the first iteration and are presented in the formalization of learning below.

5. Formalization of learning

Based on the learnings from the two ADR project iterations, a set of DPs guiding the design of open, transparent, and shareable AI solutions for the public sector is defined. We first introduce the DPs and then discuss their implications.

5.1 Design principles (DPs)

We integrated expert knowledge with prior literature and developed the following DPs using the schema created by Gregor et al. (2020). DPs can be formed as "*For Implementer I to achieve or allow for Aim A for User U in Context C, employ Mechanisms M1, M2, … Mn involving Enactors E1, E2, … En because of Rationale R*". The schema involves several components: actors such as Implementers, Users, and Enactors. In the context of this project, the enactor and implementer overlap, so we omit the enactor from the DPs. Mechanisms are broad and can be "*actions, the use of artifacts characterized by a specific form, shape, or architecture*" (Gregor et al., 2020). The rationale is the justification for the design principle and is provided following the DPs.

DP1. Transparency*: For the development team (implementer) to achieve open models and transparent data processing (aim) for the public sector (user) in data processing and response generation (context), the solution shall support open models, provide information about data processing and response generation procedures (mechanisms).* The call for transparency aims to help hold governments accountable and build trust in institutions (Young & Yin, 2020). AI practitioners, in general, are not yet invested in providing transparency and explainability in AI solutions (Denford et al., 2024) though, hence the importance of this DP,

DP2. Model Independence: *For the development team (implementer) to achieve flexible model management and query control (aim) for the public sector (user) in data analysis and decision-making (context), the solution shall be model-independent through modular design and APIs (mechanisms).* A modular design supports the ability to update and modify individual modules without affecting the entire system like the choice of LMM, which is critical for adapting to changing models and diverse analysis needs over time (Albers et al., 2019). This DP addresses diverse analysis needs and empowers the public sector to control and select the appropriate model for specific queries.

DP3. Open Source*: For the development team (implementer) to achieve an open, scalable digital platform (aim) for the public sector (user) in data management and decision-making (context), the solution shall be delivered as open-source software (mechanisms).* This DP helps prevent vendor lock-in (Persson et al., 2023), provides public sector organizations greater control over their data and systems, enhances decision-making processes, and allows for more flexible and scalable digital solutions (Ahlawat et al., 2021).

DP4. Data Ownership*: For the development team and data managers (implementer) to achieve (data and knowledge ownership (aim) for public organizations (user), regarding data and knowledge management (context), the solution shall ensure ownership of data built up by the public organization and guarantee ownership of knowledge generated with the organization's training (mechanisms).* Data ownership ensures that public organizations retain full authority over their data, preventing unauthorized usage or changes (Asswad & Gómez, 2021), and prevent vendor lock-in (Persson et al., 2023).

DP5. Interoperability*: For the development team (implementer) to achieve interoperability (aim) for the public organization (user) in software integration and automation (context), the solution shall ensure all functionality in the solution is executable via APIs, adhering to OpenAPI specifications (mechanisms).* OpenAPI facilitates automation, code generation, and consistent API design, making it easier to integrate and maintain systems across different projects (Dwayo blog, 2023).

5.2 Implications of the DPs

Below we explain the implications of the DPs. We start with a specific use case, in which citizens can ask a generative AI-powered chatbot for questions about preschool services. The data provided as input to the AI tool includes web pages describing preschool services and other publicly available documents.

The following is an example of how an answer generated by an AI service might appear when exported (i.e., in a relatively "raw" format), providing the end user with comprehensive insight into the factors influencing the decision made by the AI service at the exact time the decision was made ("LLM" means Large Language Model).

LLM	SOLAR-0-70b-16bi
LLM settings	Temperature=0.4
Embedding model	multilingual-e5-large
Question	What are the opening hours at Solskenet's preschool?
Prompt	You work at Sundsvall Municipality as an assistant and guide to residents and visitors in Sundsvall Municipality helping them by answering their questions.
Data source	sundsvall.se/solskenet
Data	Solskenet's preschool Address road 123, 123 45 Sundsvall Opens: 06.00, Closes: 18.00
LLM translation	The opening hours at Solskenets Preschool are 06.00-18.00.

Table 1. Example of data included in a decision package from the AI service.

As shown in this example, guided by DP1 Transparency, the platform provides information about which LLM model is used for the generation of the answer, what is the LLM settings, what are the data sources, and how the data is presented as the answer.

There are design considerations and implications that beyond what this simple output can illustrate. Guided by DP2 Model Independence, DP3 Open Source, and DP4 Data Ownership, in addition to giving users information about the components and settings of the process, the AI platform allows designers to design the process that specific AI tools work and have control over models and where data is processed.

The platform also provides the organization with opportunities to integrate new models, which is particularly crucial given the rapid advancements in the field of AI. *"One thing that everyone in (the field of) AI has observed, is that the release of models is occurring extremely rapidly. What we want to ensure with this framework is that customers don't buy into one singular model or one singular implementation. We need to be able to swap out models."* (Company collaborator)

Guided by DP2 Model Independence and DP5 Interoperability, the platform allows for switching among different LLM models and allows for the LLM models to be connected to different input datasets seamlessly. This platform has revolutionized how IT projects are planned, especially in terms of time scale and the way of collaboration among the IT team and the user departments. *"When I'm saying traditional approach, … we're good at planning, we're down to quarters, often we try to plan even longer period periods than a quarter. … (now) that the development is so much easier within AI. To develop a new assistant, we're talking about hours, not weeks, not months. It's hours to develop a new AI assistant."* (Project owner)

The interviewee shared that users can now approach the IT team with project ideas, such as designing an AI assistant to aid in answering citizens' questions about preschool services, assisting internal staff with the IT security handbook (e.g., can I use my work email for private purposes?) or handling invoices. In this collaborative process, users share links to shared drives containing input data sources, and they can sit together with the IT team for several hours, together they try to "plug in" different LLM models and evaluate the results. They can assess the feasibility of project ideas and refine the assistant's design on the spot. This collaborative approach relies on this open and scalable AI platform, without which such endeavors would be impossible.

Despite our aim to achieve transparency, certain aspects remain beyond our reach. LLMs are typically trained by private companies that disclose limited information about their training data, and the algorithms themselves are inherently complex. Considering these, the sources of the data used to generate answers must be provided to users, enabling them to verify the answers.

We shall point out that building an AI platform is more than just connecting the LLM model with the input dataset and output outlet. Many preparations are involved to make the data accessible. For example, it is important to engage domain experts who can together brainstorm about use case ideas, be involved in the design process, and evaluate the results.

6. Discussion

In this study, we examined the following research question: "*What design principles are needed to foster openness and transparency in the design of generative AI-platform solutions within public sector*

organizations?" To address the question, we conducted an ADR project on an innovation project at the municipality of Sundsvall. The innovation project designed and built an MVP for an open and transparent generative AI platform, producing an artifact in the form of an IT architecture, and developed five DPs.

As noted in the literature review, though openness and transparency have received a lot of attention in current literature, most studies discuss openness and transparency from theoretical perspectives and focus more on their implications. There is a lack of research regarding how openness and transparency can be achieved in the actual implementation of AI applications in public sector. This study addresses the literature gap, and the proposed DP's can guide the design of open and transparent generative AI applications in public sector with high scaling potentials. The DPs also provide a set of guidelines that future research in public sector AI can build upon.

Previous works, such as those by Smit et al. (2020) and Weisz et al. (2023), primarily discuss DPs for generative AI applications with an emphasis on ethics and trust. Our study contributes to theory by defining DPs that instead concentrate on openness and transparency as key drivers for fostering cooperation and sharing. The ADR project also demonstrates how the DPs influence design choices and achieve openness and transparency in the AI platform and use cases. The study highlights the importance of openness and transparency in the public sector and establishes a connection between these principles and the adoption of generative AI services. This connection underscores the need for transparency as a foundational element in ensuring accountability and trustworthiness in public-sector AI implementations.

The practical implications of our research are substantial. The artifact developed serves as a valuable example of open, transparent, and scalable generative AI applications for the public sector in general. The DPs of transparency, model independence, data ownership, open source, and interoperability are applicable in building generative AI solutions in various public sector domains, such as healthcare, education, governance, and social services. In practice, the open-source solution designed and built guided by these principles has the potential to be scaled up to support the organized sharing of generative AI components across all 290 Swedish municipalities. This scalability is crucial for smaller municipalities with limited resources to leverage AI technologies. Scaling the generative AI platform to other municipalities is a project we continue to follow and study, with the aim to incorporating perspectives from a wider range of stakeholders such as citizens, policymakers, and AI ethics experts.

7. References

Ahlawat, P., Boyne, J., Herz, D., Schmieg, F., & Stephan, M. (2021). Why You Need an Open Source Software Strategy.

Albers, A., Bursac, N., Scherer, H., Birk, C., Powelske, J., & Muschik, S. (2019). Model-based systems engineering in modular design. *Design Science*, *5*, e17.

Arora, A. (2020). Conceptualising artificial intelligence as a digital healthcare innovation: an introductory review. *Medical Devices: Evidence and Research*, 223-230.

Asswad, J., & Marx Gómez, J. (2021). Data ownership: a survey. *Information*, *12*(11), 465.

Avison, D. E., Lau, F., Myers, M. D., & Nielsen, P. A. (1999). Action research. *Communications of the ACM, 42(1),* 94-97.

Batarseh, F. A., & Freeman, L. (Eds.). (2022). AI Assurance: Towards Trustworthy, Explainable, Safe, and Ethical AI. *Academic Press*.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, *214*, 481-518.

Becker, F., Meyer, M., Siemon, D., & Robra-Bissantz, S. (2019). Taking action: extending participatory action design research with design thinking.

Belle, V., & Papantonis, I. (2021). Principles and practice of explainable machine learning. *Frontiers in big* Data, 4, 688969.

Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). Hello, World: Artificial intelligence and its use in the public sector. *OECD Working Papers on Public Governance.*

Bogner, J., Verdecchia, R., & Gerostathopoulos, I. (2021). Characterizing technical debt and antipatterns in AI-based systems: A systematic mapping study. *In 2021 IEEE/ACM International Conference on Technical Debt (TechDebt) (pp. 64-73). IEEE.*

Boudreau, C. (2020). Reuse of open data in Quebec: from economic development to government transparency. *International Review of Administrative Sciences, 87(4),* 855-869.

Bright, J., Enock, F. E., Esnaashari, S., Francis, J., Hashem, Y., & Morgan, D. (2024). Generative AI is already widespread in the public sector. *arXiv preprint arXiv:2401.01291*.

Coughlan, P., & Coghlan, D. (2002). Action research for operations management. *International journal of operations & production management, 22(2),* 220-240.

da Cruz, N. F., Tavares, A. F., Marques, R. C., Jorge, S., & De Sousa, L. (2016). Measuring local government transparency. *Public Management Review*, *18*(6), 866-893.

Denford, J., Dawson, G. S., & Desouza, K. (2024). National AI Strategic Plans for the Public versus Private Sectors: A Cross-Cultural Configurational Analysis. *In Proceedings of the 57th Hawaii International Conference on System Sciences*

Dwayo blog. (2023). https://blog.dwayo.com/introduction-to-openapi-specifications/

Günal CN, Peçe H. (2022). Opening the curtains: Adaption of transparency policy for eliminating in confidentiality of the faith groups. *International journal of human sciences. 2022;19(4):550–8*.

Garud, R., Jain, S., and Kumaraswamy, A. 2002. Institutional Entrepreneurship in the Sponsorship of Common Technological Standards: The Case of Sun Microsystems and Java. *Academy of Management Journal (45:1)*, pp. 196-214.

Gregor, S. and Jones, D., 2007. The anatomy of a design theory. Journal of the *Association for Information Systems, 8(5)*, p.1.

Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). Research perspectives: the anatomy of a design principle. *Journal of the Association for Information Systems, 21(6)*, 2.

Gregor, S. and Hevner, A.R., 2013. Positioning and presenting design science research for maximum impact. *MIS quarterly*, pp.337-355

Grimmelikhuijsen, S., Porumbescu, G., Hong, B., & Im, T. (2013). The effect of transparency on trust in government: A cross ‐ national comparative experiment. *Public Administration Review*, *73*(4), 575-586.

Hevner, A. R. & Chatterjee, S. (2010). Design Research in information systems. *New York, NY: Springer*

Holzner, B. and Holzner, L. (2006). Transparency in Global Change: The Vanguard of the Open Society. *University of Pittsburgh Press, Pittsburgh*.

Hood, C. (1991). A public management for all seasons?. *Public Administration, 69(1)*, 3-19.

Innerarity, D. (2017). The Improvement of Democracy Trought Transparency and Its Limits. *International Journal of Philosophy*, *5*(5), 44-49.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management*, *29*(4), 258-268.

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy, 44(6)*, 101976.

Linåker, J. and Runeson, P. (2020). Collaboration in open government data ecosystems: open cross-sector sharing and co-development of data and software. *Lecture Notes in Computer Science*, 290-303.

Lyrio, M. V. L., Lunkes, R. J., & Taliani, E. T. C. (2018). Thirty years of studies on transparency, accountability, and corruption in the public sector: The state of the art and opportunities for future research. *Public Integrity*, *20*(5), 512-533.

Maier-Rabler, U., & Huber, S. (2011). "Open": the changing relation between citizens, public administration, and political authority. *JeDEM-eJournal of eDemocracy and open Government, 3(2)*, 182-191.

Megersa, K. & Cassimon, D. (2015). Public debt, economic growth, and public sector management in developing countries: is there a link? *Public Administration and Development, 35(5)*, 329-346.

Mergel, I., Dickinson, H., Stenvall, J., & Gasco, M. (2024). Implementing AI in the public sector. *Public Management Review*, 1-14.

Persson, P., Zhang, Y., Asatiani, A., Lindman, J., & Rudmark, D. (2023). Technical Debt in the Municipality Sector - The Missing Link with Citizens and Silofication, *The 14th Scandinavian Conference of Information Systems, 2023.*

Persson, P., Zhang, Y., Asatiani, A., Lindman, J., & Rudmark, D. (2024). Toward Citizen-Centered Digital Government: Design Principles Guided Legacy System Renewal in A Swedish Municipality. *In Proceedings of the 57th Hawaii International Conference on System Sciences.*

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

Øvrelid, E. (2023). Exploring adaptive mirroring in healthcare IT architectures. *Health Systems*, 1-12.

Sein, M., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly, 35(1)*, 37-56.

Smit, K., Zoet, M., & van Meerten, J. (2020). A review of AI principles in practice.

Sundsvall Municipality. (2022). Strategi för en hållbar digital utveckling – A strategy for sustainable digital development. https://utveckling.sundsvall.se/malbild-och-strategi

Sveriges Riksdag (The Parliament of Sweden). (2023). AI-kommissionen (Fi 2023:11).

van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. *Government Information Quarterly, 39(3)*, 101714.

Weisz, J. D., Muller, M., He, J., & Houde, S. (2023). Toward general design principles for generative AI applications. *arXiv preprint arXiv:2301.05578*.

Welby, B. & E. Tan (2022). Designing and delivering public services in the digital age *Going Digital Toolkit Note, No. 22*.

Welch, E., Hinnant, C., & Moon, M. (2004). Linking citizen satisfaction with e-government and trust in government. *Journal of Public Administration Research and Theory, 15(3)*, 371-391.

Wilson, C., & Van Der Velden, M. (2022). Sustainable AI: An integrated model to guide public sector decision-making. *Technology in Society*, *68*, 101926.

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector - applications and challenges. *International Journal of Public Administration, 42(7)*, 596-615.

Wynsberghe, van, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics, 1(3)*, 213-218.

Yeung, K., & Lodge, M. (Eds.). (2019). Algorithmic regulation. *Oxford University Press*.

Young, K. G. and Yin, M. L. (2020). Role of e-government and open data to enhance transparency practices. *Journal of Enterprise Business Management*, 47-54