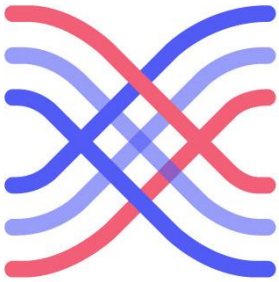


AI4Gov

Trusted AI for Transparent Public Governance
fostering Democratic Values

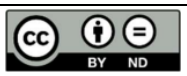
Deliverable 6.4

Stakeholders' Feedback and Evaluation of the AI4Gov Use Cases V1

31.12.2024

Version 1.0

Funded by the European Union	*Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Agency. Neither the European Union nor the granting authority can be held responsible for them.*

PROPERTIES	
Dissemination level	Public
Version	1.0
Status	Final version
Beneficiary	ViLabs LTD
License	 This work is licensed under a Creative Commons Attribution-No Derivatives 4.0 International License (CC BY-ND 4.0). See: https://creativecommons.org/licenses/by-nd/4.0/

AUTHORS		
	Name	**Organisation**
Document leader	Danai Kyrkou	ViLabs
Participants	Eri Goga	ViLabs
	Matej Kovačič	JSI
	Celia Parralejo Cano	DPB
	Dimitris Apostolopoulos	VVV
	Lila Iosif	MT
	Veroniki Diamantara	MT
	Konstantina Tsakopoulou	MT
	Kostis Mavrogiorgos	UPRC
	George Manias	UPRC
Reviewers	Fabiana Fournier	IBM
	Stroia-Vlad Iuliana	SIE

VERSION HISTORY				
Version	Date	Authors	Organisation	Description
0.1	22/10/2024	Danai Kyrkou, Eri Goga	VIL	ToC & introduction
0.2	04/12/2024	Danai Kyrkou, Eri Goga	VIL	1st draft after the partners' comments on the ToC
0.3	17/12/2024	Danai Kyrkou, Eri Goga, Matej Kovačič, Celia Parralejo Cano, Dimitris Apostolopoulos, Lila Iosif, Veroniki Diamantara, Konstantina Tsakopoulou, Kostis Mavrogiorgos, George Manias	VIL, DPB, VVV, MT, JSI, UPRC	1st draft ready for internal review
0.4	19/12/2024	Fabiana Fournier, Stroia-Vlad Iuliana, George Manias	IBM, SIE, UPRC	2nd draft with reviewers' comments
1.0	20/12/2024	Danai Kyrkou, Eri Goga	VIL	Incorporation of the comments – Final version ready for submission

Table of Contents

List of figures

List of Tables

Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
CSV	Comma-Separated-Values
DGF	Data Governance Framework
DPB	Diputación Provincial de Badajoz
DWTS (or EDAR in Spanish)	Decentralised Wastewater Treatment System
EU	European Union
GDPR	General Data Protection Regulation
GPS	Global Positioning System
HRF	Holistic Regulatory Framework
IoT	Internet of Things
JSI	Jozef Stefan Institute
KPI	Key Performance Indicator
MD	Markdown
MT	Ministry of Tourism
OECD	Organisation for Economic Cooperation and Development
OwiD	Our World In Data
PDF	Portable Document Format
RFID	Radio Frequency Identification
SAX	Situation-Aware Explainability
SCADA	Supervisory Control and Data Acquisition
SDG	Sustainable Development Goal
SME	Small Medium Enterprise
UC	Use Case
UEQ	User Experience Questionnaire
URL	Uniform Resource Locator
US	User Stories
VVV	Vari - Voula - Vouliagmeni
WP	Work Package
WWTP (or ETAP in Spanish)	Wastewater Treatment Plants
XAI	eXplainable AI

Abstract

This document, D6.4 “Stakeholders’ Feedback and Evaluation of the AI4Gov Use Cases V1”, was developed in the context of WP6 “Use Case Implementation, Validation, and Evaluation”, and more specifically, it is connected to T6.5 “Stakeholders’ Feedback and Evaluation”. The report includes the pilot methodology, the evaluation methodology, and the results of the first validation phase of the AI4Gov piloting activities. It is the first of the two versions on feedback and evaluation, covering the piloting activities up to M24.

The results of this first validation will feed WP3 and WP4 and help the technical partners improve the AI4Gov tools and release their final version by M27, where all technical tasks finish. In this deliverable, the strengths and weaknesses of the tools are demonstrated through the evaluation methodology of the project based on efficiency, usability, and trust. At this point, all tools are almost finalised, and this 1st validation phase provided some indications for final corrections. It worths mentioning that the results were mostly optimistic, showing the potential and value of the AI4Gov tools.

1 Introduction

1.1 Purpose and scope of the deliverable

This deliverable is the result of the work that has taken place under WP6 - Use Cases Implementation, Validation and Evaluation. This WP started in month 1 and ends in month 36, so it will run for the whole lifecycle of the project. The deliverable contains the evaluation results of the piloting activities so far. The purpose of D6.4 is to present the AI4Gov evaluation methodology, the evaluation tools, and the results of the first validation phase of the project, up to M24. The input will help the technical partners identify potential weaknesses in the tools and proceed optimising them before their finalisation. In addition, the evaluation methodology itself will be assessed to update it for the second validation phase.

1.2 Document structure

The deliverable is structured as follows: **Chapter 1** introduces the document, including the purpose and scope, document structure, and any updates compared to the previous version, while it presents the structure of WP6, the relation to other WPs, the target audience, and finally the data management of the evaluation process. **Chapter 2** describes the evaluation cycle, the timeline AI4Gov is following, and the evaluation methodology, as well as the contribution of the UCs to the SDGs. **Chapter 3** presents the results of the 1st validation phase for each use case (UC) up to M24, as well as some comparative reflections. **Chapter 4** summarises the key findings and next steps. **Chapter 5** includes the reference list, and finally, **chapter 6** is the appendix, with all the questionnaires that were used during this 1st validation phase.

1.3 Updates with respect to previous version

This is the first of the two versions of the deliverable on feedback and evaluation of the AI4Gov UC scenarios. The second version will be delivered in M36 (December ’25). However, a short description of the evaluation methodology has been provided in D6.2.

1.4 WP6 Structure

WP6 is the WP associated with the piloting activities, devoted to deploying, operating, validating, and evaluating the use case scenarios with the active engagement of the public organisations and policy makers of the consortium. It breaks down into five tasks:

- **T6.1** Detailed Specification of Scenarios and Use Case Preparation
- **T6.2** Data-Driven Sustainability for a Liveable Badajoz
- **T6.3** Using AI for Sustainable Development and the European Green Deal
- **T6.4** Trustworthy Data-Driven Touristic Policies
- **T6.5** Stakeholders' Feedback and Evaluation

The first task is the preparatory task that supports the design and development of the Use Case Scenarios and the user requirements. Tasks T6.2, T6.3, and T6.4 correspond to the three pilots that will test the AI4Gov technologies. Finally, T6.5 is the task that coordinates the evaluation of the pilot results and maps the gaps and needs that arise. In total, the WP6 has 5 deliverables:

- **D6.1** - Specification of UC Scenarios and Planning of Integration and Validation Activities V1 (corresponding to T6.1)
- **D6.2** - Specification of UC Scenarios and Planning of Integration and Validation Activities V2 (corresponding to all WP6 tasks)
- **D6.3** - Specification of UC Scenarios and Planning of Integration and Validation Activities V3 (corresponding to all WP6 tasks)
- **D6.4** - Stakeholders' Feedback and Evaluation of the AI4Gov Use Cases V1 (corresponding to T6.5)
- **D6.5** - Stakeholders' Feedback and Evaluation of the AI4Gov Use Cases V2 (corresponding to T6.5)

As mentioned at the beginning, WP6 is active throughout the whole lifecycle of the project.

1.5 Relation to other WPs

WP6 is related to all WPs. Given the fact that it provides the user requirements from the pilots' side, WP6 supports the work of the technical tasks in WP2, WP3, and WP4. In addition, it takes feedback from them, to better specify the needs of the Use Cases. There is a close link to T1.4 (Gender and ethics) and T1.5 (risks and threats of AI), in combination with the Holistic Regulatory Framework (HRF), while it is also related to WP5, since the training courses that are being developed, support the capacity building of the people involved in the pilots and the assessment activities. Finally, WP6 feeds WP7 both in terms of communication and dissemination activities, but also with the results to structure a solid exploitation and sustainability plan.

1.6 Target audience of the deliverable

This document constitutes the first version of the feedback and evaluation of the AI4Gov pilot activities until M18. It is an internal guide for the project's pilot manager and all project partners to use it as a reference point for understanding the UC Scenarios and the results of the 1st validation phase after the testing of the AI4Gov tools. In addition, the document can be utilised as a practical tool for "Horizon Europe" pilot managers of on-going and future projects, who will be willing to explore the AI4Gov pilot strategy and capitalise on it, as well as a control point for the reviewers of the European Commission.

1.7 Evaluation and data protection

The piloting activities of the AI4Gov UCs involved data gathering processes especially during the evaluation step. To minimise any data related risks, the partners followed the Data Governance Framework (DGF) that was developed within the AI4Gov Project under T3.2. The DGF is a structured and comprehensive set of guidelines, policies, and procedures designed to manage, share, and protect data in alignment with the EU's legal and regulatory landscape, particularly concerning data protection and privacy. The framework ensures compliance with regulations such as the Data Governance Act, GDPR, AI Regulation, EU AI Act, and ALTAI for self-assessment. It emphasises compliance with data protection laws, clear data ownership definitions, data security through measures like encryption and access controls, and maintaining data quality through standards and validation processes. Privacy by design is integral, incorporating safeguards from the outset, and data sharing agreements are established to define the terms of data access and usage. The framework also involves structured data lifecycle management, ethical AI practices to prevent bias, accountability with designated Data Stewards and a Data Governance Committee, and continuous monitoring and compliance through regular audits and reporting mechanisms.

One of the pieces structuring the DGF is the Data Management plan (DMP) designed under WP1. The DMP outlines the overarching policy and strategy for data management within the AI4Gov project, addressing both administrative and technical aspects. It encompasses topics such as application reconfiguration logs, monitoring metrics collection, the publication and deposition of open data, details about the designated data repository infrastructure, and adherence to the Open Access Infrastructure for Research in Europe (OpenAIRE). In addition, it contains dedicated sections where it monitors the UCs to highlight the usability, purpose and collection procedures that should be implemented on these datasets.

With that being said, all data management processes in the UCs have been carried out based on these two tools: The DGF and the DMP. During the pilots, all personal data that were collected, including participant feedback, demographic information, and any other identifiable details, were stored securely and used solely to evaluate and improve the project activities. Access to the data was limited to authorised personnel within the project consortium, and no identifiable information was shared with third parties or used beyond the project's scope without explicit consent from participants. Anonymised data may be used in reports or publications to ensure that individual identities are protected. Participants retain the right to access, amend, or request the deletion of their personal data at any time by contacting the project's data protection officer. By participating, individuals confirmed their understanding and agreement to these terms.

2 The Evaluation Cycle

The validation and evaluation activities of the AI4Gov project correspond to phases 3 and 4 respectively in the pilot methodology presented in chapter 2. Phase 3 is the pilot implementation spanning from M6 to M33 and phase 4 is the evaluation and optimisation, which began in M6 and will last until the end of the project (M36). To make sure that the pilot timeline is aligned with the progression of the technical WPs, all partners were involved in the design of the time plan of the AI4Gov pilot activities. In this chapter, the overall timeline of the activities is presented, along with an extended presentation of the evaluation methodology.

2.1 Validation and Evaluation timeline

The AI4Gov pilot activities are structured around two evaluation cycles which determine their demonstration and validation activities timeline. A first testing to acquire some first feedback from the initial version of the AI4Gov tools' integration to the UCs, lasted from M19 to M24. After this first testing, the pilots organised feedback workshops to examine the usability of the tools. This feedback round focused on potential technical issues, gaps, or other weaknesses of the tools, to provide the technical partners with enough information to support the fine-tuning process. The fine-tuning will last for three months, from M25 to M27 and after this period, the updated version of the AI4Gov tools will be available. Then, the second implementation phase will start, where the pilots will validate the AI4Gov tools through their UCs and proceed with the final evaluation and assessment.

The validation phase will last six months, from M28 to M33 in combination with the second cycle of the evaluation activities and the final assessment. In the last three months, this information will be analysed according to the Last but not least, with regard to the workshops, please be cautious with definitive statements about the findings, as the number of respondents for each use case is insufficient to provide any statistical significance. Please make a note of this in the report. AI4Gov evaluation methodology, and the results will be presented in M36 in D6.5.

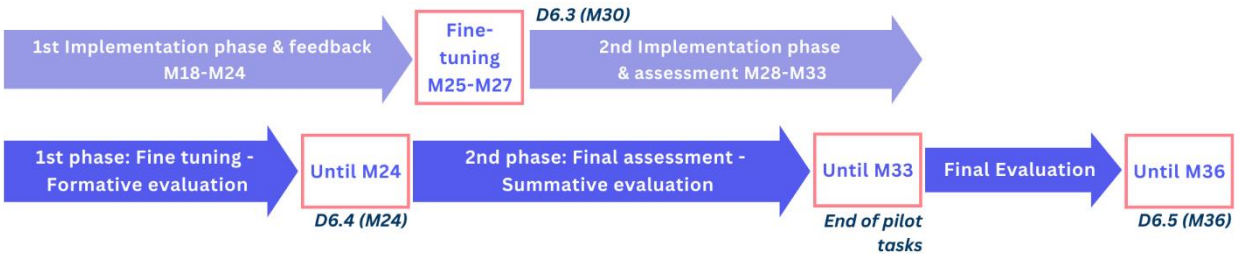


Figure 1 AI4Gov Piloting activities time plan

During this period, the pilot partners organised the first round of validation activities for their UCs, and the results of these activities are provided in Chapter 3. This initial validation phase involved a small-scale evaluation to identify and address critical issues, enabling the refinement and improvement, where feasible, of the AI4Gov tools. A second validation phase will follow, involving a broader audience to test the fine-tuned tools and gather more comprehensive feedback.

2.2 Evaluation methodology

This section presents the pathway of collecting and analysing pilot results and feedback from stakeholders to assess the effectiveness of AI4Gov solutions. The feedback analysis enhances and refines AI4Gov services, considering factors like efficiency, usability, and standards compliance.

A comprehensive approach was adopted, addressing political, technical, socioeconomic, legal, environmental, and organisational factors. From the outset, a structured evaluation framework was established, incorporating relevant metrics and tools, guided by pilot-specific KPIs. Stakeholder feedback is emphasised throughout the implementation and execution of use cases, with a series of workshops designed to actively involve participants. This continuous engagement ensures that stakeholder perspectives shape and enhance AI4Gov solutions, aligning them with practical, real-world needs. The aim is to rigorously assess AI4Gov technologies, leveraging stakeholder input for ongoing improvements that address the varied needs and expectations of all stakeholders.

The evaluation process of AI4Gov is part of the pilot methodology, as it has been described in D6.2. According to the pilot methodology, evaluation is a vital part of phases 3 and 4. The phases are not linear, and they overlap with each other so many of the activities that happen at the same time can belong to different project phases. In phase three (Pilot implementation), the evaluation process is at the point of collection. During the validation workshops, the feedback is collected directly from the participants who evaluate their experience with the AI4Gov tools. As it will be described in the next chapter, the validation workshops had a testing session and a feedback session. In the latter, the participants responded to the short version of the User Experience Questionnaire (UEQ) and a questionnaire on their trust in AI/New technologies.

The 4th phase (Evaluation & Optimisation), is focused on the evaluation process since it results in the final assessment of the AI4Gov tools. Also, since there are two validation cycles, after the 1st cycle, the feedback will help the technical partners with optimisation, to correct or enrich their tools. This refined version will be tested in the second validation phase (M28-M33) where the results of the assessment will be produced. Also, since the tools will be upgraded, so will the evaluation process, and the evaluation tools will also be refined and more detailed. After the 1st validation, some co-creation workshops will be held with the pilots so that they can share their feedback on the organisation and results of the testing and discuss ways to optimise the approach for the 2nd validation.

In this chapter, the methodology is presented, along with the tools and the KPIs of the activities.

2.2.1 Formative and Summative Evaluation

For the evaluation methodology of the AI4Gov activities, formative and summative evaluation methodologies have been chosen (Bhat & Bhat, 2019). They work complementary with each other, serving different stages of the project, and providing comprehensive insights into its development and effectiveness.

Formative Evaluation

Formative evaluation is an ongoing process used during the development and early implementation phases of a project. It is designed to provide immediate feedback that can be used to adjust, improve processes, and enhance outcomes. The primary purpose is to guide the iterative development of the project by assessing whether activities are implemented as planned and identify areas for improvement. Formative evaluation focuses on answering the following questions:

- *Are the objectives of the project being met as planned?*
- *What challenges or obstacles have emerged, and how can they be addressed?*
- *How do stakeholders perceive and engage with the intervention?* (Buelin, Ernst, Kelly, & DeLuca, 2019)

The formative evaluation employs a mixed-methods approach, combining qualitative and quantitative data collection techniques. These methods include:

- Stakeholder Interviews and Focus Groups: Engaging stakeholders in discussions to understand their experiences, needs, and suggestions for improvement. This will provide rich, qualitative data on project implementation.
- Surveys and Questionnaires: Distributing surveys to collect quantitative data on the effectiveness of the project's components and user satisfaction.
- Process Monitoring: Regularly assessing the alignment between planned activities and actual implementation to detect deviations and propose corrective actions.

The data collected during the formative evaluation are analysed continuously, allowing for iterative refinement of the project. Qualitative data are analysed using thematic analysis, while quantitative data are processed using descriptive statistics. The integration of these findings provides a comprehensive understanding of the project's progress and highlights necessary adjustments. The findings from the formative evaluation are used to inform decision-making, optimise processes, and maximise the potential impact of the project.

In the context of AI4Gov, the formative evaluation took place during the first validation phase of the project solutions. **The objective was to provide the first feedback to the technical partners to be able to identify weaknesses and strengths in the tools and proceed with the fine-tuning and refinements.** This formative evaluation happened through the use of the short version of the UEQ, and a questionnaire on participants' trust in AI/New technologies. It mostly involved people from within the pilot organisations and some external stakeholders, relevant to each UC. The objective of this phase is to adapt to new findings and ensure the project remains responsive to

emerging needs and challenges. After the end of the 1st validation phase, the feedback was gathered and analysed, to extract some lessons learned and optimise the process of the 2nd validation phase. The results of each UC and some combined conclusions are included in Chapter 3.

Summative Evaluation

Summative evaluation occurs after the piloting activities have been completed and is intended to assess the overall effectiveness and impact of the interventions. It focuses on outcomes and seeks to determine whether the project achieved its intended goals. Summative evaluation is critical for understanding the long-term value of the project and providing evidence of success.

The primary questions guiding the summative evaluation are:

- To what extent has the project achieved its intended outcomes and impacts?
- What is the overall effectiveness of the intervention?
- How do the results compare to baseline data and expectations set during the planning phase?

The summative evaluation also employs a mixed-methods approach, with a focus on robust data collection and analysis to ensure the reliability and validity of findings. This phase can include:

- Impact Assessments: Using predefined metrics and indicators to evaluate the project's outcomes on the target population and stakeholders. This will include both quantitative measures (e.g., statistical analysis of project impact) and qualitative insights.
- Comparative Analysis: Conducting a before-and-after comparison of key indicators to assess the project's overall contribution and effectiveness.
- Case Studies: Developing case studies that highlight successful outcomes, lessons learned, and best practices that can inform future initiatives.

The summative evaluation uses statistical analysis to quantify outcomes, complemented by qualitative research methods to add context and depth to the findings. Statistical significance tests will be employed to assess the validity of results, while qualitative data will be used to capture stakeholder experiences and the nuanced impact of the intervention. The results of the summative evaluation do not only showcase the project's impact but also serve as a valuable resource for scaling or replicating the intervention in other contexts. Additionally, the lessons learned can guide future projects and contribute to the body of knowledge in the field (Prince, 2015).

In AI4Gov, after the fine-tuning phase, the formative evaluation will be concluded. The summative evaluation will be employed during the last year of the project. Its primary goal will be to assess the overall effectiveness and impact of the AI4Gov solutions, determining whether the project achieved its intended outcomes and objectives. Summative evaluation will provide a comprehensive summary of the project's success and areas for future improvement. This second evaluation phase will emphasise the final outcomes of the project, providing a performance

analysis, including its strengths and weaknesses. Also, it will show the achieved results along with the insights and lessons learned that can inform future projects and broader practices.

As a result of the summative evaluation, the project will have a holistic picture of the advancements that happened after the fine-tuning phase and the improvement in knowledge, attitudes, and trust of the users. Additionally, the long-term effects of the project on the relevant stakeholders will be measured, using quantitative methods to analyse data and determine the significance of the results. A comparison between the formative and summative evaluation will make it possible to showcase the progress of AI4Gov tools and potential. Both in formative and summative evaluation, the project results will be evaluated through the UCs with the aim to achieve the dedicated pilot KPIs. The evaluation tools for the summative evaluation will be refined based on the pilots' feedback from the first validation phase to better cater to the needs of the participants.

Using a combination of formative and summative evaluations provides a balanced approach to assessing a research project. Formative evaluation helps ensure the project stays on track and adapts to challenges, while summative evaluation provides a comprehensive assessment of its overall success and impact (Dolin, Black, Harlen & Tiberghien, 2018).

2.2.2 The evaluation tools

In this 1st validation phase of the AI4Gov project, the evaluation of the UC activities took place using a variety of tools. These tools provided quantitative and qualitative data to provide a holistic view of the results of the tools' testing. At this stage, the methodology opted for simpler tools with fewer questions and attributes, since the technologies are not 100% ready. The tools will be updated for the final second iteration to be more detailed and provide a more thorough analysis of the usage and efficiency of the AI4Gov technologies. At this point, three tools were used: the short version of the User Experience Questionnaire (UEQ), a short trust questionnaire and a trust board template.

The UEQ is an established survey, focusing on obtaining feedback from software and tools. The full version includes 26 attributes, and the short version 8 attributes. The goal of the survey is to capture six factors of the user's experience: Attractiveness, Efficiency, Predictability, Stimulation, Transparency, and Originality. The survey was distributed to the participants after they tested the AI4Gov technologies along with some instructions on how to complete it. The survey along with the instructions can be found in appendix 6.1. Attractiveness is a pure valence dimension. Perspicuity, Efficiency and Dependability are pragmatic quality aspects (goal-directed), while Stimulation and Novelty are hedonic quality aspects (not goal-directed) (Schrepp, 2015).

In the short version that was used, the 8 attributes included the following: 1) Obstructive – Supportive, 2) Complicated - Easy, 3) Inefficient – Efficient, 4) Confusing – Clear, 5) Boring – Exciting, 6) Not interesting – Interesting, 7) Conventional – Inventive, 8) Usual - Leading edge. These 8 attributes correspond to 2 of the 6 factors of the full version, Pragmatic quality (attributes 1 – 4) and Hedonic Quality (attributes 5 – 8). The UEQ tool provides an excel where the researcher can analyse the data and descriptive diagrams are produced directly. The results are interpreted based on the means of the scales of pragmatic and hedonic qualities:

- Values between -0.8 and 0.8 represent a neural evaluation of the corresponding scale,
- Values > 0,8 represent a positive evaluation
- Values < -0,8 represent a negative evaluation.

The range of the scales is between -3 (horribly bad) and +3 (extremely good) (Schrepp, Hinderks, Thomaschewski, 2017).

One of the most critical aspects of the project is the trust people have in AI and new technologies. Since the AI4Gov technologies were not mature enough to assess their reliability, the methodology incorporated a side activity related to trust. The goal was to assess the trust people have in new technologies when integrated into the public operational systems they use. The tool they tested was one example of a new technology at an early stage of development. Considering their experience in the workshop along with their general knowledge and perception of AI and new technologies, they were asked to think how they would feel about using such a tool in their everyday life.

To address the aspect of trust, the pilots had two different options to choose from. The first, simpler one, was to distribute a second questionnaire after the UEQ, with 3 simple questions on their perception of the reliability and accuracy of the results of such technologies, the potential benefit to their work/everyday interaction with public services, and the security of the data. The responses reflected their personal opinion on trust, comfort, and familiarity with such technologies and it became clear to them that there were no wrong and right answers. This option was used by the UCs of VVV and JSI. The questionnaire along with the instructions can be found in appendix 6.2.

The second option for acquiring the participants' feedback on trust was the organisation of a small exercise using a template to work on the pros and cons of AI/New Technology from the perspective of technological advancements, bias, security, and trust. The idea was for the participants to split into groups of 4-5 people and work together on the template for a specific amount of time and then share their answers with the rest of the groups. The notion of the exercise was not to monitor their knowledge but understand their point of view and feeling around AI/New Technologies. This option was used by the UCs of DPB. The template can be found in appendix 6.3.

To make sure sufficient input from the pilots is acquired to proceed with a small-scale evaluation of this first validation phase, the target was to gather feedback from 100% of the participants, since the total number in the workshops was already low. However, there was a 60% minimum in case some of the participants refused to participate in the evaluation. The tables below present the number of participants per UC and the number of responses on the evaluation questionnaires.

Table 1 JSI: UC participants' overview

		Top 100 projects	SDG observatory	OECD document analysis
Participants		50	21	21
UEQ	Answers	11 (22%)	16 (76%)	13 (62%)
Trust questionnaire	Answers	11 (22%)	16 (76%)	13 (62%)

Table 2: VVV: UC participants' overview

		Traffic management	Waste management
Participants		10	10
UEQ	Answers	10 (100%)	10 (100%)
Trust questionnaire	Answers	10 (100%)	10 (100%)

Table 3: DPB: UC participants' overview

		Drinking water	Sewage water
Participants		10	10
UEQ	Answers	10 (100%)	10 (100%)
Trust template (pros/cons)	Answers	10 (100%)	10 (100%)

As presented in table 1, for the JSI use cases, the "Top 100 Projects" involved 49 participants with 22% completing the UEQ and trust questionnaires. The "SDG Observatory" and "OECD Document Analysis" use cases had 21 participants each, with response rates of 76% and 62%, respectively, for both the UEQ and trust questionnaires. Even though the target of the evaluation was not reached in the case of the Top100 projects, the nature of this UC allows for such a deviation. The audience was the Top100 reviewers, located in different places of the world and in different time zones. For this reason, they were reached only via email. Since they were not engaged though a workshop, the turnover of the responses was lower.

In the VVV use cases, the "Traffic Management" and "Waste Management" workshops each had 10 participants, achieving a 100% response rate for both the UEQ and trust questionnaires. Similarly, the DPB use cases, which focused on "Drinking Water Management" and "Sewage

Water Management," also had 10 participants in each workshop, with all of them completing the UEQ and the trust template (pros and cons).

All in all, the UCs managed to acquire valuable input by the participants during this 1st validation phase, and the results of this along with the challenges encountered during the process, are analysed in Chapter 3.

2.2.3 Metrics

To measure the success of the UCs, specific KPIs were put in place for each pilot, and were integrated into the evaluation methodology. However, it needs to be clarified that the KPIs correspond to the pilot as a whole and not to a specific UC, and that they should be achieved by the end of the project, without a specific goal for the first iteration. In this section, the KPIs are presented per pilot. Each KPI will be monitored through the evaluation process of the project during the 2nd validation phase of the UCs. DPB has 6 KPIs, VVV has 9 and JSI has 5. An overview of these KPIs is presented in table 4.

To measure the KPIs by the end of the project, several tools are being developed. These include workshops, focus groups, the User Experience Questionnaires (UEQ), trust questionnaires, and legal checklists, in addition to literature review and background theoretical work. The focus groups and workshops were selected as a direct way to approach the different stakeholders and acquire the necessary feedback from them. These activities are incorporated in different parts of the project, from T2.1 which organised focus groups targeting underrepresented groups of citizens to ensure inclusivity and support the creation of the HRF, to the validation and evaluation activities of the AI4Gov solutions in WP6. For the latter, the team will utilise the UEQ questionnaire, to cover the functionality and efficiency of the tools, while they are also developing a more detailed trust questionnaire to measure how these tools are perceived by the users on the basis of comfort and trust. In conclusion, the evaluation methodology will combine different tools and stakeholders to showcase the multi-dimensional impact of AI4Gov.

Table 4 Pilot KPIs

DPB	VVV/MT	JSI
Integrated and correlated data sources	Integrated and correlated data sources	Integrated and correlated data sources
Decrease in the citizens' taxes via sustainable water management	Visualisation dashboards	Visualisation dashboards
Increase citizens' engagement in policy development	Reduced time in resolving reported incidents	Increased communication and awareness among stakeholders

Increased trust in the policy development process	Reduction of the average cost per incident for the city	Increase geographical inclusivity
Increased number of algorithms / analytics used	Reduction of time to develop a policy	Increase gender representation
Increase efficiency: improve the success rate of new selected citizen groups	Provide real-time calculation capacity to 20% of the data	Balance performance vs. Explainability trade-offs
Detect "critical citizens" groups and increase their inclusiveness towards a fair supply of drinking water	Reduced transport operational costs for the city	
	Reduced transportation cost for the citizens	
	Increased citizens' satisfaction	

The way of monitoring each KPI is presented in the subsections below, along with the target numbers, derived from the GA of the project.

2.2.3.1 Policies for Sustainable Water Cycle Management at a Large Scale (DPB)

Integrated and correlated data sources > 3

In the context of the DPB UCs the project achieves the respective KPI by leveraging more than three data sources in the sense of historical datasets, diverse monitoring entities, and a comprehensive time-series repository. Water quality variables are combined from multiple sources, ensuring reliable and holistic data integration. This supports advanced time-series forecasting, enabling actionable insights and predictive analytics for the use cases.

Increase citizens' engagement in policy development > 20%

The AI4Gov tools target both the DPB staff in the water management department and the citizens. The citizens will have the opportunity to provide their feedback regarding the functionality of the tools and the efficiency of the service. In the second validation phase, the tools will be complete, and the representation of citizens will be the main focus. Specific questions will be asked to them to show how they perceive their role in policy developments and how much impact and influence they believe they can have through the AI4Gov tools. Indicative questions that can verify this are:

"Do you feel like you understand how the policy development process works?"

"Do you feel that the AI4Gov tools can help you be more involved in the policy development process and have an impact on the results?"

These questions are directly connected to the KPI's objective, to capture the perception of citizens towards their understanding of and impact on policy development. However, the final questionnaire is a work in progress and the questions are yet to be finalised, based on continuous literature review on AI-targeted evaluation trends.

Decrease in the citizens' taxes via sustainable water management > 20%

One of the objectives of the AI4Gov tools is to make the processes more efficient in terms of resources needed, such as time and costs. These resources are then translated into taxes for the citizens. By using the AI4Gov tools the project will make an estimation of the cost reduction and how this could affect the taxes referred to the water management. This will be a prediction based on the available data AI4Gov has.

DPB carries out the water management through a consortium. Thus, DPB manages the drinking water management in 56 municipalities and wastewater management in 53. A solidarity-based model 'Vivas donde vivas' (You live where you live) is being pursued, which proposes that all service users pay the same in all municipalities. This is done through taxes, which are composed of two elements, a fixed fee and a variable fee depending on consumption (drinking water) or volume supplied. The AI4Gov tools will allow to improve the efficiency of water treatment, distribution, and consumption. This will make it possible to anticipate events and prevent an inefficiency from occurring or being prolonged over time. Therefore, it would be reflected in citizen's costs.

Increased trust in the policy development process > 20%

In the context of the DPB UCs, one of the objectives is to also promote trust between the decision-making body (the diputación) and the citizens. In the second validation phase, where the tools will be complete, the representation of citizens will be higher, and specific questions will be asked to them to show how they perceive transparency and if they believe the AI4Gov tools offer this transparency and security that the policy development process reflects their best interests. Indicative questions that can verify this are:

"Do you trust that the policy makers in your region work in your favour?"

"Do you feel that the AI4Gov tools can create a more transparent environment and help you understand how the deputacion's services work?"

"Do you trust the information the AI4Gov tools provide are accurate and true?"

Increased number of algorithms / analytics used > 5

This KPI is addressed through the implementation of multiple advanced algorithms and techniques. Specifically, Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNNs) have been developed for time-series forecasting in both drinking water and sewage water UCs. These models incorporate explainability features via the "sufficient reasons" layer, which identifies key variables impacting predictions, enhancing transparency and interpretability. Blockchain integration ensures the integrity and traceability of both predictions and explainability reports, anchoring them securely to prevent tampering. Additionally, the integration of tasks 4.2 and 4.3 in the context of integrating XAI approaches and models (e.g., Self-Explaining Neural Networks (SENNs) with Minimal Sufficient Reasons (MSRs)) with the analytical models further enriches the analytics capabilities, enabling robust predictions and explainable insights. Collectively, these implementations contribute to the use of more than five algorithms/analytics within the project, fulfilling the KPI.

Detect "critical citizens" groups and increase their inclusiveness towards a fair supply of drinking water

The AI4Gov tools are designed to overcome bias and provide equal access to all citizens. In this vein, the DPB UCs target specific vulnerable populations in their water supply network, to make sure they have access to good quality water. The main function of DPB is to provide services which cannot be managed by Badajoz municipalities, especially the smaller ones and which are not provided directly by regional and national administrations. Theses municipalities, most of which have a rural population, are being affected by depopulation. This is meant, they cannot afford the costs of services or involved technologies. That is why DPB takes over the situation and manages it. The AI4Gov tools will be contributing to reduce the digital divide existing in our region, to improve the public services.

2.2.3.2 Tourism-driven multi-domain policy management and optimisation (VVV/MT)

Integrated and correlated data sources > 3

This specific KPI is addressed in both the waste management and traffic tickets use cases of the 3rd pilot by integrating multiple streams of data and correlating them to derive actionable insights. The Timeseries Analyzer processes data from smart garbage bins located across the Municipality of Vari-Voula-Vouliagmeni, enabling the training of an LSTM RNN for time-series forecasting. This model predicts bin fill levels and correlates them with citizen flow patterns, identifying areas with high visitor density based on the rate at which bins are filled. Moreover, with respect to the traffic tickets use case fines from the Greek police are integrated with traffic data from the municipality.

Visualisation dashboards > 2

As presented in D4.3 - "Policies Visualization Services V1" three different interfaces are available in the context of this pilot to further improve the insights derived from the analytical models applied in the two specific use cases of this pilot. More specifically, Citizens Feedback Interactive Interface, Traffic Violations Interface, Waste Management Interface are introduced that are incorporate several different visualizations fostering analytical insights understanding and explainability of the final results.

Reduced time in resolving reported incidents

The municipality of VVV implements “incident report” tools to acquire citizens’ feedback on everyday issues. After the citizens report the incident, the municipality addresses the matter in a timely manner. The AI4Gov tools aim to reduce this time to make the process more efficient. To measure that, the municipality will provide the average time needed to solve different matters and then make an estimation of the reduction of time the AI4Gov tools can offer. The goal is to identify the parts of the decision-making process that the tools can intervene in and optimise the process.

Reduction of the average cost per incident for the city

The municipality will provide information on the cost of resolving incidents of traffic violations (number of tickets issued divided by the costs incurred by the municipal police, namely staff costs, fuel, insurance, maintenance, etc) and probably of waste management and then make a prediction of the savings the AI4Gov tools can offer. Again, the goal is to identify the parts of the decision-making process that the tools can intervene in and optimise the process.

Reduction of time to develop a policy > 50%

The policy development process is also a subject of optimisation for the UCs of VVV. For this KPI, the municipality will provide information about the current process and then explain in which steps can the AI4Gov tools intervene and reduce the time needed to develop a policy. This will be an estimation and a prediction in the case of adopting the AI4Gov tools in the municipal operation systems.

The Municipality’s policy development process includes the following stages: (a) Policy issues for action (b) Develop a policy roadmap and choose the policy instruments (regulations and other policy tools), (c) Design new policy, (d) Implement and Enforce, (e) Monitor and Evaluate. The success of this process is strongly based on the interaction modalities with stakeholders based on Consultation, Communication, Co-operation and Co-ordination (4Cs).

The Municipality regulates matters within its competence by issuing Regulatory Decisions, within the framework of applicable legislation, by (a) setting rules, (b) determine the method of implementing the necessary measures, (c) determine the terms and conditions. The Regulatory Decisions are taken by the municipal council with the absolute majority of all their members, following the municipal committee. If the regulatory provision concerns exclusively the region of

a local district (Municipal community), the municipal committee formulates its recommendation, following the opinion of the relevant local council. In formulating the proposal, the municipal committee takes into account the observations and proposals of the competent social and professional bodies and groups of citizens in the region of the Municipality, with whom it is consulted, as well as any special studies that have been prepared to address the above issues.

The AI4Gov tools can reduce drastically the time needed for the initial stages to develop Municipality's touristic policy and the related policies regarding waste and traffic management, namely the (a) Preparation of studies (feasibility, technical-economical, etc.) regarding fees and pricing of waste collection services and traffic tickets, purchase of equipment and vehicles, spatial allocation of waste bins and parking spaces, etc. (b) Public consultation and engagement of the competent social and professional bodies and groups of citizens.

In addition, AI4Gov tools can reduce substantially the time needed for the other stages of policy cycle, namely: (a) Policy implementation, as they provide proactive evidence-based decisions to optimize resource allocation and usage, (b) Policy monitoring, as they allow real-time collection, analysis and visualization of the data, and (c) Policy evaluation, as they allow direct feedback regarding the level of endorsement and satisfaction of stakeholders while providing key metrics that assess its efficiency, effectiveness and impacts.

Reduced transport operational costs for the city

In the context of minimising resources needed, the AI4Gov tools will try to reduce the transport operational costs of VVV. This will be proven based on a comparison of the current costs and the prediction of the savings the AI4Gov tools can provide. These costs refer to fuel costs for the municipal police vehicles and the waste collection trucks.

The municipality has issued Regulatory Decisions, within the framework of applicable legislation, setting the rules: (a) for the regulation of traffic, the identification and operation of vehicle parking spaces, as well as for the installation and operation of meters and facilities for regulating vehicle parking in public areas (b) to maintain cleanliness in public and private outdoor areas, while managing the collection, storage and disposal of their waste. In addition, the municipality has drafted a sustainable urban mobility plan which is a ten-year horizon strategic plan for developing the sustainability of the urban space, with social, economic and environmental criteria, covering all modes and means of transport in the in the area, so that citizen mobility and urban transport become functional and sustainable.

Apart from the municipal vehicle fleet, the municipality operates a municipal transport network of buses, supplementary to the public transport, covering the needs of its inhabitants mainly in the remote areas. Finally, the municipality operates a bicycle, car sharing and electric chargers network promoting alternative modes of mobility and electromobility. All the above determine the transport operational costs for the city, which the AI4Gov tools will try to reduce. This will be proven based on a comparison of the current costs and the prediction of the savings the AI4Gov tools can provide.

Reduced transportation cost for the citizens

This KPI is more oriented towards the traffic violations UC. A part of this UC is related to the parking situation of the municipality and the time citizens loose to look for a parking spot. It can provide information on the most crowded areas and help drivers avoid them when searching for parking. In this vein, the pilot will make an effort to calculate the average fuel costs of a car while looking for a parking spot and then make a prediction of how much time and fuel can be saved as a result of the AI4Gov tools. Alternatively, the costs of other forms of transportation (bus, bike, walking) will be calculated.

Increased citizens' satisfaction

This KPI is mostly related to the overall satisfaction of the citizens when interacting with the AI4Gov tools. This will be measured during the 2nd validation activities, where more citizens will be involved, though a questionnaire where they will be able to evaluate the efficiency, user-friendliness and overall performance of the AI4Gov tools.

2.2.3.3 Sustainable Development and the European Green Deal (JSI)

Integrated and correlated data sources > 3

- Media news (SDG Observatory)
- OECD AI policy documents (SDG Observatory)
- OpenAlex (SDG Observatory)
- Patient outcome reports (SDG Observatory/Rare diseases)
- Videolectures.net - descriptions and titles of the lectures (SDG Observatory/Rare diseases)
- Top100 applications (Top100)
- Top100 reviews (Top100)
- Traffic accidents (Alcohol abuse use case)
- Traffic violations (Alcohol abuse use case)
- Breathalyser tests (Alcohol abuse use case)

Visualisation dashboards > 4

- Sentiment analysis of OECD AI policy documents (Bias board in SDG Observatory)
- Missing data about rare diseases
- Traffic accidents, alcohol abuse and breathalyser tests
- *Visualisation of OECD AI policy documents topics (in preparation)*
- *Visualisation of Top100 data (bias/ethics reviews, topics and collaboration – in preparation)*

Increased communication and awareness among stakeholders < 30%

The tools developed in the context of the JSI's pilot are oriented towards sustainability and awareness around fairness and non-bias technological solutions. In this context, the participants in the UC activities will evaluate the awareness raising aspect of the UCs and how the tools guided them towards the recognition of biases or unfairness they unintentionally ignored. In the 2nd validation phase specific questions will be distributed to the relevant stakeholders. Indicatively, some examples are:

"Were you aware of the bias/deficiencies existing in the available data for the xyz rare disease?"

"Do you believe the Ai4Gov tools helped you be more conscious towards hidden bias in the data you are working on?"

Increase geographical inclusivity > 20%

To make sure the UCs have a fair representation among different countries, the geographical coverage of the data sources will be used. The goal is for the AI4Gov tools to be trained, not only with data coming from the EU but also from non-western countries.

Increase gender representation > 20%

For this KPI there are two different approaches: The first one is similar to the previous UC. This means that except for the geographical inclusivity, the data should ensure fair gender representation in the samples. The second approach refers to the people attending the validation activities of the JSI UCs. The aim is to have a balanced sample of testers in terms of gender.

Balance performance vs. Explainability trade-offs

The project tackles the "Balance performance vs. explainability trade-offs" by integrating advanced ML and Deep Learning models, such as Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNNs), with specialised layers for explainability and LLMs. These models provide high-performance analytics while incorporating a "sufficient reasons" layer that identifies the most critical features affecting predictions. This explainability layer ensures that stakeholders not only receive accurate predictions but also understand the underlying factors driving those predictions. Furthermore, the anchoring of predictions and explainability reports to the blockchain ensures transparency and trust, preventing any tampering or loss of validity.

2.3 From Evaluation to Impact

AI4Gov is going beyond evaluating the tools and the UC activities, aiming to translate the results into impact. These results can create six impact dimensions. These six dimensions are political, socioeconomic, organisational, environmental, technological, and legal. For each one there was a research question set to explain how each dimension was approached by AI4Gov. The research questions are the following:

- Political: *How do AI4Gov tools optimise the policy development process?*
- Socioeconomic: *What are the direct and indirect benefits for the citizens?*
- Organisational: *What are the benefits for the organisation using the tools in their operational systems?*
- Technological: *What technological innovations do AI4Gov tools bring? What is the added value?*
- Environmental: *How do the tools achieve environmental sustainability?*
- Legal: *What legal and regulatory innovations does AI4Gov introduce?*

The evaluation methodology argues that the political impact is one of the most prominent, since all tools target government bodies, while crucial is also the direct impact the activities have on citizens' everyday life. In order to measure and explain how the dimensions are addressed, the pilot KPIs are used in combination with the results of the evaluation of the first validation phase, using the tools described in 2.2.2. The pilot KPIs can address more than one impact dimension, but they all have a prominent impact. Input also comes from work done in different tasks and WPs, such as T1.4 - *Gender and Ethical management*, T1.5 – *Unpacking a research methodology for identifying risks & threats*, T2.1 - *Qualitative Analysis on Fundamental Rights & Values*, and T5.4 - *Developments of (self)assessment tools on ethical and transparent AI*, ensuring a structured and integrated approach to evaluation. The methodology uses the tools and theoretical background that was produced under these tasks, but the linkage with other WPs' tasks is not limited to these three. Input also comes from the technical WPs 3 and 4, especially in the technical and environmental impact.

Political Dimension: *How do AI4Gov tools optimise the policy design and development process?*

The main impact of AI4Gov on the political dimension refers to the policy design and development process, as well as the overall decision-making processes. To address the political impact, AI4Gov focuses on the citizens' trust and engagement in policy development in an inclusive way. The goal is to raise awareness around policy making and optimise the processes for the benefit of both the citizens and the government organisations. The project takes also into consideration the implications of local, national, and European policy making procedures. In this vein, the related KPIs are aiming primarily (but not solely) towards citizens' engagement in policy development, trust in the policy development process, and the optimisation of the policy development process. This can also include societal attitudes towards the current decision-making procedures aiming for behavioural change and public engagement.

- Increase citizens' engagement in policy development (DPB) [secondary impact: social]
- Increased trust in the policy development process (DPB) [secondary impact: social]
- Reduction of time to develop a policy (VVV) [secondary impacts: organisational, social]
- Increased communication and awareness among stakeholders (JSI)

Socioeconomic dimension: *What are the direct and indirect benefits for the citizens?*

Even though the political impact holds a prominent role in the project, a vital part of the activities is to create tools that benefit the citizens either directly or indirectly. The direct impact on citizens poses a greater need for accessibility and explainability of the tools. This direct impact aims to bridge the gap between society and technology fairly and inclusively. The social dimension pertains to the cultural and demographic aspects that are integrated to the AI4Gov activities, while the economic dimension considers the broader economic benefits that can arise from the AI4Gov solutions in comparison to the solutions that are currently used. Factors addressing this socioeconomic aspect include the consideration of gender balance and geographical inclusivity, and the lower costs for certain services benefiting the citizens. This will be achieved through an inclusive approach and a focus on underrepresented groups. Social acceptance and engagement of all relevant stakeholders are crucial for the success of the AI4Gov pilot activities, as they can affect participation rates and the dissemination of results as well.

- Detect "critical citizens" groups and increase their inclusiveness towards a fair supply of drinking water (DPB) [secondary impact: political]
- Decrease in the citizens' taxes via sustainable water management (DPB) [secondary impact: political]
- Reduced transportation cost for the citizens (VVV) [secondary impacts: political, environmental]
- Increased citizens' satisfaction (VVV) [secondary impact: political]
- Increase geographical inclusivity (JSI) [secondary impacts: political, technological]
- Increase gender representation (JSI) [secondary impacts: political, technological]

Organisational Dimension: *What are the benefits for the organisation using the tools in their operational systems?*

The main focus of the organisational approach is the day-to-day work of the operational staff that use the AI4Gov tools to accelerate their efficiency and minimise operational flaws. This benefits both the staff, since their job becomes easier, and the citizens since they will experience a smoother, and faster interaction with the services. In addition, an optimised operational system, can positively affect the availability of resources. This can be measured mostly based on the reduction of time, human and financial resources needed for a specific service. In the context of the UCs, the KPIs related to such benefits refer to more efficient resource management in terms of financial, time, and personnel allocation.

- Reduced time in resolving reported incidents (VVV) [secondary impacts: political, socioeconomic, legal]
- Reduction of the average cost per incident for the city: (VVV) [secondary impact: political]
- Reduced transport operational costs for the city (VVV) [secondary impacts: political, socioeconomic, environmental]

Technological dimension: *What technological innovations do AI4Gov tools bring? What is the added value?*

In a project related to AI, it is impossible to leave behind the technological dimension, reflecting the technological advantages of the AI4Gov tools. These include research methodologies, data collection tools, and analytical techniques. In addition, the UCs will provide user-friendly and understandable visualisations of their results to ensure accessibility to all relevant stakeholders. To create tools that are state of the art and competitive, their added value will be showcased, along with some KPIs.

- Integrated and correlated data sources (all)
- Visualisation dashboards (VVV, JSI)
- Increased number of algorithms / analytics used (DPB)
- Provide real-time calculation capacity to 20% of the data (VVV)
- Balance performance vs. explainability trade-offs (JSI)

Environmental dimension: *How do the tools achieve environmental sustainability?*

The environmental dimension is addressed mostly through the technological attributes of the AI4Gov tools. Sustainability is a non-negotiable criterion, referring to issues such as energy and data efficiency, and computational complexity, among others. The implementation and integration of the AI4Gov platform offer a novel approach to the deployment, integration and utilisation of components and applications. More specifically, faster development cycles and collaboration between technical partners lead to a reduction in the time required for infrastructure provisioning and deployment, meaning also lower energy consumption. The results of this dimension will be fully available in the final deliverable D6.5, when the tools are complete.

In detail, the sustainable aspects relevant to the AI4Gov tools include Energy Efficiency (EE), Computational Complexity (CC), Model Efficiency (ME), Sustainability of Development (SD), Hardware Optimization (HO), Scalability (SC), Data Efficiency (DE), and Compliance and Standards (CS). For each of the AI4Gov technologies, different sustainability provisions are relevant, either as an already existing provision, or as a future provision once the tool is more advanced. The table below summarises the tools and the sustainability provisions.

Table 5: Overview of the AI4Gov tools and the relevant sustainability requirements

Tool/component	Relevant Requirements Addressed	Implementations So Far
Project's Infrastructure	EE, CS, SC, SD, HO	The Project Infrastructure is designed to optimize efficiency (EE), scalability (SC), and sustainability (SD) across all components. It incorporates hardware usage and energy consumption monitoring to achieve both hardware optimization (HO) and energy efficiency (EE), ensuring

		resource utilization is both effective and environmentally conscious. Through its containerized approach and the utilization of Kubernetes, the infrastructure supports streamlined workload deployment and dynamic scaling, enabling efficient management of resources. Its use of modular hardware enhances hardware optimization (HO) while promoting sustainable development (SD), and compliance with standards (CS) by also enabling easy upgrades and reuse, ensuring adaptability and long-term viability for evolving project needs.
Data Lake	DE, SC, HO, SD	The project Data Lake environment Data Lake environment enhances data efficiency (DE) by employing data compression techniques to minimize storage needs and energy consumption, ensuring faster access to critical information. With regards to scalability (SC), the containerization and auto-scaling mechanisms dynamically manage resources to accommodate fluctuating workloads, seamlessly integrating the platform with the Data Lake infrastructure. While its sustainable development (SD) is supported by reusable and interoperable metadata and data catalogs, facilitating efficient data organization and retrieval across systems. While, the hardware optimization (HO) focuses on tiered storage strategies within the Data Lake, leveraging high-speed devices for frequently accessed data while utilizing energy-efficient storage for archival purposes, ensuring a sustainable and high-performing ecosystem.
Blockchain	EE, DE, SC, DE, CS	The implementation of permissioned blockchain (Hyperledger Fabric) using PBFT (Proof of Byzantine fault-tolerant) as Consensus Mechanism reduces the carbon footprint by 99% compared to the Proof-of-Work (PoW) mechanism. Also because of permissioned blockchain no costly mining is required. The solutions enforce the energy efficiency of this solution (EE). Moreover, Hyperledger Fabric is an open-source DLT platform that allows to set up a configurable blockchain infrastructure which integrates Smart Contracts for sustainable development, compliance to standards, data efficiency and business logic (CS, SD, DE). Finally, Hyperledger Fabric blockchain can be scaled horizontally across peer nodes. Additionally, all required services are running in docker container implementation ensuring scalability (SC).

Policy Recommendation Toolkit	EE, DE, SD, SC, CS	This specific tool provides tools for energy consumption monitoring (EE) and integrates blockchain implementation for transparency and immutability of data (DE) , The definition of policies that optimize public processes of Waste Management, Traffic Violation etc. lead to the introduction of standard processes (CS) and sustainable development (SD), coupled with the container orchestration (e.g., Kubernetes) and streamlined workload deployment (SC).
Bias Detection Toolkit	EE, CC, SD, SC, HO	The Bias Detection Toolkit is designed to ensure efficiency, scalability, and sustainability in addressing bias in AI systems. It incorporates energy-efficient hardware and optimized AI models to achieve both energy efficiency (EE) and Hardware Optimization (HO), reducing computational demands while maintaining high performance in balance with low complexity (CC). Through containerization (SC), the toolkit efficiently manages resources, enabling scalable deployment and seamless integration into various workflows. The utilization of open source development and code/model Reuse (SD) promotes collaboration, transparency, and long-term sustainability, ensuring the toolkit remains adaptable and cost-effective for diverse applications.
AI Models & Policy-oriented Virtual Unbiased Framework	EE, CC, ME, SD, SC, DE, HO, CS	This framework leverages advanced technologies to ensure energy efficiency, scalability, and sustainability in AI-driven decision-making. It utilizes PyTorch (CPU-based) for energy efficiency (EE), enabling low-power inference that minimizes computational overhead. The efficiency of its models is ensured through fine-tuning pre-trained models from Huggingface with transfer learning enhances performance while reducing training costs, complexity (CC), and resource consumption (HO). The framework employs reusable and open-Source code (SD) to promote long-term development sustainability and collaboration, while its containerized architecture (SC) ensures seamless scalability and modular deployment. Finally, it integrates sustainable data storage solutions (DE) in integration with the Data Lake to optimize data handling, supporting efficient storage and retrieval with minimal environmental impact.
Situational-Awareness Explainability	ME, CC, SD, EE, DE, AUE	The SAX4BPM library has been released to the open source. This fosters a culture of open collaboration (SD); helps with reducing the need for physical meetings (EE); cutting down on transportational related carbon emissions

Library (SAX Library)		(CC); and reducing the amount of development infrastructure (HO). In addition, we promote the usage of Large Language Models (LLMs) that helps avoiding redundant training efforts, optimize fine tuning, facilitates light models' development that require less computational power to operate (ME). Moreover, the SAX4BPM library implements efficient algorithms for process mining and causal AI that are complexity bound (ME). The input required in the library consists of process executions of event logs, without any identification of any personal data adhering to GDPR (CS).
Visualization Workbench	EE, ME, SC, DE, SD, CS	The Visualization Workbench is designed with a focus on Sustainability of Development (SD), Scalability (SC), and Data Efficiency (DE) to ensure a robust, reusable, and efficient system for visualizing and managing data. More specifically, it leverages reusable and open-source code, ensuring long-term maintainability and ease of collaboration (CS). It also ensures model efficiency (ME) by streamlining visualization models, optimizing data pipelines, and leveraging adaptive, modular components that align with workload demands Built on a containerized architecture, the workbench achieves seamless scalability and deployment flexibility. Containers enable modularity, allowing developers to deploy, update, or expand individual components without affecting the entire system. The latter impacts on the implementation of a more energy efficient solution. Finally, it employs sustainable data storage solutions to optimize storage costs and minimize environmental impact.

Legal dimension: *What legal and regulatory innovations does AI4Gov introduce?*

The legal dimension encompasses the legal and regulatory frameworks that govern AI4Gov. The two main frameworks governing data, regulations, and ethics in AI4Gov are the Holistic Regulatory Framework (HRF), and the Data Governance Framework (DGF). In addition, issues such as contribution to standardisation, and Intellectual Property Rights (IPRs) are included in this dimension, while the project produces different self-assessment tools to monitor the ethical and legal aspects of the activities and ensure compliance.

The ethics and regulations lifecycle of the project follows a structured and iterative process designed to ensure compliance with ethical standards and regulatory requirements throughout the development and deployment of AI systems. The process begun with a comprehensive **literature review**, which examined and is still examining critical topics such as gender and ethics (T1.4), risks and threats posed by AI (T1.5), and fundamental rights and values alongside relevant

protocols (T2.1). This foundational stage is further enriched by expert input on AI and governance (T5.1), providing a robust knowledge base to guide subsequent phases.

Building on this groundwork, the project established essential **provisions** to align with high ethical standards and governance frameworks. These include the development of the HRF and DGF, the self-assessment tools for ethical and transparent AI (T5.4), and the ethical and organisational guidelines for ensuring trustworthy AI (T5.5). These provisions are designed to serve as both benchmarks and practical tools for implementing AI systems in alignment with fundamental ethical principles.

The project then moves into the **testing** phase, which involves a series of pilot activities designed to evaluate the AI4Gov tools developed in WP3 and WP4. These pilot activities are conducted under the framework of the established provisions to test compliance with ethical and regulatory standards. The insights gained during this phase are critical for identifying areas of improvement and ensuring that the tools meet the required benchmarks.

Finally, the lifecycle includes an **evaluation** phase, where the legal impact of the tools is assessed to verify compliance with ethical and regulatory requirements. This stage also involves fine-tuning the initial version of the tools, incorporating lessons learned from the pilot activities, and making necessary adjustments. The evaluation phase provides valuable input to enhance the provision tools, ensuring their ongoing relevance and effectiveness in fostering ethical and transparent AI practices. This iterative approach ensures that the project not only adheres to current standards but also contributes to advancing trustworthy AI governance.

By incorporating this methodology into the planning and execution of AI4Gov, the consortium provides a holistic view of the different dimensions that influence and are influenced by the project.

2.3.1 Towards sustainability: AI4Gov contribution to the SDGs

One of the main objectives of AI4Gov is to foster a sustainable and fair future, using AI applications. In this vein, special emphasis was given to identifying the contribution of the AI4Gov UCs to the 17 Sustainable Development Goals set by the United Nations. At this point, the project has set some steps to map and then report its contribution to the SDGs. The input to support this argument will come from the results of the UC activities, which will then be translated into the project's impact. Then, the impact will be associated with specific targets of the relevant SDGs, creating some policy suggestions. The final step is to gather the policy suggestions for the different SDGs and structure dedicated policy briefs, at the end of the project. This process, is presented in the figure below:

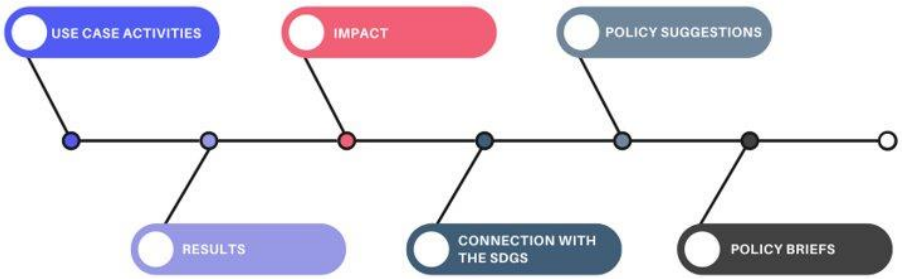


Figure 2: Contribution to SDGs: steps

All 17 SDGs are addressed through the AI4Gov UCs. Based on the mapping that was done, the IRCAI Top100 Projects, the SDG observatory, and the OECD policy document analysis address all SDGs since they focus on monitoring the SDGs and on enhancing funding for research and development purposes taking under consideration gender and geographical inclusivity. The fourth newest JSI UC on the topic of alcohol abuse addresses **goal 3 – Ensure healthy lives and promote well-being for all at all ages**, and more specifically the target *3.6 - By 2020, halve the number of global deaths and injuries from road traffic accidents.*

The two UCs focusing on water management of drinking and sewage water, led by DPB, address 3 different SDGs: **6 - Ensure availability and sustainable management of water and sanitation for all, 11 - Make cities and human settlements inclusive, safe, resilient and sustainable,** and **12 - Ensure sustainable consumption and production patterns.** The UC on drinking water contributes to target *6.1 - By 2030, achieve universal and equitable access to safe and affordable drinking water for all*, while the UC on sewage water contributes to target *6.3 - By 2030, improve water quality by reducing pollution, eliminating dumping and minimizing release of hazardous chemicals and materials, halving the proportion of untreated wastewater and substantially increasing recycling and safe reuse globally.* Both UCs contribute to the following targets:

6.4 - By 2030, substantially increase water-use efficiency across all sectors and ensure sustainable withdrawals and supply of freshwater to address water scarcity and substantially reduce the number of people suffering from water scarcity

6.5 - By 2030, implement integrated water resources management at all levels, including through transboundary cooperation as appropriate.

6.b - Support and strengthen the participation of local communities in improving water and sanitation management.

11.5 - By 2030, significantly reduce the number of deaths and the number of people affected and substantially decrease the direct economic losses relative to global gross domestic product caused by disasters, including water-related disasters, with a focus on protecting the poor and people in vulnerable situations.

12.4 - By 2020, achieve the environmentally sound management of chemicals and all wastes throughout their life cycle, in accordance with agreed international frameworks, and significantly

reduce their release to air, water and soil to minimise their adverse impacts on human health and the environment.

Lastly, the 2 UCs led by VVV address the same 3 SDGs, 6, 11 and 12. The UC on traffic violations contributes to target 3.6, same as the alcohol abuse UC, and target *11.2 - By 2030, provide access to safe, affordable, accessible and sustainable transport systems for all, improving road safety, notably by expanding public transport, with special attention to the needs of those in vulnerable situations, women, children, persons with disabilities and older persons.* The UC on waste management also contributes to target 6.b along with the water management UCs, and in addition it contributes to targets *11.6 - By 2030, reduce the adverse per capita environmental impact of cities, including by paying special attention to air quality and municipal and other waste management* and *12.5 - By 2030, substantially reduce waste generation through prevention, reduction, recycling and reuse.*

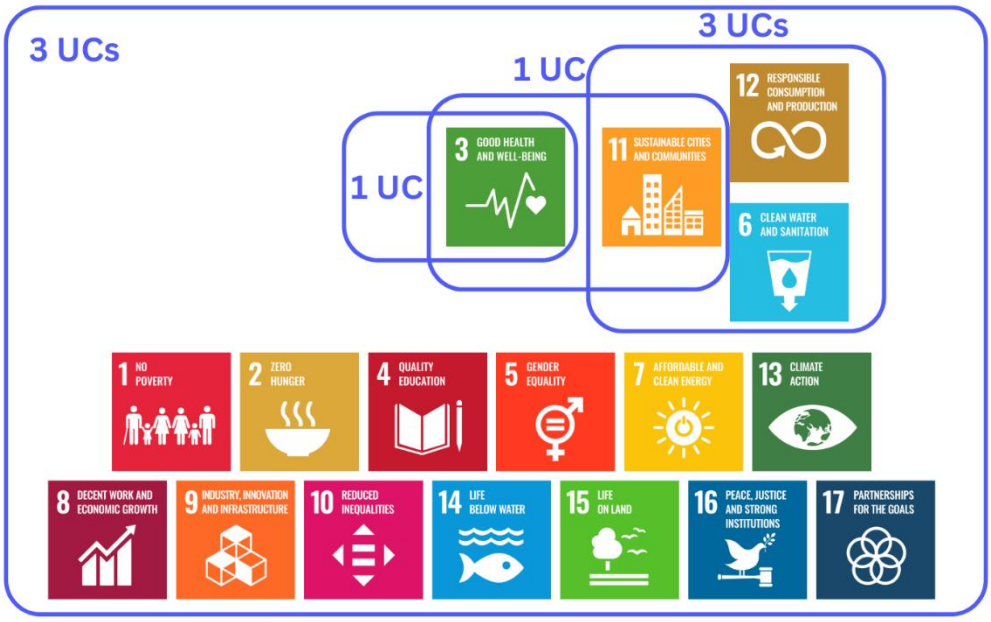


Figure 3: Overview of the SDGs addressed by the AI4Gov UCs

As the UCs progress and the AI4Gov tools are being finalised, the 2nd validation will lead to the final results of the UCs and these will be connected to the respective targets helping to produce policy recommendations directly connected and inspired by the SDGs.

In conclusion, the evaluation process stars form the formative and summative phases, where the results will be then translated into the six impacts. Lastly, these impacts will be paired with the SDGs they contribute to and produce useful suggestions and policy recommendations. This process is presented in the figure 4 below.

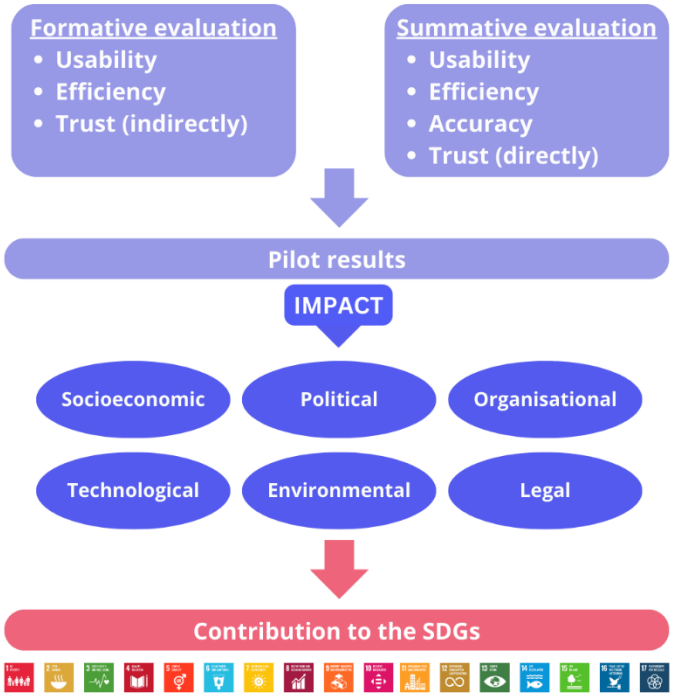


Figure 4: From evaluation to impact process

3 1st Validation phase: Evaluation results

This section provides a comprehensive overview of the organisation, execution, and outcomes of the **1st phase of the AI4Gov validation and evaluation workshops**. It details the planning and implementation of the workshops, including the initial small-scale testing of the AI4Gov tools, which served as a practical trial to gauge their functionality, relevance, and usability.

The findings from the feedback collected during these workshops are analysed to identify strengths, limitations, and areas for improvement in the tools and their deployment. This section also focuses on the **lessons learned** during this initial validation phase, offering valuable insights into how the evaluation process can be refined in anticipation of the 2nd iteration.

Finally, the section provides an **assessment of the process**, focusing on the efficacy of the methodologies employed, the robustness of the feedback mechanisms, and the overall impact of this phase in the development of AI4Gov solutions. This evaluation lays the groundwork for subsequent validation phases, ensuring the continuous improvement of the tools and their alignment with project objectives.

3.1 Small scale testing: AI4Gov Open Day

Even before the 1st validation phase, AI4Gov ran a quick evaluation of the tools during the Open Day event organised in Madrid on September 27th, 2024. The participants had the opportunity to interact with the AI4Gov tools and provide initial feedback on their experience through the completion of an EU Survey form, which can be found in Appendix 6.4. Twenty participants completed the EU survey, and their reflections are presented below.

The participants evaluated six different use cases: SDG Observatory (JSI), Top100 projects (JSI), OECD documents analysis (JSI), Traffic Violations management (VVV), Waste management – Check my Bin (VVV), and Drinking water management (DPB). The most frequently tested use case was "Traffic Violations Management (VVV)" with 5 responses. Most respondents were female (15 out of 20), with the dominant age group being 25-34 years (11 participants), followed by 35-44 years (5 participants). The participants represented six domains, with "Industry" being the most represented (6 responses).

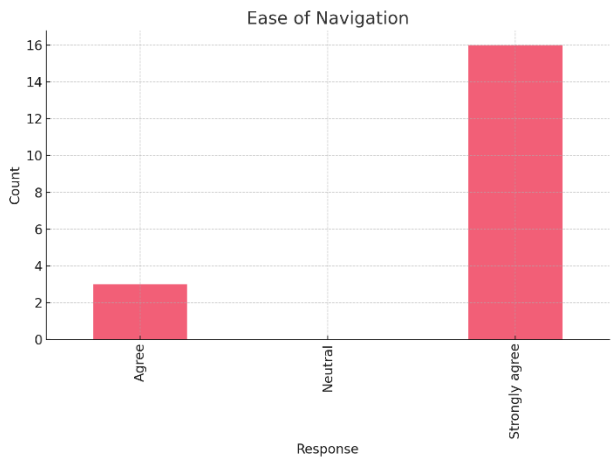


Figure 5: Ease of Navigation Insights

The majority of participants (16 out of 20) strongly agreed that the technology was easy to navigate. As concerns the usefulness of technology, 16 participants found the technology "**strongly useful**," while the remaining expressed agreement. 16 respondents **were strongly inclined** to further test the technology. In particular, most responders strongly agreed that the technology was easy to navigate and use, while a smaller proportion agreed, and no responders indicated a neutral or negative experience, as shown in Figure 5.

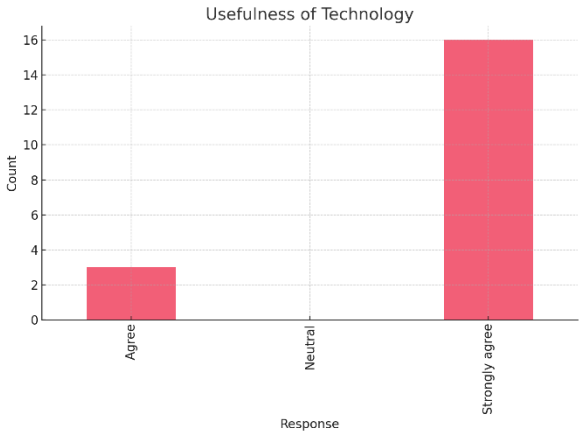


Figure 6: Usefulness of Technology Insights

As concerns the usefulness of Technology, the majority of respondents strongly agreed that the technology provided useful and relevant insights for addressing the problem at hand. A smaller group agreed, while no one indicated neutral or negative responses (Figure 6).

Lastly, most respondents strongly agreed that they would consider further testing the technology in their work or organization. A few agreed, while there is a minor representation of neutral responses.

Figure 7: Willingness for Future Testing Insights

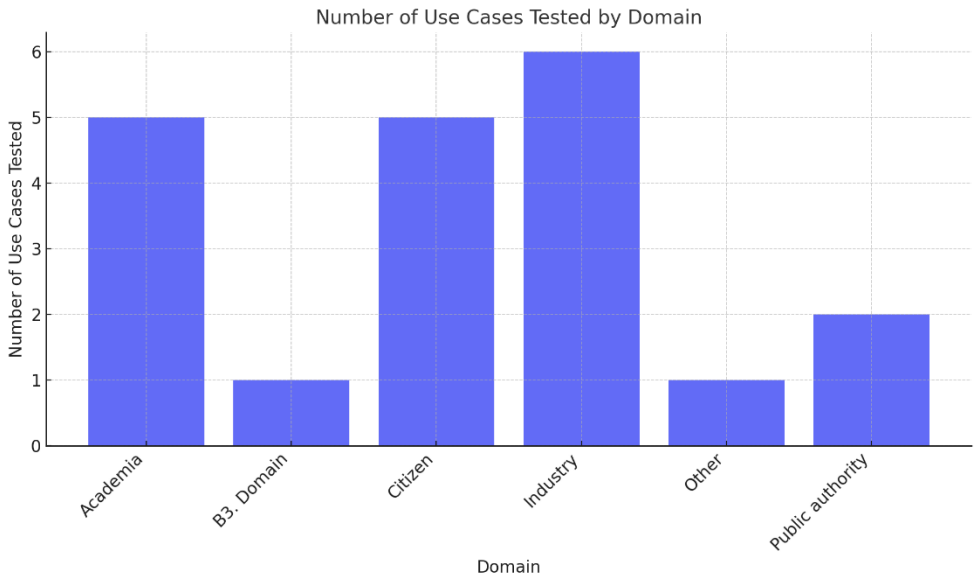


Figure 8: Overview of the Open Day feedback demographics

The domain-specific insights reveal that the "Industry" domain had the highest engagement, testing six use cases, followed closely by "Academia" and "Citizen," each testing five. Across all domains, participants overwhelmingly found the technology easy to navigate, useful, and worth further testing, with "Strongly agree" being the most common response. Comments provided were most frequent among participants from "Academia" and "Industry," indicating active engagement and constructive feedback. Demographically, the most common age group was 25-34 years, especially dominant in "Academia" and "Industry," while "Citizen" and "Other" were more represented by individuals aged 35-44. Female participants were predominant across all domains, reflecting a diverse yet gender-skewed testing cohort.

Furthermore, the participants provided qualitative feedback and suggestions for improvement:

- Comments included statements like "Useful information" and highlighted the importance of the application.
- Some provided constructive suggestions for improvement.

3.2 First round of pilot workshops

This subsection outlines the planning, execution, and outcomes of the pilot workshops, emphasising the critical insights gained and their role in the next phases of the project. The first validation phase workshops aimed to engage key stakeholders, including policymakers, researchers, and end-users, to evaluate the tools' functionality, usability, and relevance to real-world challenges. Through hands-on testing and structured feedback sessions, the workshops not only assessed the tools' technical performance but also explored participants' trust and confidence in AI-driven solutions.

3.2.1 Policies for sustainable water cycle management at a large scale

In this section the preparation, implementation of the evaluation workshops for the Spanish pilot UCs are presented: the Drinking and Sewage Water. Additionally, the results of the two workshops are presented along with the foreseen optimisation of each tool for the next evaluation period.

3.2.1.1 Drinking water

Main sector of interest: Water Management | Keywords: water management, drinking water, sustainability, Water Cycle, Real-Time Data, Efficiency

Summary: This UC is dedicated to the drinking water treatment system of the municipalities supplied by DPB. Drinking water management in the Province of Badajoz is centralised on the Drinking Water Treatment Plants and its attendant facilities, which for the past few years have been outfitted with digital monitorisation technologies. The UC is building a tool for technicians to predict possible sources of inefficiency within the system related to the quality of water and the energy consumption. In addition, the UC will create an executive reporting system for Water Management, to identify recurring problematic areas within the system, imbalances among points of service, over-time evolution of relative efficiency, etc. By analysing the data over longer periods of time, the pilot would help policymakers identify recurring problems and overall trends in Water Cycle management, to provide a tool for improving long-term investment strategies. A proper application of predictive analytics can be a useful tool for the local administration.

Target stakeholders/users:

- Technicians at the local Waste Management public consortium: these will be able to more easily access the information they use on the day-to-day management of the facilities.
- Policy-makers: Greater access to higher quality predictions and analytics will enable policy-makers to take better informed decisions on issues related to the use-case
- Consortium officials: In a similar way, high-level consortium officials can leverage these predictive capabilities for a better management of the project.
- High-level public administration workers: Proper visualisation tools will ensure that public administration workers can readily access the information they need for tasks such as reporting, resource-sharing, elaboration of open data packages, etc.
- Citizens: These will benefit both directly (greater transparency through said open data portals, eventual public dashboards, etc) and indirectly (through better management of the services they use).

AI4Gov tools to be employed

- Adaptive Analytics Framework
- XAI Library
- Visualisation Workbench

3.2.1.1.1 Workshop Organisation and Implementation

Preparatory Phase: The workshop planning process involved extensive coordination among all partners engaged in the Use Case (UC). Representatives from DPB (pilot partner), UPRC, IBM, and UBI (technical partners) actively participated in a preparatory meeting. During this meeting, it was decided that the evaluation of the Spanish UCs will be implemented in one mutual meeting including 20 participants for both UCs. The primary objective of the workshop was to demonstrate the Drinking Water Forecasting and Sewage Water Forecasting features of the Visualization Workbench for both drinking water and sewage water UCs, to gather feedback from technical and non-technical participants on usability, functionality, and trustworthiness and to validate the tools' predictive capabilities, including energy efficiency forecasting and the blockchain-based explainability mechanism.

Workshop Format and Agenda: The workshop took place on the 13th of December 2024 at the premises of DPB and involved 20 participants from DPB.

Table 6: Drinking Water Workshop Audience

Type of audience	**Number of participants**
Pilot partner employees **not** involved in the project – technical employees for Drinking Water	**10** (7 Male, 2 Female, 1 Prefer Not to Say)
Pilot partner employees **not** involved in the project – technical employees for Sewage Water	**10** (8 Male, 2 Female)
Total	**20 (**15 Male, 4 Female, 1 Prefer Not To say

The agenda was structured as follows:

- Introduction and Project Overview: A brief presentation provided an overview of the AI4Gov project, the drinking water UC, and the tools to be tested.
- Tool Demonstration and Hands-On Testing: Participants were guided through the features of the Visualisation Workbench and given time to interact with the tools, focusing on prediction capabilities and visualisations.
- Feedback Collection: Participants completed the User Experience Questionnaire (UEQ) to assess the tool's usability, reliability, and design quality.
- Focus Group Discussion: A moderated discussion explored participants' impressions of the tool, focusing on: Technological advancements, Trust and Bias and Security.

3.2.1.1.2 Evaluation Results

In this section the evaluation results from the UEQ and the exercise on trustworthiness are described, analysed and interpreted.

UEQ Results

Based on the analysis of the UEQ responses, the tools had an **overall score of 1.350,** which reflects an overall positive user experience, as shown in Table 7.

Table 7: Overview of UEQ Results for Drinking Water

Item	Mean	Variance	Std. Dev.	No.	Negative	Positive	Scale	
1	1.4	1.8	1.3	10	obstructive	supportive	Pragmatic Quality	
2	0.7	1.8	1.3	10	complicated	easy	Pragmatic Quality	
3	0.5	2.1	1.4	10	inefficient	efficient	Pragmatic Quality	
4	1.4	0.9	1.0	10	confusing	clear	Pragmatic Quality	
5	1.2	2.4	1.5	10	boring	exciting	Hedonic Quality	
6	1.9	2.1	1.4	10	not interesting	interesting	Hedonic Quality	
7	2.0	2.0	1.4	10	conventional	inventive	Hedonic Quality	
8	1.7	2.2	1.5	10	usual	leading edge	Hedonic Quality	

The provided results from the UEQ suggest an **overall positive user experience** (overall mean: 1.350), with both pragmatic quality (mean: 1.000) and hedonic quality (mean: 1.700) receiving favourable evaluations.

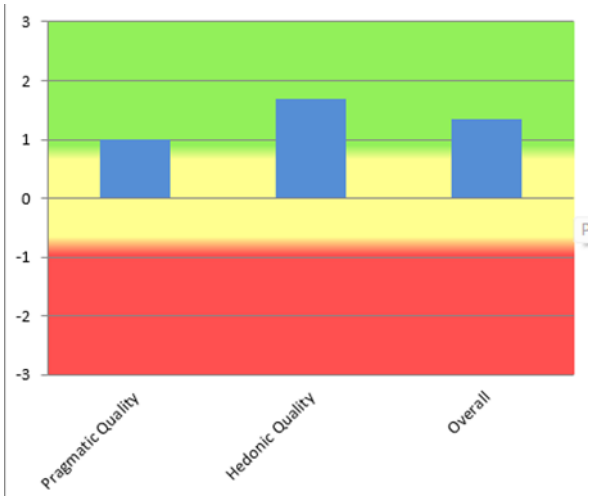


Figure 9: Overview of pragmatic, hedonic and overall quality scoring.

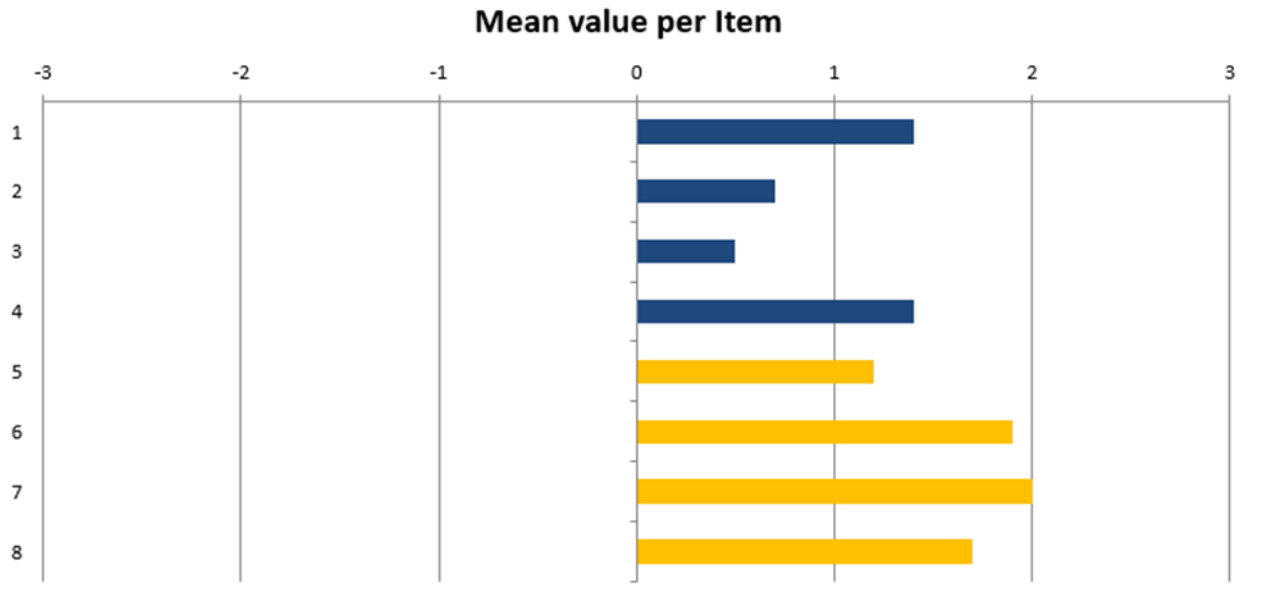


Figure 10: Mean value per Item for the Drinking Water evaluation

Pragmatic Quality Dimensions:

- For the Obstructive vs. Supportive item, a mean score of **1.4** falls within the **positive range**, indicating that users perceive the tool as **supportive** rather than obstructive. This suggests that the tool helps users achieve their goals effectively, with no major hindrances. The **standard deviation** of **1.3** reflects some variability in responses, indicating that while most users view the tool positively in this respect, there are some differing opinions.
- For the Complicated vs. Easy item, a mean score of **0.7** indicates a **neutral to positive** perception of the tool's ease of use. It is neither viewed as overwhelmingly easy nor particularly complicated. The tool likely provides a moderate level of ease, though some

users may find it more complicated than others. The **standard deviation** of **1.3** shows a moderate level of variability, meaning that user experiences with ease of use differ somewhat, with some finding it easier to navigate than others.

- For the Inefficient vs. Efficient item, a mean score of **0.5** places this item in the **neutral range**, suggesting that users perceive the tool as **neither highly efficient nor inefficient**. The tool may meet user expectations but does not stand out in terms of efficiency. The **standard deviation** of **1.4** shows significant variation in user feedback, indicating that while some users may find the tool efficient, others may feel it is lacking in efficiency.
- For the Confusing vs. Clear item, the mean score of **1.4** indicates a **positive perception** of clarity. Users generally find the tool **clear** rather than confusing, suggesting that it is easy to understand in most cases. The low score of standard deviation (1.0) indicates more consistency among user responses.

Hedonic Quality Dimensions

- For the Boring vs. Exciting item, a mean of **1.2** suggests that the tool is perceived as **exciting** rather than boring, though not overwhelmingly so. This indicates a generally positive response in terms of engagement and emotional appeal, but with a relatively high (1.5) standard deviation while some users find the tool exciting, others may feel it lacks novelty.
- For the Not Interesting vs. Interesting item, the mean score of 1.9 places this item into the positive range of evaluation, suggesting that users find the product interesting rather than dull. This reflects a strong emotional engagement, indicating that the product captures users' attention and curiosity. Although the high score, the standard deviation of 1.4 reflects some variability, meaning that some users may not feel as strongly about the tool's appeal.
- For the Conventional vs. Inventive item, the mean score is 2.0. The users see the tool as inventive rather than conventional, appreciating the product's originality and creative features.
- For the Usual vs. Leading Edge item, the mean score of **1.7** indicates that users perceive the tool as **leading edge** rather than usual. This suggests that the tool is seen as modern and innovative, which is a strong point for its hedonic appeal.

Trustworthiness Exercise

In this section, the results from the **Trustworthiness Workshop** will be analysed. Participants reflected on the pros and cons of the tested tool and AI tools in general across four key categories: **technological advancements, bias, security, and trust**. The **Trustworthiness Questionnaire** can be found in Appendix 6.3.

The results demonstrate that while AI tools are perceived **positively** for **productivity, automation, and process optimisation**, challenges remain regarding training needs and concerns about job displacement. Bias concerns arise primarily from **a lack of confidence in data**, incomplete information, and challenges in evaluating results. **Security** emerged as a major concern, especially regarding **cyber-attacks, data misuse, and vulnerabilities**. Meanwhile, **trust**

is conditional, depending on **data reliability, verification of outputs, and the gradual adoption and testing of tools**.

Pros

Technological Advancements

The majority of participants viewed AI tools as enhancing and optimising processes. The key benefits highlighted include:

- **Facilitating research** and **process automation**.
- Supporting **faster prediction and action**.
- **Improving decision-making** and supporting workflows.
- Enabling **knowledge dissemination** and process optimization.

Overall, AI tools are perceived as **time-saving** and enabling **faster workflows**.

Bias

Participants were more reserved regarding bias, with fewer references made. However, some positive remarks highlighted that:

- Bias could be mitigated through **clear decision-making processes**.
- AI tools can **deliver accurate results** when appropriately designed.

Security

Security was noted as highly dependent on **data quality**. Participants emphasised that tools can:

- **Reduce reaction times** to incidents.
- Facilitate **vulnerability detection** and ensure **data accuracy** if inputs are correct.

Trust

While participants were generally positive about trust, they emphasised its **conditional nature**. Key points included:

- Tools are reliable if **properly tested** and based on **correct data inputs**.
- Trust can be strengthened by ensuring **understanding of processes and data** used in tools.

Cons

Technological Advancements

Key challenges related to AI tools include:

- **Time investment** required for learning and operation.
- The need for **continuous training** and familiarity with AI tools.
- Concerns about **losing interest** if AI replaces manual tasks, potentially leading to **downsizing** or **job losses**.

Bias

Bias remains a significant concern, stemming from:

- **Lack of clear data** for cross-checking and evaluating results.
- **Low confidence** in the reliability of data used by AI tools.
- The risk of **incorrect or difficult-to-evaluate results**.

Security

Participants highlighted multiple security concerns, including:

- **Cyber-attacks** and communication vulnerabilities.
- **Misuse of data** due to lack of knowledge or errors in inputs.
- Concerns about **data safety** and the **reliability of results**.
- Challenges regarding **data privacy regulations**.

Trust

Trust-related concerns primarily stemmed from:

- **Initial lack of confidence** in AI results.
- The importance of **data quality** and **security** in building trust.
- Worries about **verification of outputs** and the potential for **attacks** compromising results.

The workshop results show that participants recognise the potential of AI tools to enhance productivity, optimize processes, and improve decision-making. However, significant challenges remain regarding **learning requirements, bias, security, and trust**. Addressing these concerns through **training, transparency, and robust data security measures** will be critical to building confidence in AI tools.

3.2.1.1.3 Foreseen optimisations for the 2nd validation phase

Towards the 2nd validation phase, the following optimisations are foreseen:

- The underlying technologies will be updated so that the users are able to provide their feedback with regards to the results provided by the tools.
- The design of the Visualization Workbench will be updated so that it is even more user friendly
- The implemented ML models will be finetuned

3.2.1.2 Sewage water

Main sector of interest: Water Management | Keywords: water management, sewage water, sustainability, Water Cycle, Real-Time Data, Efficiency

Summary: The Sewage Water Use Case (UC) focuses on improving sewage water management in municipalities within the DPB system. Given the region's intensive agricultural use, the UC aims to establish a real-time monitoring system to prevent sudden drops in water quality, which could lead to increased treatment demands due to agricultural water usage. The main objectives of the

Sewage Water UC are to enhance operational efficiency in water management. By analysing historical data from the treatment plants, the UC provides a tool for technicians to identify inefficiencies related to water quality and energy consumption. Additionally, the UC develops an executive reporting system to pinpoint recurring issues, imbalances across service points, and long-term trends in efficiency. Ultimately, the pilot supports policymakers in identifying recurring problems and trends, providing valuable insights to inform long-term investment strategies in water cycle management.

Target stakeholders/users:

Ref. 1 Technicians at the local Waste Management public consortium: these will be able to more easily access the information they use on the day-to-day management of the facilities.

Ref. 2 Policy-makers: Greater access to higher quality predictions and analytics will enable policy-makers to take better informed decisions on issues related to the use-case

Ref. 3 Consortium officials: In a similar way, high-level consortium officials can leverage these predictive capabilities for a better management of the project

Ref. 4 High-level public administration workers: Proper visualisation tools will ensure that public administration workers can readily access the information they need for tasks such as reporting, resource-sharing, elaboration of open data packages, etc.

Ref. 5 Citizens: These will benefit both directly (greater transparency through said open data portals, eventual public dashboards, etc) and indirectly (through better management of the services they use).

AI4Gov tools to be employed

Ref. 6 Adaptive Analytics Framework

Ref. 7 XAI Library

Ref. 8 Visualisation Workbench

3.2.1.2.1 Workshop Organisation and Implementation

One mutual workshop was conducted for both Drinking Water and Sewage Water. The workshop organisation is described in Section 3.2.1.1.1

3.2.1.2.2 Workshop Results

UEQ Results

The results of the UEQ are **generally positive**, with an overall score of 1.713, indicating that users have a favourable perception of the tool tested across both pragmatic and hedonic quality dimensions.

Table 8: Mean value per Item for the Sewage Water evaluation

Item	Mean	Variance	Std. Dev.	No.	Negative	Positive	Scale	
1	1.9	1.4	1.2	10	obstructive	supportive	Pragmatic Quality	
2	0.5	2.1	1.4	10	complicated	easy	Pragmatic Quality	
3	1.4	1.8	1.3	10	inefficient	efficient	Pragmatic Quality	
4	1.2	1.5	1.2	10	confusing	clear	Pragmatic Quality	
5	2.1	1.0	1.0	10	boring	exciting	Hedonic Quality	
6	2.3	0.9	0.9	10	not interesting	interesting	Hedonic Quality	
7	2.3	1.6	1.3	10	conventional	inventive	Hedonic Quality	
8	2.0	2.0	1.4	10	usual	leading edge	Hedonic Quality	

In terms of **pragmatic quality** dimensions, the tool scored a mean of **1.250**, which indicates that it is mostly viewed positively, but there is some neutral to mildly positive feedback about how easy it is to use. On the other hand, the score of **2.175 in hedonic quality dimensions** suggests that the tool is highly regarded for being exciting, interesting, inventive, and leading edge. It scores very positively on these dimensions, suggesting it appeals to users' emotions and engagement.

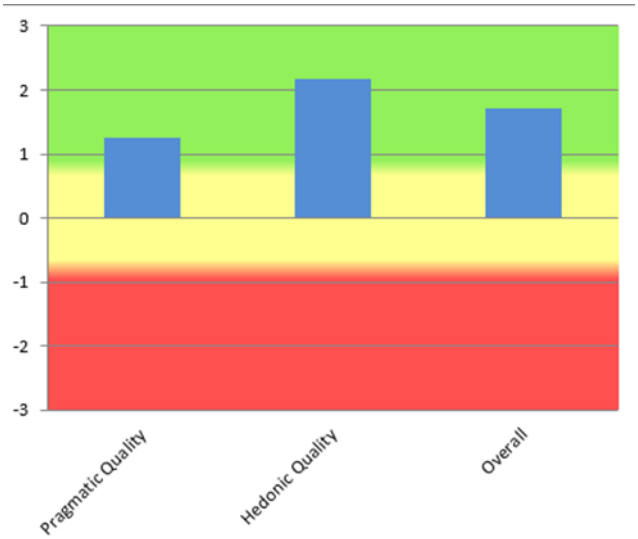


Figure 11: Overview of pragmatic, hedonic and overall quality scoring for Sewage Water

However, while the overall results are positive, there is some variation in user responses, suggesting areas where improvements could be made to further enhance the consistency of the user experience. The following analysis will delve deeper into the specific results, highlighting key strengths and potential areas for improvement.

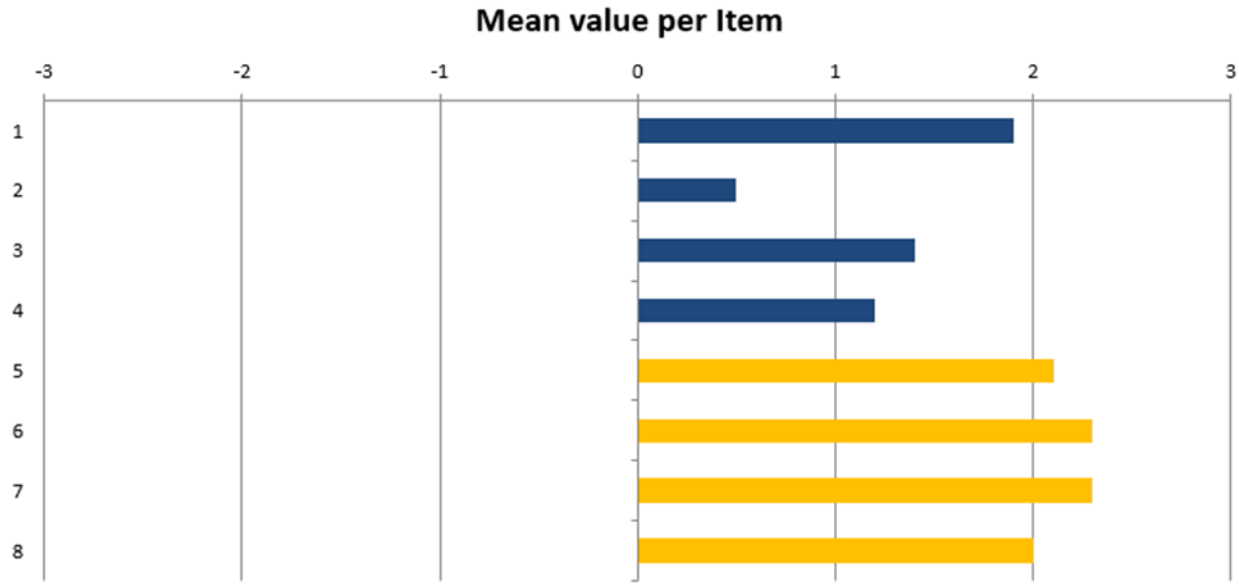


Figure 12: Mean value per Item for the Sewage Water evaluation

Pragmatic Qualities

- For the Obstructive vs Supportive item, the mean score of 1.9 falls into the **positive range** (greater than 0.8), which suggests that users find the tools **supportive** in terms of its pragmatic quality (ease of use, efficiency, etc.). This indicates that users likely feel the tool helps them accomplish their tasks effectively.
- For the Complicated vs. Easy item, the mean score of 0.5 falls within the neutral range, indicating that the tool is neither perceived as particularly complicated nor especially easy. Users seem to have mixed feelings about how easy it is to use, with a **moderate level of variance** in responses, as indicated by the **high standard deviation** (1.4).
- For the Inefficient vs. Efficient item, a mean of **1.4** indicates that users generally perceive the product as **efficient**, which is a **positive** evaluation in terms of pragmatic quality. The **standard deviation** of 1.3 shows some variability in opinions but leans positively toward efficiency.
- For the Confusing vs. Clear item, the mean score of 1.2 suggests a positive evaluation of the tool tested in terms of clarity (users find it relatively clear, but not perfectly clear. Again, the **standard deviation** (1.2) reflects a moderate degree of variation in users' views.

Hedonic Qualities

- For the Boring vs. Exciting item, the mean score of **2.1** falls in the **positive range**, indicating that the product is perceived as **exciting** rather than boring, while the **lower standard**

deviation (1.0) suggests that users have a relatively consistent positive view of the tool's ability to engage or excite them.

- For the Not Interesting vs. Interesting item, a mean of **2.3** suggests that users find the product **highly interesting,** with a low score of standard deviation of 0.9 indicating that this positive perception is relatively consistent across users. This is a **strongly positive evaluation** for hedonic quality.
- For the Conventional vs. Inventive item, the mean score of 2.3 suggests that the tool is seen as inventive rather than conventional. This is a **highly positive assessment** for the hedonic quality of creativity or originality. The **moderate standard deviation** (1.3) shows that while the response is generally positive, there is some variation.
- For the Usual vs. Leading Edge item, a mean of 2.0 indicates that the tool is considered leading edge, which reflects a positive evaluation in terms of hedonic quality, indicating it is perceived as modern or innovative. The **standard deviation** of 1.4 shows that responses are somewhat variable but still generally lean toward a positive view.
- Overall, the tool seems to have a **strongly positive** user experience, in terms of both **hedonic quality** and **pragmatic quality**. However, some users may feel neutral about how easy the tool is to use. The variability in user responses suggests that while the tool is generally well-received, there may still be areas that could be improved to make the experience even more consistent across different users. The areas of improvement are presented in the next steps section.

Trustworthiness Exercise

Participants provided feedback on the advantages and challenges of AI tools related to **Technological Advancements**, **Bias**, **Security**, and **Trust**. The data highlights both positive perceptions and concerns.

Pros

Technological Advancements

Participants emphasised the ability of AI tools to:

- **Facilitate tasks** and **save time** in decision-making.
- Enhance **optimisation, speed**, and **comfort** in processes.
- Improve **decision support** and provide **further information**.
- Enable **faster progress** and **ease of use** in operations.
- Deliver **major breakthroughs** and increased **efficiency** in workflows.
- Support **technical decision-making** and optimise **knowledge and management speed**.

Participants also noted that AI helps with:

- **Detection of vulnerabilities** in systems.
- Improving workflows that previously lacked automation.

Overall, AI tools were appreciated for their ability to **save time**, improve **efficiency**, and **optimize processes**.

Bias

Regarding bias, positive reflections were limited but included:

- AI tools enable **automatic detection** and help to **correct mistakes**.
- Participants felt they could **analyse specific data of interest**.
- Bias could be mitigated if tools allow for **clear decision-making**.

Security

Security benefits highlighted by participants included:

- AI tools can **ensure what data is used** and provide transparency.
- They are capable of **fewer errors** than human decisions.
- The tools support **vulnerability detection** in systems, enhancing preparedness.

Participants noted that **security would depend on data quality** and proper **legislative support**.

Trust

Trust in AI tools was seen as **conditional** and reliant on data quality. Key positive remarks included:

- There is **more trust** when results are checked and validated.
- AI tools are perceived as reliable if the **data used is correct**.
- AI tools help users gain **greater conviction** and teach users how to utilize them.
- Trust is present **in prediction** and **pattern recognition**.

Speed of implementation and management support were also mentioned as positive aspects of AI trust.

Cons

Technological Advancements

Participants raised concerns about:

- **Reduction in staff recruitment** and the **loss of jobs** as AI replaces manual tasks.
- The **risk of reliance** on AI solutions, which could undermine **problem-solving skills** and autonomy.
- Younger generations potentially losing knowledge of **programming logic** due to AI dominance.
- AI tools being perceived as **unethical** in certain contexts.

Bias

Bias remains a significant concern due to:

- The time-consuming nature of **defining clear goals** for AI use.
- The perception that **AI systems reflect only the owning company's interests**.
- General concerns about **lack of data** and how bias might affect results.

Participants emphasized the importance of having an **overview of the data** and clear **decision-making processes** to mitigate bias.

Security

Participants noted several security challenges, including:

- Risks of **losing control** and potential **manipulation** of outputs.
- Increased **vulnerability to attacks** and errors due to data inputs.
- Security issues being dependent on **legislation** and proper regulation.
- Concerns about **data privacy** and the inability to **control information** used by AI.

Participants also expressed concerns about **mistakes due to lack of data** and excessive reliance on AI systems.

Trust

Trust-related concerns included:

- **Mistrust** when starting data is **not clear** or reliable.
- Uncertainty regarding **data control** and **privacy**.
- Lack of knowledge about **how data is processed** or what happens to the data inputted into systems.
- Trust issues arising from **inconsistent results** and challenges in **controlling biases**.

Participants emphasized the importance of **verifying outputs** and ensuring that results are transparent and explainable.

In summary, the results reveal a mixed perception of AI tools during the Sewage Water workshop. Participants highlighted significant benefits such as **time savings**, **process optimisation**, and **vulnerability detection**. AI was seen as a useful tool for improving decision-making and enhancing efficiency. On the other hand, some participants raised valid concerns about **bias**, **data security**, and **trust**, particularly regarding **loss of control**, data privacy, and reliance on incomplete or unclear data.

3.2.1.2.3 Foreseen optimisations for the 2nd validation phase

Towards the 2nd validation phase, the following optimisations are foreseen:

- The underlying technologies will be updated so that the users are able to provide their feedback with regards to the results provided by the tools.
- The design of the Visualization Workbench will be updated so that it is even more user friendly
- The implemented ML models will be finetuned

3.2.2 Tourism-driven multi-domain policy management and optimization (VVV/MT)

For the Greek pilot, before analysing the results of the 1st validation phase, the complementary study led by MT is presented.

3.2.2.1 Primary qualitative and quantitative research on tourism flows, Municipality's services and the role of AI Survey on the quality of the touristic services of VVV & the role of AI

Within the framework of PILOT #3, the **Ministry of Tourism of Greece is conducting primary research** to gather qualitative and quantitative data on tourism flows (profile and preferences) to the Municipality of Vari Voula Vouliagmeni, as well as data on how the visitors, the permanent residents and the municipality's employees assess the municipality's services (in particular waste management and traffic management), and their attitude towards AI and how it can contribute to the improvement of public services through smart AI apps.

The aim of the research is to support the Municipality's task, by providing the data needed to plan and implement policies for the use of AI in public services, in order to respond to the citizen's needs in a more effective way. The key research questions are:

-*What are the key figures concerning inbound tourism and domestic visitor flows to the Municipality of Vari Voula Vouliagmeni?*

-*What are the key factors influencing tourists' satisfaction with tourism services provided at the Municipality?*

-*How do they assess the quality of public services provided by the Municipality?*

-*What is their opinion on Artificial Intelligence?*

-*To what extent could Artificial Intelligence contribute to the improvement of the municipality's services?*

Research is being conducted in 2 stages:

At the **1st stage**, the Ministry conducted **qualitative research via a series of interviews** with key stakeholders (hotel managers, travel agencies and tourist enterprises managers, hotel association representatives) at the municipality. The main aim of this stage was to gather information on the following

- The profile of the visitors
- Tourism flows in the municipality
- Evaluation of the services offered by the municipality focusing on the topics of the two UCs (waste and traffic management)

The interviews were conducted in the period between 27/05/2024-18/07/2024 online via zoom and in two cases by telephone. During the research process, the protection of personal data and the conditions of confidentiality were respected. Concerning the visitors' profile the main results were the following:

- The majority of the visitors come from the USA, Western Europe, Middle East countries and Cyprus or are Greek expatriates from the USA, Australia and South Africa.
- The main motivation for visiting during the summer months is leisure but during the winter months, business trips as well as trips for investing in the Athens real estate market are frequent.
- The average duration of stay ranges from 3 to 5 days.
- The most popular activities and points of interest within the municipality are visiting Lake Vouliagmeni, the temple of Poseidon in Sounio, the beaches and the coastal zone and outside the municipality the centre of Athens and the Athens Riviera.

Regarding the evaluation of the municipality's services, most of the key informants expressed satisfaction with the municipality's waste collection and management services. Nevertheless, it was pointed out that the use of Artificial Intelligence (AI) would be particularly useful to determine the necessity and timing of waste collection to avoid peak hours, so not to cause traffic problems. When asked if they knew of the VVV tools Pay As You Throw and the Novoville App, the majority stated that they were unaware of their existence and that it needs to be promoted for wider awareness. Lastly, the lack of parking spaces is one of the biggest challenges, especially in popular areas of the municipality such as Vouliagmeni & Kavouri during the hours of high visitor flows. A full report in English can be found in APPENDIX 6.5**.**

At the **2nd Stage quantitative research is being conducted on site** by experienced researchers who will assist participants in completing an online questionnaire. The target audiences are tourists, domestic day visitors from other parts of Athens and Attika, the municipality's residents as well as the municipality's employees. Visitors and residents will be reached at places of interest such as:

- Astir marina (Vouliagmeni)
- Vouliagmeni lake
- Hotels
- Public squares
- Other places of interest

The aim of the quantitative research is to investigate the attitude of tourists/visitors/residents as well as the municipality's employees towards the services of the municipality (cleaning, recycling, traffic, parking) and Artificial Intelligence (AI) and whether (or not) AI can contribute to the improvement of municipal services through the development of smart AI applications.

Two of the questionnaires (visitors and residents) are based on the key insights provided by the qualitative research conducted during the first stage and on the overview of the AI4Gov's Work Packages and deliverables such as, among others, the Holistic Regulatory Framework (WP2) developed under the project to lay the groundwork for addressing bias in AI by ensuring complying with EU regulations and facilitating the practical application of AI4Gov technologies. Input from UPRC (technical partner) was provided regarding AI questions and the Municipality's use cases (WP6). In addition, in the framework of the 1st validation workshop organised and the Municipality of VVV on the 22nd of November to evaluate the use cases tools' functionality, usability, and relevance to real-world challenges, the third questionnaire was developed for the

VVV's municipality's police and waste management personnel. Each questionnaire was tailored to the target population.

The questions included in the three questionnaires focused on the following subjects:

- The most popular places of interest at the municipality
- The effect of the arrival of tourists on services such as waste management, traffic congestion etc.
- The evaluation of the services offered by the municipality.
- The evaluation of the potential contribution of AI to the improvement of the municipality's services.

The research is being conducted both through google forms and printed questionnaires and the quantitative research is currently underway, (November 2024-March 2025). Questionnaires can be found in APPENDIX 6.6.

3.2.2.2 Traffic Management

Main sector of interest: Tourism | Keywords: tourism, traffic violations, visitors

Summary: Traffic violations, particularly related to parking, represent a significant challenge in Greece's major urban centres, including the pilot Municipality of Vari, Voula, Vouliagmeni (VVV). This issue intensifies during summer and weekends due to increased visitor and tourist flows. The Municipal administration of VVV aims to address these challenges using AI4Gov tools to improve decision-making, allocate staff and resources more efficiently, and enhance vehicle and pedestrian mobility while reducing traffic accidents and operational costs.

Currently, municipal and Hellenic police officers manually issue fines for violations such as speeding, traffic light infractions, parking violations, dangerous driving, and alcohol abuse, with data entry handled by municipal personnel. Through this use case (UC), the municipality seeks to automate these processes by leveraging AI and existing data.

Target stakeholders/users:

Municipal police staff and officers: The immediate group of stakeholders to use the AI4Gov tools in this UC are the officers of the municipal police that monitor the traffic violations in the municipality. As the responsible stakeholders for the operational part and the overall management of the day-to-day workload, the municipal police officers will be able to organise the staff's patrols in the most effective and timely manner.

Policymakers - Municipal Council: The policymakers in the municipality will use the AI4Gov tools in the context of this UC to optimise their strategic planning and decision-making process. Through this UC, the policymakers in charge of the Municipal police will be able to have a clear view of the traffic violations in order to allocate the necessary funds and resources to address the problem.

Citizens and visitors: The citizens and visitors are also a stakeholder group impacted by the optimisation of the municipal police operation leading to more effective monitoring of traffic

violations. The end goal is behavioural change by raising awareness among citizens and visitors to better respect the traffic code. In addition, the flows of visitors to the municipality, especially during summer, will be better managed causing less trouble to the permanent citizens, enterprises and hotels. As a result, the municipality will be able to take informed decisions and offer higher quality of services, in a well organised environment.

AI4Gov tools to be employed

- Policy-Oriented Analytics & AI Algorithms
- Adaptive Analytics Framework
- Visualisation Workbench

3.2.2.2.1 Workshop Organisation and Implementation

Preparatory Phase: To prepare for the evaluation workshop, tailored meetings were held involving pilot partners (VVV and MT) and technical partners (UPRC and UBI). During the meeting the pilot plan, evaluation tools to be used and the pilot KPIs were discussed. Furthermore, the tools to be used in the validation workshop were presented by the technical partner (UPRC). The workshop aimed to conduct a small-scale test with ten participants from VVV and MT. Pre- and post-activity monitoring templates and evaluation questionnaires were provided by VIL. All engaged partners in this Use Case agreed that the most suitable questionnaires for this workshop are the short version of the User Experience Questionnaire (UEQ) and a questionnaire that focus on the trust of the participants in the use of AI tools. The first phase emphasised usability and functionality, while the second phase will involve deeper analysis of the tools' potential.

Workshop Implementation: The first validation workshop for the Traffic Management UC was held on **November 22, 2024**, at the Municipality of Vari, Voula, Vouliagmeni (VVV). It involved partners from VVV, UPRC, and MT. In total 12 participants attended the workshop, seven (7) participants from municipal police directorate, 1 policy maker and 3 citizens and 1 hotel manager, taking into consideration the gender balance.

Table 9: Traffic Management Workshop audience

Type of audience	**Number of participants**
Pilot partner employees **not** involved in the project - MUNICIPAL POLICE DIRECTORATE	**7 (4 Male-3 Female)**
Policy makers - MUNICIPAL POLICE DEPUTY MAYOR	**1 (Male)**
Citizens	**3 (1 Male-2 Female)**
Hotel Managers	**1 (1 Female)**
Total	**12 (6 Male- 6 Female)**

The workshop aimed to evaluate the Traffic Violations Monitoring, Road Safety Assessment, and Traffic Density Analysis features of the Visualisation Workbench tool. It sought feedback on traffic violations within the Municipality of Vari-Voula-Vouliagmeni, as well as perceptions regarding road safety and traffic density. The goal was to understand public and municipal police perspectives on AI tools and assess their impact on addressing these specific issues. Input was gathered from the municipal police directorate, citizens, and hotel managers using two evaluation questionnaires: the UEQ and the AI & Trust Questionnaire.

The workshop had the following agenda: Short presentation of the project and the UC, the available MOOCs and a short presentation on the basics of AI. Then, training on the tools followed, during which they were presented, and their functionalities were explained in order for the participants to test them by themselves. Next on, the testing session followed, where the participants had the opportunity to use the tools and perform the available activities. After the testing, the two evaluation questionnaires were circulated to the participants in order to provide their feedback from the tools they had just tested.

3.2.2.2.2 Workshop Results

In this section the results from the UEQ and the Trust questionnaire will be presented in order to assess the feedback received after the testing of tools.

UEQ Results

Table 10: Overview of UEQ Results for Traffic Violation

Item	Mean	Variance	Std. Dev.	No.	Negative	Positive	Scale	
1	1.3	3.8	2.0	12	obstructive	supportive	Pragmatic Quality	
2	1.7	2.6	1.6	12	complicated	easy	Pragmatic Quality	
3	1.0	2.7	1.7	12	inefficient	efficient	Pragmatic Quality	
4	1.8	1.5	1.2	12	confusing	clear	Pragmatic Quality	
5	2.1	1.4	1.2	12	boring	exciting	Hedonic Quality	
6	2.2	0.9	0.9	12	not interesting	interesting	Hedonic Quality	
7	2.4	0.4	0.7	12	conventional	inventive	Hedonic Quality	
8	1.9	1.2	1.1	12	usual	leading edge	Hedonic Quality	

The overall UEQ score of 1.781 indicates a **positive evaluation** of the system. This suggests that users generally perceive the system as **supportive, clear, efficient, and exciting**. Each item is analysed below in the pragmatic and hedonic qualities.

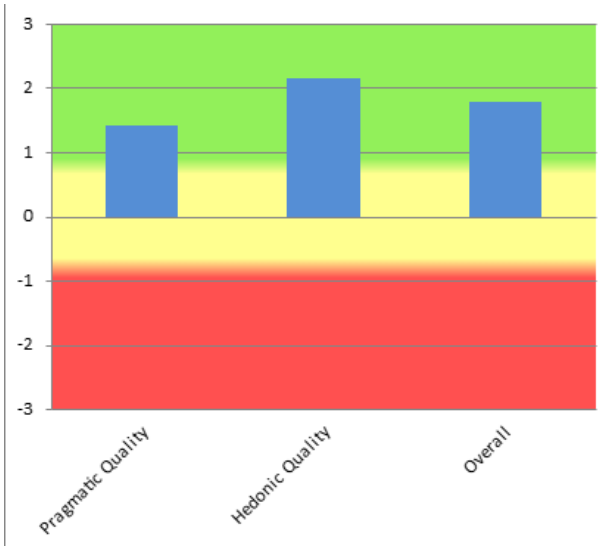


Figure 13: Overview of pragmatic, hedonic and overall quality scoring of Traffic Violation

Figure 14: Mean Value per Item for the Traffic Violation evaluation

The pragmatic quality items measure the usability and functionality of the tool. The pragmatic quality score of 1.417 indicates a **positive evaluation** of the system's **usability** and **efficiency**. Users perceive the system as **supportive, easy to use, and efficient**.

- For the Obstructive vs. Supportive item, the mean score is 1.3. The positive score indicates that users perceive the system as supportive rather than obstructive.

- For the Complicated vs. Easy item, the evaluation is positive with a mean score of 1.7, which is a positive evaluation. The positive score indicates that users find the system **easy to use**.
- For the Inefficient vs. Efficient item, the mean score is 1.0. The neutral score indicates that users perceive the system as **somewhat efficient**, but there is room for improvement in terms of speed and performance.
- For the Confusing vs. Clear item, the mean score is 1.8. The positive score indicates that users find the system **clear** and easy to understand.

The hedonic quality items assess the tool's appeal, attractiveness, and innovation. The hedonic quality score of 2.146 indicates a **strong positive evaluation** of the system's **aesthetic and motivational qualities**. Users perceive the system as **interesting, exciting, innovative, and cutting-edge**.

- For the Boring-Exciting item, the mean score is 2.1, reflecting a positive evaluation, which indicates that users find the system **exciting** and engaging.
- For the Not Interesting-Interesting item, the mean score is 2.2, highlighting a positive evaluation of the tool. The positive score indicates that users find the system **interesting**.
- For the Conventional-Inventive item, the evaluation is strongly positive with a score of 2.4. The positive score indicates that users perceive the system as **innovative** and cutting-edge.
- For the Usual-Leading Edge item, the score of 1.9 indicates that users perceive the system as **leading-edge** and ahead of its time.

Trust Questionnaire Results

The trust questionnaire included three questions focusing on the users' perceptions of the reliability, accuracy and security of AI technologies. Their responses were based not only on their experience with the AI4Gov tools, but also on their general knowledge and interaction with such technologies.

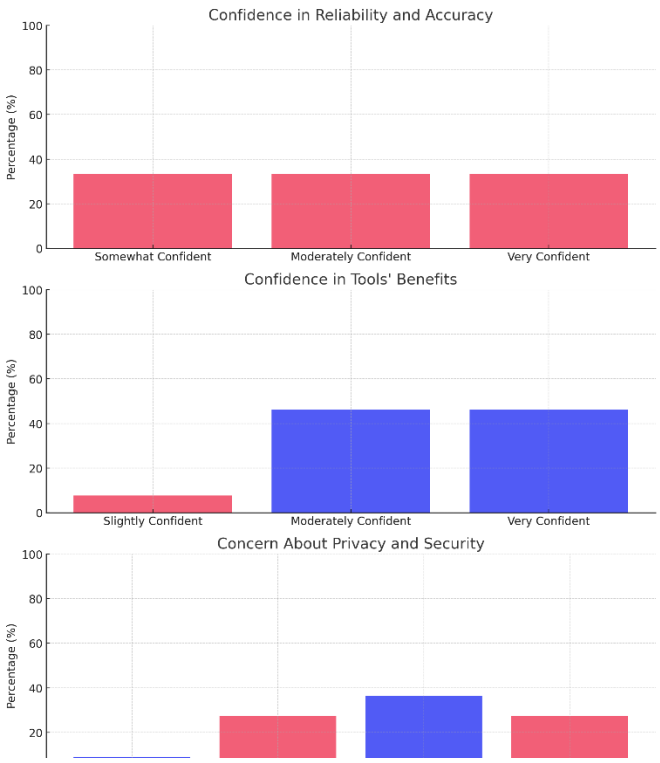


Figure 15: Responses on the Confidence in reliability and accuracy, tool's benefits and privacy and security of the Traffic Violation evaluation

The responses to the question, **"How confident are you in the reliability and accuracy of such tools' outputs in order to make decisions based on them?"**, revealed the participants' perceptions on the tool's reliability and accuracy. The respondents have a mixed level of confidence in the tool's reliability and accuracy, as the responses were evenly split among the three confidence levels: Somewhat confident (33.3%), Moderately confident (33.3%) and Very Confident (33.3%).

While there is a baseline level of trust in the tool's reliability and accuracy, the confidence levels are not overwhelmingly high. Regarding the perceptions of the participants on the **benefits of the tested tools**, the feedback was positive on the tools' utility in making work easier, indicating a high level of confidence. The majority of respondents are confident that the tools can make their jobs easier, with **92.4%** of responses falling into either **moderately confident (46.2%)** or **Very confident (46.2%)** categories. Only **7.7%** of respondents indicated they were "Slightly confident," which suggests that very few individuals have low confidence in the tool's ability to provide tangible benefits to their work.

Although the feedback on the benefits of the tools is positive, a general concern was expressed about privacy and security of the data processed by such tools. In particular, a majority of respondents (**63.7%**, combining **Somewhat concerned (36.4%)** and **Very concerned (27.3%)**) express some level of concern about the privacy and security of the data processed by the tools. This highlights a significant issue for stakeholders to address. Approximately **27.3%** of respondents selected **Neutral**, indicating that while they may not feel strongly about the issue, they are not completely confident in the security measures either, while only **9.1%** of respondents selected **Not concerned**, showing that very few participants feel entirely comfortable with the data security aspects of these tools.

3.2.2.2.3 Foreseen optimisations for the 2nd validation phase

In this section the comments and feedback received during the workshop are presented along with the optimisations foreseen for the next validation phase.

The validation workshop yielded a mix of positive feedback and constructive suggestions regarding the AI4Gov tools for traffic violation monitoring. All participants (100% response rate)

showed high levels of engagement, actively posing questions, providing suggestions, and sharing their perspectives. Many found the tools supportive, easy to use, clear, exciting, interesting, and innovative. They generally appreciated the potential of these tools to enhance decision-making and streamline workflows, though some expressed reservations about their overall efficiency.

Participants were moderately confident in the reliability and accuracy of the tools' outputs. While they recognised the potential of the tools to simplify tasks and improve decision-making, concerns were raised about the accuracy of predictions due to the representativeness of the data sources used.

Privacy and security of the processed data emerged as significant concerns among participants. They were cautious about relying on AI systems for sensitive tasks without clear assurances about data protection. This concern, coupled with the need for better data representativeness, underscored participants' reservations about the current readiness of the tools for widespread use.

Another recurring theme was the perception that the tools, while innovative, are not yet mature enough for final validation or deployment. Some participants emphasised the need to allocate more time to refine the tools, optimising their functionality for real-world application. They suggested that the tools' utility could be improved with faster data processing and enhanced capabilities for real-time analysis.

There were also mixed opinions on prioritisation, with a few participants expressing the view that predictions are based on experience and that the municipality should focus on projects with immediate impact rather than AI initiatives at this stage. Despite this, the tools were largely regarded as inventive and promising, with potential for significant improvement.

Participants provided valuable suggestions for enhancing the tools. These included:

- Adding options to filter data by specific timeframes (such as weekends or entire months) and geographic areas
- Integrating real-time updates on parking availability, especially for electric vehicles and bicycles.
- Addressing biases in the data and expanding the dataset to improve the accuracy of predictions and recommendations.

The workshop results underscored the need for further development to address the identified concerns and align the tools more closely with user needs and expectations. These insights will be instrumental in refining the tools during the next phase of development and validation.

Some users question the relevance and applicability of the AI tools to their specific context, suggesting that current efforts could be better allocated to projects with immediate and practical utility.

A significant concern lies in the representativeness and quality of the data being used. The accuracy and validity of AI-driven predictions are seen as limited due to poor or incomplete data. While users recognize the potential benefits of AI tools, they believe these tools are in an early developmental stage and require refinement and optimization before being considered reliable.

Timeliness and efficiency of data input are seen as critical for accurate analysis. Delays in data updates could reduce the tools' effectiveness.

In addition, some users emphasised the need for additional development and testing to optimise AI tools, ensuring they meet user expectations and deliver meaningful outcomes, while concerns about bias were mentioned, with users requesting more representative data and efforts to mitigate potential biases in AI predictions.

Other users highlight the importance of incorporating real-time data into the system to improve accuracy and relevance in predictions and decision-making.

Towards the 2nd validation phase, the following optimisations are foreseen:

The underlying technologies will be updated so that the users are able to provide their feedback with regards to the results provided by the tools.

Specific pages will become responsive, allowing the users to view and interact with them from mobile devices as well.

Additional data will be analysed to gain further insights with regards to parking availability.

3.2.2.3 Waste management

Main sector of interest: Tourism | Keywords: tourism, waste management, visitors

Summary: Waste management in Greece faces significant challenges, particularly with rising landfill costs. In the context of AI4Gov, the second UC of the Greek pilot will automate the operational system of the waste management making it more efficient, in terms of cost and time, employing sensors and RFID tags on the bins as well as telematics, to design the most efficient waste collection plan. The goal is to allocate staff and resources more efficiently keeping the municipality clean without wasting resources.

Target stakeholders/users:

As in the first UC, the immediate operators of the municipal staff, this time in the waste management directorate, and the policy makers in the municipal council. Lastly, the UC will positively affect citizens and visitors.

- Municipal staff and officers: The immediate group of stakeholders to use the AI4Gov tools in this UC are the municipal staff and officers in the waste management directorate. As the responsible stakeholders for the operational part and the overall management of the day-to-day workload, the waste management staff and officers will be able to better organise the garbage collection system in the most effective and timely manner. The UC will provide them with a tool to monitor overall data from telematics sensors and RFID, recommend optimum areas and resource allocation, and predict financial outcomes.
- Policymakers - Municipal Council: The policymakers in the municipality will use the AI4Gov tools in the context of this UC to optimise their strategic planning and decision-making process. Through this UC, the policymakers in the Municipal Council will be able to allocate the necessary funds and resources to efficiently monitor the overall waste management

system so that they reduce the municipality's costs and assist the transition towards a more circular and sustainable economy according to waste management legislation. The UC is also working on a tool to suggest the optimum routing for collecting garbage, to facilitate the prediction of the financial outcomes regarding the expansion of the Pay As You Throw System, so that the necessary funds and resources are effectively allocated.

- Citizens and visitors: Citizens and visitors are another stakeholder group impacted by the UC. The optimisation of the waste management will lead to a cleaner municipality. The end goal is the behavioural change of the citizens and visitors to litter less and at the same time to facilitate waste management for the municipality. As a result, the municipality will be cleaner and the negative impact of the visitors will be lower, while they will be more satisfied from their touristic experience as well. The municipality will be able to offer higher quality of services, in a well organised environment.
- Hotel Managers: The last stakeholder group is the hotel managers. This group was approached by the MT in the context of their survey, as described in the beginning of this pilot. They were identified as a valuable addition to the UC stakeholders, since they are professionals that interact with visitors and tourists and can provide input on the quality of the municipal services and their impact on the touristic sector.

AI4Gov tools to be employed

- Policy-Oriented Analytics & AI Algorithms
- Adaptive Analytics Framework
- Visualisation Workbench

3.2.2.3.1 Workshop Organisation and Implementation

Preparatory Phase: To prepare for the evaluation workshop, tailored meetings were held with pilot partners (VVV and MT) and technical partners (UPRC and UBI). During these meetings, the pilot plan, evaluation tools, and pilot KPIs were discussed in detail. The workshop's purpose was to conduct a small-scale test involving ten participants from VVV and MT. VIL provided pre- and post-activity monitoring templates and evaluation questionnaires. All participating partners agreed that the most suitable evaluation tools for the workshop were the short version of the User Experience Questionnaire (UEQ) and a questionnaire focused on participants' trust in AI tools. While the first phase emphasised usability and functionality, the second phase will delve deeper into usability and the potential of the tools.

Workshop Implementation: The first validation workshop for the Waste Management UC was held on **November 22, 2024**, at the Municipality of Vari, Voula, Vouliagmeni (VVV). It involved partners from VVV, UPRC, and MT, with nine participants in total, ensuring gender balance. The audience was consisted of 3 citizens (1 male and 2 females), 1 (female) hotel manager and 5 members of the Waste Management Directorate (4 males and 1 female).

Table 11: Waste Management Workshop audience

Type of audience	Number of participants
Citizens	**3 (1 Male-2 Female)**
Hotel Managers	**1 (1 Female)**
Pilot partner employees **not** involved in the project - WASTE MANAGEMENT DIRECTORATE	**5 (4 Male- 1 Female)**
Total	**9 (6 Male-3 Female)**

The objective of the workshop was to test the **Check my bin** and **Optimisation of garbage collection** tools and get feedback from the Waste management directorate and the visitors/Hotel managers through the two evaluation questionnaires provided to them: UEQ and AI & Trust questionnaire.

The workshop had the following agenda: Short presentation of the project and the UC, the available MOOCs and a short presentation with the basics of AI. Then, training on the tools followed, during which they were presented, and their functionalities were explained in order for the participants to test them by themselves. Next on, the testing session followed, where the participants had the opportunity to use the tools and perform the activities that were ready up to that point. After the testing, the two-evaluation questionnaires were circulated to the participants in order for them to provide their feedback from the tools they had just tested.

	Duration	Description
Welcome, Introduction to AI4Gov and UC presentation, MOOCs, AI basics	20'	Short presentation of the project and the UC, available MOOCs, short presentation with the basics of AI
Training on the tools	15'	Presentation of the tools and explanation of how they work so the participants can then test it themselves
Testing session	45'	Participants will have the opportunity to use the tools and perform the activities that are ready up to this point.
Feedback session	20'	- UEQ - AI & trust questionnaire
Wrap up	5'	Final remarks, closing of the workshop

Figure 16: Agenda of the Waste Management Workshop

3.2.2.3.2 Workshop Results

In this section, the UEQ results and results of Trust Questionnaire are presented.

UEQ Results

The tool received a strong positive evaluation, with an overall score of 1.639. This indicates that users perceive the tool as **supportive, clear, interesting, inventive, and cutting-edge**. The tool's design effectively engages users on both an emotional and experiential level.

Table 12: Overview of UEQ Results for Waste Management

Item	Mean	Variance	Std. Dev.	No.	Negative	Positive	Scale	
1	1.7	0.5	0.7	9	obstructive	supportive	Pragmatic Quality	
2	1.3	1.3	1.1	9	complicated	easy	Pragmatic Quality	
3	0.8	1.7	1.3	9	inefficient	efficient	Pragmatic Quality	
4	1.4	1.0	1.0	9	confusing	clear	Pragmatic Quality	
5	1.7	0.5	0.7	9	boring	exciting	Hedonic Quality	
6	2.0	0.8	0.9	9	not interesting	interesting	Hedonic Quality	
7	2.0	0.3	0.5	9	conventional	inventive	Hedonic Quality	
8	2.2	0.4	0.7	9	usual	leading edge	Hedonic Quality	

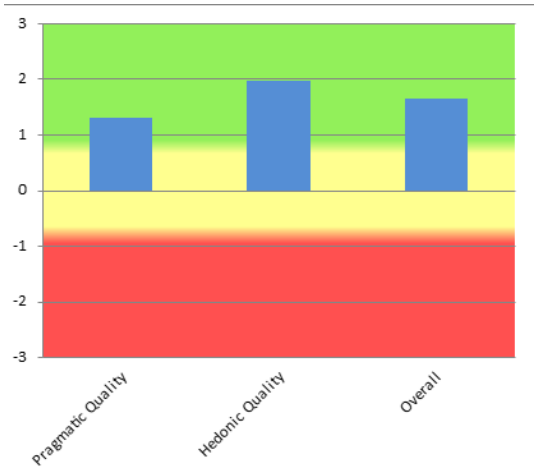


Figure 17: Overview of pragmatic, hedonic and overall quality scoring for Waste Management

The overall score of the tools tested is **1.639,** indicating a strong positive evaluation, with the pragmatic quality dimensions scoring **1.306** and hedonic quality dimensions scoring **1.972.**

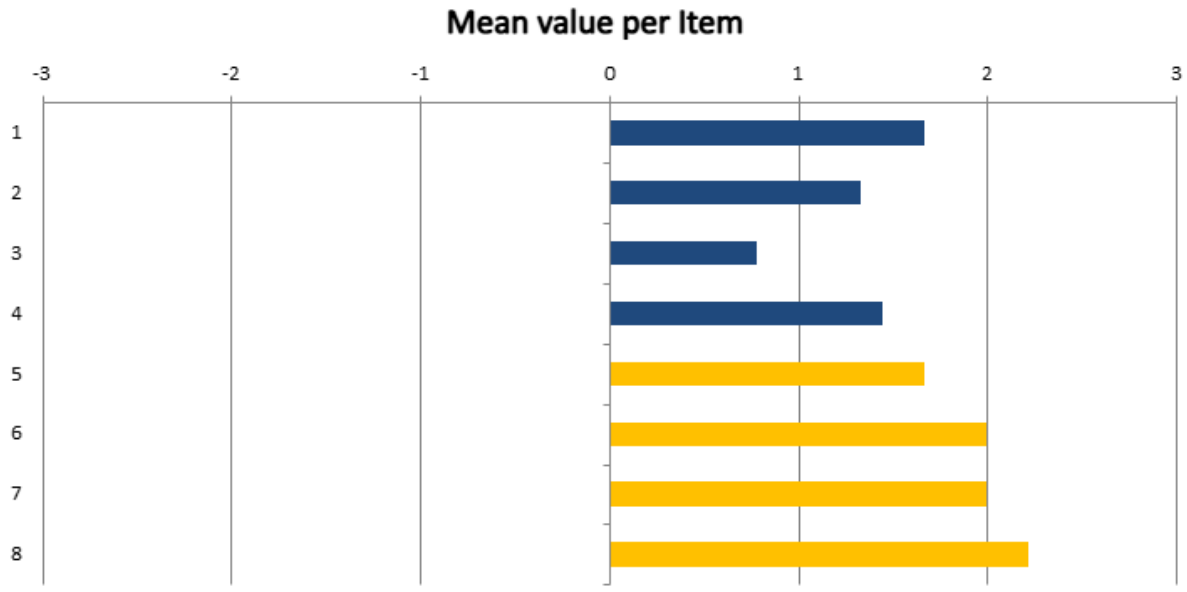


Figure 18: Mean value per Item for the Waste Management UC

For the Pragmatic Quality dimensions

- For the Obstructive vs. Supportive item, the mean is **1.7**, marking a positive evaluation. Users perceive the tool as **supportive**, with a strong positive tendency.
- For the Complicated-Easy item, the evaluation is positive (**1.3** mean), which indicates that the system is generally considered **easy to use**, but the higher variance (1.3) suggests that some users found it less easy.
- For the Inefficient-Efficient item, the evaluation is neutral with a mean score of **0.8**. This evaluation shows that the user found the system **borderline efficient**, though not highly so. This item indicates a potential area for improvement, as the score is on the lower end of positive.
- For the Confusing-Clear item, the mean score is **1.4,** marking a positive evaluation. Users find the system **mostly clear**, but the variance (1.0) shows room for improvement to ensure clarity for all users.

For the Hedonic Qualities:

- For the Boring-Exciting item, the evaluation is strongly positive (mean: **1.7**). Users find the system **exciting**, with a consistent response pattern indicated by the low variance (0.5).
- For the Not Interesting-Interesting item, is again strongly positive with a mean score of **2.0**, indicating that the system is seen as **highly interesting**, with a relatively low variance (0.8), suggesting engagement among users.
- For the Conventional-Inventive item, the evaluation is positive (mean: **2.0**). The tool was perceived by the users as **inventive,** with a strong agreement across participants (variance of 0.3).

- For the Usual-Leading Edge item, the evaluation is the highest of all the other items, with a score of **2.2.** The system is strongly viewed as **cutting-edge** and innovative, with very consistent responses (variance of 0.4)

Trust Questionnaire:

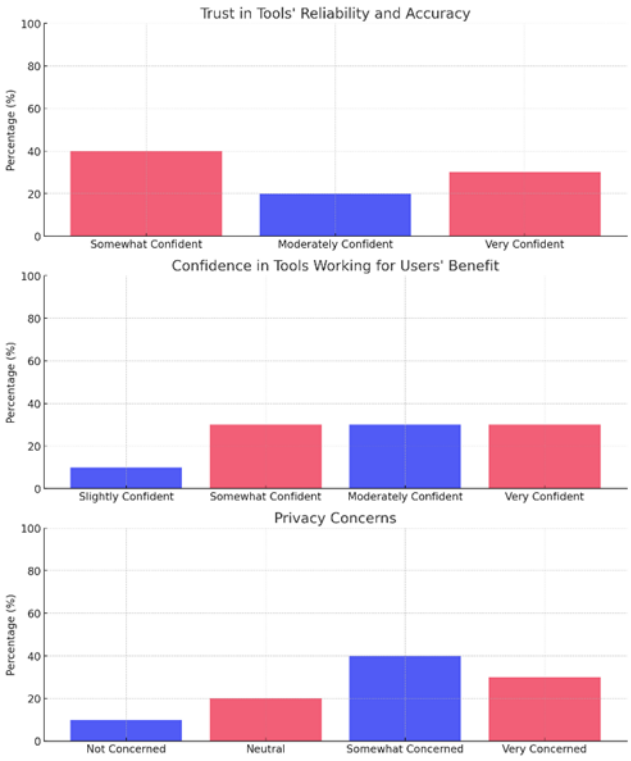


Figure 19: Responses on the Confidence in reliability and accuracy, tool's benefits and privacy and security of the Waste Management evaluation

Based on the results received on the trust questionnaire, the majority of the responders (70%) expressed **moderate to high confidence** in the **reliability and accuracy** of the tools' outputs, suggesting a positive perception of the tools' capabilities. The most common response was "Somewhat confident" (40%), followed by "very confident" (30%) and moderately confident (20%). Furthermore, regarding the question "How confident are you that such tools work for your benefit and can make your job easier?", 60% of the responders felt moderately to very confident in the tools ability to assist them. However, 40% expressed low confidence (10% slightly confident, 30% somewhat confident), primarily due to issues like slow response times and unclear answers for some queries.

Lastly, privacy concerns are significant, with 70% of responders indicating concern (40% "Somewhat concerned," 30% "Very concerned"). Those unconcerned (10%) or neutral (20%) were more focused on operational or data management challenges rather than privacy risks.

3.2.2.3.3 Foreseen optimisations for the 2nd validation

All participants (100% response rate) provided feedback. They showed high interest in the tools, offering constructive comments, suggestions, and questions.

Respondents suggested including real-time and comprehensive data for better accuracy and predictions, while also highlighted the tools' immaturity and issues like delayed responses and unclear feedback for unanswerable queries. Some participants also focused on practical projects over AI development, citing limited utility of current tools, while others emphasized the need to

optimize AI technologies further to address operational inefficiencies. Also, the need to address biased in predictions alongside diverse data use was also pointed out.

Specifically, the members of the waste management directorate were moderately confident about the reliability and accuracy of such tools' outputs in order to make decisions and that they work for their benefit, and they can make their job easier. Also, they are a little bit concerned about the privacy and security of the data processed by such tools.

The Hotel Manager find it very supportive, easy to use, efficient, clear, exciting, interesting and inventive, but not so leading edge. Also, the Hotel Manager is very confident about the reliability and accuracy of such tools' outputs to make decisions, and that they work for their benefit and can make their job easier, while is moderately concerned about the privacy and security of the data processed by such tools. Finally, the Hotel Manager suggests that that the biases are addressed, and more representative data are used for the tools to be more accurate as to the predictions and suggestions made.

Finally, the following suggestions were made:

- The drivers of the garbage trucks can visualise the routing both in maps and real time
- The routing proposed takes also into account the time needed to lift the bin and dispose the garbage into the truck
- The routing proposed takes also into account the seasonality (summer and weekends) as the quantity of garbage is double
- The routing proposed takes also into account the working hours of the staff and overtime needed
- The routing proposed takes also into account the garbage left regularly outside the bins
- The data regarding the fullness of the garbage bins need to be real time and accurate

In general, the visitors find it very supportive, exciting, interesting, inventive and leading edge, while some find less easy to use, clear and efficient. Also, they are moderately to very confident about the reliability and accuracy of such tools' outputs to make decisions, and that they work for their benefit and can make their job easier. They are more concerned about the privacy and security of the data processed by such tools. All visitors suggest that the biases are addressed, and more representative data are used in order for the tools to be more accurate as to the predictions and suggestions made.

Taking into consideration the above-mentioned feedback and suggestions, the following optimisations are foreseen for the second validation phase:

- The underlying technologies will be updated so that the users are able to provide their feedback with regards to the results provided.
- What is more, specific pages will become responsive, allowing the users to view and interact with them from mobile devices as well.
- Further analytics results will be integrated so that the users gain further insights with regards to the flows of citizens and tourists in the Municipality of Vari Voula Vouliagmeni.
- Furthermore, the routing algorithm will be updated to take into account other aspects of garbage collection, as mentioned in the provided feedback.

- Towards the 2nd validation phase, the following optimisations are foreseen:
- The underlying technologies will be updated so that the users are able to provide their feedback with regards to the results provided.
- What is more, specific pages will become responsive, allowing the users to view and interact with them from mobile devices as well.
- Further analytics results will be integrated so that the users gain further insights with regards to the flows of citizens and tourists in the Municipality of Vari Voula Vouliagmeni.
- Furthermore, the routing algorithm will be updated to take into account other aspects of garbage collection, as mentioned in the provided feedback.

3.2.3 Sustainable Development and the European Green Deal (JSI)

This section outlines the first phase of the evaluation workshop for the Sustainable Development and European Green Deal pilot. Specifically, it details the preparation, implementation, and results of the workshop for the Top 100 Projects, SDG Observatory, and OECD Policy Documents UCs, along with the suggested optimizations for the next phase.

Due to missing data, the evaluation for the Alcohol Abuse UC, the newest UC, has been postponed to early 2025. Additionally, an evaluation workshop on Rare Diseases will be conducted in early 2025 as part of the SDG Observatory. This rescheduling is necessary to address technical platform issues and ensure the involvement of relevant stakeholders.

3.2.3.1 Top100 projects

Main sector of interest: Sustainability | Keywords: SDGs, IRCAI, Top100 projects

Summary: Top100 projects is an IRCAI initiative, funding projects to address problems related to the SDGs by using Artificial Intelligence, from all five geographical regions: Africa, Europe and Americas, Asia and the Pacific, and the Middle East. The primary objective is to identify and present solutions worldwide, making a significant contribution to the SDGs through the creation of a vast platform for sustainable solutions.

The aim of the UC will be to create a document-style framework with listed services and a set of rules for bias evaluation. If possible, there will be a toolkit that allows the applicants to assess the bias in their data and models in order to self-evaluate their projects. Alternatively, the UC will create a bias evaluation platform, an API or web application where the applicants would be able to test samples of their datasets and/or their models. That application would help to assess bias and provide a report that could be used in our Top 100 project. In the long run, the UC will establish a clearinghouse for bias evaluation and have a model, rules, and toolkit for bias evaluation of any AI solution in general.

Target stakeholders/users:

- Top100 reviewers: The reviewers will have more guidance on how to evaluate bias and AI ethics concerns in submitted projects.

- Top100 applicants: The applicants will better understand the importance of bias problem and will receive feedback on how they can address those issues

AI4Gov tools to be employed

- Bias Detector Toolkit
- Training materials
- Organisational guidelines and blueprints for trustworthy AI.

3.2.3.1.1 Workshop Organisation and Implementation

Preparatory Phase: The evaluation process for the **Top100 Projects Use Case** deviates from the standard workshop structure employed in other use cases. Instead, it is conducted entirely online via email, coordinated by JSI, and targets **Top100 reviewers** from the past two years. The participant pool consists of **50 individuals not directly involved in the AI4Gov project**, with the objective of achieving a **completion rate of over 60%** for the evaluation questionnaires. Additionally, efforts have been made to maintain a **50/50 gender balance** among participants.

Objective of the evaluation process in this UC: Presentation of questionnaire about ethical considerations of AI solutions, information about inclusiveness and fairness efforts of Top100 applicants, and how they address potential biases in their data and models. The questionnaire will be presented to Top100 reviewers. The goal is to acquire feedback about the questionnaire.

Tools tested: Top100 questionnaire

Evaluation tools used: UC QUE and trust feedback questions.

Workshop Implementation: Due to the international basis of the stakeholder audience, the questionnaire was sent to via email with the relevant information and instructions. Top100 reviewers were invited to assess additional bias/ethics questions developed for Top100 applicants. A link to the online questionnaire was provided via email, asking participants to complete it. The questionnaire can be found in **Appendix 6.7**. Of the 50 participants invited, 11 provided feedback and responses.

Table 13: Top100 Projects workshop audience

Type of audience	**Number of reviewers reached**
Pilot partner employees **not** involved in the project - researchers	**8**
Other	**42**
Total	**50**

3.2.3.1.2 Evaluation Results

UEQ results

Based on the analysis of results of the UEQ responses, the Top100 questionnaire had an **overall score of 0.76,** which reflects a neutral to slightly positive user experience, as shown in Table 14.

Table 14: Overview of the UEQ results for Top100 Questionnaire

Item	Mean	Variance	Std. Dev.	No.	Negative	Positive	Scale	
1	1.0	1.8	1.3	11	obstructive	supportive	Pragmatic Quality	
2	0.5	2.9	1.7	11	complicated	easy	Pragmatic Quality	
3	0.2	2.6	1.6	11	inefficient	efficient	Pragmatic Quality	
4	0.7	3.0	1.7	11	confusing	clear	Pragmatic Quality	
5	0.6	1.3	1.1	11	boring	exciting	Hedonic Quality	
6	1.6	1.5	1.2	11	not interesting	interesting	Hedonic Quality	
7	0.5	2.3	1.5	11	conventional	inventive	Hedonic Quality	
8	0.6	2.3	1.5	11	usual	leading edge	Hedonic Quality	

The **pragmatic quality score is 0.591**, which indicates a neutral evaluation, leaning slightly positive, while the **hedonic quality score of 0.841** indicates that users generally find the tool enjoyable and engaging.

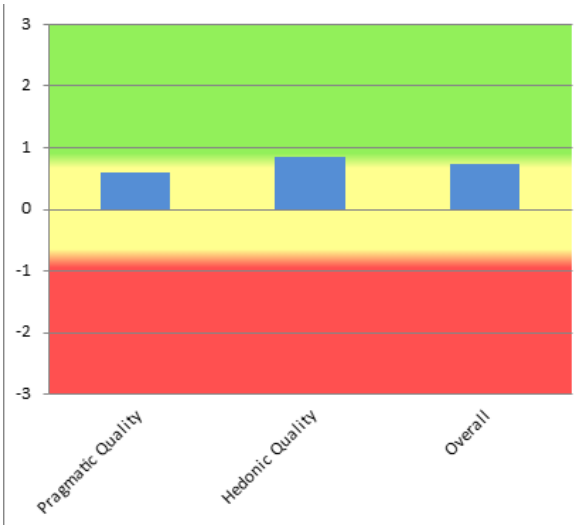


Figure 20: Overview of pragmatic, hedonic and overall quality scoring of Top100 Questionnaire

Next on the mean value per item is presenting (Figure 21), representing the average score given by all participants to the Top100 Questionnaire on the corresponding scale.

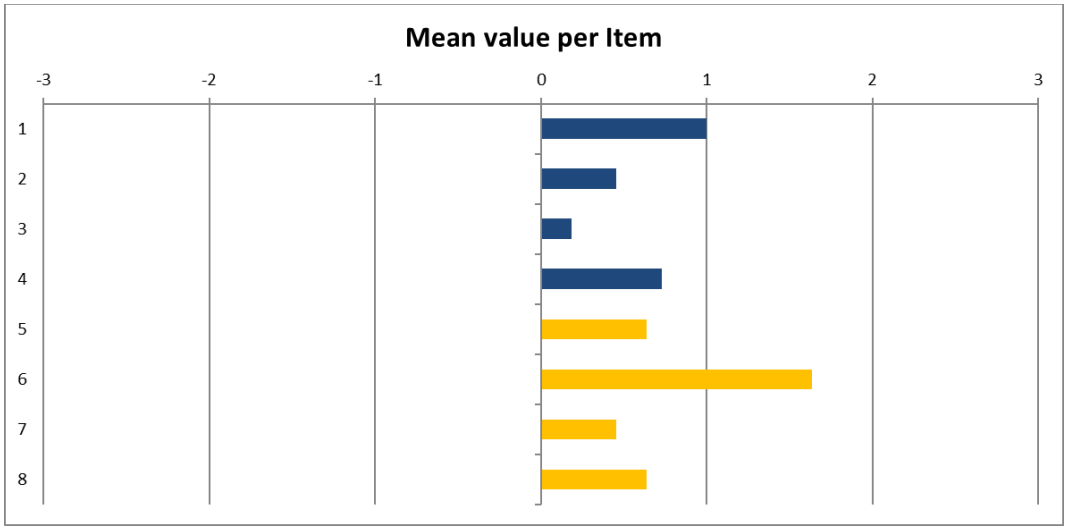


Figure 21: Mean value per item for the Top100 Questionnaire

The **Top100 Questionnaire pragmatic quality** is somewhat positive but nit overwhelmingly so, with most items either neutral or slightly leaning toward a positive evaluation. The overall score of 0.591 confirms this trend, as it lies within the neutral range but close to positive.

- For the Obstructive-Supportive item, Top100 Questionnaire had a positive evaluation with a score of 1.0. Although the score is moderate, is well above the neutral range (0.8), indicating that users found the tool **supportive rather than obstructive**.
- For the Complicated-Easy item, the evaluation was neutral. The mean score is 0.5, which is withing the range (0.8 to 0.8), indicating **no strong tendency toward either ease or complexity**.
- For the Inefficient-Efficient item, again the evaluation is neutral, with a mean score of 0.2. This reflects **a slight leaning toward efficiency**, but overall, the evaluation **remains neutral**.
- For the Confusing- Clear item the mean score is 0.7, which imposes a neutral evaluation, but close to positive. The participants perceive **some clarity** on the form of the Questionnaire, **though not definitively clear**.

The **Hedonic Quality of the Top100 Questionnaire** exhibits a slightly more positive trend than pragmatic quality, with one strong positive evaluation. The overall hedonic quality score of 0.841 is close to the positive range.

- For the Boring- Exciting item the score is 0.6, which slightly leans to positive evaluation. The participants found the tool **mildly exciting**.
- For the Not Interesting – Interesting item there is a strong positive evaluation with a score of 1.6. A mean of 1.6 indicates that the users perceive the tool as **highly interesting.**
- For the Conventional-Inventive item the score is 0.5. The tool is seen as **slightly inventive** but not distinctly so.

- For the Usual-Leading Edge item the score is 0.6, highlighting a neutral evaluation, leaning positive. Users perceive the product as **slightly innovative**.

As a summary, while the hedonic aspects of the product are generally perceived more positively, pragmatic quality evaluations remain neutral to moderately positive. Given the restricted range of commonly observed values, **the scores suggest that the Top100 Questionnaire provides a somewhat satisfactory user experience but may benefit from enhancements in clarity and ease of use to improve its pragmatic appeal.**

Trust questionnaire results

The results from the questions regarding the trust on the reliability of bias questionnaires were moderate, as the respondents showed moderate to high confidence in the reliability and utility of the bias questionnaire but less consistent opinions on privacy concerns. Additionally, it was noted variation in responses which suggests differing perceptions, possibly influenced by individual experiences or knowledge level about the tools.

For the Q5a: **Confidence in the reliability and accuracy of bias questionnaires,** most participants have moderate confidence in the reliability and accuracy of bias questionnaires for decision-making. However, only a small portion is very confident, indicating room for improvement in the perceived reliability of these tools. In Particular, 55% of responders were moderately confident, 27% were "Somewhat confident" and 18% were "Very confident".

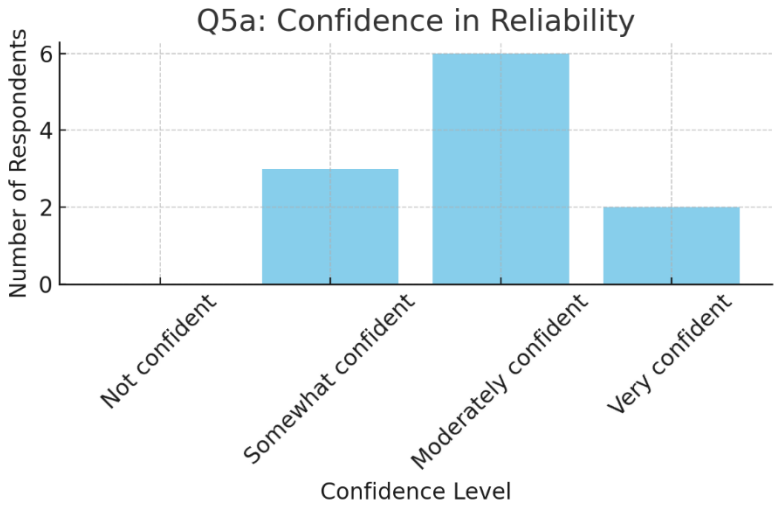


Figure 22: Confidence in the reliability of Top100 Questionnaire

For the Q6a: **Confidence in the tools' ability to benefit the user and ease their job,** half of the participants were moderately confident, while 30% have a high degree of confidence in the tools' usefulness. However, a small but notable group expressed low confidence, indicating potential skepticism or lack of clarity about the tools' benefits. In particular: 50% were "Moderately confident;" 30% were "Very confident;" and 10% each were "Slightly confident" and "Not confident at all."

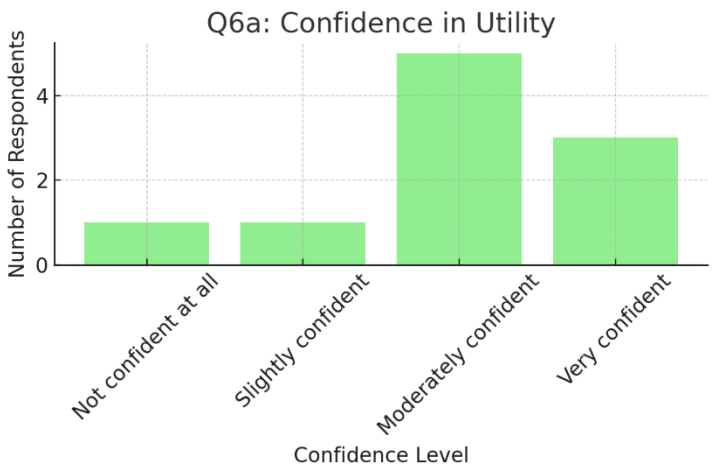


Figure 23: Confidence in Utility of Top100 Questionnaire

For the Q7a: **Concerns about privacy and security of data processed by the bias questionnaire**, while the average concern level is low to moderate, responses vary widely. A significant portion (36%) was not concerned at all, but a comparable group was somewhat concerned, suggesting that privacy and security issues might be important to address for some users. In particular, 36% each were "Somewhat concerned" and "Not concerned;" 18% were "Neutral" and 9% were "Very concerned."

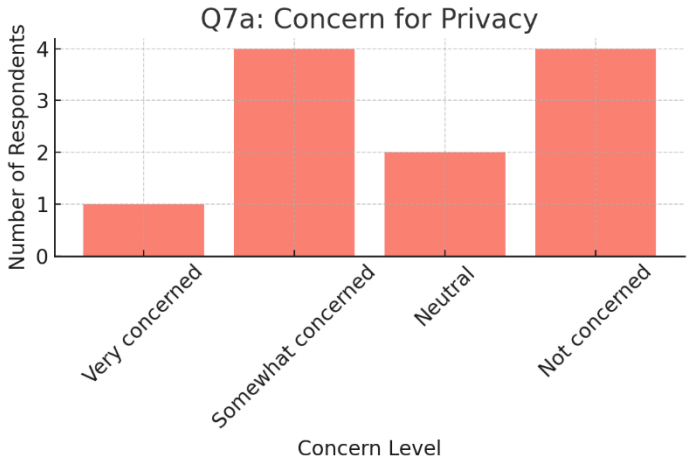


Figure 24: Concern about Privacy and security of data processed by the biased questionnaire

3.2.3.1.3 Foreseen optimisations for the 2nd validation phase

Additionally, the responders provided their feedback via email on the bias/ethics extra questions added by JSI. Below is presented an overview of the feedback received. The general feedback received highlighted the positive perception of the ethics-focused approach while noting the need for ongoing updates and refinements.

- Some participants noted the need for specificity when addressing ethics, using established frameworks such as UNESCO's *Ethical Impact Assessment Questionnaire,* as guidance, while other pointed out the concern that certain projects may not find all the questions directly relevant as they might not involve personal data.
- Furthermore, some comments were received concerning the terminology used suggesting the use of more neutral terms like “tools”, “services” or “innovations” instead of “AI Solutions” to better reflect the nature of the entries.
- Lastly, suggestions were made for modifications of some questions in order to be clearer for applicants. These suggestions included the use of open-ended or context-specific questions and the reframe of yes/no questions to elicit more detailed insights.

3.2.3.2 SDG observatory

Main sector of interest: Sustainability | Keywords: SDGs, IRCAI, SDG achievements, bias

Summary: This UC aims to create a tool that will monitor the achievements of the SDGs, in order to support policymakers. The methodology for monitoring the achievement of all SDG‘s is still being developed, and the development team is evaluating additional data sources that should be included into the SDG Observatory. The development of the SDG observatory needs to address the problem of possible bias in data and in general as much as possible. This UC will develop tools to inform developers about the bias problem and tools and methodologies to detect and eliminate biases to make the data as unbiased as possible, while identifying the topics and themes that are more prone to bias. Bias could also be in not showing some data, that are otherwise relevant, so this UC is interested in the lack of data as well (for instance, because it is not available). The UC will describe the process of how the data set selection happens, in order to locate what is missing. The UC aspires to provide visualisation and analysis of achieving SDGs in an unbiased way, taking into consideration the issue of bias in data and models.

Target stakeholders/users:

- Policy makers in the EU/Global/National level: This group would be able to identify the best practices for solving SDG’s and design policies to contribute.
- Researchers & General public/citizens interested in SDG’s: They would be able to identify the best practices for solving SDG’s and better understand if and how the policies they see being implemented are aiming towards the achievement of the SDGs. A more informed public will challenge and assess the decision makers pressuring them into a more sustainable policy making.

Updated related online infrastructure

Developers already included EventRegistry data, OECD AI policy documents and OpenAlex data.

AI4Gov tools to be employed

- Bias Detector Toolkit
- Training materials
- Organisational guidelines and blueprints for trustworthy AI
- Visualisation workbench

3.2.3.2.1 Workshop Organisation and Implementation

Preparatory Phase: The preparatory phase for the evaluation workshops focused on meticulous planning and coordination to ensure their successful execution. The workshops were structured as two distinct sessions: one centered on the SDG Observatory Use Case and the other on Rare Diseases Observatory. Key metrics and KPIs were defined to assess the effectiveness and gather meaningful feedback. The target was to include 10 participants per workshop who had no prior involvement with the AI4Gov project, ensuring an unbiased perspective. Additionally, a minimum of 60% response rate to the evaluation questionnaires was set as a benchmark for success. Gender balance among participants was emphasised, with the aim of achieving a 50/50 male-to-female ratio.

The **SDG Observatory Workshop** was planned as a hybrid event to maximise accessibility and participation, and took place on 27 December 2024, in the premises of JSI and online. The primary goal was to present the SDG Observatory tool to researchers, policymakers, and potential users, gathering feedback on its relevance, trustworthiness, and utility. The evaluation tools used for this workshop was the UEQ, the trustworthiness exercise and 4 additional questions provide to the participants regarding the fairness and inclusivity of the SDG Observatory. This additional questionnaire can be found in **Appendix 6.8.**

Table 15: SDG Observatory workshop audience

Type of audience	**No of participants**
Pilot partner employees **involved** in the project	**2**
Pilot partner employees **not** involved in the project - researchers	**13**
Pilot partner employees **not** involved in the project - other	**3**
Other	**3**
Total	**21**

3.2.3.2.2 Workshop Results

The participants filled out the UEQ questionnaire and the trust questionnaire after testing the SDG Observatory tool. In the same questionnaire five additional questions were incorporated, focusing on technical aspects of the tool and suggestions for improvement the system's fairness and inclusivity. In this section the analysis of the results received will be analyzed and interpreted.

UEQ Questionnaire

The SDG Observatory scored **0.977** in the UEQ questionnaire. This indicates a predominantly positive user experience.

Table 16: Overview of UEQ results for SDG Observatory

Item	Mean	Variance	Std. Dev.	No.	Negative	Positive	Scale	
1	1.2	3.0	1.7	16	obstructive	supportive	Pragmatic Quality	
2	0.1	3.1	1.8	16	complicated	easy	Pragmatic Quality	
3	0.8	3.1	1.8	16	inefficient	efficient	Pragmatic Quality	
4	0.8	3.3	1.8	16	confusing	clear	Pragmatic Quality	
5	1.2	3.1	1.8	16	boring	exciting	Hedonic Quality	
6	1.1	3.7	1.9	16	not interesting	interesting	Hedonic Quality	
7	1.5	2.3	1.5	16	conventional	inventive	Hedonic Quality	
8	1.1	2.1	1.4	16	usual	leading edge	Hedonic Quality	

While the hedonic aspects are a clear strength, pragmatic elements like clarity and ease of use can be further improved to enhance the overall experience. On the pragmatic quality the tool scored 0.750, signifying a moderately positive evaluation of the product's usability and functionality. The overall hedonic quality ranked higher at 1.203, indicating a strongly positive evaluation of the product's enjoyment, innovation, and engagement.

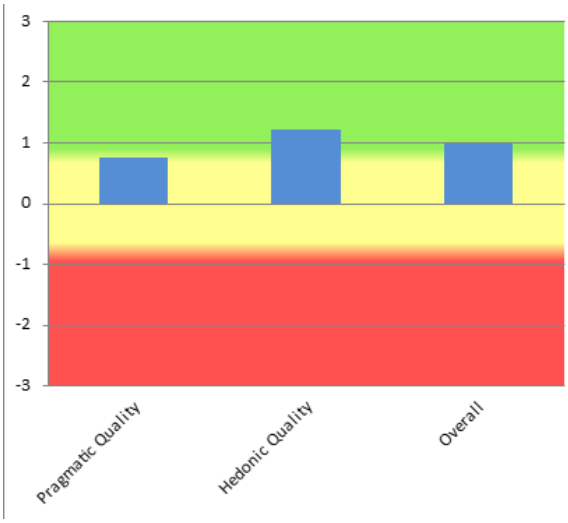


Figure 25: Overview of pragmatic, hedonic and overall quality scoring of the SDG Observatory

The mean values suggest moderate positivity in pragmatic quality, with users finding the tool somewhat supportive and efficient. However, clarity and ease of use remain in the neutral zone, reflecting room for improvement.

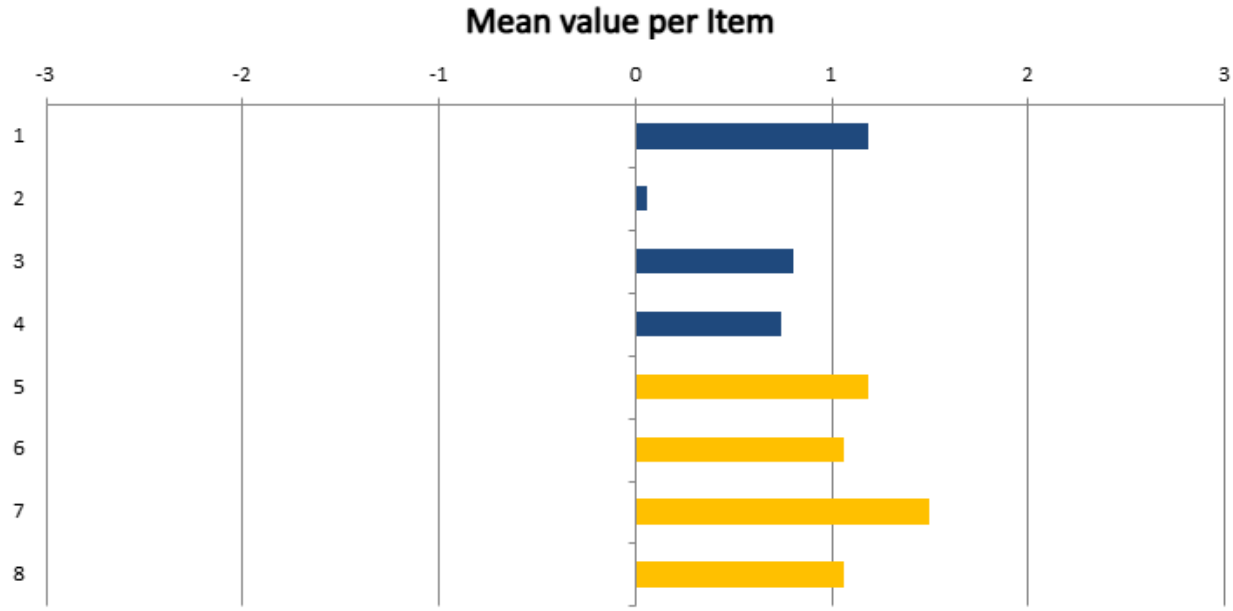


Figure 26: Mean value per Item for the SDG Observatory

The overall score for Pragmatic Quality (0.750) supports this interpretation.

- For the Obstructive vs. Supportive item, the score is 1.2. This score exceeds the positive threshold (0.8), indicating users generally find the tool **supportive**.
- For the Complicated- Easy item, the score is 0.1, which falls within the neutral range (-0.8 to 0.8), suggesting **users do not perceive the tool as particularly easy or complicated**.
- For the Inefficientvs. Efficient item, the evaluation was neutral, leaning positive. The 0.8 score is at the upper end of the neutral range, indicating some **perception of efficiency** but **not overwhelmingly** so.
- For the Confusing vs. Clear item, the score is the same as the previous item. It indicates a **slight perception of clarity**, but it remains within the neutral range.

The hedonic quality items are all rated positively, with particular strengths in perceived excitement, inventiveness, and innovation. The overall Hedonic Quality score (1.203) reflects a **clear positive trend**, indicating users generally enjoy the product's innovative and engaging aspects.

- For the Boring vs. Exciting item, the mean score is 1.2, reflecting a positive evaluation and a clear perception of the tool as **exciting** by the users.
- For the Not Interesting vs. Interesting item, the mean score is 1.1, highlighting a positive evaluation of the tool. Users generally perceive the SDG Observatory as **interesting.**
- For the Conventional vs. Inventive item, the evaluation is strongly positive. With a score of 1.5, the users view the tool as **inventive and innovative**.
- For the Usual vs. Leading Edge item, the score of 1.1 indicates that users find the tool **somewhat cutting-edge**, marking a positive evaluation.

Overall, the SDG Observatory offers **a positive user experience**, with users appreciating its innovative and engaging nature. However, slight improvements in its ease of use and clarity can help raise the perception of its pragmatic qualities to match the strong hedonic appeal.

Trust Questionnaire

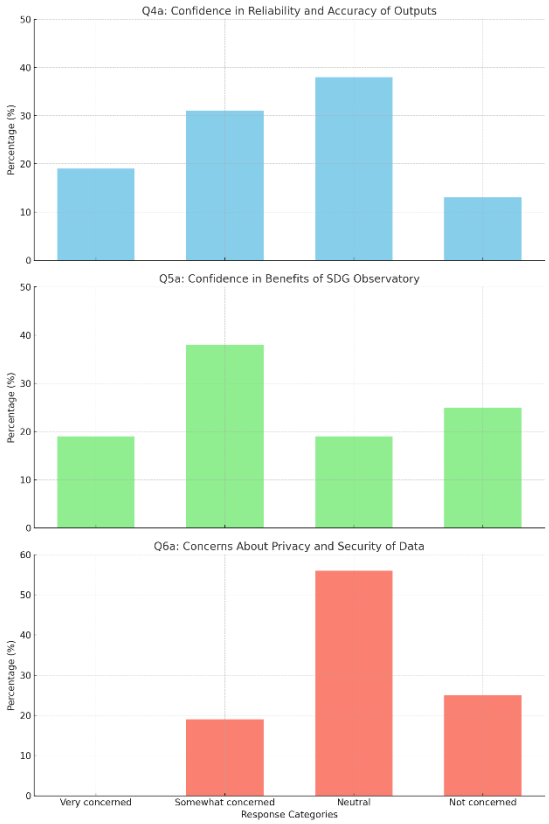


Figure 27: Responses on the Confidence in reliability and accuracy, tool's benefits and privacy and security of the SDG Observatory

The participants expressed a **moderate level of confidence** in the reliability and accuracy of the SDG Observatory's outputs. While a significant proportion (38%) expressed **moderate confidence**, a substantial number (31%) were **somewhat confident**. Additionally, 19% were **not confident** at all, highlighting potential areas for improvement in the tool's accuracy and reliability. The results indicate a **moderate level of confidence** in the tool's ability to streamline work processes. While 25% of participants expressed **high confidence**, a significant portion (38%) were **slightly confident** and 19% were **not confident at all**. This suggests that while the tool has the potential to be a valuable resource, there is still room for improvement to fully realize its benefits. The majority of participants (56%) expressed a **neutral stance** towards privacy and security concerns. Additionally, 25% were **not concerned at all**. Only 19% expressed **some concern**. The absence of responses in the "very concerned" category suggests that the tool's security measures have been effective in alleviating major privacy worries.

Participants were asked to provide their opinions on several key issues related to the fairness and inclusivity of the SDG Observatory tool. These included:

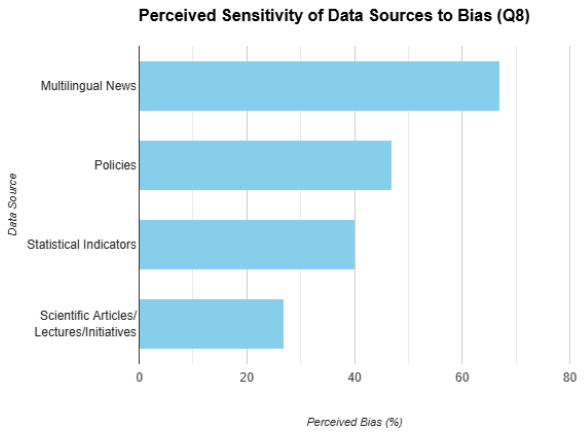
Bias and Sensitivity: Concerns about the potential for bias in the tool's algorithms and the sensitivity of the data used to train them.

Perspective and Representation: Questions about whether the tool favors specific perspectives or stakeholders, and if it might misrepresent or exclude underrepresented groups.

Participants were also given the opportunity to share suggestions for improving the tool.

When asked about the data sources most susceptible to bias during the data ingestion process, participants primarily identified **multilingual news** as the most vulnerable (67%). This was

followed by **policies** (47%), **statistical indicators** (40%), and **scientific articles, video lectures, and innovation initiatives** (27% each).



In essence, the results suggest that participants perceive multilingual news as the primary source of potential bias in the SDG Observatory tool. This could be due to factors such as language translation inaccuracies, cultural nuances, or inherent biases present in news reporting.

Figure 28: Perceived Sensitivity of Data Sources to Bias

When asked about which algorithms participants believed were most likely to introduce bias, the results indicate that **recommender systems** were perceived as the most problematic, with 54% of respondents expressing this concern. Following closely behind were **K-means clustering** and **NLP text similarity** algorithms, each perceived as potentially biased by 46% of participants. **Random forest regression** and **BERT classifier** were seen as less likely to introduce bias, with 31% of respondents identifying them as potential sources of bias. **Overall, the results highlight the perception that machine learning algorithms, particularly those used in recommendation systems and text analysis, may introduce bias into the decision-making process.**

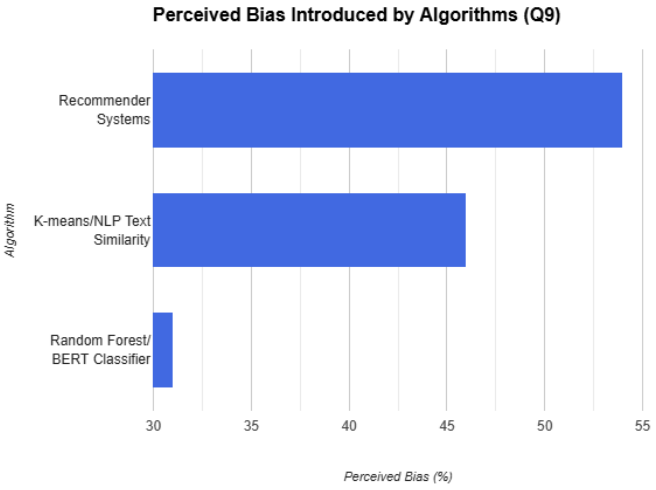


Figure 29: Perceived Bias Introduced by Algorithms

The responses to Q10a indicate a **moderate level of perceived data favoritism** in the SDG Observatory tool. While a significant proportion of respondents (50%) believe that data favoritism occurs "sometimes," a smaller group (14%) perceive it as "often." However, a notable 29% of respondents believe that data favoritism is "rarely" observed, and 7% believe it "never" occurs.

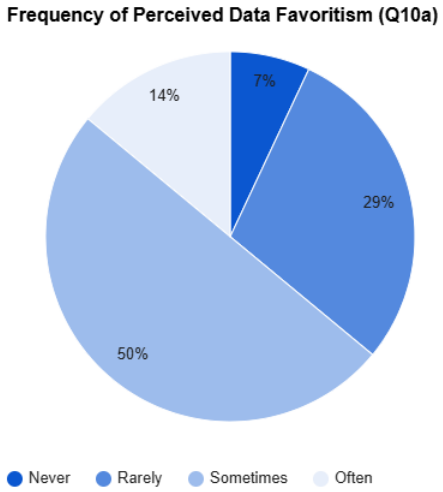


Figure 30: Frequency of Perceived Data Favoritism

Regarding the misrepresentation of underrepresented communities, the majority of respondents (63%) expressed uncertainty. However, 31% of respondents believe that the tool does **not** misrepresent underrepresented communities. No respondents indicated that the tool **does** misrepresent these groups.

The results suggest a mixed perception of the tool's fairness and inclusivity. While some respondents believe that the tool may exhibit biases and favoritism, others are uncertain or do not perceive such issues. Further investigation and improvement efforts may be necessary to address these concerns and ensure the tool's unbiased representation of diverse perspectives and communities.

3.2.3.2.3 Foreseen optimisations for the 2nd validation phase

In addition, the participants suggested several key strategies to reduce bias.

- **Diverse Data Sources and Stakeholders:** Collecting data from a variety of sources and involving diverse stakeholders can help mitigate bias.
- **Clear Data Presentation and Transparency:** Providing clear information about data sources and presentation methods can increase transparency and reduce the potential for misinterpretation.
- **Bias Disclaimers and User Feedback:** Acknowledging the potential for bias and providing mechanisms for user feedback can help identify and address biases.
- Participants also offered additional suggestions for improving the tool:
- **Enhanced Data Visualization:** Adding textual explanations to graphs can help clarify the data and make it more accessible to a wider audience.

- **"Not Sure" Option for Technical Questions:** Including a "Not Sure" option can accommodate users who may not have the technical expertise to answer specific questions.
- **Positive Reinforcement and Iterative Improvement:** Recognizing the tool's strengths while suggesting areas for improvement can foster a positive and constructive approach to development.

3.2.3.3 OECD policy documents

Main sector of interest: AI Ethics & Policies | Keywords: OECD papers, AI policies, anti-bias strategies

Summary: OECD has a collection of various national AI policies and strategies. They have an online repository with over 800 AI policy initiatives from 69 countries, territories, and the EU. The UC will analyse these documents and map the chapters referring to bias. A visual summary of how these documents approach bias and what solutions they are providing to tackle the problem of bias will be created. This will facilitate the mapping of good practices. This UC will provide visualisation and analysis of how different countries and international organisations are trying to tackle the problem of bias in AI, based on the content of the OECD policy documents. The team will analyse these documents and find the chapters presenting anti-bias policies and practices. Then a visual summary will be created presenting solutions and good practices. This UC will raise awareness of the importance of ethics in AI, and the importance of bias prevention approaches. Policy makers should be encouraged to address these issues accordingly and raise awareness among developers of AI solutions.

Target stakeholders/users:

- Policy makers: Policy makers would be able to identify best policy and legal practices dealing with bias in AI and AI ethics. Also, they can gain insights into the most effective policies that address bias in AI. This includes understanding which regulatory measures have been successful in other jurisdictions. As a result, they can integrate best practices for promoting AI ethics into the legislative process, ensuring that new regulations adequately address ethical concerns.
- Legal and ethical experts interested in AI: This group would be able to assess the current legal frameworks governing AI, identifying gaps and areas for improvement. This includes analysing existing laws on data protection, privacy, and anti-discrimination as they apply to AI technologies.
- Journalists and general public interested in AI: This group would be able to identify the current regulations and good practices, gaining a deeper understanding of the regulatory environment surrounding AI. This will enable them to report accurately and informatively on AI-related issues.

AI4Gov tools to be employed

- Visualisation workbench
- Policy oriented analytics & AI algorithms

3.2.3.3.1 Workshop Organisation and Implementation

Preparatory Phase: During the preparatory phase, a meeting was held with the primary partner involved in this Use Case, JSI (Jožef Stefan Institute). The purpose of the meeting was to organise and schedule the evaluation workshop and finalise its key elements. The workshop will include **10 participants** from diverse domains, such as JSI/IRCAI employees, researchers, and policymakers, ensuring a well-rounded representation of potential users. The participant criteria were set, including an engagement target of over 60% responding to the questionnaires and a 50/50 gender ratio will be maintained to ensure inclusivity and diversity.

The workshop was scheduled for 27 November 2024 in a hybrid format at the premises of JSI and online. The primary goal of the workshop was to present the **OECD Policy Documents Database** and the **OECD Documents Chatbot** to researchers and potential users. The workshop aimed to gather feedback on the usefulness of the solutions in their respective domains and the participants' level of trust in the chatbot and database. Furthermore, suggestions for improvements and additional features that could enhance usability, and trustworthiness would be valuable for the next phase of the project.

Table 17: OECD Policy Documents audience

Type of audience	**No of participants**
Pilot partner employees **involved** in the project	**2**
Pilot partner employees **not** involved in the project - researchers	**13**
Pilot partner employees **not** involved in the project - other	**3**
Other	**3**
Total	**21**

Tools to be tested: OECD documents chatbot

Evaluation tools to be used: UC UEQ and questions from "trustworthiness focus group".

3.2.3.3.2 Workshop Results

UEQ Questionnaire Results

The evaluation of the OECD policy chatbot received a nearly neutral overall score, which reflects a user experience with **significant room for improvement**, particularly in terms of efficiency and innovation.

Table 18: Overview of UEQ Results for OECD Policy Documents Chatbot

Item	Mean	Variance	Std. Dev.	No.	Negative	Positive	Scale	
1	-0.3	0.9	0.9	13	obstructive	supportive	Pragmatic Quality	
2	0.9	1.9	1.4	13	complicated	easy	Pragmatic Quality	
3	-1.3	1.4	1.2	13	inefficient	efficient	Pragmatic Quality	
4	0.1	2.4	1.6	13	confusing	clear	Pragmatic Quality	
5	0.4	2.1	1.4	13	boring	exciting	Hedonic Quality	
6	0.5	3.4	1.9	13	not interesting	interesting	Hedonic Quality	
7	0.2	2.5	1.6	12	conventional	inventive	Hedonic Quality	
8	0.0	1.7	1.3	13	usual	leading edge	Hedonic Quality	

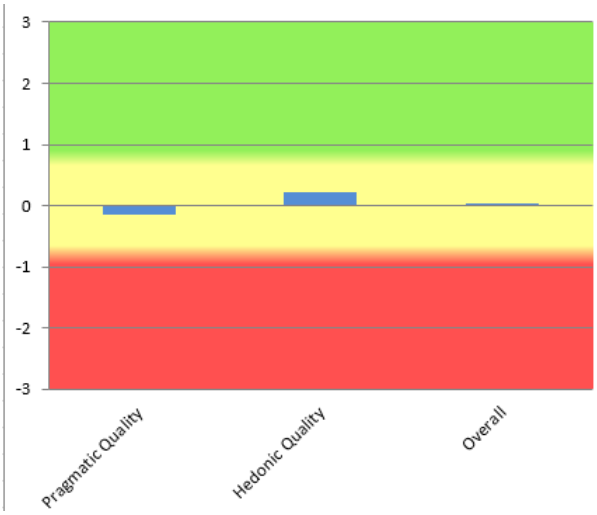


Figure 31: Overview of pragmatic, hedonic and overall scores of OECD Policy Documents Chatbot

The score of -0.154 of the pragmatic quality points out a slightly negative evaluation, driven by perceptions of inefficiency and mixed clarity, while on the opposite, the score of 0.231 indicates a neutral to slightly positive evaluation, suggesting a limited engagement or excitement with the product's innovative qualities.

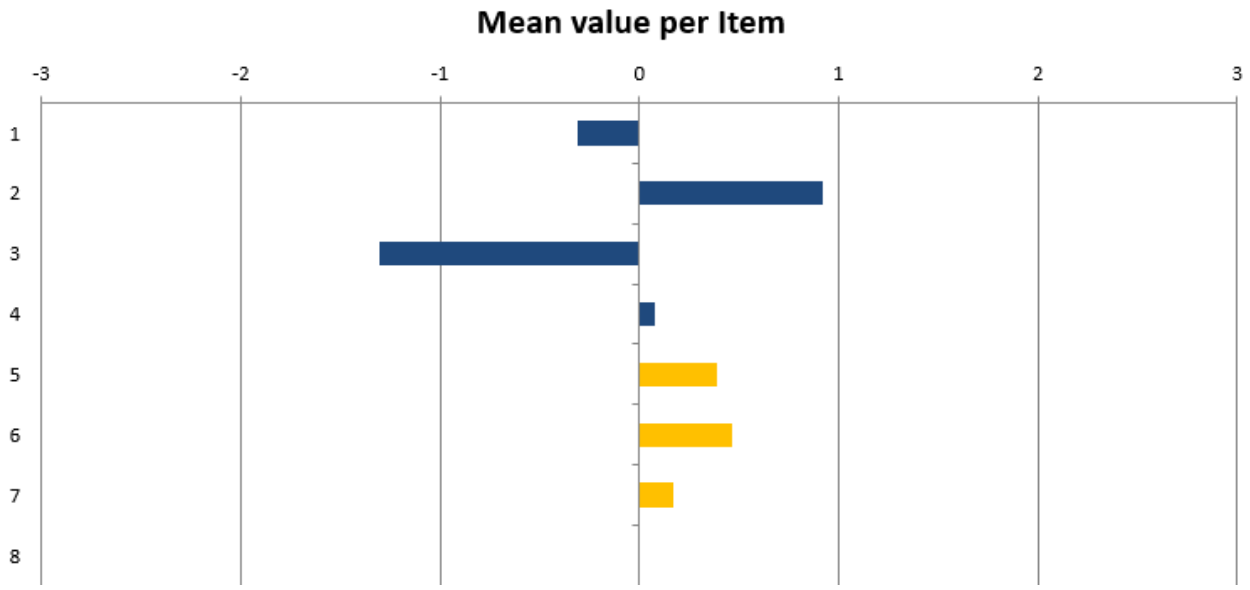


Figure 32: Mean value per item for OECD Policy Documents

The overall **Pragmatic Quality score (-0.154)** indicates a predominantly neutral to slightly negative evaluation. The perception of inefficiency (Item 3) weighs heavily on the overall pragmatic experience, even though ease of use (Item 2) is rated positively.

- For the Obstructive vs. Supportive item, the mean score is -0.3, which indicates a neutral evaluation, leaning **slightly toward obstructive**.
- For the Complicated vs. Easy item, the mean score of 0.9 slightly exceeds the positive threshold (0.8), indicating users found the tool **somewhat easy to use**.
- For the Inefficient vs. Efficient item, the evaluation is negative with a score of -1.3. This score indicates that **steps towards improvement** must be taken seriously under consideration in the fine-tuning process by the technical partners.
- For the Confusing vs. Clear item, the mean score is 0.1. The score is within the neutral range, indicating **mixed opinions** about the tool's clarity.

The hedonic quality scores are generally neutral, with no strong positive or negative evaluations. The overall **Hedonic Quality score (0.231)** reflects slight positivity but lacks strong user enthusiasm.

- For the Boring vs. Exciting item, ranked a score of 0.4, suggesting users find the tool **slightly exciting**, but not strongly so.
- For the Not Interesting vs. Interesting item, hit a score of 0.5, which indicates a neutral evaluation, leaning positive. The tool is perceived as **somewhat interesting**.
- For the Conventional vs. Inventive item, the evaluation is neutral with a mean score of 0.2. The score indicates the tool is perceived as **slightly inventive**, but overall evaluations are mixed.

- For Usual vs. Leading Edge item, the evaluation is neutral with a mean score of 0.0. Users do not perceive the tool as particularly innovative or cutting-edge.

Trust questionnaire

The results from the trust questionnaire indicated that the confidence in the chatbot's reliability, accuracy, and benefits is generally low, while privacy and security concerns are relatively moderate. During this evaluation cycle the findings suggest opportunities for improvement in the chatbot's functionality, reliability and communication of its data measures.

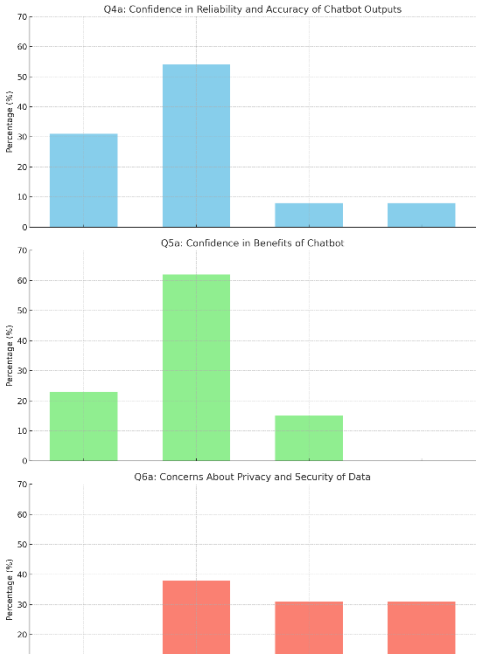


Figure 33: Responses on the Confidence in reliability and accuracy, tool's benefits and privacy and security of the OECD Policy Documents evaluation

The responses to Q4a indicate a **moderate level of confidence** in the reliability and accuracy of the Policy chatbot's outputs. A significant majority of respondents (54%) expressed **somewhat confident**, while 31% were **not confident**. Only a small percentage (8%) were **moderately confident** or **very confident,** suggesting an overall low to moderate confidence in using the chatbot for decision-making.

The responses to Q5a indicate a **limited level of confidence** in the benefits of the chatbot. A majority of respondents (62%) expressed **slight confidence**, while 23% were **not confident at all**. Only 15% were **moderately confident**, and no one was **very confident**. This reinforces the finding that users perceive the chatbot as somewhat helpful but not a highly effective tool for their tasks.

The responses to Q6a indicate a **moderate level of concern** about the privacy and security of data in the SDG Observatory. While 38% of respondents were **somewhat concerned**, a significant portion (31%) were **neutral**, and another 31% were **not concerned**. This suggests that while some users have concerns about data privacy and security, others are relatively unconcerned.

3.2.3.3.3 Foreseen optimisations for the 2nd validation phase

From the feedback received form the participants the following key insights and suggestions were extracted:

1. Chatbot Functionality and Performance:

- **Reliability:** Users reported instances of the chatbot not functioning properly.
- **Response Time:** Slow response times were identified as a major issue.
- **Accuracy and Relevance:** The chatbot often provided inaccurate or irrelevant answers.
- **2. User Interface and Experience:**
- **Clarity and Guidance:** Users expressed confusion regarding policies, background documents, and error messages.
- **Visual Feedback:** Lack of clear visual feedback during processing was identified as a usability issue.
- **3. Data Quality and Source Transparency:**
- **Database Completeness:** Users noted the need for a more comprehensive database, especially for less-represented countries.
- **Source Transparency:** Users expressed interest in accessing the sources used to generate the chatbot's responses.
- **4. Bias and Fairness:**
- **Algorithmic Bias:** Some users raised concerns about potential biases in the tool's algorithms.
- **User Feedback Mechanism:** The need for a mechanism to collect user feedback on bias and other issues was highlighted.

Towards the 2nd validation phase, the following optimisations are foreseen:

- Based on the feedback provided, the Visualization Workbench's user interface will be updated. The users will be able to view multiple documents. Thus, they will be able to compare the documents with regards to their context
- The response time of the chatbot will be improved by updating the implementation of the Policy oriented analytics & AI algorithms component.

3.3 Lessons learnt from the evaluation process

The first validation phase of the AI4Gov project provided essential insights into the strengths, challenges, and stakeholder perceptions regarding the pilot tools. Key lessons learned are summarised below:

- **Early-Stage Engagement**: Stakeholders responded positively to the tools' potential. Ensuring continuous engagement and clear communication about tool capabilities during workshops fosters trust and openness to feedback.
- **User Confidence and Skepticism**: While participants acknowledged the tools' promise in improving operational efficiency, skepticism arose around usability, accuracy, and reliability. A critical takeaway is the need for continuous refinement of AI models and the integration of real-time, high-quality data.
- **Trust and Security Concerns**: Security and trust emerged as significant themes across all pilots. Users are cautious about data privacy, the reliability of predictions, and potential biases. Addressing these concerns through transparent communication and robust security frameworks is critical.
- **Tailored Approaches for Feedback**: The mixed methods (surveys, focus groups, and interactive exercises) proved effective in capturing diverse insights. Future iterations should refine evaluation tools to balance simplicity with depth to encourage higher response rates.
- **Improvement Areas**: Stakeholders highlighted areas for enhancement, including user interfaces, data responsiveness, and inclusion of real-world contextual factors (e.g., seasonal trends or geographical nuances).

The evaluation of the first validation phase of the AI4Gov project revealed valuable insights into the tools' performance, challenges, and stakeholder engagement. Overall, the tools were positively perceived for their potential to improve operational processes and decision-making. Across all use cases, stakeholders acknowledged that the tools demonstrated significant promise, particularly in terms of their innovative and supportive nature. However, as the tools are still in their early stages of development, some limitations were identified that need to be addressed to fully meet project objectives. While the tools scored favorably in pragmatic qualities such as usability and clarity, efficiency results were more mixed, suggesting room for improvement. In contrast, hedonic qualities, such as engagement, excitement, and originality, received highly positive evaluations, reflecting the tools' ability to captivate and motivate users.

Despite the positive outlook, the evaluation highlighted several challenges that emerged during the process. Data quality and bias were central concerns, with participants emphasizing the need for accurate, real-time, and representative data to improve the reliability of predictions and recommendations. Security and privacy also surfaced as significant issues, as users expressed apprehension about the handling of sensitive data and the risks associated with cyberattacks or misuse. Moreover, while participants recognized the tools' potential utility, they also noted that certain features felt incomplete or not yet mature enough for real-world deployment. These limitations were particularly evident in areas such as efficiency, responsiveness, and ease of use.

Stakeholder engagement during the workshops was strong, with participants providing constructive feedback and actively engaging in discussions and exercises. Their perceptions of the tools were shaped by both their direct interactions and broader experiences with AI technologies. Trust in the tools was identified as conditional, heavily dependent on the reliability of the data, the transparency of the processes, and the explainability of the results. Participants appreciated the tools' innovative and leading-edge features but highlighted concerns about learning curves, job displacement, and the potential for bias.

In conclusion, the evaluation results demonstrated a generally positive reception of the AI4Gov tools, particularly in terms of their ability to excite and engage users. However, challenges related to data quality, privacy, and tool maturity must be addressed to ensure broader acceptance and usability. The feedback provided during this phase offers a clear path for improvement, focusing on enhancing data accuracy, refining user interfaces, and implementing robust security measures. These adjustments will be critical to aligning the tools with stakeholders' expectations and ensuring that the project's objectives are fully realized in the next validation phase.

4 Conclusion and next steps

The first validation phase of the AI4Gov project has provided valuable insights into the performance, usability, and perceived trustworthiness of the tools developed under each use case. Through a structured evaluation methodology that combined both formative and summative approaches, the project successfully engaged stakeholders, gathered essential feedback, and identified areas for improvement.

The results from the first phase indicate that the AI4Gov tools are progressing well toward achieving their objectives. Stakeholders recognized the tools' potential to improve operational efficiency, enhance decision-making processes, and introduce innovative solutions for policy optimization. The user experience evaluations demonstrated a generally positive perception of both the **pragmatic qualities** (e.g., usability, clarity) and **hedonic qualities** (e.g., interest, inventiveness) of the tools. While participants appreciated the tools' modern and supportive features, areas for improvement, particularly regarding efficiency and responsiveness, were highlighted.

Key findings include:

- **Positive Stakeholder Perceptions**: Stakeholders across all use cases found the tools valuable for optimizing processes, providing actionable insights, and addressing domain-specific challenges such as water management, waste reduction, and policy visualization.
- **Trust and Security Concerns**: Trust in AI tools remains conditional, influenced by data reliability, transparency, and perceived fairness. Concerns around **bias**, **data security**, and the explainability of results highlight the need for continuous improvement in these areas.
- **Engagement and Participation**: Stakeholder engagement was strong, particularly in workshops where participants actively tested and provided feedback on the tools. However, response rates varied across use cases, reflecting the need for further engagement strategies in remote or asynchronous evaluations.
- **Challenges Identified**: Technical limitations such as incomplete data integration, variability in performance, and usability barriers emerged as areas requiring targeted refinements before the second validation phase.

Building on the outcomes of this first phase, the following steps will guide the project's next phase of implementation and evaluation:

- **Tool Refinement and Optimization**:
- Address identified weaknesses related to data integration, efficiency, and clarity.
- Improve the **user interfaces** to enhance usability and streamline navigation.
- Strengthen **explainability** and transparency features to improve stakeholder trust.
- **Enhanced Engagement and Participation**:
- Increase efforts to engage a broader range of stakeholders, particularly in underrepresented groups and geographies.
- Refine feedback mechanisms to capture higher response rates during the second validation phase.

- **Focus on Trust and Security**:
- Implement robust **data security** and privacy measures to address stakeholder concerns.
- Integrate mechanisms to mitigate bias and validate the accuracy of predictions and outputs.
- **Preparation for the Second Validation Phase**:
- Incorporate lessons learned into the design and organization of the next validation workshops.
- Update evaluation tools to provide deeper insights into the tools' effectiveness and impact.
- **Alignment with Project Goals and Impact**:
- Strengthen alignment with the project's KPIs and the six impact dimensions (political, socioeconomic, organisational, environmental, technological, and legal).
- Further map the project outcomes to the **Sustainable Development Goals (SDGs)** to highlight AI4Gov's contribution to sustainable and inclusive policy development.

By addressing these next steps, the AI4Gov project will ensure that its tools are refined, trusted, and ready for in the second validation phase. The continuous engagement of stakeholders and iterative improvements will further solidify the tools' value and impact, aligning them with the needs of policymakers, citizens, and other end users.

The final version of the evaluation results of the AI4Gov project will be included in D6.5 which is to be delivered at the end of the project, in M36.

5 References

AI Observatory - 2021. AI Observatory. (n.d.). https://eujapan.ijs.si/dashboards/Main/Index?visualization=MAG--country-indicators

Bhat, B. A., & Bhat, G. J. (2019). Formative and summative evaluation techniques for improvement of learning process. *European Journal of Business & Social Sciences, 7*(5), 776-785.

Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring relations between formative and summative assessment. *Transforming assessment: Through an interplay between practice, research and policy*, 53-80.

OECD. (n.d.). https://www.oecd.org/

Prince, D. R. (2015). Approaches to summative evaluation. In *Manual of Curatorship* (pp. 690-701). Routledge.

Redefining civic engagement. Novoville. (n.d.). https://www.novoville.com/

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence, 4 (6), 103-108.*

Schrepp, M. (2015). User experience questionnaire handbook. *All you need to know to apply the UEQ successfully in your project*, 50-52.

Use the power of AI to turn news content into actionable insights. Event Registry. (n.d.). https://eventregistry.org/

6 APPENDIX

6.1 The UEQ – Short version

Thank you for taking the time to test our tools. Your feedback is valuable in helping us refine and optimise our solutions. **Please complete this brief questionnaire to evaluate your experience.** Your insights will directly contribute to the improvement of our technologies.

Instructions

For the assessment of the tool, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Example: attractive □ □ □ □ □ □ □ *unattractive*

This response would mean that you rate the application as more attractive than unattractive.

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression. Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line. It is your personal opinion that counts. Please remember: there is no wrong or right answer!

Please assess the tool by ticking one circle per line.

		1	2	3	4	5	6	7	
1	**Obstructive**								**Supportive**
2	**Complicated**								**Easy**
3	**Inefficient**								**Efficient**
4	**Confusing**								**Clear**
5	**Boring**								**Exciting**
6	**Not interesting**								**Interesting**
7	**Conventional**								**Inventive**
8	**Usual**								**Leading edge**

6.2 Questionnaire on trustworthiness of New Technologies

The purpose of this workshop is to gather feedback on the users' perspective on New Technologies, such as the tool you tested today. Your responses reflect your personal opinion on trust, comfort, and familiarity with such technologies and there are no wrong and right answers.

After receiving this short training and based on your general knowledge and opinion on new technologies used in public governance, please respond to the following questions:

Ref. 9 *How confident are you in the reliability and accuracy of such tools' outputs in order to make decisions based on them?*

[ ] Not confident

[ ] Somewhat confident

[ ] Moderately confident

[ ] Very confident

Ref. 10 *How confident are you that such tools work for your benefit and can make your job easier?*

[ ] Not confident at all

[ ] Slightly confident

[ ] Moderately confident

[ ] Very confident

Ref. 11 *How concerned are you about the privacy and security of the data processed by such tools?*

[ ] Very concerned

[ ] Somewhat concerned

[ ] Moderately concerned

[ ] Not concerned

Ref. 12 *Please provide any additional comments, concerns, or suggestions: (Open answer)*

6.3 Workshop on trustworthiness of New Technologies

The purpose of this focus group is to **assess the trust people have in new technologies** when integrated into the operational systems they use daily. The tool you tested today was one example of a new technology, at an early stage of development. Now, we are asking you to think how you would feel about using such a tool in your work.

Ref. 13 *What are the **pros** and **cons** of such a tool in the following categories: technological advancement, bias, security, and trust:*

	Technological advancements	**Bias**	**Security**	**Trust**
Pros				
Cons				

6.4 AI4Gov Open Day: Participants' feedback on the AI4Gov Use Case

Thank you for taking the time to test our tools. Your feedback is invaluable in helping us refine and optimise our solutions. **Please complete this brief questionnaire separately for each tool you have tested.** Your insights will directly contribute to the improvement of our technologies.

Ref. 14 Please choose the AI4Gov Use Case you tested:

☐ SDG Observatory (JSI)

☐ Top100 projects (JSI)

☐ OECD documents analysis (JSI)

☐ Traffic Violations management (VVV)

☐ Waste management – PAYT (VVV)

☐ Drinking water management (DPB)

Ref. 15 The technology was very easy to navigate and use

☐ Strongly disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly agree ☐ Can't say

Ref. 16 The technology provided useful and relevant insights for the problem at hand

☐ Strongly disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly agree ☐ Can't say

Ref. 17 I would consider further testing this technology in the context of my work/organisation.

☐ Strongly disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly agree ☐ Can't say

Ref. 18 Please provide any additional comments, concerns, or suggestions:

_____

Demographics

Ref. 19 Gender

☐ Male ☐ Female ☐ Nonbinary ☐ Other ☐ Prefer not to say

Ref. 20 Age group

☐ 18-24 ☐ 25-34 ☐ 35-44 ☐ 45-54 ☐ 55-64 ☐ 65+

Ref. 21 Domain

☐ Policymaker ☐ Public Authority ☐ Academia ☐ Industry ☐ Citizen ☐ Other _____

6.5 Results of Qualitative Research- AI4GOV – T6.4 "Trustworthy data-driven tourism policies"

HELLENIC REPUBLIC
MINISTRY OF TOURISM

AI4Gov
Trusted AI for Transparent Public Governance fostering Democratic Values

Results of Qualitative Research- AI4GOV – T6.4 "Trustworthy data-driven tourism policies"

Qualitative research results on:

Visitor characteristics in the Municipality of Vari-Voula-Vouliagmeni and their assessment of the municipality's services concerning the management of public spaces

Ministry of Tourism of Greece, Directorate of Research

(May-July 2024)

1. Research Identity - Methodology

In the framework of the **AI4GOV project-Pilot 3 "Trustworthy data-driven tourism policies"** of the EU-funded Horizon Europe Programme for Research and Innovation, the Ministry of Tourism of Greece is conducting primary research to collect qualitative and quantitative data on tourism flows in the Municipality of Vari-Voula-Vouliagmeni.

The aim of the research is to support the Municipality, by providing the data needed for the development of useful tools with the use of Artificial Intelligence (AI), that will contribute to the improvement of the Municipality's services towards residents and visitors.

During the first part of the primary qualitative research (May-July 2024), a series of semi-structured interviews were conducted with key stakeholders. Key stakeholders were contacted via e-mails.

All stages of the primary research were conducted in-house by the Ministry's Directorate of Research.

The purpose of the research was threefold:

- to identify the characteristics of the visitors in the Municipality of Vari-Voula-Vouliagmeni,
- to understand the preferences and the points of interest of the visitors staying at the municipality's hotels as well as daily visitors,
- to document the assessment of visitors and entrepreneurs on the Municipality's services and the usefulness of the applications used for the management of public spaces.

An Interview Guide was used during the interviews (Annex I), which was adapted to each key stakeholder category. The Interview Guide ensured that during the interviews all topics were covered, and enough flexibility was provided to allow key informants to freely express their views on the subject.

The Interview Guide attempted to gather information on the following topics:

- Customers' profile.
- Characteristics of tourism flows in the Municipality of Vari Voula Vouliagmeni.
- Assessment of municipal services and infrastructure.
- Assessment of Pay As You Throw (PAYT) pilot action and Novoville app.

In total, fourteen (14) interviews were conducted with key stakeholders (an anonymized list of key stakeholders is included in Table1). Efforts were made to ensure that key stakeholders from the following tourism business categories were included:

- Hotels
- Hotel Associations
- Tourist Enterprises
- Travel agencies based and operating in the Municipality of Vari Voula Vouliagmeni

The interviews were conducted in the period between 27/05/2024-18/07/2024 online via zoom and in two cases by telephone. During the research process, the protection of personal data and the conditions of confidentiality were respected.

Table 1: Interviews (period 27/05/2024-18/07/2024)		
Interview date	**Stakeholder**	**Code**
27/5/2024	**Hotel 5*** (Vouliagmeni)	KI01
28/5/2024	**Hotel 4*** (Vouliagmeni)	KI02
28/5/2024	**Hotel 4*** (Vari)	KI03
31/5/2024	Representative of Attica Hotels Association	KI04
5/6/2024	**Hotel 4*** (Voula)	KI05
5/6/2024	**Hotel 2*** (Voula)	KI06
18/6/2024	Tourist Enterprise (Vouliagmeni)	KI07
1/7/2024	**Hotel 3*** (Voula)	KI08
3/7/2024	Travel Agency/ Tourism Services Enterprise (Vari)	KI09
4/7/2024	**Hotel 5*** (Vouliagmeni)	KI10
5/7/2024	**Hotel 5*** (Vouliagmeni)	KI11
17/7/2024	Conference and event organization Enterprise(Voula)	KI12
17/7/2024	Restaurant-Bar-Receptions Enterprise (Vari)	KI13
18/7/2024	**Hotel 5*** (Vouliagmeni)	KI14

A detailed diary of the interviews with the participants' contact details as well as the interviews' transcription in a word file is kept in the Research Directorate's database. Informed consent of the participants was obtained prior to the interviews.

2. Research Results

2.1. Hotel Identity

In the qualitative research participated:

- Four 5* hotels that operate all year [KI01, KI10, KI11, KI14]
- Three 4* hotels, [two seasonal (KI02, KI03) and one, all year (KI05)],
- One 3* hotel that operates all year [KI08],
- One 2* hotel that operates all year [KI06],

It must be clarified that the high number of 4* and 5* hotels that participated in the research is due to the high concentration of such accommodation in the Municipality. More specifically, according to the available data[1], 56% of the hotels operating in the area belong to these categories.

Regarding hotel establishments, data on average occupancy and average room rate were collected. More specifically, the **average occupancy in all accommodation categories is particularly high**. In the case of hotels that operate throughout the year, **the average occupancy during the summer months is significantly higher than during the winter months** [KI05, KI06, KI08]. For example, in a 4* hotel during the period April-October the occupancy exceeds 75% while the rest of the year it ranges between 50-60% [KI05]. High occupancy is also recorded in the case of hotels that operate seasonally [KI01].

The average room rate varies significantly depending on the hotel category.

- In 5* hotels the average room rate amounts to 500€ [KI01, KI10, KI14].
- The average room rate in 4* hotels ranges from 80€ [KI03] or 140€ [KI02, KI03] to 210-220€ [KI05].
- In a 2* hotel the average room rate is significantly lower (60-65€) [KI06].

2.2. Visitor profile

Visitor motivation

During the summer months, the main visitor motivation is **leisure** [KI01, KI02, KI09, KI10, KI11, KI14] and specifically, the ideal weather conditions, the beaches, shopping, relaxation, nightlife and gastronomy. Visiting archeological sites or museums does not constitute the main motivation factor for the majority of visitors [KI14]. As was pointed out by one of the key informers, there are many visitors who do not even visit the center of Athens and choose to stay within the municipality to relax [KI11]. During the summer months, **incentives travel** is recorded [KI13, KI14].

In addition, many visitors participate in **conferences** [KI02, KI12, KI10, KI14] **or corporate events** [KI04, KI12, KI14].

[1] The data on hotels operating in the Municipality of Vari-Voula-Vouliagmeni were obtained from the municipality's website (https://visitvarivoulavouliagmeni.gr/index.php/el/luxury-hotels).

Business is another visitor motivation [KI09, KI11, KI12], while special mention was made to visitors whose main motivation was investing in the Athens **real estate market**. **This category of visitors consists** mostly of younger Greek expats, who choose the winter months and combine a visit to relatives and friends to their investment activities [KI01, KI09]. The number of visitors who combine business and pleasure is high [KI12].

It was also pointed out that many visitors **combine business and leisure** [KI12].

Finally, **wedding tourism** (destination weddings) has recently become a popular trend mostly by visitors from Great Britain, India and the USA. Usually, the wedding ceremony and reception take place in venues in the area and guests stay in hotels in the municipality of Vari Voula Vouliagmeni Hotels [KI02, KI10, KI13, KI14].

Demographic profile

The majority of the visitors are **families** [KI02, KI03, KI04, KI05, KI08, KI10, KI11] and **couples** [KI02, KI10, KI11] and, to a lesser extent, **older visitors** [KI08] as well as **wedding guests** [KI11].

Average duration of stay

According to the key informants, in many cases Attica is the first or last stop (stop over) before or after their visit to islands, other inland destinations or taking a cruise. In those cases, the duration of stay in the municipality of Vari-Voula-Vouliagmeni hotels:

- Usually ranges **from 3-5 days** [KI01, KI02, KI03, KI05, KI06, KI11, KI14] and is an integral part of their vacations [KI05].
- In more rare cases it **does not exceed one day** [KI03, KI05, KI11].

It was pointed out that in recent years **the average duration of stay has been steadily increasing** [KI10]. Moreover, the number of visitors who choose Athens, and particularly the southern suburbs (the so-called "Athenian Riviera") as their only destination (city break) is rising. It is estimated that after the completion of the Elllinikon infrastructure development project, the length of stay will increase by at least one day [KI14].

During the winter months the duration of stay is shorter [KI08].

For business trips the average duration of stay is 3 days, usually followed by a visit to the greek islands [KI12].

Countries of origin

According to the key informants the main countries of origin are the following:

- **USA** [KI01, KI02, KI03, KI06, KI07, KI10, KI11, KI12, KI13, KI14].
- **Western Europe** [KI03, KI07, KI09, KI12] **and more specifically Great Britain** [KI01, KI02, KI03, KI06, KI10, KI11, KI12, KI13, KI14], **France** and **Germany** [KI05, KI06, KI10, KI14].
- **South Europe** and more specifically **Italy** [KI05, KI14] and **Spain** [KI14].
- **Scandinavian countries** [KI05, KI10, KI13, KI14].
- **Eastern Europe** [KI03] **and Balkan countries** [KI02].
- **Arab countries** [KI05, KI11, KI13] and in particular **Gulf countries** [KI01, KI03, KI05, KI06, KI08, KI10, KI14].

- **Israel** [KI01, KI13, KI14].
- **India** [KI02, KI13].
- **Cyprus** [KI02].
- **African countries** [KI02].
- **New Zealand** [KI02].
- In the past, Russia was an important market, but these flows have decreased [KI01, KI07].
- **Greeks** who reside abroad and **expatriates from the USA, Australia and South Africa** especially during the winter months [KI01, KI02, KI05, KI07, KI10, KI11, KI13].

The **number of repeat visitors is high** [KI05, KI06, KI09, KI10, KI13].

Distribution channels

The majority of the visitors are independent travelers [KI01, KI03] who receive accommodation information through the following channels:

- Booking platforms such as Expedia, Booking, Hotelbeds [KI02, KI06, KI08, KI10, KI14]
- The hotel's website, especially if it has a strong brand name [KI06, KI08, KI10, KI14].

It was pointed out that positive reviews from friends and relatives ("word of mouth") plays a significant role [KI05, KI08, KI10].

In one case, the hotel cooperates with specific travel agencies that specialize in specializes forms of tourism such as **wedding tourism and conferences** [ΠΚ02].

Means of transport

The visitors use during their stay the following means of transport:

- **Taxi (uber & conventional)** [KI01, KI02, KI06, KI07, KI09, KI10, KI11, KI12, KI13].
- **Rent a car** [KI02, KI03, KI06, KI07, KI08, KI09, KI10, KI12].
- **Public Transport** (mostly tram and buses) [KI02, KI06, KI07, KI08, KI12].
- **VIP transfer services** [KI09, KI11, KI14].
- **In rare cases their own car** [KI03].

The biggest tourist flows (visitors staying at the hotels or daily visitors) are recorded on the following routes:

- From the municipality of Vari-Voula-Vouliagmeni to Glyfada [KI02, KI05, KI09, KI11, KI14].
- From the municipality of Vari-Voula-Vouliagmeni to the center of Athens [KI01, KI05, KI08, KI09, KI10, KI12, KI14].
- Inside the municipality, to popular points of interest (Vouliagmeni Lake, beaches, restaurants) [KI01, KI02, KI03, KI05, KI06, KI07, KI08, KI09, KI10, KI11, KI12] and the Temple of Poseidon in Sounion (Municipality of Lavrion).

2.3. Activities– Points of interest

Activities

The most popular activities in the municipality of Vari-Voula-Vouliagmeni **are the following**:

- Visit to Lake Vouliagmeni [KI01, KI02, KI06, KI07, KI08, KI09, KI10, KI11, KI12].
- Visit to the temple of Poseidon in Sounio [KI01, KI02, KI03, KI05, KI06, KI09, KI10, KI12].
- **Walking routes** within the municipality (Faskomilia'a path) [KI01].
- **Walks along the municipality's coastal zone** [KI02, KI14].
- **Bike or scooter rides** [KI02, KI14].
- **Open-air cinema** [KI02].
- **Agritourism activities**: a specific hotel offers its clients the possibility to visit a farm it operates within the municipality. It was reported that this is a very popular activity [KI10].

The most popular activities outside the municipality of Vari-Voula-Vouliagmeni **are the following**:

- **Visit to the center of Athens** [KI01, KI05, KI08, KI09, KI10, KI12, KI14].
- **Visit to the Acropolis** [KI01, KI02, KI12] and the **Acropolis Museum** [KI01, KI02].
- **Shopping and other activities** in the center of Athens and Glyfada [KI02, KI05, KI09, KI11, KI14].
- **Day trip to the Argolic and Saronic Gulf islands** [KI02], **like Poros** [KI01], usually with rental yachts [KI10, KI12].
- **Day trip to nearby destinations**, such as **Nafplio** [KI05], and **Saronida** [KI03]
- **Visit to points of interest in the coastal zone**, such as Glyfada [KI06] and Alimos Marina [KI06].
- **Nightlife** [KI14].

Most of the visitors have knowledge of the available activities in the area. They are mostly informed via the internet and social media and ask the hotels' employees to confirm that the information is correct or book an activity [KI01, KI05].

Gastronomy

The Municipality of Vari-Voula-Vouliagmeni is an internationally renowned gastronomic destination thanks to the high-quality restaurants that operate in the area [KI01, KI10].

In many cases, the first choice is the traditional Greek tavern so they can taste traditional Greek cuisine (i.e Loizidis tavern in Vouliagmeni, Zachos taverna, Lambros taverna) [KI01, KI02, KI06, KI08]. Many high-end visitors are attracted to restaurants with modern cuisine, such as "BeefBar" in Astir Palace Hotel, the "Ithaki Restaurant & Bar" [KI01], the "Abra Ovata" restaurant in Lake Vouliagmeni [KI02], the restaurants in the Four Seasons Hotels and Resorts [KI05] and the One and Only hotel in Glyfada [KI11], the "Barbarossa Athens" [KI11], "Nobu", in Vouliagmeni [KI14]. Finally, restaurants that have been awarded Michelin Stars such as "Pelagos" in Four Seasons Astir Palace Hotel or "Patio Dining" [KI10] are on the top of the visitors' list together with well-known restaurants in downtown Athens and Piraeus [KI14].

It was pointed out, however, that the choice of restaurants depends on the financial status of each customer [KI05, KI06].

Beaches

The municipality's beaches are **particularly popular** among visitors. For the majority of the visitors **most beaches in the area are considered clean with high quality waters,** due to the fact that most have been awarded Blue Flags [KI01, KI02, KI04, KI11]. It should be noted that, in some cases, parts of the municipality's beaches are leased by neighboring hotels which are responsible for cleaning them [KI14]. Only in one case was it reported that the problem with beach litter has deteriorated lately [KI14].

In addition, reference was made to the need to take the appropriate actions for waste collection both in the coastal zone as well as underwater [KI04]. The need to install jellyfish protection nets was also highlighted [KI01].

Regarding the **entrance fees at the municipality's private beaches**, minor complaints were recorded [KI02, KI03, KI06]. High-income visitors who stay in 4* & 5* hotels are satisfied with the cost as **the municipality offers many options at different price levels as well as numerus free beaches** [KI03, KI10].

The most popular beaches are Vouliagmeni beach, Krabo Beach (Kavouri), Astir and Varkiza beach, as well as the beaches in Lemos Vouliagmeni [KI01, KI03, KI10]. High-income visitors choose the beaches at luxury hotels such as the Four Seasons, Grecotel and Lake Vouliagmeni [KI10].

Particular mention should be made of **Lake of Vouliagmeni**, a landmark of the area with high visitor flows especially during the summer months. It attracts a large number of tourists and, since it is a controlled environment with permanent lifeguard coverage, it is chosen by families as well as couples and groups. In addition, it is popular among the elderly due to the healing properties of its waters [KI07].

Port infrastructure – Marinas

Astir Marina in Vouliagmeni is the only maritime tourism infrastructure in the Municipality of Vari-Voula-Vouliagmeni [KI01, KI02]. After the **recent upgrade of its infrastructure** and the creation of luxury retail stores and food services and facilities, it is expected that tourist flows to the Municipality as well as the promotion of luxury tourism [KI04, KI10, KI14]. The need to upgrade the road network infrastructure in the area as well as the creation of parking spaces was highlighted [KI14].

However, emphasis was put on the **urgent need for the creation of further high-quality port infrastructure** with the aim to**:**

- Address the large and steadily growing demand [KI04],
- Promote yachting [KI01]

According to the key informants, **Astir Marina in Vouliagmeni is a point of interest that attracts high income visitors** [KI02, KI04, KI06], while at the same time, the interest from lower income visitors is low [KI10].

Finally, it was mentioned that in Voula and Vouliagmeni there are rudimentary port infrastructures in the respective nautical clubs [KI01, KI02].

Athens and the Athenian Riviera

Athens and the **Athenian Riviera** in particular emerge as a distinct and complete destination in which the high-demanding visitor can participate in a multitude of activities [KI01]. The promotion of the Athenian Riviera should be a priority and should be fully integrated into the tourism product of Athens, the only coastal European capital. However, the need for the development of a large convention center in the Southern Suburbs of Attica will contribute to strengthening the tourism potential of the area [KI04].

2.3. Challenges – Assessment of municipal services

Parking

The lack of parking spaces is considered one of the main challenges, especially during peak hours and the tourist season due to increased flows towards the municipality [KI01, KI02, KI03, KI04, KI10, KI14]. This problem is particularly acute in tourist areas like Vouliagmeni and Kavouri [KI14]. Another aspect of the above challenge is the tendency of many drivers to park their cars in parking spaces reserved for people with disabilities, on the pavement or on crosswalks, which restricts the movement of pedestrians [KI11].

It's crucial that the competent authorities find effective solutions and formulate long-term planning [KI02, KI04].

Finally, some of the key informants stated that they do not face particular issues with parking because their businesses have private parking space [KI05, KI06, KI08, KI11, KI13, KI14].

Waste Management

Most of the key informants expressed satisfaction with the municipality's cleaning services (in terms of frequency and programming of waste collection hours) and the prompt response of the municipal authorities when collection issues arise [KI02, KI05, KI07, KI08, KI11].

However, a key informant expressed complaints about the placement of numerous bins and the existence of an illegal disposal garbage site near his business, as well as noise pollution from waste collection vehicles [KI06].

In addition, two key informants stated that the quality of cleaning services has been degraded recently due to the lack of personnel and the unsatisfactory frequency of waste collection [KI09]. Especially during the weekends many problems arise due to the increased flows of tourists that result to an increased volume of waste [KI13].

Finally, it was pointed out that the use of Artificial Intelligence (AI) would be particularly useful to determine the necessity and timing of garbage collection in order to avoid peak hours, not to cause traffic problems, etc. [KI14].

Assessment of Pay As You Throw (PAYT) pilot action

The majority of the participants stated they were unaware of the pilot action. They evaluate the initiative as successful and stated that they would participate, when implemented on a wider scale [KI02, KI05, KI06, KI07, KI08, KI09, KI11, KI12, KI14].

Only one of the key informants mentioned that their company is participating in the pilot program and assessed the initiative as successful [KI01]. Three participants stated that they are aware of the pilot program, but their business does not participate in it [KI03, KI10, KI13].

Novoville app assessment

Most of the participants are not aware of the novoville app [KI03, KI05, KI06, KI08, KI09, KI11, KI12, KI13, KI14]. Three of the key informants stated that in case of emergency they prefer to contact the municipality over the phone [KI05, KI06, KI14] and only two of them stated that they know and use the application [KI02, KI10]. They think that the visitors are not aware of the of the app and therefore they do not use it.

Street racing – road safety- noise pollution

The **illegal street racing** organized in the area, the frequent **speed limit violations** as well as the general **drivers' indifference to pedestrian safety** constitute one of the most pressing problems [KI04, KI05, KI08, KI09, KI10, KI11, KI12, KI13] and **cause insecurity to both residents and visitors** by putting their physical safety in danger [KI01, KI02, KI04, KI07].

The above-mentioned issues are directly related to the **noise pollution problem** which affects businesses located near major avenues [KI01, KI06, KI08, KI10].

To address the above issues, the following measures, for which the central government and not the municipality are competent, were suggested:

- Increased policing and placing cameras [KI01, KI04, KI09, KI12, KI14].
- Installation of metal or concrete safety barriers on sidewalks for pedestrian safety [KI07].
- Creation of underpasses [KI04].

Visitor's assessment of the municipality as a tourist destination

The overall evaluation of the Municipality of Vari-Voula-Vouliagmeni as a tourist destination is **positive** [KI02, KI03, KI05], however, a few complaints are listed as follows:

- **The value for money of the services** offered in the area is not good and many visitors complain about the high prices [KI07, KI09, KI12],
- **Complains were made on** the high taxi cost [KI03, KI05] and the lack of Public Transportation during the night hours [KI12].
- Complaints were expressed regarding the fact that the Municipality is not **baby friendly** [KI05].

Finally, another important issue mentioned is the different development level among the three municipal units, with Vari being considered neglected in relation to the other two municipal units and in need of interventions [KI03, KI04, KI09].

Priorities-Goals

The goal should be the promotion of **mild sustainable development** that does not affect the environment and respects the needs and perspective of the local community. Moreover, visitors themselves, during their stay, prioritize sustainability, select sustainable experiences and come in contact with the local population [KI04].

At the same time, emphasis should be placed on the utilization and absorption of funds from available European and National financial tools for the further development of the region in order to address existing and future challenges [KI04].

Finally, **promotion and marketing actions** should be carried out by the Municipality with an emphasis on important points of interest such as Lake Vouliagmeni. It was suggested that special events such as "Lake Vouliagmeni Month" or "Vouliagmeni Beach Month" could be organized [KI07].

ANNEX I

A. Identity(hotel/enterprise/association)

B. Customer's profile

Could you describe in a few words your customers' profile?

- Demographic profile: age, sex
- When they travel (winter / summer)
- Motivation of travel
- Duration of Stay
- Activities chosen at destination
- Concerns regarding destination (i.e. safety)
- Places to visit
- How do they travel? (solo travelers, couples, with their family etc.)
- Distribution channels
- Main sources of information

C. Characteristics of tourist flows in the Municipality

What are the main categories of visitors to the municipality? (e.g. foreign tourists staying in hotels, day visitors from Athens, yachts, etc.)

How would you rank them in order of size? (from most populous to least populous category)

Where are these flows moving? (points of interest)

By what means do they travel? (rent a car, public transportation, taxi, etc.)

Have you identified spots that attract an excessive concentration of tourist flows? What season? What times of the day?

D. Public services offered at the municipality

How do you personally evaluate the services offered by the municipality (cleaning, parking, traffic management, beach, etc.)?

Where do you detect problems?

Regarding parking, how do they rate the operation of controlled parking areas (if there are any)?

Application Novoville: How do you rate its effectiveness? Is it used by guests?

Cleaning - Pay As You Throw (PAYT): Are they aware of this pilot program? if they participate in the pilot? How do they evaluate effectiveness?

Do you think that your customers are satisfied with the level of service?

What are the key factors influencing their satisfaction?

For which services are complaints recorded?

Comment on marina infrastructure

6.6 Questionnaire Forms for the Quantitative and Qualitative research on tourism flows

6.6.1 Questionnaire for VVV employee

HELLENIC REPUBLIC

MINISTRY OF TOURISM

AI4Gov

Trusted AI for Transparent Public Governance fostering Democratic Values

The purpose of this research is to investigate the attitude of the employees of the Municipality of Vari Voula Voulagmeni towards AI and to what extent do employees feel AI can contribute to improve municipal services by developing smart applications.

This research is part of the HORIZON project AI4GOV, where the Ministry of Tourism of Greece and the Municipality of Vari Voula Vouliagmeni participate as partners. The project aims to put forward the potential of AI and Big Data technology for the public sector and enhance their use so that public services respond to citizens' needs in a more effective way.

Your responses will only be reported in aggregate form and will solely be used for research purposes (Law 3832/2010).

Since when have you been working at the Municipality?	

In which Municipality Department do you work?

……

To what extent do you agree or disagree with the following statements?

The arrival of tourists/visitors negatively affects:	Very	Enough	Little	Not at all	I don't know/ I don't answer
1 Waste Management					
2. Traffic congestion					
3. Parking					
4. Recycling					
5. Beaches					

Do you know what Artificial Intelligence is?	Yes	No
If so, do you trust it?	Yes	No

The application of AI in the Municipality of VVV, through the development of smart tools, could contribute towards.	Fully disagree	Disagree	Neither Disagree Nor agree	Agree	Fully agree
Improving waste management					
Improving Recycling services					
Improving traffic regulation/parking					
Reducing bureaucracy and delays in public procedures.					
improving safety in public spaces (e.g. AI cameras, automatic hazard notification).					
Enhancing decision-making based on the available data.					
The application of AI in public services should be accompanied by strong measures to protect personal data.					

Sex: □ Man □ Woman □ Other □ Prefer not to answer

Education Level: □ Secondary □ Post-secondary □ Higher □ Postgraduate □ Doctoral Diploma

Age: □ 18-25 □ 26-45 □ 45-55 □ 56-65 □ 66 plus

THANK YOU FOR YOUR TIME AND PATIENCE

6.6.2 Questionnaire for the Visitors

GREEK REPUBLIC

MINISTRY OF TOURISM

The purpose of the research is to explore the attitude of tourists/visitors towards municipal (Municipality of Vari Voula Vouliagmeni) services and artificial intelligence (AI) and to what extent AI can contribute to the improvement of municipal services in the area through the development of smart applications.

This research is part of the HORIZON project AI4GOV, where the Ministry of Tourism of Greece and the Municipality of Vari Voula Vouliagmeni participate as partners. The project aims to put forward the potential of AI and Big Data technology for the public sector and enhance their use so that public services respond to citizens' needs in a more effective way.

Your **responses will only** be reported in aggregate form and **will** solely be **used** for **research purposes**.

Is this your first visit to the area (Vari-Voula-Vouliagmeni)?	Yes	No	If NO. How many times have you visited the municipality in the past?	Number of visits

What is the purpose of your visit?

- Holidays/Leisure
- Business
- Investment in real estate market
- Other (specify)_____
- _____

How did you get to Athens?

- Directly from my country of residence by flight to Athens's airport.
- From a country other than my country of residence, as part of a multi-destination trip
- By boat, to Athens, from another Greek destination.
- Other (specify)_____
- _____ _____

How do you travel?

- Alone
- With family
- With friends
- As a couple
- With colleagues for business purposes

Other

Which of the following places of interest in the area have you visited or would you like to visit?

1. Lake of Vouliagmeni
2. Beaches in Vouliagmeni/Voula/Varkiza
3. The esplanade
4. Restaurants/Cafes
5. Astir Marina Vouliagmeni's
6. Other ……

Please rate your level of expectation VS your actual experience on the below statements using a 5-point scale where 5= Much better than expected and 1 = Much worse than expected ":

	Much better than expected	Better than expected	As expected	Worse than expected	Much worse than expected
1. Accommodation					
2. Variety of available activities					
3. Local cuisine / Gastronomy					
4. Shopping					
5. Nightlife					

6. Nature					
7. Transportation					
8. Overall quality of services					

How likely is it to visit Greece for future vacations?

1. Very unlikely
2. Unlikely
3. Neutral
4. Likely
5. Very likely

How likely is it for you to recommend Greece as destination to your friends/ acquaintances?

1. Very unlikely
2. Unlikely
3. Neutral
4. Likely
5. Very likely

To what extent do you agree or disagree with the following statements?

How do you evaluate the services offered by the Municipality, specifically:	Very good	Good	Fair	Poor	Very Poor	I don't know/ I don't answer
1. Waste Management						
2. Traffic Regulations						
3. Parking						
4. Recycling						
5. Beaches						
6. Roadworks (street surface, pavements)						
9. Accessibility (people with disabilities)						

Do you know what Artificial Intelligence is?	Yes	No
If so, do you trust it?	Yes	No

The application of AI in the area (Vari Voula Vouliagmeni) by developing smart tools, could contribute towards:	Fully disagree	Disagree	Neither Disagree Nor agree	Agree	Fully agree
Improving waste management					
Improving Recycling					
Improving traffic regulations/parking					
Reducing bureaucracy and delays in public procedures.					
Improving safety in public spaces (eg AI cameras, automatic hazard notification).					
Enhancing decision-making based on the available data.					
The application of AI in public services should be accompanied by strong measures to protect personal data.					

Sex: □ Male □ Female □ Other □ Prefer not to answer

Education Level: □ Secondary □ Post-secondary □ Higher □ Postgraduate □ Doctoral Diploma

Profession: ……

Age: : □ 18-25 □ 26-45 □ 45-55 □ 56-65 □ 66 plus

Country of residence: ……..

THANK YOU SO MUCH FOR YOUR TIME AND PATIENCE

6.6.3 Questionnaire for Permanent Residents

HELLENIC REPUBLIC

MINISTRY OF TOURISM

AI4Gov

Trusted AI for Transparent Public Governance fostering Democratic Values

The purpose of this research is to investigate the attitude of the permanent residents of the Municipality of Vari Voula Voulagmeni towards AI as well as towards the municipality's services. Also, to what extent do residents feel AI can contribute to improve municipal services by developing smart applications. This research is part of the HORIZON project AI4GOV, where the Ministry of Tourism of Greece and the Municipality of Vari Voula Vouliagmeni participate as partners. The project aims to put forward the potential of AI and Big Data technology for the public sector and enhance their use so that public services respond to citizens' needs in a more effective way. Your responses will only be reported in aggregate form and will solely be used for research purposes.

Before you answer our questionnaire, I would like to ask you:

Are you a permanent resident of the Municipality of Vari-Voula-Vouliagmeni?	Yes	No	If YES. Since what year have you lived in the municipality?	Year

Which of the following places of interest at Vari Voula Vouliagmeni do you like to visit?

1. The Lake of Vouliagmeni
2. Beaches open to the public
3. The esplanade
4. Restaurants/Cafes
5. Astir Marina Vouliagmenis
6. Other

……

To what extent do you agree or disagree with the following statements?

The arrival of tourists/visitors negatively affects:	Very	Enough	Little	Not at all	I don't know/ I don't answer
1 Waste Management					
2. Traffic congestion					
3. Parking					
4. Recycling					
5. Beaches					

Concerning municipal services, to what extent do you agree or disagree with the following statements?

How do you evaluate the services offered by the Municipality, specifically:	Very good	Good	Fair	Poor	Very Poor	I don't know/ I don't answer
1. Waste Management						
2. Traffic Regulations						
3. Parking						
4. Recycling						
5.Novoville application						
6. Beaches						
7. Roadworks (street surface, pavements)						
8. Accessibility (people with disabilities)						

Do you know what Artificial Intelligence is?	Yes	No
If so, do you trust it?	Yes	No

The application of AI in the Municipality of VVV, through the development of smart tools, could contribute towards.	Fully disagree	Disagree	Neither Disagree Nor agree	Agree	Fully agree
Improving waste management					
Improving Recycling services					
Improving traffic regulation/parking					
Reducing bureaucracy and delays in public procedures.					
improving safety in public spaces (e.g. AI cameras, automatic hazard notification).					
Enhancing decision-making based on the available data.					
The application of AI in public services should be accompanied by strong measures to protect personal data.					

Sex: □ Man □ Woman □ Other □ Prefer not to answer

Education Level: □ Secondary □ Post-secondary □ Higher □ Postgraduate □ Doctoral Diploma

Profession:……..

Age: □ 18-25 □ 26-45 □ 45-55 □ 56-65 □ 66 plus

THANK YOU SO MUCH FOR YOUR TIME AND PATIENCE

6.7 Top100 reviewers bias questions evaluation

Dear Top100 reviewers,

Our previous Top 100 reports highlighted a lack of substantive awareness and consideration for ethical criteria and in particular bias analysis in AI solutions. We also noted that reviewers emphasized the need for stronger attention toethics and potential biases in the proposed projects.

To address these gaps, our team has developed four additional questions for Top 100 applicants. We asked them to outline the ethical considerations of their AI solutions, more information about their inclusiveness and fairness efforts, and explain how they address potential biases in their data and models. Now, we would love to hear your feedback on it.

Therefore we are inviting you to assess the bias/ethics questions for Top100 applicants.

Top100 reviewers bias questions evaluation

Q1 -

Here are the new four bias/ethics questions for Top100 applicants, to be evaluated.

1. Ethical considerations and implications of the AI solution, both long and short term.

Is the application of the AI technologies ethical and equitable? Does it take consideration of potential impacts to Human Rights? Is it lawful, complying with all applicable laws and regulations?

2. Inclusiveness and fairness of the AI solution

To what extent does the initiative ensure that the AI solution does not create discriminatory or unjust impacts for different demographic and geographic groups?

3. Addresing possible bias in your data

How has your project systematically considered and addressed potential biases in the data used, irrespective of whether it was collected or obtained from external sources?

Please provide insights into the measures taken to analyze and ensure inclusivity across diverse demographic groups, outlining specific strategies implemented to identify and mitigate biases in the data.

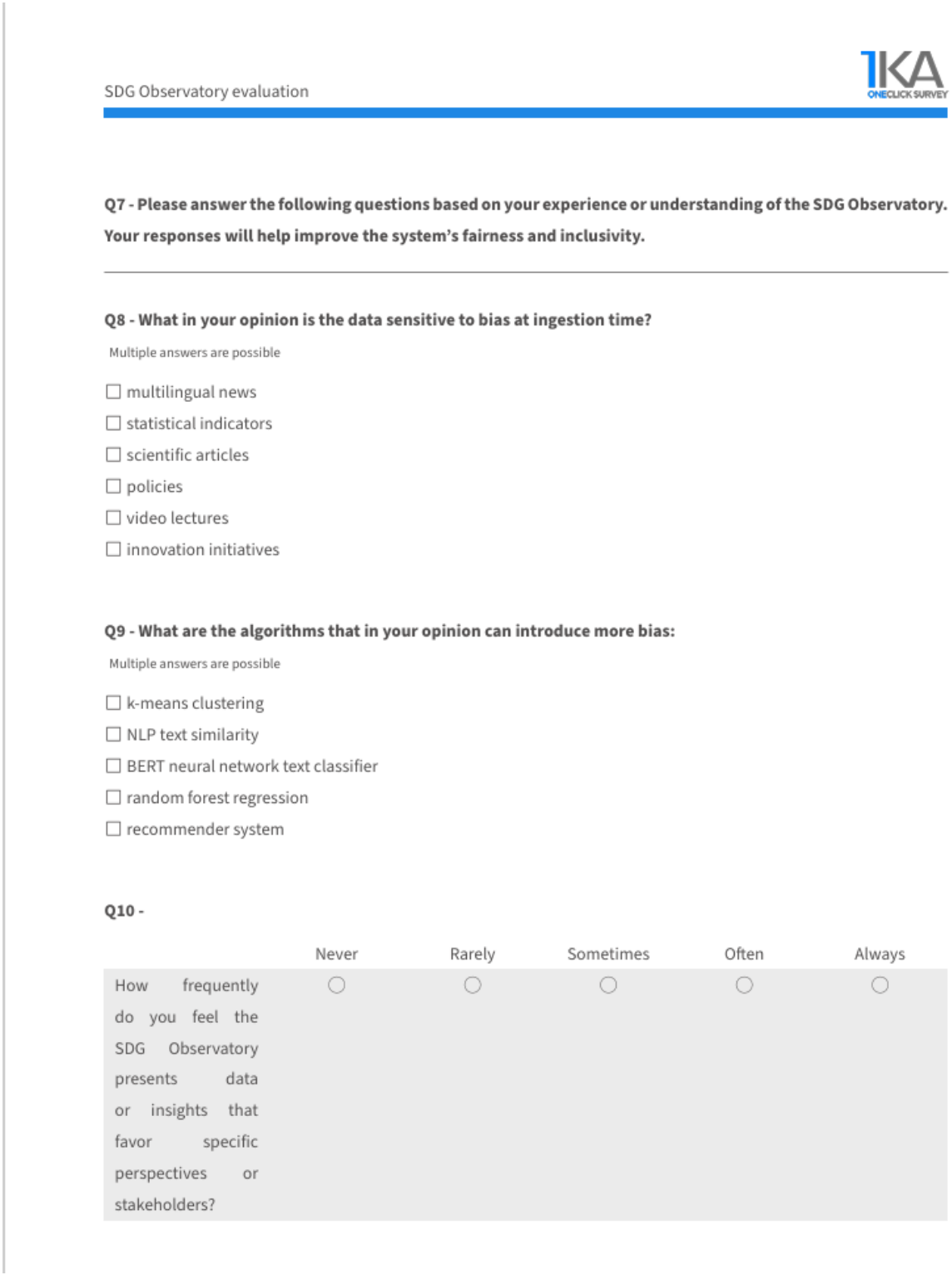
This information will help evaluate the robustness of your approach in promoting equity and minimizing biases in the context of your AI solution.

4. Addressing possible bias in your models

In the process of training and evaluating your AI models - how did your project systematically account for and verify potential biases? Please elaborate on the specific steps taken to assess bias throughout the model development lifecycle, highlighting any strategies implemented to ensure fairness, transparency, and equitable outcomes.

This information will help evaluate the effectiveness of your approach in addressing bias during the critical phases of AI model training and evaluation.

6.8 SDG Observatory Questionnaire on fairness and inclusivity



SDG Observatory evaluation

Q7 - Please answer the following questions based on your experience or understanding of the SDG Observatory. Your responses will help improve the system's fairness and inclusivity.

Q8 - What in your opinion is the data sensitive to bias at ingestion time?

Multiple answers are possible

- ☐ multilingual news
- ☐ statistical indicators
- ☐ scientific articles
- ☐ policies
- ☐ video lectures
- ☐ innovation initiatives

Q9 - What are the algorithms that in your opinion can introduce more bias:

Multiple answers are possible

- ☐ k-means clustering
- ☐ NLP text similarity
- ☐ BERT neural network text classifier
- ☐ random forest regression
- ☐ recommender system

Q10 -

	Never	Rarely	Sometimes	Often	Always
How frequently do you feel the SDG Observatory presents data or insights that favor specific perspectives or stakeholders?	○	○	○	○	○

Q11 - Have you encountered any examples where the SDG Observatory ' s outputs seemed to misrepresent or exclude underrepresented communities or perspectives?

◯ Yes (please specify)

◯ No

◯ Not sure

Q12 - What improvements would you suggest to reduce bias in the SDG Observatory?

Q13 - Do you have any other comments or suggestions regarding the SDG Observatory?