Deliverable 2.2
AI4Gov Holistic Regulatory
Framework V2

03-07-2024

Version 1.0

PROPERTIES	
Dissemination level	Public
Version	1.0
Status	Final
Beneficiary	
License	

AUTHORS		
	Name	**Organisation**
Document leader	Erilda Goga	ViLabs
Participants	Danai Kyrkou	ViLabs
	Prof.Theodore Chadjipadelis, Georgia Panagiotidou, PhD, Christos Gogos, Georgios Chatzipantelis, Alexandros Tzallas, Andreas Miltiadous	AUTH
Reviewers	Spiros Borotis	MAG
	Septimiu NECHIFOR	SIE
	George Manias	UPRC

VERSION HISTORY				
Version	Date	Author	Organisation	Description
0.1	31.05.2024	Erilda Goga, Danai Kyrkou	ViLabs	ToC
0.2	21.06.2024	Erilda Goga, Danai Kyrkou Prof.Theodore Chadjipadelis, Georgia Panagiotidou, PhD, Christos Gogos, Georgios Chatzipantelis, Alexandros Tzallas	ViLabs, AUTH	First Draft released
0.3	27.06.2024	Spiros Borotis, Septimiu Nechifor, George Manias	MAG, SIE, UPRC	Internal reviewing process and updates
1.0	03.07.2024	Erilda Goga, Danai Kyrkou Prof.Theodore Chadjipadelis, Georgia Panagiotidou, PhD, Christos Gogos, Georgios Chatzipantelis, Alexandros Tzallas, Andreas Miltiadous, Spiros Borotis, Septimiu Nechifor, George Manias	ViLabs, AUTH, MAG, SIE, UPRC	Final contributions and review for submission

Table of Contents

List of figures

List of Tables

Abbreviations

Abbreviation	Description
HRF	Holistic Regulatory Framework
AI	Artificial Intelligence
ML	Machine Learning
BIE	Blockchain-based Information Exchange
DGF	Data Governance Framework
XAI	Explainable AI
GDPR	General Data Protection Regulation
XAI Library	eXplainable AI Library
GBV	Gender-based victims
CA	Conjoint Analysis

Abstract

This document presents the final version of the Holistic Regulatory Framework (HRF), a comprehensive framework for regulating AI-based democracy. The report provides an in-depth definition and description of the HRF, outlining its objectives, goals, and the technical methodologies and specifications employed. In its theoretical framework, it explores fundamental definitions of bias and discrimination within public services, supported by an extensive literature review. Following up on the field work conducted in T2.1 and T2.2 and the focus groups with underrepresented groups, a full description of the study methodology, design, focus group structure, sampling techniques, and data collection is included. At the end, the document provides a list of guidelines and suggestions for developing inclusive and bias-free AI tools, with technical recommendations and strategies for inclusivity, to promote fairness in AI-driven democratic processes.

1 Introduction

1.1 Purpose and scope

This deliverable of Work Package 2 (WP2) of the AI4Gov project aims to define the final design of the Holistic Regulatory Framework (HRF). The purpose of this document is to identify the theoretical framework of the HRF, based on findings from the literature review and the field research conducted in T2.1 and T2.2, with a special focus on underrepresented groups. The aim is to include their perspective in the HRF and minimise as much as possible potential biases and discrimination at the source. The field work complementing the literature review and research, fed the HRF by implementing a bottom-up approach. Additionally, the deliverable provides a technical description of the HRF, along with guidelines and suggestions developed throughout the framework's creation.

1.2 Document structure

The deliverable is structured into five chapters. It begins with **Chapter 1**, the introduction, which outlines the purpose and scope of the deliverable and highlights its structure. **Chapter 2** defines the objectives of the HRF and provides a technical description and development of the HRF. **Chapter 3** presents the theoretical framework of HRF, derived from the results of participatory activities conducted during the project, along with a literature review on bias and discrimination in public services. **Chapter 4** offers technical and sociological suggestions and guidelines for developing an AI tool that adheres to EU values and human rights. The report concludes with **Chapter 5**, summarizing the key findings and insights. The References section lists the sources used for D2.2. **Annex 1** includes guidelines and the list of scenarios shared with the interested organizations during the initial phase of organizing the AI4Gov focus groups, while **Annex 2** contains the full questionnaire provided to focus groups participants.

1.3 Updates with respect to previous version

Deliverable 2.2 is a continuation of Deliverable 2.1, which laid the groundwork for addressing bias and discrimination, ensuring compliance with EU regulations, and designing the Holistic Regulatory Framework (HRF). Building on the methodology established in D2.1, D2.2 reports on a study conducted to analyse the perception of bias and discrimination faced by underrepresented groups during interactions with public services. Furthermore, this deliverable defines the HRF by outlining its purpose and objectives, along with the technical framework and specifications for its development. The overall development of the HRF will facilitate the practical application of AI4Gov technologies.

1.4 Target Audience of the Deliverable

The target audience of the deliverable in Work Package 2 (WP2) includes various stakeholders involved in the AI4Gov project and those interested in the intersection of AI, governance, and fundamental rights, including:

1. <u>Project Consortium Members</u>: The deliverable is primarily intended for the project consortium members involved in WP2 and other related work packages (mainly WP3, WP4 and WP5). The document offers detailed insights related to the qualitative analysis and the regulatory framework, in order to support the technical partners in the development of the AI4Gov platform and tools.
2. <u>Project Stakeholders</u>: Other stakeholders involved in the AI4Gov project, such as external advisors, experts, and policymakers, may be interested in this deliverable since it provides valuable information on the development of a regulatory framework (HRF).
3. <u>Researchers and Academia</u>: The deliverable can be relevant to researchers and academics working in the fields of AI, governance, ethics, and fundamental rights. The included qualitative and quantitative analysis methods and regulatory frameworks, offer potential areas for further research and study.
4. <u>Policy and Decision-Makers</u>: Policy Makers and Decision Makers: Policy Makers and decision makers involved in governance, ethics, and AI regulation may be interested in the outcome of this deliverable to incorporate ethical, non-biased and transparent AI systems in public services and to regulate the use of AI technologies in governance. This knowledge can help policy makers to connect the dots among AI, Bias, and policy.

2 AI-Based Democracy Holistic Regulatory Framework (HRF) Definition and Description.

2.1 Overview

This chapter presents the AI-Based Democracy Holistic Regulatory Framework (HRF) of the AI4Gov project, focusing on its concrete definition, objectives and goals, aiming to create a regulatory framework that supports lawful, ethical, inclusive, and transparent AI governance. Furthermore, this chapter goes beyond the theoretical conceptualization of HRF as it presents the technical description of HRF, prescribing 15 dimensions related to AI governance when AI is involved to a lesser or to a greater extend at services that are provided to the public by governmental entities.

2.2 Definition of HRF

The Holistic Regulatory Framework (HRF) developed under the AI4Gov project is a comprehensive framework designed to regulate the use of AI technologies in governance. Its primary purpose is to ensure that AI systems used in public services are developed, deployed, and operated in a manner that is ethical, transparent, and compliant with applicable laws and regulations such as the General Data Protection Regulation (GDPR) addressing in parallel the issues of bias and discrimination. The HRF also incorporates ethical recommendations from various AI bodies, including the High-Level Expert Group on Artificial Intelligence (HLEG) and aligns with regulations such as the General Data Protection Regulation (GDPR) and the AI Act.

The HRF is structured around several key dimensions that guide the governance of AI systems. These dimensions include regulations compliance, fairness, non-discrimination and inclusivity, human oversight, human-centered design, privacy, explainability and transparency, auditing and continuous monitoring, accountability, safety and security, claims and redress, public engagement, awareness, societal benefit and well-being, and sustainability. Each dimension addresses specific aspects of AI governance to ensure that AI technologies are used responsibly and equitably.

To achieve the key dimensions and finalize the HRF, a multi-level mixed-methodology approach was employed. This comprehensive process included literature reviews of relevant bibliography and legislation, Delphi rounds with experts, SWOT analysis, surveys, focus groups, and interviews. By integrating these diverse methodologies, the HRF bridges insights from literature, expert input, and perspectives from citizens with various backgrounds. Additionally, findings from other deliverables and tasks of the AI4Gov project, such as T1.4, T1.5, T2.3, T3.2, T5.1, T5.2, and T6.1, were extensively utilized in the study of the HRF. This approach ensures that the HRF is robust, inclusive, and grounded in a wide range of research and practical insights. Exploiting other tasks from the same project highlights the interconnectivity of the outcomes and the cohesive binding of the project's objectives.

2.3 Objective and Goals of HRF

The AI4Gov Holistic Regulatory Framework (HRF) aims to establish a **fair, transparent, and inclusive regulatory environment** for public organizations ensuring they serve the public good while safeguarding individual rights and societal values. More specifically, HRF's core purpose is to provide a holistic approach to AI governance which ensures the protection of citizens from potential misuse and biases associated with AI technologies in public services, and public decision-making as well as discrimination.

Its key objectives and goals are the following:

- **Ensure Regulatory Compliance**: Guarantee that AI systems adhere to all relevant laws, standards, and ethical guidelines.
- **Promote Fairness and Equity**: Design AI systems to treat all individuals equitably and reduce potential biases.
- **Foster Non-Discrimination and Inclusivity**: Prevent discrimination against any individual or group and promote inclusivity in AI development and deployment.
- **Implement Human Oversight and Control**: Establish mechanisms for human intervention in AI decision-making processes.
- **Enhance Transparency and Explainability**: Ensure AI decision-making processes are transparent and explainable.
- **Protect Privacy**: Implement robust measures to protect individual privacy and handle data securely.
- **Ensure Safety and Security**: Design AI systems with measures to prevent harm and protect against misuse.
- **Encourage Public Engagement and Awareness**: Engage stakeholders and raise awareness about AI capabilities and limitations.
- **Establish Accountability and Redress Mechanisms**: Define clear accountability mechanisms and provide channels for individuals to seek redress.
- **Promote Sustainability and Societal Benefit**: Prioritize the development of AI systems that promote societal benefits and are environmentally sustainable.

By integrating these objectives, the HRF aims to create a regulatory framework that supports ethical, inclusive, and transparent AI governance. The vision of the HRF is to create a framework that will be continuously updated and strengthened in response to advancements in AI technology that enhance public services while protecting the rights and interests of all citizens, promoting a more equitable and just society

2.4 Technical Description of HRF

The HRF prescribes certain dimensions related to AI governance when AI is involved to a lesser or to a greater extend at services that are provided to the public by governmental entities. These dimensions are the following:

1. Regulations compliance
2. Fairness
3. Non-discrimination - inclusivity

4. Human oversight
5. Human-Centered design
6. Privacy
7. Explainability - transparency
8. Auditing – continuous monitoring
9. Accountability
10. Safety and security
11. Claims and redress
12. Public engagement
13. Awareness
14. Societal benefit and well-being
15. Sustainability

A short description of each dimension follows:

Regulations compliance: Ensuring that the AI system complies with all relevant laws, standards, ethical guidelines and codes of conduct. For example, compliance with the EU Charter of fundamental rights, compliance with EU AI Act, compliance with EU Data Act, etc.

Fairness: The AI system should be designed and deployed in a way that treats all individuals fairly and equitably, even if the bias is not related to protected characteristics (gender, age, religion, disability, etc). It considers the overall distribution of benefits, burdens, and outcomes across different groups and communities, and aims to prevent or mitigate any unfair disparities or disadvantages.

Non-discrimination - inclusivity: The AI system should be designed and deployed in a way that promotes inclusivity and does not discriminate against or exclude any individual or group based on protected characteristics.

Human oversight: Appropriate human oversight and control mechanisms should be in place to monitor the AI system's decision-making processes, intervene when necessary, and ensure that the system is operating as intended and in accordance with ethical principles.

Human-Centered design: The AI system should be designed with a focus on human needs, values, and well-being, considering human factors such as usability, accessibility, and user experience.

Privacy: The AI system should be designed and deployed in a way that respects and protects the privacy of individuals, including through measures such as data minimization, anonymization, and secure data handling practices.

Explainability - transparency: The AI system's decision-making processes should be transparent and explainable to relevant stakeholders, allowing for scrutiny, accountability, and trust-building.

Auditing – continuous monitoring: There should be mechanisms in place for continuous monitoring and auditing of the AI system's performance, outputs, and impacts, enabling timely detection and correction of issues or unintended consequences.

Accountability: Clear lines of responsibility and mechanisms for holding entities accountable should be established for the development, deployment, and use of the AI system.

Safety and security: The AI system should be designed and deployed with robust measures to ensure safety and security, protecting against unintended harm, misuse, or malicious attacks.

Claims and redress: There should be clear and accessible mechanisms for individuals to seek redress and compensation in cases where they have been adversely affected by the AI system's decisions or actions.

Public engagement: Stakeholder engagement and public participation should be encouraged throughout the AI system's development and deployment, fostering transparency, trust, and alignment with societal values and needs.

Awareness: Efforts should be made to raise awareness and promote understanding of the AI system's capabilities, limitations, and potential impacts among relevant stakeholders and the public.

Societal benefit and well-being: The development and deployment of the AI system should be guided by a commitment to promoting societal benefit and well-being, considering both short-term and long-term impacts.

Sustainability: The AI system should be designed and deployed with consideration for environmental sustainability, resource efficiency, and long-term viability, minimizing negative impacts on the environment and promoting sustainable practices.

Figure 1 shows at the perimeter the above dimensions and at the center systems that should facilitate them such as the governance and administration system, the AI ethics and compliance system, the auditing and monitoring system, the explainable system and the AI registry system.
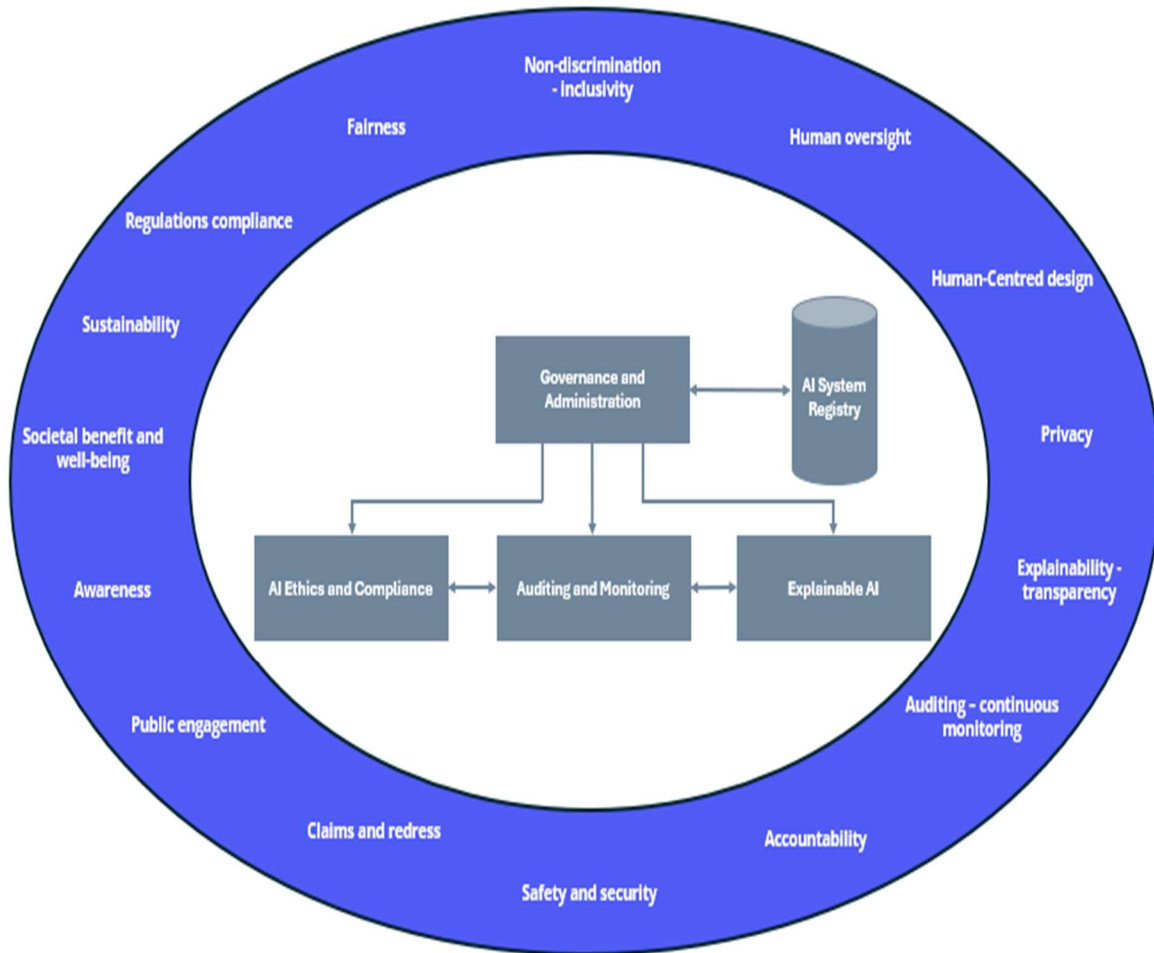
Figure 1 HRF Dimensions

3 Theoretical Framework

3.1 Overview

This chapter presents the theoretical and methodological steps that were followed to conduct a field study in the form of focus groups on the targeted underrepresented populations to complementary support the HRF of AI4Gov. It begins with a theoretical analysis defining the fundamental terms of bias and discrimination, followed by a literature review on the perception of these terms in public services. Next, it describes the study design of the case studies, methodology and description, along with the limitations. The research findings are presented using two methods: a quantitative analysis (conjoint analysis) of the questionnaire responses and a content analysis of the research results. The research results will feed into Chapter 4 as recommendations to achieve inclusivity (section 4.4).

3.2 Bias and Discrimination: Fundamental Definitions

A thorough literature review on the definition of bias and the forms of bias has already been made in D2.1 AI4Gov Holistic Regulatory Framework V1, where the definition of traditional bias and the types of conscious and unconscious biases were presented along with the impact of bias on individuals and society. Building upon the D2.1, in this section the fundamental definitions of bias and discrimination will be analyzed in order to maximise the potential of the HRF.

As described in D2.1, the definition of bias that has been chosen by AI4Gov is the following: *Bias is a term used to describe an inclination or prejudice for or against an individual or group in a way that is considered unfair. It can result from personal experiences, societal norms and expectations, or information we have absorbed from various sources such as media, education, and family (Greenwald & Krieger, 2006). Bias can be categorised into two primary types: conscious (or explicit) bias and unconscious (or implicit) bias (Banaji & Greenwald, 2013). Both types of bias can have a significant impact on individuals and society, perpetuating social inequalities and injustices.*

Discrimination occurs when individuals or groups are treated unfairly or unequally based on certain characteristics such as race, gender, age, religion, or disability. It is an action or a behaviour that arises from biased attitudes. In literature there are many types of discrimination.

1. <u>Direct discrimination</u> is defined as the unequal treatment based on protected characteristics.
2. <u>Indirect discrimination</u> describes the policies or practices that appear neutral but result in unequal treatment of certain groups.
3. <u>Systemic discrimination</u> involves the procedures, routines and organisational culture of any organisation that, often without intent, contribute to less favourable outcomes for minority groups than for the majority of the population, from the organisation's policies, programmes, employment, and services (Council of Europe, 2020).

In D1.4 Gender and Ethical Management Plan V1, an initial analysis on the definitions of bias, discrimination, fairness and equality in relation to algorithmic bias and algorithmic discrimination

was developed. D1.4 draws on Janneke Gerards and Raphaele Xenidis's Special Report 'Algorithmic Discrimination in Europe: Challenges and Opportunities for Equality and Non-Discrimination Law'. The authors distinguish between 'bias' and 'fairness' versus 'discrimination' and 'equality'.

Gerards and Xenidis note that 'bias' and 'fairness' have broader meanings in the context of algorithmic operations compared to 'discrimination' and 'equality'. 'Algorithmic bias' includes a wide range of systematic errors, not all of which constitute legal discrimination. Legal discrimination involves adverse treatment of protected groups under EU law, based on sex, racial or ethnic background, disability, sexual orientation, religion or belief, and age (Article 19 TFEU). 'Algorithmic discrimination' is thus a subset of algorithmic bias that violates these legal protections.

'Algorithmic fairness' refers to methods aimed at preventing bias and ensuring ethical outcomes, encompassing various types such as group, individual, procedural, and outcome-based fairness. This concept is broader than legal definitions of equality and non-discrimination.

The concepts of bias and discrimination are interrelated as discrimination is often a manifestation of bias. While bias is a predisposition or prejudice, discrimination is the unfair treatment that can result from biased attitudes. When biases, whether conscious or unconscious, influence decisions, they can result in discriminatory practices. Furthermore, addressing bias is a key step in preventing discrimination. By understanding and mitigating biases, organizations and individuals can create more equitable environments.

3.3   Literature Review on Bias and Discrimination in Public Services

After clearly defining the key differences and correlations between the concepts of bias and discrimination (section 3.2), this section will perform a literature review to understand discrimination specifically in public services, which will serve as the theoretical framework used to develop the methodology for the AI4Gov study as well as to feed into the technical development of the HRF.

Bias and discrimination in public services remain critical issues globally, influencing access, quality, and outcomes for diverse populations. Public services, including healthcare, education, law enforcement, and social services, are essential for societal well-being. Yet, systemic biases and discriminatory practices within these sectors can perpetuate inequalities and undermine trust in public institutions. This review synthesizes the current literature on bias and discrimination in public services, exploring the nature, causes, and impacts of these phenomena.

Bias and Discrimination in Healthcare

Healthcare is a critical area in which bias and discrimination manifest with significant consequences. As strongly pointed by Ledoux (2015), "*the differences in socioeconomic positions should not be considered as autonomous but produced by the long history of discrimination and racism.*" Studies have shown that racial and ethnic minorities often receive lower quality care compared to their white counterparts.

The World Health Organization (2001) has linked health disparities to sociostructural inequalities that stem from discrimination in societies. Experiences of discrimination within healthcare settings may present a barrier to healthcare for people that are socially disadvantaged due to gender, immigration, race/ethnicity, or religion (Rivenbark & Ichou, 2020). Furthermore, research indicates that Black patients are less likely to receive pain management, appropriate cardiovascular interventions, and necessary diagnostic tests (Green et al., 2007).

Implicit bias among healthcare providers has been extensively documented. A meta-analysis by Hall et al. (2015) found that healthcare providers exhibit implicit biases favoring white patients over minority patients, which adversely affects clinical decision-making and patient-provider interactions. Moreover, language barriers and cultural misunderstandings enhance healthcare disparities for non-English speaking patients (Samkange-Zeeb et al, 2020).

In addition, research in the UK by Nazroo (2003) reveals that ethnic minorities experience significant healthcare inequalities, with institutional racism cited as a major factor. In Sweden, a study by Wamala et al. (2007) found that immigrants and ethnic minorities reported higher instances of discrimination in healthcare settings, affecting their overall health outcomes.

Bias and Discrimination in Education

In education, bias and discrimination affect both access to resources and educational outcomes. Studies highlight that students of colour, particularly Black and Latino students, face harsher disciplinary actions compared to white students for similar infractions (Skiba et al., 2011). These practices contribute to the school-to-prison pipeline, disproportionately affecting minority students.

Teacher biases also impact student performance. Research by Gershenson et al. (2016) shows that teachers' expectations for students can vary significantly based on race, with lower expectations for Black students potentially leading to lower academic performance. Additionally, biases in special education placements and gifted programs further perpetuate educational inequalities (Harry & Klingner, 2006).

In the UK, a study by Gillborn et al. (2012) found that Black students are systematically disadvantaged by racial biases in disciplinary practices and academic expectations. In France, research by Oberti (2007) shows that immigrant students are often relegated to lower-quality schools and tracks, perpetuating educational inequalities. Furthermore, in the Netherlands, studies highlight significant disparities in educational attainment and school placements among students of migrant backgrounds, often linked to teacher biases and systemic discrimination (Crul & Doomernik, 2003).

Bias and Discrimination in Law Enforcement

Law enforcement is another domain where bias and discrimination have profound effects. Numerous studies have documented racial profiling and disparate treatment of minority communities by police. Black and Latino individuals are more likely to be stopped, searched, and

arrested compared to their white counterparts, even when controlling for relevant variables (Gelman et al., 2007).

A field experiment conducted in the United States aimed to examine whether racial discrimination exists in access to public services. The results showed that racial discrimination exists in local public services, as emails from putatively Black senders were almost 4 percentage points less likely to receive an answer compared to emails signed with a white-sounding name. Moreover, responses to queries from Black names were less likely to have a cordial tone (Giulietti et al., 2019).

Implicit bias training for law enforcement officers has been implemented as a strategy to reduce discriminatory practices. However, the effectiveness of such training remains contested, with some studies suggesting limited impact on actual behavior (Worden et al., 2020). The need for comprehensive reforms addressing structural issues within the criminal justice system is increasingly emphasized (Alexander, 2010).

In the UK, research by the Equality and Human Rights Commission (2013) indicates that Black and minority ethnic individuals are disproportionately targeted by stop-and-search practices. Studies in France show that racial profiling by police is a significant issue, particularly affecting Black and Arab youth (Open Society Justice Initiative, 2009). In Germany, there is evidence of ethnic profiling in police practices, with studies indicating that people of Turkish and Middle Eastern descent are more likely to be stopped and searched (Amnesty International, 2016).

Bias and Discrimination in Social Services

In social services, bias and discrimination can influence the delivery of welfare benefits, child protection services, and housing assistance. Minority families often face differential treatment in child welfare investigations (Dettlaff et al., 2011). Similarly, access to housing assistance can be hindered by discriminatory practices by landlords and housing agencies (Silver et al., 2019).

Sexual orientation and gender identity that differ from the heteronormative standard of most societies have been targets for discrimination (Rivers & D'Augelli, 2001). Research highlights that LGBTQ+ individuals encounter bias and discrimination in social services. This population often faces barriers to accessing appropriate services, including homeless shelters, due to stigma and lack of inclusive policies (Mottet & Ohle, 2003).

A study by Van Lancker (2013) in Belgium revealed that migrants and ethnic minorities face significant barriers in accessing social housing and welfare benefits, often due to bureaucratic hurdles and discriminatory practices. In the UK, research by Hudson and Lowe (2009) showed that minority ethnic groups are disproportionately represented in child protection cases, often due to cultural misunderstandings and biases. Additionally, the European Union Agency for Fundamental Rights (FRA) report (2013) highlighted widespread discrimination against LGBTQ+ individuals in accessing social services across EU member states.

3.4 Case studies

To examine the forms of non-AI discrimination faced by underrepresented citizens in non-AI public services, **three case studies** were set in **Greece**, **Spain**, and **Slovenia**. The focus groups were conducted in these countries as they are the AI4Gov pilot countries, and the aim was to gather in-depth insights on experiences of vulnerable groups of citizens when interacting with public services. It has to be noted that the results of the focus groups that are presented below, are complementary to the research and function under the umbrella of the theoretical background of the HRF, but there are in no way a generalisation for the case study countries.

The study methodology, outlined in detail in Deliverable D2.1, served as the foundation for the AI4Gov focus groups. In this section, the study design is elaborated, focusing on the key objectives and structure, sampling techniques, and data collection procedures applied. The research results and limitations are also discussed.

3.4.1 Study Design

The main goal of the study was to **explore the challenges and discriminatory practices faced by vulnerable or underrepresented citizen groups** when accessing government services, whether online or in person. The central research question guiding the study was: ***How do traditional biases (non-AI) currently impact the rights and values of underrepresented groups?***

To address this question, the project aimed to map the challenges encountered by these groups when attempting to access various government services and information. The overarching goal of the study was to incorporate the perspectives and experiences of marginalized groups into the **Holistic Regulatory Framework** of AI4Gov in the form of recommendations, thereby complementing existing literature reviews in sections 3.2 and 3.3 and Deliverable 2.1.

To achieve this, **a series of focus groups** were conducted with underrepresented groups and relevant organizations. These groups included: Students, Young unemployed adults, People with disabilities, Elderly individuals and Gender-based Violence Victims.

3.4.1.1 Key Objectives Examined

In order to address the main research question, some key objectives were set to guide the interview series and help us structure the content of the focus groups. These objectives covered the **why, how**, and **when** of discrimination to provide a holistic view of the participants' experiences. They are categorised below:

Perceived Causes of Discrimination: Investigate perceived causes such as lack of interest, information, awareness, and potential gaps in the skills or experience of employees.

Forms of Discrimination: Identify forms of discrimination faced by citizens, such as negative assessments for loans, differential pricing for services, extended waiting times, etc.

Underlying Discrimination: Identify underlying forms of discrimination, including microaggressions.

Given the vulnerability of the research groups, the study was designed as **scenario-based focus groups** to enhance participation. The aim was to elicit rich narratives and insights from participants regarding their encounters with discrimination in various service contexts. This way, the interviewee didn't feel the pressure of sharing directly personal stories, rather than reflecting on scenarios based on their experience.

3.4.1.2   Focus Group Structure

- ➢ **Introduction and Icebreaker (10 minutes):** The facilitator welcomed participants, introduced the research team, and established rapport through an icebreaker activity to create a comfortable atmosphere.
- ➢ **Explanation of Scenario-Based Approach (5 minutes):** The facilitator explained the scenario-based methodology, emphasizing that participants would be presented with hypothetical situations related to accessing public services and encouraged to share their thoughts and experiences.
- ➢ **Presentation of Scenarios (15 minutes):** Participants were presented with predefined scenarios depicting common situations where discrimination in accessing public services might occur. Each scenario was read aloud, and participants were given time to reflect on their own experiences related to the scenario.
- ➢ **Group Discussion (30 minutes):** Participants engaged in an open discussion prompted by the presented scenarios. The facilitator encouraged participants to share personal anecdotes, feelings, and perceptions related to discrimination in accessing public services. Probing questions were used to explore the nuances of each scenario and encourage deeper reflection.
- ➢ **Reflection and Debrief (10 minutes):** The facilitator guided participants in reflecting on the themes and insights that emerged during the discussion. Participants had the opportunity to share any additional thoughts or concerns they had regarding discrimination in accessing public services. The participants had to complete a questionnaire provided by the facilitator with overall questions and demographic information.
- ➢ **Closing and Appreciation (5 minutes):** The facilitator thanked participants for their valuable contributions and provided information about any follow-up steps or support resources available. Participants were reminded of the confidentiality of the discussion and encouraged to reach out if they had any further questions or concerns.

3.4.2   Sampling Techniques

The target population identified for the AI4Gov focus groups comprised by underrepresented or vulnerable groups that interface with public institutions, both face-to-face and online in the three case studies set. Specifically, participants were recruited from diverse backgrounds, including marginalized communities, students, persons with disabilities, and individuals facing socio-economic challenges.

Participants were recruited through community organizations, social service agencies, and online platforms known to serve vulnerable populations. The goal was to ensure a **diverse and**

inclusive sample representing the experiences and perspectives of different vulnerable groups in each country studied.

A **purposive sampling technique** was employed for this study, as the aim was to explore challenges faced by specific underrepresented groups, including young unemployed adults and students, people with disabilities, elderly individuals, and low-income households. Purposive sampling was chosen as it allows researchers to **deliberatively select participants** who meet specific criteria relevant to the study objectives. This method, also known as judgemental, selective, or subjective sampling, it commonly used in qualitative research and involves selecting participants based on characteristics or qualities that align with the research objectives. It is a **non-probability sampling method** where the researcher decides who to include in the sample based on their specialist knowledge of the research issue and the participants' capacity and willingness to engage in the study (Rai et al., 2015).

Therefore, initial research identified potential organizations and services that align with the study's objectives. In terms of geographic spread, it was decided the focus groups to be conducted in the three (3) AI4Gov pilot countries: **Greece, Spain and Slovenia.** Over 40 organizations in these countries were contacted via email, many of which were suggested by the AI4Gov pilot partners in those countries. Due to limitations in availability, time, and resources, the participants recruited primarily came from various vulnerable groups in Greece, as well as from Spain and Slovenia. Consequently, the objective is to conduct an initial analysis of citizens' perceptions of public services across the three case studies, with the intention of performing a more in-depth analysis in future research. In line with the collaboration between work packages (WPs), these organizations will be contacted again for potential future participation in project activities, validation, and evaluation processes. In the subsequent phase, a meeting was held with interested organizations to present the study's objectives and the focus groups methodology.

During the sampling and recruitment phase, efforts were made to ensure diversity and inclusion while addressing ethical concerns. This included obtaining informed consent, ensuring confidentiality and anonymity, and being sensitive to the participants time and circumstances.

3.4.2.1 Participating Organizations

Six organizations and social agencies agreed to conduct focus groups:
- **IASIS at Work** (Greece) is a day center supporting workers and the unemployed, reaching out to young adults, permanent employees, and the permanently unemployed. Conducted **two focus groups**: the first with **12 participants** and the second with **9 participants.**
- **EDRA Social Cooperative Activities for Vulnerable Groups** (Greece) is a civil, non-profit cooperative organization dedicated to promoting mental health services and safeguarding the rights of socially vulnerable groups. EDRA conducted 2 focus groups:
  - One focus group with **7 mental health service recipients**.
  - One focus group with **8 beneficiaries of the social meal program and grocery store**.
- **Aristotle University of Thessaloniki**, Department of Political Sciences (Greece): Conducted two focus groups with **51 participants**.
- **REUSE Center (CPU)** - Slovenia is an epicenter of sustainability and social innovation, creating employment opportunities for marginalised populations. CPU conducted 1 focus group with **12 employees with disabilities**.

o **ALMA Association (Spain)** is dedicated to fighting gender violence and achieving real equality between men and women. Conducted 2 focus groups with **3 employees** and **2 victims of gender-based violence**.

o **Autism Federation of Extremadura (FAE)** - Spain is a federation that protects people with autism spectrum disorder. FAE conducted a focus group with **9 participants.**

3.4.3   Data Collection Procedures

The focus groups were conducted face-to-face, moderated by personnel from each participating organization, with members of ViLabs present or e-present to offer assistance and guidance when needed. Before conducting the focus groups, a small-scale pilot test of the focus groups' questionnaire was conducted with a sample of the target population consisted of students from the Aristotle University of Thessaloniki and members of the Greek organisation "IASIS at work". The aim of the pilot was to evaluate the clarity, comprehensibility, and cultural appropriateness of the questions. Based on the feedback received, the necessary revisions and refinements were made.

The IASIS at work focus groups were moderated by ViLabs members. In some cases (e.g. EDRA and CPU focus groups), the presence of ViLabs members were intentionally avoided to enhance participants' trust and encourage them to speak confidently and freely without the influence of "unfamiliar" individuals.

Recording and Transcription: Most focus groups were audio-recorded and transcribed verbatim by the moderators. These transcriptions were then shared with ViLabs, which conducted a thematic analysis (see section 3.6.1) to identify recurring patterns, themes, and insights related to experiences of discrimination in accessing public services.

Timeline: The focus groups and data collection took place from May 2024 until early June 2024. The specific schedule was as follows:

1. **IASIS at Work:**
   ➢ 1st Focus Group: May 15, 2024
   ➢ 2nd Focus Group: May 16, 2024
2. **EDRA:**
   ➢ 1st Focus Group: May 13, 2024
   ➢ 2nd Focus Group: May 14, 2024
3. **CPU:**
   ➢ Focus Groups conducted during the week of May 27-31, 2024
4. **ALMA Association:**
   ➢ 1st Focus Group: June 4, 2024
   ➢ 2nd Focus Group: June 7, 2024
5. **FAE:**
   ➢ Focus Group: June 5, 2024
6. **Aristotle University of Thessaloniki**
   ➢ 1st Focus Group: May 20, 2024
   ➢ 2nd Focus Group: May 25, 2024

Table 1 AI4Gov Focus Groups

Organization	Focus Group 1	Focus Group 2	Focus Group 3
IASIS at Work	May 15, 2024	May 16, 2024	
AUTH	May 20, 2024	May 25, 2024	
EDRA	May 13, 2024	May 14, 2024	
CPU	May 27-31, 2024		
ALMA	June 4, 2024	June 7, 2024	
FAE	June 5, 2024		

3.5   Study Description

The structure of the Study was designed in two main parts:

a) **Scenario-based discussion:** Participants engaged in an open-ended discussion where they were invited to share their perspectives and experiences with public services. In order to enhance the sharing of information, the moderators read to the participants tailored scenarios adapted to their specific characteristics, on which they elaborated with their experiences. This part of the study allowed for in-depth insights and qualitative data collection.

b) **Questionnaire:** At the end of the focus group, participants completed a questionnaire featuring closed questions. This questionnaire included demographic information and inquiries about their engagement with public services. Additionally, it was used to assess their interest in specific aspects of public services based on the factors set for conjoint analysis. The full questionnaire form can be found in Annex 1.

We identified four key measuring factors to be correlated with other variables in the study:

o **Ease of Communication** with Government Services: This factor evaluated how easily participants could communicate with government services.

o **Issues** Encountered Requiring Service Visits: This factor assessed the types and frequencies of issues that necessitated participants' visits to public services, choosing from personal or social issues.

o **Clarity** of Service Responses: This factor examined whether participants found the responses from public services regarding issue resolution to be understandable and clear. It focused on the transparency and comprehensibility of the information provided by the services.

o **Trust** in Issue Resolution: This factor measured participants' trust in the public services' ability to resolve or address their issues effectively. It explored their confidence in the resolution process and the perceived reliability of the services.

3.5.1 Study Questions

The following questions were meant to guide the conversation specifying the kind of information the research team seeks to gather. There were two separate sets of questions:

1. For the vulnerable population who have felt discriminated against,

2. For the staff members that can give their input as observers.

Group 1: Vulnerable citizens

The questions **were not asked directly** and were covered through scenarios to make the participants feel more confident and comfortable sharing their experiences.

- *Have you ever felt discriminated against when accessing public services, either in person or online? - If yes, please explain why you think they discriminated against you.*
- *Can you describe the situation where you felt discriminated against while accessing public services? (Which services and what type of difficulties did you encounter? Was it an online service or a f2f?)*
- *Were there any specific actions or behaviours by public service staff or officials that made you feel discriminated against?*
- *How has discrimination impacted your access to services or information? Can you provide examples of outcomes of discrimination, (e.g. denial of service, inflated prices, extended waiting times)*
- *Do you think that the discrimination you experienced was intentional, or unintentional/unconscious?*
- *Have you ever reported instances of discrimination when accessing public services? If yes, what was the outcome of your report?*
- *Are you aware of the fundamental rights and protections granted to you under EU law?*

Group 2: Observers

For the second group their input was gathered through the scenarios or by asking direct questions.

- *Through your work, have you noticed any patterns or trends in how biases manifest themselves in public service interactions? (For example, in decision-making processes, resource allocation, or service provision.)*
- *When accessing public services online, do you perceive any biases in the design or functionality of digital platforms or applications? If yes, please describe.*
- *Have you ever witnessed/experienced microaggressions or subtle forms of bias when interacting with public service providers? If so, please provide examples.*
- *In your opinion, what measures could public service providers take to ensure that all individuals are treated fairly and without discrimination when accessing their services, whether face-to-face or online?*

3.5.2   Participant Demographics

The focus groups conducted during the AI4Gov project involved a total of **113 participants** in the three case studies set. The demographic breakdown of these participants is detailed below:

Geographic Distribution

Greece: 78%   (87 participants)
Slovenia: 10%   (12 participants)
Spain: 12% (14 participants)
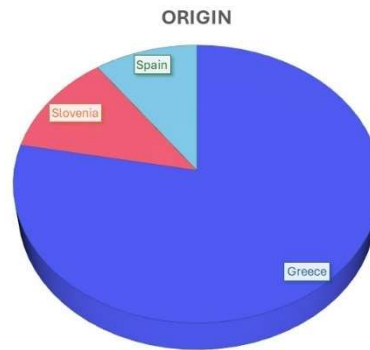
Figure 2: Geographic Distribution of AI4Gov Focus Groups Participants
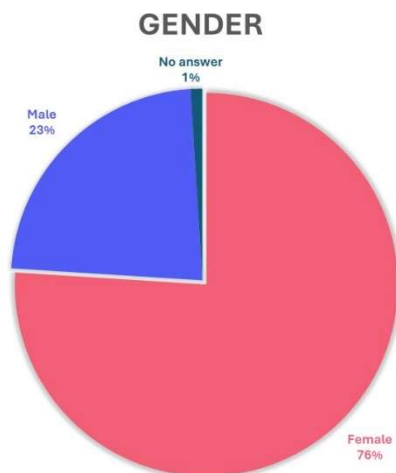
Gender Distribution

Female: 76%   (86 participants)
Male: 23% (26 participants)
Rather Not Say: 1% (1 participant)

Figure 3: Gender Representation of AI4Gov Focus Groups Participants

Age Distribution

Younger than 25 years old: 22% (25 participants)
Between 26-40 years old: 25% (28 participants)
Between 41-65 years old: 51% (58 participants)
No answer: 2% (2 participants)
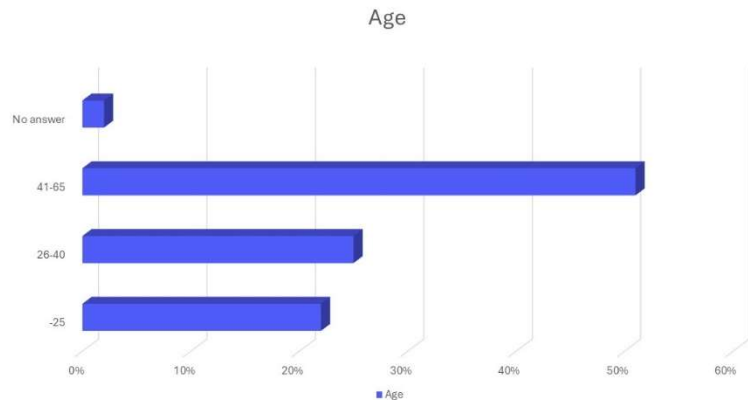
Figure 4 Age Distribution of the AI4Gov Focus Groups Participants
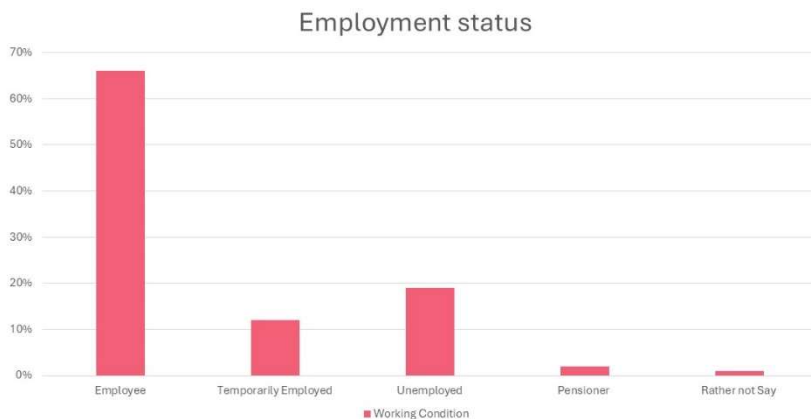
Employment Status

Employees: 75 participants (66%)
Unemployed: 22 participants (19%)
Temporarily Employed: 13 participants (12%)
Pensioners: 2 participants (2%)

Figure 5 Employment Status of AI4Gov Focus Groups Participants

The demographic information reveals a majority presence of females and a significant proportion of middle-aged individuals. Although the predominant representation is from Greece, these participants come from various vulnerable groups, reflecting diverse backgrounds such as pensioners, people with mental health issues, individuals living below the poverty line. In the Spanish case study, the participants consist of different vulnerable groups, namely gender-based victims and people with autism and families with autistic members, while the Slovenian case study focuses on employees with various disabilities. The employment status indicates that most participants are currently employed, with smaller fractions being unemployed, temporarily employed, or retired.

3.5.3 Limitations

While the focus groups conducted in this study provide valuable insights, several limitations must be acknowledged. The theoretical framework of this survey was designed at a European level. However, the workshops were designed as mini case studies targeting vulnerable citizens in the three pilot countries of the AI4Gov project: Greece, Slovenia, and Spain. This focus on specific countries limits the generalizability of the findings across the broader European context. Thus, the study findings provide a methodology for the development of inclusive AI tools and will be complementarily used for the development of the HRF.

The chosen study sampling technique was consisting of participants from Greece, Slovenia, and Spain. In the case of Greece, the sample is the larger of all three countries. This was due to an easier communication with the local organisations and the participation of students that added the perspective of (mainly) non-vulnerable citizens in comparison to the vulnerable groups participating. In Spain and Slovenia, the sample is limited only to vulnerable citizens or employees in the respective organisations, making the results more context specific. As a result, the participants may not adequately represent all demographic and socio-economic groups within the participating countries.

The insights gained from the focus groups and workshops are context specific. They are based on the particular social, economic, and political environments of Greece, Slovenia, and Spain at the time of the study. These environments may differ significantly from other European regions, limiting the transferability of the findings to different contexts.

The study was conducted within a specific timeframe. Social and political changes that occur after the study period may influence the relevance and applicability of the findings in different contexts. Future studies should consider longitudinal approaches to capture evolving dynamics in the context of bias and discrimination in public services.

The analysis was primarily qualitative, providing rich, detailed insights but potentially lacking the statistical power and generalizability of quantitative studies. Future research could benefit from a mixed methods approach to triangulate findings and enhance robustness.

With all the above said, it is crucial to make clear that the focus group results are not a generalisation of the situation vulnerable citizens experience in Europe, rather than shorter case studies that allowed us to identify the most problematic areas in public services. Also, even though the focus group results feed the HRF, they are not the main input of the framework and function complementary to the extensive European-level research that has been conducted.

3.6 Research Findings

As previously mentioned, the focus groups were conducted in the three AI4Gov Use Case countries: Greece, Spain, and Slovenia. Although a large number of participants are from Greece, various vulnerable groups are represented in the three countries. Their perceptions are presented and analysed in this section.

The research findings will be analysed along two main axes:
- **Statistical Analysis**: This will involve quantitative data collected from the questionnaires provided to the focus groups' participants, presented on a per case study basis. The analysis will offer an initial methodology and examination of each vulnerable group's perception of public authorities in each of the participating countries.
- **Content Analysis**: This will focus on key insights gathered from the qualitative data during the focus groups.

The statistical analysis measured the factors, along with the most important variable for each factor, that impacts citizens' attitudes towards public services in each of the cases studied. Additionally, content analysis will shape these perceptions by identifying the main issues faced by citizens in each participating country, particularly those that conflict with EU values and human rights.

The research findings will be categorized based on the four measuring factors initially established:

- ❖ Ease of Communication
- ❖ Issues Encountered
- ❖ Clarity of Service Responses
- ❖ Trust in Issue Resolution

3.6.1 Statistical Analysis

The statistical method used for this study was **conjoint analysis (CA),** which does not require samples to be representative of the target population. We applied CA in the three case studies, and this section presents the results based on questionnaires collected during the focus groups.

Conjoint analysis is a survey-based statistical technique widely used in market research to determine how people value different attributes (features, functions, benefits) of a product or service. This approach combines real-life scenarios with statistical techniques to model actual market decisions. The term "conjoint analysis," coined by Green and Srinivasan (1978), refers to various paradigms in psychology, economics, and marketing that quantitatively describe consumer preferences or value trade-offs. The conceptual framework of conjoint analysis in marketing is derived from consumer theory, which involves search and learning processes as consumers identify available market options (Louviere, 1994).

Conjoint analysis is powerful for understanding the relative importance of different features in decision-making processes. The term "conjoint" signifies the consideration of multiple factors jointly. This statistical method was chosen to analyse the four factors defined as study objectives and their relative importance when associated with other variables, such as gender, age, education level, and interest in politics. It is important to note that this is a work in progress, and the deliverable presents the initial results from the conjoint analysis performed in focus groups conducted in Greece.

Based on the conjoint analysis, the importance of each measurement factor was assessed alongside other variables in each case study. The results are presented separately for the three case studies: Greece, Slovenia, and Spain. It is important to note that due to time and resource limitations, as well as differing availabilities, the number of participants varied in each country, which led to some limitation in the breakdown of key factors by demographic characteristics for Slovenia and Spain. Consequently, the following analysis should not be viewed as a generalization of the perceptions of vulnerable groups regarding public services. Instead, we propose this methodology to provide recommendations that will support the development of the HRF.

3.6.1.1 Greece

The analysis identified **four key factors influencing perceptions of public services in Greece**, and each factor's impact on individual attitudes was evaluated (Table 2):

Type of Issue (35%): The most influential factor on individuals' attitudes towards public institutions is the type of issue they need to address with a public service. Personal issues are the most significant variable within this category.

Ease of Communication (23.5%): The second most important factor is the ease of communication. The primary variable influencing this factor is the difficulty in communicating with public services.

Trust in Public Services (22%): The third factor affecting public opinion is the trust in public services to solve the problem at hand. Low trust in public services leads to dissatisfaction among participants.

Clarity of Responses (19.6%): The factor with the least influence on public perception is the clarity of responses received from public bodies. The primary variable affecting this factor is the difficulty in understanding the responses.

	Importance		Most Important Point Factor
Ease of communication	23.5%	Easy	
		Difficult	█████
Issues encountered	34.9%	Personal	█████
		Economic/ Employment	
		Social	
Clarity of response	19.6%	Relatively Understandable	
		Difficult to Understand	█████
Trust in Resolving the Issue	22%	Low	█████
		High	

Table 2 CA results - Greece

Overall, the factors that influence Greek citizens' attitudes towards public services are: 35% the issue with the most important factor being the "personal" issue, 23.5% the communication that they find difficult, 22% the trust that is low and 19.6% the response that is difficult to understand.

Breakdown of Key Factors by Demographic Characteristics

To better understand the results, the breakdown of the 4 key factors based on the demographic characteristics of the participants follows below.

Employment Status

The unemployed (26.6%) visit public services for personal issues mainly, communication with public communications is difficult, and they have low trust in the services to sort out their issues.

For temporary workers (9.4%) visit public services for personal issues mainly, they find it easy to communicate with officials, but have little trust in public services to solve their problems.

For long-term employed (64.1%), communication with public services is difficult, they find it hard to understand the response they usually receive and have little trust to resolve the issue.

Sex

Women (79.5%) participants find it difficult to communicate with public services and the responses they receive from those responsible are hard to understand and have little confidence that they will take appropriate action to resolve issues.

On the contrary, male (20.5%) participants visit the public services about a personal issue mainly, while they have low trust in the ability of the public services to find a solution to the problems.

Age
The under-25 age group (25%) turn to public services for personal issues and find it difficult to communicate with them.

In the 26-40 age group (26.5%) they visit public services for a social issue, find it easy to communicate with public services, but continue to have little confidence in public services to resolve issues.

The 41-65 age group (50%) visit public services for issues related to work or finances, find it difficult to communicate with them, and also find the answers they usually receive from public services difficult to understand.

Educational level
People with a high school/technical education level are familiar with public services for issues related to work or finances, find it easy to communicate with public services, have enough confidence in the authorities and consider the answers they receive to be relatively understandable.

Participants with a University/College level of education (70.5%) that approach the authorities on social issues, find it difficult to communicate and find the answers they receive from public services difficult to understand.

Participants with an educational level higher than a university degree (25.5%) turn to the authorities for personal issues, find it difficult to communicate and the answers they receive from public services difficult to understand, and have little trust in public services.

Way of contact with public services
Furthermore, the analysis concludes that most citizens (67.2%) choose to communicate with public services both face-to-face and electronically, with communication being difficult and trust in public services being low. The next most preferred mode of communication is online, with 21.9% of citizens choosing it mainly for work/economic issues, finding communication and understanding difficult and trust in the authorities low. Finally, 10.9% of participants communicate interpersonally on a social issue, but with little trust in public services.

Familiarity with the internet and applications in general
In terms of familiarity with the use of the internet and applications in general, 46.9% responded that they are very familiar, while only 3.1% have little familiarity with online applications. In addition, 78.1% frequently contact a public service about a social issue but with difficulty in communicating and understanding the answer and little confidence towards the services.

Finally, when asked about their attitude towards an important issue in the local community, the answers were fragmented across all possible answers. The answer "I mobilize through social networks (FB, Instagram etc.)" prevailed by a small margin (29.7%), followed by the answer "I personally address the responsible persons" with 25% and "I let the responsible persons do their job" with 23.4%. The mode of action with the fewest responses was "I address a TV channel, a newspaper", which received only 1.6%.

3.6.1.2   Slovenia

		Most Important Point Factor
Ease of communication	Easy	
	Difficult	
Issues encountered	Personal	
	Economic/Employment	
	Social	
Clarity of response	Relatively Understandable	
	Difficult to Understand	
Trust in Resolving the Issue	Low	
	High	

Table 3 CA Results - Slovenia

In the Slovenian case study, the conjoint analysis identifies key variables influencing citizens' attitudes towards public institutions and services. The findings reveal that the most significant factor is the ease of communication, which the majority of participants found satisfactory. The second most influential variable is the low trust in public authorities to effectively resolve issues.

The other two variables analyzed—the clarity of responses and the type of issue—did not receive sufficient responses to draw meaningful conclusions. Given that the participants in the Slovenian case study were predominantly individuals with disabilities and the limited number of participants, a detailed breakdown of the key factors by demographic characteristics is not feasible.

3.6.1.3   Spain

The conjoint analysis conducted in the Spanish case study identified the most important factor that affects the citizens' attitude towards public authorities being the **type of issue.** In particular:

Type of Issue (47%): The most influential factor on individuals' attitudes towards public institutions is the type of issue they need to address with a public service. Personal issues are the most significant variable within this category.

Trust in Public Services (23%): The second most important factor is the trust in public services to solve the problem. Responders from the Spanish focus groups have enough trust to the public authorities that their issue will be resolved.

Ease of Communication (16%): The third most important factor is the ease of communication. The primary variable influencing this factor is the difficulty in communicating with public services.

Clarity of Responses (14%): The factor with the least influence on public perception is the clarity of responses received from public bodies.

	Importance		Most Important Point Factor
Ease of communication	16%	Easy	
		Difficult	▰
Issues encountered	47%	Personal	▰
		Economic/Employment	
		Social	
Clarity of response	14%	Relatively Understandable	▰
		Difficult to Understand	
Trust in Resolving the Issue	23%	Low	
		High	▰

Table 4 CA results - Spain

Similarly to the Slovenian case study, given that the participants in the Spanish case study were predominantly individuals with disabilities and due to the limited number of participants, it is not feasible to perform a detailed breakdown of the key factors by demographic characteristics.

Overall, the conjoint analysis performed showcased the key factors that have the strongest impact on the attitudes of citizens and their perceptions towards public services in each of the three countries that the focus groups were conducted, highlighting also the most important variable for each of the factors. From this analysis, it is important to note the differences on the importance of variables that affect the citizens' attitude towards the public institutions and authorities. In Greece and Spain, the most important factor is the issues that encountered for the citizens and need to be addressed by the public institutions, with personal issues ranking as the most important factor that will affect the attitude of the citizens towards the public services. On

the contrary, in Slovenia the participants ranked the ease of communication as the most important factor, considering that they communicate easily with the public authorities, while for Greece and Spain, the communication with the public authorities is considered difficult. Furthermore, for the Greek and Slovenian participants, the trust in public authorities was considered low, while on the Spanish focus groups the trust in the public authorities is considered high.

3.6.2 Key Insights

The feedback from the focus groups highlights several recurring themes and insights regarding the challenges faced by underrepresented groups accessing public services. Although the feedback pertains to public sectors in three different countries, the key insights are presented generally to address common problems faced by individuals based on the four measuring factors set during the survey design. Below are the key insights coming from the 9 focus groups conducted in May 2024.

- ❖ Ease of Communication
- ❖ Issues Encountered
- ❖ Clarity of Service Responses
- ❖ Trust in Issue Resolution

3.6.2.1 Ease of Communication

In all the focus groups conducted, the majority of feedback were related to the easiness of communicating with the public services and public institutions. Many participants expressed **difficulties in reaching public services**. In particular, many participants reported that communication with public services is often challenging, citing unavailability and unhelpfulness of employees. One student mentioned the persistent delays and bureaucracy reminiscent of past decades, despite recent efforts to modernize services.

In addition, students expressed **frustration over poor service** and inefficiency during **face-to-face interactions**, particularly citing delays and bureaucracy in Public Employment Services (PES) for unemployment benefits, contrasting with more efficient online health services.

A participant recounted a negative experience with court services after her father's death, where despite arriving before closing time, she was **rudely asked to leave** and eventually locked out. Furthermore, participants noted that older employees often offload work onto newer ones, exhibiting a condescending attitude. From the feedback received during the Greek focus groups, a significant problem identified was the **employees' lack of technological proficiency**, making it difficult for them to assist citizens effectively.

Many participants reported instances of misunderstanding and inconsistency in the public sector, such as employees being unaware of proper procedures, leading to **inefficiencies and frustration among citizens**.

There were many complaints from gender-based victims (GBV) about how women's situations are often devalued by public services, lacking proper empathy and professional handling. In particular, an employee of an organization dealing with gender-based victims reported: "*Sometimes the women are treated as if they did not have sufficient capacity or if they were not sufficiently coherent in their testimony. They don't keep in mind the state of the victim, how they describe the incidents or how vulnerable they are. On the part of some professionals, training and empathy are scarce. They **are not capable of addressing the issues** professionally. The services limit themselves to carrying out protocols. The protocols are not always effective because not all victims are in the same situation in terms of abuse and the experience they have lived. The institutions and the state devalue a lot the incidents and the reports of GBV. It is important to remember that GBV itself is not an invention. It is something that is there and should not be devalued at all. And the problem is that what this type of behaviour does is devalue something serious that is continually happening.*"

Furthermore, the employees of GBV organizations reported to create specific difficulties in the communication such as **prolonged processing times**, **unclear eligibility criteria** and **difficult technical jargon** especially for non-natives and non-EU citizens. An GBV participant stated: "*There's inadequate support for non-native speakers, with poor translations. In Catalonia if you don't speak Catalan, you cannot navigate the public websites and you won't be able to proceed, you have to go in person. The translation to Spanish is not guaranteed and is very poor*".

For people with autism and their families, which participated in the AI4Gov focus group, there are **communication gaps** especially in public healthcare, with **difficulties in reaching doctors** for follow-up appointments, even in emergencies. A parent who has a kid with autism reported: "*The doctor prescribed a medication that was very, very bad for him, and I found it impossible to contact her, even in emergencies. I was given many excuses, and in the end, I had to decide to stop the medication myself.*" Another one noted that there is no customised care for people with autism that are already struggling with people they don't know, referring to the replace of the paediatricians in the consultation who are not familiar with the children or their conditions.

Apart from the negative feedback, there were also some **positive feedback** collected, such as the **efforts made to expedite processes**. In particular, a participant shared an experience where a tax office director personally addressing people in the queue to handle simpler requests more quickly, while another one noted that the online services provided in the post-Covid era have simplified the procedures. Another participant reported that public employees are generally willing to help but issues like system failures still necessitate in-person resolution.

3.6.2.2 Issues encountered

In general, most of the participants of the focus groups reach a public institution in order to receive a public service in order to address a personal issue such as booking an appointment with a health institution or applying for social benefits and legal documents.

Although the option of online services has improved the experience with public services, this has brought some struggles. Electronic services often introduce additional bureaucracy and delays. Participants cited examples where applications were rejected outright due to missing documents,

causing unnecessary delays. This problem was reported also by participants that visited the public services for a social welfare and faced prolonged bureaucratic processes for simple requests, such as repairing damaged pavement, for which they were encountered with lengthy approval processes and lack of communication.

During the Greek focus groups, many participants described the Greek public sector as **impersonal,** with **employees** being **less willing to help** due to the shift to electronic transactions, exacerbating issues for citizens unfamiliar with technology.

The **lack of technological literacy** was also mentioned as a major struggle by **older participants**, who reported that they struggle with online services, often leading to delayed medical attention and a sense of being unvalued by the system, highlighting the **ageism** that **digital inequalities** have caused.

Another issue that was frequently reported by the participants was **gender-specific biases in vocational training and employment services.** In particular, an employee working with GBV reported: "*My sister, now she got a vocational training, and said to her counsellor that there was a mechanic vocational training in Villa Franca, [...]and the counsellor said they better put first hairdressing and food and then mechanics.*" While when applying for housing, many GVB participants reported that they face **unrealistic criteria and insufficient support for housing and financial independence**, highlighting the need for better resources and consideration of their unique circumstances.

Another important issue mentioned was the **discrimination in employment facing GBV survivors**, who are often offered training for low-paid and low-quality jobs under the assumption that this type of violence is common only in women working in such employment sectors or not working at all. Specifically, a participant cited: "*[...] But the women who suffer gender violence, come from all employment sectors. So, every time we have tried to find resources for these women and so that they can embark on their new path, the only thing they offer is cleaning, caregiving, home help assistant, or similar. It is like it's unconceivable that a woman who has her professional training, degree, master's degree, doctorate, can suffer gender violence. But she does*".

Additionally, all the underrepresented groups from the 3 participating countries (GR, SL, ES) that took part during the AI4Gov focus groups frequently mentioned that they **face difficulties in booking medical appointments via apps or phone**, especially for urgent needs, with **long waiting times** and **poor system functionality**. Especially for the people with autism the prolonged waiting hours were cited as a strong challenge. A parent of an autistic child reported that long waiting times for doctors make the experience very uncomfortable for both the child and the family. This often leads to preferring visits to the paediatrician without the child present to avoid the discomfort and stress of waiting.

3.6.2.3 Clarity of Service Responses

One of the main issues that many participants reported were the inconsistent procedures, unwilling or inability of public servants to help the citizens, the miscommunication and the delayed responses.

Many participants felt that public services often reject application **without providing clear guidance on corrections**, leading to **repeated visits and frustration**, while the t**echnological incompetence** of public servants was also cited as a **cause of unclear service responses**, particularly affecting **elderly and technologically inexperienced citizens.** A student stated: "*There is a general misunderstanding between the services without knowing all the necessary information about the procedures that the citizen must follow, as a result of which considerable time is lost. In the end, the citizen tries to find a way out on their own*", while an GVB participant shared the following experience: "I*n the social services office they said she [her mother] didn't have enough years to qualify for the pension but in the unemployment office they said she has sufficient years to qualify for the pension. She waited for 5 months to clarify the issue and with my support managed to get approval for 200 euros*". All the above-mentioned feedback led to the conclusion that the public services are characterized by **bureaucracy** that leads to **delayed responses and unresolved issues**, **with participants frequently needing to follow up multiple times**. Specifically, a participant with disability stated that "*they had to be consistent and go again and again and bother them until they finally manage to solve the problem*."

Many participants characterized the responses from various public services as "**undetermined**" and "**complicated**" and the process "***very bureaucratic.***" The services especially for vulnerable groups are **not simple and accessible**, while also the process of seeking assistance and resources is also limited. Even in online application services, which have improved the situation in many ways, inconsistencies remain. For example, a participant mentioned that the Spanish Health Service app for appointments was not up to date with available free slots for doctor appointments.

In the cases of GVB participants, many of them reported that public institutions such as legal experts, police authorities and social support authorities **do not provide proper protection and guidance**, but on the contrary in many incidents they **dissuaded the victims** from denouncing their abusers. Furthermore, the lack of support of for non-native speakers, which includes **poor translation services** and difficult technical jargon, creates **barriers for non-native speakers**, leading to **confusion and unclear service responses**.

In general, there were reports of public services not adhering to consistent procedures, such as varying requirements for document submissions and unclear eligibility criteria, highlighting **inconsistent procedures** that affect the **clarity and quality of services provided**.

3.6.2.4 Trust in Issue Resolution

Most of the participants expressed a general **lack of trust in public services' ability to resolve issues**, often citing repeated negative experiences and unhelpful attitudes from employees, which led to **prolonged bureaucratic processes and unresponsive public services**.

Although there is a **general disbelief** towards public services, many participants reported empathetic and proactive assistance from public servants. A student cited: "*To my great surprise, when I needed to make an appointment for the covid vaccine, the process was particularly simple, not requiring specialized computer knowledge, and giving me many options for location and time*".

Although the feedback received from the general population included some positive aspects, the feedback from GVB participants and family members of individuals with autism was more negative. Vulnerable populations often receive **inadequate support, further diminishing trust in public services' ability to address their needs effectively**. These population feel that they are not treated with respect from the public authorities and lack the support from social services and legal authorities. A participant working with GBV cases reported: "*if a victim of gender violence arrives at a health center with evident signs that she has been mistreated [...] although by protocol and by law they have to do it [complain], some of the professionals do not take responsibility to go deep into the incident and give it the credibility it deserves. A person comes looking for help, they come looking for support and what they have to find is that, help and support. Not a person who wants to move on as soon as possible and finish the shift as soon as possible. Social responsibility has a lot to do here, and if we had it there would be much fewer problems both with addressing GBV and in general.*"

Another issue that was frequently reported was the **inadequate training of professionals** in handling specific cases, such as autism, people with disabilities or victims, which contributes to the perception that public services are **ill-equipped to resolve issues effectively**. Specifically, the **lack of proper training and preparation of public servants on GVB** were frequently mentioned by the participants as the reason behind the **insufficient guidance** by the public legal experts. An employee working with GVB victims stated: "*there is no training or awareness in the [police] stations. There are some people who try to help but most of them don't care. There are disagreements between the court, the police, and the lawyers, and they "exchange" paperwork and responsibilities without results. Whether it is violence, psychological abuse, whether it is at work, whether it is with a child in a school, whether it is with a woman, there is no preparation, they tend not to believe, they tend to minimise, and no aid is provided.*" Families with autistic members also confirm the **lack of preparedness of public institutions**, especially health institutions' staff to deal with people with disabilities, highlighting the lack of trust in public servants. A family member reported that: *In the emergency, my husband had to intervene [...] not only the doctors but all the staff working in the health centres, they are not prepared. And of course, the red dot is not fulfilled most of the times either".*

Additionally, another issue that was mentioned as a reason of untrust to the public services, and specifically the health system, was the fact that many **health professionals are unaware and uninformed about the specific protocols** they have to follow when facing special cases, such as people with autism. A participant stated: "*Not only the red dot [indicator that the patient has autism and needs priority care] as such that appears if the professional doesn't really know what autism is.*"

3.6.3　Implications

The study conducted revealed that underrepresented groups face challenges, difficulties, and on some occasions, they are being discriminated in terms of access and quality of public services provided to them. Based on the results and the qualitative and quantitative analysis, the following implications regarding the experiences and perceptions of underrepresented groups in relation to public services can be drawn:

Communication Barriers and Inefficiencies

The difficulties in reaching and interacting with public services indicate significant **communication barriers**. The reported unavailability and unhelpfulness of employees, delays, and bureaucratic procedures suggest a need for systemic changes to improve access and efficiency.

The **lack of technological proficiency among employees**, particularly older ones, hinders effective communication and service delivery. This gap highlights the necessity for enhanced training programs to boost digital literacy within public institutions and older citizens.

Furthermore, the **frustration over inefficiencies in face-to-face interactions**, contrasts sharply with the relatively smoother experiences with online health services, which for most of the participants have improved the quality of service that they receive.

Issues Encountered by Vulnerable Groups

As concerns the underrepresented groups, the **inadequate empathy** and **professional handling** of GBV cases underscore a **systemic failure to address the needs of vulnerable populations.** The **lack of proper training and understanding** among public service employees aggravates the trauma faced by GBV victims.

In addition, for non-native speakers and non-EU citizens, the **prolonged processing times**, **unclear eligibility criteria**, and **inadequate support for non-native speakers** reveal significant **accessibility issues**. The poor translation services and the requirement to navigate services in Catalan, for instance, create **substantial barriers for non-native speakers**.

As concerns the people with autism, the **communication gaps** in public healthcare services, particularly in emergency situations, highlight the urgent need for **specialized training and protocols**. The difficulties faced by families in reaching doctors and the lack of customized care for autistic individuals emphasize the **system's unpreparedness**.

Clarity and Consistency of Service Responses

The frequent reports of **inconsistent procedures**, **miscommunication**, and **delayed responses** suggest deep-rooted bureaucratic inefficiencies. This inconsistency affects the clarity and quality of services, leading to repeated visits and frustration among citizens.

Furthermore, the **technological incompetence** of public servants further complicates service responses, particularly affecting **elderly and technologically inexperienced citizens**. This issue necessitates comprehensive training programs to improve service delivery.

Trust in Issue Resolution

The widespread **lack of trust in public services' ability** to resolve issues effectively, often due to unhelpful attitudes and inadequate training, highlights the **need for systemic reforms**. The empathetic and proactive assistance reported by some participants is an exception rather than the norm.

Also, the inadequate support and respect for GVB victims and families with autistic members indicate a significant **gap in the training and preparedness of public servants**. This gap further diminishes trust in public services among vulnerable populations.

4 Specifications, guidelines and suggestions on Inclusive and Bias-free AI Tools

4.1 Overview

As explained in the previous chapters, the HRF is a framework that incorporates different methodologies, including an extensive theoretical literature review on Human Rights, EU values, and how biases can compromise them. It also examines current regulations and ethical standards that AI tools should follow to comply with European provisions. These are then translated into technical specifications. In addition, field research has been conducted on citizens, with a focus on underrepresented groups, to provide insights into how people perceive biases in their everyday encounters with government services. This chapter gathers the guidelines, specifications, and recommendations produced during the creation of the HRF to set the basis for developing lawful, unbiased (as much as possible), inclusive, and ethical AI solutions

4.2 HRF's technical specifications

In order to build an AI tool that follows the provisions of the HRF, it is crucial to have a list of concrete specifications. This list is presented below and has been structured in accordance to reach one of the 15 dimensions identified for the HRF:

1. Regulations compliance

- Establish a comprehensive compliance framework that identifies and addresses all relevant laws and regulations.
- Implement processes for regularly reviewing and updating compliance measures as regulations evolve.
- Conduct independent audits and assessments to verify compliance.

2. Fairness

- Develop algorithms and models that are trained on diverse and representative data sets to reduce potential biases.
- Implement techniques for detecting and mitigating unfair biases in the AI system's decision-making processes.
- Conduct rigorous testing and validation to ensure fair and equitable treatment across different groups and communities.

3. Non-discrimination - inclusivity

- Prohibit the use of protected characteristics (e.g., race, gender, age, disability) in the AI system's decision-making processes unless strictly necessary and justified.
- Ensure accessibility and usability for individuals with disabilities or other specific needs.
- Promote diversity and inclusivity in the development and deployment teams, and actively seek input from underrepresented groups.

4. Human oversight

- Establish clear roles, responsibilities, and decision-making authorities for human oversight of the AI system.
- Implement mechanisms for human intervention, review, and override of the AI system's decisions when necessary.

5. Human-Centered design

- Conduct user research and usability testing throughout the design and development process to ensure the AI system meets human needs and expectations.
- Incorporate principles of accessibility, inclusivity, and user-friendly interfaces in the design.
- Regularly solicit feedback from end-users and stakeholders to inform iterative improvements.

6. Privacy

- Implement data minimization practices to collect and process only the personal data necessary for the AI system's intended purpose.
- Anonymize or pseudonymize personal data whenever possible to protect individual privacy.
- Establish robust security measures, such as encryption and access controls, to safeguard personal data.

7. Explainability - transparency

- Develop interpretable and explainable AI models that can provide clear and understandable rationales for their decisions and outputs.
- Implement mechanisms for providing explanations and interpretations to relevant stakeholders upon request.
- Conduct regular audits and assessments to verify the transparency and explainability of the AI system.

8. Auditing – continuous monitoring

- Establish a comprehensive monitoring and auditing framework that covers the AI system's inputs, processes, outputs, and impacts.
- Implement automated monitoring tools and dashboards to track key performance indicators and detect anomalies or deviations.

9. Accountability

- Clearly define roles, responsibilities, and lines of authority for the development, deployment, and operation of the AI system.
- Implement robust governance structures and processes for decision-making, oversight, and reporting.
- Establish mechanisms for addressing failures, incidents, or misconduct, including corrective actions and potential consequences.

10. Safety and security

- Conduct thorough risk assessments and impact analyses to identify potential safety and security vulnerabilities.
- Implement robust security measures, such as access controls, encryption, and secure development practices, to protect against malicious attacks or misuse.
- Establish incident response and business continuity plans to mitigate and recover from safety or security breaches.

11. Claims and redress

- Establish clear and accessible channels for individuals to submit claims or grievances related to the AI system's decisions or impacts.
- Implement fair and impartial processes for reviewing and adjudicating claims or grievances.

12. Public engagement

- Conduct public consultations and stakeholder engagement activities throughout the AI system's development and deployment phases.
- Establish advisory boards or committees that include representatives from diverse stakeholder groups and communities.
- Implement mechanisms for regularly communicating updates, developments, and potential impacts to the public.

13. Awareness

- Develop educational materials and resources to promote understanding of the AI system's capabilities, limitations, and potential impacts.
- Conduct awareness campaigns and training programs for relevant stakeholders, including end-users, operators, and decision-makers.

14. Societal benefit and well-being

- Establish a framework for assessing the AI system's potential societal impacts, both positive and negative, in the short and long term.
- Prioritize the development and deployment of AI systems that have clear and demonstrable societal benefits.
- Implement mechanisms for monitoring and evaluating the AI system's actual societal impacts over time.

15. Sustainability

- Conduct environmental impact assessments to identify and mitigate the AI system's potential negative effects on the environment.
- Prioritize energy-efficient and resource-optimized algorithms, hardware, and infrastructure.

4.3 Guidelines for creating self-assessment tools

After following the HRF to build an AI tool, it is crucial to set a motoring and assessment plan and ensure compliance. This plan will contain specific tools that will allow the user to check if the AI

tool operates according to the provisions of the HRF and keep it in track. Starting from a general overview of how these assessment tools should be structured, a set of guidelines is provided. .

Clear Objective: A clear objective of each tool should be documented, defining goals and outcomes that it aims to achieve in alignment with HRF.

Coverage: The tools should cover all relevant dimensions defined in HRF. A checklist that documents the coverage of every dimension must exist.

Easily interpretable questions and results: All questions, instructions, and result explanations should be written in straightforward, non-technical language. Moreover, visualization aids (e.g., bar graphs, pie charts, spider charts) should be exploited to make data easier to understand immediately.

Promote transparency: The assessment process and criteria should be transparent to all stakeholders. The assessment methodology and criteria should be published with clear explanations on how results are derived.

Enable regular updates: Each tool should be able to be updated on a regular basis, to reflect new developments and emerging best practices. This will be implemented by setting up a review process and incorporating feedback loops to keep the tools current and relevant.

Promote ease of use: The tools should be designed to be user-friendly and accessible to non-experts. Clear instructions should be provided, intuitive interfaces and examples to guide users through the assessment process.

Provide actionable insights: The tools should deliver clear, actionable recommendations based on assessment results. One way of implementing this is by including specific guidance on how to address identified issues and improve practices.

Highlight compliance with standards: The tools should be compliant with relevant standards and let users be aware of this.

Encourage accountability: The tools should foster a culture of accountability and continuous improvement, which can be implemented by incorporating mechanisms for holding individuals or teams accountable for follow-up actions.

Follow State of the Art, aim to Beyond State of the Art: Relevant tools that are considered State of the Art (SOTA) should be studied identifying best practices that should be followed and pinpointing flaws that should be mitigated. Such tools are the Assessment List for Trustworthy AI (ALTAI), the South Korea's Trustworthiness Verification self-diagnosis system (https://aitrustops.or.kr/common/aitrust/start.do) and many others including the several Human Rights Impact Assessment Tools (HRIA) that exist and usually have narrower scope.

4.3.1   Technical Suggestions of the self-assessment tools

From a technical point of view, the self-assessment tools that will be implemented should adhere to modern technologies and application development practices resulting in robust, effective and

user-friendly applications that can adapt to the evolving needs of their users. So, the following technical suggestions are given:

Modular architecture: The tools should be designed with a modular architecture so as to allow easy updates and customization. The usage of microservices or plugin-based architecture should enable adding or modifying features with minimal disruption of the entire system.

User friendly interface: Apply user-centered design principles, conduct usability testing, and iterate based on feedback to enhance the user experience.

Data encryption and security: Security and privacy should be ensured by employing robust encryption and security measures. Regular security audits should be undertaken to protect sensitive information.

User authentication: Provide an intuitive user experience for registration and login processes, supporting modern authentication methods such as multi-factor authentication (MFA) and single sign-on (SSO).

Design for scalability: Design the tools to be scalable to handle varying amounts of data and user interactions.

Open-source preference: Utilize free/open-source software (FOSS) to build, manage, and secure the self-assessment tools, ensuring flexibility and cost-effectiveness. Avoid vendor lock-in.

4.4 Recommendations to achieve inclusivity

Leveraging the insights from the case studies we ran in the three pilot countries (Greece, Slovenia, Spain), we extracted some recommendations to guide the public services through the most problematic areas that were identified during the focus groups. These recommendations either arose during the interaction with the citizens or were directly proposed by them, complementary fed into the HRF. The aim was to indicate the most common issues encountered by the citizens that require extra attention when building a public service tool, in order to minimise as much as possible potential biases at the source. Since the focus groups were conducted in three different countries, the recommendations generally address the problems faced by vulnerable groups in these regions. The suggestions are categorized by policy areas, aiming to prioritize issues and suggestions in a structured and focused manner, which will lead to more effective and equitable design and delivery of the HRF.

Training and Professional Development

Enhanced Training Programs: Implement comprehensive training programs for public service employees to improve their communication skills, technological proficiency, and understanding of the needs of vulnerable populations.
Specialized Training for Handling GBV Cases: Develop specialized training programs for public service employees to handle GBV cases with empathy and professionalism.
Customized Care for Autistic Individuals: Establish specialized training for healthcare providers to understand and effectively manage the needs of autistic individuals.

Improving Technological Literacy: Enhance the technological literacy of public servants through regular training and support.

Service delivery and Process Improvement

Streamlined Face-to-Face Services: Review and revise in-person service protocols to reduce delays and bureaucratic hurdles. Establish clear guidelines and accountability measures to ensure efficient and respectful service delivery.

Standardizing Procedures: Standardize procedures across public services to reduce inconsistencies and ensure clear, transparent processes.

Customized Care for Autistic Individuals: Implement protocols for priority care and ensure consistent, familiar healthcare providers for these patients.

Accessibility and Inclusivity

Improved Translation Services: Invest in high-quality translation services and ensure that all public service information is accessible in multiple languages. Simplify the navigation of public websites for non-native speakers.

Customized Care for Autistic Individuals: Ensure that digital services are user-friendly and accessible to all citizens, including the elderly and technologically inexperienced.

Addressing Specific Needs of Vulnerable Populations: Ensure that public servants are adequately trained and prepared to address the specific needs of vulnerable populations. Implement feedback mechanisms to continuously improve service delivery based on citizen experiences.

Cultural Change and Public Trust

Empathy and Professionalism: Foster a culture of empathy and professionalism within public services. Recognize and reward employees who demonstrate exceptional service and support for citizens.

Addressing Specific Needs of Vulnerable Populations: Ensure that public servants are adequately trained and prepared to address the specific needs of vulnerable populations. Implement feedback mechanisms to continuously improve service delivery based on citizen experiences.

Technological Enhancement and Digital Services

Improved Translation Services: Invest in high-quality translation services and ensure that all public service information is accessible in multiple languages. Simplify the navigation of public websites for non-native speakers.

Improving Technological Literacy: Enhance the technological literacy of public servants through regular training and support.

Customized Care for Autistic Individuals: Ensure that digital services are user-friendly and accessible to all citizens, including the elderly and technologically inexperienced.

5 Conclusions

This report has comprehensively detailed the final version of the HRF, a holistic framework designed to support lawful, ethical, inclusive, and transparent AI governance. The HRF provides a well-defined and detailed structure aimed at promoting fairness and inclusivity in AI-driven democratic processes. It includes clear objectives, goals, and technical description is provided, focusing on the 15 dimensions related to AI governance when AI is involved to a lesser or to a greater extent at services that are provided to the public by government entities.

Through an extensive theoretical analysis and literature review, the report analyses the definitions and perceptions of bias and discrimination in public services, which sets the foundation for understanding the challenges and necessities of regulating AI in democracy. Furthermore, the fieldwork conducted in T2.1 and focus groups with underrepresented groups provided valuable insights in the perceptions of citizens on non-AI public services, with the dual approach of quantitative (conjoint analysis) and qualitative (content analysis) methods offering a comprehensive understanding of the data collected. The last section of this report builds on the previous sections and provides practical guidelines and suggestions for developing inclusive and bias-free AI tools. These include concrete specifications for the HRF's 15 dimensions and recommendations for building self-assessment tools to ensure trustworthy AI. Specific technical suggestions and strategies for promoting inclusivity in AI applications are provided. These recommendations are crucial for fostering an equitable and just AI-driven democratic environment.

The field study conducted in the three case studies—Greece, Spain, and Slovenia—revealed several key challenges that citizens face when interacting with public services. A pervasive issue across all three countries is communication barriers and inefficiencies. Participants frequently reported the unavailability and unhelpfulness of employees, intensifying frustration during face-to-face interactions. A significant challenge is the lack of technological proficiency among public service employees at some public institutions, which obstruct effective communication and service delivery, especially in public healthcare services. This technological incompetence is particularly problematic for elderly and technologically inexperienced citizens. Another critical challenge reported is the inadequate empathy and professional handling of gender-based violence (GBV) cases, highlighting a systemic failure to meet the needs of vulnerable populations. The lack of proper training and understanding among public servants leads to inconsistent procedures, miscommunication, and delayed responses, which are indicative of deep-rooted bureaucratic inefficiencies. Furthermore, the inadequate support and respect for GBV victims and families with autistic members emphasize a significant gap in the training and preparedness of public servants, further diminishing trust in public services. Overall, these challenges illustrate the pressing need for specialized training, improved communication protocols, and enhanced technological competencies among public service employees to better serve and support vulnerable populations. These key challenges complementary supported the development of HRF, especially through the provision of the HRF's 15 dimensions.

To build an AI tool compliant with the HRF, it is essential to follow the list of specifications addressing the 15 dimensions identified in the framework. For regulations compliance, a robust framework must be established to identify and address relevant laws, with regular reviews and independent audits to ensure ongoing adherence. Fairness involves developing diverse, unbiased algorithms and conducting rigorous testing to ensure equitable treatment. Non-discrimination and inclusivity require prohibiting the use of protected characteristics in decision-making unless justified, ensuring accessibility, and promoting diversity in development teams. Human oversight is maintained by defining roles and implementing intervention mechanisms. A human-centered design approach ensures the system meets user needs, incorporating accessibility and iterative feedback. Privacy is safeguarded through data minimization, anonymization, and robust security measures. Explainability and transparency are achieved with interpretable models and regular audits. Continuous monitoring and auditing frameworks track the system's performance and detect anomalies. Accountability is ensured through defined roles, robust governance, and mechanisms for addressing failures. Safety and security involve thorough risk assessments, security measures, and incident response plans. Claims and redress mechanisms allow for grievance submission and fair adjudication. Public engagement includes consultations, advisory boards, and regular communication updates. Awareness efforts promote understanding of the AI system's capabilities and impacts through educational materials and campaigns. Societal benefit and well-being are prioritized by assessing and monitoring the system's impacts, while sustainability focuses on mitigating environmental effects and prioritizing energy-efficient algorithms and infrastructure.

By incorporating these objectives, the HRF seeks to establish a regulatory framework that ensures lawful, ethical, inclusive, and transparent AI governance. The HRF envisions a framework that will be regularly updated and reinforced in line with AI technological advancements, improving public services while safeguarding the rights and interests of all citizens and fostering a more equitable and just society.

6 References

Alexander, M. (2010). *The color of justice. The new Jim crow: Mass incarceration in the age of colorblindness, 59-62.*

Amnesty International. (2016). *Living in insecurity: How Germany is failing victims of racist violence. Amnesty International.*

Boogaard, B., & Roggeband, C. (2010). *Paradoxes of Intersectionality: Theorizing Inequality in the Dutch Police Force through Structure and Agency. Organization, 17(1), 53-75. https://doi.org/10.1177/1350508409350042*

Crul, M., & Doomernik, J. (2003). *The Turkish and Moroccan Second Generation in the Netherlands: Divergent Trends between and Polarization within the Two Groups 1. International Migration Review, 37(4), 1039-1064.*

Equality and Human Rights Commission. (2013). *Stop and think again: towards equality in police PACE stop and search.*

FRA, E. (2013). *EU LGBT Survey: European Union lesbian, gay, bisexual and transgender survey– Results at Glance. Luxembourg: Publications Office of the European Union. 31 s.*

Gerards, J., & Xenidis, R., *Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law (European Commission, 2021).*

Gillborn, D., Rollock, N., Vincent, C., & Ball, S. J. (2012). *Gender intersections in the educational experiences of the Black middle class.*

Giulietti, C., Tonin, M., & Vlassopoulos, M. (2019). *Racial discrimination in local public services: A field experiment in the United States. Journal of the European Economic Association, 17(1), 165-204.*

Harry, B., & Klingner, J. (2014). *Why are so many minority students in special education?: Understanding race and disability in schools. Teachers College Press.*

Hudson, J., & Lowe, S. (2009). *Understanding the policy process: analysing welfare policy and practice. Policy Press.*

Louviere, J. J. (1994). *Conjoint analysis. Advanced methods of marketing research, 223-259.*

Mottet, L., & Ohle, J. M. (2003). *Transitioning our shelters. National Gay and Lesbian Task Force Policy Institute. Retrieved January 8, 2022.*

Oberti, M. (2007). *Social and school differentiation in urban space: Inequalities and local configurations. Environment and Planning A, 39(1), 208-227.*

Open Society Institute. (2009). *ETHNIC PROFILING IN THE EUROPEAN UNION: Pervasive, Ineffective, and Discriminatory.*

Rai, N., & Thapa, B. (2015). A study on purposive sampling method in research. Kathmandu School of Law, 5(1), 8-15. Retrieved from https://d1wqtxts1xzle7.cloudfront.net/48403395/A_Study_on_Purposive_Sampling_Method_in_Research-libre.pdf?1472466369=&response-content-disposition=inline%3B+filename%3DA_STUDY_ON_PURPOSIVE_SAMPLING_METHOD_IN.pdf&Expires=1717421102&Signature=RDczqfF~4QHWrldZMi6zAsp1YBeaXvWst8pERuLzOTGrawL3zcIy9fWaZM6faxx7uMUUjOOweCUYipzaSTyt7~JkMRW7tBYno8ctRxwfYGYwprnCAHkKIp6hfTn9D7yU7jFKvGA4buFSnZWSty07A5gjOw49gG6HV4yLYfMfgocNWeGkICIKC7lum5QiarWYyOBMy7S83~sbkOocMD5OX6v3C0lPF-kfFaJet3bXy61Fsie8kjkj4WFM4iWjMcbbO7dvNkZb5UyV2ugwnTP6N2OGMcsO7O7cri99dv0kJXdJk-9ZBFxWXcYwmKXn0NHNTBLuZA0kRNNBhp~iO5u-DQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Rivers, I., & D'Augelli, A. R. (2001). The victimization of lesbian, gay, and bisexual youths. Lesbian, gay, and bisexual identities and youth: Psychological perspectives, 9, 187-210.

Rivenbark, J. G., & Ichou, M. (2020). Discrimination in healthcare as a barrier to care: experiences of socially disadvantaged populations in France from a nationally representative survey. BMC public health, 20(1), 31.

Silver, H., & Danielowski, L. (2019). Fighting Housing Discrimination in Europe. Housing Policy Debate, 29(5), 714-735. https://doi.org/10.1080/10511482.2018.1524443

Van Lancker, W. (2013). Putting the child-centred investment strategy to the test: evidence for the EU27. European Journal of Social Security, 15(1), 4-27.

ANNEX 1. Focus Groups Guidelines and Scenarios

For the Moderator:

Present the general topic: "Access to public services by citizens, recording their experience, and evaluating the process and outcomes. Proposals for optimization and topics to avoid."
Ask each participant to briefly introduce themselves and share their experience with public services.
– "Please tell us your name and the first word that comes to mind when you hear 'public service.' We use this question to engage the participants."
Then you can discuss the scenarios you have chosen.
No one should monopolize the conversation. If you see the discussion veering off track, gently steer it back on course.

You can delete the scenarios that do not interest you.

It is important for participants to share their experiences and how those experiences made them feel.

The scenarios are meant to start the conversation. If participants want to share experiences with other services, let the conversation flow.

Scenarios

Online services

Scenario A: Applying for Unemployment Benefits Online

You recently lost your job and applied for unemployment benefits through the government's online portal. However, you encountered numerous technical glitches and errors during the application process. Despite your repeated attempts to seek assistance through the provided helpline or chat support, you felt ignored and left without proper guidance.

Scenario B: Applying for Disability Benefits Online

As a person with a disability, you attempted to apply for disability benefits through the government's online platform. However, you found that the website was not accessible for individuals with certain disabilities, making it nearly impossible for you to complete the application process independently. Despite reaching out to request accommodations, you received no response or assistance.

Scenario C: Accessing Online Healthcare Services

You attempted to schedule a health appointment with a government healthcare provider through their online portal. However, you encountered barriers such as complex registration requirements and inaccessible features for individuals with limited internet access or technological literacy. Despite your efforts to seek alternative means of accessing healthcare services, you felt like the system was not designed to be accesses easily by all people.

Scenario D: Applying for Social Assistance Programs Online

You applied for social assistance programs such as food assistance or housing subsidies through the government's online application portal. However, you encountered issues such as lengthy

processing times, unclear eligibility criteria, and inadequate support for non- native speakers or individuals with limited English proficiency. Despite meeting the necessary qualifications, you felt like the system was stacked against you, resulting in delayed or denied assistance.

F2f services

Scenario A: Applying for Identity Documents

You visited the government's local office to apply for a new identification card or passport. Despite having all the required documents, you were met with skepticism and subjected to excessive questioning about your identity, nationality, or immigration status. You felt like you were critisised based on your appearance or background.

Scenario B: Seeking Legal Assistance

You went to a government-funded legal aid clinic to seek assistance with a legal issue. However, when you arrived for your appointment, you were met with skepticism and disbelief about the validity of your case. Despite trying to explain your situation, you were met with condescension and reluctance to provide assistance.

Scenario C: Accessing Social Services

You visited a government social services office to inquire about financial assistance or support programs due to a recent hardship. However, you found that the staff were unresponsive and unwilling to listen to your concerns, treating you with indifference and neglect. Despite being eligible for assistance, you felt like your needs were being disregarded.

Scenario D: Attending Job Training Programs

You enrolled in a government-sponsored job training program aimed at enhancing employability skills. However, during the training sessions, you noticed that certain participants were given preferential treatment and more opportunities for advancement, while you were overlooked and sidelined. This unequal treatment made you feel like you were being treated differently.

Scenario E: Seeking Housing Assistance

You visited a government housing office to apply for affordable housing options. However, you encountered barriers such as long wait times, bureaucratic hurdles, and unhelpful staff who seemed indifferent to your housing needs. Despite your urgent situation, you felt like your concerns were trivialized and overlooked.

Scenario F: Accessing Disability Services

As a person with a disability, you sought assistance from a government disability services center to access accommodations or support services. However, you encountered physical barriers such as inaccessible facilities and lack of accommodations for your specific needs. Despite advocating for your rights, you felt like your disability was being disregarded.

Scenario G: Accessing Healthcare Services

You recently visited a public hospital for a health check-up. While waiting to be seen, you noticed that other patients were called in before you, even though you arrived earlier. You felt like the staff ignored you, even though you tried to explain your situation to the healthcare professionals.

ANNEX 2. Focus Groups Questionnaire

E1. A citizen contacts a public service for a matter concerning him/her.

Please carefully read the following situations and mark with the number one (1) the one that you think happens most often. Then, for the next most frequent, mark the number two (2), and so on until the number eight (8).

	Citizen A communicates **relatively easily** with the service for an economic or work-related issue. They **struggle to understand the response** and have **little trust** regarding the resolution of the issue.
	Citizen B communicates **relatively easily** with the service for a **social issue**. They **relatively understand** the response and have **considerable trust** regarding the resolution of the issue.
	Citizen C communicates **relatively easily** with the service for a **personal issue**. They **relatively understand** the response but have **little trust** regarding the resolution of the issue.
	Citizen D communicates **relatively difficultly** with the service for an **economic or work-related issue**. They **relatively understand** the response and have **considerable trust** regarding the resolution of the issue.
	Citizen E communicates **relatively difficultly** with the service for a **personal issue**. They **struggle to understand** the response but have **considerable trust** regarding the resolution of the issue.
	Citizen F communicates **relatively difficultly** with the service for **a social issue**. They **struggle to understand** the response and have **little trust** regarding the resolution of the issue.
	Citizen G communicates **relatively difficultly** with the service for a **personal issue**. They **relatively understand** the response but have **little trust** regarding the resolution of the issue.
	Citizen H communicates **relatively easily** with the service for **a personal issue**. They **struggle to understand** the response but have **considerable trust** regarding the resolution of the issue.

D1. Employment Status

Unemployed	Long-term Unemployed	Temporarily Employed	Employed	Retired

D2 Gender

Male	Female	Prefer not to answer

D3 Age

-25	26-40	41-65	66+

D4 Educational Level

Compulsory Education	High School/Technical School	University/College	Postgraduate Degree

E1.1/ E1.2 Public Issues

	Personal/Close Personal or family environment/Work/Health	Residential Area/Public Transport/Parking/Green Spaces	Broader Political/Economic/Social Issues
Priority **[1/2/3]**			
Severity Level **[Low/ Medium/HIgh]**			

Knowledge of Applications

E2. Are you aware of applications/services for Citizens via the internet?

1		YES
2		SOMEWHAT
3		NO
9		NA

Frequency of Application Use

E3.1/E3.2 Have you used applications/services for Citizens provided by your Municipality or any other entity (such as gov.gr) via the internet? How often? [1/2/3/4/5]

		Frequency [FOR EACH]	
1			YES [my Municipality]
2			YES [my Municipality and another entity]
3			YES [another entity]
4			NO
9			NA

Means of Communication

E6. What is the primary form of communication with Citizen services provided by your Municipality or any other entity?

1		In Person
2		Online
3		Both in Person and Online
9		NA

Familiarity with the Internet and Applications in General

E4. How familiar are you with using the internet and electronic applications?

1		VERY
2		QUITE
3		SOMEWHAT
4		LITTLE
5		NOT AT ALL

Frequency of Communication

E5. How often have you needed to contact Citizen services provided by your Municipality or any other entity?

1		NEVER
2		RARELY

3		OFTEN
9		NA

Evaluation of Applications [Have you conducted an evaluation?]

E7. Have you used any service evaluation application?

1		YES
2		NOT OFTEN
3		NEVER
9		NA

Interest in Politics

E8. How interested are you in politics in general?

1		VERY
2		QUITE
3		SOMEWHAT
4		NOT AT ALL

Actions for an Area Issue

E9. For a pressing issue in your area that concerns you, which of the following actions do you usually choose?

1		NOTHING, I DO NOT GET INVOLVED
2		I LET THE AUTHORITIES DO THEIR JOB
3		I PERSONALLY CONTACT THE AUTHORITIES
4		I CONTACT A TV CHANNEL, NEWSPAPER
5		I MOBILIZE THROUGH SOCIAL NETWORKS (FB, Instagram, etc.)

6		I PARTICIPATE WITH OTHERS IN PROTESTS

Level of Information

E10. How informed are you about politics in general?

1		VERY
2		QUITE
3		SOMEWHAT
4		NOT AT ALL