AI & Intersectionality

A TOOLKIT FOR FAIRNESS & INCLUSION

FOR CIVIL SOCIETY ORGANISATIONS

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 CSOs, additional kits targeting the industry and policy sector have also been developed under the DIVERSIFAIR project.

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

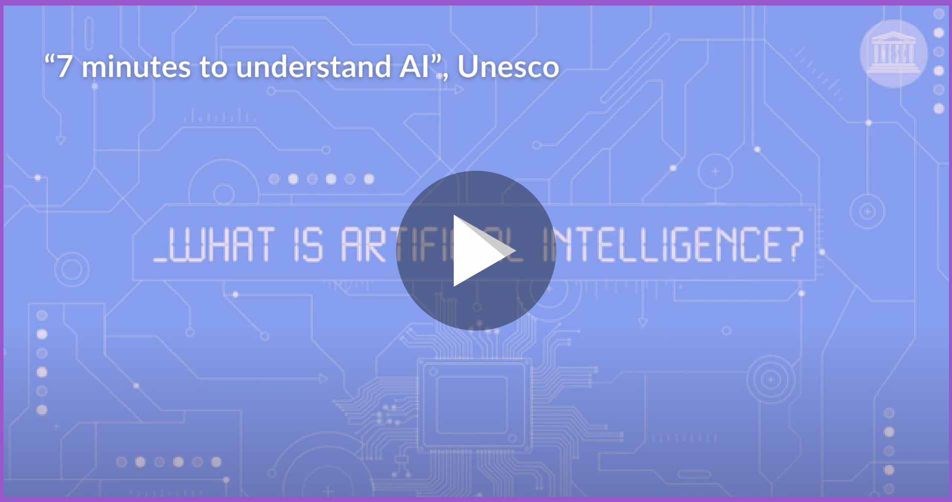
1.1 KEY CONCEPTS DEFINED

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



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.

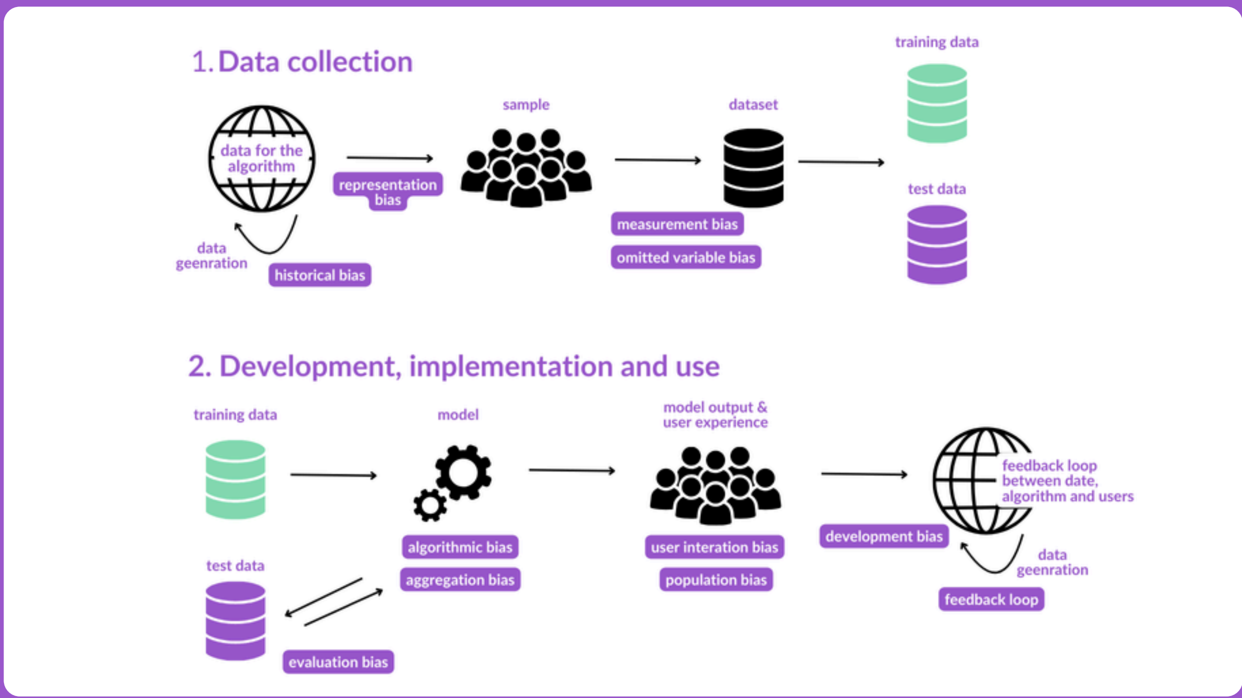


Diagramme taken from the online course ["Basics of Bias & Fairness in AI systems"]

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).

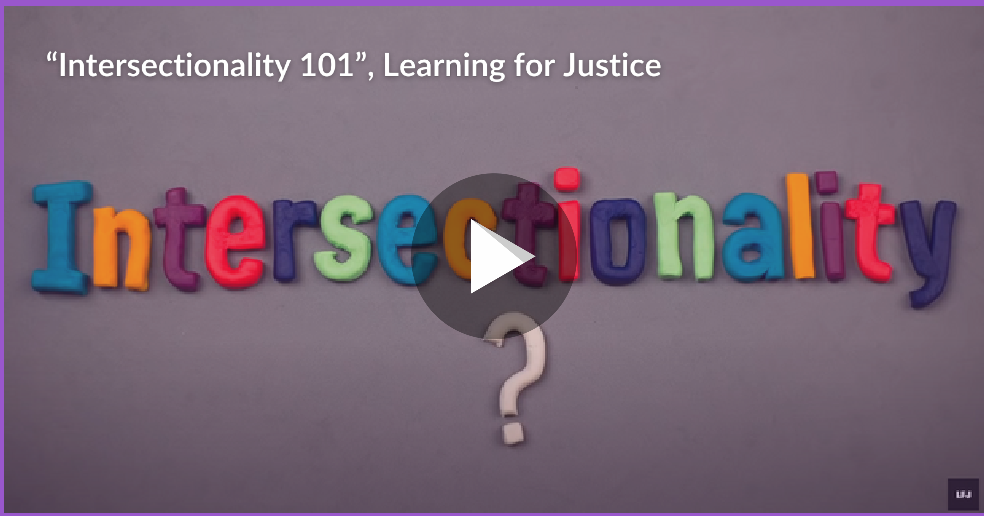
INTERSECTIONALITY

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.

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 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

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

ACT FOR ACCOUNTABILITY

Civil society organisations are uniquely positioned to address intersectional bias. You can advocate for policy changes, raise awareness, and demand transparency and fairness from AI developers.

By understanding these core concepts, CSOs can become powerful agents of change, working towards the development of ethical and inclusive AI systems.

1.3 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.

DEBUNK AND DEFY

- **What other myths have you observed in your community? How would you debunk them?**

SUPPORTING MATERIALS FOR THIS SECTION

NEWS ARTICLES

- "Automating (In)Justice: An Adversarial Audit of RisCanvi", Eticas Foundation (July 2024) https://eticas.ai/automating-injustice-an-adversarial-ai-audit-of-riscanvi/
- "How AI-powered welfare systems fuels mass surveillance and risks discriminating", Amnesty International, November 2024: https://www.instagram.com/p/DCTrNCmPC8I/?hl=fr

REPORTS & POLICY DOCUMENTS

- Gender Equality Strategy (2024-2029), Council of Europe, available at: 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
- National Institute of Standards and Technology (NIST), "There's More to AI Bias than Biased Data: NIST Report Highlights," 10 March 2022, https://www.nist.gov/news-events/news/2022/03/theres-more-ai-bias-biased-data-nist-report-highlights

RESEARCH PAPERS & SCHOLARLY ARTICLES

- **Ulnicane, Inga. (2024)**. Intersectionality in Artificial Intelligence: Framing Concerns and Recommendations for Action. Social Inclusion. 12. 10.17645/si.7543.
- **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

VIDEOS & MULTIMEDIA RESOURCES

- "Intersectionality 101", Learning for Justice: https://www.youtube.com/watch?v=w6dnj2lyYjE&t=24s&ab_channel=LearningforJustice
- "7 minutes to understand AI", Unesco: https://www.youtube.com/playlist?list=PLWuYED1WVJIPHJLk84wWQbzeZcWLt5rwU
- Institute of Business Analytics, University of Ulm, Bias & Fairness in AI Systems: Basics, https://bias-and-fairness-in-ai-systems.de/en/basics/

02. EMPOWERING COMMUNITIES

As AI continues to influence decisions across many sectors, promoting community engagement and adopting intersectional perspectives are crucial to ensuring technology serves everyone equitably. Civil society organisations (CSOs), educators, and advocates can use this toolkit to empower marginalised groups, advocate for policy changes, and support the development of inclusive AI systems.

This section provides **actionable strategies for incorporating intersectional perspectives into advocacy**, fostering cultural sensitivity in AI development, and enhancing AI literacy.

You can also adapt materials from Section 1, which introduces AI fundamentals and highlights biases, to raise awareness and mobilise communities. By **combining these resources** with the strategies outlined here, CSOs can work towards a more equitable technological future.

2.1 INTEGRATING INTERSECTIONAL INSIGHTS INTO ADVOCACY

ENGAGING MARGINALISED COMMUNITIES

Participatory design is essential for creating fair and inclusive AI systems. CSOs can lead efforts by ensuring underrepresented communities are actively involved in shaping technology.

1 CENTRE LOCAL VOICES

Marginalised groups often experience AI's harms but rarely have input into its development. Involving their perspectives ensures that solutions address real-world challenges.

> *Co-developing algorithmic accountability interventions in communities supports outcomes that are more likely to address problems in their situated context and re-center power with those most disparately affected by the harms of algorithmic systems.*
>
> - Katell et al., "Toward Situated Interventions for Algorithmic Equity." (2020)

2 VALUE EVERYDAY EXPERTISE

Community members, even without technical knowledge, can provide critical insights. Their feedback often highlights social factors and user needs that technical approaches may overlook.

3 RETHINK "BIAS FIXES"

Addressing AI harm goes beyond improving data. It involves questioning whether certain technologies should exist at all and exploring alternative approaches to ethical design.

> *(Framing) harms around algorithmic bias suggests that more accurate data is the solution, at the risk of missing deeper questions about whether some technologies should be used at all. More broadly, we found that community-based methods are important inroads to addressing algorithmic harms in their contexts.*
>
> - Katell et all, "Toward Situated Interventions"

ADVOCATE FOR CULTURAL SENSITIVITY IN AI DEVELOPMENT

AI systems must reflect the diverse cultures they impact. CSOs can play a key role in pushing for inclusive datasets and culturally aware practices.

1 GO LOCAL

Prioritise tools and solutions tailored to local needs rather than one-size-fits-all approaches. Generic systems risk erasing cultural nuances.

> ***(Community) organizations derive the most value from localized materials as opposed to what is "scalable" beyond a particular policy context.***
>
> - Katell et all, "Toward Situated Interventions"

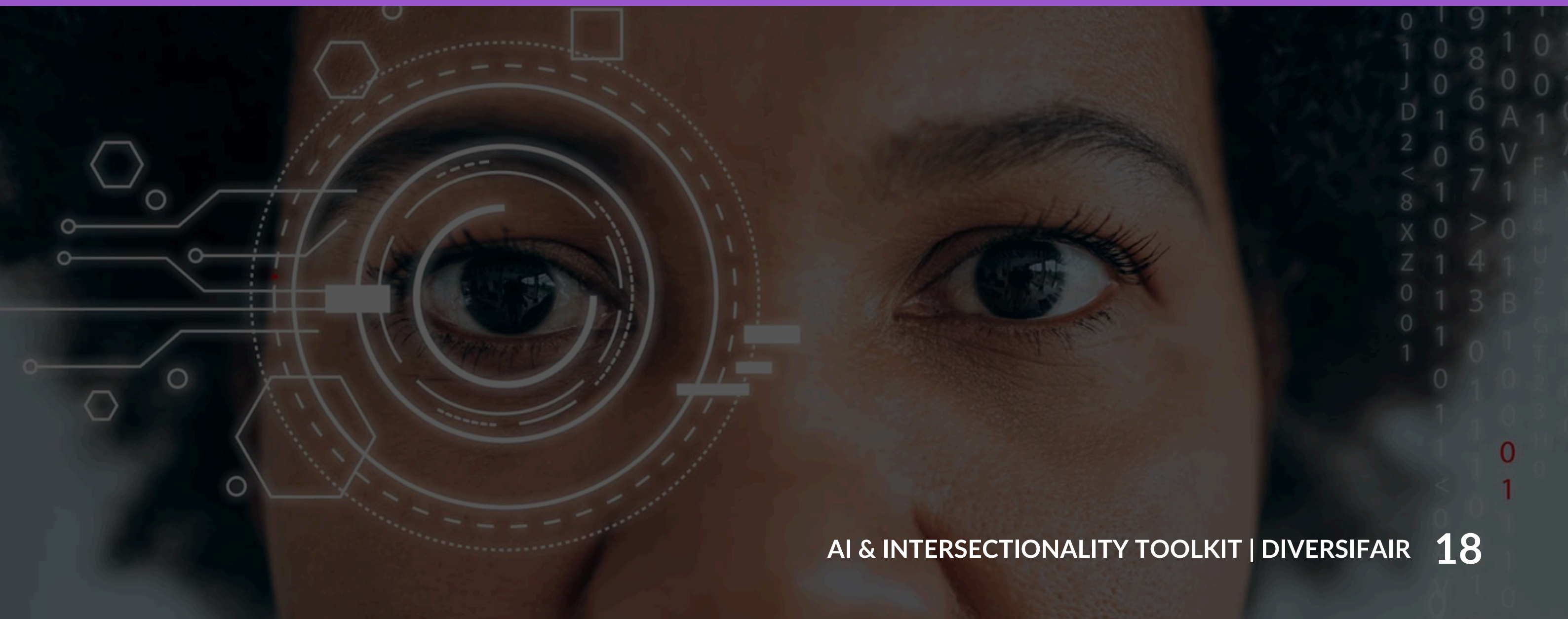
2 REPRESENTATION MATTERS

Ensure AI models include intersectional identities to avoid perpetuating harm. Overlapping factors, such as race and gender, must be addressed to create equitable systems.

> ***A commitment to intersectionality means that systems should work not only for the mainstream, majority use case, but also for those on the margins.***
>
> - Suresh et all. "Towards Intersectional Feminist and Participatory ML" (2022)

3 DESIGN FOR EQUITY

AI systems should prioritise fairness and cultural values, even if it requires additional time and resources.



2.2 INCREASING AI LITERACY AND AWARENESS

AI literacy equips individuals with the knowledge to:

- **Understand AI fundamentals:** Grasp basic concepts, such as machine learning and data ethics.
- **Navigate AI in daily life:** Recognise AI applications like chatbots, recommendation systems, and automation.
- **Critique and use AI ethically:** Evaluate AI tools for fairness, intersectional bias, and privacy concerns.

Promoting AI literacy builds public trust, supports ethical AI adoption, and prepares individuals for AI-driven workplaces. It also ensures diverse participation in discussions about AI governance.

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
Basics is a comprehensive resource that provides an accessible introduction to understanding bias and fairness in AI systems. It's ideal for CSOs and advocates aiming to build foundational knowledge. Visit here.

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. For detailed insights, refer to the original article here and the accompanying kit here.

RESOURCES TO BUILD AWARENESS

This toolkit includes resources designed to raise awareness about intersectional bias in AI. These materials can also support advocacy efforts and encourage critical discussions about fairness and equity in AI systems.

DEFINITIONS

Clear explanations of essential concepts such as AI, bias, fairness, and intersectionality.
How to use them? Share these during workshops or campaigns to help audiences grasp the foundational principles of AI fairness.

CASE-STUDIES

Real-world examples, like the Dutch childcare benefits scandal and Denmark's welfare surveillance system, show the tangible impacts of biased AI systems.
How to use them? Use these examples in campaigns or discussions with policymakers and stakeholders to illustrate the societal consequences of AI bias.

ATLAS OF RISKS

This tool showcases 380 real-world AI bias cases, highlighting issues in hiring, healthcare, image search, and platform algorithms.
How to use them? Explore documented cases to understand AI bias impacts and inform advocacy, education, or ethical AI development efforts.

AI INCIDENT TIMELINE

A visual resource highlighting key instances where AI systems caused harm, underlining the importance of vigilance and ethical oversight.
How to use them? Integrate this timeline into presentations to stress the importance of continuous education and accountability in AI development.

SUPPORTING MATERIALS FOR THIS SECTION

RESEARCH PAPERS & SCHOLARLY ARTICLES

- **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.

- **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.

- **Brianna Blaser, Christopher Lynnly Hovey, Vidushi Ojha, and Manuel A. Pérez Quiñones. 2023**. Engaging with Identity, Inclusion, & Intersectionality: Videos that Spark Conversations.

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.

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

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

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

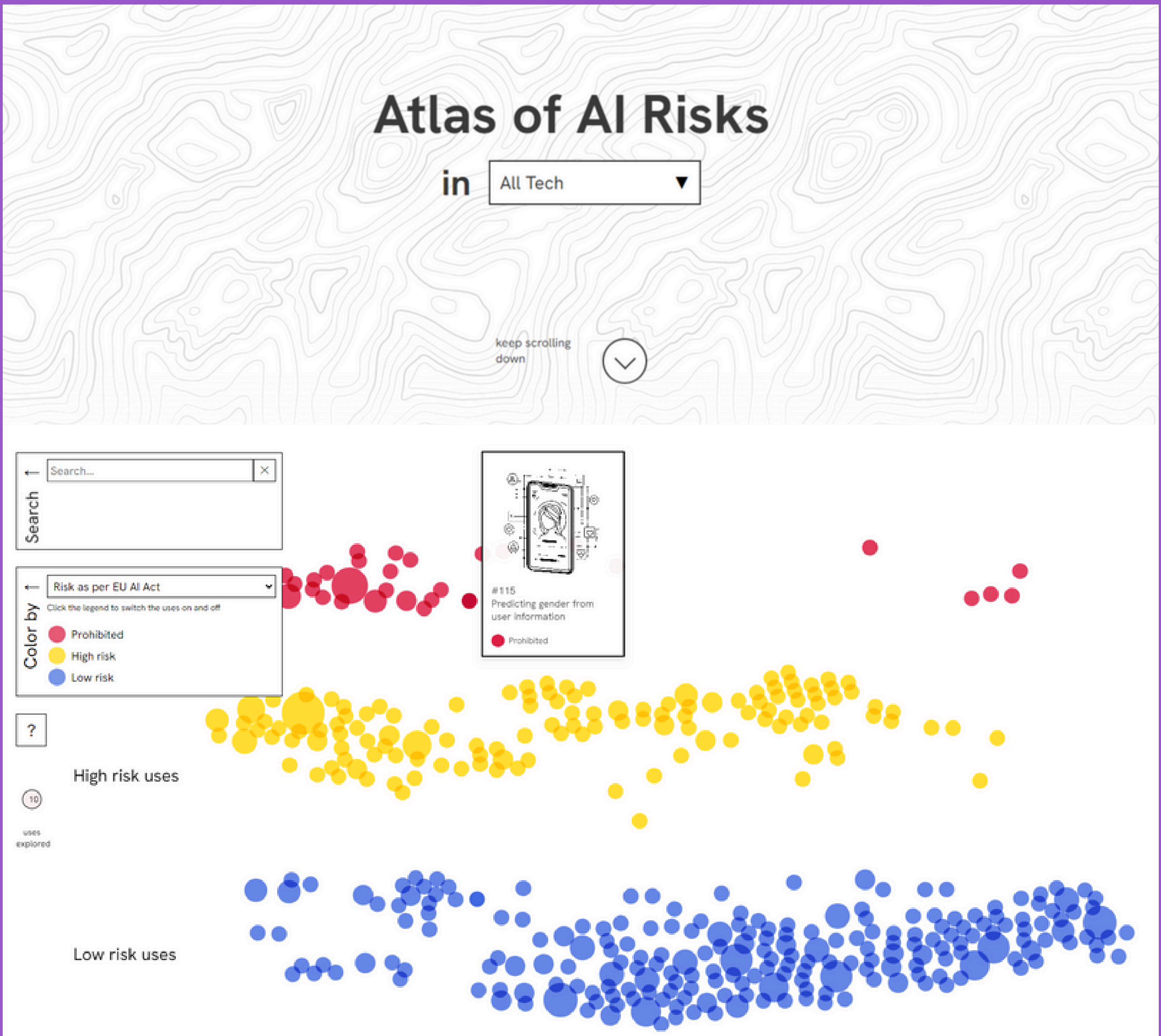
The intersection of race, disability, and non-traditional family structures makes certain individuals more vulnerable to being flagged by the system. For example, a Black person with a disability who is part of a single-parent household might face compounded discrimination, as the algorithm may flag 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



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

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.