FORSEE

Forging Successful AI Applications
for European Economy and Society

D3.2 analysis of Civil Society Organisations' perspectives on AI impact on gender imbalance

Project Information
Project Number: 101177579
Project Title: Forging Successful AI Applications for European Economy and Society - FORSEE
Funding Scheme: HORIZON-CL2-2024-TRANSFORMATIONS-01-06
Project Start Date: 1 February 2025

Deliverable Information	
Title	Analysis of Civil Society Organisations' perspectives on AI impact on gender imbalance
Work Package	WP 3 - Mapping of social expectations: understanding of success from a lifeworld perspective (society and economy)
Lead Beneficiary	NUID UCD
Due Date	31/01/2026
Revision Number:	V2
Authors	Elizabeth Farries, University College Dublin[1] Alexandros Minotakis, University College Dublin[2]

[1] https://orcid.org/0000-0002-6524-2671

[2] https://orcid.org/0009-0001-9294-4564

Funded by
the European Union

	Pierre Ratinaud, Université Toulouse[3] Loredana Bucseneanu, DigitalSME Sandra Sieron, WZB Berlin Social Science Center[4] Julia Pohle, Wissenschaftszentrum Berlin für Sozialforschung[5]
Dissemination Level	PU - Public
Deliverable Type	Report

Revision History				
Version	Changes implemented by	Revision Date	Reviewed by:	Brief description of changes
V1.0	Alexandros Minotakis, Elizabeth Farries, Loredana Bucseneanu	22/01/2026	Pat de Brún, Johannes Mikkonen	Revised according to internal peer-review

Approval Procedure			
Version	Deliverable Name	Approved by	Approval Date
<Fx.x>	<.....>	<Name>	<dd/mm/yyyy>

[3] https://orcid.org/0000-0002-8386-436X
[4] https://orcid.org/0009-0005-0988-1259
[5] https://orcid.org/0000-0002-9442-4626

Forging Successful AI Applications for European Economy and Society

Mapping of social expectations: understanding of success from a lifeworld perspective

The capabilities of artificial intelligence (AI) are advancing rapidly, yet understanding what constitutes successful AI for society and the conditions that enable its effective deployment remains limited. AI promises economic growth, knowledge creation, and broader societal benefits, but realising this potential depends on developing and integrating applications that are successful not only technologically and economically but also socially, and ethically. AI applications are embedded within complex social contexts, reflecting and shaping aspirations, biases, and inequalities; thus, understanding AI success requires attention to these broader dimensions.

The FORSEE project (*Forging Successful AI Applications for European Economy and Society*) adopts a sociological perspective to examine these dynamics, focusing on how different stakeholders define success and how controversies, and unequal distributions of risks and benefits are articulated and potentially resolved.

This cluster of research papers maps social expectations of AI success across stakeholders, adopting a lifeworld perspective that situates understandings within societal and economic contexts. It comprises four interrelated reports. The first examines **digital small and medium-sized enterprises' (SMEs) success narratives**, identifying recurring themes and operational challenges. The second addresses **civil society organisations (CSOs) perspectives on gendered risks related to AI**, examining potential paths to advocate for gender vulnerable communities. The third investigates **criteria for awards and prizes**, providing an external perspective on standards of AI success. The fourth applies a **gendered lens**, exploring SMEs’ perspectives on AI and gender bias. Together, these reports link SMEs’ and CSOs’ viewpoints with societal concerns, offering a multidimensional understanding of AI success in Europe.

The present report focuses on how civil society organisations perceive artificial intelligence broadly, with particular attention to issues of gender bias in AI systems, and analyses the strategies, operational practices, and collaborative arrangements CSOs employ to counter such bias and support gendered communities, while also highlighting the systemic obstacles that limit the effectiveness of these initiatives.

Table of Contents

Executive Summary

The present study examines the perspectives of civil society organisations (CSOs) on artificial intelligence (AI) in general and on AI gender bias in particular. It further explores existing CSO strategies, practices and partnerships aimed at addressing AI gender bias and empowering gendered communities, while identifying the structural barriers that constrain these efforts.

Within this framework, CSOs are conceptualised as key instruments of democratic accountability, with the potential to contribute meaningfully to AI governance in accordance with EU values and principles. At the same time, CSOs occupy a privileged position that enables direct engagement with gender marginalised communities and the articulation of their concerns regarding instances of algorithmic discrimination.

To address the study's research questions, a mixed-methods approach was adopted, combining qualitative data from participatory workshops with quantitative data collected through an online survey. The target sample consisted of CSO workers and activists, who were invited to reflect on AI-related risks and benefits, engage with recent European regulatory initiatives such as the EU AI Act, and describe organisational activities undertaken to address AI gender bias.

Key findings of present research indicate that CSOs demonstrate high awareness of AI-related risks, including AI gender bias, but their organisations face significant constraints that limit their capacity to act. AI is frequently associated with concerns regarding the erosion of democratic processes, intensified surveillance, and a lack of accountability on the part of technology companies. In particular, AI gender bias is widely recognised but rarely prioritised organisationally due to resource scarcity, shrinking civic space, diminished presence in the online public sphere and exclusion from regulatory deliberation and consultation processes.

Furthermore, algorithmic platform governance practices, including content moderation and reduced visibility of civil society content, further undermine CSO engagement with AI governance. Finally, participants expressed low confidence in the EU's current capacity to effectively address these challenges.

Based on these findings, the study highlights the need for stronger and more inclusive AI governance frameworks at both EU and national levels. Policy measures should prioritise the meaningful inclusion of CSOs in regulatory consultations, provide targeted funding and capacity-building support for CSO engagement with AI, and address platform governance practices that limit civil society visibility and participation. Without such interventions, existing gendered inequalities risk being reproduced and amplified through AI systems, while civil society's role in democratic oversight of AI remains structurally weakened.

Section 1: Introduction

> In Europe and around the world, AI systems are used to monitor and control us in public spaces, predict our likelihood of future criminality, facilitate violations of the right to claim asylum, predict our emotions and categorise us, and to make crucial decisions that determine our access to public services, welfare, education, and employment.
>
> Without strong regulation, companies and governments will continue to use AI systems that exacerbate mass surveillance, structural discrimination, centralised power of large technology companies, unaccountable public decision-making, and environmental damage
>
> - 148 Civil Society Organisations, (Access Now, 2023)

The policy world watched as the landmark regulation EU Artificial Intelligence (AI) Act was ratified in 2024. We continue to watch as its enforcement is negotiated. EU Civil Society Organisations (CSOs), at supranational and national levels, have an important and recognised role in this AI governance space. FORSEE's approach understands AI not as a fixed technological object but as a shifting socio-technical field that is continuously contested. More concretely, this research starts with an assumption derived from existing scholarship that controversies surrounding AI coalesce around both recurring tensions as well as new types of controversies. The goal of our study here is to better understand EU CSOs' awareness and perspectives on AI gender bias, a concept informed by researcher Timnit Gebru (2020) who has famously articulated bias as not just a technical error, but a structural and systematic issue where AI systems reinforce existing societal prejudices, including bias against gender. Connected with AI gender bias is discrimination, where bias acts as a cause or mechanism and discrimination becomes the real world consequence of differential treatment along categories of identity including gender. For example, an AI hiring tool gender bias that preferentially selects men, leads to a discriminatory consequence for applicants who are women.

Despite their accepted role as 'instruments of accountability' in governance, the perspectives of EU CSOs on AI success are not clearly mapped. FORSEE's present research addresses this gap by analysing how EU CSOs themselves understand AI, its benefits and risks from a lifeworld perspective, with particular attention to AI gender bias. In doing so, this research does not assume that all CSOs should or can engage with AI gender bias, and indeed acknowledges the diversity among CSOs and their mandates thereby precluding the drawing of broad conclusions across this cohort. However, understanding how CSOs conceptualise AI benefits and risks, particularly in relation to AI gender bias, offers crucial insight for shaping inclusive and democratically accountable AI systems.

This report is structured as follows. **Section 2** introduces the theoretical framework on AI and bias with a particular focus on gender bias and the different ways it manifests across the AI lifecycle. **Section 3** engages with CSOs perspectives, their potential contribution to AI governance and the limitations to their capacity. **Section 4** outlines the methodological approach, describing how both quantitative (online survey) and qualitative (workshops) methods were used to address the research questions. **Section 5** presents the empirical findings, detailing survey results and workshop transcripts. **Section 6** discusses the findings in relation to CSOs perspectives and practical experiences in addressing AI and gender bias as well as broader limitations that AI imposes on their organisations. Furthermore, it presents limitations of present research as well as pathways for potential future research. Finally, **Section 7** concludes the present report, reiterates key findings and reflects on their broader implications for AI governance and CSOs.

Section 2: AI, Bias, and AI Gender Bias

2.1 AI and lifecycle biases

Scholars have long established that the way contemporary AI systems operate poses significant risks through biased deployments that enable discrimination, while AI outputs are often cloaked as objective, impartial or neutral data-driven systems (Xavier, 2025). The concept of bias is not new; it precedes AI development and reflects the tendency to "promote prejudiced results due to erroneous assumptions" (Mavrogiorgos et al., 2024, p.1). However, as AI systems become integrated into more and more social spheres and decision-making processes, the issue of bias is widely discussed and is a significant concern. AI is trained, validated and tested on large datasets - the raw information or material from which AI systems learn to recognise patterns, make predictions, and perform tasks. Consequential to this training is the ability of AI to detect connections and patterns in the datasets that are not explicitly identified by the people who input or label the data. However, dataset bias arises through systematic errors or prejudices that skew the data, leading AI to develop flawed interpretations of the patterns that emerge (see for example, Yang et al, 2024; Ferrara, 2024). This can lead to poor performance outside of narrowly constructed norms.

Crenshaw (1989) has written seminally of intersecting discriminations between race and gender and, indeed, recent AI studies have shown that poor AI performances can be amplified in particular at intersections of marginalised identities rather than being reducible to a single demographic axis (Gohar & Cheng, 2023). In social realms, AI is trained on large datasets and can identify patterns across a range of demographic characteristics: gender, sexuality, race, age, income, residence, and more. In many cases, AI performance declines in intersectional scenarios targeting simultaneously gender and other demographic characteristics (see Noble, 2018) leading to intersectional bias. For example, LLM-based resume scoring systems have been shown to systematically favor white-associated names while disadvantaging black male candidates, even when all applicants had identical qualifications (Wilson & Caliskan, 2024).

AI gender bias is not a function of flawed datasets alone and so cannot be solved simply with more diverse or larger datasets. The systemic literature reveals how categories of bias can emerge at any stage of the "AI lifecycle" (Chen et al., 2023), which includes design: dataset collection and selection, model design, development and validation (Yilmaz, Yorgancioglu & Koyutürk, 2025); however, bias also occurs in deployment and monitoring and through opportunities for redress. Bias can be introduced at every stage of this cycle and can interact cumulatively to affect final outcomes. Bias surfacing via these lifecycle stages can overrepresent, underrepresent, and misrepresent findings or influence decisions. While most frameworks evaluate bias in models rather than datasets, it is important that datasets are also a focus of importance because model bias can often be traced back to dataset shortcomings (Dai et al., 2025). Similarly, it is important to evaluate how AI is deployed in real world settings, how it impacts diverse populations, how it is monitored and restrained, and what opportunities for redress and change exist.

2.2 Systemic discriminations as a function of AI bias

Established and critical findings for bias and discrimination in AI applications have been summarised by Ferrara (2024) in key public policy areas including healthcare, employment and hiring practices, criminal justice, migration, warfare, credit scoring, and Generative AI models. Prominent examples in public policy discourse increasingly reflect these risks. For example, social welfare AIs are promoted by governments as tools of efficiency. However, civil society has often criticised these systems for relying on flawed datasets within underfunded and inadequate social protection systems, ultimately resulting in harmful automation of benefit denial (Banerji & Satija, 2025). A widely cited private-sector case is Amazon's recruitment AI, which through its selected datasets trained on male-dominated work forces, built an AI model that when deployed prioritised resumes reflecting historically male hiring patterns, systematically excluding women (Buolamwini, 2019; Dastin, 2018).

Furthermore, while sometimes unintentional, biased deployment can reinforce biased policy. In Hungary, biometric surveillance technologies such as facial recognition are now deployed to monitor peaceful LGBTQ+ demonstrations, including Budapest Pride (Hungarian Civil Liberties Union et al., 2025). In such scenarios, discriminatory results can be recorded as vindication of new AI technologies: if social welfare allowances are reduced, if a pool of applicants are surfaced, if a policing system identifies LGBTQ+ protest content for sanction, then this can be characterised as an AI "success" despite the bias and discrimination propelling systemic exclusion. Eubanks (2018, p. 7) describes this phenomenon as a "feedback loop of injustice," in which AI systems not only reproduce but reinforce structural oppression under the guise of neutrality or precision.

2.3 The limits of techno-solutionist responses to bias

In line with techno-solutionist narratives, complex social problems are often reframed as technological and addressed through technical fixes (Lindgren & Dignum, 2023). Dataset bias mitigation strategies supported by published research include interventions in the preprocessing, processing, post processing stages. Preprocessing solutions include data training set balancing (costly) and augmentation (synthetic) (Yang et al, 2024). Processing techniques have included causal model-based dataset creation (González-Sendino et al., 2024) or multi-attribute bias mitigation (Duong & Conrad, 2024). Post processing techniques include the validation of dataset representativeness and performance through audits, which are prescribed regularly and for model updates (Chadha, 2024). These purely technical measures however tend to omit what Birhane (2021) and Noble (2018) have emphasised: social and structural conditions that create bias are not merely an error but rather a function of social and cultural processes that amplify structural inequalities. Such sociotechnical analysis, i.e. considering the interdependent relationship between social elements (people, culture, rules, values) and technical systems (tools, processes, technology), surfaces the reality that by focusing on technical bias metrics alone, larger facets of exclusionary design are missed, including the power, context, and relational harms that AI performs in everyday life and the way that systems harm social groups through exclusionary and unjust outcomes.

The point of inflection in this debate is how intersectionally biased results in the AI lifecycle discriminate against marginalised populations. Bias does not result from targeting of individuals although, expectedly, they experience discrimination in a personal manner. Rather, people "are targeted as members of social groups, not as individuals" (Eubanks, 2018, p.6). Therefore, as preeminent scholars have pointed out (see Benjamin, 2019; Noble, 2018; Buolamwini & Gebru, 2018; D'Ignazio & Klein, 2020; and Birhane, 2021), bias in AI can lead to social harms for members of social groups according to categories including gender, but also sexuality, and race.

2.4 Sociotechnical drivers for bias: Stakeholder roles

The explanation for biased and discriminatory AI lies not just in technical explanations but at the nexus of technical and sociotechnical issues. Here we categorise biases within the AI lifecycle including dataset, design and development choices, together with deployment and monitoring decisions. Such choices carry assumptions about which data points matter, and which do not. These assumptions can embed cultural norms and privilege certain outputs or conceptualisations over others. For example, a 2025 peer-reviewed study found that popular AI text-to-image models generate from their selected datasets leadership images that significantly overrepresent men and white individuals relative to real-world demographics in U.S. hospitals (Gisselbaek et al, 2025). Despite actual leadership distributions showing substantial female and minority representation, the AI outputs skewed heavily toward men and white leaders, reflecting design assumptions embedded in the dataset facilitating a model in which leadership roles are predominantly male and white. Evidently, defining what is normal and what is an outlier relates to design decisions and reflects developers' focus and omissions.

Discrimination also results in intersectional ways when stakeholders enforce biased AI systems outputs. For example, Wilson et al. (2025) showed that when humans acted on biased AI recommendations in hiring simulations, their selection patterns mirrored the AI's intersectional biases in discriminatory ways, even when the participants were aware of potential AI bias. Conversely, when they made screening decisions without AI guidance, choices were relatively balanced across groups. Such findings support classic research on automation bias, showing how people often over-rely on algorithmic recommendations (Mosier & Skitka, 1996; Parasuraman & Riley, 1997). Automation bias belies the understanding that the integration of human oversight, judgment, or intervention at key stages of an AI system's lifecycle is sufficient to detect, prevent, and mitigate bias that the system might otherwise perpetuate (the "human-in-the-loop" solution).

Indeed, human oversight may actively shape systems in ways that amplify biased outcomes in pursuit of specific objectives. A 2025 investigation into the UK's national facial recognition deployment found that police forces adjusted the system's confidence thresholds to reintroduce a more biased configuration, as it produced more "useful" suspects, despite higher misidentification rates for Black and Asian people and for women compared with white men (Wilding & Boffey, 2025). This deliberate adjustment of algorithmic thresholds to reinforce, rather than mitigate, bias reflects organisational priorities that override ethical considerations and underscores that AI bias is not a purely technical failure.

AI systems are thus shaped and driven by sociotechnical choices that often reflect narrow assumptions about what (and who) matters. AI lifecycle decisions that are made without diverse perspectives can embed cultural norms and disproportionately privilege dominant identities to the exclusion of other demographics. These exclusions are acute at the intersections of gender, race and other social identifiers (Noble, 2018). Bias does not stop at deployment or monitoring phases; rather, it is compounded when human actors uncritically enforce AI bias, thereby reinforcing intersectional discrimination (Wilson et al., 2025; Wilding & Boffey, 2025).

Section 3: CSO Perspectives on AI

3.1 CSOs as instruments of democratic accountability

CSOs are recognised as essential to democratic EU governance, serving as the voices of public interest, defenders of fundamental rights and Charter values, and key instruments of democratic accountability to both states and technology companies. EU institutions affirm the centrality of CSOs in upholding Union values, with the European Commission's Recommendation (2023/2836 at para 4) urging Member States to ensure a safe and enabling environment for CSOs. Scholars similarly underscore the normative intermediary function CSOs play in bridging citizens and EU institutions, and enhancing the legitimacy of multilevel governance by injecting expertise and representing marginalised perspectives in policy processes (Smismans, 2003; Greenwood, 2007). Regarding the perspectives of marginalised communities, CSOs are recognised to play a role in identifying and responding to gender and LGBTQ+ marginalisation by translating lived experiences of discrimination into advocacy, monitoring, and rights-based interventions that inform EU law and policy (European Union Agency for Fundamental Rights, 2020). Normative perceptions of CSOs' role in the EU's policy architecture is intended as not merely consultative but structural, thereby ensuring that democratic principles are continuously negotiated and defended.

3.2 CSOs' engagement with AI

Policy responses to new technologies including AI require the specialised perspectives of EU CSOs who engage AI's implications for democracy, fundamental rights, and accountability within the EU. An indicative example is the open letter signed by a coalition of 31 civil society organisations in September 2025, urging the European Commission and EU member states to speed up the national implementation of the AI Act (EDRi, 2025). This was preceded by a letter signed by 119 CSOs (EDRi, 2021) that highlighted how AI can contribute to high-risk domains such as policing and border control. Other CSO coalitions have emerged in EU member states, see for example in Ireland the open letter in relation to policing facial recognition technologies (UCD Centre for Digital Policy, 2023) or in Hungary the open letter in relation to AI surveillance of marginalised LGBTQ+ populations (Hungarian Civil Liberties Union et al., 2025). CSOs applying such sociotechnical analysis consider the interdependent relationship between social elements (people, culture, rules, values) and technical systems (tools, processes, technology).

Overall, CSOs play a counterbalancing role in relation to industry lobbyists. Big Tech companies devote significant resources trying to influence lawmaking and regulation across Europe. CSOs often occupy the same spaces, seeking to defend against this profit-focused influence and promote the protection of fundamental rights. They assist layered understanding of AI success beyond technical efficiencies in their recognition of intersectional biases against marginalised populations, including AI gender bias, and systemic effects, including discrimination, together with impacts on democratic values and fundamental rights (See 3.4 for elaboration of sociotechnical impacts of AI). EU CSOs therefore play a vital role by interpreting and applying EU norms and principles to AI while shaping debates around regulation.

3.3 Fragmented capacity concerns

However, not all EU CSOs engage with AI impacts or possess expertise in AI. To date CSOs have reported feeling marginalised in EU AI regulatory discussions (WISERD, 2024). While many CSOs possess deep expertise in rights-based advocacy, their familiarity with AI and new technologies systems varies significantly, often depending on personal perspectives and/or their organisation's mandates and thematic priorities. As with other governance stakeholders, these perspectives can be partially limited by rapid tech innovation given that technical language can restrict understanding to a limited group of privileged insiders and experts.' (Murphy, 2018). We can see that understanding difficulty elaborated at the level of EU jurisprudence built around CSO interventions. In the notable 2015 case of *Zakharov v Russia* (2015), the European Court of Human Rights observed that core rights principles attached to new technologies experiencing rapid advancement can remain opaque and abstract.

Many organisations may therefore not address AI, or if they do, they approach AI through the lens of their existing advocacy priorities. Specific organisations may lack in-house technical capacity or specialised knowledge here, instead bring thematic specialisations, such as gender, feminism, or children's rights to AI policy discussions (Siapera & Farries, 2025). Such orientations may result in advocacy that can be contradicting their organisational mandates. For example, children's rights organisations can promote surveillance-based AI to protect children online; while well-intentioned, such measures often involve invasive data collection on children and monitoring of their activity (see discussion in Siapera & Farries, 2025). Therefore, special interest CSOs can operate within a context of conflicting goals and priorities while simultaneously possessing limited capacity, which may constrain their influence in regulatory spaces dominated by technical language and industry lobbying resources.

3.4 Further data required

Given the recognised role of CSOs in shaping democratic EU governance, it is important to gather more comprehensive data on EU CSOs perspectives of AI, both in general terms and specifically in relation to AI gender bias. Our research seeks to better understand CSOs perspectives on AI. We asked, how do CSOs perceive AI risks and benefits, how AI implicates gendered and LGBTQ+ marginalisations, and their understanding of the evolving EU regulatory landscape. Such data also enabled our identification of organisational actions, such as campaigns, policy submissions, or internal training, aimed at addressing bias and discrimination in AI systems. Without this insight, democratic EU policymaking risks overlooking both the challenges CSOs face in navigating AI debates and the valuable contextual knowledge they contribute from the perspective of marginalised communities (Ada Lovelace Institute, 2023; European Union Agency for Fundamental Rights, 2020).

Section 4: Methodology

We have argued that it is important to gather more comprehensive data on EU CSOs perspectives on AI in general terms and specifically in relation to gendered and other marginalisations. This area of study is important given CSOs structural role in holding governments and industry accountable in relation to new technologies. The research question of present study is formulated as follows: *How do civil society organisations (CSOs) perceive AI in general and AI gender bias in particular, what strategies have they developed to address it, and what limitations constrain their efforts?* This research question is organised into the following themes presented in the findings section:

a) What are the main risks and benefits that CSOs associate with AI?
b) What is their engagement with AI systems and how do they report their own understanding of it?
c) How do CSOs perceive EU initiatives to regulate AI systems?
d) How important is AI gender bias within the setting of CSOs?

e) What strategies, initiatives and partnerships have CSOs deployed to raise awareness on and address AI gender bias?

To address our research question, a dual approach was developed in order to collect both quantitative and qualitative data, through the dissemination of an online CSO survey and the organisation of three CSO workshops, respectively. The goal of the survey was to provide a general overview of CSOs perspectives on AI and AI gender bias, while the workshops would capture the breadth and nuances of their lived experiences, advocacy priorities, and institutional capacities. This approach would allow our research to explore how personal perspectives, organisational missions, and thematic expertise (e.g., in gender or LGBTQ+ rights) intersect and shape understandings of AI risks and benefits.

4.1 Design and dissemination of survey

FORSEE designed and executed a survey among workers and activists of EU CSOs. Online anonymous surveys are helpful for reaching populations across wide geographic areas and where participants are diverse and topics are sensitive (Wright, 2005; Bhattacherjee, 2012), including CSOs populations across EU member states addressing AI. Bhattacherjee (2012) describes how online surveys support a variety of question types and sequencing logics, which here allowed nuanced questions within a CSO context. The survey presented multiple choice, ranking, and open-ended questions asking participants to complete 26 questions which took approximately 10 minutes. The survey was structured around 7 distinct sections:

1) Personal perspectives on AI
2) Familiarity with AI systems through personal and professional use
3) Assessment of risks and bias in AI systems
4) Awareness of AI and gendered and LGBTQ+ marginalisations
5) Awareness of AI regulatory landscape
6) CSOs actions and practices to address gendered and/or LGBTQ+ marginalisations
7) Participant demographics

To deliver this survey, we referred to Samarin et al. 's (2020) articulation of the standardised questionnaire design guidelines that are specific to CSOs advocacy considerations. These included minimising survey response time in recognition of CSOs time capacity constraints, protecting the confidentiality of responses given the sensitivity of CSOs work, establishing trust towards managing recruitment challenges. To establish trust towards effective recruitment, we leveraged our Project Coordinator's UCD Centre for Digital Policy's established partnerships. Advocates from these institutes together with our FORSEE partners, distributed the survey broadly to CSOs across Europe.

In organising the survey, FORSEE received ethics clearance from UCD to disseminate an anonymous online questionnaire under the provision that participant anonymity and safe data storage will be guaranteed. The research team contacted organisations and CSO networks, explained the nature and goal of the survey in advance and then they distributed the link across their personnel and affiliated networks. The platform where the survey was hosted collected no data on the participants. Therefore, anonymity was ensured.

A total of 91 fully completed survey responses were collected. The relatively limited sample size is addressed and discussed in the study's limitations section.

4.2 Preparing and organising the workshops

Workshops provide a flexible, relational research practice well suited to AI as a complex, socially embedded issue (Eidenskog, Andersson, and Glad, 2024). They can allow CSOs and researchers to co-construct knowledge that blends academic research with grassroots understanding. In the case of present study, while the survey provided a broad overview, the workshops enabled a deeper exploration of the structural dimensions shaping CSOs' engagement with AI. In particular, regarding organisational capacities, constraints, and modes of participation could not be captured through the survey alone.

FORSEE organised three two hour, in person workshops in Berlin, Germany (October 14, 2025); Budapest, Hungary (October 28, 2025), and Dublin, Ireland (November 20, 2025). We sought to cohost these workshops with our partners: WZB in Berlin, Irish Council for Civil Liberties (ICCL) in Ireland, and the Hungarian Civil Liberties Union (HCLU) in Hungary[6]. Attendance included n = 17 Berlin, n = 10 Hungary, n = 10 Dublin. Attendees were workers in CSOs that primarily focused on gender, technology or LGBTQ+ issues. The majority identified as female, with smaller proportions of male, non-binary, and undisclosed respondents. We ran the workshops in two hours in-person formats using inclusive and participatory principles, seeking to ensure all voices have space. One to three neutral facilitators were present in each session, with note-takers and tone setting for respectful and open engagement. We deployed brainstorming sessions and small group breakout sessions adaptively and reflexively to adapt surface concepts and issues, allowing the structure to shift depending on the dynamics in the room.

The objective was to share research findings and delve deeper into the impact of AI applications on gender marginalisation. Workshops proceeded through a brief presentation of the ongoing survey result combined with case studies pertinent for each national environment, encouraging participants to bring in their lived experiences, tacit knowledge, and contextual insight to AI impacts and governance issues. Following in broad strokes the themes of the survey, workshop participants were asked how they define or understand AI, querying what concerns or hopes shape their advocacy, research or knowledge around AI.

We then presented participants with case studies from participating CSOs jurisdictions including:

- In Berlin, the social media campaign "Too queer for the algorithm? Of clicks and classes - how the digital market discriminates" (Superrrlab)", which reports that

[6] The CSOs involved (ICCL and HCLU) assisted with recruiting survey and workshop participants through their affiliated networks. No organisation received funding for its involvement in the project.

algorithms used in recommendation systems, automated decisions, and digital services can reproduce or amplify discrimination (@superrrnetwork, 2025)

- In Budapest, the CSOs open letter (Hungarian Civil Liberties Union, et al., 2025) in relation to Hungary's outlawing of LGBTQ+ peaceful assemblies and government's use of facial recognition technology to surveil protestors.
- In Dublin, the CSO's campaign around policing facial recognition technology was used as a case study.

We asked participants about their understanding and concerns in relation to various types of AI systems and bias. We queried how confident they feel about the regulatory frameworks. In discussing their understanding, participants were presented with case studies of AI bias that is discriminatory and asked to identify:

- Which are the main responsible actors for governing AI systems (users, providers, mediators-platforms)?
- How should CSOs and academic institutions engage with them?
- Where do you see CSOs and researchers engaging effectively in (gender/sexuality/racialised) settings?
- What barriers do CSOs and researchers face in responding to harmful AI systems? What strategies or partnerships have worked or could work?

FORSEE circulated a briefing document and consent forms to participants. All events were audio-recorded and transcribed, and transcripts were anonymised prior to analysis, removing any personal information or concrete reference to organisations. Data were analysed using NVivo 15 qualitative analysis software. The coding process followed a deductive content analysis approach, guided by a set of predefined analytical categories derived from the survey's conceptual framework (e.g. perceptions of AI risks and benefits, perspectives on AI gender bias, regulation and organisational priorities). These categories were used to ensure comparability between survey and workshop findings. During coding, transcripts were first coded against these predefined categories. During transcript analysis, additional sub-codes were developed to capture themes that emerged from the discussions but were not fully anticipated by the survey framework; a prominent example of this (as discussed in the following sections) are cases where algorithmic content moderation limits CSOs' online activity. This iterative process allowed the analysis to remain grounded in the empirical material and simultaneously flexible enough to adapt to emerging themes.

Section 5: Findings

5.1 Survey results

Overall, survey results indicate that CSOs identify significantly more risks than benefits regarding AI. At the same time, they use AI systems frequently and feel confident in their ability to personally identify related risks and biases. They express great concern about AI impact on gender equality and consider the EU regulatory framework to be largely inadequate. With regards to their own organisations, they do not consider AI gender bias as a priority and report low levels of activity and advocacy on the subject matter. In the following subsections, results will be comprehensively presented, following the order of sections in the survey.

5.1.1 CSOs perspectives on AI

The first section of the survey employs a three-step free-association exercise to get insights into respondents' immediate cognitive and emotional reactions toward AI. Participants were invited to provide up to five words or expressions that spontaneously came to mind when hearing the term "AI", to rank these associations based on their perception, and to assign each a positive, negative, highly positive, highly negative, or neutral evaluation (5-point scale for sentiment evaluation for each word). As free associations produce diverse and unexpected responses, the next step involves relating the responses and their associations to pertinent terms and integrating them into broader conceptual clusters.

The resulting eight categories emerged as heuristic constructs to identify and group recurring patterns.

Politics and governance:

- Expressions relating to unregulated power, danger, lack of accountability, and governance gaps were prominent among the respondents. This cluster reframes AI not merely as an "ethical" dilemma, but as a political project marked by power asymmetries (e.g., *"power asymmetry," "wild west," "unregulated"*). Crucially, this category also captures concerns regarding epistemic integrity and the degradation of truth, with frequent references to *"disinformation," "deepfakes," "lies,"* and *"hallucinations."*

AI bias and structural discriminations:

- Concerns regarding unfairness, inequality, stereotyping, and racism formed another distinct and structurally embedded cluster. Unlike general public surveys that may view bias as a "glitch," we find these associations to point to an awareness of AI's potential to automate and entrench existing systems of oppression, including sexism and racism (e.g., *"past discrimination propelled," "male-oriented"*)

Environmental and material impact:

- We also clustered responses around another important emerging theme, focusing on the material footprint of AI. Respondents explicitly cited *"energy consumption," "water waste," "ecocide,"* and *"fossil fuel guzzler,"* highlighting the tangible environmental consequences of AI which is becoming an increasingly critical topic.

Surveillance and privacy:

- We also noted how respondents used a lot of terms relating to monitoring, tracking, profiling, and loss of control, all of which suggest a view of AI as a technology that may be used as a means for pervasive surveillance, which remains a crucial point for many CSOs.

Data and concentration:

- Many respondents referred to data and Big Tech firms (e.g., *"Big Tech," "Meta," "Google," "Grift"*), foregrounding a perception of AI as inseparable from industry concentration and the importance of data in the AI and platform economy (e.g., *"widening wealth gap")*

Automation and efficiency:

- Respondents frequently invoked speed, efficiency, process automation, and productivity gains, which presented the positive aspects, or potential, of AI.

Labour impact:

- We also noted some respondents associating AI with issues such as automation of labour and job displacement (e.g., *"job loss"*).

Innovation and technological progress:

- Although less frequent than critical associations, some respondents referenced innovation, progress, or broadly the future, pointing to a wider optimistic view toward AI.

Sentiment evaluation:

First, sentiment should not be read as a stable dimension; rather, it reflects the emotional value respondents attach to the specific associations they selected. For instance, a negative sentiment marking for "risk" may express criticism directed at the *lack of regulation* rather than the technology itself, which requires a careful reading of the responses.

Nonetheless, the sentiment evaluations assigned by respondents suggest a predominantly critical orientation toward AI. The clusters associated with political risk, discrimination, surveillance, and environmental impact carry clear and intense negative sentiment ("--" or "-" in the survey). We find these evaluations to signal some apprehension regarding the negative implications of AI, including but not limited to bias and structural discriminations, risk to democratic governance processes and environmental harms. By contrast, positive sentiment is largely concentrated around the automation and efficiency cluster, reflecting a perception of AI's instrumental utility in improving processes. The data and concentration cluster shows some ambivalence as some respondents view AI as enabling, while others mark it negatively. Finally, we combined respondents' ranking of associations with their sentiment markings and ended up with four levels of evaluation:

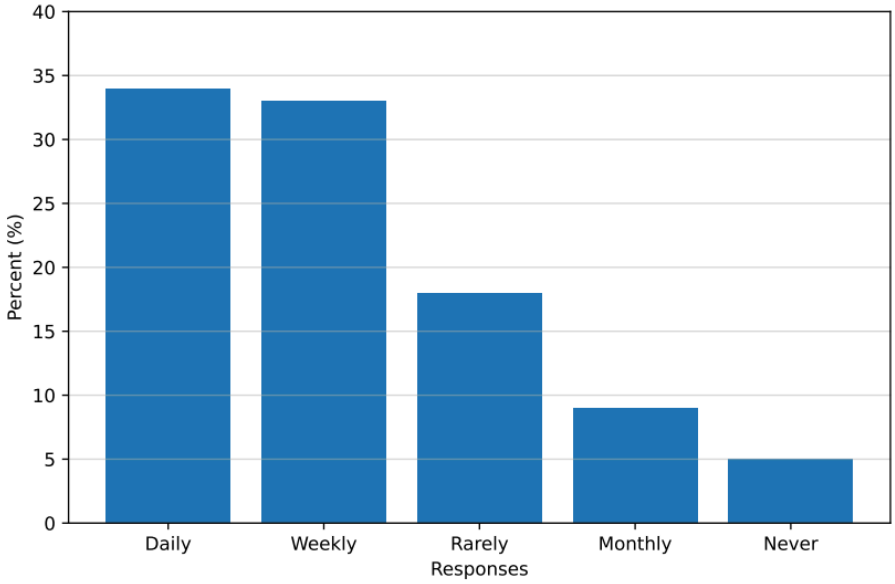
- High importance with strong negative sentiment: *Politics and governance*; *Bias and structural discrimination*; *Surveillance and privacy*; *Environmental and material impact*.
- High importance with positive sentiment: a*utomation and efficiency*.
- High importance with mixed sentiment: d*ata and concentration*.
- Medium importance with mixed sentiment: *innovation and progress* (positive-leaning); *labour impact* (negative-leaning).

5.1.2 Familiarity with AI systems through personal and professional use

This section inquires into how often respondents use AI for personal or professional reasons, as well as how confident they feel in the output itself as well as in understanding how said output is produced. These findings are included in tables 1 through 5.

Table 1

How often do you engage with AI systems?

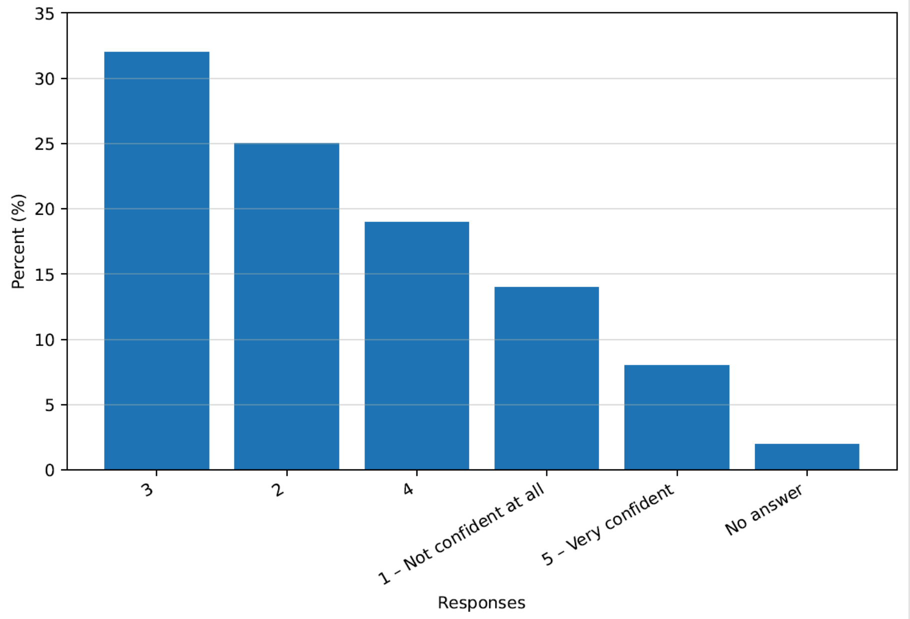


As indicated on Table 1, CSO workers and activists use AI systems quite often, either daily or weekly.

Table 2

How confident do you feel in the output of AI systems for personal use?

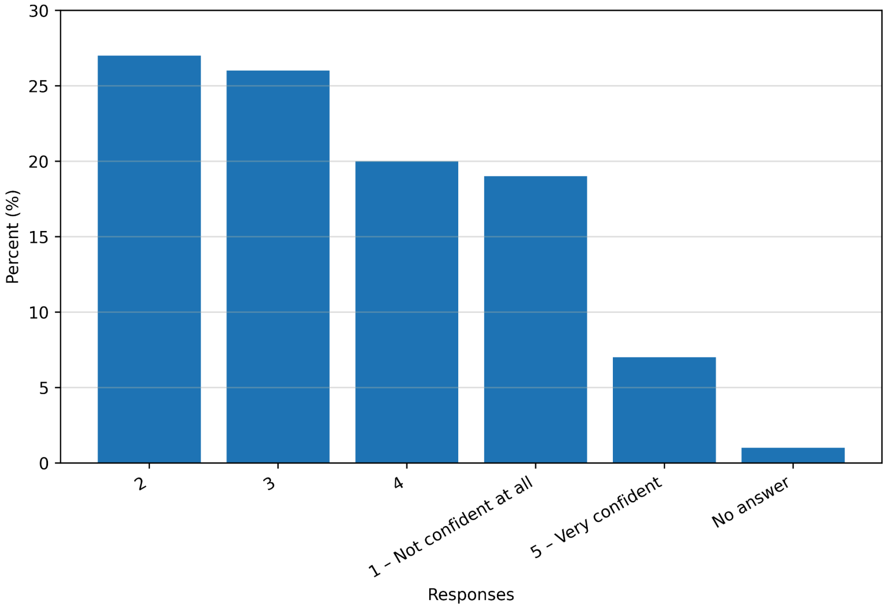
Funded by the European Union



As shown in Table 2, responses concentrate mainly in the middle categories, indicating overall moderate confidence in AI output for personal use.

Table 3

How confident do you feel in the output of AI systems for professional use? (1 = Not confident at all, 5 = Very confident).

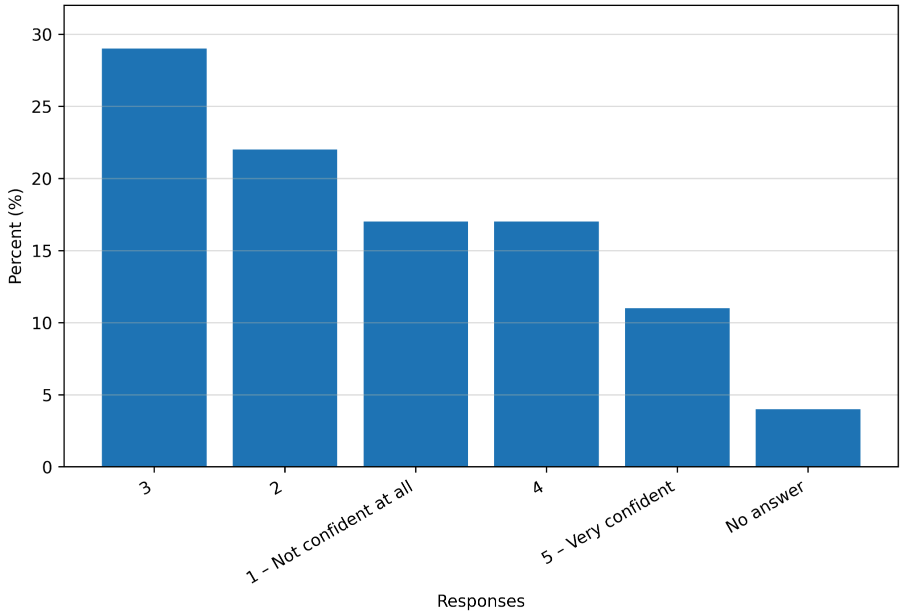


Funded by the European Union

As indicated in Table 3, CSOs workers and activists report mainly moderate levels of confidence in the output of AI systems for professional use.

Table 4

How confident do you feel in understanding how AI systems produce their outputs for your personal use? (1 = Not confident at all, 5 = Very confident).

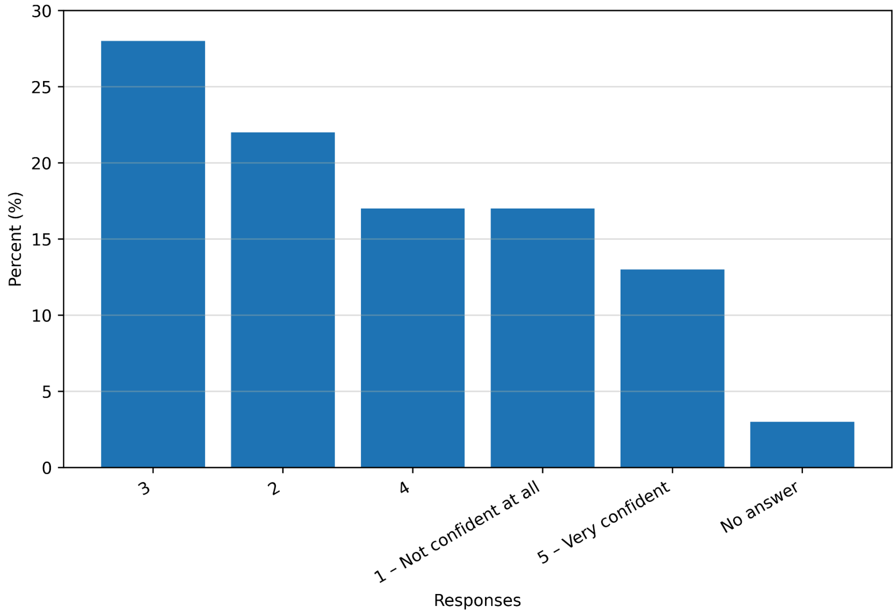


As shown in Table 4, CSOs workers and activists report mainly moderate levels of confidence in understanding how AI systems produce their outputs for personal use.

Table 5

How confident do you feel in understanding how AI systems produce their outputs for professional use? (1 = Not confident at all, 5 = Very confident).

Funded by the European Union



Similarly, CSOs workers and activists report in Table 5 moderate levels of confidence in understanding how AI systems produce their outputs for professional use.

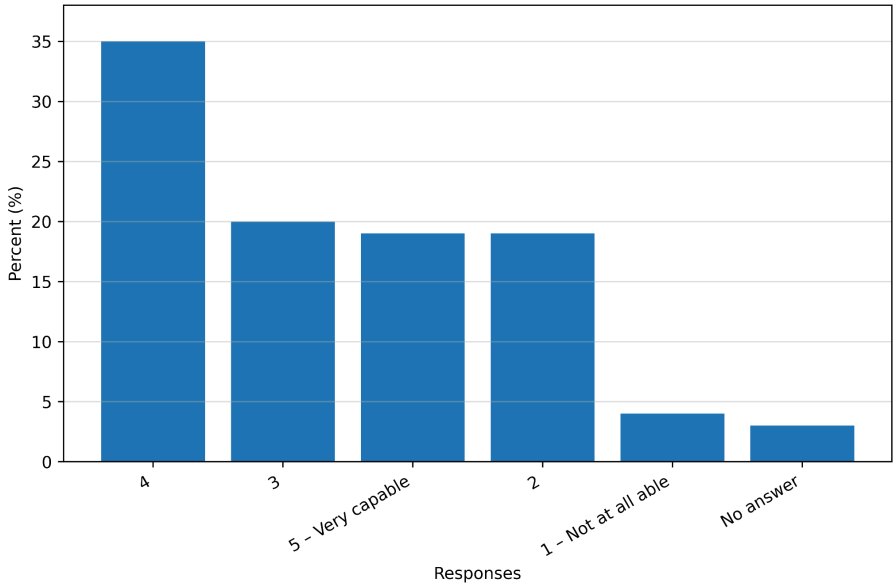
Overall, and despite the predominance of risks and negative associations highlighted in Section 1, CSO workers and affiliated activists report frequent use of AI. They express moderate confidence in its outputs, combined with moderate to low confidence in understanding how these outputs are generated.

5.1.3 Assessment of Risks and Bias in AI systems

In section 3 of the survey, respondents were asked to rate their own ability to assess risks and biases for the AI systems they use either for personal or professional activities. Findings are included in tables 6 through 9. Within this context, risks refer to privacy, security, misinformation, job displacement, ethical concerns, or unintended consequences of AI outputs.

Table 6

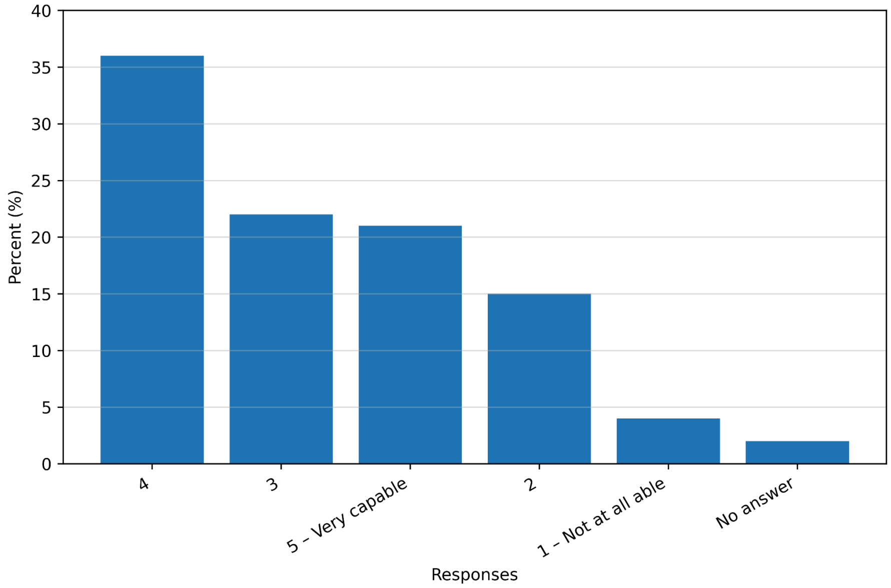
Rate your ability to critically assess the potential risks of the AI tools you use, for personal activities. (1 = Not at all able, 5 = Very capable).



As indicated in Table 6, CSO workers and activists report moderate to relatively high self-assessed ability to critically assess the potential risks of the AI tools they use for personal activities.

Table 7

Rate your ability to critically assess the potential risks of the AI tools you use, for professional activities. (1 = Not at all able, 5 - Very capable).



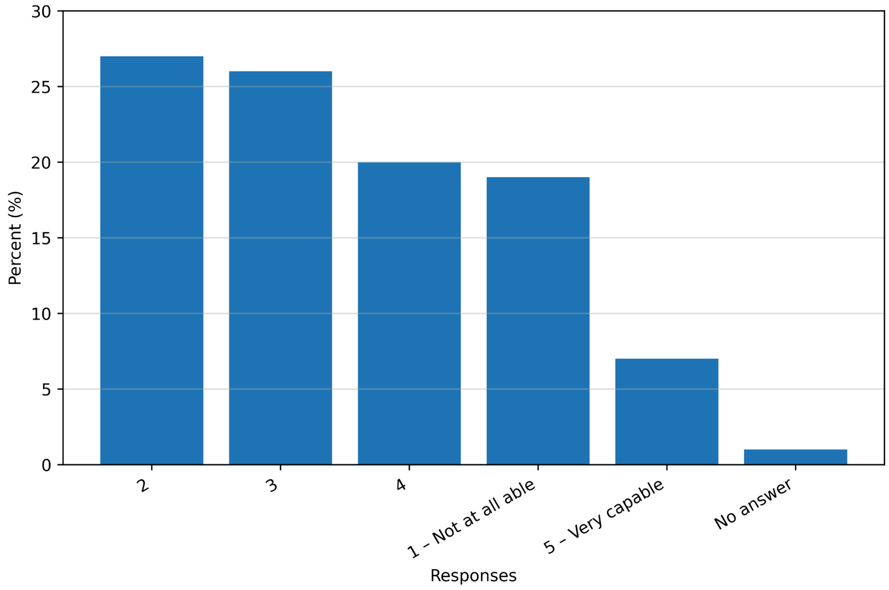
Funded by the European Union

Similar results are yielded in Table 7 with regard to AI used for professional activities as CSOs workers and activists report moderate to relatively high self-assessed ability to critically assess the potential risks of the AI tools they use for professional activities.

On the other hand, when discussion shifts to potential biases in AI systems used for personal or professional activities, respondents' confidence in their ability to assess them is lower as indicated in Tables 8 and 9.

Table 8

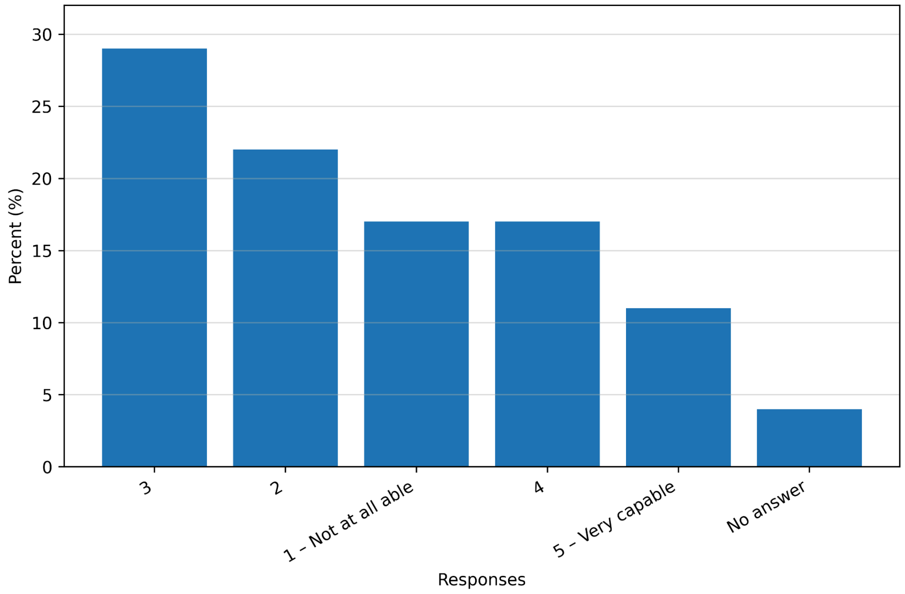
Rate your ability to critically assess the potential biases of the AI tools you use, for personal activities. (1 = Not at all able, 5 = Very capable).



As shown in Table 8, CSOs workers and activists report mainly moderate to low self-assessed ability to recognise potential biases in the AI tools they use. In this case, the percentage that reports low confidence is higher, approaching 20%. Biases in this context are related to gender, race, culture, socioeconomic status, or underrepresentation of certain groups in AI-generated content or decisions.

Table 9

Rate your ability to critically assess the potential biases of the AI tools you use, for professional activities. (1 = Not at all able, 5 = Very capable).



As indicated in Table 9, CSO workers and activists report mostly moderate self-assessed ability to critically assess the potential biases of the AI tools they use for professional activities, with fewer respondents selecting the highest confidence category.

Overall, respondents are more confident in their ability to identify risks (as they relate to issues of privacy, security, misinformation, job displacement, ethical concerns, or unintended consequences of AI outputs) than to identify biases (as they relate to gender, race, culture, socioeconomic status, or underrepresentation of certain groups in AI-generated content or decisions). This may reflect a perception that risks like data privacy are more tangible and easier to detect or guard against, whereas detecting bias requires deeper knowledge of AI systems that are already opaque and difficult to interpret.

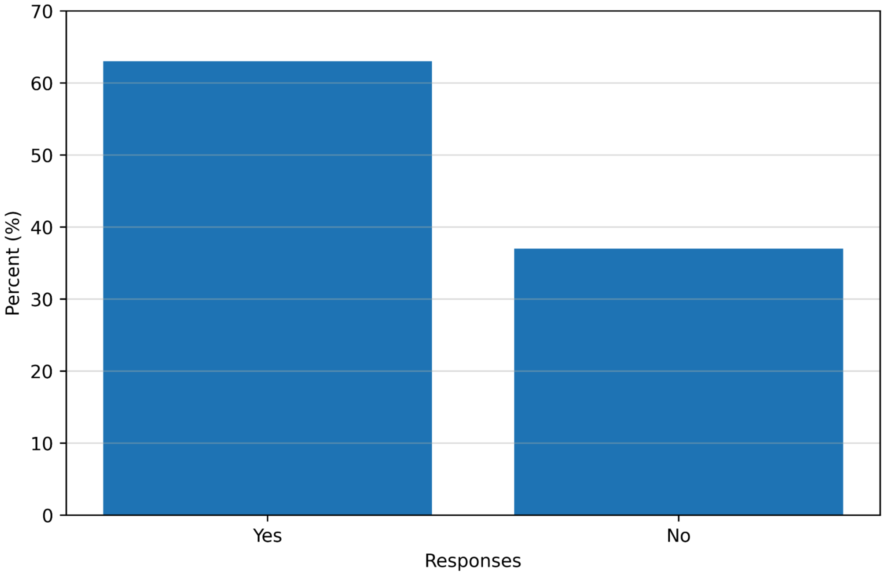
The next section shifts to discuss awareness of AI gender bias.

5.1.4 Awareness of AI and gendered and LGBTQ+ marginalisation

Within section 4, respondents were asked to answer whether they are aware of cases of AI bias relating to gender and LGBTQ+ communities and how concerned they feel about AI gender bias in general. Furthermore, they were provided 7 examples of AI systems (immigration technologies, applications in healthcare, education and student assessment tools, deepfakes and other generative outputs, facial recognition technology, hiring algorithms, content moderation systems) and were asked to rank them from the least to most potentially harmful in terms of bias. Results are included in table 10 through 12.

Table 10

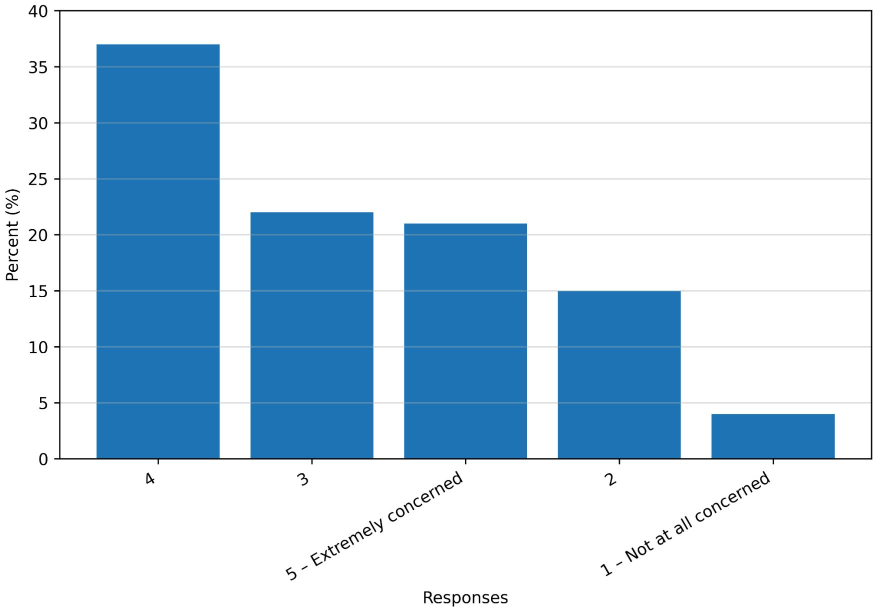
Are you aware of cases where AI systems have exhibited gender or LGBTQ+ bias?



As indicated in Table 10, almost ⅔ of respondents are aware of cases involving gender or LGBTQ+ bias in AI systems.

Table 11

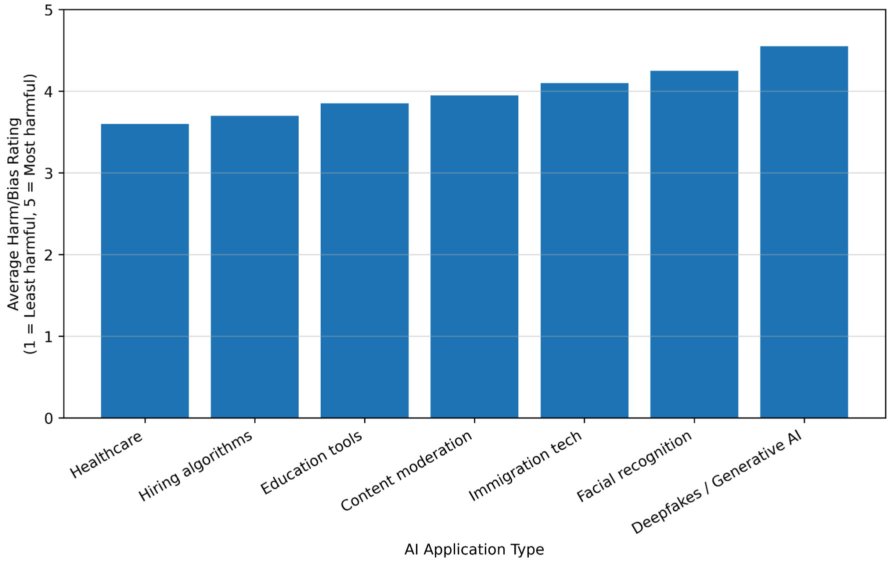
How concerned are you about the impact of AI systems, in general, on gender and LGBTQ+ equality? (1 = Not at all concerned, 5 = Extremely concerned).



The awareness recorded in Table 10 is also reflected in a high concern over the impact of AI systems on gender and LGBTQ+ equality.

Table 12

Ranking of AI systems from least to most potentially biased



As shown in Table 12, respondents perceived healthcare applications and hiring algorithms as the least likely to produce AI gender bias or LGBTQ+ bias, assigning them the lowest average harm rankings. By contrast, facial recognition systems and especially deepfakes or other generative AI tools were viewed as the most potentially harmful, reflecting concerns about hallucinations, their capacity to reinforce discriminatory outcomes and their overall impact on public debate. This topic will be reiterated and further explored in the findings from the workshops.

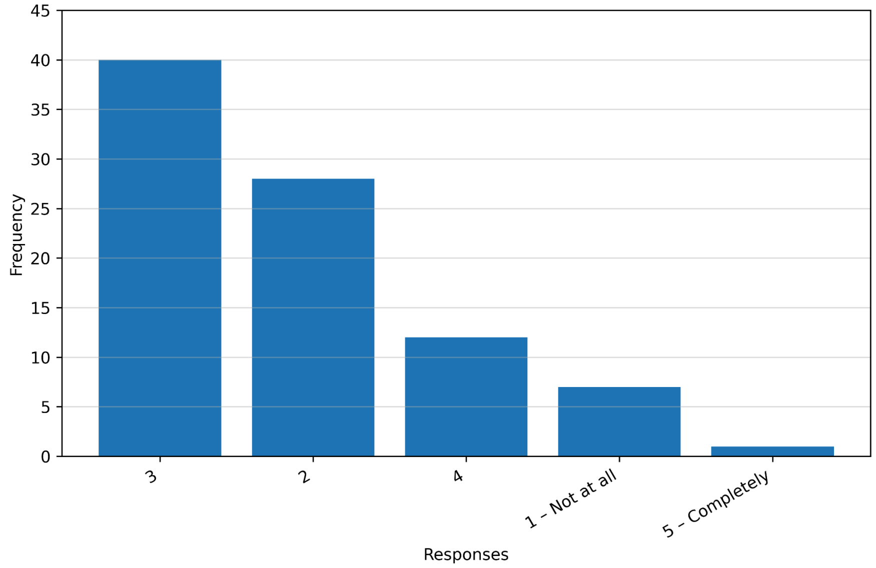
In the next section, the AI regulatory landscape will be discussed.

5.1.5 Awareness of AI regulatory landscape

Respondents were asked to evaluate their confidence in the potential of EU regulatory initiatives (like the AI Act) to address AI gender-based discrimination. Then, they were provided with a list of 7 stakeholders involved in AI regulation (national governments, tech companies, civil society organisations, international-supranational institutions, academic and research institutions, a coalition of the above and other) and they were asked to select 2 of them as the primarily responsible actors of AI governance. Results can be found in Tables 13 and 14.

Table 13

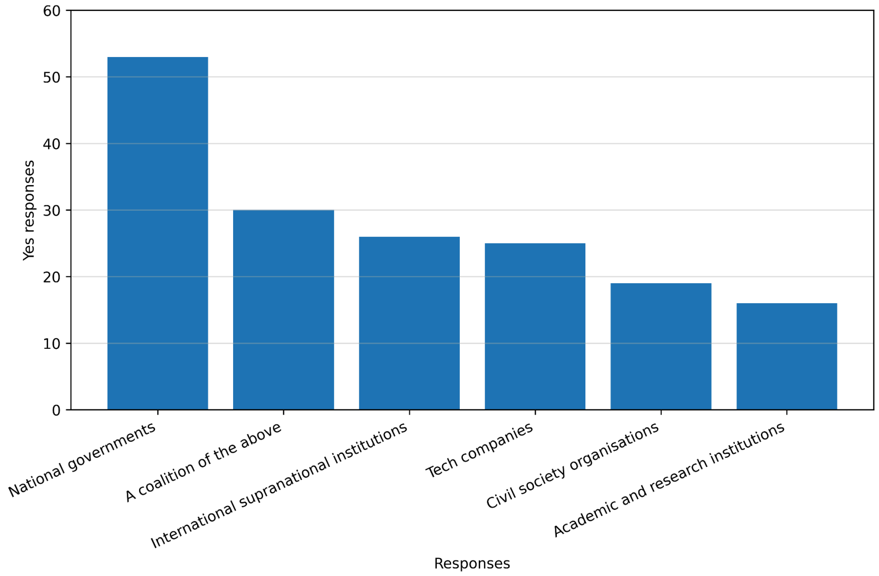
To what extent do you believe supranational AI regulations like the EU AI Act adequately address the potential of AI for harmful use or discrimination based on gender and sexual orientation in the EU? (1 = Not at all, 5 = Completely).



What is shown in Table 13 is a moderate to low trust in the EU's capability to mitigate AI gender bias, while very few see it as fully adequate.

Table 14

There are many stakeholders involved in and contributing to AI regulation. In your opinion, which are the primarily responsible actors in regulating AI systems to protect women and LGBTQ+ rights? (Select up to 2).



Funded by
the European Union

Table 14 shows that, as respondents are not particularly confident on international and supranational institutions as regulators of AI, national governments occupy a prominent position in the responses, followed by a coalition of actors.

The next section shifts to actions and initiatives by CSOs themselves to address AI gender-related bias.

5.1.6 Organisational actions and practices to address gendered and/or LGBTQ+ marginalisation

Respondents were presented with a list of possible organisational actions and educational activities related to AI bias and were asked to select all those implemented within their organisations. They were also asked how frequently their organisations collaborate with women's rights or LGBTQ+ advocacy groups to address the gendered impacts of AI. Then, participants were invited to identify the two groups of actors they consider most central when examining AI bias. Finally, they were asked to assess the extent to which AI gender bias is a priority for their organisation. Results are summarised in Tables 15 through 20.

Table 15

When addressing gender and LGBTQ+ bias in AI, which groups do you consider your main focus or priority? (Select up to 2).

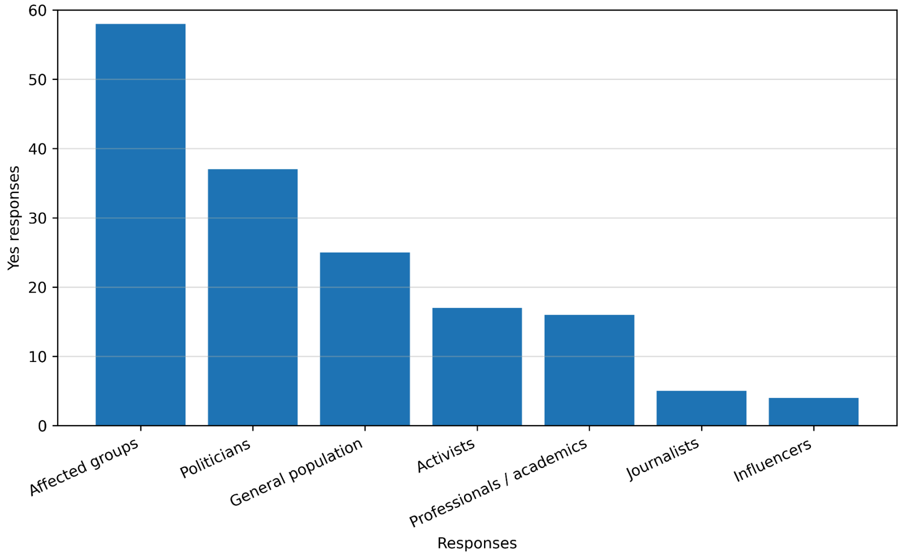
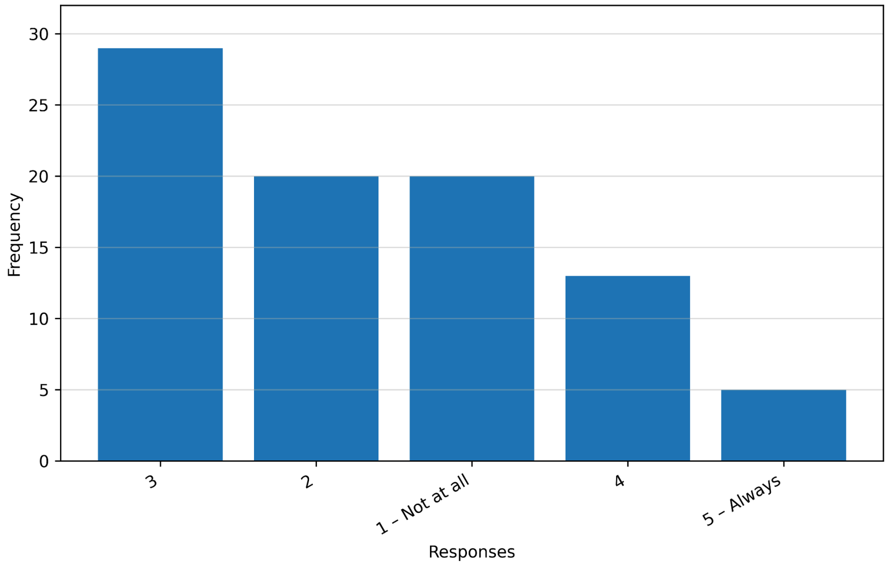


Table 15 shows that CSOs focus primarily on affected groups, followed by politicians and the general population, while journalists and influencers are the least prioritised.

Funded by the European Union

Table 16

When examining AI, to what extent does your organisation focus on its impact on women's rights and/or LGBTQ+ perspectives? (1 = Not at all, 5 = Always).

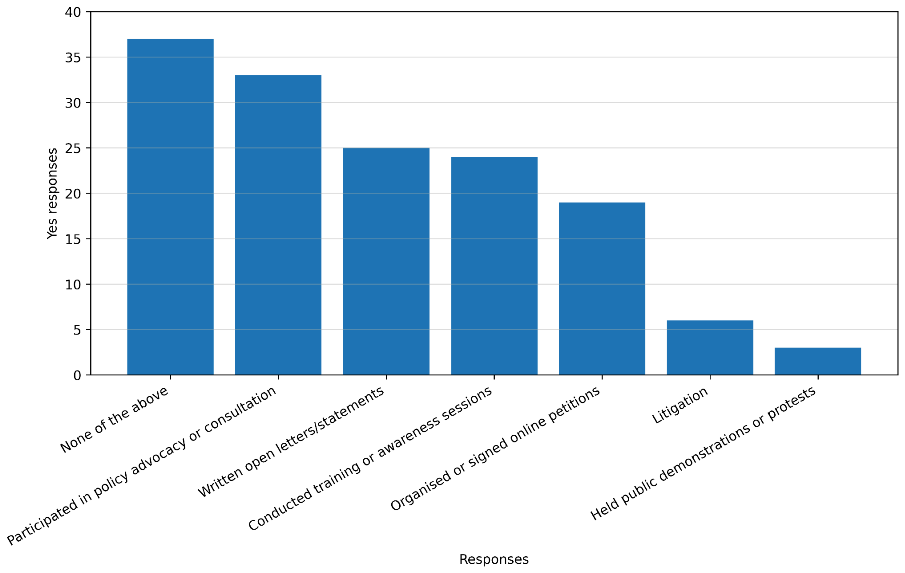


As Table 16 indicates, CSOs exhibit a moderate to low focus on AI's impact on women's and LGBTQ+ rights with very few respondents reporting that their organisation always treat this issue as a priority.

Table 17

Have you or your organisation engaged in any of the following actions to address gender and /or LGBTQ+ bias in AI systems? (Select all that apply).

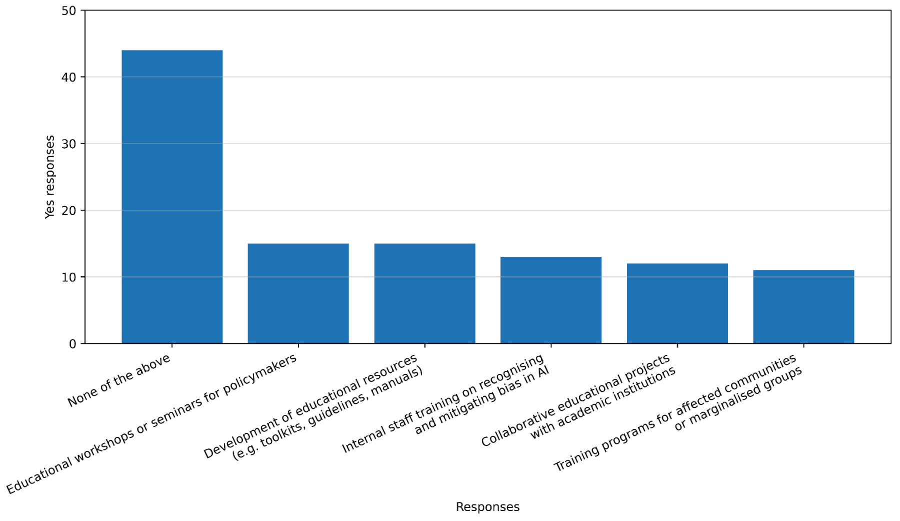
Funded by the European Union



Most respondents indicated in Table 17 stated that they and their organisations have not engaged in any of the listed actions; this inaction is partly interpreted through the workshops that will be presented in section 5.2. Among those who have, the most common activities are policy advocacy/consultation, open letters/statements, and training or awareness sessions, while litigation and public demonstrations remain rare.

Table 18

Has your organisation engaged in any of the following educational activities to address gender and/or LGBTQ+ bias in AI systems? (Select all that apply).

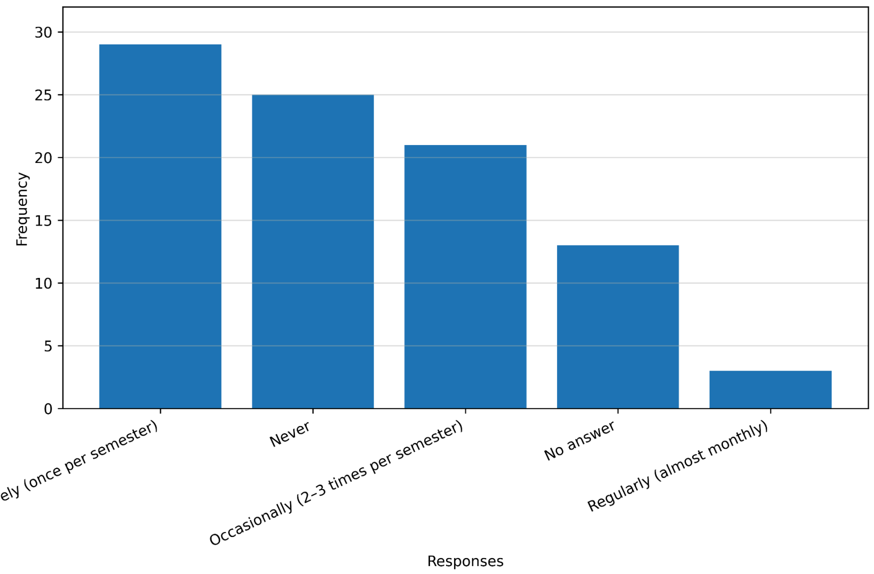


Funded by the European Union

Reiterating the results of Table 17, table 18 demonstrates that most organisations had no engagement in educational activities, while those who did mainly focused on workshops for policymakers, creation of educational resources, and internal staff training.

Table 19

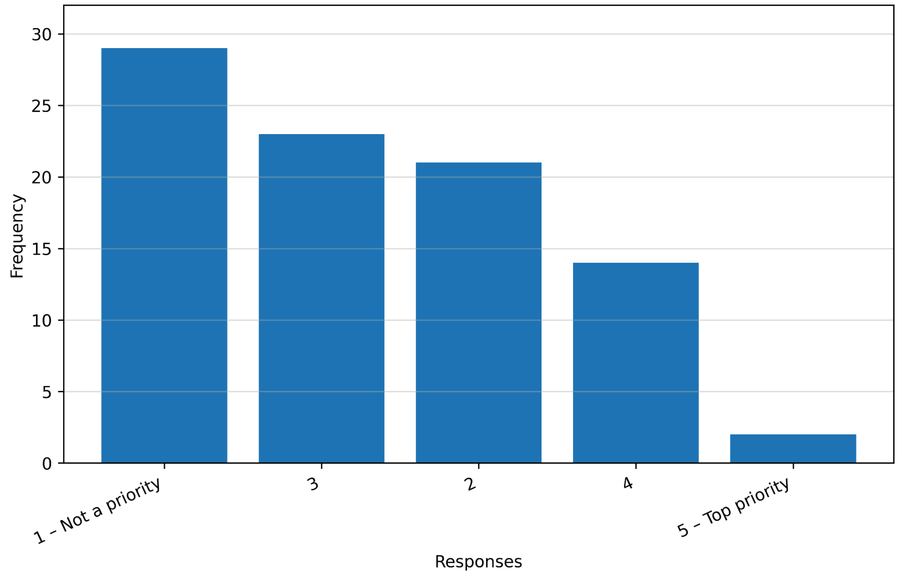
How often does your organisation collaborate with women's rights and/or LGBTQ+ advocacy groups to address the gendered impacts of AI?



Keeping in line with the previous findings, Table 19 shows that most organisations collaborate rarely or never with women and LGBTQ+ advocacy groups to address AI bias, while regular collaboration is extremely limited.

Table 20

How much of a priority is addressing AI-related gender and LGBTQ+ bias within your organisation's agenda? (1 = Not a priority, 5 = Top priority).



The findings on Table 20 seem as a logical conclusion of the previous ones; respondents state that their organisations rate addressing AI-related gender and LGBTQ+ bias as a low to moderate priority, with very few considering it a top priority.

5.1.7 Demographics of participants

The final section of the survey inquires about respondents' demographics on education, gender and age. Findings are captured in tables 21 through 23.

Table 21

What is your highest level of education completed?

Funded by the European Union

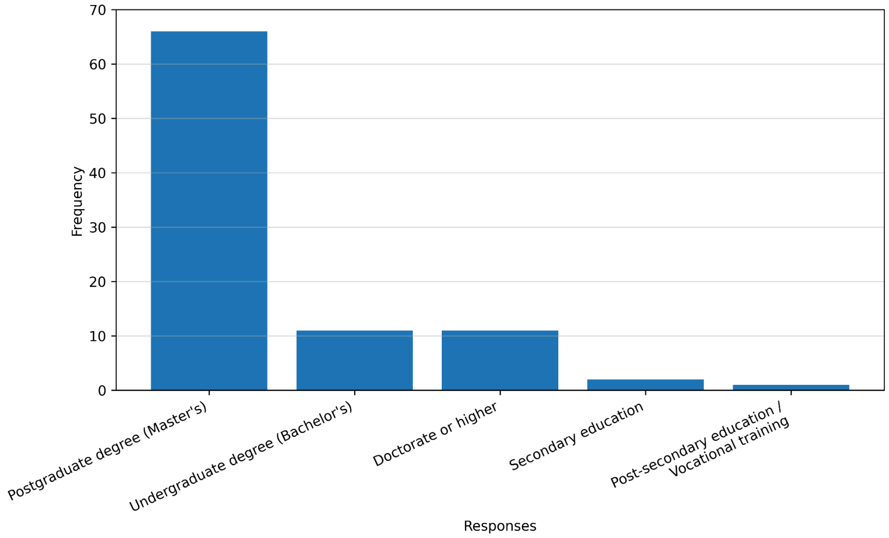


Table 22

Which of the following best describes your gender?

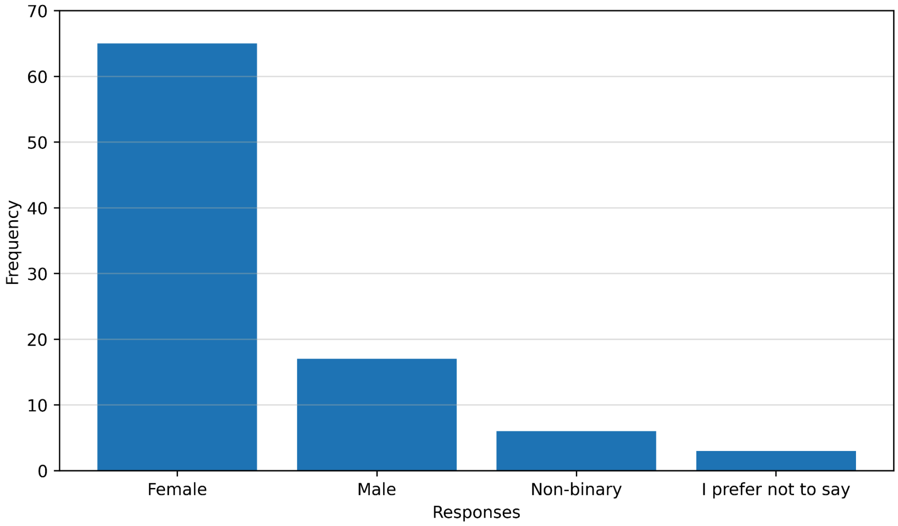
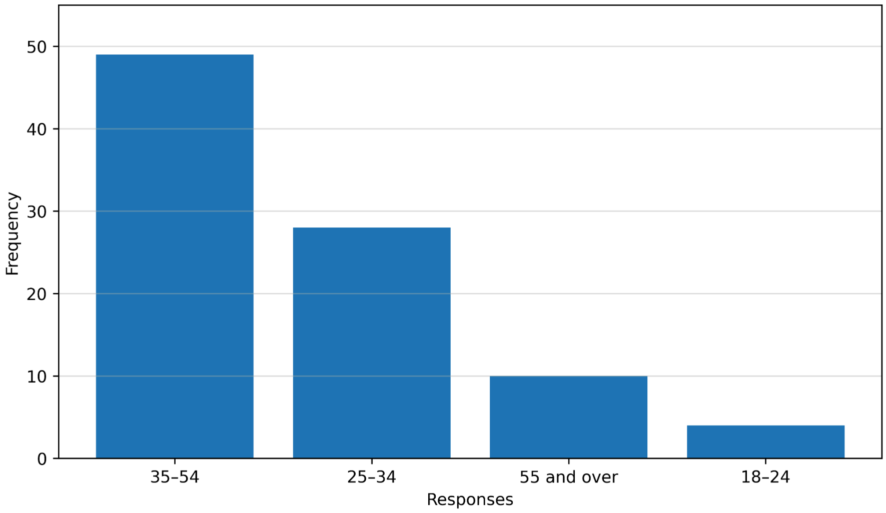


Table 23

What is your age group?



Survey respondents form a highly educated and predominantly mid-career group, with the large majority holding a Master's degree and most falling within the 35–54 age range. The sample is also predominantly female, with smaller proportions of male, non-binary, and undisclosed respondents.

5.2 Workshop findings

Workshop discussion confirmed the main findings identified through the survey. Participants reported themselves as users of AI systems, while at the same time they were largely associating AI systems with potential risks. They were conscious of AI gender bias and were aware of specific cases of discrimination as well the mechanisms involved. More importantly, they were wary of their own organisations' capacity to address AI gender bias and, keeping in line with the survey findings, reported limited engagement with actions and educational initiatives on this topic, with notable exceptions.

Furthermore, as will be discussed, workshops allowed us to observe tendencies and perspectives that were not captured in the survey, such as the concern that AI systems hinder CSOs' activity both online and offline. Participants demonstrated an overall pessimistic stance and worry for the prospect of individual CSOs and the future of civil society in general due to a number of factors.

5.2.1 CSO perspectives on AI-related risks

In discussing AI in general terms, participants in all workshops tended to focus significantly more on the associated risks than the potential benefits. In the beginning of each workshop, they were prompted to write down words that they associate with AI (either positive or negative or neutral); participants consistently wrote more words and phrases with negative connotations. In the discussion that followed, participants identified a wide array of AI-related risks and linked them with a variety of AI systems. However, three topics occupied a prominent position:

a) Erosion of democracy - undermining of public debate
b) Intensified surveillance both politically and commercially
c) Lack of transparency and accountability.

Other, less discussed, risks included exacerbation of economic inequalities, loss of creative and critical thinking, undermining of intellectual rights, and the role of AI in modern warfare.

In a similar vein, participants mentioned a number of AI systems and potential uses of AI. However, they were mostly focused on three types of AI

a) Generative AI
b) Content moderation algorithms particularly prominent in social media platforms
c) Facial Recognition Technology (FRT).

Other case studies included AI-enabled surveillance in immigration and border control, algorithms determining remuneration and task allocation in food delivery platforms and the use of AI in both healthcare and warfare.

In the following subsections, the three types of risks will be discussed and correlated with different AI systems.

5.2.1.1 Erosion of democracy - undermining of public debate

When discussing AI as a threat to democracy, participants often referred to "echo chambers" as a stepping stone to introduce this issue. As it was mentioned:

> Algorithms need training data, and that alone can shape what gets shown. If the algorithm is trained on existing Instagram data, say posts that historically got more likes, then it creates a loop where it keeps reinforcing the same things that were successful before. It's basically reproducing existing aesthetic or popularity criteria. And that leads to more of the same and at the same time the quality drops, because everything becomes more generic over time.

Other participants introduced the potential of AI to be used for hate speech content to be disseminated through social media:

> AI content targets first impressions and leaves a lasting mark on audiences. Most content from these parties is targeted against marginalised groups, making it a significant issue that threatens democratic processes.

In that sense, participants highlighted the dangers arising from the intersection of AI-generated hate speech with the tendency of social media platforms to abandon content moderation policies shifting to a "free speech first" orientation:

> The entire concept of content moderation has been pushed back. Both Meta and X have moved toward a "free speech first" approach. They've said, essentially, "Let's return to free expression," which in practice means less content moderation. That's a remarkable shift. We're seeing less human moderation, and more space for what they call "free opinions," including narratives that are extremely toxic and that previously were regulated.

In a similar vein, another participant stated that:

> This is something we can observe happening, and it's part of the rise of the far right. People are only being fed the same content repeatedly by algorithms — what was referred to earlier as "personalisation." This algorithm keeps showing you the same things you've seen before, and it can radicalise people, both positively and negatively. In this context, we are primarily concerned about the radicalisation of hatred.

More specifically, in the workshop held in Budapest, participants drew from their experience with the government threat to use FRT against Pride participants last June, to note a "chilling effect" to public participation resulting from AI systems:

> What happened in Hungary shows this clearly. There is an ongoing debate about the "chilling effect", even in democratic societies. The government used this "chilling effect" in its communications. They essentially threatened people: if you go to the Pride march, you will be identified, because it is impossible to hide from facial recognition technology. This becomes more like an authority figure reaching out in a coercive way, which is problematic. It creates a serious issue.

This is also a concern for intensified surveillance (see 5.2.1.2 below).

Drawing from the same case study, participants echoed these concerns while also pointing to the relative power of the technology industry relative to democratically elected state leaders.

> This [what happened in Hungary] ties into the powers of authoritarian surveillance states, the fragility of democratic safeguards, and the transfer of control to entities that can suddenly target vulnerable communities. These tech oligarchs, acting for profit, can even influence the appointment or removal of state leaders.

5.2.1.2 Intensified surveillance both politically and commercially

Unsurprisingly, the quotes presented above were tightly connected with the second major risk discussed by participants: intensification of surveillance, both politically and commercially. Predictive policing and FRT was the starting point in relevant discussions:

> Another concern regarding AI and related tools is predictive policing and facial recognition. These technologies have increasingly been used to target vulnerable people, especially at borders, resulting in profiling and biased assumptions about who may be committing crimes. There is an inherent bias in the people included in these databases through predictive technologies.

In a similar note regarding travel restrictions, it was mentioned that:

> Relating this back to AI, when entering a country (like the United States) everything is likely scanned using AI or algorithmic systems. This means people must be very careful about what they post or share on social media, as it could affect their ability to enter other countries

Recent political developments in the United States were often mentioned as indicative of the risks posed by AI systems:

> Many tools are convenient at first, but they can become tools of surveillance. For example, number plate recognition: you go to the airport, park your car, and it gets you in quickly. That seems efficient. But on the other extreme, if everyone is linked to a central system that can track all movements, that becomes a huge problem, as we see in the US with ICE. The question then is: what are you trading off? What are you giving up?

Concerns were also expressed regarding commercial profiling through Big Tech platforms.

> You know those cookies we all encounter online? Even if I reject them twenty times, people say, "You just accept." I say, no, I'm resisting. Still, I know they track me. For example, on Facebook, if someone writes to me or I'm talking with someone about yoga, suddenly I start seeing ads for yoga mats or online yoga classes. This is the ad tech ecosystem at work, which is heavily reliant on AI.

The most pessimistic version of this issue was offered by a Dublin workshop participant:

> There's so much information available about us, and even though companies say it's anonymous, there's always some back-end way to de-anonymize it. They can know everything about us, our preferences, who we are, where we fit. Our privacy is basically non-existent.

Participants also linked surveillance with inequalities and authoritarian forms of management in the workplace. As they noted:

> Algorithmic management tools in the workplace are a concern. In the platform economy - such as for people working at Lieferando, Uber, or care platforms- data about workers is collected and processed in ways that can be used against them, without their control. There is this particular big concern called algorithmic management, which we see most in digital platforms. I am much less worried about losing jobs…

5.2.1.3 Lack of transparency and accountability

The third major risk identified by workshop participants relates to lack of transparency and accountability. This issue needs to be understood in a complementary manner with the two identified above: erosion of democracy and intensified surveillance pose even greater threats when AI systems operate in an opaque manner. Connecting the different risks highlighted above, one participant noted the role that platform economics play:

> I think the problem is the liability and accountability for that cannot be clearly established. But I would also say that the platform itself has responsibility, because they have never succeeded in implementing a real set of rules to prevent this from happening. I think this is, of course, also part of their business model.

Another participant shifted to questions of accountability when AI is deployed by the public sector:

> I don't fully know the extent of AI use in public services yet, but I did notice Tara, the chatbot for the Department of Justice. The way it works can shift the responsibility of processing information onto the user. This is specifically an immigration chatbot. The stakes are very high, and under the EU AI Act, it's considered low risk because it's just answering queries. However, this removal of responsibility effectively puts the blame on people.

Painting an even more bleak picture, a participant in the Berlin workshop highlighted our collective inability to understand how AI systems operate, effectively linking the lack of transparency with a large scale threat to democratic participation:

> We move towards a point where our confusion is being weaponised—what we know, what limits are in place, and how all of this is being put into the world and into the things we're working on.

These are the main concerns discussed by workshop participants, indicating the inclination of CSOs workers and activists to examine the state of democracy, public debate and civil rights. However, a series of other risks were also identified. For example, with regards to creative and critical thinking and the use of Generative AI, one participant noted that:

> We won't be able to tell now, but the data 10 years from now might be interesting to see how young people are adapting to large language models that enable them to take a break in terms of certain decision-making or certain cognitive processes. I'm not sure the data's there yet.

5.2.2 CSOs Perspectives on AI-related benefits

As already mentioned, participants identified potential benefits of AI systems to a significantly smaller degree. The main topics that emerged here were efficiency and access to information. Participants, drawing from their work experience, highlighted certain advantages offered by AI systems. Their accounts indicate that Generative AI is the major type of AI systems deployed within CSOs, unsurprisingly since it facilitates research and creation of multimedia content.

> In our organisation, we use AI tools for research and other tasks to speed up certain processes and to connect with different types of knowledge. It is really helpful because there is a gap between how we articulate and discuss AI at the policy level and how we integrate it into our everyday work.

Another participant, drawing from their experience as a PhD candidate, noted that:

> If you're working on a PhD and want to test ideas, especially if your topic is niche or no one around you is particularly interested, you can put your thoughts into the system and it talks back. It helps you parse things through conversation. Even if you don't use the output, it helps you digest ideas. I actually like it for that. Some people use AI tools in research as well. For instance, you can ask a model to scrutinise something "from the lens of a lawyer" and it will pick the argument apart once you instruct it that way.

Overall, many participants recounted experience from using AI within their own organisations but were cautious to identify benefits as they wanted to highlight the importance of context. As one participant stated:

> For me, "efficient" isn't a neutral word. Even when it is meant neutrally, I'm cautious using it, especially in policy contexts, because efficiency is subjective. Whose efficiency are we talking about? From what perspective? For example, if I ask AI to produce a LinkedIn post, I then have to make sure it doesn't obviously look AI-generated. I need to clean it up and double-check that everything in it is accurate. So am I really saving time? If efficiency is measured in minutes or hours, I'm not always convinced it's actually efficient. That's why I challenge the idea that these tools automatically create efficiency

5.2.3 AI and Gender Bias

When discussion shifted to AI gender bias, participants demonstrated understanding of the technical aspects of bias in training datasets as well as a broad conceptualisation of how gendered, racialised and heteronormative stereotypes can be replicated within AI systems.

One of the participants, used Generative AI to illustrate how stereotypes persist:

> For example, if you ask it to create an image of an IT person, you usually get a male with a beard and glasses. These AI tools have some built-in biases; some, I would say, a lot. One of

> our members presented an AI hiring tool, which was built by an EU digital SME and used to select resumes and conduct interviews. Can you guess what this AI depicted the hiring person as? Actually, not what you might expect—she was HR, young, female, very blonde, and very conventionally "HR-looking." It's an interesting illustration of how biases appear in AI systems.

Another participant linked biased datasets to AI used in healthcare to demonstrate potential risks for patients belonging to marginalised and underrepresented communities:

> When it comes to data, and when it comes to gender, I think we are definitely missing a lot of diverse data. For example, in healthcare, people who are not cis, white, and male are recognised less by different medical schools and systems, and diagnoses get missed. People who do not identify as cis male are not represented in the data, which means that in healthcare we receive fewer diagnoses, which delays treatment and can actually have severe consequences

In a similar note, another participant noted how "edge cases" are often not considered when designing AI systems, resulting in discrimination based on cis heteronormative concepts of gender binaries:

> Another major issue is the continued binariness of gender throughout the entire sector. This relates to data in a huge way. Gender data is vast, yet it does not consider differences related to sex assignment, socialisation, identity, expression, and so on. Not all of these aspects are relevant in every circumstance, and I am not saying that every form should ask all of them. Rather, systems should consider what is actually needed, what they are interested in, and why.

Linking this issue with the lack of accountability discussed above, another participant mentioned that:

> AI could widen the gender gap. It makes us invisible because technology's neutral, and we're supposed to accept it when it comes from technology, but it feels like we're taking steps back. And the lack of a responsible person, there are so many different players that have a role in how AI is deployed.

The issue of design was often discussed, indicating an understanding that extends beyond biased datasets:

> Gender bias is about design. For example, women get injured or killed in car accidents at higher rates because most crash-test dummies were modeled on male anatomy. This has been known for a long time, yet little has changed. It feels like a vicious cycle, AI is just going to keep feeding on these same biases. How do we stop it when we haven't been able to stop it in the real world?

In offering a justification for this, participants often pointed to a broader lack of diversity within the tech sector (which we discuss in detail for Digital SMEs in FORSEE D3.4) For example one mentioned that:

> There is a systemic nature to these issues. We do a lot of work on gender-based violence, and we always talk about how systemic measures at every level of the system affect this. We all know about the lack of representation in senior positions, and it is hard to address

Overall, the workshops confirmed the survey's findings on gender bias, indicating that CSOs workers and affiliated activists are concerned about gendered distortions in AI. At the same time, they possess adequate knowledge to relate these biases to issues such as the composition of training datasets, design omissions in AI systems, and the broader lack of diversity within the tech sector. Participants demonstrated a considerable level of expertise, reflecting their professional roles in CSOs that, among other issues, study and address challenges arising in the digital environment. Additionally, their identities as women and members of LGBTQ+ communities may further motivate them to scrutinise and address the relationship between AI and gender.

The next subsection examines how participants perceived the role of CSOs with regard to AI gender bias; whether it constitutes a priority, what strategies have been developed and what kind of coalitions yield results.

5.2.4 CSOs activity to mitigate AI bias

CSO activity confronts two concerns in particular: CSO participation and visibility in an era of AI surveillance and mediated visibility, and also AI gender bias.

5.2.4.1 AI as a risk to CSO participation and visibility

In discussing CSOs themselves, workshops provided an interesting finding that was not addressed in the survey: AI poses not only a risk to democracy in general but to CSOs organisations and workers in particular in areas of civic participation and visibility.

CSOs participation

The most evident case for civic participation was the use of FRT against Pride protesters in Budapest. Unsurprisingly, this created distinct political, civil, and criminal sanction risks for CSOs, their workers, and affiliated activists involved in organising the Pride. One participant recounted what transpired:

> Hungary, very famously, has a government that banned their Pride parade. In June of this year, when the parade was supposed to be banned, a record number of people showed up. There was something about facial recognition being used to track the people who attended.

> The scary reality is that identities are being criminalised in Hungary. A new act was rushed through Parliament in April without due process, stating that if you attend a protest that acknowledges the right of LGBTQ people to exist, you can face civil fines. Organising such a protest carries criminal sanctions, including jail time. Despite this, people came out in incredible numbers. The Hungarian Civil Liberties Union was involved, and it became a massive and powerful celebration. It's alarming what the government is doing, but also inspiring what the people are doing in response.

The risks of persecution based on one's gender or sexuality status, despite EU safeguards for protected identities, extends concerns about the erasure of identity, personal or organisational, by AI systems empowering governments and other institutions seeking authoritative policies.

CSOs visibility

While distinct from state repression of free expression and participation, platform governance and related algorithmic restrictions emerged in the workshops independent of the survey. These findings proved particularly valuable, as the algorithmic regulation of civil society's visibility is often overlooked in broader discussions of AI-related risks.

Workshop participants expressed concerns that algorithmically moderated platforms may target political content for reduced visibility or outright removal, often without providing clear justification or explanation for such actions. This significantly harms the reach and impact of CSOs as their online presence is effectively limited:

> When society labels something as political, it's part of public debate; people can argue about it, challenge it, or reinterpret it. But when an algorithm labels something as political, it becomes much harder to contest. If the system says your content was mislabeled, you still need to go through a long and difficult appeal process, and many organisations simply don't have the resources for that.

One participant discussed a specific case study where the visibility of their organisation's online activity was constrained:

> We want to reach the right people, including marginalised people and people from different groups. But we face the issue that sometimes our content gets filtered or flagged. In one case, our campaign was deleted by a major video platform, fully deleted. The platform flagged it as containing inappropriate content, which was clearly a mistake. There was no warning and no request to double-check the video. It was immediately blocked, and the entire account was gone for several weeks. We issued a press release, and only then was the issue escalated to a higher level within the company, allowing us to finally get into the proper content moderation channel. Afterward, the company offered us an invitation for educational content so that our social media team could "improve," even though we had not done anything wrong in the first place

Even when CSOs content is not directly undermined, AI summaries in search engines reduce website visits; the notorious "death of the click".

> Another concern is that people are increasingly using chatbots and AI summaries for information consumption, which may reduce the visibility of civil society and its diversity. People no longer visit websites as much, but large civil society organisations rely on website visits to provide information about their work, fundraising, or newsletters. Even though they are experts and still provide information, they are not being recognised as a source anymore. This is a massively unsolved problem at this point.

These issues were not included in the survey design and provide a different perspective on the challenges that civil society faces that should be addressed in future research.

5.2.4.2 CSOs strategies and partnerships to address AI gender bias

Regarding the mitigation strategies of CSOs, participants reiterated the importance of advocacy and finding the best ways to communicate with the broader public. For one of the participants that meant strengthening people's awareness of their rights:

> An essential role for us is that people understand themselves as rights holders, already being protected, for example, before something even goes wrong. And to challenge AI systems mostly online, let's say, through platforms, or through an employer using them. So that's one narrative we subscribe to and communicate a bit more

Social media remain crucial fields for CSO in their effort to reach larger audiences, despite new platform policies blocking "political" content:

> It gets much, much harder now, and with the blocking of political ads, it has become almost impossible to do our work through these communication channels. Yet social media remains a hugely important way to reach people where they are.

Advocacy is often linked with AI literacy initiatives:

> My hope is actually to create more AI literacy, in a way similar to media literacy, because I think that has a huge influence on how people perceive AI systems, how they engage with them, and how they use them.

In a similar note on literacy, another participant stated that:

> Speaking about civil society organisations, there is another important aspect that goes beyond regulation, and that is providing literacy. A big part of the problem is that people using social media often don't understand where the information is coming from. It's important to create literacy, educate people, and help them understand how to fact-check and always verify sources.

5.2.4.3 Regulatory initiatives and SMEs strategies of intervention

With respect to regulation, participants did not identify one clear point of pressure but described multi-level strategies. Given the recent developments in Hungary, participants from Budapest felt that the national government does not represent a favorable point of pressure:

> In Hungary, the national government is disregarding our rights. So you're saying: let's address the EU, let's address the UN, let's build coalitions. And I think the Hungarian perspective is very valuable because of this. Working in Hungary in this field is extremely difficult. But at the same time, we do have an AI Coalition in Hungary. So if the government is not open to these issues, we need to mainstream them through other entities.

On the other hand, frustration was also expressed with regard to the EU framework:

> I think it would be useful to send petitions as a norm, especially regarding regulation on how our data is stored and how we are tracked. One of the fundamental issues that needs to be raised is that, if you check the existing regulations,the AI Act, EU General Data Protection Regulation (GDPR), child rights conventions, women's rights conventions, and so on, none of them mention bias. Bias simply does not exist in these documents. I cross-checked and researched this, and I think it is fundamental.

Overall, participants noted that CSOs are excluded from regulatory initiatives and not included in the relevant debates. The following quotes are indicative of the feeling of being "left out":

> When attending high-level events with government officials or parliamentarians, the conversation often only happens between the private sector and public sector. Other organisations are either not invited or not genuinely welcome. This creates a feeling of always running behind. By the time CSOs gather information, decisions or conversations are already underway.
>
> We need more experts, more research, more fundamental grants for lobbying, and we need to be invited to the table. When discussions about the AI Act are happening, civil society is barely represented. The for-profit sector, banks, the military, and others are all there, but civil society hardly appears.

Regarding effective strategies to mitigate bias, the importance of building alliances and collaborations across different actors, groups and sectors was often highlighted:

> Civil society organisations, in particular, can build around these connections. Even small initiatives can be valuable, like creating a code of conduct among parties willing to engage on topics such as migration or the use of AI. These efforts don't necessarily rely on regulation. Starting with conversations, bringing a group together to brainstorm, and maybe drafting a document can already be a first step.

One participant noted how alliances can be built around knowledge sharing and pooling resources to create shared datasets on platform activity:

> There are two initiatives I'd like to highlight. The first involves a number of civil society organisations trying to issue a mass data request to the major platform providers. The goal is to gather, for example, the thousand most viral posts per day in each member state, and to create a kind of knowledge base or database. This would allow them to vet the content and see which posts are actually being widely disseminated each day. The second initiative involves different civil society research organisations that actively engage with platform data. They are trying to pool their knowledge and requests through the platforms to create some form of shared space, whether a database or similar, where they can see what has already been researched and what data has been made available. This would allow organisations to reuse and share this kind of data. Of course, there are many barriers and restrictions, but this is primarily about knowledge sharing.

In contrast with survey findings, workshop participants were more familiar with existing campaigns to address AI gender bias and some of them deployed a number of interventions:

> So far, our strategy includes advocacy, providing independent information, education from experts, and building coalitions between CSOs and academics.

On the other hand, some participants reiterated the survey findings, stating that necessary collaborations are lagging behind:

> In this field [AI bias] there is still a lack of collaboration between, for example, consumer-protection actors and anti-discrimination actors. Bringing these groups together—working jointly, coordinating pressure, and developing shared strategies—could be a really important next step.

Once again, in discussing potential alliances, the notion of "being political" emerges as a barrier, hindering CSO activity:

> There is also a lot of confusion for us. When we try to join or collaborate with larger organisations, they sometimes question whether this issue is "too political." People are afraid of anything political. I usually say that everything can be political, it doesn't mean we are acting against the government.

In discussing barriers to CSOs activity, participants often noted that the past years have been characterised by a series of budget cuts and the concomitant struggle for funding. Two participants noted the impact of budget cuts in creating a sense of uncertainty.

> We received funding from USAID and also had another project funded by SCBO, with a budget for these kinds of initiatives. So, yes, definitely. In the last 8 or 10 years, UNHCR used to provide support, but now they don't. I thought the UN was one of the last safe places, but apparently not.
>
> There are capacity issues. We have our daily work, so how can we go out and learn about everything on top of that? Especially if funding is shrinking, CSO workers are having to do more and more. It's even harder when there's a crisis situation every other second. What do you mean by crisis? To run a family. They come to our office and tell us that they don't have anywhere to go and no income. In terms of hierarchy of needs, the immediate priority is ensuring people have housing before worrying about how technology is impacting them.

The final quote perhaps sheds light on the issue identified in the survey (CSOs not treating AI bias as a priority). Within a context of austerity and rising uncertainty, AI genders bias risks being treated as of secondary importance.

Section 6: Discussion

In contrast to the AI industry (where FORSEE's D3.4 research identified predominantly positive expectations, with risks acknowledged but treated as secondary), CSO respondents adopt a far more critical stance. In the survey and even more clearly in the workshops, participants devoted only limited attention to potential benefits, instead foregrounding complex and multifaceted risks. While respondents expressed a degree of cautious optimism; this was largely confined to technical aspects of AI such as efficiency, speed, and cost reduction.

By comparison, respondent concerns were considerably stronger and more diverse. While survey and workshop findings. in terms of identified risks were not identical and the methodologies themselves do not allow for a direct comparison of findings, a common thread ties them together. For CSOs workers and affiliated activists, AI is negatively associated with politics and democratic participation, surveillance, lack of accountability as well as certain material (environmental and labour) impacts which were more prominent in the survey. The predominance of risk-related categories, both in the workshops and in the surveys, suggests that CSOs form a critically engaged sample that approaches AI through a lens of power asymmetries and rights-based concerns.

A notable tension emerges in the survey: respondents are highly critical of AI yet report using it frequently. This apparent contradiction can be partly explained by the nature of CSO work, which often involves tasks that Generative AI systems can support, including communications, a point that surfaced repeatedly in the workshops. Such usage may also clarify the survey pattern: respondents feel reasonably confident in the outputs they supervise directly, but far less confident in detecting underlying bias, which requires insight into the opaque processes through which AI systems operate. Workshops assisted in further illuminating this issue as they indicated that participants are familiar with Generative AI and use it frequently. Possibly the nature of CSO work (involving generation of multimedia content) creates an environment favourable for Generative AI systems. At the same time, this raises an additional question for future research regarding whether CSOs have established policies for the internal use of AI, and if so, what form these policies take.

With respect to AI gender bias, concern about its impact is widespread across CSOs workers and affiliated activists. Participants in the survey and in the workshops were aware of cases of gender-based discrimination. In particular, workshop participants discussed case studies of how gender marginalised groups were targeted by AI systems, while their experience illuminated the potential use of algorithms on social media (Berlin case study) FRT against the Pride Parade in Hungary (Budapest case study) or widespread adoption of policing FRT in Ireland (Dublin case study).

Overall participants across our study demonstrated a significant level of expertise on the subject matter and a familiarity with academic literature, specific case studies and policy challenges. AI gender bias was linked in the workshops with all three categories of risks: the way AI erodes democracy, either through industrial platform mediation of public debate or government deployment of AI against marginalised gender populations, leading to intensified surveillance, targeting or erasure in both public and industry spaces, and a lack of transparency about how AI operates or is deployed or how industry and government account for such actions. The depth of knowledge of CSO workers and affiliated activists (as captured in the survey and reflected in the workshops) further highlighted the contradiction between awareness and action, making this one of the key findings of the present research; CSOs are highly aware of AI risks and the potential for AI gender bias but lack the capacity, tools, or strategies to translate knowledge into sustained action.

On the subject of regulation, the survey yielded more interesting results as it indicated a limited confidence towards EU capabilities as a regulator of AI and, at the same time, an increased focus on national governments as actors responsible for AI governance. This finding requires further research as the AI Act comes into effect. Conversely, survey responses showed very little confidence in the capability of CSOs and academic institutions to act as regulators; this finding is partly corroborated by references in the workshops as participants highlighted the marginalisation of said actors from policy consultation.

6.1 Structural barriers to CSOs and potential ways forward

When prompted to discuss the activity of their own organisations, CSO workers an activists express frustration that, despite the recognised importance of AI-related gender bias, their organisations rarely take concrete action, build alliances, or engage in educational initiatives. Our initial findings point to a discrepancy between CSO workers and affiliated activists and the capacities of their respective organisations, reflecting a divergence in priorities. While workers and activists express a willingness to engage with AI-related gender bias, their organisations tend to implicitly treat it as a secondary concern. This does not imply that all CSOs should be mandated to engage with AI and gender bias issues; rather, it highlights the need for further research into why this willingness is not proportionately reflected in CSOs' activities.

The workshops allowed us to engage fruitfully with survey findings through a new lens. The issue is not a lack of willingness to practically engage with AI bias but a lack of resources, a limited presence in policy discussions and, at times the direct use of AI systems in a repressive manner against CSOs, as was the case with the Hungary Pride. This is compiled by a larger concern for civil society regarding capacity generally. The United Nations Special Rapporteur on the Rights to Freedom of Peaceful Assembly has in 2025 summarised internal CSO consensus that Civil Society is at risk in part due to structural crisis in the global aid system (Romero, 2025). Regarding funding generally, we are witnessing the rapid termination and dismantling of major aid institutions, requiring CSOs to shrink operations and transform their advocacy strategies simply to keep their doors open. As a broader political shift is underway, contesting civil liberties, CSOs feel that their role is contested and their engagement with new technologies and their implications is effectively undermined (INCLO, 2025).

This situation is further exacerbated by participants' comments on the exclusion of CSOs from consultations on the European AI regulatory framework, an issue that has been identified by Siapera and Farries (2025) in relation to the Digital Services Act. These findings raise serious questions about the viability of the multi-stakeholder engagement envisaged in EU policy documents. When coupled with budget cuts, this trend points to the growing incapacitation of CSOs and a further decline in their engagement with AI-related gender bias, which survey results indicate is already limited.

Platform governance and the precarious position of CSOs

However, the potentially most significant finding is the link participants drew between the erosion of civil society and democratic spaces and the weakening of their capacity to address AI bias through a prominent presence in online public spheres. Workshop conversations were particularly illuminating in this regard, connecting contemporary developments, such as extreme moderation of political content in social media platforms and the "death of the click" (see the Berlin case study), to a diminished public sphere that restricts open and democratic debate on AI's societal impacts. As social media content advocating for gender-marginalised communities is targeted as "political content" and is restricted, the question emerged whether algorithmic content moderation treated hate speech in the same manner; further research is necessary to clarify this point. In any case, as CSOs experience limited visibility for their online presence, their capability to address AI bias or organise educational activities to raise their internal capacity, is significantly restricted. This tendency, coupled with broader austerity and lack of funds, raises the concern whether CSOs are under threat of being marginalised and significantly weakened.

However, our findings also point to a different direction; the persistence of civil society mobilisation in the face of adversity and repression.

The persistence of civil society

A notable source of optimism was identified in partnerships developed with and around gender-marginalised communities, aimed at increasing AI literacy while empowering these groups to engage publicly with AI systems and policymakers. With some national exceptions (notably Hungary), participants in both the workshops and the survey identified national legislators and governance mechanisms as critical pressure points for intervention. Although synergies in this area were reported to be limited, practical experiences suggest that engagement through this pathway has proven to be effective.

Finally, instances of the undemocratic use of AI should not be understood solely through the lens of "chilling effects." In reference to the deployment of FRT against Pride Parade participants in Budapest, workshop participants highlighted not only the associated risks but also the continued mass participation in the event. This finding can be read as an argument against techno-pessimism; while AI may possess the technical capacity to undermine democratic participation, its effects are ultimately mediated by political context. Civil society and its organisations therefore continue to retain the potential to contest the repressive use of AI. Participants in both the Dublin and Budapest workshops interpreted such uses of AI and the corresponding public mobilisation as a potential "stepping stone": beyond generating concern, these practices were seen as opportunities to raise awareness about AI-related risks and to mobilise and empower affected communities.

Nonetheless, the overall study findings paint a bleak picture on the prospects of civil society to engage meaningfully with AI systems and gender-related biases. Policymakers need to reconsider the role of CSOs as stakeholders in AI governance both with regards to resources and access to deliberation channels. Furthermore, an issue arises for future research: if public advocacy of CSOs and LGBTQ+ communities is treated in social media platforms as "political content" to be banned or to have its visibility limited, then AI gender-bias needs to be expanded as a research field to also examine and address this.

6.2. Limitations

Several limitations of the present study should be acknowledged regarding the workshops, participants and surveys.

Regarding the workshops, our participants reflected a unique sample in that they were relatively informed and active in mitigating AI gender bias. They also were largely speaking contextually at the domestic state level. Future research could draw on a broader range of EU CSOs, potentially capturing perspectives that are less specialised and more reflective of the wider population of CSO workers and affiliated activists. Future studies could continue to gather valuable data from CSOs working at the EU member state level, beyond Germany, Hungary and Ireland.

Second, interviewees participated as representatives of their organisations rather than as private individuals, which likely influenced how openly they addressed AI gender bias or their own use. Several responses suggested a tendency toward perceived organisationally acceptable or reputationally cautious framing This raises concerns about social desirability bias (see Bisto JP, 2022's explanation of these dynamics in qualitative interviews). Such dynamics may have led interviewees, particularly those in senior or client-facing roles, to downplay structural gender bias in AI systems they develop or deploy.

Regarding the survey, first, it should be noted that our methodological approach carries rather inherent constraints. Overall, word associations are less precise than semi-structured interviews (though easier to deploy across large, disseminated samples). Their analysis is potentially subject to the researcher's subjectivity and their results can potentially be biased by the preceding information, which is why they were placed at the very beginning of a questionnaire. Furthermore, free associations, which reveal immediate (and often impromptu) reactions rather than deeply articulated reasoning; and asking respondents to rank or evaluate their responses imposes a certain hierarchy even where respondents may not necessarily have or experience one (Moliner & Monaco, 2017).

Second, while we were optimistic that we would see 400 participants respond to our survey given the PI's fellowship in established CSO research networks we only received 91 valid responses. After analysing the valid responses and feedback from the organisations supporting the delivery of our survey, the primary challenge we've identified in this process, and discussed in our conclusion, is a resource bound one: the current fragility of CSO capacity.

Section 7: Conclusion

This study set out to examine civil society organisations' (CSOs) perspectives on artificial intelligence (AI), with a particular focus on AI-related gender bias, existing mitigation practices, and the structural constraints shaping CSO engagement. The findings demonstrate that CSO workers and affiliated activists possess a high level of awareness of AI systems and their associated risks, particularly in relation to democratic erosion, intensified surveillance, and the lack of accountability of technology companies. Concerns regarding gender-based algorithmic discrimination are widespread, yet confidence in the ability of existing EU regulatory frameworks to effectively address these issues remains low.

Despite this critical awareness, a clear gap emerges between individual expertise and organisational capacity. AI gender bias is rarely treated as a strategic priority within CSOs, with limited public advocacy, educational initiatives, and sustained partnerships observed.

Participants consistently linked this gap to structural constraints, including limited financial and human resources, the impact of algorithmic content moderation on civil society visibility, and the systematic exclusion of CSOs from regulatory consultations and policy-making processes. These factors collectively contribute to the marginalisation of civil society voices within AI governance, undermining the multi-stakeholder model promoted in EU policy discourse.

The study's findings point to the need for targeted and structural policy interventions. Meaningful inclusion of CSOs in AI regulatory deliberation must be strengthened, moving beyond symbolic consultation towards sustained participation. Dedicated funding streams and capacity-building mechanisms are required to enable CSOs to engage with AI governance in a consistent and informed manner, particularly in relation to gender and LGBTQ+ marginalisations. In parallel, platform governance practices (often overlooked in debates on AI governance) limit the visibility and reach of civil society content and therefore warrant closer regulatory scrutiny.

Overall, the present report highlights the rethinking of AI governance as a participatory process that enables substantive civil society engagement through funding, access to deliberation, and safeguards against platform practices that marginalise CSO voices.

Abbreviations and definitions

AI Gender Bias: AI systems' reinforcement of existing technical errors together with structural, systematic and societal prejudices in relation to gender.

CSO: Civil Society Organisation

Discrimination: The real world consequence of AI bias leading to the differential treatment of different categories of people, for example race, LBTQ+ status or gender.

EDRi : The European Digital Rights Initiative

FRT: Facial recognition technologies

GDPR: EU General Data Protection Regulation

Intersectional bias. AI performance declines across intersecting attributes., for example, gender and race and other demographic characteristics

Sociotechnical analysis: Analysis that considers the interdependent relationship between social elements (people, culture, rules, values) and technical systems (tools, processes, technology)

Section 8: References

Access Now et al. (2023, July 12). EU Trilogues: The AI Act must protect people's rights https://www.accessnow.org/press-release/eu-trilogues-ai-act/

Ada Lovelace Institute. (2023, March). Inclusive AI governance: Enabling meaningful civil society participation in standards development. [Report] https://www.adalovelaceinstitute.org/wp-content/uploads/2023/03/Ada-Lovelace-Institute-Inclusive-AI-governance-Discussion-paper-March-2023.pdf

Banerji, L., & Satija, D. (2025, February 7). AI as double speak for austerity. *TechPolicy.Press.* https://www.techpolicy.press/ai-as-double-speak-for-austerity/

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity.

Bhattacherjee, A. (2012). *Social science research: Principles, methods, and practices* (2nd ed.). University of South Florida Scholar Commons. https://usq.pressbooks.pub/socialscienceresearch/chapter/chapter-9-survey-research/

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, 2(2), 1-9. https://doi.org/10.1016/j.patter.2021.100205

Buolamwini, J. & Gebru, T., (2018) *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. In S. Friedler & C. Wilson (Eds.). Proceedings of the 1st Conference on Fairness, Accountability and Transparency (pp. 1 - 15), Proceedings of Machine Learning Research http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Buolamwini, J. (2019, January 25). Response: Racial and Gender bias in Amazon Rekognition – Commercial AI System for Analyzing Faces. *Medium.* https://medium.com/@Joy.Buolamwini/response-racial-and-gender-bias-in-amazon-rekognition-commercial-ai-system-for-analyzing-faces-a289222eeced

Chadha, K. S. (2024). Bias and fairness in artificial intelligence: Methods and mitigation strategies. *International Journal for Research Publication and Seminars*, 15(3), 36–49. https://doi.org/10.36676/jrps.v15.i3.1425

Chen, Y., Clayton, E., Novak, L., Anders, S., & Malin, B. (2023). Human-centered design to address biases in artificial intelligence. *Journal of Medical Internet Research,* e43251. https://doi.org/10.2196/43251

Crenshaw, K (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination, *University of Chicago Legal Forum*, 1(8), 139-168. https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1052&context=uclf

Dai, K., Kim, J., , Džeroski, S., Wicker, J., Dobbie, G., & Dost, K. (2025). Assessing the risk of discriminatory bias in classification datasets. *Machine Learning*, 114(9), 1 - 119. https://doi.org/10.1007/s10994-025-06843-9

Dastin, J. (2018, October 11). Insight - Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT

Duong, M. K., & Conrad, S. (2024). Measuring and mitigating bias for tabular datasets with multiple protected attributes, arXiv preprint, https://arxiv.org/abs/2405.19300

Eidenskog, M., Andersson, R., & Glad, W. (2024). Workshops as a relational material research practice: Creating space for shared knowledge. *International Journal of Qualitative Methods*, 23, Article 16094069241297427. https://doi.org/10.1177/16094069241297427

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

EDRi. (2025, September 2023). Open Letter: The European Commission and Member States must keep AI Act national implementation on track. https://edri.org/our-work/open-letter-european-commission-member-states-keep-ai-act-national-implementation-on-track/

EDRi. (2021, November 30). Civil society calls on the EU to put fundamental rights first in the AI Act https://edri.org/our-work/civil-society-calls-on-the-eu-to-put-fundamental-rights-first-in-the-ai-act/

European Union Agency for Fundamental Rights. (2020, May 14). A long way to go for LGBTI equality. https://fra.europa.eu/en/publication/2020/eu-lgbti-survey-results

European Commission. (2023). Commission Recommendation (EU) 2023/2836 of 12 December 2023 on promoting the engagement and effective participation of citizens and civil society organisations in public policy-making processes. C/2023/8627. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32023H2836

Ferrara, E. (2024). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci,* 6(1), 1-15. https://doi.org/10.3390/sci6010003

Gebru, T. (2020). Race and Gender. In Dubber M. D., Pasquale F. and Das S.(Eds.), *The Oxford handbook of ethics of AI* (1st ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.001.0001

Gisselbaek, M., Berger-Estilita, J. M., Minsart, L., Saxena, S., et al. (2025). Gender disparities in artificial intelligence-generated images of hospital leadership in the United States. *Mayo Clinic Proceedings: Digital Health*, 3(2), 100218. https://doi.org/10.1016/j.mcpdig.2025.100218

Gohar, U., and Cheng L. (2023). *A survey on intersectional fairness in machine learning: Notions, mitigation, and challenges.* In E. Elkind (Ed.) Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, (pp. 6619–6627), International Joint Conferences on Artificial Intelligence https://www.ijcai.org/proceedings/2023/0742.pdf

González-Sendino, R., Serrano, E., and Bajo, J. (2024). Mitigating bias in artificial intelligence: Fair data generation via causal models for transparent and explainable decision-making. *Future Generation Computer Systems*, 155, 384–401. https://doi.org/10.1016/j.future.2024.02.023

Greenwood, J. (2007). Review article: Organized civil society and democratic legitimacy in the European Union. *British Journal of Political Science,* 37(2), 333–357. https://doi.org/10.1017/S0007123407000166

Hungarian Civil Liberties Union, Amnesty International, Hungarian Helsinki Committee, & Hatter Society. (2025, March 21). Legislating Fear: Banning Pride is the latest assault on fundamental rights in Hungary. https://tasz.hu/wp-content/uploads/2025/03/AIHU_Hatter_HCLU_HHC_Pride_03202025.pdf

INCLO. (2025). *Hard conversations for the future of human rights.* https://inclo.net/wp-content/uploads/2025/12/Hard-Conversations.pdf

Lindgren, S., & Dignum, V. (2023). Beyond AI solutionism: Toward a multi-disciplinary approach to artificial intelligence in society. In S. Lindgren (Ed.), *Handbook of critical studies of artificial intelligence* (pp. 163–172). Edward Elgar Publishing. https://doi.org/10.4337/9781803928562.00019

Mavrogiorgos, K., Kiourtis, A., Mavrogiorgou, A., Menychtas, A., & Kyriazis, D. (2024). Bias in Machine Learning: A Literature Review. *Applied Sciences*, 14(19), 8860. https://doi.org/10.3390/app14198860

Moliner, P & Lo Monaco G. (2017). *Méthodes d'associations verbales pour les sciences humaines et sociales*, Grenoble: PUG

Mosier, K. L., & Skitka, L. J. (1996). *Human decision makers and automated decision aids: Made for each other?* In *R. Parasuraman & M. Mouloua (Eds.), Automation and human performance: Theory and applications* (pp. 201–220). Lawrence Erlbaum.

Murphy, M. H. (2018). *Surveillance and the law: Language, power, and privacy* (2nd ed.). Routledge.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 230–253.

Romero, G. (2025, November 20). Civil society is at risk—and tech is part of the problem. *Tech Policy Press.* https://www.techpolicy.press/civil-society-is-at-riskand-tech-is-part-of-the-problem/

Samarin, N., Frik, A., Brooks, S., Cheshire, C., & Egelman, S. (2020). Surveying vulnerable populations: A case study of civil society organizations, arXiv preprint, https://arxiv.org/pdf/2003.08580v2.pdf

Smismans, S. (2003). European civil society: Shaped by discourses and institutional interests. *European Law Journal*, 9(4), 473–495.

Siapera, E., & Farries, E. (2025). Platform governance and civil society organisations: Tensions between reform and revolution continuum. *Internet Policy Review,* 14(1). https://doi.org/10.14763/2025.1.2002

superrrnetwork (2025, October 2) Zu queer für den Algorithmus? 🏳️‍🌈
Von Klicks und Klassen – wie der digitale Markt diskriminiert. Das Forum Digitale Fairness und. Instagram. https://www.instagram.com/p/DPTiTmeDLin/?

UCD Centre for Digital Policy. (2023, October 5). Ireland's experts draw a red line on Garda facial recognition tech. https://digitalpolicy.ie/ireland-experts-red-line-on-garda-facial-recognition-tech/

Funded by the European Union

Wilding, M., & Boffey, D. (2025, December 10). UK police forces lobbied to use biased facial recognition technology. *The Guardian*. https://www.theguardian.com/technology/2025/dec/10/police-facial-recognition-technology-bias

Wilson, K., & Caliskan, A. (2024). *Gender, race, and intersectional bias in resume screening via language model retrieval*. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES). arXiv preprint, https://arxiv.org/abs/2407.20371

Wilson, K., Sim, M., Gueorguieva, A.-M., & Caliskan, A. (2025). No thoughts, just AI: Biased LLM hiring recommendations alter human decision making and limit human autonomy. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society,* 8(3), 2692–2704. https://doi.org/10.1609/aies.v8i3.36749

WISERD. (2024, February 12). Civil society perspectives on AI in the EU. Wales Institute of Social and Economic Research and Data. https://wiserd.ac.uk/blog/civil-society-perspectives-on-ai-in-the-eu/

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, 10(3), JCMC1034. https://doi.org/10.1111/j.1083-6101.2005.tb00259.x

Xavier, B. (2025). Biases within AI: challenging the illusion of neutrality. *AI & Society*, 40(3), 1545–1546. https://doi.org/10.1007/s00146-024-01985-1

Yang, Y., Lin, M., Zhao, H., Peng, Y., Huang, F., & Lu, Z. (2024). A survey of recent methods for addressing AI fairness and bias in biomedicine. *Journal of Biomedical Informatics,* 154,104646. https://doi.org/10.1016/j.jbi.2024.104646

Yilmaz, S., Yorgancioglu, K., and Koyutürk, M. (2025) Bias-aware training and evaluation of link prediction algorithms in network biology, *PNAS*, 122 (24) e2416646122, https://doi.org/10.1073/pnas.2416646122

Zakharov v Russia [2015] EHCR 47143/06