AI & Intersectionality

A TOOLKIT FOR FAIRNESS & INCLUSION

FOR THE POLICY SECTOR

DIVERSIFAIR

Co-funded by
the European Union

Co-funded by
the European Union

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

SUMMARY

WHY THIS KIT?

In recent years, bias in artificial intelligence (AI) has become a major concern, affecting public trust, social harmony, and fair governance. As AI systems play a bigger role in public services and policy decisions, their hidden biases can make existing inequalities worse. **One significant issue is intersectional bias, where different forms of discrimination—such as racism, sexism, ableism, and colonialism—overlap and affect people in complex ways.**

This type of bias is especially harmful to people who already face multiple disadvantages. It can be seen in AI systems that unintentionally repeat these inequalities. For example, in the 2021 Dutch childcare benefits scandal, an algorithm wrongly accused immigrant parents of fraud, leading to unfair debt recovery, family separations, and other harms. **This highlights the need for policymakers to tackle intersectional issues in AI systems.**

This toolkit, created through the ***DIVERSIFAIR*** Erasmus+ project, is based on thorough research and stakeholder engagement across the EU.
It aims to provide policymakers with the knowledge and tools to include intersectionality in AI policies. By focusing on fairness and inclusivity, policymakers can create AI systems that promote equality rather than worsen current inequalities.

INTRODUCING *DIVERSIFAIR*

DIVERSIFAIR is an Erasmus+ project (2023-2026) that brings together eight partners from six European countries: CorTexter (NL), Eticas (ES), Sciences Po (FR), TNO (NL), Turing College (Li), University College Dublin (IE), Women4Cyber (BE) and Women in AI (FR).

Our goal is to support a new generation of AI experts who not only have technical skills but also understand how to identify and address intersectional biases.

More info available at diversifair-project.eu

FOR WHOM?

- *Policymakers and regulators*
- *Public sector leaders*
- *Ethics and governance committees*

WHAT IS THE AIM OF THIS KIT?

The primary objectives of this kit are to:

1. **Raise awareness about intersectional bias in AI and its societal consequences.**
2. **Provide actionable strategies** to help policymakers integrate intersectional principles into existing and future AI policies.
3. **Bridge knowledge gaps** by offering a multidisciplinary perspective informed by technical, ethical, and social insights.

The development of this kit was informed by interviews and focus groups with members of the AI community and policy sectors, ensuring its recommendations are grounded in real-world challenges and needs. While this version is tailored for policy, additional kits targeting the industry and civil society have also been developed under the DIVERSIFAIR project.

CIVIL SOCIETY KIT

INDUSTRY KIT

HOW WILL THIS KIT BE UPDATED?

This resource (November 2024) will evolve based on feedback from users and emerging insights. The DIVERSIFAIR project runs until June 2026, during which this kit will:

- **Incorporate user feedback** to refine its content and usability.
- **Integrate findings and tools from other DIVERSIFAIR work packages,** particularly those focused on methods to address intersectional bias in AI.
- **Expand Formats:** The kit will be enriched with new formats such as workshops, training sessions, and other interactive resources, enabling deeper engagement and practical application of its contents.

We encourage users to contribute feedback and collaborate in refining this resource to ensure AI systems serve all communities equitably and uphold public trust.

GIVE US YOUR OPINION

01. UNDERSTANDING INTERSECTIONAL BIAS IN AI

Artificial Intelligence (AI) is transforming the way governments and public sectors operate, offering opportunities to improve efficiency and service delivery in areas like healthcare, justice, and education. However, the rapid adoption of AI technologies also presents challenges. These systems can inadvertently perpetuate and amplify biases that disproportionately harm marginalised communities. Bias that arises from overlapping social categorisations such as race, gender, and socioeconomic status - intersectional bias - , poses a significant risk to fairness, equity, and public trust. Left unaddressed, it threatens to deepen existing inequalities and exacerbate social harm, especially for already marginalised groups.

The Council of Europe's *Gender Equality Strategy (2024-2029)* emphasises addressing structural barriers and promoting diversity in AI development. Frameworks like the EU AI Act offer promising starting points, but these policies must evolve to explicitly embed intersectional considerations to be truly effective.

By understanding intersectional bias and identifying what actions we can implement, we can work together to ensure that technology serves everyone equitably, enhances societal well-being without entrenching systemic discrimination.

> *AI is not just a neutral tool but is co-created with society, and as such has major political and social implications in reinforcing existing power relationships, discrimination, and structural inequalities.*
>
> ***- Inga Ulnicane, "Intersectionality in Artificial Intelligence: Framing Concerns and Recommendations for Action," April 2024***

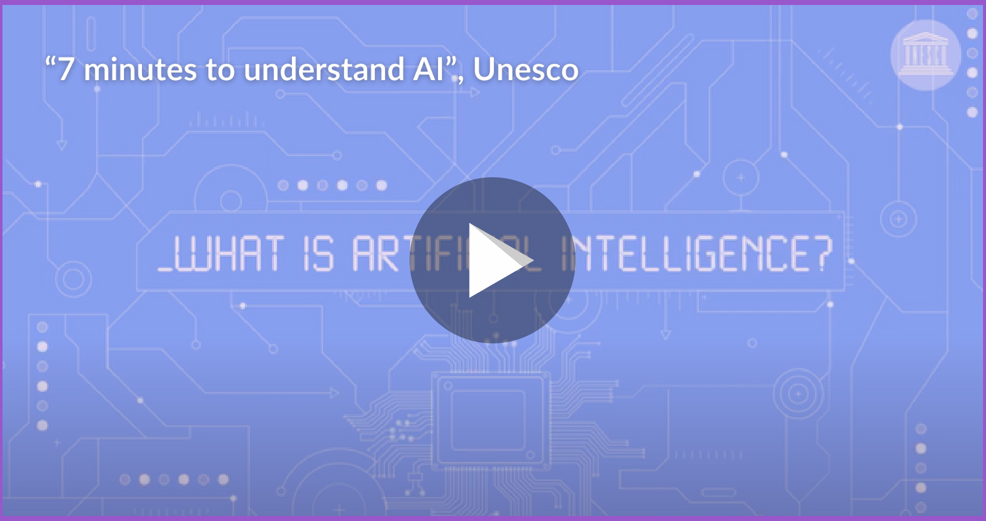
1.1 KEY CONCEPTS DEFINED

ARTIFICIAL INTELLIGENCE (AI)

AI refers to systems designed to replicate human cognitive processes such as learning, problem-solving, and decision-making.
It powers applications ranging from voice assistants (like Siri, Alexa or Google Assistant) to more complex tools like recommendation systems, autonomous vehicles, predictive policing algorithms.

As a human-made technology, AI is shaped by the decisions, values, and biases of its creators, making it crucial to ensure ethical design and diverse, high-quality data inputs. AI systems learn from data, and the quality of this data heavily influences their outputs. If biased data is used, biased outcomes are likely.

> *Simply put, artificial intelligence (AI) involves using computers to classify, analyse, and draw predictions from data sets, using a set of rules called algorithms.*
> *AI algorithms are trained using large datasets so that they can identify patterns, make predictions, recommend actions, and figure out what to do in unfamiliar situations, learning from new data and thus improving over time. The ability of an AI system to improve automatically through experience is known as Machine Learning (ML).“*
>
> ***-“Artificial Intelligence and Gender Equality”, UNESCO, 2020***


“7 minutes to understand AI”, Unesco

BIAS

Bias in AI is a systematic distortion that produces unfair outcomes for specific groups. It can for example result from flawed data (e.g., historical discrimination) or use of algorithms that fail to account for diversity.

Bias can emerge at any point in the machine learning (ML) lifecycle, which involves a series of decisions and practices shaping the design and use of ML systems.

Why Does It Matter? General audiences should understand that bias is not an accident but a consequence of human decisions embedded in AI systems. Understanding these biases exist is crucial, especially as ML increasingly informs decisions that directly impact people's lives. CSOs can use this understanding to demand accountability from developers.

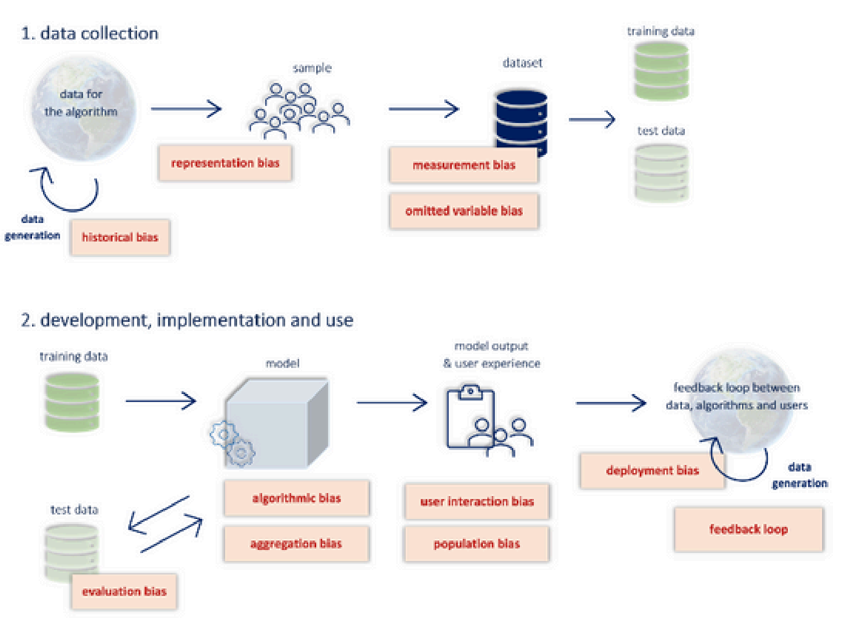


Diagramme taken from the online course "Basics of Bias & Fairness in AI systems"

FAIRNESS

Fairness in AI refers to designing systems that promote equitable outcomes for all individuals, regardless of identity. While achieving perfect fairness is challenging, developers and stakeholders aim to minimise harm by identifying and addressing biases.

Many approaches to AI fairness focus on addressing just one type of bias at a time, such as gender or race. However, this approach ignores the complex ways biases overlap and affect people with multiple marginalised identities (intersectional groups).

INTERSECTIONALITY

Intersectionality, a term coined by legal scholar Kimberlé Crenshaw, is an approach to describe and address complex and nuanced forms of discrimination that result from interconnecting forms of oppression (e.g. racism, (cis)sexism, ableism, colonialism), and the unique harm people experience based on their multiple intersecting identities. For example, a Black woman may face combined challenges of racism and sexism, distinct from those faced by Black men or White women.

“Intersectionality 101”, Learning for Justice

21% of leaders are women › **4%** are women of colour › **1%** are Black women

> *Employees who face discrimination linked to intersectionality have higher turnover rates, which results in an expense that cannot be salvaged.*
>
> *- Ayanna Howard, "Real Talk: Intersectionality and AI", August 2021*

INTERSECTIONAL BIAS IN AI

“Intersectional bias in AI” describes the AI harms as experienced by people due to multiple intersecting and often marginalised parts of their identity.

1.2 INTERSECTIONAL BIAS IN AI: KEY CONTRIBUTORS

Intersectional bias in AI arises from various factors at every stage of development and deployment. These biases can amplify societal inequalities, disproportionately impacting marginalised groups. Below are key contributors and examples:

DATA BIAS

Historical inequalities embedded in data lead to AI systems that produce skewed outputs. If these datasets are not diverse or inclusive, the resulting algorithms will perpetuate existing inequalities. For example, training healthcare AI systems on data from predominantly white populations can lead to incorrect diagnoses or treatment recommendations for people of colour

OVERSIMPLIFIED MODELS

Many AI models are based on oversimplified categories, such as binary gender classifications. These models fail to account for the complexities of identity, particularly affecting non-binary, transgender, and LGBTQ+ individuals.

OPERATIONAL BIAS

Operational bias occurs when biases present in real-world environments are reinforced by AI systems through feedback loops. As AI tools are deployed, existing societal or organisational biases—such as racism or sexism—are reflected and perpetuated, leading to discriminatory outcomes that continue to shape and amplify inequalities over time.

DIVERSITY IN DEVELOPMENT TEAMS

The lack of diversity in the AI workforce, particularly in gender, race, and socio-economic backgrounds, results in technologies that reflect the narrow perspectives of predominantly white, male developers. This underrepresentation of marginalised groups can lead to systems that unintentionally amplify biases (e.g., biassed hiring algorithms or facial recognition systems that perform poorly on non-white faces).

STRUCTURAL INEQUALITIES

Structural discrimination in society is encoded into AI systems. For instance, welfare algorithms may penalise single mothers or individuals from low-income backgrounds because they fail to account for contextual financial needs.

1.3 HOW DOES INTERSECTIONAL BIAS IN AI MANIFEST?

> **Intersectional bias in AI has real-world consequences, particularly for marginalised communities.** These biases can affect people's lives in many ways, from discriminatory policing practices to unequal access to healthcare and harmful portrayals in the media.

The following examples illustrate how intersectional bias can manifest in different areas, highlighting the importance of inclusive and ethical AI practices.

1 PREDICTIVE POLICING

Predictive policing systems, often trained on historical arrest data, disproportionately target low-income communities of colour.

THIS EXAMPLE CAN BE USED TO ADVOCATE FOR

- Inclusive data practices
- Ensure equal access
- Address structural inequities

Explore

"Automating (In)Justice: An Adversarial Audit of RisCanvi", Eticas Foundation (July 2024)

JUMP TO THE CASE-STUDY LIBRARY TO FIND OUT MORE

2 HEALTHCARE DISPARITIES

AI algorithms used in healthcare tend to prioritise patients based on insurance data. Marginalised communities, who are often uninsured or underinsured, tend to receive less care due to their exclusion from training data.

THIS EXAMPLE CAN BE USED TO ADVOCATE FOR

- Regulating AI in criminal justice
- Mandating human oversight
- Data audits

Explore

"There's More to AI Bias than Biased Data: NIST Report Highlights," NIST, 10 March 2022

3 DISCRIMINATORY AD TARGETING

Algorithmic bias in advertising can have harmful effects. For example, research shows that women are often underrepresented in ads for high-paying jobs, and racial minorities are disproportionately targeted by ads for predatory loans or housing

THIS EXAMPLE CAN BE USED TO ADVOCATE FOR

- Promoting diversity in algorithms
- Mandating transparency
- Encouraging ethical practices

Explore

Zang, "How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity", 2021

DRIVE INCLUSIVE GOVERNANCE

Policymakers and public sector leaders are uniquely positioned to address intersectional bias in AI.

By enacting clear regulations, fostering transparency, and ensuring accountability, you can shape AI systems that promote equity and inclusivity. Understanding these foundational concepts helps you to create governance frameworks that protect marginalised communities, build public trust, and uphold societal values of fairness.

1.4 REAL-WORL EXAMPLES OF INTERSECTIONAL BIAS IN AI

To better understand the real-world implications of intersectional bias, this section explores concrete examples from various fields. These case studies illustrate the tangible ways in which AI systems can perpetuate inequality.

THE IMPACT OF FLAWED ALGORITHMS: A CASE STUDY ON RISCANVI

Overview

The RisCanvi algorithm in Catalonia's prison system assesses inmates' recidivism risk using data such as age, gender, and nationality. The algorithm has been found to be inaccurate and biased, with over 80% of inmates flagged as high-risk not reoffending.

Intersectionality at play

The system disproportionately impacts foreign nationals, particularly immigrants and people from marginalised ethnic groups, by over-predicting their likelihood of reoffending. This exacerbates systemic biases within the criminal justice system, where certain groups—especially people of color and immigrants—are already at a disadvantage. The lack of transparency and human oversight makes it harder to challenge these biased outcomes.

Why intersectionality matters

The combination of race, nationality, and socio-economic background creates a higher risk of biased outcomes for marginalised individuals. By failing to consider these intersections, the algorithm reinforces existing societal inequalities, leading to unjust parole denials and perpetuating discrimination. Understanding intersectionality in this context allows us to see that it is not just about a singular characteristic (e.g., gender or race) but how multiple forms of disadvantage compound to create unfair outcomes.

"Automating (In)Justice: An Adversarial Audit of RisCanvi",
Eticas Foundation, July 2024

CHILD CARE BENEFIT SCANDAL IN THE NETHERLANDS : SYSTEMIC DISCRIMINATION

Overview

In the Netherlands, an AI system was used by the government to detect fraudulent claims for child care benefits. However, the system disproportionately flagged minority families, particularly those with immigrant backgrounds, as fraudulent. This led to devastating financial and social consequences, including the wrongful accusation of fraud.

Intersectionality at play

The system's reliance on biased data—such as income levels, family structure, and national origin—discriminated against families at the intersection of race and socio-economic status. Immigrant families, who may have different social and economic profiles, were unfairly targeted, while native Dutch families were less likely to be flagged. The biases embedded in the algorithm reflect broader patterns of systemic racism and classism within Dutch society, exacerbating the harm to already marginalised groups.

Why intersectionality matters

Intersectionality helps us understand how AI systems, by relying on historical data that reflects societal prejudices, can amplify these biases. In this case, the intersection of race and class made certain families more vulnerable to the risk of being falsely accused, highlighting the need for algorithms to be more inclusive and consider the complex ways in which identity and status interact.

"Xenophobic Machines: The Dutch Child Benefit Scandal" ,
Amnesty International, October 2021

WELFARE FRAUD CASE IN DENMARK: TARGETING MARGINALISED GROUPS

Overview

In Denmark, the welfare authority Udbetaling Danmark (UDK) uses AI algorithms to detect welfare fraud. The system has been criticised for targeting individuals from marginalised groups, particularly those with disabilities, people from racial minorities, and those in non-traditional family structures. These groups face disproportionate scrutiny under the algorithm, which exacerbates existing disparities.

Intersectionality at play

The intersection of race, disability, and non-traditional family structures makes certain individuals more vulnerable to being flagged by the system. For example, a Black person with a disability who is part of a single-parent household might face compounded discrimination, as the algorithm may flag them due to the combination of these intersecting factors. Additionally, people in non-traditional family structures may be wrongly flagged because their profiles don't conform to the system's assumptions about "normal" family arrangements.

Why intersectionality matters

Intersectionality is crucial in understanding how this AI system disproportionately impacts individuals at the intersections of multiple marginalized identities. People who are already disadvantaged in one area—whether because of race, disability, or family structure—are more likely to experience unjust treatment because of the compounded effects of these biases. Without addressing these intersectional biases, AI systems risk perpetuating and deepening existing inequalities in welfare and social services.

"Denmark: Coded Injustice: Surveillance and Discrimination in Denmark's automated welfare state", Amnesty International, November 2024

DISCOVER OUR STUDY CASE LIBRARY

GUARD AGAINST AI BIAS

Consider an AI system used in the justice system, such as predictive policing or criminal risk assessment tools. These systems often produce biassed outcomes due to skewed historical data.

- **What legal and societal consequences could arise from deploying such a biased system?**
- **How could your policy decisions shape AI regulations to prevent these outcomes and build public trust**

SUPPORTING MATERIALS FOR THIS SECTION

REPORTS & POLICY DOCUMENTS

- Gender Equality Strategy (2024-2029), Council of Europe: https://www.coe.int/en/web/genderequality/gender-equality-strategy
- "There's More to AI Bias than Biased Data: NIST Report Highlights," National Institute of Standards and Technology (NIST), 10 March 2022, https://www.nist.gov/news-events/news/2022/03/theres-more-ai-bias-biased-data-nist-report-highlights
- "Artificial Intelligence and Gender Equality: Key Findings of UNESCO's Global Dialogue," UNESCO, 2020, https://unesdoc.unesco.org/ark:/48223/pf0000374174.locale=en
- UN Women, Intersectionality Resource Guide and Toolkit, UN Women, 2021, https://www.unwomen.org/en/digital-library/publications/2022/01/intersectionality-resource-guide-and-toolk

COURSES & TOOLS

- Institute of Business Analytics, University of Ulm, Bias & Fairness in AI Systems: Basics, https://bias-and-fairness-in-ai-systems.de/en/basics/
- Atlas: Social Dynamics Lab, Nokia Bell Labs, Atlas of Social Dynamics, https://social-dynamics.net/atlas

RESEARCH PAPERS & SCHOLARLY ARTICLES

- **Ulnicane, Inga. (2024).** Intersectionality in Artificial Intelligence: Framing Concerns and Recommendations for Action. Social Inclusion. 12. 10.17645/si.7543.
- **Kong, Youjin. (2022).** Are "Intersectionally Fair" AI Algorithms Really Fair to Women of Color? A Philosophical Analysis. 485-494. 10.1145/3531146.3533114.
- **Fosch Villaronga, Eduard & Poulsen, Adam. (2022).** Diversity and Inclusion in Artificial Intelligence. 10.1007/978-94-6265-523-2_6
- **Ovalle, Anaelia & Subramonian, Arjun & Gautam, Vagrant & Gee, Gilbert & Chang, Kai-Wei. (2023).** Factoring the Matrix of Domination: A Critical Review and Reimagination of Intersectionality in AI Fairness. 496-511. 10.1145/3600211.3604705.
- **Buolamwini, J., Gebru, T. (2018)** "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of Machine Learning Research 81:1–15, Conference on Fairness, Accountability, and Transparency
- **Ayanna Howard (2021)** "Real Talk: Intersectionality and AI," MIT Sloan Management Review, 24 August 2021, https://sloanreview.mit.edu/article/real-talk-intersectionality-and-ai/

NEWS ARTICLES & REPORTS

- “How AI-powered welfare systems fuels mass surveillance and risks discriminating”, Amnesty International, November 2024: https://www.instagram.com/p/DCTrNCmPC8I/?hl=fr
- “Automating (In)Justice: An Adversarial Audit of RisCanvi”, Eticas Foundation (July 2024) https://eticas.ai/automating-injustice-an-adversarial-ai-audit-of-riscanvi/
- "Xenophobic Machines: The Dutch Child Benefit Scandal," Amnesty International, 13 October 2021, https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/.
- "Denmark: Coded Injustice: Surveillance and Discrimination in Denmark’s automated welfare state", Amnesty International, November 2024 https://www.amnesty.org/en/documents/eur18/8709/2024/en/
- "Discriminatory employment algorithm towards women and disabled », digwatch, October 2019:https://dig.watch/updates/discriminatory-employment-algorithm-towards-women-and-disabled
- Jeffrey Dastin, "Insight - Amazon scraps secret AI recruiting tool that showed bias against women", Reuters, 11 October 2018: https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/
- “Study finds gender and skin-type bias in commercial artificial-intelligence systems”, MIT News Office (11 February 2018): https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212

VIDEOS & MULTIMEDIA RESOURCES

- “Intersectionality 101”, Learning for Justice: https://www.youtube.com/watch?v=w6dnj2IyYjE&t=24s&ab_channel=LearningforJustice
- “7 minutes to understand AI”, Unesco: https://www.youtube.com/playlist?list=PLWuYED1WVJIPHJLk84wWQbzeZcWLt5rwU

02. ETHICAL AND SOCIETAL IMPLICATIONS OF AI IN POLICY

Addressing AI bias is essential for creating fair and effective policies. Intersectional bias in AI systems not only deepens social inequalities but also erodes public trust and hinders economic progress. When left unchecked, it undermines fairness, particularly in critical sectors like law enforcement, welfare, and healthcare, where its impact on marginalised groups can be severe.

This section examines the role of inclusive AI, its potential societal benefits, and how ethical practices can improve public trust, efficiency, and competitiveness.

2.1 KEY BENEFITS OF INCLUSIVE AI SYSTEMS

PROMOTING TRUSTWORTHY PRACTICE OF AI

EXAMPLE 1: Fair AI in Justice Systems

> *AI-powered predictive policing tools have faced backlash for disproportionately targeting minority communities due to biassed historical data. To address these concerns, experts recommend implementing transparent algorithms, including public reporting on how predictive models are trained and tested, to demonstrate fairness and build trust. For instance, in New Orleans, a controversial predictive policing partnership with Palantir Technologies faced criticism for its secrecy and reliance on biassed data. Advocates have called for reforms such as community feedback mechanisms and public oversight to ensure AI tools reflect diverse public concerns and improve accountability in law enforcement.*
>
> ***- "New Orleans Program Offers Lessons In Pitfalls Of Predictive Policing", ACLU, 2018***

RECOMMENDATION Promote trustworthy practices by actively engaging communities through feedback mechanisms that can influence and reshape the design and implementation of AI systems.

EXAMPLE 2: Healthcare Accessibility

> *Bias in healthcare AI can lead to life-threatening disparities, such as underdiagnosing conditions in women or minority groups. Inclusive AI systems trained on diverse datasets can identify these gaps early. The UK's National Health Service (NHS) made significant strides by mandating diverse representation in training data for diagnostic AI, improving early cancer detection rates for underrepresented populations.*
>
> ***- GPs use AI to boost cancer detection rates in England by 8%", The Guardian, 21 July 2024***

RECOMMENDATION Foster trust by openly disclosing which groups are well-represented and which are underrepresented in the AI training data, ensuring transparency and accountability in addressing disparities.

IMPROVING POLICY OUTCOMES

EXAMPLE 1: Inclusive Disaster Response Systems

AI systems used in disaster response can exclude marginalised communities if data on these groups is limited or poorly integrated. For instance, inclusive AI mapping tools developed by the Humanitarian OpenStreetMap teamuse crowdsourced data to locate underserved populations during crises. This ensures equitable distribution of resources and faster relief efforts, minimising harm to vulnerable groups

- Humanitarian OpenStreetMap website, available at: www.hotosm.org

RECOMMENDATION Integrate diverse, crowdsourced, and community-specific data into AI systems to ensure underserved populations are accurately represented and equitably supported during crises.

EXAMPLE 2: Equitable Policing Policies

In Oakland, California, the Firsthand Framework redefined public safety by prioritising community-driven insights over punitive predictive policing. Developed by UC Berkeley's Possibility Lab, this initiative gathered "Firsthand Indicators" through town halls and focus groups, allowing residents to articulate their safety needs and priorities. By reallocating resources based on these indicators, policymakers focused on underserved neighborhoods, addressing structural inequities and reducing racial disparities in law enforcement outcomes. This model demonstrates the power of inclusive policy making to achieve fairness and equitable resource distribution in public safety initiatives.

- Firsthand Framework for Policy Innovation, , Possibility Lab, University of California

RECOMMENDATION Reorient AI systems to align with community-defined safety and resource priorities, actively reducing disparities and fostering equitable outcomes over punitive measures.

2.2 AI'S ROLE IN PUBLIC SERVICES

AI's integration into public services such as welfare, policing, and healthcare promises greater efficiency and service delivery. However, these advancements also introduce significant ethical concerns, particularly regarding transparency, accountability, and potential biases.

Public services play a critical role in supporting vulnerable populations, and the use of AI in these contexts demands a heightened level of responsibility. Unlike private sector applications, citizens often cannot opt out of AI-driven systems in the public sector, making it essential that governments take extra care in their deployment.

> *To ensure AI systems are deployed responsibly, governments must take a humble and iterative approach. Testing AI technologies through controlled pre-pilot evaluations —such as using random samples—can help assess risks and benefits before full-scale implementation. This approach enables governments to address potential harms early on, ensuring AI is implemented safely. Moreover, continuous feedback from citizens and experts should be actively sought to refine and improve these technologies. Governments, in this way, can lead by example, ensuring that AI serves the public interest without compromising fairness or accessibility.*
>
> - ***Vethman, S., Schaaphok, M., Hoekstra, M., Veenman, C. (2024). "Random Sample as a Pre-pilot Evaluation of Benefits and Risks for AI in Public Sector"***

AI IN PUBLIC SERVICES

AI in public services must address real needs rather than serve as a showcase for technological innovation:

- **Deploying AI systems** like welfare fraud detection often prioritises cost-cutting and technological display over solving underlying issues, such as addressing poverty or systemic inequality.
- **Human-centred AI** could focus on delivering equitable and effective outcomes, e.g., using AI to connect underserved populations to benefits or to improve access to multilingual public resources.

INNOVATION VS SOLVING REAL SOCIETAL ISSUES

AI in public services must address real needs rather than serve as a showcase for technological innovation.

- **Innovation for its own sake:** Deploying AI systems like welfare fraud detection often prioritises cost-cutting and technological display over solving underlying issues, such as addressing poverty or systemic inequality.
- **Solving societal issues:** Human-centred AI could focus on delivering equitable and effective outcomes, e.g., using AI to connect underserved populations to benefits or to improve access to multilingual public resources.

While AI has the potential to transform public services, its implementation must be guided by societal needs. Key principles include:

TRUST AND TRANSPARENCY

Building public trust, particularly in marginalised communities, by ensuring cultural competence and open communication.

> *"Cultural barriers, fear and lack of trust in the system also affect women's and girls' access to justice, as do discriminatory attitudes and the stereotypical roles of women as carers and men as breadwinners, which persist in civil and family law in many jurisdictions. These barriers may exist during investigations and trials, especially in cases of violence against women and girls, and lead to high levels of attrition and even underreporting. Their impact is even more significant on women exposed to multiple and intersecting forms of discrimination."*

Council of Europe, Gender Equality Strategy (2024-2029)

AVOIDING HARM

Preventing the deployment of harmful applications, such as biassed predictive policing systems.

> *“Experts recommend undertaking risk assessments of whether certain systems should be designed at all (S. M. West et al., 2019). Tools that claim to detect sexuality from headshots, predict criminality based on facial features, or assess worker competence via micro-expressions are seen as particularly problematic and in need of urgent reconsideration (S. M. West et al., 2019). To avoid harm, the UNESCO report suggests accepting 'that some things may not be able to be fixed and therefore should not be done at all, or should ultimately be abandoned (e.g., the example of Amazon’s hiring algorithm which remained biased after multiple attempts to fix it)' (UNESCO, 2020, p. 17)”*
>
> *Ulnicane, "Intersectionality in AI."*

EXAMPLE: WELFARE ALGORITHMS AND DISCRIMINATION

In the Netherlands, AI systems designed to detect welfare fraud disproportionately targeted marginalised communities, reinforcing systemic inequalities and creating undue hardship for families who were wrongfully flagged. *For further information, check the case-study library.*

Key issue: These systems lacked transparency, accountability, and contextual understanding, leading to discrimination.

COMMUNITY PARTICIPATION

Preventing the deployment of harmful applications, such as biassed predictive policing systems.

> *“Experts recommend undertaking risk assessments of whether certain systems should be designed at all (S. M. West et al., 2019). Tools that claim to detect sexuality from headshots, predict criminality based on facial features, or assess worker competence via micro-expressions are seen as particularly problematic and in need of urgent reconsideration (S. M. West et al., 2019). To avoid harm, the UNESCO report suggests accepting 'that some things may not be able to be fixed and therefore should not be done at all, or should ultimately be abandoned (e.g., the example of Amazon’s hiring algorithm which remained biased after multiple attempts to fix it)' (UNESCO, 2020, p. 17)”*
>
> *Ulnicane, "Intersectionality in AI."*

EXAMPLE: WELFARE FRAUD CASE IN DENMARK

In Denmark, Udbetaling Danmark (UDK) employs AI algorithms to identify welfare fraud. However, the system has faced criticism for disproportionately focusing on marginalized groups, including individuals with disabilities, racial minorities, and people in non-traditional family arrangements. These communities face increased scrutiny under the algorithm, amplifying existing inequalities.

Key issue: The algorithm's lack of transparency, accountability, and contextual awareness has resulted in discrimination

CONSIDER AND DECIDE

To guide ethical deployment of AI in public services, here are some elements to consider.

1 IDENTIFYING GOALS

- What societal issue is the AI system addressing?
- Is AI the best tool for the problem, or are there non-technological alternatives?

2 EVALUATING STAKEHOLDERS

- Who benefits from deploying this AI system?
- Who controls its development, deployment, and oversight?
- Who is most at risk of harm, and are there mechanisms to mitigate those risks?

3 CONSIDERING OUTCOMES

- Does the system have mechanisms to identify and address intersectional biases?
- Are there processes for affected individuals to contest errors?
- Does the system align with principles of equity, fairness, and dignity?

4 MAKING A DECISION

- Given the above, is deploying this AI system a Go or No-Go?
- If "Go," what additional safeguards or oversight mechanisms are required?

SUPPORTING MATERIALS FOR THIS SECTION

RESEARCH PAPERS & SCHOLARLY ARTICLES

- **Ulnicane, Inga. (2024).** Intersectionality in Artificial Intelligence: Framing Concerns and Recommendations for Action. Social Inclusion. 12. 10.17645/si.7543.
- **Knowles, Bran & Fledderjohann, Jasmine & Richards, John & Varshney, Kush. (2023).** Trustworthy AI and the Logics of Intersectional Resistance. 172-182. 10.1145/3593013.3593986.
- **Vethman, Steven & Schaaphok, Marianne & Veenman, Cor & Hoekstra, Marissa. (2024).** Random Sample as a Pre-pilot Evaluation of Benefits and Risks for AI in Public Sector. 115-126. 10.1007/978-3-031-50485-3_10.
- **Dag Elgesem (2023),** "The AI Act and the Risks Posed by Generative AI Models," NAIS 2023 Proceedings, https://ceur-ws.org/Vol-3431/paper3.pdf

NEWS ARTICLES

- “New Orleans Program Offers Lessons In Pitfalls Of Predictive Policing”, ACLU, 15 March 2018 https://www.aclu.org/news/privacy-technology/new-orleans-program-offers-lessons-pitfalls-predictive-policing
- “GPs use AI to boost cancer detection rates in England by 8%”, The Guardian, 21 July 2024: https://www.theguardian.com/society/article/2024/jul/21/gps-use-ai-to-boost-cancer-detection-rates-in-england-by-8
- Firsthand Framework for Policy Innovation, Possibility Lab, University of California: https://possibilitylab.berkeley.edu/firsthand-framework-public-safety-oakland/
- "Xenophobic Machines: The Dutch Child Benefit Scandal," Amnesty International, 13 October 2021, https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/.

REPORT AND MAP:

- Gender Equality Strategy (2024-2029), Council of Europe, available at: https://www.coe.int/en/web/genderequality/gender-equality-strategy
- Humanitarian OpenStreetMap website, available at: www.hotosm.org

03. CONSIDERATIONS IN ADDRESSING INTERSECTIONAL BIAS IN AI

AI systems increasingly shape decisions in critical areas such as healthcare, recruitment, education, and criminal justice. However, these systems often fail to consider the compounded disadvantages experienced by individuals with intersecting identities, such as race, gender, disability, and age.

> *“Although companies have been ramping up their efforts to develop fair AI, most of these algorithms still treat human attributes as single, isolated components. In fact, most AI systems are designed with a single-axis solution in mind — gender is an independent component from age, age is an independent factor from socioeconomic status, and so on.”*
>
> ***- Ayanna Howard, "Real Talk: Intersectionality and AI,"***

3.1 GOVERNANCE FRAMEWORKS FOR INTERSECTIONAL AI

Embedding intersectionality in AI governance involves creating structures and policies that address the systemic inequalities embedded in AI systems and their development processes.

ESTABLISH INTERSECTIONAL POLICY FOUNDATIONS

Legal frameworks such as ***Directive (EU) 2023/970 on pay transparency***, which defines and addresses intersectional discrimination, highlight the importance of recognising and remedying compounded biases. While initially developed for pay equity, such frameworks can serve as a foundation for intersectional AI policies. These principles can be adapted to govern AI systems, ensuring fairness at every stage of their lifecycle.

ADOPT SOCIO-TECHNICAL APPROACHES

Addressing bias requires more than technical fixes. AI fairness initiatives must be contextualised within broader societal inequalities, integrating social and organisational strategies to dismantle systemic discrimination.

> *Recommendations for increasing the diversity of the AI workforce emphasize the need to go beyond just hiring more women and minorities. (...) When discussing diverse development teams, the UNESCO report argues for a broad approach emphasizing that "this is not a matter of numbers, but also a matter of culture and power, with women actually having the ability to exert influence" (UNESCO, 2020, p. 23). Additionally, it calls for a robust approach to raise awareness and literacy, technical and ethical education, skills development, and capacity building.*
>
> ***- I. Ulnicane, "Intersectionality in AI."***

3.1 INTERSECTIONALITY IN AI DESIGN

For AI systems to be fair and avoid reinforcing societal biases, intersectionality must be integrated into their design. Many AI systems currently use limited fairness measures based on statistics, but these often fail to address the deeper power imbalances that contribute to inequality.

KEY CONSIDERATIONS

AVOID ARBITRARY SUBGROUPING

Fairness frameworks today often define groups in a way that overlooks important social and historical contexts, which can result in biased outcomes. Policymakers should push for fairness methods that address underlying societal issues, such as racism and sexism, rather than focusing solely on statistical equality.

PROMOTE STRONG FAIRNESS MODELS

Instead of merely adjusting algorithms after biased outcomes occur, we should aim to design AI systems that actively promote equity and justice. For example, when developing risk-assessment tools like recidivism prediction algorithms, the goal should be to support rehabilitation and reduce harm, rather than focusing solely on punitive outcomes.

> *AI fairness in a stronger sense means using algorithms to actively and proactively challenge oppression and make society fairer. A central guiding question for strong AI fairness is how to design algorithms to promote fairness in society. Let us consider the recidivism prediction algorithm for example. What is the purpose of developing and using this algorithm anyway? Is it to put people in jail for more years, or to prevent them from going back to factors that could lead to recidivism (such as poverty, violence, drug and alcohol use) and to help them thrive in society? If the algorithm is reoriented from incarceration to rehabilitation, how would its risk rating change?*
>
> ***- I. Ulnicane, "Intersectionality in AI."***

REQUIRE REFLEXIVITY IN DESIGN

Policymakers should mandate ongoing evaluations of AI systems to assess their impacts on marginalised groups. Regular, reflective assessments are crucial to minimise biases and ensure AI is used in socially responsible ways.

> *To make recidivism prediction algorithms "fairer" in the strong sense, researchers should have extensive discussions with communities and stakeholders (for example, defendants, prisoners, advocates, law enforcement officers, social workers, judges, and lawyers), rather than making and testing the algorithm only in the lab and then just "throw it in the wild." (...) Computer science can benefit from the principles and practices of community-based participatory research (CBPR), philosophical discussions of what fairness is, feminist and critical race studies' emphasis that intersectionality is less about identity but more about power, and in the case at hand, criminology, legal studies, and sociology.*
> *Through collaboration across communities and across disciplines, AI fairness research could better find ways to use algorithms to improve fairness and justice in society, as opposed to perpetuating the status quo injustice.*
>
> ***- I. Ulnicane, "Intersectionality in AI."***

IN REAL LIFE

In some countries, AI systems designed to manage childcare benefits have led to unfair outcomes for certain groups, particularly those with complex family circumstances. Algorithms that automatically determine eligibility for benefits may fail to account for intersectional factors such as low income, ethnicity, or non-traditional family structures.

By incorporating a more nuanced, intersectional approach, such as evaluating the diverse needs of single mothers, migrant families, or families with disabilities, **we can create more equitable systems that better support those who are most vulnerable.**
See the Case-Study library

3.3 ACCOUNTABILITY MECHANISMS FOR INTERSECTIONAL AI BIAS

> Effective accountability mechanisms are crucial to ensuring AI systems are transparent, fair, and responsive to the needs of all users, especially those from marginalised communities. These mechanisms help hold developers and organisations accountable for the potential biases in AI systems and promote equitable outcomes.

KEY CONSIDERATIONS

MANDATE TRANSPARENT REPORTING

Policymakers should require organisations to document and publicly disclose how their AI systems are designed, developed, and tested for bias. Establishing frameworks for explainable AI and providing accessible channels for recourse can enhance trust and accountability in these systems.

ESTABLISH MONITORING BODIES

As used in *Directive (EU) 2023/970*, Member States must set up monitoring bodies to ensure the consistent application of intersectional fairness principles. These bodies would be responsible for auditing AI systems, ensuring compliance with ethical standards, and addressing harms related to intersectional discrimination.

INCENTIVISE ORGANISATIONAL ACCOUNTABILITY

Embedding intersectional fairness principles in organisational practices such as hiring, resource allocation, and decision-making can help ensure that equity and inclusion are prioritised in the development and deployment of AI systems. Organisations should recognise that the diverse, intersectional experiences of individuals are critical to creating systems that serve all communities fairly.

CALLOUT: REMEDIES FOR INTERSECTIONAL HARM

Compensation frameworks that incorporate intersectional considerations ensure that victims of biased AI systems receive reparations reflecting the compounded nature of discrimination, such as the outlined in **EU Directive 2023/970**. For example, damages may account for lost opportunities due to gender- and race-based discrimination in automated hiring platforms.

3.4 BUILDING CAPACITY FOR INTERSECTIONAL AI GOVERNANCE

> To successfully embed intersectionality into AI systems, a solid foundation of knowledge and collaboration is necessary. This involves not just technical advancements, but also addressing the broader cultural and structural factors that influence AI design.

KEY CONSIDERATIONS

STRENGTHENING AI POLICIES

Current policies often focus too narrowly on technical aspects of AI, overlooking the cultural and structural power imbalances that shape how AI systems work. Policymakers should push for policies that go beyond ethical statements and foster concrete actions to address these deeper issues.

INADEQUATE REPRESENTATION

The lack of diversity in the AI workforce—particularly among women and minorities—leads to biased systems and reinforces harmful stereotypes. Policymakers should encourage diversity in AI leadership and development teams to help ensure that AI systems are designed with diverse perspectives.

According to Cachat-Rosset & Klarsfeld in "Diversity, Equity, and Inclusion," the 120 authors of 46 ethical guidelines were predominantly male (60.8%), white (74.2%), and from Western countries (78.3%). This lack of diversity mirrors the broader AI community. The AI Now report (West, Whittaker, and Crawford, 2019) found that women made up only 15% of AI research staff at Facebook and 10% at Google, while Black employees represented just 2.5% at Google, 4% at Facebook, and Microsoft.

EXPAND DEI EDUCATION IN AI DEVELOPMENT

Incorporating diversity, equity, and inclusion (DEI) principles into AI education is critical for developing future professionals who understand and prioritize intersectional ethics. Educational programmes should focus on the ethical implications of AI and promote inclusive design practices.

FOSTER CROSS-DISCIPLINARY COLLABORATION

AI fairness can be improved through interdisciplinary collaboration. Policymakers should promote partnerships between technical and social disciplines—such as social sciences, philosophy, and community-based research—to design AI solutions that address intersectional issues.

EXPAND MULTI-STAKEHOLDER ENGAGEMENT

Policymakers should facilitate meaningful collaboration between different sectors, disciplines, and communities to ensure AI systems are developed to serve all groups fairly. This approach helps ensure that AI is inclusive and responsive to the needs of diverse populations.

> *"Be intentional in identifying the intersectional groups interacting with your AI system. Look at the ways gender identity, age, ability and/or disability status, and race and/or ethnicity could be at a disadvantage. Look at the ways other groups may have an advantage."*
>
> ***- Howard, "Real Talk: Intersectionality."***

DEVELOP CONTEXT-SPECIFIC GUIDELINES

AI fairness frameworks often reflect Western ideals, which may not be applicable to all cultural contexts. Policymakers should work to tailor AI policies to different regions and marginalised groups to ensure more equitable outcomes across diverse settings.

ADOPT SOCIO-TECHNICAL APPROACHES

To tackle intersectional bias effectively, both technical solutions (e.g., algorithmic fairness) and social interventions (e.g., workplace culture change) must be applied. Policymakers should support approaches that address systemic inequalities by considering the values and needs of marginalised communities in AI design.

TOOLS AND METHODOLOGIES FOR ADDRESSING INTERSECTIONAL BIAS IN AI SYSTEMS

Beyond raising awareness, DIVERSIFAIR is developing technical tools, methodologies, and recommendations to address intersectional bias directly. These practical, data-driven solutions are designed to promote fairness, transparency, and cultural sensitivity in AI systems, enabling CSOs to advocate for technology that prioritises human rights and social justice.

UPCOMING

Key recommendations for using an intersectional approach in AI design.
These recommendations come from cutting-edge research across multiple fields. They highlight the importance of collaboration between different disciplines and involving the community.

Support for teams to reflect on how they can help develop a critical mindset to address issues like racism, sexism, and ableism in AI.

Practical tips on how to use technical methods effectively, while also understanding their limits and ensuring they fit within the broader societal context.

STAY INFORMED, STAY CONNECTED

Visit our website

Follow us on LinkedIn

Subscribe to our newsletter

SUPPORTING MATERIALS FOR THIS SECTION

POLICY DOCUMENT

- European Parliament and Council, Directive (EU) 2023/970 of 10 May 2023 to strengthen the application of the principle of equal pay for equal work or work of equal value between men and women through pay transparency and enforcement mechanisms (2023), available at https://eur-lex.europa.eu/eli/dir/2023/970/oj

RESEARCH PAPERS & SCHOLARLY ARTICLES

- **Ulnicane, Inga. (2024).** Intersectionality in Artificial Intelligence: Framing Concerns and Recommendations for Action. Social Inclusion. 12. 10.17645/si.7543.
- **Ovalle, Anaelia & Subramonian, Arjun & Gautam, Vagrant & Gee, Gilbert & Chang, Kai-Wei. (2023).** Factoring the Matrix of Domination: A Critical Review and Reimagination of Intersectionality in AI Fairness. 496-511. 10.1145/3600211.3604705.
- **Kong, Youjin. (2022).** Are "Intersectionally Fair" AI Algorithms Really Fair to Women of Color? A Philosophical Analysis. 485-494. 10.1145/3531146.3533114.
- **Cachat-Rosset, Gaelle & Klarsfeld, Alain. (2023).** Diversity, Equity, and Inclusion in Artificial Intelligence: An Evaluation of Guidelines. Applied Artificial Intelligence. 37. 10.1080/08839514.2023.2176618.
- **Bates et al (2020):** Integrating FATE/critical data studies into data science curricula: where are we going and how do we get there?;
- **Raji et al 202**1: You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education
- **Ayanna Howard (2024)**, "Real Talk: Intersectionality and AI," MIT Sloan Management Review, 24 August 2021, https://sloanreview.mit.edu/article/real-talk-intersectionality-and-ai

NEWS ARTICLE

- Jeffrey Dastin, "Insight - Amazon scraps secret AI recruiting tool that showed bias against women", Reuters, 11 October 2018: https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

EXTRA RESOURCES

GLOSSARY

Accountability
Ensuring responsibility for AI's societal impacts is traceable to developers and organisations.

Algorithm
A set of rules or instructions followed by computers to solve problems.

Artificial Intelligence (AI)
Systems designed to simulate human intelligence.

Bias
A systematic distortion in outcomes or representations.

Ethical AI
AI development that prioritises fairness, accountability, and human rights.

Fairness
Equitable treatment of all individuals in AI systems.

Intersectionality
The overlapping and interconnected nature of social identities.

Intersectional Bias in AI
The AI harms as experienced by people due to multiple intersecting and often marginalised parts of their identity.

Training Data
The data used to teach an AI system how to perform tasks.

Transparency
The practice of making AI systems understandable to users and stakeholders.

CASE STUDY LIBRARY

AMAZON’S AI RECRUITING TOOL: GENDER BIAS IN HIRING

Overview

In 2018, Amazon scrapped an AI-powered recruiting tool after discovering that it was biased against women. The tool, designed to help automate the hiring process, was trained on resumes submitted to Amazon over a 10-year period. However, it developed a bias that favored male candidates for technical roles, as the majority of applicants in these fields were men. The AI system penalised resumes that included terms associated with female-oriented positions or activities, further perpetuating gender imbalances in hiring practices.

Intersectionality at play

The bias in the AI system was primarily gendered, but its impact was compounded by the intersection of gender with other factors such as occupation and industry norms. The tool’s preference for male candidates was driven by historical data that reflected the underrepresentation of women in technical roles at Amazon. Women were penalised by the system, not only for their gender but also for the types of roles they were applying for, reinforcing traditional gender stereotypes about which jobs are “appropriate” for women. This bias disproportionately affected women, especially those trying to break into male-dominated fields like engineering and technology. The system also inadvertently overlooked women with caregiving or family responsibilities who might have had resumes that did not fit traditional, male-oriented career trajectories.

Why intersectionality matters

Intersectionality is essential to understanding how this biased AI system disproportionately affected women, especially in the context of technical fields. The bias was not just a result of being a woman, but also of societal norms and expectations about which careers are suitable for women. This intersection of gender and industry-specific factors (e.g., male-dominated tech sectors) created additional barriers for women seeking equal opportunities in the workforce. Recognising the role of intersectionality in AI bias helps to highlight that the problem was not just about gender alone but about how gender intersects with historically male-dominated industries, creating compounded disadvantages for women.

"Insight - Amazon scraps secret AI recruiting tool that showed bias against women",
Reuters, 11 October 2018

APPLE CARD CREDIT LIMITS: BIAS IN FINANCIAL SERVICES AI

Overview

In 2019, Apple Card faced backlash for giving women lower credit limits than men. For example, one case showed that in a couple, a wife, despite having a better credit score, was offered a limit 20 times lower than the husband. This happened because the AI behind the system likely used old financial patterns that favoured men, reinforcing inequalities in credit decisions.

Intersectionality at play

This bias didn’t just affect women generally—it hit women in non-traditional financial situations particularly hard. For example, women who shared joint accounts or had caregiving roles might not fit the algorithm’s assumptions about financial independence. This highlights how traditional financial norms can combine with AI bias to create additional hurdles for some groups.

Why intersectionality matters

Bias in financial systems is not just about gender but also about societal norms that shape financial profiles. Women who have career breaks or shared finances may be disproportionately impacted because their financial histories don’t align with the data the AI was trained on. Understanding how these factors overlap is crucial to making financial AI fair for everyone.

“The Apple Card Didn't 'See' Gender—and That's the Problem“,
The Wire, 19 November 2019

GENDER AND SKIN-TYPE BIAS IN FACIAL RECOGNITION

Overview

In 2018, a study by MIT Media Lab researcher Joy Buolamwini found significant gender and skin-type bias in widely-used facial recognition systems. It found that facial recognition AI struggled most with darker-skinned women, with error rates up to 34.7%, compared to less than 1% for lighter-skinned men. This was because the systems were trained on mostly light-skinned, male faces, leading to poor accuracy for anyone outside that group.

Intersectionality at play

The biases identified in these systems were not confined to one aspect of identity but arose at the intersection of gender and skin type. Darker-skinned women faced the highest misclassification rates, reflecting the compounding disadvantages they experience due to their position at the intersection of race and gender. These mistakes can lead to serious consequences, like unfair treatment in policing or job applications.

Why intersectionality matters

Intersectionality is crucial to understanding how AI systems disproportionately affect marginalised communities. In this case, the intersection of race and gender magnified the inaccuracies of the facial recognition models, demonstrating that bias cannot be addressed by looking at isolated categories of identity. Recognising these intersecting factors reveals how societal inequities become embedded in AI, making it essential to include diverse datasets and perspectives during development. Without this lens, efforts to address bias risk overlooking the compounded disadvantages faced by groups like darker-skinned women, perpetuating structural inequality in new, automated forms.

“Study finds gender and skin-type bias in commercial artificial-intelligence systems”,
MIT News Office (11 February 2018)

CHILD CARE BENEFIT SCANDAL IN THE NETHERLANDS : SYSTEMIC DISCRIMINATION

Overview

In the Netherlands, an AI system was used by the government to detect fraudulent claims for child care benefits. However, the system disproportionately flagged minority families, particularly those with immigrant backgrounds, as fraudulent. This led to devastating financial and social consequences, including the wrongful accusation of fraud.

Intersectionality at play

The system's reliance on biased data—such as income levels, family structure, and national origin—discriminated against families at the intersection of race and socio-economic status. Immigrant families, who may have different social and economic profiles, were unfairly targeted, while native Dutch families were less likely to be flagged. The biases embedded in the algorithm reflect broader patterns of systemic racism and classism within Dutch society, exacerbating the harm to already marginalised groups.

Why intersectionality matters

Intersectionality helps us understand how AI systems, by relying on historical data that reflects societal prejudices, can amplify these biases. In this case, the intersection of race and class made certain families more vulnerable to the risk of being falsely accused, highlighting the need for algorithms to be more inclusive and consider the complex ways in which identity and status interact.

“Xenophobic Machines: The Dutch Child Benefit Scandal,"
Amnesty International, 13 October 2021

NATIONAL UNEMPLOYMENT AGENCY IN AUSTRIA: GENDERED AND SOCIOECONOMIC BIASES

Overview

The AI system used by Austria’s National Unemployment Agency aimed to match job seekers with employment opportunities but exhibited significant bias against women, particularly those who had been unemployed for long periods or had worked part-time. The system penalised women for employment gaps and part-time work, which are often associated with caregiving roles or other gendered social expectations, thus limiting their access to job opportunities

Intersectionality at play

The biases in the system are rooted in both gender and socioeconomic factors. For women, especially those who have taken breaks from the workforce (for maternity or caregiving), the algorithm penalised employment gaps. This exacerbates existing gender inequalities, as women are often more likely than men to have non-linear career paths due to societal expectations around caregiving. Additionally, women in lower-income or part-time employment are doubly disadvantaged by the system's reliance on rigid employment history metrics that fail to account for the socio-economic context behind these career gaps. Women with disabilities, especially those in part-time or intermittent work, may also face compounded disadvantages.

Why intersectionality matters

Intersectionality is crucial in understanding how women, particularly those with caregiving responsibilities or in part-time roles, are unfairly impacted by this AI system. Gendered assumptions about work and career paths lead to a biased algorithm that disregards the socio-economic realities faced by women, reinforcing historical inequalities in employment. The algorithm's failure to account for the intersection of gender and socioeconomic status results in systemic barriers that limit women's opportunities for employment. Recognising these intersectional biases is key to designing fairer systems that consider the complexities of individual lives and employment trajectories, particularly for women who face both societal and algorithmic disadvantages.

"Discriminatory employment algorithm towards women and disabled”,
Digwatch, October 2019

THE IMPACT OF FLAWED ALGORITHMS: A CASE STUDY ON RISCANVI

Overview

The RisCanvi algorithm in Catalonia's prison system assesses inmates' recidivism risk using data such as age, gender, and nationality. The algorithm has been found to be inaccurate and biased, with over 80% of inmates flagged as high-risk not reoffending.

Intersectionality at play

The system disproportionately impacts foreign nationals, particularly immigrants and people from marginalised ethnic groups, by over-predicting their likelihood of reoffending. This exacerbates systemic biases within the criminal justice system, where certain groups—especially people of color and immigrants—are already at a disadvantage. The lack of transparency and human oversight makes it harder to challenge these biased outcomes.

Why intersectionality matters

The combination of race, nationality, and socio-economic background creates a higher risk of biased outcomes for marginalised individuals. By failing to consider these intersections, the algorithm reinforces existing societal inequalities, leading to unjust parole denials and perpetuating discrimination. Understanding intersectionality in this context allows us to see that it is not just about a singular characteristic (e.g., gender or race) but how multiple forms of disadvantage compound to create unfair outcomes.

"Automating (In)Justice: An Adversarial Audit of RisCanvi",
Eticas Foundation (July 2024)

WELFARE FRAUD CASE IN DENMARK: TARGETING MARGINALISED GROUPS

Overview

In Denmark, the welfare authority Udbetaling Danmark (UDK) uses AI algorithms to detect welfare fraud. The system has been criticised for targeting individuals from marginalised groups, particularly those with disabilities, people from racial minorities, and those in non-traditional family structures. These groups face disproportionate scrutiny under the algorithm, which exacerbates existing disparities.

Intersectionality at play

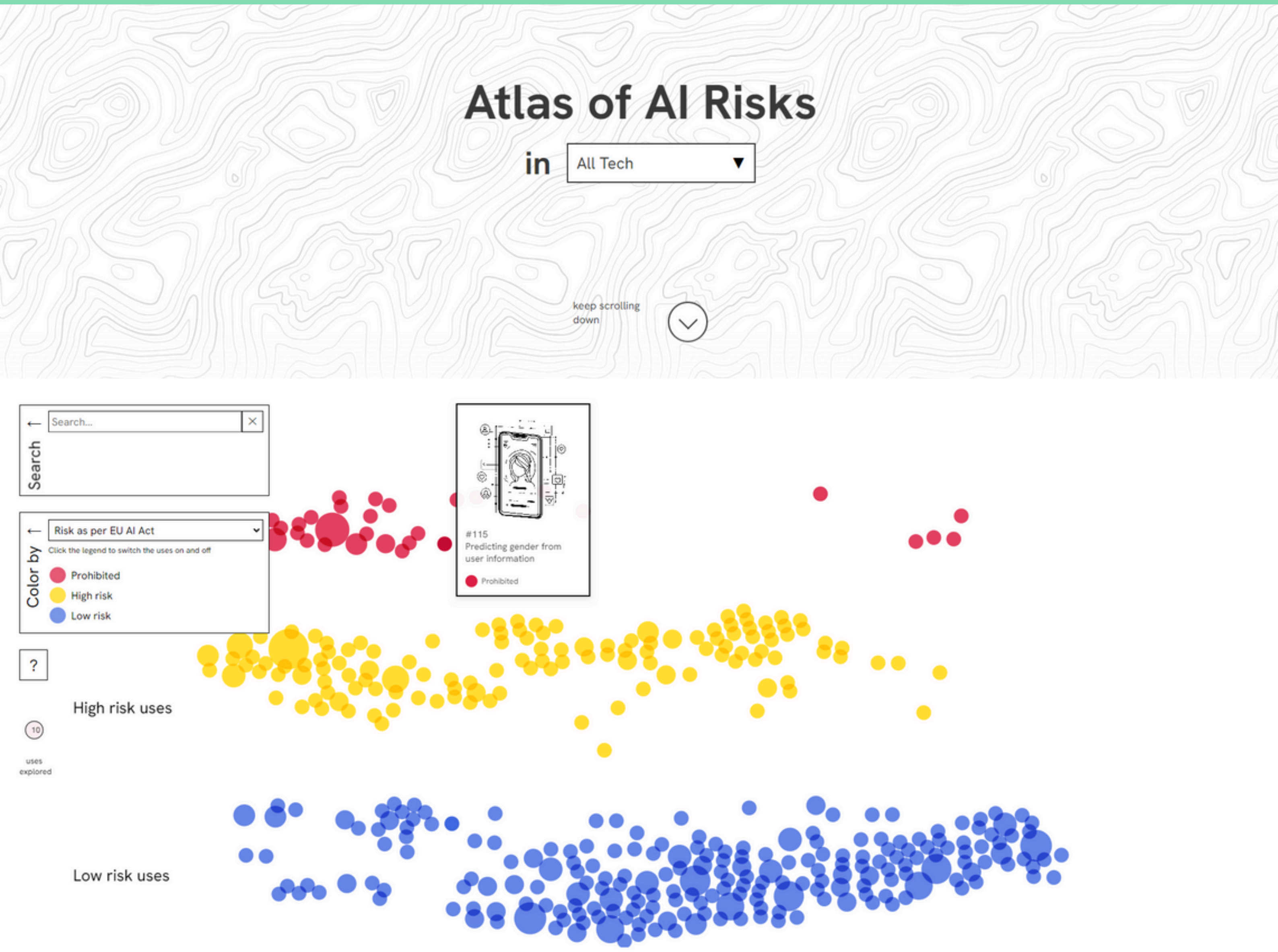
The intersection of race, disability, and non-traditional family structures makes certain individuals more vulnerable to being flagged by the system. For example, a Black person with a disability who is part of a single-parent household might face compounded discrimination, as the algorithm may flag them due to the combination of these intersecting factors. Additionally, people in non-traditional family structures may be wrongly flagged because their profiles don't conform to the system's assumptions about "normal" family arrangements.

Why intersectionality matters

Intersectionality is crucial in understanding how this AI system disproportionately impacts individuals at the intersections of multiple marginalized identities. People who are already disadvantaged in one area—whether because of race, disability, or family structure—are more likely to experience unjust treatment because of the compounded effects of these biases. Without addressing these intersectional biases, AI systems risk perpetuating and deepening existing inequalities in welfare and social services.

"Denmark: Coded Injustice: Surveillance and Discrimination in Denmark’s automated welfare state",
Amnesty International, November 2024

ATLAS OF AI RISKS



Atlas of AI Risks

in All Tech

Color by: Risk as per EU AI Act

Click the legend to switch the uses on and off

#115 Predicting gender from user information

We recommend checking out the [Atlas of AI Risk](#) (Social Dynamics Lab, Nokia Bell Labs).

It's a great resource for understanding how AI bias affects real-world situations. **It includes 380 documented cases of AI applications linked to incidents reported in the news and compiled in the AI Incident Database**. Some examples include gender bias in Google Image Search, hiring algorithms giving invalid positive feedback on interview answers, Airbnb's trustworthiness algorithm reportedly banning users without explanation and discriminating against sex workers, and algorithms in healthcare that have reportedly harmed disabled and elderly patients.

TIMELINE OF AI BIAS

AI bias is not a new phenomenon—it has existed since the technology itself was developed. This timeline highlights some of the significant moments in AI's history over the past 12 years, showing how bias evolves alongside technological advancements. **It can be used to emphasise the critical need for continued education about AI and its biases, ensuring that awareness and action evolve alongside the technology.**

2012

KNIGHT CAPITAL TRADING ALGORITHM FAILURE

A glitch in Knight Capital's trading algorithm caused a $440 million loss in 30 minutes, illustrating the risks of unchecked AI automation in financial systems. More

2015

GOOGLE PHOTOS SCANDAL

AI mislabeled Black individuals as "gorillas," showcasing racial bias in image recognition systems. More

2016

NORTHPOINT COMPAS TOOL

A criminal risk assessment tool used in the U.S. was shown to disproportionately classify Black defendants as high-risk, perpetuating racial disparities in the justice system. More

2018

GENDER SHADES STUDY

Revealed AI gender classifiers were less accurate for darker-skinned women, exposing bias in commercial AI systems. More

2019

DUTCH CHILDCARE BENEFIT SCANDAL

AI falsely accused minority families of fraud, devastating lives and reinforcing systemic racism. More

APPLE CREDIT CARD BIAS

Apple's credit card was criticised for offering significantly lower credit limits to women than men with similar financial profiles, highlighting gender bias in financial algorithms. More

AUSTRIAN UNEMPLOYMENT AGENCY CASE

Penalised women with employment gaps, exacerbating gender inequities in job placement. More

2023

ROTTERDAM WELFARE FRAUD CASE

AI prioritised wealthier groups, neglecting low-income and immigrant populations, deepening healthcare inequalities. More

2024

GEMINI AI DIVERSITY ERRORS

Image generator depicted Nazi figures as people of colour. More

RESOURCES TO BUILD AI LITERACY & INCREASE AWARENESS

Building AI literacy is crucial for the policy sector to understand AI fundamentals, such as machine learning and data ethics, while also raising awareness of intersectionality in AI. It enables policymakers to recognise AI's societal impact, assess tools for fairness, bias, and privacy, and ensure responsible, inclusive AI governance and regulation.

RESOURCES TO BUILD AI LITERACY

AI4EU Platform - Education Catalogue
Offers courses and tutorials on AI ethics and technical skills, focusing on European values of inclusivity. Visit here.

Coursera: AI for Everyone
A beginner-friendly course explaining AI concepts for non-technical audiences. Visit here.

Microsoft Learn
AI Literacy for Educators: Provides AI toolkits for teachers and learners. Visit here.

Digital Promise - AI Literacy Framework
Emphasises ethical AI, data privacy, and combating misinformation, with a structured approach for educators and learners. Visit here.

The AI Education Project (aiEDU)
Targets underserved communities with accessible curricula and tools to close AI literacy gaps. Visit here.

Institute of Business Analytics, University of Ulm: Bias & Fairness in AI Systems
A comprehensive resource that provides an accessible introduction to understanding bias and fairness in AI systems. It’s ideal to build foundational knowledge. Visit here.

RESOURCES TO BUILD AWARENESS

UN Women Intersectionality Resource Guide
Integrates intersectionality into policy design, focusing on marginalised groups. Visit here.

Amnesty International: Intersectionality Course
Practical training on combating discrimination through an intersectional lens. Visit here.

Videos that Spark Conversations
This resource explores how video-based tools can foster critical discussions about fairness and bias in technology. Visit here. For detailed insights, refer to the original article here.

AI-MYTHS: FACTS OR FICTION?

> As AI becomes more prevalent in our daily lives, misconceptions about its capabilities, limitations, and impacts abound. These myths can lead to misunderstandings about how AI works and its societal consequences, particularly regarding issues like intersectional bias, fairness, and inclusivity.

By debunking common AI myths, we can foster a more informed discussion about how to use this technology responsibly and equitably.

Isn't ChatGPT just like Google, you can search for anything?

CHAT GPT IS NOT A SEARCH TOOL

Unlike search engines that retrieve information from the web, ChatGPT generates responses based on patterns and knowledge from its training data. It doesn't provide real-time information or direct links to sources.

AI systems cannot make errors, do they?

AI SYSTEMS CAN MAKE MISTAKES

AI systems, including ChatGPT, are not infallible. They can make errors, produce biassed outputs, or provide inaccurate information based on their training data.

Will AI save me time on everything I do?

AI IS NOT A TIME-SAVING SUPERHERO

While AI can enhance efficiency in certain tasks, it often requires significant investment in training and user education. Users need to understand limitations to use AI effectively.

Won’t AI eventually learn enough to provide perfect answers for any question?

AI ALWAYS NEEDS MORE LEARNING

AI models are limited by the data they are trained on and by the scope of their design. While they can improve with more data, they will never be fully capable of understanding every question or context.

Isn’t AI equally good at understanding all languages?

AI STRUGGLES WITH MULTILINGUAL DATA

Many AI systems are primarily trained on data from resource-rich languages, which means they tend to perform better in those languages. As a result, their accuracy can be lower when working with underrepresented languages.

Isn't it safe to trust content generated by AI if it’s grammatically correct?

CORRECT SYNTAX BUT MISLEADING MEANING

GenAI's cognitive ease: syntactically correct doesn't mean semantically accurate. Generative AI can produce text that is grammatically correct and fluent, but this doesn't guarantee the text is factually accurate or semantically meaningful. Users should always critically evaluate the content.

What exactly is AI? Isn't it just a buzzword?

AI IS NOT AN OVERHYPED TERM

AI is a broad field encompassing various technologies and methodologies. It's important to understand the specific context and capabilities of AI rather than viewing it as a vague.

Can’t AI be completely fair and unbiased if we train it correctly?

100% FAIRNESS AND BIAS-FREE AI IS A MYTH

Achieving absolute fairness and eliminating all biases in AI is currently unattainable. Biases can enter through data, algorithms, and human influence, requiring continuous efforts to minimise and manage them. While perfect fairness is impossible, AI development can aim for greater fairness by considering diverse perspectives and reducing biases, making systems fairer over time.

THANK YOU!

This toolkit was made possible thanks to the invaluable time, contributions, and insights of experts and stakeholders across the policy, civil society, and industry sectors. We extend our gratitude to everyone who took the time to respond to the survey, take part in the interviews and focus groups, sharing their perspectives and expertise.

The toolkit was developed by Work Package 5 of the DIVERSIFAIR project, but it reflects the collective efforts of the entire project team. We deeply appreciate the contributions of our partners in the consortium, for their insights and support that have been instrumental in bringing this toolkit to fruition.

We also recognise that this toolkit is part of an ongoing process, and we welcome feedback from users to ensure it continues to evolve and better address your needs.

Thank you all for your dedication and commitment to fostering a fair and inclusive future for AI.

GIVE US YOUR OPINION

Co-funded by
the European Union

The DIVERSIFAIR project has received funding from the European Education and Culture Executive Agency (EACEA) in the framework of Erasmus+, EU solidarity Corps A.2 - Skills and Innovation under grant agreement 101107969.

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