AI & Intersectionality

A TOOLKIT FOR FAIRNESS & INCLUSION

FOR THE INDUSTRY 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?

Artificial intelligence (AI) is reshaping industries, driving innovation, and unlocking new possibilities. However, alongside its transformative potential, AI carries risks that can undermine its credibility and impact. **Bias in AI - particularly intersectional bias - has become a critical issue, challenging the fairness, inclusivity, and trustworthiness of AI systems.**

Intersectional bias arises when overlapping forms of discrimination, such as racism, sexism, ableism, and classism, intersect within AI systems, amplifying inequalities. This doesn't just affect individuals—it can harm organisations' reputation, alienate key customer groups, and increase exposure to legal and regulatory risks. A notable example is the Dutch childcare benefits scandal of 2021, where an algorithm unfairly targeted immigrant parents, leading to systemic harm and public distrust.

This toolkit, created through the ***DIVERSIFAIR*** Erasmus+ project, is based on thorough research and stakeholder engagement across the EU.
It provides an introduction to understanding and addressing intersectional bias in AI. AI tools were used to streamline the writing process.
While it highlights foundational concepts, key risks, and emerging strategies, it also sets the stage for more detailed tools and practical resources, which will be developed and refined in the future. By engaging with these principles now, organisations can start laying the groundwork for more inclusive, ethical, and effective AI systems

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?

Industry professionals

Executives

AI governance teams

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 implications for businesses and society.**
2. **Provide actionable strategies for identifying and mitigating bias in AI systems.**
3. **Demonstrate how fairness and inclusivity in AI can enhance brand reputation, increase market share, and ensure compliance with regulatory frameworks.**

The development of this kit was informed by interviews and focus groups with members of the AI community and the industry sector, offering practical insights tailored to the challenges and opportunities faced by businesses today. While this version is tailored for the industry sector, additional kits targeting the policy and the civil society sectors have also been developed under the DIVERSIFAIR project.

POLICY KIT

CIVIL SOCIETY 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:

- **Incorporating feedback:** Adapting content based on insights from industry professionals, executives, and other users.
- **Integrating new research:** Adding case studies and tools as our understanding of intersectional bias advances.
- **Expanding resources:** Developing supplementary materials to deepen engagement and support implementation**.**

We invite all users to contribute to this iterative process, helping us create a more robust and impactful resource to ensure AI systems serve everyone fairly and equitably.

GIVE US YOUR OPINION

01. UNDERSTANDING INTERSECTIONAL BIAS IN AI

Artificial Intelligence (AI) is changing many parts of our lives, from healthcare and education to media and law enforcement. While AI has the potential to improve our world, it often reflects and reinforces biases that exist in its design and the data it uses. For civil society organisations, educators, and the general public, understanding these biases—especially through the lens of intersectionality—is key to creating fair and inclusive technology.

If left unchecked, these biases can worsen existing inequalities and cause harm, particularly for already marginalised groups. The Council of Europe's *Gender Equality Strategy (2024-2029)* calls for tackling structural barriers and encouraging diversity in AI development. Policies like the EU AI Act are steps in the right direction, but for these efforts to truly succeed, they must fully consider how different forms of discrimination intersect.

But what exactly is intersectional bias in AI? Are concepts like "intersectionality" merely trendy terms, or do they carry genuine importance? This section will break down these ideas, offering a clear explanation of what they mean and why they matter.

> *By addressing intersectional bias, industry leaders can ensure AI systems enhance equity and inclusion while aligning with organisational values and regulatory standards.*
>
> *- Inga Ulnicane, "Intersectionality in Artificial Intelligence: Framing Concerns and Recommendations for Action," April 2024*

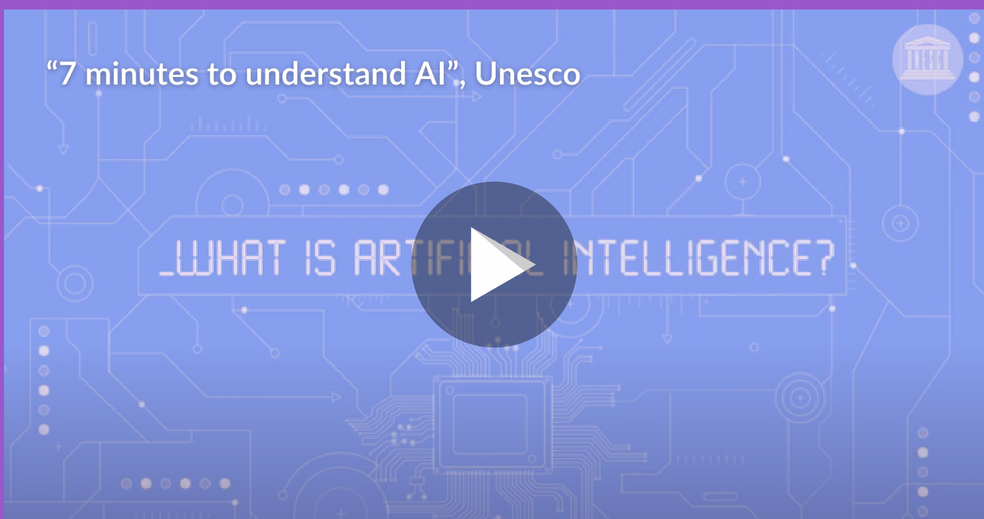
1.1 KEY CONCEPTS DEFINED

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***



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? Bias in AI can lead to flawed insights and missed opportunities. For example, biased hiring algorithms may exclude qualified candidates, reducing workforce diversity and innovation. Businesses that actively identify and mitigate bias position themselves as ethical leaders, enhancing trust among stakeholders.

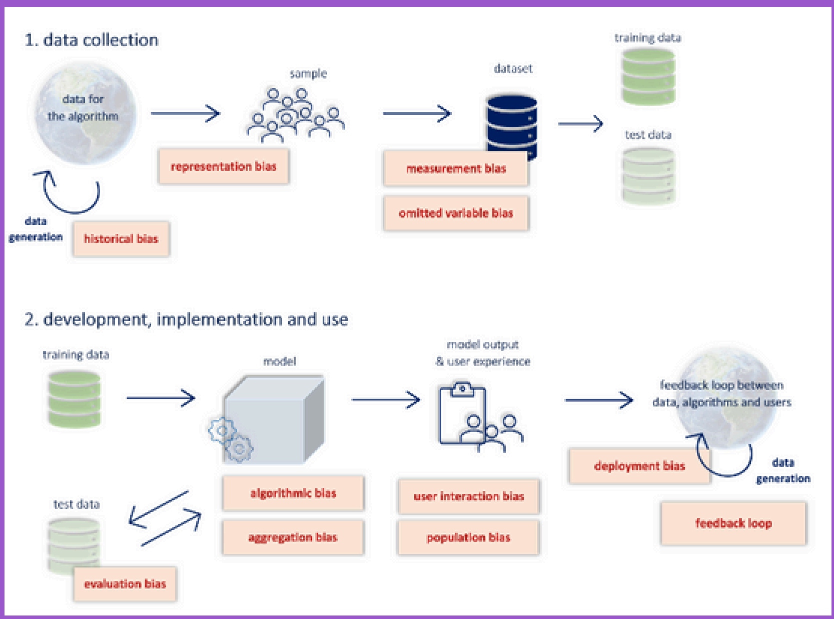


Diagramme taken from the online course "Basics of Bias & Fairness in AI systems"

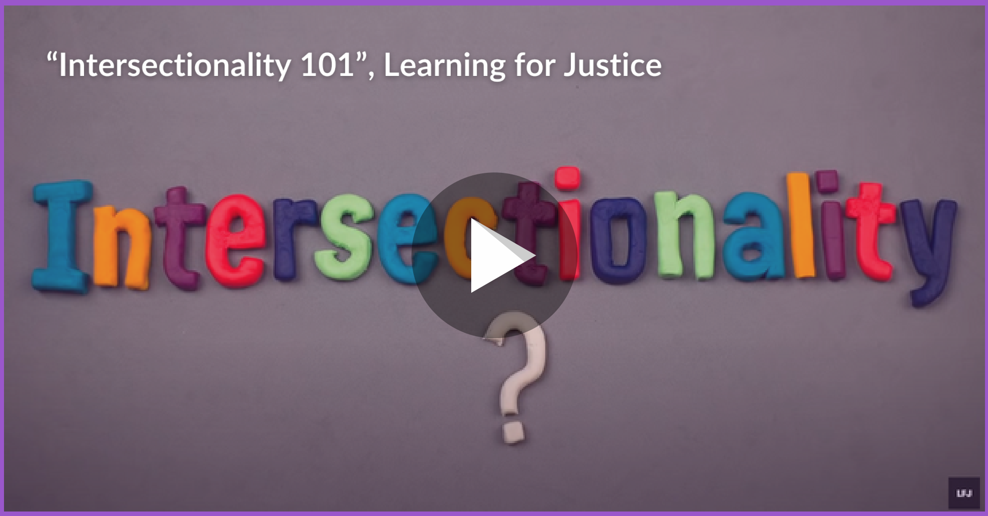
FAIRNESS

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.



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.

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 BIAS IN HIRING

AI hiring tools trained on biased data can disadvantage certain groups, such as penalizing resumes that reference women-focused organisations. This highlights how biased training data can reinforce inequalities, especially for underrepresented groups. To address this, companies should audit algorithms, use diverse datasets, and involve external experts to promote fair and inclusive hiring practices.

THIS EXAMPLE EMPHASISES THE NEED FOR

- Auditing algorithms
- Incorporating diverse datasets
- Engaging external experts

Explore

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

JUMP TO THE CASE-STUDY LIBRARY TO FIND OUT MORE

2 STEREOTYPES IN NATURAL LANGUAGE PROCESSING

AI chatbots often struggle with recognising different dialects and may respond inappropriately when interacting with women or underrepresented groups. This is partly due to training data that doesn't encompass the diversity of human language and behavior. For example, voice assistants like Siri and Alexa have been critiqued for reinforcing gender stereotypes by presenting female voices as "helpful" or "submissive" in service roles.

THIS EXAMPLE CAN BE USED AS A REMINDER TO

- Prioritise diversity and inclusivity in AI training data

Explore

"I'd blush if I could: closing gender divides in digital skills through education", Unesco, 2019

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

- Establishment of robust, transparent auditing processes for all AI-driven advertising algorithm.

Explore

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

BUILD TRUST, DRIVE GROWTH

Industry professionals and leaders have a critical role in shaping the future of ethical AI. By addressing intersectional bias, businesses can enhance trust, innovation, and inclusivity while ensuring regulatory compliance. The benefits go beyond risk mitigation—ethical AI systems unlock new markets, foster customer loyalty, and attract diverse talent, driving long-term business growth.

Now is the time to act. Businesses should prioritise diversity in AI development teams, establish regular bias audits, and embed fairness as a core principle in their AI strategy. By taking these steps, your organisation can lead the way in building inclusive technologies that resonate with global audiences.

1.3 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.

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

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

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)

DISCOVER OUR STUDY CASE LIBRARY

SUPPORTING MATERIALS FOR THIS SECTION

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/
- **Social Dynamics Lab, Nokia Bell Labs, Atlas of Social Dynamics,** https://social-dynamics.net/atlas

NEWS ARTICLES

- 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/
- “The Apple Card Didn't 'See' Gender—and That's the Problem“, The Wire, 19 November 2019: https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/
- “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
- “Apple promised an expansive health app, so why can't I track menstruation?”, The Verge, 2014: https://www.theverge.com/2014/9/25/6844021/apple-promised-an-expansive-health-app-so-why-cant-i-track
- “Google builds equity into the Pixel 6 with Real Tone photos and new voice features”,CNET, 2021: https://www.cnet.com/tech/mobile/google-builds-equity-into-the-pixel-6-with-real-tone-photos-and-new-voice-features/

RESEARCH PAPERS & SCHOLARLY ARTICLES

- **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.
- **Zang (2021),** How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity
- **Buolamwini, J., Gebru, T. (2018)** "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency
- **Fioravante (204)** "Beyond the Business Case for Responsible Artificial Intelligence: Strategic CSR in Light of Digital Washing and the Moral Human Argument"
- **Suresh, Harini & Movva, Rajiv & Dogan, Amelia & Bhargava, Rahul & Araujo Cruxen, Isadora & Martinez Cuba, Angeles & Taurino, Guilia & So, Wonyoung & D'Ignazio, Catherine. (2022).** Towards Intersectional Feminist and Participatory ML: A Case Study in Supporting Feminicide Counterdata Collection. 667-678. 10.1145/3531146.3533132.
- Ayanna Howard, "Real Talk: Intersectionality and AI," MIT Sloan Management Review, 24 August 2021, https://sloanreview.mit.edu/article/real-talk-intersectionality-and-ai

REPORTS & POLICY DOCUMENTS

- Gender Equality Strategy (2024-2029), Council of Europe: https://www.coe.int/en/web/genderequality/gender-equality-strategy
- "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
- "I'd blush if I could: closing gender divides in digital skills through education", Unesco, 2019: https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1

LEAD WITH INCLUSIVE AI

Your company deploys an AI-driven hiring tool to screen candidates. After a year, you notice fewer women and minority candidates being hired compared to previous methods.

- **How could this affect your company’s reputation and compliance with diversity goals?**
- **What steps can your team take to ensure your AI systems reflect your company’s commitment to inclusivity and ethical practices?**

02. THE BUSINESS CASE FOR FAIR AI

Fair and inclusive AI is not just an ethical imperative—it is a strategic advantage that benefits businesses in multiple ways. Addressing bias and fostering inclusivity in AI systems enable businesses to unlock new opportunities, align with corporate social responsibility (CSR) goals, and build stronger relationships with consumers. This section highlights the tangible benefits of fair AI and actionable insights for leveraging its potential.

TAPPING INTO NEW MARKETS

Ignoring bias in AI systems can result in substantial missed opportunities, restricting access to diverse talent and untapped markets. By failing to address the needs of marginalised communities, companies limit innovation and growth potential. Inclusive AI, designed with diverse voices and perspectives, not only drives fairness but also resonates with broader audiences, fostering market expansion and brand loyalty.

Expanding Market Reach through Inclusive Product Design
Apple initially overlooked menstrual tracking in its Health app, sparking backlash for failing to meet the needs of half its user base. Inclusive updates to address such oversights not only enhance functionality but also broaden product appeal to diverse customer segments.

Apple promised an expansive health app, so why can't I track menstruation?
The Verge, 2014

Google Pixel's Real Tone Feature
Google enhanced its Pixel phone cameras and voice recognition systems by incorporating insights from diverse teams. The Real Tone feature better represents a variety of skin tones, making the product more appealing to people of color and setting it apart from competitors.

Google builds equity into the Pixel 6 with Real Tone photos and new voice features,
CNET, 2021

EMBEDDING FAIR AI FOR SUSTAINABLE SUCCESS

Adopting fair AI practices strengthens corporate social responsibility efforts and demonstrates a company's commitment to societal well-being. In a landscape where consumers and investors increasingly value ethical practices, addressing bias and inclusivity becomes a business imperative. Companies that prioritise fairness in AI signal responsibility and innovation, securing competitive advantages. Conversely, those that neglect these practices risk reputational harm, loss of trust, and diminished profitability.

Ethical AI as CSR Strategy

Companies that fail to integrate fairness into AI risk accusations of "bias washing" or superficial ethics claims. By genuinely addressing bias, businesses can build trust with stakeholders.

Fioravante "Beyond the Business Case for Responsible Artificial Intelligence: Strategic CSR in Light of Digital Washing and the Moral Human Argument", 2024

Addressing Social Challenges with AI

Feminist participatory AI initiatives, such as those that support feminicide counterdata collection, showcase how businesses can use AI to address social challenges while strengthening brand values.

Suresh et al., "Towards Intersectional Feminist and Participatory ML: A Case Study in Supporting Feminicide Counterdata Collection," 2022.

03. STRATEGIC APPROACHES TO REDUCING INTERSECTIONAL BIAS IN AI

Mitigating intersectional bias in AI requires a collaborative effort that goes beyond technical solutions to embrace intentional strategies, diverse perspectives, and accountability. Bias can arise at any stage—whether in the datasets that train AI, the design of algorithms, or the deployment and usage decisions—making it essential to address it holistically.

This section draws on insights from research and recommendations from industry stakeholders, including participants from the DIVERSIFAIR expert interviews and focus groups.

- It offers tailored, **actionable tools and guiding questions** for developers, executives, governance teams, and HR professionals, empowering them to embed fairness and inclusivity into their AI systems and workflows.

- To ensure ongoing relevance and impact, **we intend to further evaluate and fine-tune these recommendations.** This iterative approach reflects the dynamic nature of bias in AI and incorporates the reflective element that is central to working through the lens of intersectionality. By doing so, we aim to adapt and improve these strategies to address emerging challenges and insights effectively.

FOR DEVELOPERS

Design AI systems with intersectionality in mind

- Avoid arbitrary subgrouping by incorporating social and historical contexts.
- Develop fairness-driven models that align with societal goals.
- Apply reflexivity in design with regular evaluations and community input.

Improve data practices

- Audit datasets for gaps and diversify data to ensure fair representation.
- Document and disclose data practices to enhance transparency.
- Tailor datasets to local or cultural needs to improve contextual relevance.

Embed inclusivity and cultural sensitivity

- Prioritise localised solutions tailored to specific cultural or regional needs.
- Plan for marginal use cases, allocating resources to support the most vulnerable groups.
- Rethink bias mitigation as holistic design, challenging whether certain AI systems should exist.

Foster collaboration across disciplines

- Partner with social scientists, ethicists, and community leaders to account for lived experiences.
- Engage with external groups to bring fresh perspectives and critical insights into development

FOR EXECUTIVES

Build diverse and inclusive teams

- Foster workforce diversity by recruiting underrepresented groups and creating inclusive cultures.
- Engage marginalised voices through participatory design workshops and collaborations.
- Use crowdsourcing to gather diverse input early in the development cycle

Build awareness and AI literacy

- Provide AI literacy training across teams, focusing on ethics, intersectionality, and societal impact.
- Debunk AI myths, such as the notion of AI neutrality, through internal workshops and communication.
- Promote intersectionality awareness in decision-making and internal policies.

Establish accountability and transparency mechanisms

- Mandate transparent reporting of AI design, deployment, and outcomes.
- Adopt an ethical risk framework to assess benefits, control, and risks for all stakeholders.
- Advocate for culturally sensitive policies and fairness audits to ensure equity.

Advocate for fair AI policies

- Push for regulations that prioritise fairness, inclusivity, and cultural values over efficiency.
- Collaborate with policymakers to ensure systems are designed for societal equity.

FOR GOVERNANCE TEAMS

Establish accountability and transparency mechanisms

- Mandate and oversee transparent reporting practices across the organisation.
- Set up independent oversight committees to audit AI systems for fairness and intersectionality.
- Develop ethical risk frameworks to guide decision-making and manage power dynamics.

Foster collaboration across disciplines

- Promote partnerships between technical teams and social science experts.
- Support external collaborations with advocacy groups and researchers to bring diverse expertise.

Advocate for fair AI policies

- Work with external regulators and internal teams to align AI practices with cultural and societal equity goals.
- Redesign datasets and systems to prioritise inclusivity and fairness.

FOR HR PROFESSIONALS

Build diverse and inclusive teams

- Recruit individuals from underrepresented groups to enhance diversity within AI teams.
- Create workplace cultures that encourage inclusive decision-making and amplify marginalised voices.

Build awareness and AI literacy

- Incorporate AI literacy and ethics training into employee development programmes.
- Raise awareness about intersectionality and its relevance in the AI lifecycle.

Foster collaboration across disciplines

- Facilitate partnerships between internal technical teams and external experts to ensure holistic development.

SUPPORTING MATERIALS FOR THIS SECTION

RESEARCH PAPERS & SCHOLARLY ARTICLES

- **Ayanna Howard**, "Real Talk: Intersectionality and AI," MIT Sloan Management Review, 24 August 2021, https://sloanreview.mit.edu/article/real-talk-intersectionality-and-ai
- **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.
- **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.
- **Suresh, Harini & Movva, Rajiv & Dogan, Amelia & Bhargava, Rahul & Araujo Cruxen, Isadora & Martinez Cuba, Angeles & Taurino, Guilia & So, Wonyoung & D'Ignazio, Catherine**. (2022). Towards Intersectional Feminist and Participatory ML: A Case Study in Supporting Feminicide Counterdata Collection. 667-678. 10.1145/3531146.3533132.
- **Bondi, Elizabeth & Xu, Lily & Acosta-Navas, Diana & Killian, Jackson.** (2021). Envisioning Communities: A Participatory Approach Towards AI for Social Good. 10.48550/arXiv.2105.01774.
- **Katell, Michael & Young, Meg & Dailey, Dharma & Herman, Bernease & Guetler, Vivian & Tam, Aaron & Bintz, Corinne & Raz, Daniella & Krafft, P M.** (2020). Toward situated interventions for algorithmic equity: lessons from the field. 45-55. 10.1145/3351095.3372874.

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

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.

CHECKLIST

HOW READY IS YOUR ORGANISATION TO ADDRESS BIAS?

This checklist is designed for startups and SMEs, offering practical steps for organisations with limited resources. It provides a foundation for evaluating readiness and commitment to addressing bias in AI, encouraging broader discussions on fairness and inclusivity. Share it with teams and stakeholders to align and initiate equitable AI practices.

LEADERSHIP COMMITMENT

- [ ] Appoint a fairness champion or ethics officer, can be part-time or shared across teams.
- [ ] Include fairness goals in strategic plans, even for smaller-scale projects.

DIVERSE TEAM REPRESENTATION

- [ ] Leverage community networks to recruit diverse candidates for project teams.
- [ ] Offer mentorship or collaboration opportunities to underrepresented voices, especially in technical roles.

RESOURCE ALLOCATION

- [ ] Dedicate a portion of your project budget to tools or services that identify and mitigate AI bias.
- [ ] Partner with academic or advocacy groups for low-cost training or audits.

REGULAR MONITORING AND ACCOUNTABILITY

- [ ] Start with lightweight bias audits using free or open-source tools
- [ ] Create a simple impact log: Track incidents of bias and corrective actions taken.
- [ ] Engage customers or community stakeholders to review and provide feedback on AI systems.

TRANSPARENCY

- [ ] Publicly share your fairness journey, even if it’s imperfect. Start with quarterly blog posts or team updates highlighting challenges and progress.

CASE STUDY LIBRARY

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

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

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

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)

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

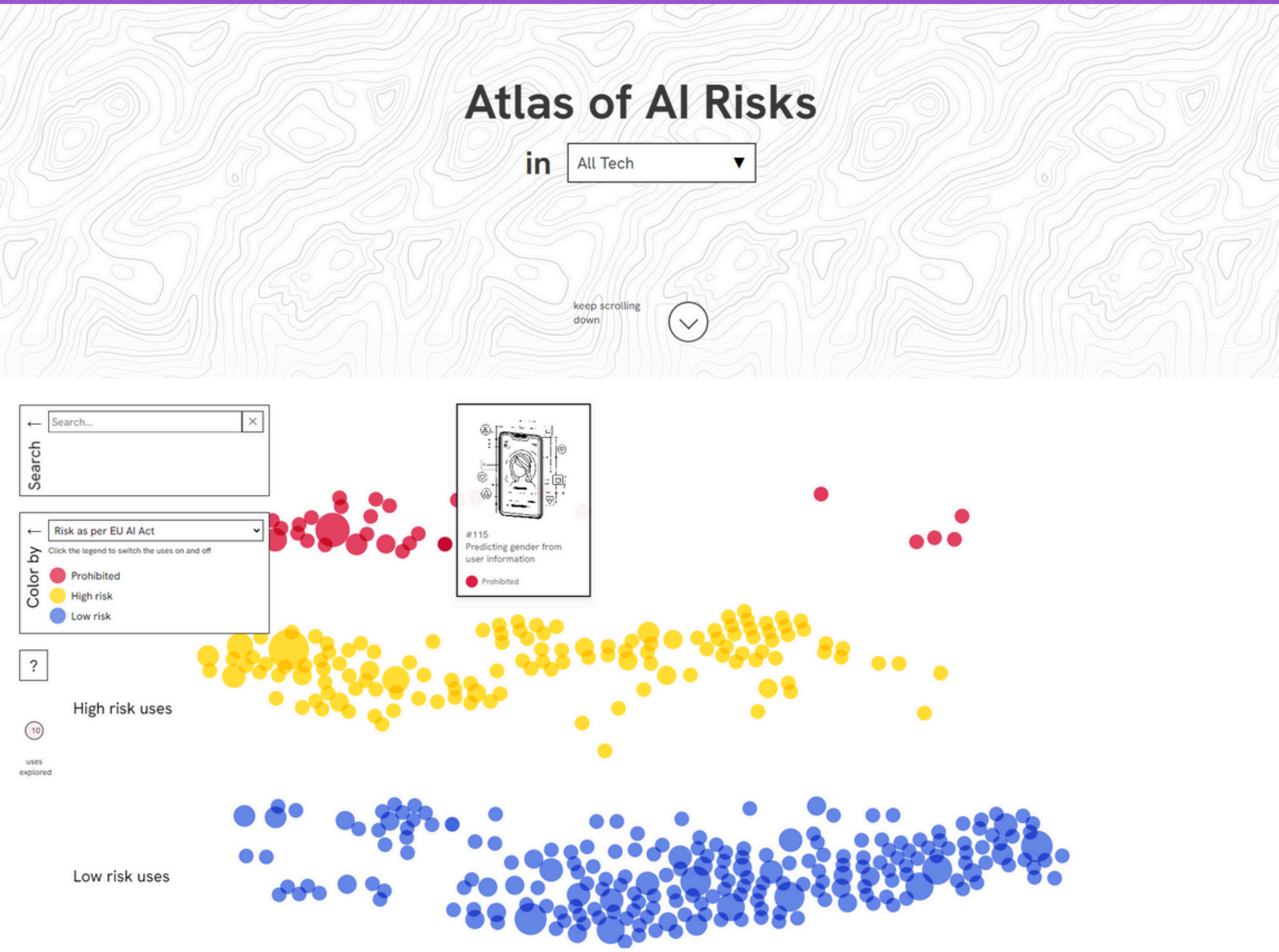
The biases identified in these systems were not confined to one aspect of identity but arose at the intersection of gender and skin type. Darker-skinned women faced the highest misclassification rates, reflecting the compounding disadvantages they experience due to their position at the intersection 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)

ATLAS OF AI RISKS



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
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
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
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.

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.