Data & Policy (2026), 8: e40
doi:10.1017/dap.2026.10088

RESEARCH ARTICLE

AI Ethics Canvas: A graphical tool to design and deploy ethical artificial intelligence for public administrations

Riccardo Nanni[1], Pietro Giovanni Bizzaro[2], Albana Celepija[2], Munazza Usmani[2] and Maurizio Napolitano[2]

[1]Centre Internet et Société, CNRS, France
[2]Digital Society Centre, Fondazione Bruno Kessler, Italy
Corresponding author: Riccardo Nanni; Email: riccardo.nanni@cnrs.fr

Received: 19 August 2025 UTC; **Revised:** 24 December 2025 UTC; **Accepted:** 02 July 2026 UTC

Keywords: artificial intelligence; data protection; European Union; ethics; public administration

Abstract

This article presents the AI Ethics Canvas: a graphical tool to guide public administrations in mainstreaming ethics in the development, deployment, and monitoring of artificial intelligence (AI) tools for their own everyday use. Our canvas is the result of an action-research effort conducted within the scope of a project involving the Autonomous Province of Trento (PAT), Italy. We initially developed it based on existing ethical and legal frameworks. Then, we conducted qualitative interviews with PAT functionaries and technologists developing software on behalf of the PAT to ensure the canvas's content was clear and applicable to public administration work. While originally elaborated for the PAT, the canvas presented is applicable to all public administrations within the European Union who wish to adopt AI in their operations, as it references to core European legislation such as General Data Protection Regulation, AI Act, Data Act, and Data Governance Act. The article presents the AI Ethics Canvas both as a final product and in its creation process to enhance transparency in AI ethics tool development.

Policy Significance Statement

The AI Ethics Canvas supports public administrations in considering ethical implications in AI adoption. Any EU-based public administration can use it to assess ethical concerns while adopting AI to (partially) automate its internal bureaucratic procedures. The canvas seeks not to replace human scrutiny of ethical implications in AI adoption, but to support it. AI adoption for public administrations is intended to be in support of functionaries' work rather than as a replacement for it.

1. Introduction

While Artificial Intelligence (AI) development has been ongoing since the 1950s, only recently have governments all over the world started considering how to regulate and deploy it in the everyday running of public administration (PA). Applying AI to PAs promises to accelerate bureaucratic processes while providing decision-makers with instruments to elaborate ad hoc solutions oriented to citizens' needs.

This research article was awarded Open Materials badge for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

However, slow decision-making mechanisms, lack of skills, and the novelty of AI-focused legislation constitute bottlenecks in the process (Wirtz et al., 2020).

In the European Union (EU), supranational institutions are adopting a number of regulatory measures to regulate the (mis)use of AI and data. These include the Data Governance Act, the Data Act, and the Artificial Intelligence Act (European Parliament, 2021; European Union, 2023a, 2023b).

We present the AI Ethics Canvas (Appendix A, Figures A1–A7) to guide PAs in the development, deployment, and monitoring of AI tools for their own everyday use. It is a graphical tool containing guiding questions to provide step-by-step support in the implementation of AI tools, which has its roots in design thinking literature (Ojasalo and Ojasalo, 2018). Our canvas is the result of an action research project involving the Autonomous Province of Trento (PAT), Italy. In its development, we adopted "value-sensitive design" (Van Der Hoven and Manders-Huits, 2020), an approach that mainstreams human values in the design of technological systems since their inception. To ensure both value-sensitivity and legal coherence, the canvas was developed based on existing European ethical and legal frameworks and validated through interviews with PAT functionaries. While originally elaborated for the PAT, the canvas is applicable to all PAs who wish to adopt AI in their operations as it is based on EU legislative requirements. We present the AI Ethics Canvas both as a final product and in its creation process to enhance transparency on AI ethics tool development.

The next section presents a literature review. Section 3 illustrates the methods adopted in the elaboration of the canvas. Section 4 presents the canvas and its application, while Section 5 discusses its use, its strengths, and weaknesses as emerged during the interviews, along with technical instruments that can support the mainstreaming of ethics in AI instruments. Section 6 draws conclusions.

2. Literature review

2.1. AI adoption in public administration

Currently, AI adoption in the public sector is fragmented and primarily concentrated in three government functions: improving policy-making processes, optimising public service delivery, and strengthening internal administrative management (van Noordt and Misuraca, 2022; Organisation for Economic Co-operation and Development, 2024). In policy making, AI supports the early identification of emerging social issues, enhances the quality of decisions, monitors implementation, and evaluates existing measures (Organisation for Economic Co-operation and Development - Observatory of Public Sector Innovation, 2025). In service delivery, AI is used to increase efficiency and effectiveness, often by enabling more responsive and personalised interactions with businesses and citizens, while optimising the allocation of financial and human resources (Manzoni et al., 2022).

However, AI implementation remains a challenging problem. On the one hand, the technical complexity of AI-based solutions stems from factors such as continuously changing data and the non-deterministic behaviour of AI models. On the other hand, AI adoption in PAs is constrained by several hindering factors, such as insufficient AI-related expertise, organisational implementation challenges, limited system and data integration, and data availability and ethical concerns (Mikalef et al., 2019). In addition, organisational challenges in implementing AI solutions frequently arise from uncertainty about how to initiate adoption, combined with resistance to modifying established work practices (Medaglia et al., 2023).

Consequently, despite AI potential, implementation is still lagging. Ethical implications are among the main issues to address when implementing AI. To this end, the EU's Ethics Guidelines for Trustworthy AI outline seven requirements (human agency, technical robustness, privacy, transparency, non-discrimination, societal well-being, and accountability) that directly address government responsibilities in service delivery and regulation by grounding the principles in practice (High-Level Expert Group on Artificial Intelligence, 2020).

However, AI's inherent probabilism demands procedural safeguards such as expert review panels or algorithmic impact assessments to contextualise outputs before decisions become binding (Ada Lovelace

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

Institute, 2022). Existing ethics guidelines define what PAs should achieve but lack accessible, domain-specific methods for operationalising these principles, highlighting the need for a tailored AI Ethics Canvas.

Overall, opportunities and challenges emerge in adopting AI in PAs. Namely, the challenges emerge from the biases contained in the algorithms that constitute an AI system, in the data used to train it, in the data analysed, as well as in the viewpoint (positionality) of the person(s) in charge of making AI-based decisions (Teo, 2022).

2.2. Canvas as an instrument

The use of a canvas to design a thematic management model draws back to the work on business models by Osterwalder et al. (2010). Drawing from the theory of design thinking in business, the aforementioned work opened the way for the elaboration of graphic schemes (known as canvases) for the elaboration of a management model. A canvas can be devised as a conceptual map or a form to check the actions needed to conduct a process (e.g., establishing a business) according to a recognised benchmark (Kronsbein and Mueller, 2019).

The use of canvases has then expanded to other areas. Canvases are employed as visual tools to help in establishing management models and strategies for data-driven policymaking. For example, a variety of canvases are targeted to the needs of different groups of actors in building and managing their own dataspace (Ojasalo and Ojasalo, 2018). The Open Data Institute has elaborated a Data Ethics Canvas to help organisations in establishing ethically sound data governance mechanisms, reducing potential harm to others (Open Data Institute, 2021), whereas the International Data Spaces Association (2021) published a business model for a sovereignty-based dataspace.

These are just a few examples of how different management models can be identified depending on the objectives one has. In particular, the International Dataspace Association (IDSA) model is recognised in literature and in practice through academic publications. Initially launched by the Fraunhofer Society, the International Data Space has gathered industrial interest that led to the birth of IDSA to update and elaborate international data space management models (International Data Spaces Association, 2021). This instrument is a canvas that provides practical, user-friendly guidance in a technical process.

The use of canvases has also been applied to AI implementation in the 2020s. Business canvases to integrate AI in business processes have been presented, thus setting the scene for AI implementation canvases like the one presented in this article. Furthermore, researchers have elaborated canvases for machine learning implementations in fake online products detection (Nurcahyo et al., 2022). A canvas for public good-oriented AI applications has also been developed, observing both the technical implementation aspects and assessing the social value of it (Fatima et al., 2022). Indeed, the latter has quite a few things in common with the canvas presented in this article. First, the public good orientation. Second, there is an eye for both technical implementations and socio-cultural elements, such as social value and trust. However, the canvas presented in this article is specifically oriented to EU-based PAs, thus encompassing PA legal requirements along with both technical and social considerations. This is in line with the emerging use of graphical tools to guide non-technical AI users through the adoption of AI instruments (Celepija et al., 2024).

3. Methods

The canvas is the result of an action-research effort. Action-research is an approach through which the researcher engages with their research participants to try and deliver change. In doing so, those experiencing the issues being studied are involved first-hand in envisaging solutions (Cornish et al., 2023). Specifically, we engaged with the PAT's PA system to implement AI solutions in its everyday work. The PAT is a province of over 500 thousand inhabitants, mostly concentrated in the capital, Trento, lying in the river Adige valley. While small and mountainous, it is one of the wealthiest Italian provinces. Although its Digital Economy and Society Index (DESI) is higher than the Italian average thanks to high

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

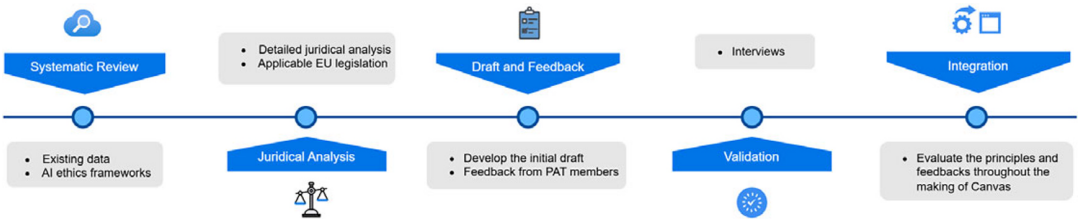


Figure 1. *Canvas co-creation process.*

human capital and relatively good connectivity, public services are not notably digitalised compared to the rest of the country (Carbone et al., 2023). These characteristics make it a suitable pilot case to explore the potential and bottlenecks for AI implementation in the PA.

The canvas (Appendix A) is the result of a process to provide the PAT's PA system with instruments to guide it through the creation and implementation of AI solutions. Its staff members have been involved in the elaboration of the canvas through an ongoing validation process as summarised in Figure 1. The project focused on PAT functionaries rather than lay citizens as the AI instruments in question are for functionaries' use, thus making them the main target in terms of need and preparedness. Appendix A shows the English-language canvas. However, the GitHub repository mentioned also includes an Italian version, which is more closely oriented to the Italian PA, and the PAT specifically.

To elaborate on the first draft of the canvas, we conducted a twofold analysis.

First, we conducted a systematic review of the existing data and AI ethics frameworks, distinguishing between those originating in the EU and, therefore, straightforwardly applicable to an EU PA and those originating elsewhere but still (partially) applicable to an EU context. Nine such frameworks were included in the analysis and their key points were included in the canvas. These nine frameworks were chosen as they represent either the main Italian and European documents in relation to data and AI ethics or because of their recognised contribution in literature and practice, inside or outside the EU. These frameworks are:

1. EU Ethics guidelines for trustworthy AI (European Commission, 2019)
2. Italian deontological rules for data treatment for statistical or scientific research purposes (Garante per la Protezione dei Dati Personali, 2018)
3. EU Data protection and privacy ethical guidelines (Experts Working Group on Data Protection and Privacy, 2009)
4. EU Ethics for Researchers (European Commission, 2008)
5. Open Data Institute's (2021) Data Ethics Canvas
6. The Toronto Declaration. Protecting the right to equality and non-discrimination in machine learning systems (Amnesty International et al., 2018)
7. Ada Lovelace Institute (2022) Algorithmic impact assessment: a case study in healthcare
8. Ethics as a Service (Digital Catapult, 2020)
9. National Institute of Standards and Technology (2023) Artificial Intelligence Risk Management Framework

The systematic review of ethical AI frameworks allowed us to identify the key themes to address in the canvas. In particular, we identified the fivefold thematic partition of the canvas: data, algorithms, methods of analysis, sociocultural elements, and functional requirements. This partition, its meaning, and its functionality are described in the next section.

Second, we analysed the applicable EU legislation. We included legislation that was adopted as the analysis was conducted as well as new regulations such as the AI Act, of which the practical effects are still unknown at the time of writing. This was to try and anticipate the consequences of emerging rules. As the effects will not be fully visible before a few years' time, the canvas will undergo constant monitoring and

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

adaptation. Through this analysis, we identified the legal elements that needed to be addressed in developing an AI tool. The full list of analysed legislative instruments includes the following.

10. AI Act
11. Data Act
12. Data Governance Act
13. Open Data Directive
14. General Data Protection Regulation

Such twofold analysis allowed us to elaborate on the first draft of the canvas.

We then drafted the first version of the canvas and requested feedback from the key PAT informants. Such feedback supported the creation of the canvas version that was submitted to PAT functionaries and directors during the following interview phase for validation. At this later stage, PAT functionaries and directors were approached for qualitative interviews to ensure that the canvas was clear for both technical and lay functionaries, as well as coherent in content with their everyday work. As we adopted value-sensitive design, we ensured that the ethical requirements were mainstreamed in AI development from the start and we identified such requirements both based on legal obligations and through PAT functionaries' foreseen needs and AI-based activities. We interviewed twelve people between 3 September 2024 and 28 January 2025 (Table 1). Of these, two had professional profiles that fit under the AI scientist definition, two under the AI engineer definition, while eight were categorised as AI users. An AI scientist is a software developer or researcher who develops new AI algorithms and models. This role often involves designing novel approaches and advancing the field of AI through theoretical and applied research. An AI engineer oversees fine-tuning, maintaining, and implementing the AI tool. It includes data engineers and machine learning engineers. Finally, the AI user is the final utiliser of the developed AI tool. AI users include public decision-makers such as policymakers, PA directors, and domain experts. Their participation in the development process is essential to render an AI tool accurately applicable to a given domain. All three roles must be involved in the development of the AI tool within a PA to ensure legal compliance, coherence between the software functioning and the administrative purpose it is built for, and full comprehension of the functioning of the tool by the final users. The identification of these three roles is based on the tripartition provided in the broader project in which the canvas fits (see

Table 1. *Interview participant profiles*

Interview date	Participant pseudonym	Profile	AI scientist/engineer/ user
03/09/2024	P1	PAT family services expert	AI engineer
05/09/2024	P2	Software developer and researcher	AI scientist
10/09/2024	P3	PAT director	AI user
11/12/2024	P4	PAT cartographer	AI user
11/12/2024	P5	PAT territorial planning engineer	AI user
12/12/2024	P6	PAT research funding programmes functionary	AI user
13/01/2025	P7	PAT cartography conservatory	AI user
13/01/2025	P8	PAT cadastre functionary	AI user
14/01/2025	P9	PAT ICT functionary	AI user
21/01/2025	P10	Trentino digitale - innovation area	AI engineer
21/01/2025	P11	Trentino digitale - innovation area	AI scientist
28/01/2025	P12	PAT IT/digitalisation functionary	AI user

Note: Trentino Digitale is a PAT-owned company developing technical IT solutions for PAT. Info available here: https://www.trentinodigitale.it/ (accessed 21 January 2025. Italian only).

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

Acknowledgements section). These three roles are to be understood as macro-categories that incorporate a diverse set of professional profiles. The canvas follows the same tripartition as better described in Section 4. This tripartition was also maintained in the interviews, ensuring that the sample featured at least two participants for each profile. Participants were recruited through snowball sampling. The recruiting process stopped at twelve participants as the information collected in the latest interviews did not add anything significantly new to the canvas design.

Interviews lasted from 30 to 75 minutes. Interview participants responded individually or in pairs (as indicated by the dates in Table 1). After each interview, the canvas was edited to incorporate suggestions that emerged from it. This way, the following participant could anonymously address previous participants' proposed changes and we could seek validation for them. Changes were retained if, based on the aggregate twelve interviews, they were considered needed. Appendix B features the interview questionnaire we set up. As the interviews were semi-structured, the interview structure could vary.

3.1. Methodological limitations

The canvas presented in this article is the result of an action research project and, as such, it is built based on the experiences and needs of the PAT, while being grounded in the existing European legislation and existing ethical framework for AI implementation in the public sector.

This entails three main sets of limitations. First, while aimed at being applicable across the EU, the canvas was built on the experience of one specific PA. Therefore, it may require adaptation to other contexts. To this end, the canvas is available under the CC BY-SA 4.0 licence to ensure that other actors can reuse it and change it according to their needs.

Second, while we sought to include a variety of PAT functionaries and directors from different sectors in the interviews, the canvas may be more applicable in certain sectors and situations rather than others, such as sectors that deal with individual- or family-level service provision.

Third, the EU legislation on AI and data is still developing at the time of writing, especially in terms of jurisprudence. This may require adaptation to the canvas content in the forthcoming years.

4. The canvas

The canvas (Appendix A) is a five-page graphic document that poses questions to drive the development of AI tools for PAs, accompanied by a two-page explanation on how to use it.

When the answer to a question is negative or unknown, the actors involved should take measures to ensure that the requirement is satisfied. For example, a question such as "did you adopt anonymisation or pseudonymisation techniques before data use?" should push the actors involved in adopting anonymisation or pseudonymisation strategies, if not already adopted. If the use case in question does not require anonymisation or pseudonymisation, this should be motivated. Either way, the canvas has raised a key privacy question and triggered reasoning among the actors involved in the making of an AI tool for PAs.

As anticipated in Section 3, the canvas identifies three macro-profiles, each of whom should be involved in filling the canvas and developing AI tools. These three macro-profiles are: AI scientist, AI engineer, and AI user.

Graphically, the canvas presents questions in three distinct colours: green, blue, and yellow. These colours represent AI scientists, AI engineers, and AI users, respectively. When questions present a mixed colour, it means the question is transversal to all the three profiles. Each profile is also represented by a symbol: square (AI scientist), circle (AI engineer), and triangle (AI user) to make the distinction among roles visible to people with colour-blindness. All these rules are illustrated in the introductory pages of the canvas as per Appendix A.

Each of the five pages of the canvas addresses a different key aspect of AI development, namely: data, algorithms, methods of analysis, socio-cultural elements, and functional requirements. These five elements have been elaborated based on the literature and the ethics frameworks analysed.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

Before addressing the questions posed in the five sections, the subjects involved in elaborating the AI tool need to define the use case. That is, they need to define the scope of application of the AI tool they are elaborating. In this phase, the canvas invites its users to think about whether AI is necessary to achieve the objectives. The canvas is meant to provide general guidelines, but each field and theme of application will have its own characteristics and features. Therefore, subject matter experts will need to be involved in AI tool development among AI scientists, AI engineers, and AI users.

As per the instructions contained in the canvas, this instrument is to be used at the beginning of the development phase to brainstorm about the characteristics of the AI instrument and reason around its ethical implications.

4.1. Data

As for the first section of the canvas, data constitutes the raw material for the functioning of AI. Good training and input data facilitate the obtainment of good algorithmic output. On the other hand, the treatment of input and training data can raise privacy issues (Open Data Institute, 2021), which are addressed in the canvas through appropriate questions such as: "Are there other datasets the use of which can yield unwanted consequences (e.g., re-identification)?" If the answer is "yes," the canvas prompts the AI scientist, engineer, and user to find a technical solution to prevent it. The same applies to social biases that can be incorporated into the data. The canvas openly addresses ethnic, gender, and class biases, but leaves the door open to addressing other biases depending on the context for which the given AI tool is developed. The canvas also addresses the question of data stewardship, that is, the management, update, and maintenance of data (Mons, 2018).

4.2. Algorithms

Second, algorithms constitute the core technical element for the functioning of AI. Despite good input data, wrong weights and optimisation parameters in the algorithm can perpetuate biases and/or yield suboptimal results (Ada Lovelace Institute, 2022). This could also be the result of the misuse of an algorithm for intents that differ from those for which it was originally programmed for. For this reason, the canvas puts the AI engineer in charge of defining the algorithm's purpose and the AI user in charge of checking if the use matches the purpose. Furthermore, this section seeks to anticipate the AI Act and address the aspect of certifying a system's risk level: minimum, limited, high, unacceptable (European Parliament, 2021).

This section also has a focus on social biases and their embeddedness in algorithms, which follows up strictly on the same questions asked in the data section.

4.3. Methods of analysis

This section deals with understanding the algorithmic output. That is, once an algorithm is created, trained, and validated on the chosen training datasets, and fed the data to analyse, it produces an output in terms of suggested decisions for the PA in question. In this context, it is important that AI users (that is, PA functionaries and directors) are capable of interpreting the algorithmic process and defining who takes responsibility for the algorithm-driven decision (Reddy, 2022). This entails understanding why and through which process the AI has suggested a certain course of action and establishing whether to follow the course suggested by the algorithm as well as taking legal responsibility in case of discrimination, copyright violations, or privacy violations among other aspects. To ensure these competences are possessed by AI users, PAs must establish ad hoc training activities for their personnel (Liao et al., 2020). Ideally, this includes training in both interpreting algorithmic outcomes and the basics of software programming. The latter is to ensure that AI users know the gist of how an AI is programmed and, potentially, can operate ad hoc tweaks to the AI instrument they are using.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

4.4. Social and cultural elements

This section introduces questions for the AI user to ensure a common understanding of data treatment terms between the AI user and the subjects of the data treatment (Pangrazio and Selwyn, 2019). Furthermore, it poses questions on the possibility of making data available for reuse: "Can the collected data be released as open data in aggregated form?" Conversely, this section also poses questions on the risks that come from the possibility of re-identifying data subjects through the use of third-party data: "Are there third-party data the use of which can yield unethical consequences (e.g., re-identification)?" The latter question is meant to trigger measures for data protection as needed.

Therefore, this section is designed to foreground how AI tools interact with the specific institutional and cultural context of the PA in which they are deployed, rather than treating ethics as a purely technical add-on. The canvas explicitly prompts users to reflect on the differentiated impact of AI systems on distinct social groups, and possible forms of exclusion or marginalisation. Fairness is therefore not reduced to computational metrics, but is framed as a contextual and qualitative concern that must be interpreted in dialogue with domain experts and, where feasible, with representatives of potentially affected communities.

4.5. Functional requirements

This section is for the user to define which characteristics the elaborated AI tool should possess. Based on the questions answered in the previous four sections, the user can define the desiderata for the AI tool. In other words, they can elaborate the requirements that the AI scientist and the AI engineer can then implement.

This section is willingly left at a high level. It is for AI users to specify the functions they would like to see incorporated in the AI instruments. For detailed documentation, developers can use tools such as AI product cards and data/model cards as indicated in the canvas. Such instruments are also further discussed in the next section.

5. Discussion

The canvas is the result of a systematic analysis of ethical frameworks and EU legislation, as well as qualitative interviews with AI scientists, engineers, and users from (or working with) the PA in the PAT (see Section 3).

This section discusses interview results and how they shaped the final form of the canvas as it appears in Appendix A. Second, it introduces other instruments that can be used in conjunction with the canvas further down the line of AI development to build clearer documentation. Third, it discusses the limits of the intended uses of the canvas. Finally, it discusses the deployment of technical tools to mainstream ethics in AI product development.

5.1. Interview results to validate the canvas

As per Table 1 in Section 3, we interviewed twelve people who work at (or with) the PAT.

To begin with, through our qualitative investigation we observed that PAT functionaries hold no expertise in terms of AI development. All software development work is delegated to third parties, chiefly Trentino Digitale, a PAT-owned company from which we interviewed an AI engineer and an AI scientist (P10 and P11). This is typical of many PAs and raised one main issue in terms of devising the canvas: for ethical principles to be mainstreamed in AI use—other than being embedded by design in the technology, which is only partially possible—the canvas had to address the problem of training and educating AI users. This is addressed in the Section 4.3.

Most AI users we interviewed expressed appreciation for the canvas graphics, which help to make a number of technical questions easier to address. In the words of an early interview participant (P3): "The

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

canvas is very easy to the eye at first glance, but quite technical. However, its pleasant aesthetics helps to 'digest' its technical nature." Nonetheless, many AI users struggled to familiarise themselves with technical terms. Indeed, AI-related technical terms in Italian are either English words or anglicisms, which makes them hard to grasp for lay users. Upon suggestion by interview participant P6, the Italian version of the canvas contains an extra page featuring a glossary. Indeed, P6 pinpointed that in their experience they had developed technical expertise in their domain, but no such expertise in AI and therefore needed support. Most participants stressed the fact that technical AI knowledge is unavailable inside the PAT as AI is merely introduced as an instrument developed outside the institution.

Furthermore, many AI users make a direct association between AI and chatbots in the interviews. Others, while being aware of the distinction, mostly interact with AI in the form of GPTs in their everyday jobs. For example, when inquired about AI usage, P8 said they use ChatGPT, but mostly for private rather than professional purposes. This led us to make an explicit distinction in the canvas between chatbots and other instruments. Section 4.2 requires the compilers to state which type of instrument they are seeking to create, distinguishing chatbots from the other types of AI instruments.

Technologists (that is, AI scientists and AI engineers) contributed by making questions more practice-oriented. For example, P2 contributed to the Section 4.5 by helping us to define the main implementation requirements. This participant stressed the need for questions in this last section that help developers to operationalise the principles and desiderata that emerge throughout the canvas. Furthermore, P2 helped us to construct a linkage between the canvas and documentation instruments such as AI Product Cards and Data Cards to be used further down the development pipeline (Celepija et al., 2024).

Altogether, these interviews allowed us to elaborate a canvas that reflects the experiences and needs of a local PA. The interviews covered both the content and the graphics of the canvas to ensure usability by a wide range of PA functionaries. Coherently with the literature on design thinking, in which canvas literature is nested (Brown, 2008), this instrument stands out as a user-friendly support for AI technologists and lay users alike when it comes to implementing AI in PAs.

5.2. Other instruments to use in connection with the canvas

The AI Ethics Canvas presents a high-level instrument for collectively planning the development, adoption, and deployment of AI instruments in PAs. It is not meant to provide detailed technical guidance for software development, but rather to aid the mainstreaming of ethics in the development of AI. The levels of ethical abstractions formulated in the canvas need to be translated into the AI instrument's technical functioning. To do so, one can adopt other support tools along with the canvas.

In recent years, various approaches to documentation practices have been developed. These methods primarily aim to improve the transparency and quality of creating, implementing, and assessing reliable AI systems. For example, in response to the increasing need for transparency and accountability in the development of AI applications, the concept of the Model Card (Mitchell et al., 2019) was introduced as a viable solution. Other notable examples include DataSheets for Datasets (Gebru et al., 2021) and Dataset Nutrition Label (Hallinan et al., 2020).

Recent advancements in AI systems documentation include System Cards (See, for example: https://ai.meta.com/research/publications/system-level-transparency-of-machine-learning/ [accessed 23 July 2025].) and AI Product Cards (Celepija et al., 2024), offering a holistic approach for developing and operationalising an AI solution. In particular, AI Product Cards are designed to enhance the reuse and adaptation of AI products within PA by utilising machine-readable specifications derived from model and data cards. This framework integrates essential descriptive and formal specifications to coordinate the execution of associated code references and other artefacts, enabling the seamless reuse of the complete AI product.

From the perspective of the procurement of AI in the public sector, ethical frameworks and methodologies have been developed to identify ethical risks and to prevent, assess, and mitigate their potential impact on individuals' lives (Kawakami et al., 2024) (This includes the HUDERIA method adopted by the Council of

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

Europe: https://www.coe.int/en/web/artificial-intelligence/huderia-risk-and-impact-assessment-of-ai-systems (accessed 23 July 2025).). The enforcement of the EU AI Act will establish a legal obligation for public institutions and organisations to document their AI products and associated risk management information (Golpayegani et al., 2024), which will be essential in proving compliance with the regulation.

On a different note, one can accompany the use of the canvas with these ethical evaluation instruments to balance out potential biases and shortcomings embedded in each such instrument. For example, the Assessment List for Trustworthy Artificial Intelligence elaborated by the EU High-level Expert Group on Artificial Intelligence along with the EU ethics guidelines for trustworthy AI is a powerful tool to conduct a self-assessment of the ethics and lawfulness of an AI instrument (High-Level Expert Group on Artificial Intelligence, 2020).

The list of existing instruments to facilitate software documentation and development increases day by day. Nonetheless, this subsection sought to provide a general overview of the existing tools to aid AI development and deployment that can be used along with the canvas to achieve a deeper level of technical detail.

5.3. Scope and limits. Or: What the canvas is not for

The canvas presented in Appendix A and discussed above in this section should not be construed as a checklist waiving individual and institutional responsibility. It is rather an instrument that guides PA functionaries and their providers (e.g., software developers from outsourced services) to mainstream ethics in the functional requirements of an AI tool for public services.

While canvases offer a streamlined visual approach to co-design, they are not without drawbacks. Chiefly, presenting too many fields at once can induce cognitive overload, reducing participants' ability to focus on key trade-offs (Osterwalder et al., 2010). Nonetheless, canvas use actually improves project outcomes in governmental or complex technical settings. Likewise, efforts to co-create a "fit-for-purpose" canvas for open data initiatives found initial versions promising but provisional, highlighting the need for systematic outcome assessments in real-world use (Gouscos et al., 2022).

The canvas poses questions and seeks to trigger reasoning over legal and ethical grey areas where they exist. Nonetheless, it is general, and specific ethical questions that arise in niche cases may have fallen out of its scope. In those cases, it is essential for the PA to address those aspects independently of whether they are included in the canvas. Furthermore, the canvas refers to the types of algorithmic and data biases most often addressed in literature. However, specific biases may arise in given fields: it is up to domain experts to identify them. In other words, PA functionaries are expected to be experts in their own field capable of recognising specific liabilities and elements of danger that apply to their own area of expertise, tackling these issues alongside AI scientists and engineers to mainstream ethical requirements in the AI product.

Furthermore, it is up to each PA to define responsibilities within their own structures. Indeed, the Section 4.3 asks "who takes responsibility in the following cases?," thus pointing towards the need for PAs to define responsible subjects within their hierarchy.

On a final note, it must be re-stressed that this canvas was devised for EU-based PAs only. While it can be adaptable to use by private actors or non-EU PAs, the AI Ethics Canvas demonstrates strong potential for EU-wide transferability through its alignment with harmonised regulations like the AI Act and General Data Protection Regulation. These frameworks ensure baseline applicability across member states, transcending local peculiarities. To enhance adaptability, an open-source licensing framework has been adopted for future iterations of national/regional guidelines, allowing PAs to append jurisdiction-specific prompts without altering core content.

5.4. Technical tools in connection with the canvas

5.4.1. Tools for data

The canvas developed in this study emphasises the need to address data biases as part of ethical AI development for PAs. To support this goal, several tools have emerged that specifically tackle data bias during the collection and preprocessing stages, aligning closely with the Section 4.1 of the canvas. AI

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

Fairness 360 (AIF360) exemplifies a robust tool designed specifically to tackle data bias before training (Trusted-AI, 2018). Through a suite of over seventy bias detection metrics and various preprocessing algorithms, AIF360 enables users to assess fairness across different demographic groups by re-weighting or transforming data to reduce bias, promoting a balanced representation in datasets. This tool is particularly useful in fields such as healthcare and finance, where ethical considerations are critical. Complementing AIF360, Fairlearn provides a preprocessing algorithm which effectively adjusts data weights for sensitive attributes, promoting a more balanced dataset (Weerts et al., 2023). This adjustment is crucial, as it helps mitigate the risk of propagating data bias through the machine learning pipeline. Fairlearn's data balancing techniques ensure that sensitive attributes do not overly influence predictions, thereby promoting fairness during the training phase of AI models.

These tools exemplify actionable strategies for addressing data biases, providing practical avenues for PAs to operationalise the ethical data principles embedded in the canvas.

5.4.2. Tools for algorithms

Following the emphasis on addressing data biases, the canvas also underscores the importance of tackling biases embedded in algorithms. Section 4.2 of the canvas emphasises the need for fair and transparent algorithmic design to ensure ethical AI outcomes in PAs. Several tools directly support these objectives by focusing on algorithmic fairness during design, training, and post-training phases.

Randomized Threshold Optimizer (RTO) stands out as a post-processing tool that modifies the decision thresholds of a trained model to achieve statistical parity, effectively mitigating biases present after the model has been trained (Alabdulmohsin and Lucic, 2021). By introducing a degree of randomness in decision-making near these thresholds, RTO can ensure that different demographic groups are treated fairly in terms of output distribution without significantly compromising accuracy.

FAT (Fairness, Accountability and Transparency) Forensics directly integrates fairness, accountability, and transparency into predictive systems. It enables practitioners to inspect predictions across population subgroups, supporting the debugging and analysis of potential algorithmic biases (Sokol et al., 2020). The canvas encourages the use of such transparency-driven tools to enhance accountability and build trust among public sector stakeholders.

By leveraging these versatile tools, PAs can effectively address the challenges outlined in the canvas's "Algorithms" section, equipping them to design and implement AI systems that uphold fairness, transparency, and accountability.

5.4.3. Tools for methods of analysis

Ensuring that AI models operate ethically and transparently requires rigorous methods of analysis, and several tools support this critical stage by allowing practitioners to explore model behaviours in hypothetical scenarios, visualise biases, and assess societal impacts. The What-If Tool (WIT) is a powerful open-source tool that facilitates interactive model exploration by allowing users to adjust feature values in individual data instances and observe the effects on model predictions (Wexler et al., 2020). Through this probing, users can compare model behaviours across different demographics, thereby identifying disparities and potential biases.

FAT Forensics, previously introduced for its bias detection capabilities, extends its functionality to support comprehensive transparency evaluations during analysis. By integrating seamlessly, it allows practitioners to rigorously scrutinise model predictions and detect algorithmic biases, making it an indispensable tool for ensuring responsible deployment in public sector applications.

The Ethics & Algorithms Toolkit further addresses the socio-technical aspects of AI deployment (Wong and Madaio, 2023). Designed specifically for public sector use, this toolkit includes guides and worksheets to help practitioners navigate the ethical implications of algorithm management, fostering transparency and accountability throughout the AI lifecycle.

Together, these tools align with the canvas's emphasis on the Section 4.3, equipping PAs with practical resources during the critical stages of AI model evaluation and deployment.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

6. Conclusions

This paper presented a canvas to aid the implementation of AI tools in PAs. The canvas available in Appendix A is a finalised version that underwent further validation through a two-hour workshop, involving participants from the project who ethically assessed three use cases against public values and EU regulatory principles. Participants worked systematically through all five sections of the canvas, with more time allocated to Data, Algorithms, and Methods of Analysis, using printed canvases and post-its to document reflections, risks, and recommendations. Unlike pre-existing ethical frameworks, which mainly provide high-level principles and checklists, the canvas offered a concrete, visual, and structured workflow that operationalised those principles. The workshop discussions showed that the canvas supports compliance-oriented reflection and fosters systematic evaluation of AI tools by prompting participants to identify biases, data and model limitations, governance gaps, and mitigation strategies in a comparable way across cases.

These activities fit within a broader framework of further validation and maintenance of the canvas. Indeed, the canvas will need to be adapted as legislation emerges.

Altogether, the canvas represents a useful working tool for PAs who plan to take their first steps into developing and/or adopting AI in their everyday work. Furthermore, the canvas can provide guidance to PAs in choosing external AI providers (e.g. companies) and could support the latter in approaching PAs to co-develop AI instruments.

In conclusion, the canvas presented in this article represents a tool to support the digitalisation of the public sector.

Data availability statement. The canvas is available here: https://github.com/dclfbk/canvasaiethics/or https://doi.org/10.5281/zenodo.21282489.

The analysed documents (ethical frameworks and legal texts) are listed in the methodological section (Section 3) and publicly available here: https://github.com/dclfbk/canvasaiethics/tree/main/docs/documents%20for%20analysis.

Interviews are confidential, thus pseudonyms are used and the transcripts are not shared. The interview questionnaire is, however, available in Appendix B, while the Methods section illustrates the participants' profiles.

Acknowledgements. We thank Paolo Massa, Nadia Mana, Stefano Micocci, Giulia Deppieri, Raman Kazhamiakin, Luisa Bentivogli, Beatrice Savoldi and Alberto Amaduzzi for reviewing and commenting on the canvas in their different roles and capacities. We also thank Marco Pistore for supervising the work as the Director of the Digital Society Centre of FBK and Samantha Riccadonna from the ICT and Digital Transformation Service of the Autonomous Province of Trento for commenting on the guidelines and helping in the validation process. Finally, we thank Andrea Nicolini, Silvia Valentini, and Matteo Gerosa for commenting on the canvas and the guidelines at various stages of the process, as well as helping with reaching out to interview participants. A special thanks goes also to Danilo Giampiccolo for his technical support. Finally, we thank our twelve anonymous interview participants for their contribution and the anonymous reviewers for their comments.

Author contributions. Conceptualization: R.N, M.N.; Data curation: R.N., P.G.B., A.C., M.N.; Formal analysis: R.N., P.G.B., A.C.; Funding acquisition: M.N.; Investigation: R.N., P.G.B., A.C.; Methodology: R.N., M.N.; Project administration: M.N.; Supervision: M.N.; Validation: M.U., M.N.; Visualisation: M.U.; Writing - original draft: R.N., P.G.B., A.C., M.U., M.N.; Writing - review & editing: R.N., P.G.B., A.C., M.U., M.N.

Funding statement. This work is part of the project "Progetto Bandiera – PNC (Piano Nazionale per gli investimenti Complementari) - "AI data: Intelligenza artificiale nel sistema della PA," CUP C49G22001020001. Flagship Project – PNC (Piano Nazionale per gli investimenti Complementari - National Plan for Complementary Investments) - 'AI data: Artificial Intelligence in the Public Administration'" CUP C49G22001020001. The Autonomous Province of Trento serves as the projects' Responsible Entity, with FBK acting as a leading implementation partner.

Competing interests. The authors report no conflict of interest.

References

Ada Lovelace Institute (2022) *Algorithmic Impact Assessment: A Case Study in Healthcare.* Available at https://tinyurl.com/mucfc6pc.

Alabdulmohsin I and Lucic M (2021) *A Near-Optimal Algorithm for Debiasing Trained Machine Learning Models.* https://doi.org/10.48550/arXiv.2106.12887.

Amnesty International and Access Now (2018) *The Toronto Declaration: Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems.*

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

Brown T (2008) Design thinking. *Harvard Business Review 86*(6), 84–92.

Carbone F, **Alivernini A**, **Bascietto M**, **Oreti L and Barbarese F** (2023) The digitalization framework of the national forest system at 2020. . *Forest - Rivista di Selvicoltura ed Ecologia Forestale 20*, 30–38. https://doi.org/10.3832/efor4265-020.

Celepija A, **Palmero Aprosio A**, **Lepri B and Kazhamiakin R** (2024) *AI Product Cards: A Framework for Code-Bound Formal Documentation Cards in the Public Administration*. Cambridge: Data & Policy, pp. 1–17.

Cornish F, **Breton N**, **Moreno-Tabarez U**, **Delgado J**, **Rua M**, **de-Graft Aikins A and Hodgetts D** (2023) Participatory action research. *Nature Reviews Methods Primers 3*(1), 34. https://doi.org/10.1038/s43586-023-00214-1.

Digital Catapult (2020) *Lessons in Practical AI Ethics Taking the UK's AI Ecosystem from 'What' to 'How'.* Available at https://tinyurl.com/4wz9vahf

European Commission (2008) *Ethics For Researchers: Facilitating Research Projects Involving Animals, Humans, or Ethical Implications.* Available at https://tinyurl.com/3hxcyf33.

European Commission (2019) *Ethics Guidelines for Trustworthy AI.* Available at https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

European Parliament (2021) *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.* Available at https://tinyurl.com/mpf9dyzu.

European Union (2023a) *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act).* Available at https://www.zotero.org/google-docs/?mBpniF.

European Union (2023b) *Regulation of the European Parliament and of the Council on Harmonised Rules on Fair Zccess to and Use of Data (Data Act).* Available at https://tinyurl.com/4u8ker97.

Experts Working Group on Data Protection and Privacy (2009) *Data Protection and Privacy Issues Ethical Guidelines.* Available at https://tinyurl.com/yc8j6wcd.

Fatima S, **Desouza KC**, **Buck C and Fielt E** (2022) Public AI canvas for AI-enabled public value: A design science approach. *Government Information Quarterly 39*(4), 101722. https://doi.org/10.1016/j.giq.2022.101722.

Garante per la Protezione dei Dati Personali (2018) *Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica effettuati nell'ambito del Sistema Statistico nazionale*. https://tinyurl.com/5n8ywtrn.

Gebru T, **Morgenstern J**, **Vecchione B**, **Vaughan JW**, **Wallach H and Crawford K** (2021) Datasheets for datasets. *Communications of the ACM 64*(12), 86–92. https://doi.org/10.1145/3458723.

Golpayegani D, **Hupont I**, **Panigutti C**, **Pandit HJ**, **Schade S**, **O'Sullivan D and Lewis D** (2024) AI cards: Towards an applied framework for machine-readable AI and risk documentation inspired by the EU AI act. *arXiv.* https://doi.org/10.48550/arXiv.2406.18211.

Gouscos D, **Melidis A and Tambouris E** (2022) ECCOdata: A proposal of an Empirical Co-created Canvas for Opening up Data of Public Interest (short paper). EGOV-CeDEM-ePart-*. https://ceur-ws.org/Vol-3399/paper20.pdf

Hallinan D, **Leenes R**, **Gutwirth S and and De Hert P** (eds) (2020) *Data Protection and Privacy: Data Protection and Democracy.* Oxford: Hart Publishing. https://doi.org/10.5040/9781509932771.

High-Level Expert Group on Artificial Intelligence (2020) *The Assessment List for Trustworthy Artificial Intelligence.* Available at https://altai.insight-centre.org/.

International Data Spaces Association (2021) *New Business Models for Data Spaces Grounded in Data Sovereignty.* Available at https://internationaldataspaces.org/wp-content/uploads/IDSA-Position-Paper-New-Business-Models-sneak-preview-version.pdf.

Kawakami A, **Coston A**, **Zhu H**, **Heidari H and Holstein K** (2024) The situate AI guidebook: Co-designing a toolkit to support multi-stakeholder early-stage deliberations around public sector AI proposals. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–22. New York: Association for Computing Machinery. https://doi.org/10.1145/3613904.3642849.

Kronsbein T and Mueller RM (2019) Data thinking: A canvas for data-driven ideation workshops. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, pp. 561–570. http://hdl.handle.net/10125/59496.

Liao QV, **Gruen D and Miller S** (2020) Questioning the AI: Informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–15. https://doi.org/10.1145/3313831.3376590.

Manzoni M, **Medaglia R**, **Tangi L**, **Van Noordt C**, **Vaccari L and Gattwinkel D** (2022) *AI Watch Road to the Adoption of Artificial Intelligence by the Public Sector: A Handbook for Policymakers, Public Administrations and Relevant Stakeholders*. Luxembourg: European Union. https://doi.org/10.2760/288757.

Medaglia R, **Gil-Garcia JR and and Pardo TA** (2023) Artificial intelligence in government: Taking stock and moving forward. *Social Science Computer Review 41*(1), 123–140. https://doi.org/10.1177/08944393211034087.

Mikalef P, **Fjørtoft SO and Torvatn HY** (2019) Artificial intelligence in the public sector: A study of challenges and opportunities for Norwegian municipalities. In Pappas IO, Mikalef P, Dwivedi YK, Jaccheri L, Krogstie J and Mäntymäki M (eds), *Digital Transformation for a Sustainable Society in the 21st Century.* Cham: Springer, pp. 267–277. https://doi.org/10.1007/978-3-030-29374-1_22.

Mitchell M, **Wu S**, **Zaldivar A**, **Barnes P**, **Vasserman L**, **Hutchinson B**, **Spitzer E**, **Raji ID and Gebru T** (2019) Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 220–229. https://doi.org/10.1145/3287560.3287596.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

Mons B (2018) *Data Stewardship for Open Science: Implementing FAIR Principles*, 1st Edn. Boca Raton, Florida: Chapman and Hall/CRC. https://doi.org/10.1201/9781315380711.

National Institute of Standards and Technology (2023) *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Gaithersburg, US: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1.

Nurcahyo A, **Suroso J and Wang G** (2022) The artificial intelligence (AI) model canvas framework and use cases. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika 8*(1), 1. https://doi.org/10.26555/jiteki.v8i1.22206.

Ojasalo J and Ojasalo K (2018) Service logic business model canvas. *Journal of Research in Marketing and Entrepreneurship 20*(1), 70–98. https://doi.org/10.1108/JRME-06-2016-0015.

Open Data Institute (2021) *Data Ethics Canvas.* Available at https://www.zotero.org/google-docs/?mBpniF.

Organisation for Economic Co-operation and Development (2024) *Governing with Artificial Intelligence: Are Governments Ready?*. https://doi.org/10.1787/26324bc2-en.

Organisation for Economic Co-operation and Development - Observatory of Public Sector Innovation (2025) *Case Study Library.* Available at https://oecd-opsi.org/case_type/opsi/.

Osterwalder A, **Pigneur Y and Clark T** (2010) *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. Hoboken, N.J.: Wiley.

Pangrazio L and Selwyn N (2019) 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society 21*(2), 419–437. https://doi.org/10.1177/1461444818799523.

Reddy S (2022) Explainability and artificial intelligence in medicine. *The Lancet Digital Health 4*(4), e214–e215. https://doi.org/10.1016/S2589-7500(22)00029-2.

Sokol K, **Hepburn A**, **Poyiadzi R**, **Clifford M**, **Santos-Rodriguez R and Flach P** (2020) FAT forensics: A python toolbox for implementing and deploying fairness, accountability and transparency algorithms in predictive systems. *Journal of Open Source Software 5*(49), 1904. https://doi.org/10.21105/joss.01904.

Teo SA (2022) How artificial intelligence systems challenge the conceptual foundations of the human rights legal framework. *Nordic Journal of Human Rights 40*(1), 216–234. https://doi.org/10.1080/18918131.2022.2073078.

Trusted-AI (2018) *AI Fairness 360.* Available at https://github.com/ibm/aif360.

Van Der Hoven J and Manders-Huits N (2020) Value-sensitive design. In Miller KW and Taddeo M (eds), *The Ethics of Information Technologies*, 2nd Edn. Routledge. https://doi.org/10.4324/9781003075011.

van Noordt C and 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. https://doi.org/10.1016/j.giq.2022.101714.

Weerts H, **Dudík M**, **Edgar R**, **Jalali A**, **Lutz R and Madaio M** (2023) *Fairlearn: Assessing and Improving Fairness of AI Systems.* Available at https://fairlearn.org.

Wexler J, **Pushkarna M**, **Bolukbasi T**, **Wattenberg M**, **Viegas F and Wilson J** (2020) The what-if tool: Interactive probing of machine learning models. *IEEE Transactions on Visualization and Computer Graphics 26*(1), 56–65. https://doi.org/10.1109/TVCG.2019.2934619.

Wirtz BW, **Weyerer JC and Sturm BJ** (2020) The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration 43*(9), 818–829. https://doi.org/10.1080/01900692.2020.1749851.

Wong RY and Madaio MA (2023) *45 Seeing Like a Toolkit: How Toolkits Envision the Work of AI Ethics.* https://doi.org/10.1145/3579621.

Appendix A. AI Ethics Canvas

Updated versions in Italian and English are here: https://github.com/dclfbk/canvasaiethics (12 November 2025). This repo contains the AI Ethics Canvas presented in this article, as well as a fork oriented towards companies.

The AI Ethics Canvas is licensed under CC BY-SA 4.0.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

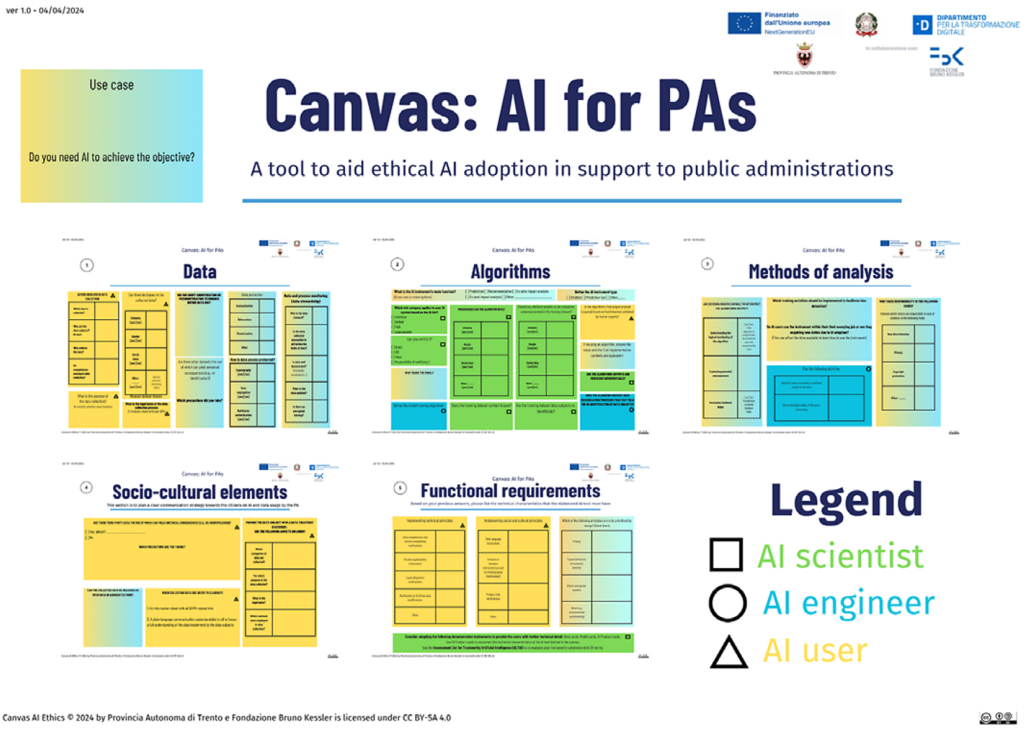


Figure A1. *Canvas index.*

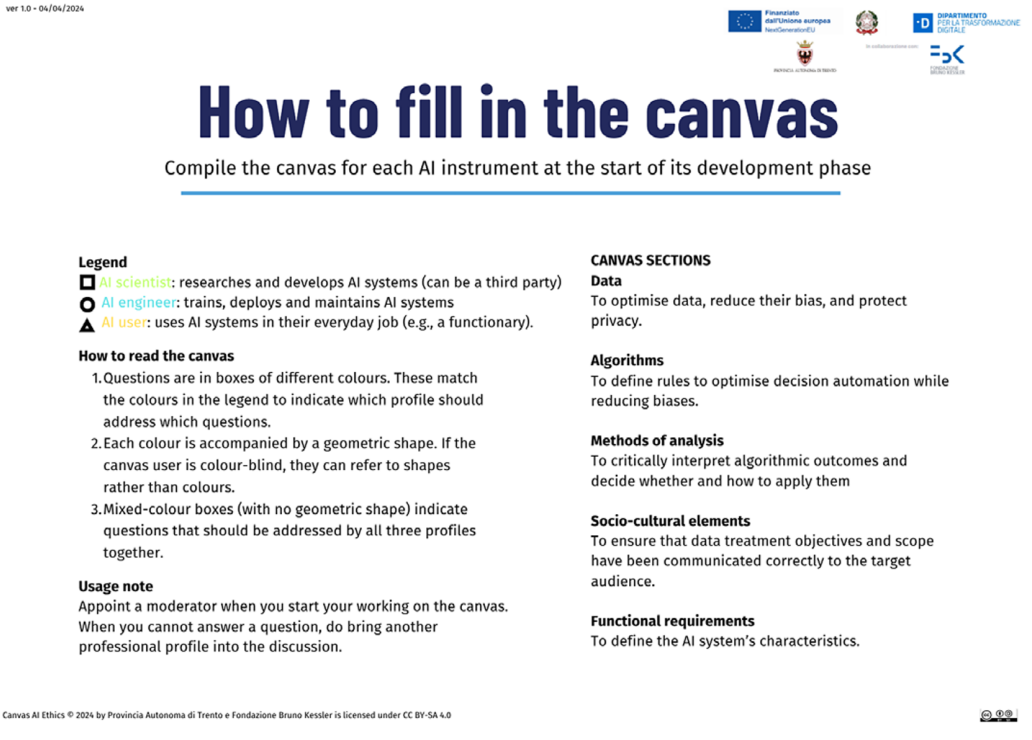


Figure A2. *Canvas instructions.*

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

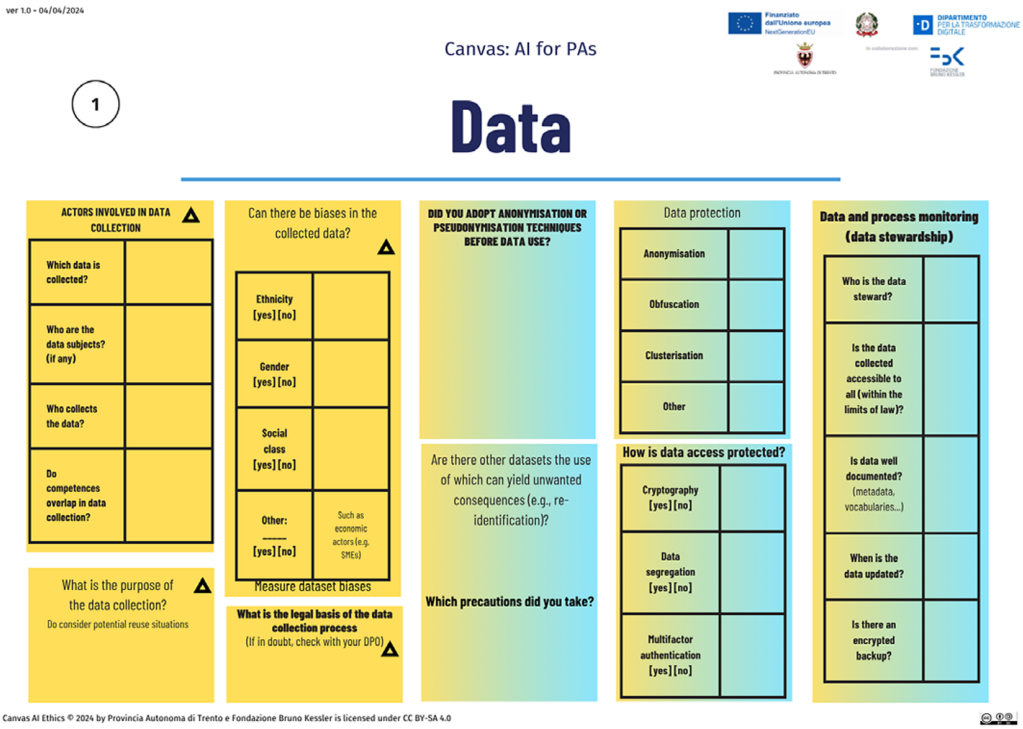


Figure A3. *Canvas section: Data.*

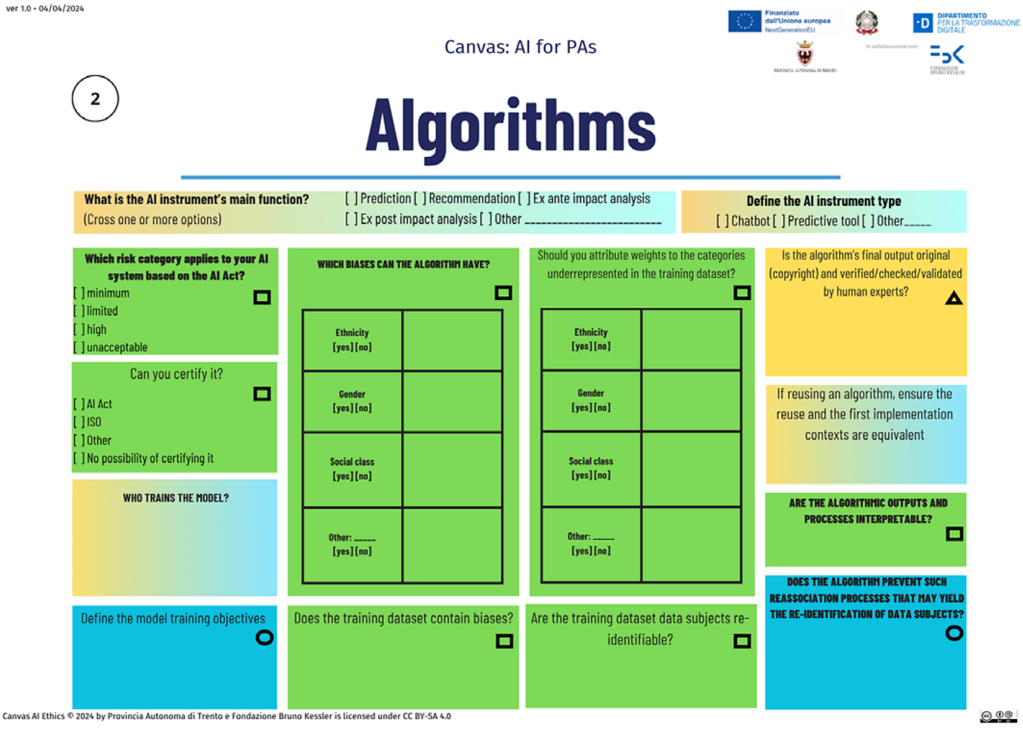


Figure A4. *Canvas section: Algorithms.*

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

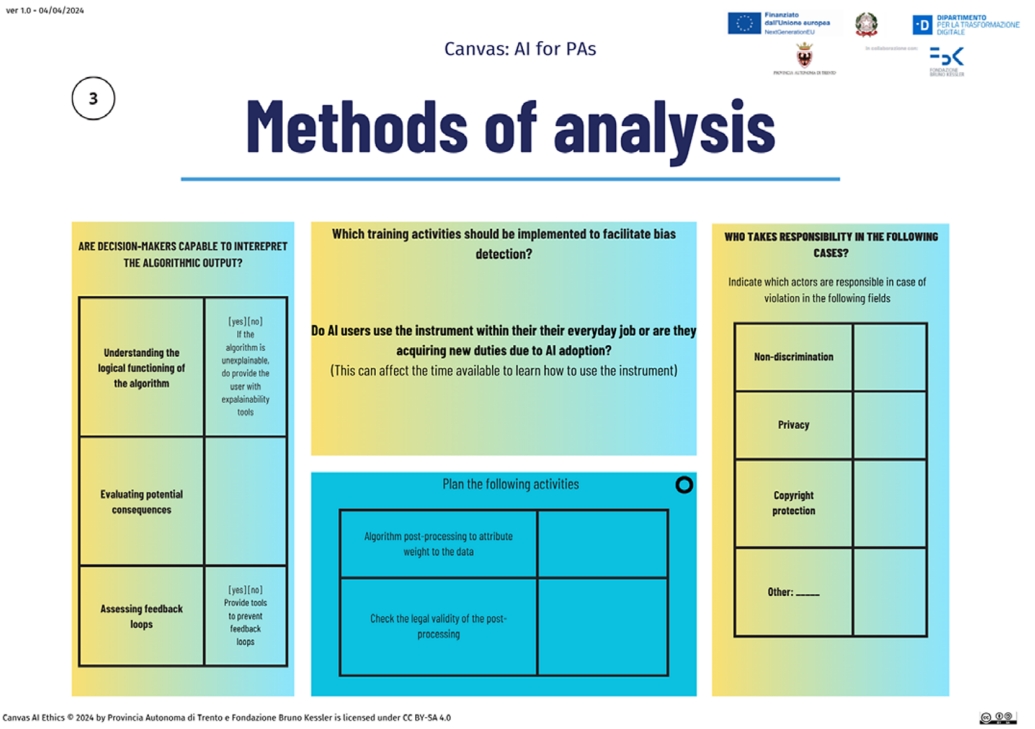


Figure A5. *Canvas section: Methods of analysis.*

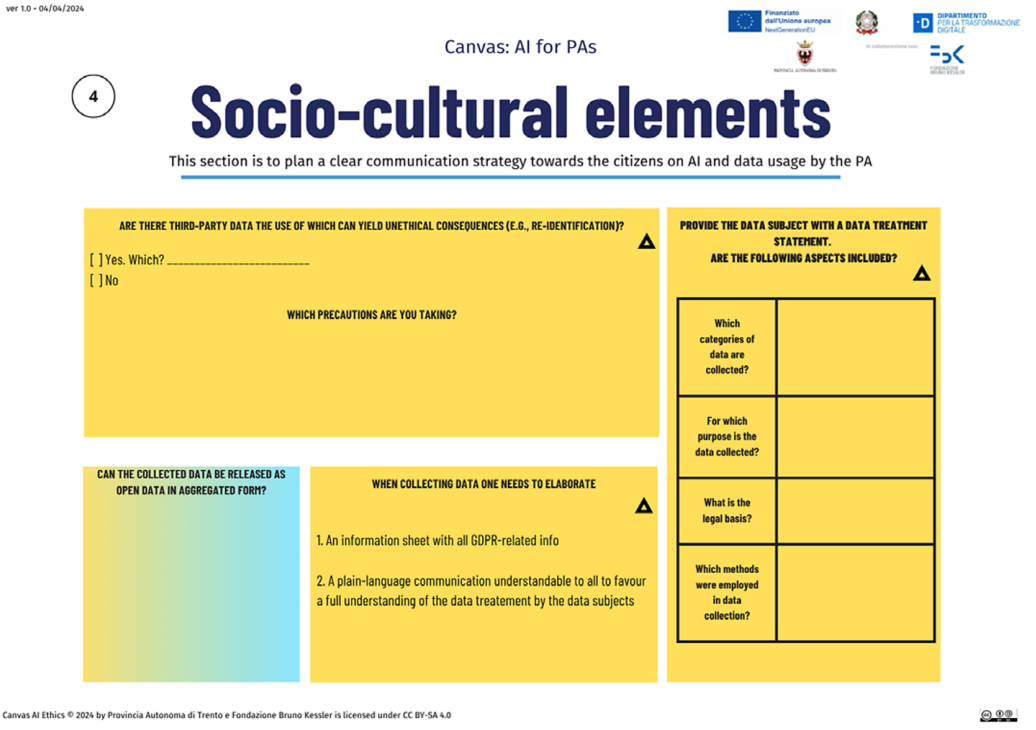


Figure A6. *Canvas section: Socio-cultural elements.*

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.

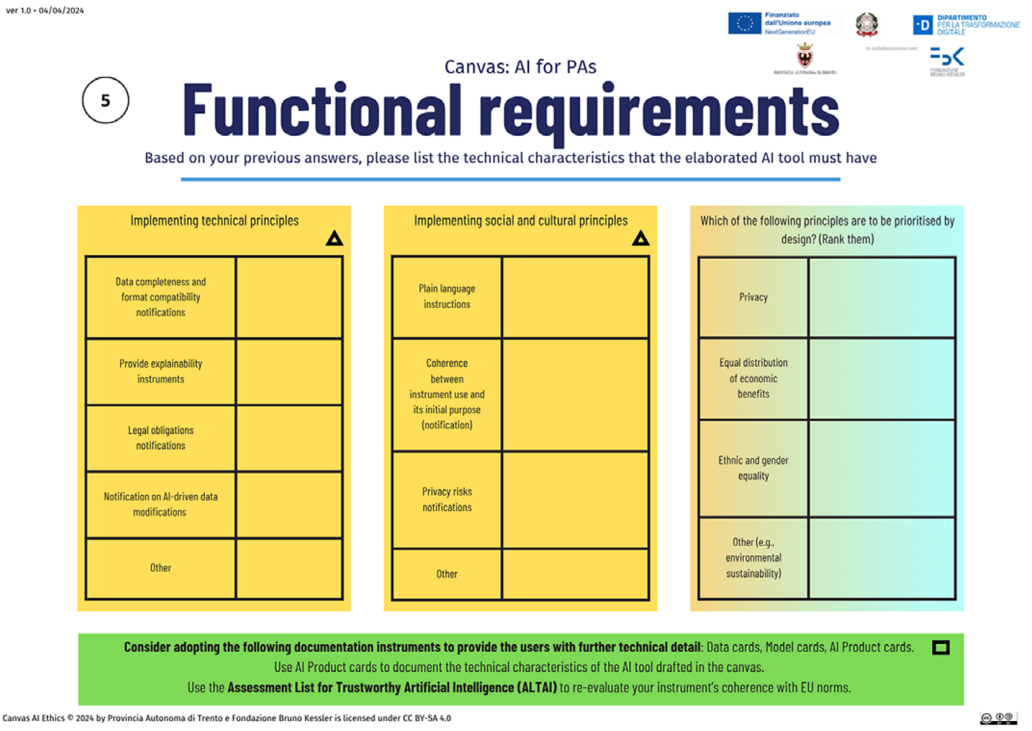


Figure A7. *Canvas section: Functional requirements.*

Appendix B. Interview Questionnaire

(Original questions were in Italian, as the PA functionaries interviewed were based in the Autonomous Province of Trento)

1. How do you use AI in your everyday work? (if used)
2. How do you think you will use AI in the future, in light of the PAT's ongoing AI-oriented projects?
3. About the canvas:
   a. Which parts are less clear?
      i. How do you suggest modifying them?
   b. Which aspects are inapplicable to your everyday work, if any?
      i. How do you suggest modifying them?
   c. Are there aspects you think were not sufficiently addressed in the canvas?

Cite this article: Nanni R, Bizzaro PG, Celepija A, Usmani M and Napolitano M (2026). AI Ethics Canvas: A graphical tool to design and deploy ethical artificial intelligence for public administrations. *Data & Policy*, 8: e40. doi:10.1017/dap.2026.10088

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 05:58:17, subject to the Cambridge Core terms of use.