



# CERTAIN

**Certification for Ethical and Regulatory Transparency  
in Artificial Intelligence**

## **D3.3: SOCIAL ANALYSIS & AI ACCEPTANCE MODELLING**

*Social Acceptance of AI: Cross-Country Qualitative  
Insights*



Co-funded by  
the European Union

Project funded by



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,  
Education and Research, EAER  
State Secretariat for Education,  
Research and Innovation SERI

<b>Work package</b>	WP 3
<b>Task</b>	T3.3
<b>Due date</b>	28/02/2026
<b>Submission date</b>	01/03/2026
<b>Deliverable lead</b>	University of Tartu (UT)
<b>Version</b>	1.0
<b>Authors</b>	Irina Urumova (INCO), Georgios Myrogiannis (INCO), Victoria Sovatzoglou (INCO); Fenia Giannakopoulou (ANAD)
<b>Reviewers</b>	Fenia Giannakopoulou (ANAD), Matic Kadiš (UM), Zouhair Haddi (NVISION)
<b>Abstract</b>	This study examines social acceptance of artificial intelligence (AI) across five countries based on focus group discussions with diverse participant groups. The analysis shows that AI is evaluated primarily through its institutional use rather than as a standalone technology. Across contexts, acceptance was consistently conditional and depended on trust in institutions, meaningful human oversight, data protection, accountability, and perceived fairness. While these expectations were shared, countries differed in which risks were most salient, reflecting variations in institutional experience and governance contexts. Overall, the findings indicate that social acceptance of AI is shaped less by technological performance than by how AI is governed, deployed, and embedded in decision-making processes affecting individuals and communities.
<b>Keywords</b>	Social acceptance, AI, Comparative Approach, Inclusion, Data handling, Governance, Data protection, Fairness

### Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1	04/02/2026	Draft	Irina Urumova (INCO)
V0.2	05/02/2026	Internal Review	(INCO) Georgios Myrogiannis, Victoria Sovatzoglou
V0.3	09/02/2026	1st edit/section 4.1 added by ANADELTA Integrated corrections after 1st review by ANAD	Fenia Giannakopoulou (ANAD), Irina Urumova (INCO)
V0.4	11/02/2026	2nd edit/ Integrated corrections based on the second review by UM.	Irina Urumova (INCO); Victoria Sovatzoglou (INCO); Fenia Giannakopoulou (ANAD)
V0.5	18/02/2026	3 <sup>rd</sup> edit/ Integrated corrections based on the third review by NVISION.	Irina Urumova (INCO); Fenia Giannakopoulou (ANAD)
V0.6	02/03/2026	Finalization of the deliverable/Fine tuning	Irina Urumova (INCO); Victoria Sovatzoglou (INCO); Georgios Myrogiannis (INCO)

**Grant Agreement No:** 101189650 | **Topic:** HORIZON-CL4-2024-DATA-01-01  
**Call:** HORIZON-CL4-2024-DATA-01. | **Type of action:** HORIZON-IA

## REVIEWERS COMMENTS

Fenia Giannakopoulou	ANADELTA	Date of review: 09.02.2026
<p><b>General comments</b></p> <p>The deliverable was very interesting to read and reflected a high level of quality in both its design and execution. The results of the focus group discussions were presented in a clear, coherent, and well-structured manner, making the findings easy to follow and understand. The analysis demonstrated thoughtful interpretation and provided meaningful insights into the views of the individuals involved. As a suggestion for further strengthening the deliverable, it might be useful to consider including one or more tables or visual figures to better highlight the key commonalities and differences among the countries that conducted the FGDs. Such visual elements could enhance clarity, facilitate comparison, and make the main messages even more accessible to the reader.</p>		

Matic Kadiš	UM	Date of review: 11.02.2026
<p><b>General comments</b></p> <p>The deliverable presents high-quality qualitative work, particularly in the "Social Analysis" components (Sections 6 &amp; 7). However, there are some administrative and structural discrepancies that should be resolved before submission, detailed in the specific comments below.</p> <p>My main observation is that the report is currently text heavy. I suggest improving clarity by using <b>summary tables</b> where applicable. For example, in <b>Section 6 (Cross-Cutting Observations)</b>, a "Comparative Summary Matrix" (listing Countries vs. Key Themes like Trust, GDPR, etc.) would allow the reader to instantly visualize the "EU vs. Non-EU" differences and significantly enhance the readability of the findings.</p> <p><b>Specific comments</b></p> <p><b>Cover Page</b></p> <ul style="list-style-type: none"> <li>Lead Beneficiary: The Grant Agreement lists 10 - UT (University of Tartu) as the Lead Beneficiary for D3.3. Please verify if this is still the case and update the "Deliverable Lead" field accordingly. If INCOM is the lead, please disregard.</li> <li>Reviewers List: A reminder that the list of reviewers is currently incomplete/contains placeholders. Please update this before final PDF conversion.</li> </ul> <p><b>Acronyms</b></p> <ul style="list-style-type: none"> <li>Add missing acronyms used in the text: QETAM, TAM, UTAUT, BIU, PT, PU, PEOU, SI, FC, GI, GDPR, SME, e-ID, ETIS, EVK, LSE.</li> <li>Correct "CERI" to "SERI" (Swiss State Secretariat) to match the disclaimer.</li> </ul> <p><b>Section 4.1 (QETAM)</b></p> <ul style="list-style-type: none"> <li>The text states the study "follow-up will subsequently deploy" the QETAM tool but lacks detail on when. My suggestion would be to explicitly state in the text that while D3.3 establishes the theoretical framework for the model; the actual quantitative data collection and validation will be executed in a future task. This creates a clear link to ongoing work and justifies why quantitative results are not included in this report.</li> </ul> <p><b>Section 6 &amp; 7 (Analysis)</b></p> <ul style="list-style-type: none"> <li>The synthesis is good but lacks "voice". I suggest inserting anonymized, verbatim quotes from the FGD participants (e.g., illustrating specific fears regarding "loss of control" or "fairness"). This validates the themes and strengthens the empirical evidence.</li> </ul> <p><b>Limitations Section</b></p> <ul style="list-style-type: none"> <li>The text implies broad non-representation of vulnerable groups. However, the Spain protocol explicitly included "Older adults." Please clarify that older adults were represented, distinguishing them from "persons with disabilities" (who were not). If this was not the case in final implementation of Spain FGD, please ignore.</li> </ul> <p><b>Figure 1</b></p> <ul style="list-style-type: none"> <li>While generating examples using AI is acceptable, I recommend creating figures such as this one manually to ensure the best possible representation and accuracy. If retaining the AI image, please add "using AI" along with the specific model version used, if that information is available to you (e.g. GPT 5.2).</li> </ul> <p><b>References (p. 69)</b></p> <ul style="list-style-type: none"> <li>This section is currently a placeholder. Please remove or add the actual citations used</li> </ul> <p><b>Appendix A (p. 70)</b></p> <ul style="list-style-type: none"> <li>"Appendix A" is labeled twice (p. 33 and p. 70). Delete the placeholder on p. 70 and ensure the Table of Contents points to the Participant Breakdown (p. 33) as the correct Appendix.</li> </ul> <p><b>Appendix B</b></p> <ul style="list-style-type: none"> <li>Please change Slovenian FGD to reflect the latest version that we used. I've uploaded the document to the Martel SharePoint.</li> </ul> <p><b>General Formatting</b></p> <ul style="list-style-type: none"> <li>The text alignment is currently set to the left. Please change to "Justify" for a cleaner, professional appearance.</li> </ul>		

Zouhair Haddi	NVISION	Date of review: 17.02.2026
---------------	---------	----------------------------

This deliverable provides a clear and well-structured analysis of how lived institutional experiences shape social acceptance of artificial intelligence. The findings show that acceptance is conditional and consistently linked to expectations of accountability, fairness, transparency, and human oversight across countries and participant groups. Overall, it is a high-quality contribution that offers a strong basis for informing future research and policy. However, a few minor recommendations could further strengthen the deliverable before submission:

- It would be advisable to specify how the outputs of D3.3 will contribute to CERTAIN ongoing project activities especially WP5 and WP7.
- Participant information is provided in Appendix A, but a brief summary in the methodology introduction would facilitate comprehension.
- D3.3 suggests that QETAM deployment is planned as part of a future task; however, additional detail regarding its timeline and implementation context (i.e. under which WP or task) would be helpful.
- Correct the abbreviation of Perceived Ease of Use in the abbreviation table to PEOU
- The term AI Act is mentioned only once in D3.3 (excluding annexes), although the deliverable implicitly addresses several principles central to the regulation. Given that this is a public deliverable, making the connections to the EU AI Act more explicit would further enhance its policy relevance and strategic positioning.
- Adding tables in sections 6 and 7 to summarise the findings and provide clear comparisons across countries, participant groups, and participant groups within each country would enhance the clarity and interpretability of the deliverable.
- Figure 1 is present but not referenced in the text; adding a citation would improve clarity.
- The reference to the links between WP3 deliverables in the conclusion is valuable; introducing this clarification earlier in the document, such as in Section 2 (Purpose and Scope), would further enhance reader orientation.

Draft

## DISCLAIMER



The CERTAIN project received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No 101189650. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

## COPYRIGHT NOTICE

© 2024 – 2027 CERTAIN

Project funded by the European Commission in the Horizon Europe Programme		
<b>Nature of the deliverable:</b>	Social analysis and AI acceptance modelling -R-	
<b>Dissemination Level</b>		
<b>PU</b>	Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)	<input type="checkbox"/>
<b>SEN</b>	Sensitive, limited under the conditions of the Grant Agreement	
<b>Classified R-UE/ EU-R</b>	<i>EU RESTRICTED</i> under the Commission Decision <a href="#">No2015/ 444</a>	
<b>Classified C-UE/ EU-C</b>	<i>EU CONFIDENTIAL</i> under the Commission Decision <a href="#">No2015/ 444</a>	
<b>Classified S-UE/ EU-S</b>	<i>EU SECRET</i> under the Commission Decision <a href="#">No2015/ 444</a>	

- \* *R: Document, report (excluding the periodic and final reports)*
- DEM: Demonstrator, pilot, prototype, plan designs*
- DEC: Websites, patents filing, press & media actions, videos, etc.*
- DATA: Data sets, microdata, etc.*
- DMP: Data management plan*
- ETHICS: Deliverables related to ethics issues.*
- SECURITY: Deliverables related to security issues*
- OTHER: Software, technical diagram, algorithms, models, etc.*

## TABLE OF CONTENTS

<b>REVIEWERS COMMENTS</b> .....	3
DISCLAIMER.....	5
Copyright notice.....	5
<b>1. STRUCTURE OF THE DELIVERABLE</b> .....	<b>11</b>
<b>2. PURPOSE AND SCOPE</b> .....	<b>12</b>
<b>3. CONCEPTUAL STARTING POINTS</b> .....	<b>14</b>
<b>4. METHODOLOGY</b> .....	<b>15</b>
<b>5. THEMATIC ORGANISATION OF THE ANALYSIS</b> .....	<b>17</b>
<b>6. CROSS-CUTTING OBSERVATIONS</b> .....	<b>19</b>
6.1. Literature contextualisation of cross-cutting findings .....	19
6.2. Relevance to AI Act.....	19
6.3. Trust in Institutions .....	20
6.4. Human control over AI.....	22
6.5. Data protection and misuse concerns .....	24
6.6. Perceived fairness .....	25
<b>7. COMPARATIVE PERSPECTIVE</b> .....	<b>28</b>
7.1. Shared patterns across all countries and groups.....	28
7.2. Differences in emphasis across country contexts.....	29
7.3. Overall comparative insight by country and group .....	30
<b>8. QUANTITATIVE EXTENSION</b> .....	<b>33</b>
8.1. Measurement of AI acceptance through QETAM .....	33
8.2. Preliminary Results of the Technology Acceptance Survey .....	33
8.2.1 Survey Demographics .....	34
8.2.3 Overall BIU.....	37
8.2.4 Average Score Distribution by Factor.....	38
8.2.5 Factor Correlation Matrix .....	39
<b>9. REFLECTIONS FOR POLICY AND PRACTICE</b> .....	<b>41</b>
9.1. Implications for public communication.....	41
9.2. Implications for institutional use of AI.....	41
9.3. Implications for future regulation .....	42
<b>10. WAYS FORWARD</b> .....	<b>43</b>
<b>11. CONCLUSIONS</b> .....	<b>45</b>
<b>REFERENCES</b> .....	<b>46</b>
<b>APPENDIX A – FGD PARTICIPANT BREAKDOWN BY COUNTRY, TARGET GROUP AND GENDER</b> ...	<b>47</b>

<b>APPENDIX B - DATA COLLECTION TOOLS</b> .....	<b>48</b>
Armenia: FGD Protocol.....	48
Estonia: FGD Protocol .....	54
Kazakhstan: FGD Protocol .....	60
Slovenia: FGD Protocol.....	66
Slovenia: Survey questionnaire .....	71
Spain: FGD Protocol .....	76
<b>APPENDIX C – TECHNOLOGY ACCEPTANCE AND PERCEPTION OF AI: END-USER SURVEY</b> .....	<b>83</b>

Draft

## EXECUTIVE SUMMARY

This study examines social acceptance of Artificial Intelligence (AI) as it is perceived and discussed by members of the public and selected stakeholder groups across five country contexts: Armenia, Estonia, Kazakhstan, Slovenia, and Spain. Rather than assessing AI from a technical or legal compliance perspective, the study focuses on how people understand, assess, and respond to AI as it is embedded in institutional practices, public services, and decision-making processes that shape everyday life.

In doing so, the study strengthens the CERTAIN Project by complementing the legal and ethical analyses in D3.1 and D3.2 with empirically grounded insights into social perceptions of AI. By examining how trust, fairness, human oversight, and governance expectations shape acceptance across countries and groups, it adds a critical social dimension to the project's framework. These findings will inform the final guidelines and compliance tools, helping ensure that AI governance and deployment under CERTAIN are not only legally robust, but also socially responsive, thereby contributing to more fair and responsible AI across the EU and its partner contexts.

The analysis is based on Focus Group Discussions (FGDs) and adopts a qualitative, comparative approach. It explores how trust, perceived risks, and expectations of human oversight shape acceptance of AI across different institutional and social environments. The inclusion of both EU and non-EU countries reflects the transnational nature of AI governance and the EU's wider regulatory, neighbourhood, and partnership context, including concerns related to cross-border data flows and jurisdictional accountability.

Across all country FGDs, participants did not approach AI as a neutral or autonomous technology. Instead, AI was consistently evaluated through its institutional use and social consequences. Acceptance was framed as conditional, depending on who uses AI, for what purposes, and under what safeguards. No FGD expressed unconditional trust in AI, nor support for fully autonomous decision-making in contexts affecting rights, services, education, or care.

A strong cross-cutting finding is the expectation of meaningful human control. Participants across all countries emphasized that AI may assist or support decision-making, but should not replace human judgment. Human involvement was seen as essential for maintaining responsibility, contextual understanding, and the ability to review, correct, or challenge outcomes.

Trust in AI was closely linked to trust in institutions, rather than to confidence in the technology itself. Participants evaluated AI as part of institutional practice, expressing concern that AI could obscure responsibility or amplify existing institutional weaknesses if deployed without clear accountability. Where institutions were perceived as predictable, rule-bound, or effectively governed, discussions of AI were more pragmatic and focused on conditions rather than rejection.

Concerns about data protection and potential misuse featured prominently across all FGDs. Participants expressed anxiety about loss of control over personal data once it enters AI-supported systems, particularly regarding secondary use, unclear access, and cross-border data handling. These concerns were articulated in everyday terms and were not offset by references to efficiency or innovation alone.

Perceived fairness emerged as a central evaluative criterion. Participants questioned whether AI-supported decisions could be applied consistently, accounting for individual circumstances, and be meaningfully contested. Automated or rigid systems were widely viewed as unfair where they limited contextual judgment or made outcomes difficult to challenge.

Across participant groups, differences were primarily a matter of emphasis rather than normative standards. Professionals tended to focus on discretion, responsibility, and procedural clarity, while participants speaking from a general citizen perspective highlighted vulnerability, power imbalance, and loss of agency. Despite these differences, expectations regarding human oversight, data protection, fairness, and accountability converged strongly.

Overall, the findings indicate that social acceptance of AI depends less on technological capability than on governance, accountability, and institutional practice, and suggest that variation in social acceptance lies not in what people expect AI governance to achieve, but in what they fear may go wrong. Shared expectations provide a common normative baseline, while contextual differences reflect lived experience with institutions, services, and power. These insights are directly relevant for EU internal policy, external partnerships, and future research on socially grounded AI governance.

## LIST OF FIGURES

Figure 1. Non-hierarchical representation of cross-cutting themes emerging from fgds on social acceptance of ai (image generated with chatgpt 5.2) .....	32
Figure 2. Age groups of respondents .....	35
Figure 3. Gender distribution of respondents .....	35
Figure 4. Self-identification of disability among respondents .....	35
Figure 5. Highest level of education of respondents .....	36
Figure 6. Area of employment/study of respondents.....	36
Figure 7. Country of workplace of respondents.....	37
Figure 8. Overall biu value.....	37
Figure 9. Average score distribution by factor.....	38
Figure 10. Factor correlation matrix.....	39

Draft

## ABBREVIATIONS

<b>AI</b>	Artificial Intelligence
<b>EPCA</b>	Enhanced Partnership and Cooperation Agreement
<b>EU</b>	European Union
<b>FC</b>	Facilitating Conditions
<b>FGD</b>	Focus Group Discussion
<b>FRA</b>	Fundamental Rights Agency (EU)
<b>GI</b>	Gender Influence
<b>QETAM</b>	Quantitative Effect and Technology Acceptance Modelling
<b>PEOU</b>	Perceived Ease of Use
<b>PT</b>	Perceived Trust
<b>PU</b>	Perceived Usefulness
<b>SERI</b>	Swiss State Secretariat for Education, Research and Innovation
<b>SI</b>	Social Influence
<b>TAM</b>	Technology Acceptance Modelling
<b>UTAUT</b>	Unified Theory of Acceptance and Use of Technology
<b>WP</b>	Work Package

Draft

## 1. STRUCTURE OF THE DELIVERABLE

This deliverable is structured to guide the reader from conceptual framing and methodological grounding to empirical findings and forward-looking implications.

The report opens with a Purpose and Scope section, outlining why social acceptance of AI is examined, the country coverage, and the analytical limits of the study. This is followed by Conceptual Starting Points, which situate social acceptance in relation to trust, perceived risks, and expectations of human oversight.

The Methodology section describes the qualitative research design, including participant selection, focus group implementation, analytical approach, and acknowledged limitations. Particular attention is paid to the purposive sampling strategy and to transparency regarding representation gaps.

The core of the report presents the FGD findings, organised around a set of Cross-Cutting Thematic Observations. These sections examine institutional trust, human control over AI, data protection and misuse concerns, and perceived fairness, drawing exclusively on FGD material to identify patterns that recur across countries and participant groups.

This is followed by a Comparative Perspective section, which builds on the cross-cutting analysis to systematically examine similarities and differences across country contexts and participant groups. This section highlights where expectations converge and where emphasis varies, reflecting differences in institutional experience, governance contexts, and social positioning.

The report then moves to Reflections for Policy and Practice, drawing out implications for public communication, institutional deployment of AI, and regulatory design, followed by a Ways Forward section identifying priority avenues for further research.

The deliverable concludes with a Conclusion, summarising key insights and outlining next steps, supported where relevant by annexes detailing methodological tools and supplementary materials.

## 2. PURPOSE AND SCOPE

This study explores **social acceptance of artificial intelligence (AI)** as it is perceived and discussed by members of the public and selected stakeholder groups across multiple country contexts. Rather than assessing the technical performance or legal compliance of AI systems, the study focuses on how people understand, evaluate, and respond to the use of AI in everyday institutional settings, including public services, education, healthcare, and other domains affecting rights, opportunities, and access to services.

Social acceptance is examined because emerging AI governance debates increasingly recognise that technical safeguards and formal regulation alone are insufficient if AI systems are perceived as untrustworthy, unfair, or unaccountable in practice. Understanding how different groups interpret AI use, articulate risks, and define acceptable conditions provides essential insight into the societal foundations of effective and legitimate AI governance.

The analysis is based on focus group discussions (FGDs) conducted in **Armenia, Estonia, Kazakhstan, Slovenia, and Spain**. These countries were selected to capture variation in institutional context, governance traditions, and public-sector experience with digital systems, rather than to provide statistical representation.

The inclusion of **both EU and non-EU countries** is intentional and analytically important. It reflects the reality that AI governance is not confined to the EU's internal market, but increasingly shaped by cross-border data flows, regulatory influence, institutional cooperation, and policy alignment beyond EU borders.

EU Member States (Estonia, Spain, and Slovenia) provide insight into social expectations within the EU regulatory space, where AI governance is increasingly shaped by common legal and policy frameworks.

Non-EU countries included in the study (Armenia and Kazakhstan) are politically and strategically significant to the EU. They are either part of the European Neighbourhood or have concluded an Enhanced Partnership and Cooperation Agreement (EPCAs) with the EU. Such cooperation involves commitments related to governance, digitalisation, and regulatory approximation, making public perceptions of AI in these contexts directly relevant to EU external policy objectives.

Including both EU and non-EU contexts allows the study to examine how shared governance principles, such as human control, fairness, accountability, and data protection, are interpreted across different institutional environments, and how expectations converge or diverge where regulatory frameworks, state capacity, and institutional trust differ.

Across all country contexts, the FGDs explored participants' perceptions of AI in public and social domains, with particular attention to:

- institutional use of AI in decision-making and service provision
- conditions under which AI is considered acceptable or unacceptable
- concerns related to human control, data protection, fairness, and accountability
- perceived risks for vulnerable groups, including children and older adults.

The study adopts a comparative qualitative approach, identifying both shared patterns and context-specific differences across countries and participant groups.

This study is exploratory and qualitative in nature. Its findings are grounded entirely in participant discourse and reflect how AI is perceived, discussed, and evaluated within the FGDs. The analysis does not seek to measure prevalence of opinions, predict behaviour, or assess the effectiveness of specific AI systems.

Accordingly, the study **can claim**:

- insight into dominant interpretive frames through which AI is socially understood
- identification of shared conditions for acceptance and key sources of concern
- comparative understanding of how emphasis on risks and safeguards varies across institutional contexts.

The study **cannot claim**:

- representativeness of national populations or stakeholder groups

- empirical assessment of AI technologies, governance frameworks, or regulatory compliance
- causal conclusions about the impact of AI on trust, fairness, or institutional performance.

Findings should therefore be read as **context-sensitive indicators of social meaning and concern**, rather than as generalisable measures of public opinion. Their value lies in informing EU internal and external policy dialogue, governance design, and future research on socially grounded AI regulation.

Together with Deliverables D3.1 and D3.2, which examined the legal and ethical dimensions of AI, this deliverable introduces a complementary social perspective to the CERTAIN Project's analytical framework. By grounding governance considerations in empirically observed expectations around trust, fairness, accountability, transparency, and human oversight, D3.3 strengthens the foundation for Deliverables D3.4 and D3.5, which will develop comprehensive guidelines and compliance tools. In doing so, it helps ensure that CERTAIN's final outputs are not only legally and ethically robust, but also socially responsive and aligned with lived experience.

The findings of D3.3 further support the operational objectives of WP5 and WP7 by integrating socially grounded insights into technical development and demonstrator validation. In WP5, the emphasis on human oversight, contestability, transparency, and accountability informs the design of trustworthiness testing methods, lifecycle modelling, and AI compliance guidance, ensuring that technical tools reflect both regulatory standards and public expectations. In WP7, insights into institutional trust, perceived risks, and cross-border governance dynamics contribute to stakeholder engagement strategies, co-creation processes, and impact evaluation within demonstrator environments. Through this integration, D3.3 bridges social acceptance research with technical validation and real-world implementation, reinforcing CERTAIN's holistic approach to responsible and trustworthy AI.

Draft

### 3. CONCEPTUAL STARTING POINTS

This study approaches social acceptance of AI as a relational and contextual phenomenon, shaped by how people assess **trust**, **perceived risks**, and the **expectations they hold regarding human oversight**. Social acceptance is not treated as a binary judgment for or against AI, but as a conditional and situational evaluation that depends on who uses AI, for what purpose, and with what safeguards. In this framing, acceptance emerges through participants' assessments of institutional credibility, anticipated consequences of AI use, and the degree to which human judgment remains visible and effective in decision-making processes.

Trust functions as a central mediating concept, not in relation to AI as a technical system, but in relation to the institutions and actors deploying it. Perceived risks – such as loss of control over personal data, unfair treatment, or diffusion of responsibility – shape whether AI is seen as acceptable or problematic in specific contexts. Expectations of human oversight provide the key condition through which trust and risk are balanced: participants consistently assess AI more favourably where humans retain decision-making authority, responsibility is identifiable, and outcomes can be reviewed or challenged. Together, these three elements provide the conceptual lens through which FGD discussions are analysed, enabling comparison across countries and groups while remaining grounded in participants' own frames of reference.

Draft

## 4. METHODOLOGY

Methodologically, the study adopted a qualitative, comparative design. Its primary objective was not to measure prevalence or predict behaviour, but to identify drivers and barriers of AI acceptance, analyse trust dynamics across institutions and sectors, and surface shared conditions under which AI is considered acceptable or unacceptable. The analysis supported the development of inclusive policy and business reflections by tracing how acceptance varies across contexts while remaining anchored in a common normative baseline.

Participants for the focus group discussions (FGDs) [1] were selected using a purposive and stratified sampling approach, designed to capture a wide range of perspectives relevant to the social acceptance of AI rather than to achieve statistical representativeness. Selection focused on individuals whose experiences, professional roles, or social positions were likely to shape their interaction with, exposure to, or assessment of AI-enabled systems in everyday life and institutional settings.

To this end, participant recruitment was organized around two broad clusters: institutional and corporate stakeholders, and communities affected by AI use. Within each cluster, participants were selected to reflect variation in institutional proximity, sectoral experience, age, gender, and familiarity with digital technologies, as relevant to the AI-related contexts under discussion. Group composition was informed by practical relevance to AI deployment in areas such as governance, public services, employment, education, and digital infrastructure.

A summary of participant composition across country contexts and focus group types is provided in Appendix A. In brief, the study included participants from both institutional and corporate stakeholder groups (e.g. academic and business communities) and from communities directly affected by AI use (including youth, older adults, and migrants and displaced persons). While the sample was not statistically representative, it was designed to ensure diversity in institutional proximity, professional background, age, and digital experience across the five country contexts.

Across FGDs, participants were invited to speak both from their professional or social position and from their personal experience as individuals affected by AI-supported systems. Where participants self-identified with particular roles, backgrounds, or perspectives during the discussions, this positioning was taken into account analytically. However, no rigid demographic categories were imposed *ex ante*, and group-based distinctions in the analysis reflect how participants framed their own experiences rather than predefined classifications. This approach enabled the study to surface shared patterns of concern as well as context-specific differences, while remaining grounded in participants' own frames of reference.

While efforts were made to ensure diversity of perspectives, women were overrepresented in most FGDs. This reflects both the composition of some participant pools and the voluntary nature of participation, and is taken into account when interpreting the findings.

Despite targeted outreach efforts, recruitment of persons with disabilities proved challenging across the selected country contexts, resulting in their non-representation in the FGDs. This reflects practical barriers to participation, including reliance on intermediary organisations, and is acknowledged as a limitation of the study.

For each group, researchers employed tailored focus group discussion guides, including scenario-based prompts that situate AI use in concrete contexts such as public administration, employment screening, education, healthcare, or security-related decision-making. Instruments were adapted to each country to ensure cultural sensitivity and accessibility, recognising differences in institutional experience and digital familiarity.

---

[1] See Appendix A for the gender and target group breakdown of each country's sample. See Appendix B for the FGD protocols.

Analytical attention is structured around recurring dimensions that emerged inductively across discussions, including perceptions of fairness (procedural, distributive, and experiential), trust in institutions and system operators, empowerment versus control, transparency and visibility of decision-making, human agency and dignity, harm minimisation, and perceptions of cross-border or externally governed AI systems. These dimensions provide a consistent analytical backbone while allowing for contextual variation in emphasis.

Geographically, the study deliberately included both EU and non-EU countries, reflecting the transnational nature of AI governance, data flows, and regulatory influence. EU Member States offer insight into expectations within a shared regulatory space, while non-EU partner countries – particularly those engaged with the EU through neighbourhood policy or Enhanced Partnership and Cooperation Agreements – provide a critical perspective on how AI legitimacy is negotiated in contexts of regulatory approximation, export, and institutional cooperation. Research instruments and implementation are adapted in partnership with local actors to reflect each context's regulatory environment, digital maturity, and civic space.

Overall, the methodology was designed to balance comparability and contextual sensitivity, enabling the study to identify shared patterns of social acceptance while remaining attentive to how institutional histories, governance capacity, and lived experience shape perceptions of AI in different settings.

Draft

## 5. THEMATIC ORGANISATION OF THE ANALYSIS

A central organising principle emerging from the FGDs is that participants assessed artificial intelligence primarily through **institutional purpose and geopolitical context**, rather than through sector-specific applications alone. Across country discussions, AI was not treated as a neutral technical tool, but as a capability embedded in state power, corporate strategy, and cross-border dynamics. Participants' evaluations, therefore, focused less on specific AI functions and more on **who controls AI, whose interests it serves, and how its use intersects with national security, economic competition, and political influence**.

Across all country FGDs, participants spontaneously differentiated between **domestic administrative use, corporate deployment, and strategic or security-related use of AI**. This differentiation emerged without prompting and indicates that perceptions of AI are shaped by perceived **levels of power asymmetry and strategic risk**, rather than by technical characteristics of AI systems themselves.

In discussions related to public administration and state use, participants in Kazakhstan, Estonia, and Armenia acknowledged potential efficiency gains from AI in internal processes, data management, and decision support. However, these acknowledgements were consistently accompanied by reservations about concentration of control, accountability, and opacity. Participants expressed discomfort with AI being used to justify decisions without clear human responsibility, particularly where such decisions affect citizens' rights or access to services. These concerns form the empirical basis for treating institutional accountability and responsibility as core analytical themes.

Healthcare and education were discussed less extensively than in earlier drafts of the analysis. Where mentioned, AI in these domains was generally framed as acceptable **only as a support tool**, contingent on professional oversight. Participants did not engage in sustained discussion of children as a distinct analytical category, nor did child-specific risks dominate the FGDs. References to children tended to arise indirectly, mainly in relation to future societal impacts or long-term consequences of data accumulation, rather than as a standalone focus.

Law enforcement likewise did not constitute a dominant thematic strand across the FGDs. While isolated references were made to surveillance technologies or facial recognition, these were not elaborated as a primary concern. Instead, such references were often subsumed within broader discussions about **state power, national security, and cross-border data use**, rather than policing practices per se.

Notably, **national security and geopolitical considerations emerged as a recurring but uneven theme**, particularly in the Armenia and Estonia FGDs and, to a lesser extent, in Spain. Participants raised concerns about AI being developed, trained, or controlled outside national borders, and about dependence on foreign technologies in strategically sensitive domains. AI was discussed as part of global technological competition, with participants questioning whether smaller or less technologically advanced states could retain sovereignty, control, or bargaining power in relation to AI systems developed elsewhere.

In these discussions, risks were framed less in terms of individual harm and more in terms of **strategic vulnerability**, including exposure of sensitive data, reliance on external providers, and asymmetries between states with different technological capacities. Participants expressed concern that AI could deepen existing global inequalities or be used as a tool of influence, pressure, or leverage in international relations. These concerns were articulated without technical detail, but were grounded in participants' broader understanding of geopolitics and security.

Across all domains where AI was discussed, acceptance was consistently framed as **conditional rather than absolute**. Participants used qualifying language ("it depends," "only if," "as long as") to articulate boundaries of acceptability. These conditions centred on human oversight, clarity of purpose, limits on scope, and identifiable responsibility. This pattern reflects participants' own reasoning structures and is directly grounded in FGD discourse.

Trust emerged as a cross-cutting theme, with participants repeatedly linking trust in AI to trust in the institutions, states, or actors deploying it. In contexts where institutions were perceived as accountable and predictable, participants were more willing to discuss conditional or limited uses of AI. Where institutional trust was lower, AI was framed as amplifying existing risks rather than mitigating them.

Concerns about accountability and responsibility were explicitly raised when participants were asked who would be responsible if AI-supported decisions caused harm. Across countries, participants struggled to

identify a single accountable actor, often referring to a diffusion of responsibility between developers, institutions, and decision-makers. This uncertainty constitutes a core empirical finding rather than a normative overlay. Participants also linked accountability to explainability, noting that without clear, accessible explanations of how AI-informed decisions are reached, responsibility becomes even more difficult to locate or contest.

Expectations regarding human oversight were among the most consistent findings across FGDs. Participants repeatedly stated that AI should assist, inform, or support human decision-making, but not replace it. This expectation was articulated across countries and contexts and did not rely on references to specific sectors such as policing or child protection.

Finally, issues of participation and voice emerged inductively in several FGDs, particularly in Armenia and Spain. Participants expressed frustration at decisions about AI being made “above” or “elsewhere,” without meaningful public consultation. These concerns were framed in terms of democratic legitimacy and national agency, rather than sector-specific governance. While not present in all FGDs, their recurrence justifies cautious analytical inclusion.

Draft

## 6. CROSS-CUTTING OBSERVATIONS

### 6.1. LITERATURE CONTEXTUALISATION OF CROSS-CUTTING FINDINGS

The FGD findings strongly align with existing literature emphasising institutional trust and governance credibility as central determinants of AI acceptance. The FRA's analysis of artificial intelligence and fundamental rights underscores that public trust depends on transparency, accountability, and the availability of remedies where harm occurs. [1] Similarly, the European Commission's guidance under the AI Act highlights that legal compliance alone does not guarantee legitimacy without clear allocation of responsibility and oversight. [2] Participants in the FGDs echoed this logic by consistently evaluating AI not as a neutral technology but as embedded in institutional practice. Their emphasis on who controls AI, for what purposes, and under what safeguards reflects broader scholarship arguing that governance arrangements, rather than technical capability, shape public confidence in AI. [3] [4] The empirical findings therefore reinforce the literature's conclusion that social acceptance is institutionally mediated.

The strong expectation of meaningful human control and an emphasis on accountability expressed in the FGDs also resonates with debates in the literature on human agency and oversight. In healthcare, WHO [5] and Council of Europe [6] stress that AI systems must preserve the patient-physician relationship and avoid displacing professional judgment. Similar concerns are documented in studies of generative AI in medicine, which caution against over-reliance on automated outputs without human oversight. [7] [8] In law enforcement and security contexts, concerns about opacity, bias, and accountability gaps are similarly documented. [9] [10] [11] FGD participants' insistence that AI should "assist but not decide," and their discomfort with blurred responsibility, mirror these documented risks. The convergence between participant discourse and the literature suggests that expectations of human oversight are not merely intuitive reactions but reflect broader normative principles emerging in policy and academic debate.

Finally, participant concerns about fairness, data use, and cross-border governance correspond closely to established research on inequality, bias, and regulatory fragmentation. The FRA [12] and Europol [13] identify risks of disproportionate impacts and biased outcomes in both public administration and law enforcement contexts. In education, Council of Europe [14] [15] highlights concern regarding profiling, opacity, and long-term developmental consequences: themes echoed in FGDs discussing rigidity and early categorisation. Cross-border data flows and jurisdictional uncertainty, raised particularly in Armenia and Estonia, resonate with broader discussions about global AI governance and regulatory alignment. [16] Taken together, the literature confirms that the themes emerging inductively from the FGDs (such as institutional trust, human oversight, fairness, and data control) are consistent with internationally recognised legal, ethical, and governance challenges associated with AI deployment.

### 6.2. RELEVANCE TO AI ACT

Across all country contexts, participants' expectations closely mirror several core principles embedded in Regulation (EU) 2024/1689 (AI Act), even where the Act itself was not explicitly referenced in discussions. The repeated emphasis on meaningful human oversight, transparency of use, accountability, fairness, and contestability aligns directly with the Act's foundational requirements for high-risk AI systems. In particular, Article 14's requirement for effective human oversight mechanisms corresponds closely with participants' insistence that AI should assist rather than replace human decision-making, especially in contexts affecting rights, access to services, or sanctions. Similarly, transparency obligations under Articles 13 and 50 reflect participants' expectations that individuals should be informed when AI is used, understand its role in decision-making, and be able to identify the responsible actor. The FGDs therefore demonstrate that these principles are not abstract regulatory constructs but reflect socially grounded expectations regarding how AI should function within institutional settings.

Participants' concerns about data protection, secondary use of information, cross-border governance, and diffusion of responsibility also resonate strongly with the AI Act's risk-based structure and its provisions on

accountability, documentation, and traceability. The Regulation places clear obligations on providers and deployers to ensure proper risk management (Article 9), maintain technical documentation (Article 11), enable record-keeping (Article 12), and support post-market monitoring (Article 72). While FGD participants did not frame their concerns in legal terminology, their repeated questions about who is responsible, how decisions can be challenged, and what happens to data once collected correspond closely with these governance mechanisms. In this respect, the empirical findings provide tangible social validation of the AI Act's architecture, particularly its emphasis on traceability, oversight, and clearly allocated responsibility.

Importantly, the FGDs also highlight a critical dimension for AI Act implementation: formal compliance alone is unlikely to secure social legitimacy unless safeguards are visible, comprehensible, and enforceable in practice. Participants consistently evaluated AI systems not only on whether rules exist, but on whether those rules operate meaningfully and can be experienced by affected individuals. This suggests that implementation efforts under the AI Act should prioritise procedural clarity, accessibility of information, and practical avenues for contestation and redress. The cross-country findings therefore enhance the policy relevance of D3.3 by demonstrating that the Regulation's core principles correspond to public expectations, while simultaneously underscoring that institutional practice, communication, and enforcement will ultimately determine whether those principles translate into sustained trust and acceptance (see also section 8.3. *Implications for future regulation* below).

### 6.3. TRUST IN INSTITUTIONS

Across all focus group discussions, participants did not assess artificial intelligence as an autonomous or self-standing technology. Instead, trust in AI was consistently articulated through trust in the **institutions and actors responsible for its deployment and use**. Participants evaluated AI in relation to institutional behaviour, accountability, and decision-making practices, rather than in terms of technical performance or innovation alone.

In all country FGDs, participants explicitly linked their willingness to accept AI-supported decisions to whether they trusted the institution involved to act fairly, transparently, and in the public interest. Where institutions were perceived as lacking accountability or clarity, AI use was approached with caution. In these discussions, AI was not framed as a neutral or corrective instrument, but as something that could **inherit or amplify existing institutional practices**, for better or worse.

This institutional framing was particularly visible in discussions involving public authority and service provision. Participants repeatedly raised concerns about whether AI might be used to obscure responsibility, justify decisions without adequate explanation, or make it more difficult to challenge outcomes. In this context, **explainability** emerged as a key safeguard: participants stressed that when AI informs or shapes decisions, individuals must be able to understand in clear, accessible terms how and why a particular outcome was reached. Importantly, these concerns were directed primarily at **institutional decision-making structures**, with AI perceived as an additional layer within those structures rather than as an independent source of risk.

Differences emerged, however, in how participants across countries interpreted the institutional role in mitigating or exacerbating these risks.

In the **Kazakhstan** FGDs, participants expressed concern about the risks associated with AI use by public authorities, particularly in relation to personal data protection, responsibility for decisions, and the potential for errors or abuse in the absence of clear rules. These concerns were not framed as a rejection of state involvement in AI governance. On the contrary, participants repeatedly emphasised that **the state should play a central and active role** in regulating, overseeing, and setting boundaries for AI use.

Across academic, business, and student discussions, participants highlighted the need for **clear regulations, defined responsibility, enforceable standards for data protection, and visible state accountability** as prerequisites for trust. AI use in public services was often described as acceptable – or even desirable – if embedded within transparent procedures and supported by explicit legal frameworks. Trust-related concerns in Kazakhstan reflected a **demand for stronger, more explicit state governance**, rather than scepticism toward institutional involvement as such. Where uncertainty or mistrust was expressed, it was linked primarily to regulatory gaps and weak enforcement, not to opposition to public oversight.

In the **Armenia** FGDs, participants raised sustained concerns about how AI might be used by public institutions, particularly in relation to accountability, fairness, and responsibility for decisions. These

discussions reflected **ambivalence rather than blanket distrust** toward institutional practice. Participants frequently expressed concern that AI could reinforce existing problems – such as lack of transparency, selective application, or diffusion of responsibility – if deployed without clear rules and oversight.

At the same time, some participants articulated the view that AI could **reduce discretionary abuse or corruption** if embedded within transparent and well-defined procedures. References to internal rules, clearer governance frameworks, and institutional responsibility were present, indicating that AI was not viewed as inherently unacceptable, but as highly contingent on how institutions design, control, and take responsibility for its use.

In addition, the Armenia FGDs introduced **explicit concerns related to national security and cross-border use of AI systems**. Participants questioned where data are stored, who ultimately controls AI systems developed or hosted abroad, and how cross-border data flows might affect national sovereignty and security. These concerns were particularly salient in discussions of government use of AI, where participants expressed unease about reliance on foreign platforms or systems whose governance lies outside national control. In this framing, AI-related risks were linked not only to domestic institutional practice, but also to geopolitical dependence and loss of control over sensitive data and decision-making processes.

In **Estonia**, while the overall emphasis on procedural safeguards, clarity of roles, and oversight remains unchanged, discussions place greater weight on **jurisdictional alignment and cross-border AI governance** as necessary conditions for trust, indicating that confidence in AI use is shaped not only by domestic institutional arrangements but also by the legal and regulatory environments in which AI systems are embedded.

Generally, in the Estonia FGD, participants approached trust primarily through a **procedural and governance-oriented lens**. Rather than focusing on potential misuse or abuse, discussions centred on how AI systems should be designed, regulated, and overseen. Participants emphasised the importance of clear rules, transparency of processes, and mechanisms to review or correct AI-supported decisions. References to public authorities, including law enforcement, were present but limited and framed in terms of safeguards and accountability requirements rather than expressions of concern about surveillance or coercive use.

In the **Spain** FGDs, trust in institutions was discussed primarily through **professional proximity, responsibility, and regulation**, rather than as generalized confidence in public authorities. Participants across academia, business, and older adult groups repeatedly framed AI as a **powerful but immature tool**, whose acceptability depends on who uses it, for what purpose, and under what rules. Trust was rarely attributed to “institutions” in the abstract; instead, it was mediated through **professional competence, transparency of use, and regulatory clarity**.

In education and research contexts, participants expressed relative confidence when AI was used as a **supporting instrument under human supervision**, particularly by trained professionals who could critically assess outputs. Trust was linked to the ability to review, correct, and contextualise AI-generated content, rather than to institutional authority as such. At the same time, participants voiced concern about widespread, uncritical use – especially by students – highlighting fears of dependency, loss of critical thinking, and misuse driven by lack of understanding rather than malicious intent.

Across business and older adult FGDs, trust concerns shifted more clearly toward **governance and accountability**. Participants frequently stated that AI deployment is advancing faster than clear rules, and that this regulatory lag undermines trust. Several participants explicitly argued that **stronger regulation is necessary**, not to block innovation, but to clarify responsibility, data use, and liability. In this context, mistrust was not framed as opposition to AI itself, but as discomfort with ambiguity—particularly regarding who is responsible when AI-informed decisions cause harm.

Importantly, Spanish participants also articulated a **countervailing concern**: that excessive or poorly designed regulation could **stifle innovation** or unfairly penalise responsible users. This tension – between the need for regulation and fear of overregulation – was explicitly raised, particularly in business and academic discussions. Trust in institutions was therefore conditional and dynamic: participants were willing to trust AI use where regulation is clear, proportionate, and intelligible, but sceptical where governance appears either absent or disconnected from real-world practice.

“They are like roadblocks on the road to development — they come at a cost, which is that we cannot move at the same pace as China or the United States, or places where transparency and justice are not prioritised. Those are obstacles.

Today, if you want to develop a business in Barcelona, all of this can feel like an obstacle — but for me, it's a welcome one. In fact, up to now I see it as something positive. Still, we cannot run the same race that other countries are running.”

**Quote, Spain FGD participant**

Overall, the Spanish FGDs depict trust as **procedural rather than institutional**: AI is trusted when its use is understandable, contestable, professionally mediated, and governed by rules perceived as fair and realistic.

In the **Slovenia** FGDs, trust was discussed in a more **practice-oriented and relational manner**, with less explicit focus on state authority or regulation. Participants tended to evaluate AI primarily through its **impact on everyday decision-making**, particularly in education and child-related contexts. Trust was associated with **retention of human judgment**, sensitivity to individual circumstances, and avoidance of rigid or automated categorisation.

Rather than expressing strong views about institutional credibility, Slovenian participants focused on whether AI systems could appropriately account for context, development, and nuance. Concerns about trust emerged most clearly where AI was perceived as distancing decisions from human understanding, especially in areas involving learning, assessment, or long-term consequences for children. AI was more readily accepted when framed as a background tool supporting professionals, and more critically assessed when associated with standardisation or loss of discretion.

Overall, the Slovenia FGDs suggest that trust is less explicitly tied to institutional authority and more closely linked to **how closely AI remains embedded in human-scale practices**. While concerns about fairness and control were present, they were articulated through the lens of everyday experience rather than through explicit discussion of governance or regulation.

Importantly, FGDs across all countries demonstrated that expressions of concern did not amount to a general rejection of technological innovation. Participants frequently distinguished between technological capability and institutional intent, emphasising that the key issue was not AI itself, but **who controls it, for what purposes, and under what rules**. This distinction recurred across country contexts and sectors and formed a shared interpretive frame for evaluating AI.

Overall, the FGD findings indicate that trust operates as an institutional filter for AI acceptance. Where institutions were perceived as accountable, transparent, and responsive, or capable of being made so through clear oversight, participants were more open to discussing conditional or limited uses of AI. Where institutional practices were perceived as opaque or untrustworthy, AI deployment was viewed as inherently risky, regardless of the specific application.

## 6.4. HUMAN CONTROL OVER AI

Across all focus group discussions, participants consistently stressed the importance of **maintaining clear human control over AI-supported processes**. AI was generally discussed as a tool that may assist or support human decision-making, but not as something that should replace human judgment in decisions affecting rights, access to services, education, care, or sanctions. Participants repeatedly drew a distinction between AI as an aid and AI as a decision-maker, with the latter viewed critically across all country contexts.

“But I can say this: the main advantage of artificial intelligence is the ability to generate ideas. It truly helps in finding excellent topics for research papers and project work. I've experienced this myself — at school I wrote many projects and term papers. When I didn't know which topic to choose, I simply described my interests and received good suggestions. Then I formulated the research questions and objectives myself. AI is also useful for finding sources. For example, if I need specific statistics, I ask where to find them and receive recommendations for reliable websites. This saves time and helps locate high-quality information. I think it is already impossible to fully control how people use AI.”

**Quote, Kazakhstan FGD participant**

“I guess I use ChatGPT a lot for different tasks, mostly for any of coding I need to do because I'm not really, it's not my expertise, or maybe some other informations, umm, I can find them easily there. Not some deep research, but like the general things that I don't need to search on Google.

“Yeah, same goes for me as well. Mostly, yeah. Mostly is for ChatGPT, for brainstorming, maybe help with the writing, summarisation.”

**Quote, Slovenia FGD participant**

In all FGDs, participants articulated boundaries around acceptable uses of AI based on the **degree of autonomy attributed to the system**, rather than on its technical complexity. AI use was described as acceptable when humans retained the authority to review, interpret, override, or disregard AI outputs. By contrast, systems perceived as operating automatically or determinatively – particularly where human involvement was unclear or invisible – were met with discomfort and resistance. Participants associated such systems with loss of responsibility and reduced ability to question or correct decisions.

While expectations regarding human control were shared across all countries, the **concerns associated with loss of control differed in emphasis**.

In the **Kazakhstan and Armenia FGDs**, participants frequently linked loss of human control to concerns about **responsibility and accountability**, but with different emphases. In Kazakhstan, discussions highlighted the risk that officials could rely on AI outputs without clear rules or legal frameworks, leading to uncertainty about who is responsible for decisions. Human control was therefore framed as essential to ensure that decisions remain attributable to identifiable actors and subject to oversight. In Armenia, similar concerns were raised, particularly regarding the possibility that AI could be used to shift or obscure responsibility. At the same time, Armenian participants also articulated that AI could reduce discretionary abuse if embedded within transparent procedures, underscoring that loss of human control was viewed as problematic primarily in the absence of clear governance and responsibility.

“I'm really afraid that even if we have all the regulations in the world, people will still find ways to misuse AI. At the same time, I agree with [...] — and I think this is a very important point — that regulation needs to be in place so that when we identify people who are breaching the rules, they can at least be held accountable under the law.”

**Quote, Armenia FGD participant**

In the **Estonia FGD**, participants likewise emphasised the necessity of human control, but approached the issue predominantly through a **procedural and design-oriented lens**. Discussions focused on ensuring that humans retain final decision-making authority, that AI outputs can be reviewed and corrected, and that roles and responsibilities are clearly defined within system design. Rather than emphasising misuse or avoidance of responsibility, participants concentrated on governance arrangements that would operationalise human oversight in practice.

Also in Estonia, participants described AI as increasingly operating as an **invisible infrastructure**, embedded in everyday systems and processes without always being explicitly noticed. While this pervasiveness was not viewed negatively in itself, participants emphasised that invisibility becomes problematic when it undermines awareness, transparency, or the ability to question how decisions are made. This reinforces the importance of procedural transparency and user awareness as safeguards alongside formal governance arrangements

In the **Spain FGDs**, concerns about human control emerged strongly in relation to **education, professional practice, and learning environments**. Participants emphasised that AI should support, rather than replace, teachers', researchers', and professionals' judgment. Loss of human control was associated with risks of over-reliance, reduced critical thinking, and diminished professional discretion, particularly where AI tools are used uncritically. Concerns were framed less around avoidance of responsibility and more around the erosion of judgment and understanding in complex or developmental contexts.

In the **Slovenia FGDs**, concerns about human control were also most visible in discussions involving **education and children** but were articulated primarily through the lens of **rigidity and standardisation**. Participants expressed concern that automated processes could fail to account for individual circumstances or change over time. Loss of human control was associated with distance from everyday human understanding rather than with accountability avoidance or procedural failure.

Across all country FGDs, participants rejected the idea of AI functioning as an autonomous decision-maker in high-impact contexts. Differences across countries lay not in whether human control is necessary, but in **why its absence was perceived as problematic** – whether because it risks diffusing responsibility, undermining procedural clarity, weakening professional judgment, or limiting contextual sensitivity. Taken together, the FGDs indicate that meaningful human control is viewed as a **non-negotiable condition** for the acceptable use of AI across countries and sectors, grounded in participants’ concerns about accountability, fairness, and the ability to question or challenge decisions.

## 6.5. DATA PROTECTION AND MISUSE CONCERNS

Across all focus group discussions, participants consistently raised concerns about **data protection and the potential misuse of personal information** in relation to artificial intelligence. These concerns were expressed in practical, everyday terms rather than through technical or legal concepts. Participants spoke about data in terms of what information is collected, who might have access to it, and what could happen to it once it enters AI-supported systems.

In all country FGDs, participants expressed unease about the **amount and sensitivity of data** required for AI systems to operate, particularly in public-sector contexts. Participants questioned what kinds of personal data are collected, who can see or use them, and whether individuals are informed about how long such data are retained or how they may be used in the future. Lack of clarity around these issues was repeatedly cited as a reason for mistrust toward AI-supported services.

Concerns were especially evident where AI was associated with **personal, behavioural, or identifying data**. Participants referred to facial images, location data, and digital traces as examples of information that could be collected and later reused. These risks were described in terms of **loss of privacy and loss of control**, with participants expressing discomfort at the idea of being monitored, tracked, or assessed without their knowledge or meaningful consent.

In discussions touching on **social services and healthcare**, participants emphasised the sensitivity of personal information shared in contexts of care or support. Participants questioned whether data provided for one purpose – such as receiving assistance or medical care – could later be accessed by other institutions or used to make unrelated decisions. These concerns were voiced even where participants acknowledged potential efficiency gains, indicating that perceived benefits did not outweigh anxieties about data protection.

Across several FGDs, participants expressed concern not only about accidental data breaches, but about **intentional misuse of data**. Participants raised fears that information collected through AI-supported systems could be deliberately repurposed or accessed by authorities, private actors, or third parties for purposes beyond those originally stated. Misuse was described broadly, including unauthorised access, secondary use, and use of data for control or pressure rather than for service provision.

Participants frequently linked data protection concerns to **trust in institutions**. Where participants doubted institutions’ ability or willingness to handle data responsibly, AI systems were viewed as increasing exposure to harm rather than improving outcomes. Assurances about technical performance or efficiency were not seen as sufficient to mitigate these concerns in the absence of clear responsibility and oversight.

“Here, in my opinion, the state should approach this issue very carefully and more frequently develop policies on data protection. It should also, in some cases, impose conditions on private companies — including fines — and treat the dissemination of personal data with great sensitivity.

Which companies am I referring to? In the context of artificial intelligence, I mean companies that provide public services, such as financial services. Our banking applications are very well developed. We even have a government services platform. But data leaks still happen. Even in simple situations: we enter our data into some government database, and shortly afterwards we start receiving phone calls. The consequences follow almost immediately, and it is clearly from fraudulent groups. You don’t know where the leak occurred.

That is why the state must act responsibly here — not only by developing these systems and promoting digitalisation, but also by introducing accountability for companies. When they develop AI-based services, they

must pay serious attention to the protection of personal data, and there should be sufficiently strict requirements in this area.”

**Quote, Kazakhstan FGD participant**

Importantly, FGDs showed that participants tended to understand data protection **holistically**, rather than distinguishing between stages such as collection, storage, sharing, or reuse. Participants expressed the view that once personal data enters an AI-supported system, individuals may lose meaningful control over how that data circulates, is interpreted, or is reused over time. This perception contributed to a general sense of vulnerability associated with AI-supported data practices.

While concerns about data protection were present across all country FGDs, **differences emerged in the risks participants emphasised most strongly.**

In the **Kazakhstan FGDs**, participants focused on gaps in regulation and enforcement, expressing concern that weak or unclear rules could expose personal data to misuse. These concerns were frequently accompanied by expectations that stronger state regulation and clearer standards could mitigate risks. In the **Armenia FGDs**, concerns about data protection were more often framed around **intentional misuse**, including fears that data could be repurposed for monitoring, pressure, or control. Armenian participants also raised explicit questions about **cross-border data flows and national security**, expressing unease about where data are stored and who ultimately controls AI systems developed or hosted abroad.

In the **Estonia FGD**, participants focused primarily on **data handling practices and safeguards**. Rather than assuming misuse, discussions centred on who has access to data, how access is controlled, how long data are retained, and how improper use would be prevented. Concerns were framed mainly in terms of procedural clarity and governance.

Echoing the concerns of participants from Armenia about AI and national security, Estonian FGD articulated that trust in AI is contingent not only on domestic institutional arrangements, but also on **where AI systems are developed, hosted, and legally governed**. AI systems operating within EU legal frameworks or under comparable regulatory standards were generally perceived as more trustworthy, while systems developed or governed outside such frameworks raised concerns related to accountability, data protection, and long-term democratic control. These concerns were framed explicitly in terms of **sovereignty and national security**, rather than individual harm or misuse, reflecting awareness of AI as part of broader geopolitical and regulatory ecosystems.

In the **Spain FGDs**, data protection concerns were closely linked to **regulatory clarity and proportionality**, particularly in education and digital contexts. Participants emphasised the sensitivity of personal and developmental data and expressed concern about long-term consequences of reuse or reinterpretation over time. At the same time, some participants cautioned that overly rigid regulation could constrain beneficial uses, highlighting the need for clear but adaptable data governance.

In the **Slovenia FGDs**, data protection concerns were articulated primarily through the lens of **everyday impact and long-term consequences**, especially for children and learners. Participants focused on the risk that stored or reused data could shape future opportunities in ways that are difficult to foresee or correct, rather than on surveillance or coercive use.

Across all countries, participants shared the view that AI-supported data practices carry a risk of diminishing individual control over personal information. Cross-country variation lay not in whether data protection matters, but in whether participants primarily feared **intentional misuse, insufficient safeguards, cross-border loss of control, or long-term unintended consequences.**

## 6.6. PERCEIVED FAIRNESS

Across all focus group discussions, participants consistently referred to **fairness** as a central criterion for evaluating the acceptability of AI-supported systems. Fairness was discussed in practical terms, linked to whether decisions were experienced as reasonable, understandable, and applied in ways that do not

disadvantage particular individuals or groups. Participants did not approach fairness as an abstract principle, but as something reflected in everyday outcomes and interactions with institutions or services.

In all country FGDs, participants questioned whether AI systems would **treat people in the same way in practice**, particularly in decisions affecting access to services, education, or other consequential outcomes. Participants expressed concern that AI could reproduce or reinforce existing differences in treatment, especially where decision-making processes are not visible or cannot be explained. These concerns were articulated through concrete examples of everyday situations rather than through technical discussions of bias.

Concerns about **categorisation and rigidity** featured prominently across several FGDs. In discussions touching on social services, education, and authority-driven decision-making, participants expressed discomfort with the idea that AI systems might label, score, or categorise individuals in ways that are difficult to contest or revise. Participants noted that once such categorisation occurs, individuals may have limited opportunities to explain their circumstances or correct errors, which was widely perceived as unfair.

Perceptions of fairness were closely linked to **human involvement in decision-making**. Across FGDs, participants stated that decisions are experienced as fairer when a human can take context, intent, or exceptional circumstances into account. By contrast, decisions that appeared to rely heavily on automated processes were described as unfair when they seemed inflexible, overly standardised, or detached from individual situations. This concern was particularly pronounced in discussions involving children, learners, and individuals perceived as more vulnerable.

FGDs also showed that participants understood fairness in **procedural terms**. Participants emphasised the importance of knowing how decisions are made and having the ability to ask questions, seek clarification, or challenge outcomes. Where AI-supported processes were perceived as opaque or difficult to contest, participants described them as inherently unfair, regardless of whether such systems might improve efficiency or consistency.

While concerns about fairness were shared across all country FGDs, **differences emerged in how unfairness was framed and what risks were foregrounded**.

In the **Kazakhstan FGDs**, participants most often discussed fairness in relation to **consistency and equal application**, expressing concern that AI-supported systems could be applied unevenly in the absence of clear rules and enforcement. Fairness was closely linked to the expectation that AI should operate under transparent standards and be used in the same way for everyone.

In the **Armenia FGDs**, fairness concerns similarly focused on unequal application, but were also tied to risks affecting **vulnerable or non-standard cases**, such as individuals with incomplete documentation or different linguistic or social backgrounds. Participants emphasised that rigid systems could exacerbate exclusion if contextual factors are not considered.

In the **Estonia FGD**, fairness was discussed primarily through a **procedural and governance-oriented lens**. Participants focused on whether AI-supported decisions would follow clear and predictable rules, be applied uniformly, and allow for correction in case of error. Rather than emphasising selective misuse, participants highlighted the importance of process clarity and effective mechanisms to contest outcomes.

The Estonia FGD also introduced a clearer focus on **cultural and linguistic adequacy** as a dimension of acceptable AI use. Participants expressed concern that AI systems developed for larger or different linguistic environments may inadequately reflect Estonian language use or social nuance. This was not framed as technical bias in the abstract, but as a risk of structural mismatch, where systems may formally comply with governance requirements yet fail to function appropriately in local contexts. In this sense, contextual fit emerged as an additional condition for maintaining trust.

In the **Spain FGDs**, fairness concerns were most prominent in discussions related to **education, professional practice, and learning environments**. Participants emphasised the risk that AI-based assessment or categorisation could disadvantage students or learners by failing to account for individual circumstances, development, or creativity. Fairness was framed in terms of flexibility and professional judgment, alongside concern that rigid systems could have long-term consequences for opportunities. In the **Slovenia FGDs**, fairness concerns were also closely linked to education and children, but were articulated more through the lens of **standardisation and loss of nuance**. Participants focused on the risk that automated processes could oversimplify complex situations and apply uniform logic where sensitivity and adaptation are needed.

Across all country FGDs, participants agreed that fairness is undermined when AI-supported decisions are **opaque, difficult to challenge, or detached from human judgment**. No FGD reflected an assumption that AI would inherently improve fairness. Instead, participants consistently framed fairness as **conditional** on transparency, human involvement, and the ability to question or appeal decisions. Cross-country variation lay in which risks to fairness were most salient, rather than in fundamentally different understandings of what fair treatment entails.

Draft

## 7. COMPARATIVE PERSPECTIVE

A comparative analysis of the focus group discussions (FGDs) reveals a strong degree of convergence across countries in how participants approach artificial intelligence, alongside clear differences in emphasis shaped by national context and discussion focus. Across all FGDs, participants demonstrated nuanced and conditional reasoning about AI, evaluating it primarily through its relationship with institutions, decision-making processes, and everyday impacts rather than as a standalone technology.

### 7.1. SHARED PATTERNS ACROSS ALL COUNTRIES AND GROUPS

Across all countries (Armenia, Estonia, Kazakhstan, Slovenia, and Spain), FGD participants approached artificial intelligence primarily through its **institutional context of use**. AI was not discussed as an abstract or autonomous technology, but as something embedded in public authorities, professional practice, and service systems. Participants' evaluations focused consistently on who controls AI, for what purposes it is deployed, and how its use affects individuals in concrete situations.

A defining feature across all discussions was the **conditional nature of acceptance**. In every FGD, participants avoided absolute judgments and instead framed their views using qualifying language such as “it depends,” “only if,” or “as long as.” Acceptance of AI was consistently linked to the presence of safeguards, including clearly defined purposes, limits on the scope of use, identifiable responsibility, and visible human involvement. AI was not rejected as such, but its acceptability was treated as contingent and context-dependent.

Expectations regarding **human control** emerged uniformly across all countries and groups. Participants rejected the idea of AI operating autonomously in decisions affecting rights, access to services, education, care, or other high-impact outcomes. Human involvement was framed as essential not only for decision quality, but for responsibility, contextual judgment, and the ability to review, correct, or challenge AI-supported decisions. This expectation formed a shared baseline across all discussions.

Concerns related to **data protection and potential misuse** were likewise present in all country FGDs and across participant groups. Participants raised recurring questions about what data are collected, who has access to them, and whether data might be reused beyond their original purpose. These concerns persisted even where participants acknowledged potential benefits of AI-supported systems, indicating that perceived utility did not outweigh anxieties about loss of control over personal information.

Across all discussions, **perceived fairness** emerged as a central criterion for acceptability. Fairness was discussed in practical terms, including equal application, sensitivity to individual circumstances, and the ability to question or appeal decisions. Participants consistently expressed concern that AI could make unfair outcomes harder to detect or contest, particularly where systems are opaque, rigid, or difficult to challenge. No FGD reflected an assumption that AI would inherently improve fairness without accompanying safeguards.

A similar convergence was evident when examining the FGDs across **different participant groups**. Regardless of professional background, age, or how participants positioned themselves during the discussions, the same core concerns – conditional acceptance, human control, data protection, and fairness—were repeatedly raised. While participants from different groups sometimes emphasised different aspects, such as procedural clarity, professional discretion, personal vulnerability, or long-term consequences, these differences reflected **variation in emphasis rather than divergence in underlying expectations**.

Taken together, the FGDs demonstrate a strong cross-country and cross-group convergence around the conditions under which AI may be considered acceptable. Participants did not articulate competing normative standards, but rather a shared set of expectations focused on human oversight, institutional responsibility, protection of personal data, and fair treatment.

## 7.2. DIFFERENCES IN EMPHASIS ACROSS COUNTRY CONTEXTS

While participants across all FGDs shared core concerns regarding human control, data protection, fairness, and accountability, the **relative emphasis placed on specific risks and expectations varied across country contexts**, reflecting differences in how AI use is situated within institutional and social experience.

In the **Kazakhstan and Armenia FGDs**, participants devoted considerable attention to the role of public authorities in the deployment of AI, particularly in relation to accountability, scope of use, and responsibility for decisions. In both contexts, discussions frequently raised concerns about selective application, unequal treatment, and the potential diffusion of responsibility when decisions are supported or justified by AI systems. References to public-sector use, including law enforcement and administrative decision-making, formed a significant part of these discussions.

In the **Kazakhstan FGDs**, such concerns were often accompanied by explicit expectations that the state should play a **stronger and more visible role** in regulating and overseeing AI use. Participants emphasised the need for clear rules, defined responsibility, enforceable standards, and state-led oversight as conditions under which AI could be considered acceptable. In this context, concern reflected not rejection of state involvement, but a demand for more robust, transparent, and enforceable governance.

In the **Armenia FGDs**, concerns were framed more ambivalently. Participants expressed unease about institutional practice, particularly regarding transparency, accountability, and the treatment of vulnerable or non-standard cases. AI was frequently described as likely to reinforce existing institutional problems if deployed without safeguards. At the same time, some participants articulated that AI could reduce discretionary abuse or corruption if embedded within transparent procedures. Armenian discussions also introduced explicit concerns about **national security and cross-border AI use**, including where data are stored and who controls AI systems developed or hosted abroad.

“And regarding the fact that they are developed abroad — that could potentially endanger the country’s internal stability [...], especially if national databases end up in the hands of other countries, which I don’t think is something we would want right now.”

“Do you mean that this could, at times, endanger national security?”

“Yes, national security — exactly.”

**Quote, Armenia FGD participant**

The **Estonia FGD** also echoed concerns about national security and cross-border AI use. Moreover, in the Estonia FGD, participants approached AI primarily through a **procedural and governance-oriented lens**. Rather than focusing on potential misuse or abuse by authorities, discussions centred on system design, clarity of roles, transparency of decision-making processes, and the availability of mechanisms to review or correct AI-supported decisions.

“I think that the problem is not per se that things are developed or even data stored outside of Estonia, but whether that outside of Estonia complies with the same rules as Estonia. So let me put it this way. I think that provided that we have the same rules and legal frameworks, if the database is based in France, Germany, Italy or Ireland, that's not the problem. But when the database is based in China, based in the United States and then the rules that apply there are different and not only different, but incompatible with our rules. [...] So then I think it can become a problem because, in a way, it might interfere with the functioning of our societies and, in the long term, might also undermine our democracies.”

**Quote, Estonia FGD participant**

In the **Spain FGDs**, discussions focused predominantly on **education, professional practice, and learning environments**, particularly where AI might affect children and young people. Concerns about fairness and human control were framed around risks of rigidity, over-reliance, and reduced professional judgment. Data protection concerns were linked to the sensitivity of educational and developmental data and to long-term consequences of reuse. At the same time, the Spain FGDs uniquely introduced a concern that **overly rigid or poorly calibrated regulation could constrain beneficial innovation**, especially in education and digital

contexts. This perspective did not challenge the need for safeguards, but highlighted a tension between protection and regulatory flexibility.

In the **Slovenia FGDs**, discussions similarly centred on education, but with a different emphasis. Participants focused less on regulation and more on the **everyday impact of standardisation and loss of nuance**. Concerns about fairness and human control were articulated primarily in terms of whether AI-supported processes could account for individual circumstances, development, and change over time, rather than in terms of authority-driven misuse or regulatory design.

### Differences in Emphasis across Participant Groups

Beyond country context, the FGDs also revealed differences in emphasis linked to **how participants positioned themselves during the discussions**, rather than to predefined demographic categories.

Participants who spoke from **professional or expert perspectives** (including education, healthcare, administration, or business) tended to focus on practical and procedural implications of AI use. Discussions emphasised how AI might affect professional discretion, responsibility, and day-to-day decision-making, with particular attention to clarity of roles, accountability, and the ability to override or correct AI outputs.

Participants who spoke primarily from a **general citizen perspective**, without referencing professional proximity to AI systems, were more likely to frame concerns in terms of power imbalance, loss of control, and vulnerability. AI was often discussed as something imposed by institutions, with questions about who benefits from its use and whether individuals would have meaningful recourse if harmed.

Age-related differences were less explicitly articulated. However, where participants referred to **children or young people** – most notably in the Spain and Slovenia FGDs – concerns focused on long-term consequences, rigidity of categorisation, and the impact of early AI-based assessment on future opportunities. Where **older participants** referenced their own position, concerns centred on accessibility, understandability, and the ability to navigate AI-supported systems, particularly in interactions with public services.

Participants with **business or private-sector experience**, where identifiable, were more likely to acknowledge potential efficiency or organisational benefits of AI, especially for administrative tasks. However, these participants also emphasised responsibility, reputational risk, and the need for clear rules, indicating that perceived benefits did not override concerns about misuse, accountability, or loss of trust.

## 7.3. OVERALL COMPARATIVE INSIGHT BY COUNTRY AND GROUP

Taken together, the FGDs indicate a **high degree of convergence across countries and participant groups** regarding the conditions under which AI is considered acceptable (see Figure 1 for a non-hierarchical representation of cross-cutting themes emerging from FGDs on social acceptance of AI), alongside **systematic variation in which risks are most salient** depending on national context and participant perspective.

Across all countries and groups, participants did not approach AI as a neutral or autonomous technology. Instead, AI was evaluated through its **institutional embedding and social consequences**, with acceptability framed as conditional on human control, data protection, fairness, and identifiable responsibility. No FGD reflected unconditional trust in AI, nor support for fully autonomous decision-making in contexts affecting rights, access to services, education, or care. This shared baseline cuts across both country and group differences and constitutes the core comparative finding of the study.

At the **country level**, variation lay primarily in how participants interpreted the principal risk associated with AI use. In the Kazakhstan and Armenia FGDs, discussions foregrounded concerns about how AI might be used by public authorities, particularly in relation to accountability, selective application, and responsibility for decisions. In Kazakhstan, these concerns were frequently accompanied by explicit expectations that the state should assume a stronger and more visible regulatory and oversight role. In Armenia, concerns were articulated more ambivalently, reflecting unease about institutional practice and capacity, alongside recognition that AI could reduce discretionary abuse if embedded within transparent procedures. Armenian discussions also introduced concerns related to cross-border control and national security, shaping perceptions of risk beyond domestic governance alone.

In the Estonia FGD, participants approached AI predominantly through a procedural and governance-oriented lens. Rather than emphasising misuse or abuse, discussions focused on system design, clarity of rules, transparency of processes, and the availability of mechanisms to review or correct AI-supported decisions. Risk was framed less in terms of institutional intent and more in terms of whether governance arrangements function as intended in practice.

In the Spain and Slovenia FGDs, discussions were oriented less toward authority-driven use and more toward education, children, and long-term consequences. Participants focused on risks related to rigidity, loss of contextual judgment, and the cumulative effects of AI-based assessment or categorisation over time. In Spain, this was accompanied by reflections on the need for regulatory approaches that protect against harm without unduly constraining beneficial innovation, particularly in educational contexts. In Slovenia, emphasis was placed more strongly on everyday impact and the risk of standardisation overriding sensitivity to individual circumstances.

At the group level, differences similarly reflected variation in emphasis rather than opposing standards. Participants speaking from professional or expert perspectives tended to focus on procedural clarity, responsibility, and the impact of AI on discretion and decision-making within their fields. Participants speaking primarily from a general citizen perspective more often framed concerns in terms of power imbalance, loss of agency, and vulnerability in interactions with institutions. Where participants referred to children or older adults, concerns centred on long-term consequences, accessibility, and the ability to understand or challenge AI-supported decisions.

Despite these differences, no group articulated a fundamentally different standard of acceptability. Across groups, AI was consistently viewed as potentially useful but inherently risky if introduced without strong safeguards. Human involvement, transparency, and the ability to question or contest outcomes were repeatedly identified as non-negotiable conditions.

Overall, the comparative insight emerging from the FGDs is that variation across countries and groups lies in what participants fear most may go wrong, not in what they expect AI governance to achieve. Shared expectations regarding human control, fairness, data protection, and accountability provide a common normative baseline, while differences in emphasis reflect participants' lived experience with institutions, services, and decision-making processes. This combination of convergence and contextual variation is a defining feature of social acceptance of AI as expressed across the FGDs.

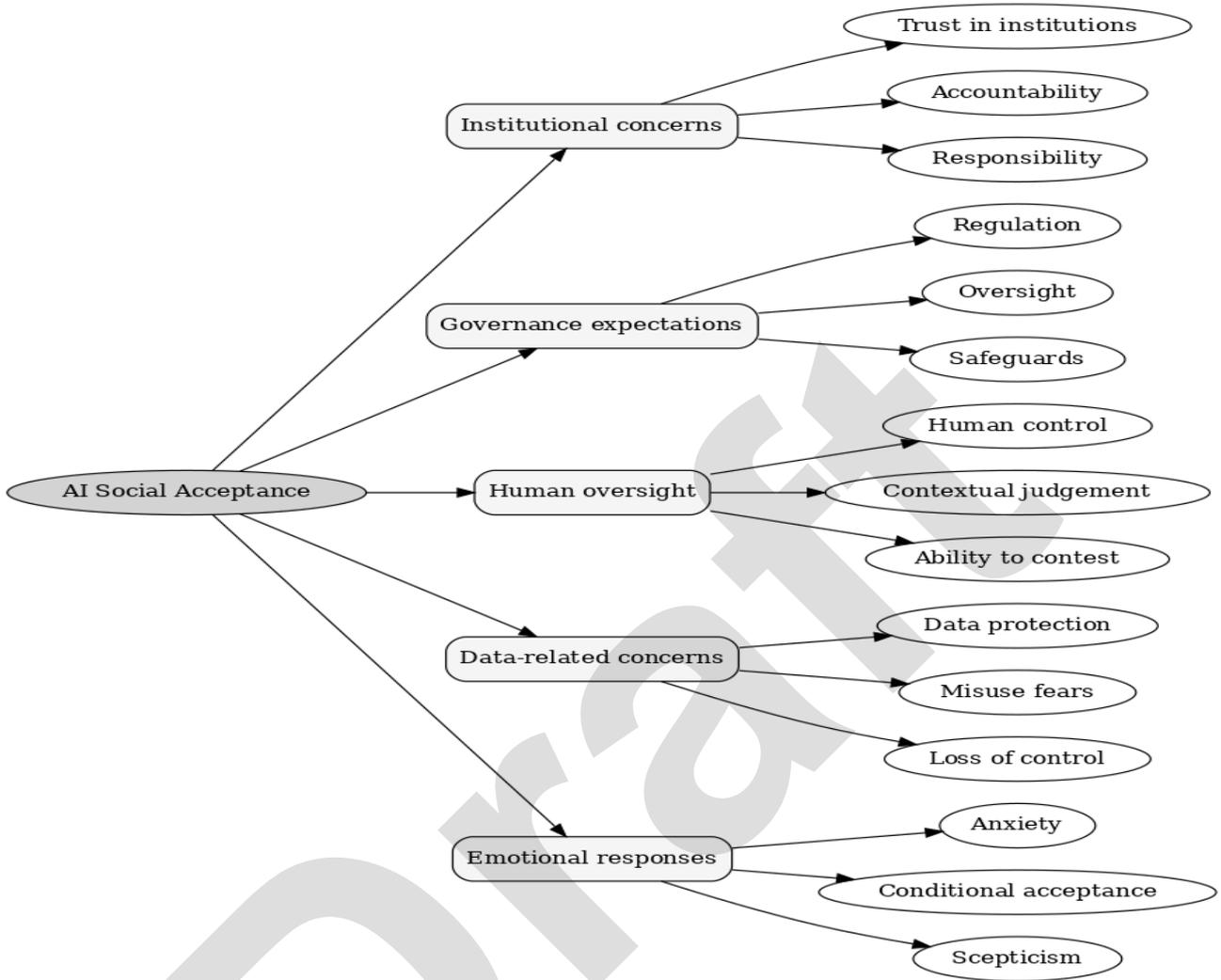


Figure 1. Non-hierarchical representation of cross-cutting themes emerging from FGDs on social acceptance of AI (image generated with ChatGPT 5.2)

## 8. QUANTITATIVE EXTENSION

### 8.1. MEASUREMENT OF AI ACCEPTANCE THROUGH QETAM

To complement and extend the qualitative insights generated through the focus group discussions (FGDs), the study deployed the **QETAM (Quantitative Effect and Technology Acceptance Modelling)** [17] tool across broader and more diverse target groups. While the FGDs provide depth and contextual understanding of perceptions, values, and trust dynamics surrounding AI, the QETAM framework enables systematic measurement of technology acceptance patterns across larger populations, supporting triangulation and comparative analysis.

The QETAM framework applies a blended analytical model that integrates core constructs from the **Technology Acceptance Model (TAM)** and the **Unified Theory of Acceptance and Use of Technology (UTAUT)**, enriched with a gender-sensitive dimension. Specifically, the model examines **Perceived Trust (PT)**, **Perceived Usefulness (PU)**, and **Perceived Ease of Use (PEOU)**, alongside **Social Influence (SI)** and **Facilitating Conditions (FC)**, while incorporating **Gender Influence (GI)** as an additional analytical factor. This integrated structure allows for a multidimensional assessment of behavioural intention to use AI technologies, capturing not only functional and social determinants of acceptance but also gender-related variations in perception, confidence, and engagement with digital systems.

Operationally, the **QETAM** tool was implemented through a **structured questionnaire** that collected both demographic information and scaled responses (1–10) across the six core dimensions (PU, PEOU, PT, SI, FC, and GI). Based on these responses, a composite **Batch Index Unit (BIU)** score was calculated to estimate **overall behavioural intention to use** a given AI technology, with lower scores in any single dimension reducing overall acceptance. In addition, the model enables **comparative analysis** of gender-related differences across all dimensions, supporting examination of how gender dynamics shape AI acceptance patterns.

By combining qualitative interpretation from FGDs with quantitative measurement through QETAM, the methodological design strengthens both contextual depth and analytical breadth. This mixed approach allows the research to identify which factors – such as perceived usefulness, trust, social endorsement, or enabling conditions – most strongly influence AI acceptance across different groups. This understanding, in turn, can help develop specific policies, designs, and communication methods to foster better inclusion, reliability, and fair adoption of AI systems, making certain that these technologies truly serve the broader public. The preliminary results of this study are presented in the following section.

### 8.2. PRELIMINARY RESULTS OF THE TECHNOLOGY ACCEPTANCE SURVEY

Within the framework of task T3.3: Social responsibility and acceptance of AI and data regulations, a tailored survey was conducted to explore how people from diverse backgrounds perceive, understand, and experience Artificial Intelligence (AI) technologies in everyday life, work, and society. The questionnaire was distributed through a dedicated Google Forms link to a diverse audience both within and outside the CERTAIN consortium, as well as to the FGD participants. The questionnaire included questions covering the following areas:

1. **Respondents' demographics**, such as gender, level of education, and workplace location, etc.
2. **General knowledge and familiarity** with the relevant technologies
3. **Key influencing factors for technology acceptance**, including:
  - **Perceived Usefulness (PU)** - Measure of the usefulness of a particular technology
  - **Perceived Ease to Use (PEU)** - Measure of the usability of a particular technology
  - **Perceived Trust (PT)** - Measure of a person's trust in a particular technology

- **Social Influence (SI)** - The degree to which someone is influenced by social norms and their social environment
- **Facilitating Conditions (FC)** - The degree to which an individual believes that an organizational and technological infrastructure exists, and also the degree to which they have the appropriate knowledge and resources to use the system
- **Gender Influence (GI)** - The extent to which an individual's gender identity, roles, norms, or societal expectations related to gender affect their preferences, behaviours, or interactions with a system or technology. This includes how gender-specific expectations may shape attitudes toward adoption, usage, or perceived usefulness of a tool, platform, or practice.

Except for the demographic section, all responses were collected using a **1–10 Likert scale**. The analysis of the responses was conducted using the **QETAM model**, whose methodological framework presented in the previous section.

Following data collection, a **correlation matrix** was generated to calculate the **Behavioural Intention to Use (BIU)** value, which represents the level of acceptance of the respective technology. The BIU number was calculated using the following equation:

$$BIU_i = PU_i \times PEU_i \times PT_i \times SI_i \times FC_i \times GI_i$$

This formulation indicates that a user's **behavioural intention to use a technology** increase when the technology is perceived as useful, easy to use, trustworthy, socially supported, and well facilitated, while also being influenced by gender-related factors such as differences in technology perception, confidence, and usage patterns. A low score in any single factor negatively impacts the overall behavioural intention.

Once calculated, the BIU values were categorized according to the classification presented in the table below.

BIU Number	Behavioural Intention to Use
0-1.000	Low
1.000-10.000	Medium
10.000-50.000	High
50.000-100.000	Very High

Table 1: BIU Categorization

### 8.2.1 Survey Demographics

In the Technology Acceptance and Perception of AI End-User Survey, a total of 55 individuals participated. Almost half of the respondents (47,3%) were aged 25-34, identified as females, and held a master's degree. Most participants reported that their employment or field of study was related to "professionals" (e.g., scientists, engineers, educators, healthcare, ICT professionals). Additionally, 56% of the respondents indicated that they resided in Greece.

Detailed demographic information is illustrated in the figures below.

- Age Group

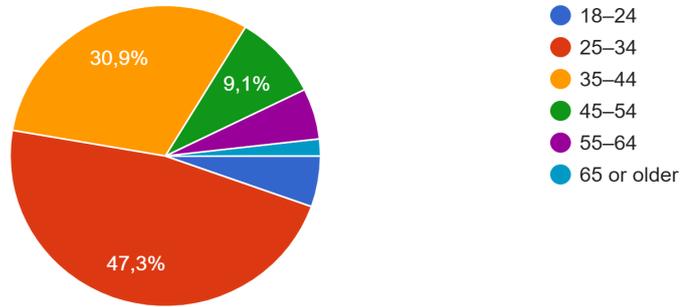


Figure 2. Age Groups of Respondents

- Gender

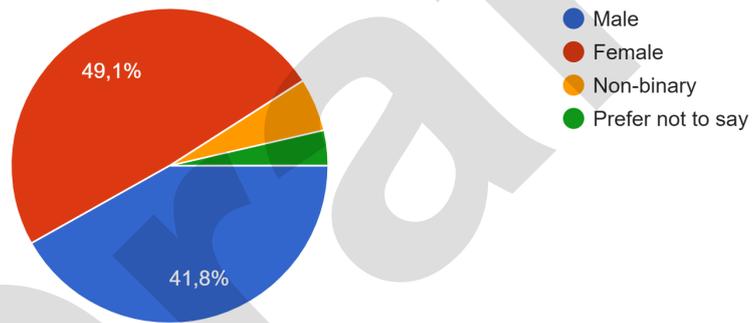


Figure 3. Gender Distribution of Respondents

- Disability:

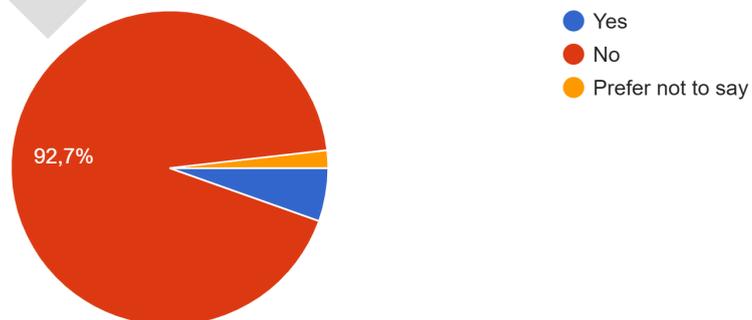


Figure 4. Self-Identification of Disability Among Respondents

▪ Highest completed level of education:

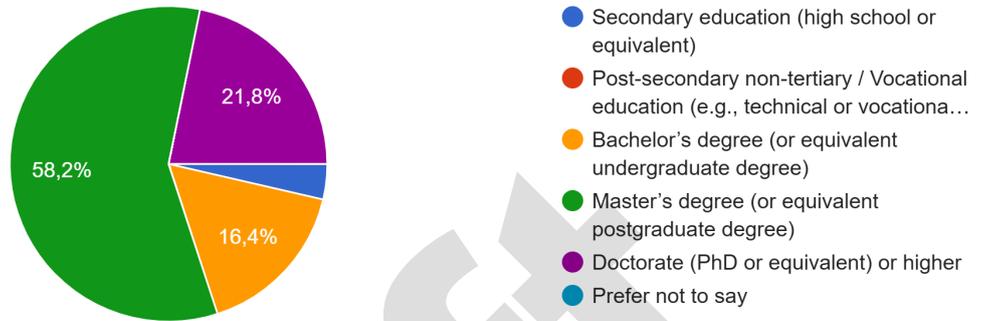


Figure 5. Highest Level of Education of Respondents

▪ Area of Employment/Study:

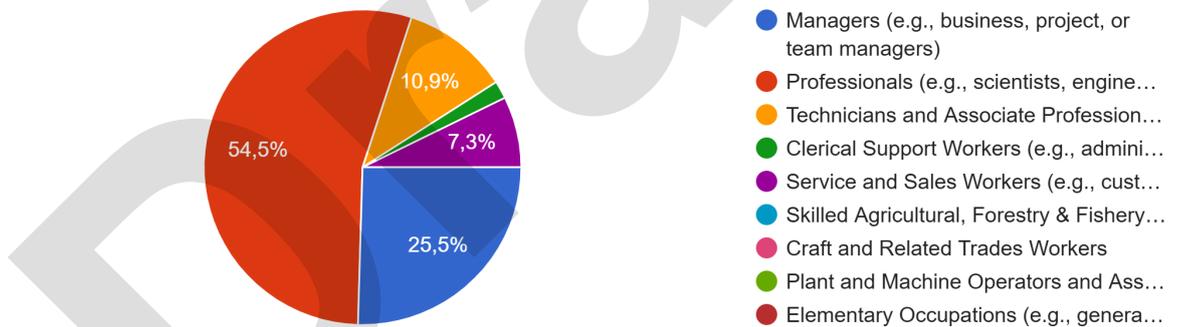


Figure 6. Area of Employment/Study of Respondents

▪ Country of Workplace:

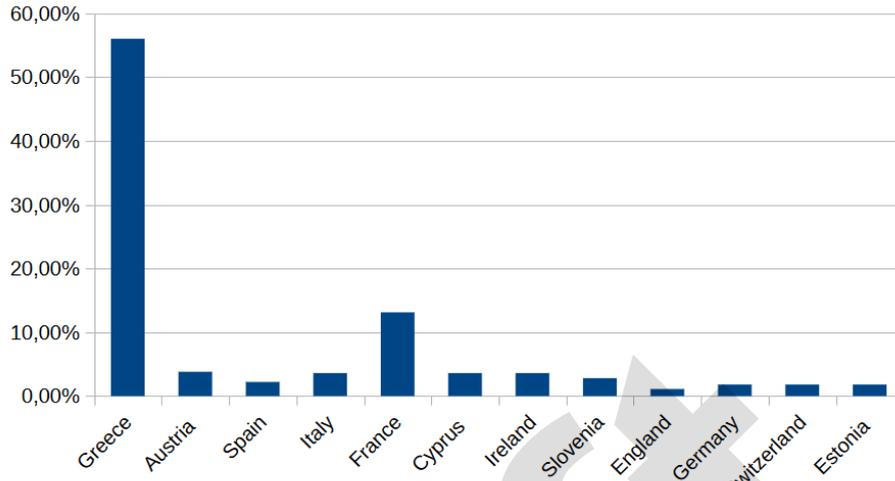


Figure 7. Country of Workplace of Respondents

### 8.2.3 Overall BIU

Among the participants, a very high overall Behavioural Intention to Use (BIU) score was observed. The high BIU value of 134,830 indicates that respondents reported a strong likelihood of engaging with emerging AI technologies in the specific use cases outlined in the survey. Responses to the BIU items consistently reflected agreement with statements about future use, suggesting readiness and openness toward adoption within this sample. These results capture participants' stated intentions at the time of the study and should be interpreted as reflecting self-reported acceptance rather than confirmed future behaviour.

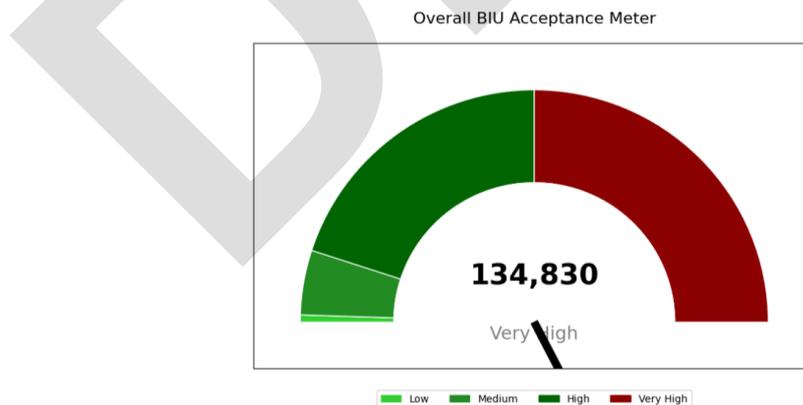


Figure 8. Overall BIU value

### 8.2.4 Average Score Distribution by Factor

The figure below presents the average score distribution across six acceptance factors: Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Perceived Trust (PT), Social Influence (SI), Facilitating Conditions (FC), and Gender Influence (GI). The results show relatively high mean scores across all constructs, suggesting generally favourable evaluations of the AI technology within this sample.

**Perceived Ease of Use (PEOU)** recorded one of the highest average scores (7.6; 17.7%). As a measure of usability, this indicates that respondents generally found the AI system easy to learn and operate. The prominence of PEOU suggests that usability is a strong contributor to positive attitudes toward the technology in this group.

**Facilitating Conditions (FC)** showed the highest overall average score (7.9; 18.3%). This implies that participants perceived adequate organizational and technical infrastructure, as well as sufficient knowledge and resources, to support system use. The strength of FC suggests that respondents feel supported in their potential adoption of the AI system. **Perceived Usefulness (PU)** and **Social Influence (SI)** both yielded relatively high and nearly identical mean scores (7.3; 17.0% and 16.9%, respectively). The PU result indicates that participants consider the AI technology beneficial for their tasks or performance. Meanwhile, the SI score

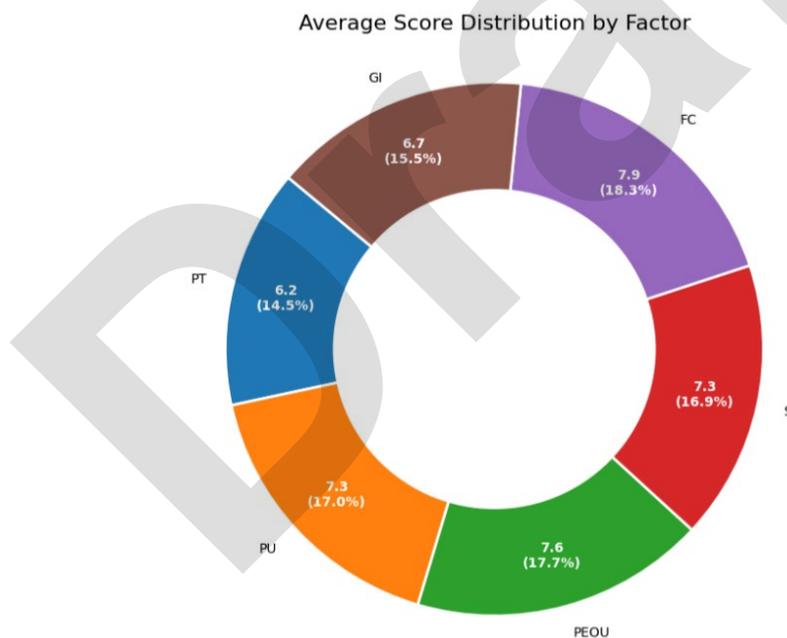


Figure 9. Average Score Distribution by Factor

suggests that social norms, peer expectations, or professional environments may play a meaningful role in shaping respondents' attitudes toward using the system.

**Gender Influence (GI)** demonstrated a moderate average score (6.7; 15.5%). This suggests that gender-related norms or expectations may have some influence on perceptions or adoption intentions, but this effect appears less pronounced compared to structural or usability-related factors.

**Perceived Trust (PT)** recorded the lowest average score among the constructs (6.2; 14.5%). Although still relatively positive, this comparatively lower value indicates that trust-related considerations—such as reliability, transparency, or data security—may represent an area where perceptions are less strong than other acceptance drivers.

Overall, the distribution suggests that structural support (FC) and usability (PEOU) are the most prominent factors in shaping acceptance within this sample, while trust and gender-related influences appear comparatively weaker. The consistently high averages across constructs indicate a generally favourable perception of the AI system, with variation in the relative strength of specific determinants.

### 8.2.5 Factor Correlation Matrix

The correlation matrix illustrates the strength and direction of relationships among the six constructs: Perceived Trust (PT), Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Social Influence (SI), Facilitating Conditions (FC), and Gender Influence (GI). All reported correlations are positive, indicating that increases in one factor are associated with increases in others, although the strength of these relationships varies.

The strongest correlation observed is between **Perceived Ease of Use (PEOU)** and **Facilitating Conditions (FC)** ( $r = 0.80$ ). This suggests a strong association between respondents' perceptions of system usability and their belief that adequate organizational and technical support exists. Participants who perceive sufficient infrastructure and resources also tend to view the system as easier to use. A relatively strong relationship is also evident between **Perceived Trust (PT)** and **Perceived Usefulness (PU)** ( $r = 0.72$ ). This indicates that higher levels of trust in AI technology are associated with stronger perceptions of its usefulness. Similarly, **PU** shows substantial correlations with **Social Influence (SI)** ( $r = 0.68$ ) and **PEOU** ( $r = 0.58$ ), suggesting that perceived benefits are linked both to usability perceptions and to social or normative pressures.

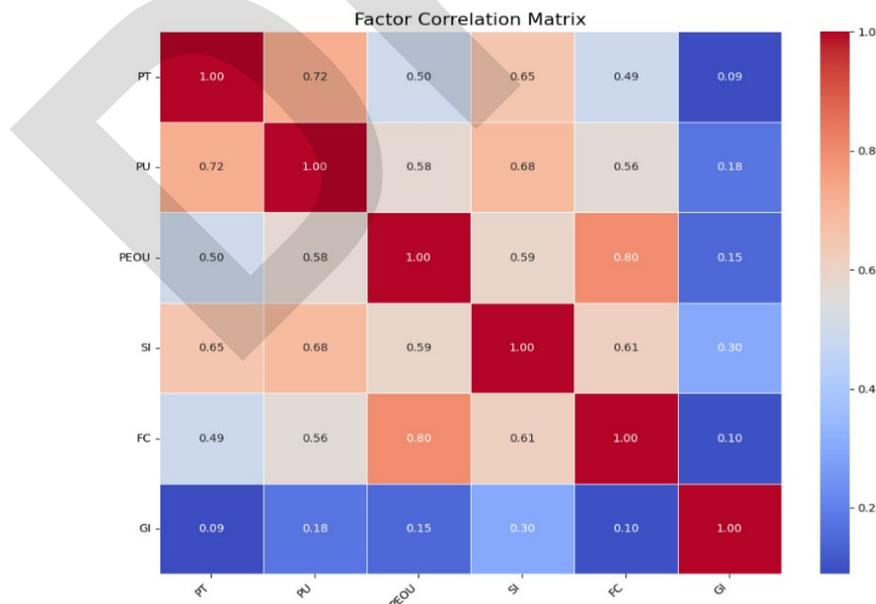


Figure 10. Factor Correlation Matrix

**Social Influence (SI)** demonstrates moderate correlations with most constructs, including **PT** ( $r = 0.65$ ), **FC** ( $r = 0.61$ ), and **PEOU** ( $r = 0.59$ ). This pattern indicates that individuals who report stronger social or normative

influences also tend to report higher trust, better usability perceptions, and stronger beliefs in supporting conditions.

In contrast, **Gender Influence (GI)** exhibits comparatively weak correlations with the other variables. Its strongest association is with **Social Influence (SI)** ( $r = 0.30$ ), while correlations with PT ( $r = 0.09$ ), FC ( $r = 0.10$ ), and PEOU ( $r = 0.15$ ) are minimal. This suggests that, within this sample, gender-related norms or expectations show limited association with the other acceptance determinants.

Overall, the matrix indicates that the core technology acceptance of AI technologies constructs—usefulness, ease of use, trust, social influence, and facilitating conditions—are moderately to strongly interrelated, forming a coherent network of associated perceptions. Gender influence, however, appears more independent, with weaker connections to the primary acceptance factors.

Draft

## 9. REFLECTIONS FOR POLICY AND PRACTICE

The focus group discussions reveal a coherent set of public expectations and concerns regarding artificial intelligence that have direct implications for how AI is communicated about, deployed by institutions, and regulated in the future. While participants across countries and groups differed in emphasis, their perceptions converged around a shared baseline: AI is evaluated primarily through its institutional use, its effects on people, and the safeguards surrounding it. These perceptions point to several implications for policy and practice.

### 9.1. IMPLICATIONS FOR PUBLIC COMMUNICATION

FGD findings suggest that public communication about AI should focus less on technical capability and more on **institutional responsibility and practical consequences**. Participants did not engage with AI as a technological abstraction; instead, they asked who uses it, for what purposes, and with what consequences. Communication strategies that emphasise efficiency, innovation, or technical sophistication without addressing these questions are unlikely to build trust.

Participants consistently expressed uncertainty about how AI systems operate in practice, particularly regarding data use, decision-making authority, and accountability. This indicates a need for communication that is **clear, concrete, and procedural**, explaining when AI is used, what role it plays in decisions, and how human oversight is ensured. Importantly, participants did not expect technical explanations of algorithms; rather, they wanted to understand how AI affects them and what recourse exists if something goes wrong.

FGDs also show that public concern is not driven by hostility to technology, but by fear of misuse, opacity, and loss of control. Communication that acknowledges these concerns – rather than dismissing them or framing AI as inevitable – may be more credible. Participants' repeated emphasis on fairness, human judgment, and accountability suggests that messaging framed around safeguards and responsibility is more likely to resonate than messaging focused on innovation alone.

### 9.2. IMPLICATIONS FOR INSTITUTIONAL USE OF AI

Participants' perceptions imply that institutional use of AI should be **explicitly bounded and visibly governed**. Across FGDs, acceptability depended on whether AI was perceived as assisting humans or replacing them. Institutions that deploy AI in ways that obscure responsibility or reduce opportunities for human judgment risk undermining trust, regardless of the system's technical performance.

FGDs highlight that institutions are expected to retain **clear responsibility for AI-supported decisions**. Participants expressed strong resistance to situations in which decisions could be attributed to "the system" rather than to identifiable actors. This suggests that institutional practices should ensure that AI outputs are treated as inputs into decision-making, not as final determinations, particularly in high-impact areas such as public services, education, healthcare, or enforcement.

Data handling practices emerged as a critical area of concern. Participants were less focused on isolated data breaches than on the broader risk that data could be reused, shared, or repurposed beyond their original context. This implies that institutions should be cautious not only about how data are collected, but also about how future use is constrained and communicated. Visible limits on secondary use and clear separation of functions may be as important for public trust as technical security measures.

FGDs also suggest that institutional sensitivity to context matters. Where participants believed institutions could take individual circumstances into account—through professional judgment or discretionary review—AI use was more readily accepted. Where AI was associated with rigidity or standardisation, it was viewed as unfair. This points to the importance of embedding AI within practices that allow contextual judgment rather than replacing it.

### 9.3. IMPLICATIONS FOR FUTURE REGULATION

Participants' perceptions indicate that regulation is expected to focus less on endorsing AI use and more on **setting boundaries, safeguards, and accountability mechanisms**. Across FGDs, participants did not call for unrestricted adoption of AI; instead, they emphasised the need for rules that define where AI may be used, under what conditions, and with what oversight.

A recurring theme across countries was concern about loss of control once AI systems are in place, particularly regarding data use, decision-making authority, and the ability to challenge outcomes. This suggests that regulatory frameworks that prioritise **contestability, transparency of use, and human oversight** may align more closely with public expectations than those focused solely on risk classification or technical compliance.

FGDs also point to the importance of regulation that is **institutionally credible**. In contexts where participants expressed mistrust of institutional practice, regulation was not assumed to be effective unless it was accompanied by visible enforcement and responsibility. This implies that the perceived legitimacy of regulatory frameworks may matter as much as their formal content.

Participants in several FGDs (notably Armenia and Estonia) linked regulatory adequacy not only to national rules, but to **where AI systems are developed, hosted, and governed**. This suggests that public expectations of regulation extend beyond domestic frameworks to include safeguards for cross-border data flows, jurisdictional accountability, and alignment with trusted legal regimes.

Some participants expressed concern that **overly rigid or slow-moving regulatory approaches could constrain beneficial innovation**, particularly in education-related and digital contexts. This perspective was not framed as opposition to regulation, but as a call for regulatory approaches that balance protection with flexibility and responsiveness to change. From a policy perspective, this suggests that regulation perceived as proportionate, adaptive, and clearly linked to concrete risks may be more readily accepted than frameworks seen as overly restrictive or disconnected from practical use.

Finally, cross-country differences in emphasis suggest that regulation may need to be **sensitive to context**. While core expectations – human control, fairness, data protection – were shared, participants differed in which risks they considered most salient. Regulatory approaches that allow for contextual application, sector-specific safeguards, or differentiated oversight may therefore be better aligned with public perceptions than one-size-fits-all solutions.

The findings suggest that future regulatory efforts should not focus solely on classification of risk or technical conformity assessment, but also on **reinforcing contestability, human oversight, and institutional accountability in ways that are visible to users**. Participants consistently evaluated AI systems through the lens of who is responsible, how decisions can be challenged, and whether safeguards operate in practice. This indicates that regulatory legitimacy depends not only on compliance structures but on procedural clarity and enforceability, which are areas directly addressed by the AI Act's provisions on human oversight, transparency, and post-market monitoring.

At the same time, cross-country differences in emphasis – particularly regarding national security, cross-border data governance, and regulatory proportionality – underscore the importance of **situating AI governance within the EU's broader neighbourhood and partnership framework**. For countries linked to the EU through accession processes, Eastern Partnership engagement, or Extended Partnership and Cooperation Agreements, alignment with EU AI standards carries both regulatory and geopolitical significance. The external dimension of the AI Act therefore becomes central: sustaining trust will require not only harmonised internal rules, but also **coherent cross-border governance, jurisdictional clarity, and proportionate implementation** in partner contexts. In this respect, regulatory alignment under the EU's neighbourhood policy is not simply a matter of legal convergence, but of maintaining social legitimacy and reinforcing shared standards of responsible AI governance across interconnected governance environments.

## 10. WAYS FORWARD

The FGDs highlight a set of shared expectations and context-specific concerns that point to several priority areas for further research. These avenues are intended to deepen understanding of social acceptance of AI, refine governance responses, and address gaps that cannot be resolved through perception data alone.

### From perceptions to practices of human control

While participants across all FGDs emphasised human control as a non-negotiable condition for acceptable AI use, discussions remained largely at a normative level. Further research is needed to examine how human oversight is operationalised in practice across different sectors, including what forms of review, override, and accountability are meaningful to users. Comparative case studies of AI-supported decision-making processes could explore how responsibility is assigned and exercised, and where gaps emerge between formal rules and lived experience.

### Institutional trust and AI governance trajectories

FGD findings suggest that perceptions of AI are filtered through prior experience with institutions, but the mechanisms through which institutional trust shapes AI acceptance remain underexplored. Future research could examine how different governance models, oversight arrangements, and transparency practices affect public confidence over time. Longitudinal or comparative institutional studies would help clarify whether and how trust can be built or restored through AI governance reforms.

### Cross-border and jurisdictional dimensions of AI governance

The findings point to the importance of examining cross-border and jurisdictional dimensions of AI governance. Participants' concerns about data flows, external control, and legal alignment indicate that social acceptance of AI cannot be understood solely within national regulatory frameworks. Further work is needed to explore how people perceive accountability and protection when AI systems are developed, hosted, or governed across borders, and how international cooperation or regulatory alignment shapes trust in practice.

### Data practices and perceived loss of control

Participants consistently articulated data protection concerns in holistic terms, expressing anxiety about loss of control once data enter AI-supported systems. Further research is needed to explore how people understand data flows, reuse, and cross-sector or cross-border data sharing in practice. User-centred research could investigate which forms of information, consent, and control meaningfully address these concerns, and which fail to do so despite formal compliance.

### Fairness, contestability, and lived experience of AI decisions

While fairness emerged as a central evaluative criterion, FGDs primarily captured anticipated risks rather than lived encounters with AI-supported decisions. Further empirical research could focus on how individuals experience fairness and contestability in real AI-mediated interactions, particularly in education, social services, and public administration. Such research would help distinguish between perceived and experienced unfairness and identify where procedural safeguards succeed or fall short.

### Adolescents and youth, long-term impacts, and developmental trajectories

Concerns related to adolescents and young people were prominent in several FGDs, particularly regarding rigidity, early categorisation, and long-term consequences. Further research should examine how AI use in education and child and youth-related services shapes trajectories over time, including impacts on opportunity, self-perception, and institutional pathways. This includes exploring children's and young people's own perspectives, which were largely discussed indirectly in the FGDs.

### AI accessibility and inclusion

The limited engagement with AI accessibility issues in the FGDs reflects, in part, the fact that persons with disabilities were not adequately represented in the discussions due to significant recruitment challenges. As a result, perspectives on how AI-supported systems are experienced by individuals with sensory, cognitive, or

physical accessibility needs remain underexplored in this study. Further research is therefore needed to specifically examine how AI affects accessibility, agency, and the ability to understand, navigate, and challenge AI-supported decisions for persons with disabilities. Dedicated, accessibility-adapted research designs would be required to capture these experiences meaningfully, including alternative recruitment channels, tailored facilitation methods, and engagement with disability-focused organisations.

### Balancing regulation and innovation in practice

FGDs introduced a nuanced concern about regulatory approaches that may unintentionally constrain beneficial innovation. Further research could examine how different regulatory models balance protection and adaptability in practice, particularly in fast-evolving sectors such as education and digital services. Comparative analysis of regulatory sandboxes, pilot schemes, or adaptive governance mechanisms could shed light on how safeguards and innovation coexist.

Taken together, these avenues point toward the need for **interdisciplinary, user-centred, and context-sensitive research** that bridges perception studies with institutional analysis and real-world practice. Building on the shared normative baseline identified in the FGDs, future research can support more responsive, trustworthy, and socially grounded approaches to AI governance.

Draft

## 11. CONCLUSIONS

This study set out to explore how social acceptance of artificial intelligence is shaped by people's lived experience with institutions, decision-making, and public services. Across all focus group discussions, participants did not assess AI as a neutral or purely technical innovation. Instead, AI was consistently evaluated through its institutional embedding and its consequences for accountability, fairness, data protection, and human agency. Acceptance was uniformly framed as conditional rather than absolute, with meaningful human oversight, transparency of purpose, and the ability to question or challenge AI-supported decisions emerging as non-negotiable expectations. These shared patterns formed a common baseline across countries and participant groups, despite differences in national context and emphasis.

At the same time, the analysis revealed that variation in social acceptance is driven less by disagreement over principles than by differing perceptions of institutional risk. Participants across countries expressed similar expectations of what responsible AI governance should achieve, but foregrounded different concerns depending on their experience with public authorities, regulatory practice, and decision-making environments. In some contexts, fears of misuse or diffusion of responsibility dominated; in others, concerns centred on procedural clarity, rigidity, or long-term societal effects. These differences underscore that social acceptance of AI cannot be understood in isolation from institutional trust, governance capacity, and historical experience with public power.

Taken together, the findings suggest that efforts to foster social acceptance of AI must move beyond technical assurances or generic ethical commitments. Next steps should focus on translating shared expectations into concrete governance practices, including clearer accountability arrangements, visible human control, accessible information about data use, and effective mechanisms for review and redress. Further research should deepen understanding of how these conditions operate in specific sectors and decision-making contexts, and how public engagement can be meaningfully integrated into AI governance over time. Strengthening the alignment between technological deployment and institutional practice will be critical to ensuring that AI systems are not only innovative, but socially legitimate and trusted.

Together with D3.1 and D3.2, which addressed the legal and ethical dimensions of AI, this deliverable adds a social perspective to the CERTAIN Project's analytical foundation. By grounding governance considerations in empirically observed expectations around trust, fairness, accountability, and human oversight, it strengthens the basis for Deliverables D3.4 and D3.5, which will develop comprehensive guidelines and compliance tools. In doing so, D3.3 helps ensure that CERTAIN's final outputs are legally robust, ethically sound, and socially responsive.

The findings of D3.3 will also directly enhance WP5 and WP7 by embedding socially grounded expectations into the project's technical tools and demonstrator activities. For WP5, the emphasis in the FGDs on human oversight, fairness, transparency, and accountability provides critical input for the development of trustworthiness testing algorithms (T5.2), lifecycle ontology modelling (T5.3), and AI compliance guidance (T5.6), ensuring that synthetic test data, evaluation protocols, and compliance pathways reflect not only regulatory requirements but also lived expectations around contestability and responsibility. For WP7, the insights into trust, perceived risks, and cross-border governance concerns have the potential of strengthening the co-creation methodology (T7.1) and demonstrator evaluation framework (T7.4) by informing stakeholder engagement, KPI selection, and the assessment of societal impact during pilot deployment. In this way, D3.3 helps bridge social acceptance research with technical validation and real-world testing, reinforcing CERTAIN's integrated approach to responsible and trustworthy AI.

## REFERENCES

- [1] European Union Agency for Fundamental Rights (FRA), *Getting the future right – Artificial intelligence and fundamental rights* (Luxembourg: Publications Office of the European Union, 2021).
- [2] European Commission, *Annex to the Communication to the Commission: Guidelines on the scope of the obligations for general-purpose AI models established by Regulation (EU) 2024/1689 (AI Act) (2025)*.
- [3] AI Now Institute, *AI Now 2025 Landscape Report (2025)*.
- [4] United Kingdom Department for Science, Innovation and Technology, *International AI Safety Report (2025)*.
- [5] World Health Organization, *Ethics and governance of artificial intelligence for health: WHO guidance* (Geneva: WHO, 2021).
- [6] Council of Europe, *Report on the application of artificial intelligence in healthcare and its impact on the “patient–doctor” relationship (2024)*.
- [7] T. M. Maddox et al., ‘Generative AI in Medicine – Evaluating Progress and Challenges’, *New England Journal of Medicine*, 392(24) (2025), 2479–2481.
- [8] P. Zhang and M. N. Kamel Boulos, ‘Generative AI in Medicine and Healthcare: Promises, Opportunities and Challenges’, *Future Internet*, 15(9) (2023), 286.
- [9] United Nations Interregional Crime and Justice Research Institute (UNICRI), *Not just another tool: Public perceptions of police use of artificial intelligence (2023)*.
- [10] Europol Innovation Lab, *AI bias in law enforcement – A practical guide* (Luxembourg: Publications Office of the European Union, 2025).
- [11] Policing Foundation and Royal United Services Institute, *Policing and AI: Opportunities, risks and governance* (RUSI Occasional Paper, 2024).
- [12] European Union Agency for Fundamental Rights (FRA), *Getting the future right – Artificial intelligence and fundamental rights* (Luxembourg: Publications Office of the European Union, 2021).
- [13] Europol Innovation Lab, *AI bias in law enforcement – A practical guide* (Luxembourg: Publications Office of the European Union, 2025).
- [14] Council of Europe, *Regulating the Use of Artificial Intelligence Systems in Education: Preparatory Study for the Development of a Legal Instrument (2024)*.
- [15] Council of Europe, *The State of Artificial Intelligence and Education Across Europe (2024)*.
- [16] See United Kingdom Department for Science, Innovation and Technology, *International AI Safety Report (2025)* and World Health Organization, *Ethics and governance of artificial intelligence for health: WHO guidance* (Geneva: WHO, 2021). See also SaferAI, *Methodology: Risk management ratings for general-purpose AI models* (AI Seoul Summit, 2025).
- [17] K. Karathanasopoulou K, Tsaousi E., Mitsiogianni D, Dimitrakopoulos G., Georgiadis D., Konstantina, et al. Technology Acceptance Modelling for Investigating the Uptake of Electric, Connected, Autonomous, and Shared Mobility Technologies, in Proc. 11th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS) 2025, Portugal, April 2025.

## APPENDIX A – FGD PARTICIPANT BREAKDOWN BY COUNTRY, TARGET GROUP AND GENDER

### Armenia

Academia: 6 females, 2 males

Business: 8 females, 1 male

Immigrants and displaced persons: 6 females, 3 males

Youth: 11 females, 3 males

### Estonia

Youth: 3 females, 3 males

Migrants: 2 females, 4 males

Academia: 1 female, 4 males

### Kazakhstan

Academia: 9 females, 3 males

Business: 6 females, 7 males

Youth: 7 females, 0 males

### Slovenia

Youth: 3 females, 3 males

### Spain

Academia: 4 females, 6 males

Business: 5 females, 6 males

Older adults: 5 males

## APPENDIX B - DATA COLLECTION TOOLS

### Armenia: FGD Protocol

#### PURPOSE OF THE FGD

The aim is to gather nuanced, context-specific insights on how different segments of Armenian society perceive, interpret, and respond to AI integration, especially in high-stakes contexts such as public services, employment, public order, and safety.

The FGD findings will:

Identify **drivers and barriers** to AI acceptance.

Explore **trust dynamics** between citizens, institutions, and developers.

Inform **inclusive policy recommendations** for government, industry, and education.

#### TARGET GROUPS

Separate FGDs for:

**Academia** – researchers, educators, policy analysts in AI, ethics, and governance.

**Business associations** – industry chambers, sectoral unions, tech-driven SMEs, and enterprises using AI.

**Youth (18 - 30)** – students and early-career professionals from diverse fields, ideally representing both urban and rural residents.

**Migrants and displaced persons** – Indian and Iranian nationals resident in Armenia, “relocants” from Ukraine, Belarus and Russia (following the full-scale invasion) as well as Armenians from Karabakh, focusing on their experiences with digital services, access barriers, and perceptions of fairness and inclusion in AI systems.

**Size:** 6 - 8 participants per FGD.

**Sampling:** Stratified purposive sampling (gender, age, region, ethnicity, field/sector diversity).

#### DURATION & FORMAT

**Length:** 60-90 minutes

**Mode:** In-person or hybrid (depending on logistics and participant preference)

**Languages:** Armenian.

**Recording:** Audio (and video if consented) with simultaneous note-taking.

#### ETHICAL CONSIDERATIONS

Informed consent

Participants receive an **information sheet** explaining:

Purpose of the research

Expected duration

Voluntary nature of participation

Confidentiality protections

Use of data in anonymized form

Participants sign a **consent form** (available in Armenian, and/or in English/Russian in the case of migrants, relocates and displaced persons) before starting.

Right to withdraw

Participants may skip any question or leave at any point without penalty.

Recording & data handling

Audio recording used solely for transcription and analysis using GDPR-compliant tools with anonymisation.

Stored securely on encrypted, password-protected devices.

All identifiers removed in transcripts; pseudonyms or codes used in reporting.

Confidentiality

Facilitator reminds participants that while the research team will protect anonymity, confidentiality among participants cannot be guaranteed — request that they do not share others' statements outside the group.

Minimizing harm

Avoid leading or judgmental questions.

Allow breaks if participants become uncomfortable.

Provide contact details for follow-up or support if the discussion triggers concerns.

## FACILITATION ROLES

**Lead facilitator:** Guides discussion, probes for depth, ensures balanced participation.

**Co-facilitator/Observer (where feasible):** Observes group dynamics, monitors time, and co-facilitates as need be.

## FACILITATOR SCRIPT

Welcome & introduction (10 min)

*(Facilitator reads)*

Good [morning/afternoon]. Thank you for joining today's discussion. My name is [name] and I will be facilitating our conversation. This session is part of an international research project exploring how different groups in Armenia view the growing use of artificial intelligence in various areas of life, such as public services, employment, and safety.

There are no right or wrong answers — we are interested in your personal views and experiences. Your participation is voluntary, and you may choose not to answer a question or leave at any time.

We will record the discussion to make sure we capture your words accurately, but your name will not appear in any report. Please also respect the confidentiality of others here and avoid sharing their comments outside this group.

Before we start, please review and sign the consent form. [Pause for consent collection]

Warm-up & rapport building (10 min)  
(Facilitator prompts each participant)

Please introduce yourself: first name, your field of work/study, and one word that comes to mind when you hear “artificial intelligence.”

Main discussion questions (60-80 min)

### **Theme 1 – Perceptions and experiences**

How do you see AI being used in Armenia today?

What AI applications have you personally interacted with?

Which AI uses feel most helpful? Which raise concerns?

### **Theme 2 – Fairness, trust, and transparency**

Do you think AI decisions can be fair? Why or why not?

How much do you trust the public sector, businesses, and universities to develop and use AI responsibly?

What would increase your trust?

### **Theme 3 – Empowerment vs. control**

Do you feel AI increases people’s control over their lives or reduces it?

How does AI affect human dignity and autonomy?

### **Theme 4 – Cross-border AI**

How do you feel about AI technologies developed abroad but used in Armenia?

What do you expect from the implementation of the EU AI Act? Will it improve the situation/alleviate your concerns about AI? Why/why not?

Scenario-based discussion (20-25 min)

#### **Scenario A – Academia:**

##### **Scenario A1**

*An Armenian university adopts an AI-assisted grading and plagiarism detection system trained primarily on English and Russian texts, which fails to accurately process the structure and vocabulary of Armenian-language essays. Students report unfair results and low confidence in the system’s accuracy.*

What could be positive or harmful about such a system?

Should students be informed or allowed to challenge the algorithm’s conclusions?

##### **Scenario A2**

*A consortium of universities pilots an AI-based student engagement platform that tracks attendance, time spent on learning portals, and emotional tone in online discussions to “identify disengaged students early.” Some professors appreciate the data insights, while others feel it promotes surveillance and pressure rather than genuine learning.*

How would you feel about universities collecting and analyzing this kind of behavioral data?

Where should the line be drawn between helpful support and excessive monitoring?

What kind of consent or transparency should be required from students and teachers?

### **Scenario B – Business Associations:**

#### **Scenario B1**

*A local ride-hailing app introduces AI-based “dynamic risk scoring” to manage drivers and passengers, adjusting visibility or priority in the app based on behavioral data. Some drivers claim the system unfairly penalizes them for cancellations or low ratings.*

What are the ethical challenges here?

Should workers have a right to human review or appeal when AI decisions affect their income?

#### **Scenario B2**

*Large Armenian supermarket chains (e.g. Yerevan City) adopt AI-powered dynamic pricing, adjusting food prices in real time based on demand, time of day, and online shopping behavior.*

How do you feel about supermarkets using AI to change prices this way?

Should Armenia set limits on how far AI can alter prices?

### **Scenario C – Youth:**

#### **Scenario C1**

*A popular Armenian social media app introduces AI filters that automatically flag and hide “inappropriate content.” Many young users notice and complain about the fact that posts in regional dialects of Armenian diverging from the standard literary language (e.g. Gyumri, Gavar) and in minority languages (Kurmanci/Ēzdikî) are frequently misclassified and hidden by mistake.*

How do you feel about AI moderating online content?

Should there be more human moderation or more user control over what AI hides?

#### **Scenario C2**

*A social media platform introduces an AI mental-health chatbot for teenagers, offering advice on stress, anxiety, and relationships.*

Would you trust an AI chatbot with sensitive personal issues?

Should Armenia establish special rules for AI tools targeting youth well-being?

### **Scenario D – Migrants and displaced persons**

#### **Scenario D1**

*The Armenian government launches a new digital identification system with AI-powered verification for access to healthcare, social assistance, and education services. The system relies on facial recognition and document scanning to confirm identity. However, relocants and displaced persons frequently experience authentication failures – for example, the system rejecting foreign passports*

*or documents with transliteration inconsistencies – forcing them to travel to service centers for manual verification, which can be particularly burdensome for those living in rural areas..*

How would you feel about using such a system for essential public services?

What risks might arise from relying on automated identity verification?

How could authorities ensure that digital ID systems remain accessible and trustworthy for all residents, regardless of citizenship or document type?

## Scenario D2

*A new municipality-run online portal uses AI to review rental applications and allocate affordable housing to vulnerable groups. However, foreign nationals, relocants, and displaced persons report receiving vague or inconsistent rejection notices, as well as long processing delays, which they attribute to errors in the automated system.*

Would you trust such a digital system to allocate housing or benefits?

How can Armenia ensure that AI-based eligibility checks do not discriminate based on language, nationality, or data gaps?

Closing round (5-10 min)  
(Facilitator asks)

If you could give one recommendation to policymakers or AI developers in Armenia, what would it be?

Thank participants, explain next steps, and share contact information for follow-up.

## DATA MANAGEMENT AND ANALYSIS

Transcribe recordings verbatim; translate into English where necessary.

Apply thematic coding based on: fairness, trust, empowerment/control, transparency, dignity, harm minimization, cross-border views.

Compare patterns between academia, business associations, and youth.

## ANNEX – Consent form

English version

### Project context

This study is being conducted as part of the **Horizon Europe CERTAIN Project**. The project brings together researchers from multiple countries to better understand public perceptions of AI and inform policies for responsible AI governance.

### Purpose of the study

You are invited to take part in a research study about how people in Armenia view and experience artificial intelligence (AI) in different areas of life. The study aims to understand opportunities, concerns, and trust in AI, and will help inform policy and practice in Armenia, wider EU and internationally.

### What participation involves

If you agree to participate, you will join a group discussion lasting about 1 - 1.5 hours. The discussion will be audio recorded (and video recorded if you agree) so we can accurately capture your comments. Notes will also be taken.

### **Voluntary participation**

Your participation is voluntary. You may refuse to answer any question and may withdraw at any time without any negative consequences.

### **Confidentiality**

Your name will not appear in any report. We will use a code or pseudonym instead. While the research team will protect your anonymity, we cannot guarantee that other participants will not share what is said. We ask that you also respect the confidentiality of others in the group.

### **Risks and benefits**

There are no known physical risks. Some topics may make you feel uncomfortable; you can skip any question or take a break. There is no payment for participation, but your views will contribute to better AI-related policies and practices in Armenia and other countries.

### **Use of data**

The data will be used for research purposes only, stored securely on password-protected devices, and retained for the duration of the project implementation before being destroyed.

### **Consent statement**

By signing below, you confirm that:

You have read (or listened to) this information and understand it.

You have had the opportunity to ask questions.

You agree to participate in this focus group discussion.

### **Audio Recording:**

- I agree to be audio recorded
- I do not agree to be audio recorded

### **Video Recording (if applicable):**

- I agree to be video recorded
- I do not agree to be video recorded

**Name:** \_\_\_\_\_

**Signature:** \_\_\_\_\_

**Date:** \_\_\_\_\_

## Estonia: FGD Protocol

### PURPOSE OF THE FGD

The aim is to gather nuanced, context-specific insights on how different segments of Estonian society perceive, interpret, and respond to AI integration, especially in high-stakes contexts such as public services, employment, public order, and safety.

The FGD findings will:

Identify **drivers and barriers** to AI acceptance.

Explore **trust dynamics** between citizens, institutions, and developers.

Inform **inclusive policy recommendations** for government, industry, and education.

### TARGET GROUPS

Separate FGDs for:

**Academia** – researchers, educators, policy analysts in AI, ethics, and governance.

**Business associations** – industry chambers, sectoral unions, tech-driven SMEs, and enterprises using AI.

**Youth (18 - 30)** – students and early-career professionals from diverse fields, ideally representing both urban and rural residents.

**Size:** 6 - 8 participants per FGD.

**Sampling:** Stratified purposive sampling (gender, age, region, ethnicity, field/sector diversity).

### DURATION & FORMAT

**Length:** 60-90 minutes

**Mode:** In-person or hybrid (depending on logistics and participant preference)

**Languages:** Estonian.

**Recording:** Audio (and video if consented) with simultaneous note-taking.

### ETHICAL CONSIDERATIONS

Informed consent

Participants receive an **information sheet** explaining:

Purpose of the research

Expected duration

Voluntary nature of participation

Confidentiality protections

Use of data in anonymized form

Participants sign a **consent form** (available in Estonian and English) before starting.

Right to withdraw

Participants may skip any question or leave at any point without penalty.

#### Recording & data handling

Audio recording used solely for transcription and analysis using GDPR-compliant tools with anonymisation.

Stored securely on encrypted, password-protected devices.

All identifiers removed in transcripts; pseudonyms or codes used in reporting.

#### Confidentiality

Facilitator reminds participants that while the research team will protect anonymity, confidentiality among participants cannot be guaranteed — request that they do not share others' statements outside the group.

#### Minimizing harm

Avoid leading or judgmental questions.

Allow breaks if participants become uncomfortable.

Provide contact details for follow-up or support if the discussion triggers concerns.

#### FACILITATION ROLES

**Lead facilitator:** Guides discussion, probes for depth, ensures balanced participation.

**Co-facilitator/Observer (where feasible):** Observes group dynamics, monitors time, and co-facilitates as need be.

#### FACILITATOR SCRIPT

Welcome & introduction (10 min)

*(Facilitator reads)*

Good [morning/afternoon]. Thank you for joining today's discussion. My name is [name] and I will be facilitating our conversation. This session is part of an international research project exploring how different groups in Estonia view the growing use of artificial intelligence in various areas of life, such as public services, employment, and safety.

There are no right or wrong answers — we are interested in your personal views and experiences. Your participation is voluntary, and you may choose not to answer a question or leave at any time.

We will record the discussion to make sure we capture your words accurately, but your name will not appear in any report. Please also respect the confidentiality of others here and avoid sharing their comments outside this group.

Before we start, please review and sign the consent form. [Pause for consent collection]

Warm-up & rapport building (10 min)  
(Facilitator prompts each participant)

Please introduce yourself: first name, your field of work/study, and one word that comes to mind when you hear "artificial intelligence."

Main discussion questions (60-80 min)

#### **Theme 1 – Perceptions and experiences**

How do you see AI being used in Estonia today?  
What AI applications have you personally interacted with?  
Which AI uses feel most helpful? Which raise concerns?

### **Theme 2 – Fairness, trust, and transparency**

Do you think AI decisions can be fair? Why or why not?  
How much do you trust the public sector, businesses, and universities to develop and use AI responsibly?  
What would increase your trust?

### **Theme 3 – Empowerment vs. control**

Do you feel AI increases people's control over their lives or reduces it?  
How does AI affect human dignity and autonomy?

### **Theme 4 – Cross-border AI**

How do you feel about AI technologies developed abroad but used in Estonia?  
What do you expect from the implementation of the EU AI Act? Will it improve the situation/alleviate your concerns about AI? Why/why not?

Scenario-based discussion (20-25 min)

#### **Scenario A – Academia: Scenario A1**

*The University of Tartu pilots an AI-supported exam system integrated with Estonia's national e-ID and X-Road. The system authenticates students remotely, grades assignments, and predicts performance trends.*

What benefits and risks do you see in linking AI tools directly to Estonia's state digital infrastructure?  
How can fairness and accountability be guaranteed if students challenge the AI's decisions?

#### **Scenario A2**

*Tallinn University establishes an AI-driven research evaluation system that analyses publications, citations, and collaboration networks using Estonia's national research information system (ETIS). The results will be used to decide promotions and allocate funding.*

What opportunities and risks do you see in relying on AI for academic career progression and funding in Estonia?

How can transparency and accountability be ensured if researchers disagree with the AI's assessment?

#### **Scenario B – Business Associations: Scenario B1**

*A Tallinn-based fintech start-up launches an AI tool for automated credit scoring of SMEs, using not only financial data but also social media activity and online reviews.*

Would you trust such an AI-driven credit evaluation for your business? Why or why not?

Should Estonia regulate which kinds of data AI can use for financial decisions?

### **Scenario B2**

*Estonian logistics companies adopt an AI system for route optimisation and workforce management. Trade unions raise concerns that algorithms may cut staff hours or put pressure on delivery workers.*

What opportunities and risks do you see in AI managing work schedules?

How should Estonia balance business innovation with labour rights and fairness?

### **Scenario C – Youth:**

#### **Scenario C1**

*A popular Estonian e-learning platform links with the national e-School system (E-Kool) and integrates AI study analytics. The system tracks performance, behaviour, and attendance across schools via X-Road.*

Would you feel comfortable if an AI linked to national platforms monitored your learning behaviour so closely?

Should there be limits on how much data Estonia's education system can collect about young people?

#### **Scenario C2**

*A new AI career guidance app supported by the Ministry of Education and Research analyses social media, browsing patterns, and e-School results to recommend jobs in Estonia's digital economy.*

Would you use such a service? Why or why not?

What risks could arise if AI predictions shape young people's career choices too narrowly in a small labour market like Estonia's?

### **Scenario D – Older adults:**

#### **Scenario D1**

*The Estonian Social Insurance Board introduces an AI-powered pension advice service integrated with e-ID and X-Road. Older adults can log in to see personalized recommendations about retirement age, pension amount, and tax planning.*

What opportunities does this create for older adults in Estonia?

What risks might arise if pension advice is automated?

How should the government ensure that older adults who are less digitally skilled are not excluded?

#### **Scenario D2**

*Estonian hospitals begin using an AI-driven telemedicine system to monitor elderly patients with chronic illnesses at home. The system automatically alerts doctors if health indicators (blood pressure, glucose, mobility) deviate from the norm.*

Would you trust AI systems to monitor older patients remotely? Why or why not?

How can Estonia ensure that elderly patients understand and consent to data-sharing?  
Should human doctors always review AI alerts before action is taken?

## Scenario E – People with disabilities

### Scenario E1

*Tallinn city government pilots an AI-based public transport assistant designed to help persons with disabilities plan barrier-free travel routes using real-time data (e.g. elevators, ramps, bus accessibility).*

What benefits could this bring for mobility and independence?  
What risks might arise if the AI system gives wrong or incomplete information?  
Should Tallinn ensure human support is still available for travellers with special needs?

### Scenario E2

*An Estonian start-up develops an AI-powered sign language interpretation tool for online education and public services. It automatically translates between Estonian Sign Language (EVK) and written/spoken Estonian.*

How could this improve access for persons with hearing impairments?  
What risks might occur if the AI misinterprets meaning in sensitive contexts (e.g. health, legal)?  
Should Estonia establish quality standards for AI accessibility tools?

Closing round (5-10 min)  
(Facilitator asks)

If you could give one recommendation to policymakers or AI developers in Estonia, what would it be?

Thank participants, explain next steps, and share contact information for follow-up.

## DATA MANAGEMENT AND ANALYSIS

Transcribe recordings verbatim; translate into English where necessary.

Apply thematic coding based on: fairness, trust, empowerment/control, transparency, dignity, harm minimization, cross-border views.

Compare patterns between academia, business associations, and youth.

## ANNEX – Consent form

English version

### Project context

This study is being conducted as part of the **Horizon Europe CERTAIN Project**. The project brings together researchers from multiple countries to better understand public perceptions of AI and inform policies for responsible AI governance.

### **Purpose of the study**

You are invited to take part in a research study about how people in Estonia view and experience artificial intelligence (AI) in different areas of life. The study aims to understand opportunities, concerns, and trust in AI, and will help inform policy and practice in Estonia, wider EU and internationally.

### **What participation involves**

If you agree to participate, you will join a group discussion lasting about 1 - 1.5 hours. The discussion will be audio recorded (and video recorded if you agree) so we can accurately capture your comments. Notes will also be taken.

### **Voluntary participation**

Your participation is voluntary. You may refuse to answer any question and may withdraw at any time without any negative consequences.

### **Confidentiality**

Your name will not appear in any report. We will use a code or pseudonym instead. While the research team will protect your anonymity, we cannot guarantee that other participants will not share what is said. We ask that you also respect the confidentiality of others in the group.

### **Risks and benefits**

There are no known physical risks. Some topics may make you feel uncomfortable; you can skip any question or take a break. There is no payment for participation, but your views will contribute to better AI-related policies and practices in Estonia and other countries.

### **Use of data**

The data will be used for research purposes only, stored securely on password-protected devices, and retained for the duration of the project implementation before being destroyed.

### **Consent statement**

By signing below, you confirm that:

You have read (or listened to) this information and understand it.

You have had the opportunity to ask questions.

You agree to participate in this focus group discussion.

### **Audio Recording:**

- I agree to be audio recorded
- I do not agree to be audio recorded

### **Video Recording (if applicable):**

- I agree to be video recorded
- I do not agree to be video recorded

**Name:** \_\_\_\_\_

**Signature:** \_\_\_\_\_

**Date:** \_\_\_\_\_

## Kazakhstan: FGD Protocol

### PURPOSE OF THE FGD

The aim is to gather nuanced, context-specific insights on how different segments of Kazakh society perceive, interpret, and respond to AI integration, especially in high-stakes contexts such as public services, employment, public order, and safety.

The FGD findings will:

Identify **drivers and barriers** to AI acceptance.

Explore **trust dynamics** between citizens, institutions, and developers.

Inform **inclusive policy recommendations** for government, industry, and education.

### TARGET GROUPS

Separate FGDs for:

**Academia** – researchers, educators, policy analysts in AI, ethics, and governance.

**Business associations** – industry chambers, sectoral unions, tech-driven SMEs, and enterprises using AI.

**Youth (18 - 30)** – students and early-career professionals from diverse fields, ideally representing both urban and rural residents.

**Size:** 7 - 10 participants per FGD.

**Sampling:** Stratified purposive sampling (gender, age, region, ethnicity, field/sector diversity).

### DURATION & FORMAT

**Length:** 90 - 120 minutes

**Mode:** In-person or hybrid (depending on logistics and participant preference)

**Languages:** Kazakh and/or Russian, depending on group composition. Facilitator should be bilingual or co-facilitated.

**Recording:** Audio (and video if consented) with simultaneous note-taking.

### ETHICAL CONSIDERATIONS

Informed consent

Participants receive an **information sheet** explaining:

Purpose of the research

Expected duration

Voluntary nature of participation

Confidentiality protections

Use of data in anonymized form

Participants sign a **consent form** (available in Kazakh and Russian) before starting.

Right to withdraw

Participants may skip any question or leave at any point without penalty.

Recording & data handling

Audio recording used solely for transcription and analysis using GDPR-compliant tools with anonymisation.

Stored securely on encrypted, password-protected devices.

All identifiers removed in transcripts; pseudonyms or codes used in reporting.

Confidentiality

Facilitator reminds participants that while the research team will protect anonymity, confidentiality among participants cannot be guaranteed — request that they do not share others' statements outside the group.

Minimizing harm

Avoid leading or judgmental questions.

Allow breaks if participants become uncomfortable.

Provide contact details for follow-up or support if the discussion triggers concerns.

#### FACILITATION ROLES

**Lead facilitator:** Guides discussion, probes for depth, ensures balanced participation.

**Co-facilitator/Observer (where feasible):** Observes group dynamics, monitors time, and co-facilitates as need be.

#### FACILITATOR SCRIPT

Welcome & introduction (10 min)

*(Facilitator reads)*

Good [morning/afternoon]. Thank you for joining today's discussion. My name is [name] and I will be facilitating our conversation. This session is part of an international research project exploring how different groups in Kazakhstan view the growing use of artificial intelligence in various areas of life, such as public services, employment, and safety.

There are no right or wrong answers — we are interested in your personal views and experiences. Your participation is voluntary, and you may choose not to answer a question or leave at any time.

We will record the discussion to make sure we capture your words accurately, but your name will not appear in any report. Please also respect the confidentiality of others here and avoid sharing their comments outside this group.

Before we start, please review and sign the consent form. [Pause for consent collection]

Warm-up & rapport building (10 min)

(Facilitator prompts each participant)

Please introduce yourself: first name, your field of work/study, and one word that comes to mind when you hear "artificial intelligence."

Main discussion questions (60-80 min)

### **Theme 1 – Perceptions and experiences**

How do you see AI being used in Kazakhstan today?

What AI applications have you personally interacted with?

Which AI uses feel most helpful? Which raise concerns?

### **Theme 2 – Fairness, trust, and transparency**

Do you think AI decisions can be fair? Why or why not?

How much do you trust the public sector, businesses, and universities to develop and use AI responsibly?

What would increase your trust?

### **Theme 3 – Empowerment vs. control**

Do you feel AI increases people's control over their lives or reduces it?

How does AI affect human dignity and autonomy?

### **Theme 4 – Cross-border AI**

How do you feel about AI technologies developed abroad but used in Kazakhstan?

Should Kazakhstan align with international AI standards like the EU AI Act? Why or why not?

Scenario-based discussion (20-25 min)

#### **Scenario A – Academia:**

##### **Scenario A1**

*A Kazakh university introduces an AI-based student evaluation system that grades assignments and predicts future performance, reducing human grading time by 70%.*

What opportunities do you see? What are the risks?

How to ensure fairness?

##### **Scenario A2**

*A Kazakh national research council launches a grant competition for AI-driven public health tools. The winning project proposes to create an algorithm that predicts which patients are likely to develop chronic diseases, using both medical records and lifestyle data from wearable devices.*

How do you feel about academic researchers having access to such personal data for health prediction?

What safeguards should be in place to protect privacy while enabling innovation?

How might public trust be affected if such tools are used in Kazakhstan?

#### **Scenario B – Business Associations:**

##### **Scenario B1**

*An AI-powered financial risk scoring tool is introduced by a Kazakh bank to decide on small business loan approvals. The algorithm's logic is not fully disclosed for "commercial confidentiality reasons."*

Would you trust such a system to make fair loan decisions?

Should there be a legal requirement for AI transparency in financial services?  
How might lack of transparency impact business confidence?

### **Scenario B2**

*A major logistics company in Kazakhstan adopts AI-based route optimization and driver monitoring. The system tracks drivers' facial expressions to detect fatigue and automatically alerts supervisors if it detects signs of drowsiness."*

Does this strike you as a positive safety measure or excessive surveillance?  
How could employers balance safety benefits with employee privacy?  
Should employee consent be required for such systems?

### **Scenario C – Youth: Scenario C1**

*A large online marketplace in Kazakhstan introduces an AI system that tracks your shopping habits – such as what you search for, how long you look at a product, and what you put in your cart but don't buy. Based on this data, it automatically changes the prices you see, offers you special discounts, or promotes certain products over others."*

How do you feel about AI using your shopping behavior to personalize prices and offers?  
Would this make you trust the platform more, or less? Why?  
Should there be rules about how much a company can change prices based on individual behavior?  
Would you want to know exactly how the AI decides what prices or products you see?

### **Scenario C2**

*A new AI-powered career guidance app for students analyses your social media posts and online behavior to suggest suitable professions and employers.*

Would you use such a service? Why or why not?  
What risks could arise from basing career advice on personal online behavior?  
Should there be rules about how AI interprets or uses your online history?

Closing round (5-10 min)  
(Facilitator asks)

If you could give one recommendation to policymakers or AI developers in Kazakhstan, what would it be?

Thank participants, explain next steps, and share contact information for follow-up.

## DATA MANAGEMENT AND ANALYSIS

Transcribe recordings verbatim; translate into English where necessary.

Apply thematic coding based on: fairness, trust, empowerment/control, transparency, dignity, harm minimization, cross-border views.

Compare patterns between academia, business associations, and youth.

## ANNEX – Consent form

English version

### **Project context**

This study is being conducted as part of the **Horizon Europe CERTAIN Project**. The project brings together researchers from multiple countries to better understand public perceptions of AI and inform policies for responsible AI governance.

### **Purpose of the study**

You are invited to take part in a research study about how people in Kazakhstan view and experience artificial intelligence (AI) in different areas of life. The study aims to understand opportunities, concerns, and trust in AI, and will help inform policy and practice in Kazakhstan and internationally.

### **What participation involves**

If you agree to participate, you will join a group discussion lasting about 1.5 - 2 hours. The discussion will be audio recorded (and video recorded if you agree) so we can accurately capture your comments. Notes will also be taken.

### **Voluntary participation**

Your participation is voluntary. You may refuse to answer any question and may withdraw at any time without any negative consequences.

### **Confidentiality**

Your name will not appear in any report. We will use a code or pseudonym instead. While the research team will protect your anonymity, we cannot guarantee that other participants will not share what is said. We ask that you also respect the confidentiality of others in the group.

### **Risks and benefits**

There are no known physical risks. Some topics may make you feel uncomfortable; you can skip any question or take a break. There is no payment for participation, but your views will contribute to better AI-related policies and practices in Kazakhstan and other countries.

### **Use of data**

The data will be used for research purposes only, stored securely on password-protected devices, and retained for the duration of the project implementation before being destroyed.

### **Consent statement**

By signing below, you confirm that:

You have read (or listened to) this information and understand it.

You have had the opportunity to ask questions.

You agree to participate in this focus group discussion.

**Audio Recording:**

- I agree to be audio recorded
- I do not agree to be audio recorded

**Video Recording (if applicable):**

- I agree to be video recorded
- I do not agree to be video recorded

**Name:** \_\_\_\_\_

**Signature:** \_\_\_\_\_

**Date:** \_\_\_\_\_

Draft

## Slovenia: FGD Protocol

### Purpose of the FGD

The aim is to gather nuanced, context-specific insights on how different segments of Slovenian society perceive, interpret, and respond to AI integration, especially in high-stakes contexts such as public services, employment, public order, and safety. The FGD will be conducted alongside questionnaire data to provide complementary qualitative insights.

The FGD findings will:

- Identify drivers and barriers to AI acceptance.
- Explore trust dynamics between citizens, institutions, and developers.
- Inform inclusive policy recommendations for government, industry, and education.

### Target groups

Primary target group:

**Doctoral Candidates and Academic Assistants** The study population comprised doctoral candidates who in many cases concurrently work as teaching and research assistants, thereby encompassing the responsibilities and characteristics of both the students and junior academic staff.  
**Size:** 2-4 participants per FGD (adapted to available participants).

**Sampling:** Stratified purposive sampling (gender, age, region, ethnicity, field/sector diversity).

### Duration and format

Length: 90 - 120 minutes

Mode: MS Teams (online)

Languages: English

Recording: Audio and video (if consented) with MS Teams transcription enabled. Transcription will be anonymized using GDPR-compliant AI tools with anonymous recording and transcription.

Ethical considerations

### Informed consent

Participants receive an information sheet explaining:

- Purpose of the research
- Expected duration
- Voluntary nature of participation
- Confidentiality protections
- Use of data in anonymized form
- Use of AI transcription tools (MS Teams) with GDPR compliance

Participants sign a consent form (available in Slovenian and English) before starting.

### Right to withdraw

Participants may skip any question or leave at any point without penalty.

### Recording & data handling

Audio and video recording (if consented) used solely for transcription and analysis.

Transcription will use GDPR-compliant AI tools (MS Teams) with anonymous recording and transcription, stored securely on encrypted, password-protected devices.

All identifiers removed in transcripts; pseudonyms or codes used in reporting.

### **Confidentiality**

Facilitator reminds participants that while the research team will protect anonymity, confidentiality among participants cannot be guaranteed --- request that they do not share others' statements outside the group.

### **Minimizing harm**

- Avoid leading or judgmental questions.
- Allow breaks if participants become uncomfortable.
- Provide contact details for follow-up or support if the discussion triggers concerns.

## **WELCOME AND INTRODUCTION (Facilitator script, 5 min)**

(Facilitator reads)

Hi everyone. Thank you for joining today's discussion. I will be facilitating our conversation that is part of an international research project exploring how young people in Slovenia view the increasing use of artificial intelligence in different areas of life, focusing on research/academia where applicable. There are no right or wrong answers. We are interested in your personal views and experiences, and you may skip any question or leave at any point. We will record the session for transcription so we capture your words accurately, but your name will never appear in any report and everything will be anonymized.

Please also respect the privacy of everyone here and avoid sharing others' comments outside the group.

## **WARM-UP AND RAPPORT BUILDING (3–5 min)**

(Facilitator prompts each participant)

Please introduce yourself with your study field, and one experience or one word that comes to mind when you hear "AI."

## **SHORT MAIN DISCUSSION (5–10 MIN)**

(Facilitator explains)

Before we move into the scenarios, I will go over some themes that emerge when we think about and use AI in our daily life. I will give a few examples to set the scene, and then I will invite your views.

## **THEME 1 — EVERYDAY AI USE**

(Facilitator script)

AI appears in many parts of daily academic and student life, from the recommendation systems that shape what we see on platforms such as TikTok, Instagram, or YouTube, to plagiarism detection and writing-analysis tools used within the university. Many people also rely on AI-based search and summarization tools for literature review, or coding assistants for debugging and analysis. In everyday routines, AI shows up through translation tools, photo filters, and various chatbots used to support work or study.

With these examples in mind, my question is: **Where do you notice AI affecting your daily life as a student or in your academic work?**

## THEME 2 — BENEFITS AND CONCERNS

(Facilitator script)

AI tools can be extremely helpful. They speed up reading, summarizing, and analysis; they support fast translation and terminology checking; and they automate routine tasks such as scheduling or organizing. At the same time, AI systems can introduce new concerns. Deepfake or manipulated content is becoming more common, surveillance and data tracking occur across many platforms, automated systems often make decisions that are difficult to challenge, and AI-generated misinformation spreads quickly.

Thinking about both sides, the question is: **What feels most useful, and what concerns you the most about AI in your daily work or student life?**

(Allow brief answers.)

## THEME 3 — TRUST

(Facilitator script)

Trust often depends on how transparent a system is. Some platforms clearly explain what data they collect and how they use it, while others remain vague. Previous incidents where tools mishandled data or produced misleading results can also influence trust. In general, people tend to trust systems they understand and distrust those that behave unpredictably or without explanation.

With that in mind, the question is: **What would an AI tool need to show you — such as how it handles your data, how reliable it is, or how its outputs are generated — for you to trust it in your academic or personal work?**

## TRANSITION TO SCENARIOS

(Facilitator script)

Thank you. This gives us a shared starting point. We will now review some concrete scenarios, and I will ask for your reactions to each one.

### SCENARIO 1 — ACADEMIC INTEGRITY

**Full facilitator script**

In the first example, imagine that universities in Slovenia introduce an AI system designed to detect plagiarism in student work. The tool automatically compares assignments and theses with large databases of online content and previous submissions, and it highlights possible overlaps or sections that appear suspicious. The intention is to support academic integrity and reduce cheating, but there is also a possibility of false positives, incorrect matches, and added pressure on students.

My questions for you are:

- How do you feel about universities relying on AI to detect plagiarism?
- What benefits do you see, and what concerns would you have if you were evaluated by such a system?
- And what kinds of safeguards would you expect universities to put in place?

### SCENARIO 2 — AI IN MENTAL HEALTH

**Full facilitator script**

In the next example, picture a popular social media platform introducing an AI chatbot for mental-health support. The chatbot provides stress-management tips, helps users reflect on their emotions, and occasionally recommends seeking professional help. Young people often use such tools

privately because they feel anonymous, but there are concerns about data collection, privacy, accuracy, and whether an AI should take on roles that require expert judgment.

My questions here are:

- Would you personally feel comfortable using a chatbot for stress, anxiety, or emotional problems?
- What do you see as the main risks and limitations?
- And what rules or protections should exist when an AI tool targets teenagers or young adults with mental-health advice?

### **SCENARIO 3 — EMPLOYMENT SCREENING**

#### **Full facilitator script**

In this example, a large employer in Slovenia uses AI tools to screen job applications. The system automatically evaluates CVs, ranks candidates, and recommends who should be invited for interviews. This can speed up the process and reduce workload, but concerns arise about how the AI judges applicants, whether it introduces biases, whether it disadvantages certain groups, and whether people can challenge the decision.

My questions are:

- How do you feel about companies using AI to shortlist candidates for jobs?
- Do you see it as fair and efficient, or as something that could reinforce bias or reduce transparency?
- And what level of human involvement would you expect in the final decisions?

### **SCENARIO 4 — PREDICTIVE POLICING**

#### **Full facilitator script**

Here, Slovenian police agencies deploy an AI system that analyzes past crime data to predict where crimes are likely to occur. The system then recommends how police should allocate patrols or resources. Supporters argue that this helps prevent crime and improves efficiency, while critics worry about reinforcing existing bias in data, reducing accountability, and affecting certain neighborhoods more than others.

My questions are:

- What is your initial reaction to police using AI to guide patrol routes?
- Do you see potential benefits, and what are the main concerns that come to mind?
- How important is transparency or public oversight in such a system?

### **SCENARIO 5 — HEALTHCARE DIAGNOSIS**

#### **Full facilitator script**

In the final example, hospitals in Slovenia adopt AI diagnostic tools to help doctors detect early signs of cancer from imaging data. These systems can process scans rapidly and notice patterns that humans might miss, but they can also make errors, and the way they reach their decisions is not always easy to interpret. The question becomes how hospitals balance technological support with clinical judgment, and how patients understand and trust this process.

My questions are:

How do you feel about hospitals using AI to assist in diagnosing diseases such as cancer?

Would you see this as increasing safety and accuracy, or would you worry about mistakes or lack of explanation?

And who should be responsible when an AI-assisted diagnosis turns out to be wrong?

### CLOSING ROUND (3–5 min)

(Facilitator asks)

If you could give one recommendation to policymakers or AI developers in Slovenia, what would it be?

Thank participants and explain next steps.

### Annex - Consent forms

#### (English version)

**Project context:** This study is being conducted as part of the Horizon Europe CERTAIN Project. The project brings together researchers from multiple countries to better understand public perceptions of AI and inform policies for responsible AI governance.

**Purpose of the study:** You are invited to take part in a research study about how people in Slovenia view and experience artificial intelligence (AI) in different areas of life. The study aims to understand opportunities, concerns, and trust in AI, and will help inform policy and practice in Slovenia, wider EU and internationally.

**What participation involves:** If you agree to participate, you will join an online group discussion via MS Teams lasting about 1.5 - 2 hours. The discussion will be audio and video recorded (if you agree) using MS Teams transcription so we can accurately capture your comments. The transcription will be anonymized using GDPR-compliant AI tools. Notes may also be taken.

**Voluntary participation:** Your participation is voluntary. You may refuse to answer any question and may withdraw at any time without any negative consequences.

**Confidentiality:** Your name will not appear in any report. We will use a code or pseudonym instead. While the research team will protect your anonymity, we cannot guarantee that other participants will not share what is said. We ask that you also respect the confidentiality of others in the group.

**Risks and benefits:** There are no known physical risks. Some topics may make you feel uncomfortable; you can skip any question or take a break. There is no payment for participation, but your views will contribute to better AI-related policies and practices in Slovenia and other countries.

**Use of data:** The data will be used for research purposes only, stored securely on password-protected devices, and retained for the duration of the project implementation before being destroyed.

**Consent statement:** By signing below, you confirm that:

- You have read (or listened to) this information and understand it.
- You have had the opportunity to ask questions.
- You agree to participate in this focus group discussion.
- You understand that MS Teams AI transcription will be used and that transcripts will be anonymized.

Audio Recording:  I agree to be audio recorded  I do not agree to be audio recorded

Video Recording:  I agree to be video recorded  I do not agree to be video recorded

Name: \_\_\_\_\_ Date: \_\_\_\_\_ Signature: \_\_\_\_\_

## Slovenia: Survey questionnaire<sup>2</sup>

### Section A: Background Information

1. Age: \_\_\_\_
2. Gender: \_\_\_\_
3. Region of residence: \_\_\_\_
4. Occupation/Field of study: \_\_\_\_
5. Disability status:  I identify as a person with a disability /  I do not identify as a person with a disability /  Prefer not to say

### Section B: Digital skills baseline

How confident are you in ...

6. ... installing or setting up new software / mobile apps?

- Not confident at all
- Slightly confident
- Moderately confident
- Very confident
- Extremely confident

7. ... managing online privacy and security settings?

- Not confident at all
- Slightly confident
- Moderately confident
- Very confident
- Extremely confident

8. ... using online banking / e-government services?

- Not confident at all

---

<sup>2</sup> Survey was deployed in Slovenia only due to participants' preferences.

- Slightly confident
- Moderately confident
- Very confident
- Extremely confident

9. ... assessing if the information I find online is trustworthy?

- Not confident at all
- Slightly confident
- Moderately confident
- Very confident
- Extremely confident

### Section C: Perceptions and Experiences

10. *AI is already influencing my work or studies.*

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

11. Have you personally interacted with AI applications?

- Yes (please specify): \_\_\_\_\_
- No

12. Which AI uses do you find most helpful?

- [Open text box]

13. Which AI uses raise concerns for you?

- [Open text box]

### Section D: Fairness, Trust, and Transparency

14. *AI are capable of making fair decisions.*

- Strongly disagree

- Disagree
- Neutral
- Agree
- Strongly agree

15. *I trust universities, businesses, and the public sector in Slovenia to use AI responsibly.*

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

16. *I trust AI developers to prioritize ethical use of AI.*

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

17. *I believe that AI systems should always include a possibility for human appeal/override.*

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

18. *AI systems treat people from different groups equally and without bias.*

- Strongly disagree
- Disagree
- Neutral

- Agree
- Strongly agree

19. What would increase your trust in AI?

- [Open text box]

### Section E: Empowerment and Control

20. *AI increases people's control over their lives.*

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

21. *AI reduces people's autonomy and dignity.*

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

22. *People like me have a say in how AI is designed and used.*

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

### Section F: Cross-Border AI

23. *I am comfortable with AI technologies developed abroad being used in Slovenia.*

- Strongly disagree

- Disagree
- Neutral
- Agree
- Strongly agree

24. *I believe the EU AI Act will improve trust in AI.*

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

25. *Please explain your response.*

- [Open text box]

### Section G: Scenarios

(Short scenario descriptions mentioned in FGD followed by rating scales and comments.)

#### Example: Academic Integrity

*An AI tool is introduced at universities to detect plagiarism in student work.*

- How beneficial do you find this? (1 = Not at all, 5 = Very much)
- How concerning do you find this? (1 = Not at all, 5 = Very much)
- Comments: [Open text box]

#### Example: AI in Mental Health

*A social media platform offers an AI chatbot that provides stress management tips for teenagers.*

- How beneficial do you find this? (1–5)
- How concerning do you find this? (1–5)
- Comments: [Open text box]

[Other scenarios from the FGD guide can be adapted similarly.]

### Section H: Final Recommendation

26. If you could give one recommendation to policymakers or AI developers in Slovenia and the EU, what would it be?

- [Open text box]

## Spain: FGD Protocol

### PURPOSE OF THE FGD

The aim is to gather nuanced, context-specific insights on how different segments of Spanish society perceive, interpret, and respond to AI integration, especially in high-stakes contexts such as public services, employment, public order, and safety.

The FGD findings will:

Identify **drivers and barriers** to AI acceptance.

Explore **trust dynamics** between citizens, institutions, and developers.

Inform **inclusive policy recommendations** for government, industry, and education.

### TARGET GROUPS

Separate FGDs for:

**Academia** – researchers, educators, policy analysts in AI, ethics, and governance.

**Business associations** – industry chambers, sectoral unions, tech-driven SMEs, and enterprises using AI.

**Youth (18 - 30)** – students and early-career professionals from diverse fields, ideally representing both urban and rural residents.

**Older adults (60+)**

**People with disabilities**

**Migrants** – representing different backgrounds (migrants from Hispanoamerica, migrants from EU candidate countries such as Moldova and Ukraine, migrants from North Africa, etc.) with approximately equal representation of genders.

**Size:** 6 - 8 participants per FGD.

**Sampling:** Stratified purposive sampling (gender, age, region, ethnicity, field/sector diversity).

### DURATION & FORMAT

**Length:** 60-90 minutes

**Mode:** In-person or hybrid (depending on logistics and participant preference)

**Languages:** Spanish.

**Recording:** Audio (and video if consented) with simultaneous note-taking.

### ETHICAL CONSIDERATIONS

Informed consent

Participants receive an **information sheet** explaining:

Purpose of the research

Expected duration

Voluntary nature of participation

## Confidentiality protections

Use of data in anonymized form

Participants sign a **consent form** (available in Spanish) before starting.

Right to withdraw

Participants may skip any question or leave at any point without penalty.

Recording & data handling

Audio recording used solely for transcription and analysis using GDPR-compliant tools with anonymisation.

Stored securely on encrypted, password-protected devices.

All identifiers removed in transcripts; pseudonyms or codes used in reporting.

Confidentiality

Facilitator reminds participants that while the research team will protect anonymity, confidentiality among participants cannot be guaranteed — request that they do not share others' statements outside the group.

Minimizing harm

Avoid leading or judgmental questions.

Allow breaks if participants become uncomfortable.

Provide contact details for follow-up or support if the discussion triggers concerns.

## FACILITATION ROLES

**Lead facilitator:** Guides discussion, probes for depth, ensures balanced participation.

**Co-facilitator/Observer (where feasible):** Observes group dynamics, monitors time, and co-facilitates as need be.

## FACILITATOR SCRIPT

Welcome & introduction (10 min)

*(Facilitator reads)*

Good [morning/afternoon]. Thank you for joining today's discussion. My name is [name] and I will be facilitating our conversation. This session is part of an international research project exploring how different groups in Spain view the growing use of artificial intelligence in various areas of life, such as public services, employment, and safety.

There are no right or wrong answers — we are interested in your personal views and experiences. Your participation is voluntary, and you may choose not to answer a question or leave at any time.

We will record the discussion to make sure we capture your words accurately, but your name will not appear in any report. Please also respect the confidentiality of others here and avoid sharing their comments outside this group.

Before we start, please review and sign the consent form. [Pause for consent collection]

Warm-up & rapport building (10 min)  
(Facilitator prompts each participant)

Please introduce yourself: first name, your field of work/study, and one word that comes to mind when you hear “artificial intelligence.”

Main discussion questions (60-80 min)

### **Theme 1 – Perceptions and experiences**

How do you see AI being used in Spain today?

What AI applications have you personally interacted with?

Which AI uses feel most helpful? Which raise concerns?

### **Theme 2 – Fairness, trust, and transparency**

Do you think AI decisions can be fair? Why or why not?

How much do you trust the public sector, businesses, and universities to develop and use AI responsibly?

What would increase your trust?

### **Theme 3 – Empowerment vs. control**

Do you feel AI increases people’s control over their lives or reduces it?

How does AI affect human dignity and autonomy?

### **Theme 4 – Cross-border AI**

How do you feel about AI technologies developed abroad but used in Spain?

What do you expect from the implementation of the EU AI Act? Will it improve the situation/alleviate your concerns about AI? Why/why not?

Scenario-based discussion (20-25 min)

### **Scenario A – Academia:**

#### **Scenario A1**

*The University of Barcelona introduces an AI-based grading platform that automatically marks essays in Castilian Spanish, Catalan, and English, and flags potential plagiarism.*

What benefits and risks do you see for students and professors?

How can fairness and accountability be guaranteed if students challenge the AI’s decisions?

#### **Scenario A2**

*The Spanish National Research Council (CSIC) launches a project where AI analyses public hospital data to predict outbreaks of chronic diseases. Data comes from Spain’s national health system and wearable devices.*

How do you feel about researchers in Spain using personal health data for predictive tools?

What safeguards should be in place to protect patient trust while enabling innovation?

### **Scenario B – Business Associations:**

#### **Scenario B1**

*A Madrid-based fintech start-up introduces an AI-driven credit scoring model for SMEs, which includes not only financial history but also online ratings and supplier feedback.*

Would you trust such an AI-driven evaluation for your business? Why or why not?

Should Spain regulate what kinds of non-financial data can be used in credit scoring?

### **Scenario B2**

*Large Spanish supermarket chains (e.g. Mercadona) adopt AI-powered dynamic pricing, adjusting food prices in real time based on demand, time of day, and online shopping behavior.*

How do you feel about supermarkets using AI to change prices this way?

Should Spain set limits on how far AI can alter prices?

### **Scenario C – Youth:**

#### **Scenario C1**

*The Spanish Ministry of Labour and Social Economy pilots an AI-powered career guidance platform for young people. It uses academic records and labour market data to recommend job opportunities. Many students and graduates notice the system frequently suggests short-term contracts or paid internships, while long-term stable positions are rarely highlighted.*

Would you use such a platform to guide your career choices? Why or why not?

What risks might arise if AI systems normalize precarious employment conditions for young people in Spain?

#### **Scenario C2**

*A Spanish social media platform introduces an AI mental-health chatbot for teenagers, offering advice on stress, anxiety, and relationships.*

Would you trust an AI chatbot with sensitive personal issues?

Should Spain establish special rules for AI tools targeting youth well-being?

### **Scenario D – Older adults:**

#### **Scenario D1**

*The Spanish Social Security system introduces an AI-based digital assistant to help retirees manage pensions and healthcare appointments online. While it reduces waiting times for tech-savvy users, many older adults report difficulties using it without personal assistance.*

What opportunities and risks does this create for older adults in Spain?

How can Spain ensure that people with lower digital skills are not excluded from accessing essential services?

#### **Scenario D2**

*Hospitals in Madrid and Málaga pilot AI telecare systems to monitor elderly patients at home (heart rate, mobility sensors, medication reminders). While families appreciate the support, some worry about privacy and the replacement of human contact.*

Would you trust AI systems to monitor older relatives at home? Why or why not?

Should Spanish hospitals guarantee that human doctors or nurses always review AI alerts before action is taken?

## Scenario E – People with disabilities

### Scenario E1

*The City of Barcelona launches an AI-powered mobility app to help persons with disabilities navigate public transport and city streets. It uses real-time data on elevators, ramps, and traffic. However, users report the system sometimes directs them to routes that are blocked or not fully accessible.*

What benefits could this bring for mobility and independence in Spain's urban areas?

What risks might occur if the system gives incomplete or inaccurate information?

### Scenario E2

*A Spanish start-up develops an AI tool for Spanish Sign Language (LSE) to be used in schools and universities, automatically translating between LSE and spoken Spanish. Teachers see it as a breakthrough, but disability organizations raise concerns about accuracy in complex discussions.*

How could this improve access to education for persons with hearing impairments?

What safeguards are needed to ensure quality and reliability in sensitive contexts such as classrooms or exams?

## Scenario F - Migrants

### Scenario F-1

*The Spanish Ministry of Inclusion, Social Security and Migration introduces an AI system to process residence and work permit applications. The system automatically prioritizes cases and flags incomplete files. Migrants' organizations report that the system sometimes rejects applications due to small errors (e.g. missing translations or minor inconsistencies), which delays access to legal status.*

Would you trust an AI system to decide on your migration paperwork? Why or why not?

What safeguards should Spain introduce to ensure that AI does not unfairly disadvantage immigrants in accessing legal rights?

### Scenario F2

*A Spanish job-matching platform uses AI to connect job seekers with available positions. Migrant women report that the system overwhelmingly directs them to caregiving roles, even when they have higher qualifications or different career goals.*

How do you feel about AI systems that seem to reinforce gender stereotypes in the jobs offered to migrant women?

Should Spain set rules to ensure AI employment platforms promote equal opportunities for all job seekers, regardless of gender?

Closing round (5-10 min)  
(Facilitator asks)

If you could give one recommendation to policymakers or AI developers in Spain, what would it be?

Thank participants, explain next steps, and share contact information for follow-up.

#### DATA MANAGEMENT AND ANALYSIS

Transcribe recordings verbatim; translate into English where necessary.

Apply thematic coding based on: fairness, trust, empowerment/control, transparency, dignity, harm minimization, cross-border views.

Compare patterns between academia, business associations, and youth.

#### ANNEX – Consent form

English version

##### **Project context**

This study is being conducted as part of the **Horizon Europe CERTAIN Project**. The project brings together researchers from multiple countries to better understand public perceptions of AI and inform policies for responsible AI governance.

##### **Purpose of the study**

You are invited to take part in a research study about how people in Spain view and experience artificial intelligence (AI) in different areas of life. The study aims to understand opportunities, concerns, and trust in AI, and will help inform policy and practice in Spain, wider EU and internationally.

##### **What participation involves**

If you agree to participate, you will join a group discussion lasting about 1 - 1.5 hours. The discussion will be audio recorded (and video recorded if you agree) so we can accurately capture your comments. Notes will also be taken.

##### **Voluntary participation**

Your participation is voluntary. You may refuse to answer any question and may withdraw at any time without any negative consequences.

##### **Confidentiality**

Your name will not appear in any report. We will use a code or pseudonym instead. While the research team will protect your anonymity, we cannot guarantee that other participants will not share what is said. We ask that you also respect the confidentiality of others in the group.

##### **Risks and benefits**

There are no known physical risks. Some topics may make you feel uncomfortable; you can skip any question or take a break. There is no payment for participation, but your views will contribute to better AI-related policies and practices in Spain and other countries.

##### **Use of data**

The data will be used for research purposes only, stored securely on password-protected devices, and retained for the duration of the project implementation before being destroyed.

##### **Consent statement**

By signing below, you confirm that:

You have read (or listened to) this information and understand it.

You have had the opportunity to ask questions.

You agree to participate in this focus group discussion.

**Audio Recording:**

- I agree to be audio recorded
- I do not agree to be audio recorded

**Video Recording (if applicable):**

- I agree to be video recorded
- I do not agree to be video recorded

**Name:** \_\_\_\_\_

**Signature:** \_\_\_\_\_

**Date:** \_\_\_\_\_

Draft

## APPENDIX C – TECHNOLOGY ACCEPTANCE AND PERCEPTION OF AI: END-USER SURVEY

### Description:

This questionnaire is part of research conducted within the Horizon Europe co-funded project [CERTAIN](#). It aims to explore how people from diverse backgrounds perceive, understand, and experience Artificial Intelligence (AI) technologies in everyday life, work, and society. The questionnaire examines key factors such as perceived usefulness, ease of use, trust, facilitating conditions, and social and gender influences, and how these shape individuals' willingness to use AI technologies. All responses are anonymous and confidential. The data collected will be used solely for research purposes and will be analyzed in aggregated form, ensuring that no individual participant can be identified.

### A) Demographic Information

#### What is your age group?

18–24

25–34

35–44

45–54

55–64

65 or older

#### Which gender do you identify most with?

Male

Female

Non-binary

Prefer not to say

Other (please specify): \_\_\_\_\_

#### Do you self-identify as a person with a disability?

Yes

No

Prefer not to say

#### In which city and country is your workplace located?

[short text]

#### What is your highest completed level of education?

Secondary education (high school or equivalent)

Post-secondary non-tertiary / Vocational education (e.g., technical or vocational school, professional diploma)

Bachelor's degree (or equivalent undergraduate degree)

Master's degree (or equivalent postgraduate degree)

Doctorate (PhD or equivalent) or higher

Prefer not to say

### **What is your field of Employment/Study?**

*(Please select the closest match from the list below. Options are based on ISCO/ESCO classifications.)*

Managers (e.g., business, project, or team managers)

Professionals (e.g., scientists, engineers, educators, healthcare, ICT professionals)

Technicians and Associate Professionals (e.g., lab technicians, technical specialists)

Clerical Support Workers (e.g., administrative assistants, office support)

Service and Sales Workers (e.g., customer service, sales roles)

Skilled Agricultural, Forestry & Fishery Workers

Craft and Related Trades Workers

Plant and Machine Operators and Assemblers

Elementary Occupations (e.g., general labor roles)

B) Awareness and Knowledge of AI Technologies

**(Scale: 1 = Not at all, 10 = Very much)**

How familiar are you with Artificial Intelligence (AI) in general?

How well do you understand how AI systems learn from data (machine learning)?

How aware are you of AI technologies used in everyday life (e.g., chatbots, recommendations, smart assistants)?

How confident are you in explaining what AI does to another person?

Perceived Trust (PT)

*Measure of a person's trust in a particular technology*

**(Scale: 1 = Very low, 10 = Very high)**

How much do you trust AI systems to provide accurate results?

How confident are you when relying on AI to support decisions (e.g., recommendations, suggestions)?

Do you believe AI systems are used responsibly and ethically?

How much do you trust AI to reduce human mistakes in tasks?

How confident are you in the long-term reliability of AI technologies?

Perceived Usefulness (PU)

*Measure of the usefulness of a particular technology*

**(Scale: 1 = Not useful, 10 = Very useful)**

How useful do you think AI is in improving daily tasks or work efficiency?

To what extent do you believe AI helps people make better decisions?

How much do you think AI saves time and effort?

Do you believe AI improves consistency and quality in tasks?

How valuable is AI for innovation and solving complex problems?

Perceived Ease of Use (PEU)

*Measure of the usability of a particular technology*

**(Scale: 1 = Very difficult, 10 = Very easy)**

How easy is it for you to interpret AI-generated outputs in a meaningful way?

How easy are AI tools and applications to use?

How confident do you feel using AI technologies?

How intuitive is interacting with AI systems (e.g., chatbots, assistants, smart tools)?

How likely are you to use AI without needing technical help or external guidance?

Social Influence (SI)

*The degree to which someone is influenced by social norms and their social environment*

(Scale: **1 = Not at all**, **10 = Very much**)

To what extent do people around you encourage the use of AI?

To what extent does seeing others use AI influence your own attitude toward AI?

Does the growing use of AI affect how valuable you think it is?

How important are expert opinions or research in shaping your trust in AI?

How likely would you be to use AI tools if they became a standard or commonly accepted practice in your field of work/study or in society?

Facilitating Conditions (FC)

*The degree to which an individual believes that an organizational and technological infrastructure exists, and also the degree to which they have the appropriate knowledge and resources to use the system*

(Scale: **1 = Very low**, **10 = Very high**)

To what extent do you have access to the devices, internet connectivity, and financial resources (e.g., subscription fees) needed to use AI tools?

How easy is it to find learning resources about AI (tutorials, videos, courses)?

Do you feel you have enough support when learning to use AI tools?

How accessible are AI technologies to you?

Do you feel prepared to adapt to increasing AI use in everyday life?

Gender Influence (GI)

*The extent to which an individual's gender identity, roles, norms, or societal expectations related to gender affect their preferences, behaviors, or interactions with a system or technology.*

(Scale: **1 = Not at all**, **10 = Very much**)

To what extent does gender influence opportunities to participate in the development, design, or governance of AI systems?

To what extent do people of different gender identities have equal opportunities to access AI education and tools?

Have you observed gender-related bias in AI and/or digital technology fields?

To what extent do stereotypes influence participation in AI and digital technology development and deployment?

How important is promoting diversity and inclusion across both the use and development of AI and digital technologies?