



# **Bias, Fairness and Privacy in AI**

**October 28, 2025**

**Seminar kick-off**

**[seminar@m-chair.de](mailto:seminar@m-chair.de)**

**Chair of Mobile Business & Multilateral Security**

**Goethe University Frankfurt**

- Organizational Information
- Introduction to Bias in AI
- Introduction to Fairness in AI
- Introduction to Privacy in AI
- Investigate Research Topics
- Presentation Topics

<p><b>E-Finance</b></p> <p>Prof. Dr. Peter Gomber</p>	<p><b>Business Informatics (Informatics)</b></p> <p>Prof. Dr. Mirjam Minor</p>	<p><b>Business Informatics &amp; Information Management</b></p> <p>Prof. Dr. Oliver Hinz</p>
<p><b>Business Ethics &amp; Business Education</b> (associated)</p> <p>Prof. Dr. Gerhard Minnameier</p>	<p><b>Game-Theoretic and Causal AI Business and Economics</b></p> <p>Prof. Dr. Kevin Bauer</p>	<p><b>Economic and Business Education</b> (associated)</p> <p>Prof. Dr. Eveline Wuttke</p>
<p><b>Business Education</b> (associated)</p> <p>Prof. Dr. Helmut Niegemann</p>	<p><b>Information Systems &amp; Information Management</b></p> <p>Prof. Dr. Wolfgang König</p>	<p><b>Business Informatics</b></p> <p>Hon. Prof. Dr. Matthias Zieschang</p>
<p><b>Information Systems Engineering</b></p> <p>Prof. Dr. Roland Holten</p>	<p><b>Business Informatics &amp; Microeconomics</b></p> <p>Prof. Dr. Lukas Wiewiorra</p>	<p><b>Mobile Business &amp; Multilateral Security</b></p> <p>Prof. Dr. Kai Rannenber</p>

# Chair of Business Administration, especially Business Informatics, Mobile Business and Multilateral Security

Chair of Mobile Business & Multilateral Security

Theodor-W.-Adorno-Platz 4  
Campus Westend  
RuW Building, 2<sup>nd</sup> Floor

Phone: +49 69 798 34701

Email: [info@m-chair.de](mailto:info@m-chair.de)

URL: [www.m-chair.de](http://www.m-chair.de)



# Team & External PhD Students



**Kai  
Rannenberg**



**Narges  
Arastouei**



**Diana  
Weiss**



**Basharat  
Mubashir  
Ahmed**



**Arman  
Khan**



**Ann-Kristin  
Lieberknecht**



**Sascha  
Löbner**



**Ahad  
Niknia**



**Atiyeh  
Sadeghi**



**Peter  
Hamm**



**Tim  
Schiller**



**Michael  
Schmid**



**Frédéric  
Tronnier**



**Adrian  
Völker**

## Selected Alumni



Prof. Dr. Jan  
Muntermann  
*Augsburg  
University*



Dr. Stefan  
Figge  
*Microsoft*



Dr. Mike  
Radmacher  
*Techem  
Energy  
Services  
GmbH*



Dr.  
Andreas  
Albers  
*Deutsche  
Telekom*



Dr. Stefan  
Weiss  
*Swiss Re*



Prof. Dr. Denis  
Royer  
*Ostfalia  
Hochschule für  
angewandte  
Wissenschaften*



Dr. Markus  
Tschersich  
*Continental*



Dr. Ahmad  
Sabouri  
*Harman  
Kardon*



Dr. Falk  
Wagner  
*T-Mobile*



Dr. Christian  
Kahl  
*CyberSolutions  
GmbH*



Dr. Gökhan Bal  
*Helaba*



Dr. André  
Deuker  
*KfW*



Dr. Shuzhe  
Yang  
*GLS*



Dr. Ahmed  
Yesuf  
*FARO*



Dr.  
Welderufael  
Tesfay  
*Deutsche  
Telekom*



Dr. Fatbardh  
Veseli  
*E.ON Digital  
Technology*



Dr. Majid  
Hatamian  
*Mastercard*



Dr. habil.  
Sebastian  
Pape  
*Continental*



Dr. David  
Harborth  
*Capgemini  
Invent*



Dr.  
Christopher  
Schmitz  
*Deutsche  
Börse*

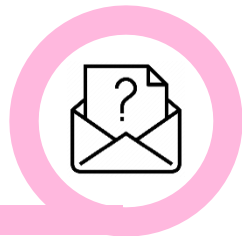


**Arman Khan**

**RuW Building, Office 2.220**

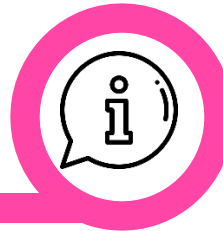


- **arman[dot]khan[at]m-chair[dot]de**
- **seminar[at]m-chair[dot]de**



All questions to:

arman[dot]khan[at]m-chair[dot]de



More information

All relevant information will be published on

[www.m-chair.de](http://www.m-chair.de)



Appointment

You will always get a personal/online appointment ASAP, since a regular

exchange is important for a very good

work (arranged by mail)

Date	What	Where/When/How
28.10.2025	Kick-off and distribution of topics	<a href="#">RuW - RuW 2.202</a>
04.11.2025 (12:00 German time)	Submit Chosen Topic	Email to: <a href="mailto:seminar@m-chair.de">seminar@m-chair.de</a> <a href="mailto:arman.khan@m-chair.de">arman.khan@m-chair.de</a>
19.01.2026 (by 12:00 Midday German time)	Final Paper submission	MS-word AND PDF to: <a href="mailto:seminar@m-chair.de">seminar@m-chair.de</a>
19.01.2026 (by 12:00 Midday German time)	Presentation submission	PPT or PDF as Email to: <a href="mailto:seminar@m-chair.de">seminar@m-chair.de</a>
26.01.2026	Presentation	11:00 - 17:00 <a href="#">Seminarhaus SH-3.102</a>
28.01.2026	Presentation	09:00 - 18:00 <a href="#">Seminarhaus SH-3.103</a>
29.01.2026	Presentation	09:00 - 18:00 <a href="#">Seminarhaus SH-3.103</a>

- **Course agenda is online.**
  - Please keep yourself updated!
  - Check the website of the course:
    - <https://www.m-chair.de/teaching?view=article&id=289:seminar-bias-fairness-and-privacy-in-ai-systems-winter-2025-2026&catid=15:lectures>
- Exam:
  - <https://www.wiwi.uni-frankfurt.de/en/faculty/deans-office/operational-divisions/examination-office>

## Program:



Presentations from 11:00 - 18:00, 26.01.2026 - 30.01.2026



Presentation length: 10 - 12 min  
Paper length: max 10 pages



Presentation discussion: 10 - 15 min, students will be assigned to ask questions before the presentation of another topic.



Template on: <https://m-chair.net/teaching/theses>  
Citation: APA Style

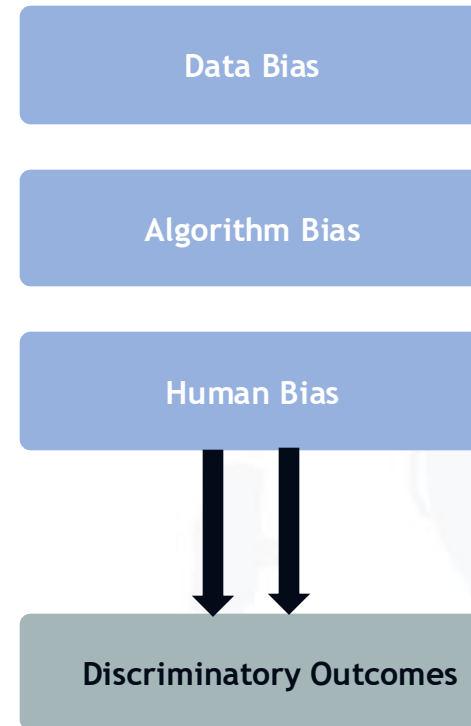


Grading: 40% presentation (incl. participation), 60% paper

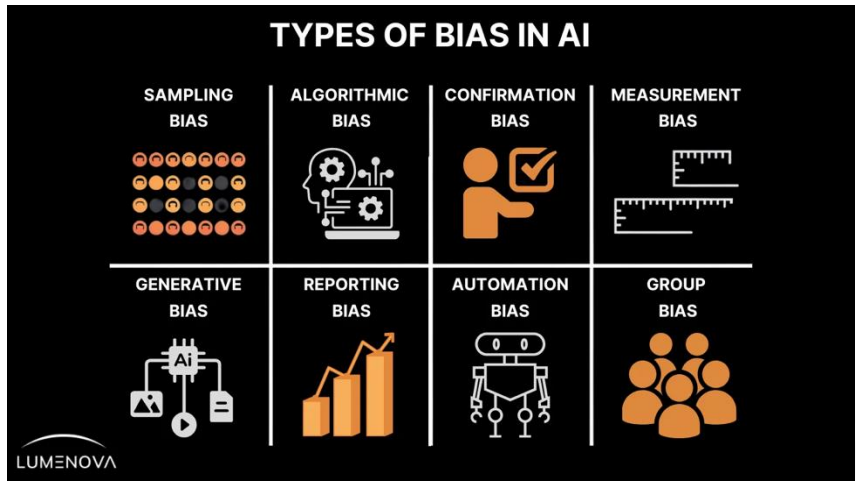
- Organizational Information
- Introduction to Bias in AI
- Introduction to Fairness in AI
- Introduction to Privacy in AI
- Investigate Research Topics
- Presentation Topics

- AI bias refers to systematic deviations from normative standards, often reflecting societal prejudices.
- Bias arises from multiple sources: biased data, model design choices, and human decision-making.
- Algorithms trained on imperfect or prejudiced data can reproduce and even amplify discrimination.

## Understanding AI Bias (Origin)



General Source of Bias	Specific Types (from Image)	Explanation / Link
Data Bias	<ul style="list-style-type: none"> <li>- Sampling Bias</li> <li>- Measurement Bias</li> <li>- Group Bias</li> <li>- Reporting Bias</li> </ul>	These arise due to flaws in the datasets: under-representation, skewed measurement tools, omitted variables, or selectively reported outcomes.
Algorithmic Bias	<ul style="list-style-type: none"> <li>- Algorithmic Bias</li> <li>- Automation Bias</li> <li>- Generative Bias</li> </ul>	These emerge from how the AI system processes inputs, makes decisions, or generates content (e.g., a generative model mimicking biased patterns).
Human Bias	<ul style="list-style-type: none"> <li>- Confirmation Bias</li> <li>- Reporting Bias (again)</li> <li>- Group Bias</li> </ul>	Reflects cognitive and societal biases from developers, annotators, or users that are baked into design, labelling, or interpretation.



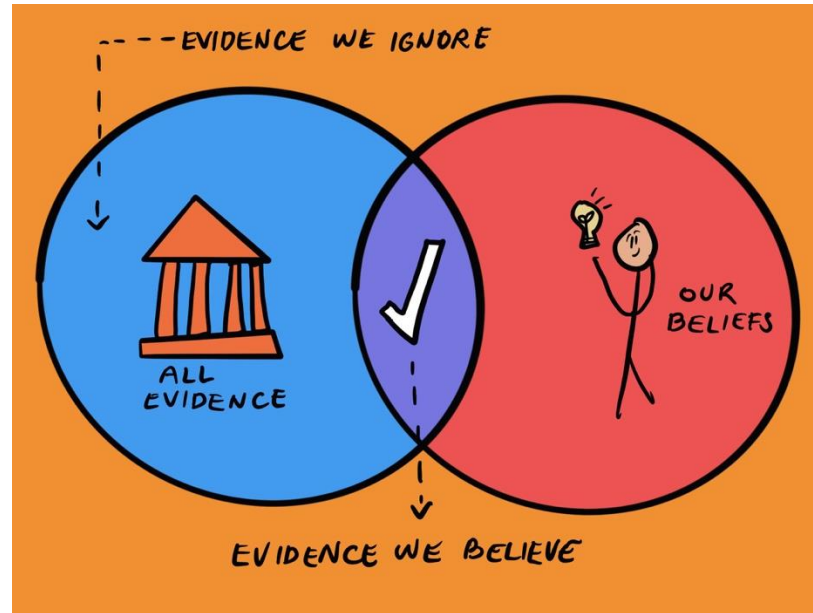
- Occurs when the training dataset is not representative of the full population
- Leads to skewed predictions and unfair outcomes, favouring overrepresented groups
- Example: Facial recognition systems trained mostly on light-skinned faces perform poorly on darker-skinned individuals
- Ensuring fairness and accuracy requires inclusive and representative data collection across all demographic groups



- Arises when an AI system **prioritizes certain attributes or patterns** due to limitations in data or algorithm design.
- Models often **reproduce historical biases** embedded in training datasets.
- Example: A hiring algorithm trained mostly on male resumes may **favor male candidates**, reinforcing gender bias.
- Addressing this requires **representative data** and **careful algorithmic design** to avoid amplifying existing inequities



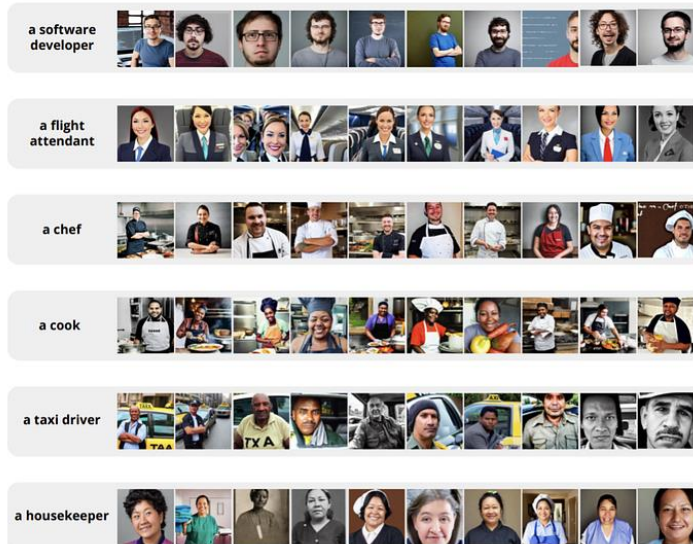
- Occurs when an AI system **amplifies pre-existing biases** in data or its creators.
- Reinforces patterns that **align with prior assumptions**, rather than challenging them.
- Example: An AI system that predicts job candidates' success based on biases held by the hiring manager.
- Consequences include **reduced innovation** and **misalignment with evolving social norms and regulations**.



- Occurs when **data is inaccurately captured**, leading to **overrepresentation or underrepresentation** of certain groups or scenarios.
- Example: A survey collecting more responses from urban residents, leading to an underrepresentation of rural opinions.
- Leads to **distorted predictions and decisions** when applied to real-world contexts.
- Avoiding this bias requires **comprehensive and accurate data collection practices**.



- Occurs when **AI-generated content** reflects **imbalances in the training data**.
- Common in **text, image, or speech generation** tasks using **biased datasets**.
- Example 1: A model trained mostly on **Western literature** may **overlook non-Western cultural perspectives**.
- Example 2: An image generation model trained on datasets with **limited diversity in human portraits** may struggle to accurately represent a **broad range of ethnicities**.
- Results in **unfair or culturally narrow outputs**, especially problematic when representing **global viewpoints**.
- Requires **inclusive and diverse training data** to ensure fair, representative content generation.



- Occurs when the **frequency or nature of events** in training data **doesn't match real-world patterns**.
- Leads to **skewed AI outputs**, especially in tasks like sentiment analysis or trend detection.
- Example 1: Overrepresentation of positive reviews causes AI to **overestimate customer satisfaction**.
- Example 2: A medical dataset that under-represents women, leading to less accurate diagnosis for female patients.
- Results in **inaccurate insights and poor decision-making**.
- Requires **balanced, representative data** that reflects the **true distribution** of events and opinions.



- Refers to the **tendency to over-rely on AI decisions**, even when they may be flawed or uncertain.
- Common in high-stakes domains like **medical diagnostics** or **quality control**.
- Example: A defect detection system may **miss subtle errors** that a human inspector would catch.
- Over-trusting automation can lead to **critical mistakes being overlooked**.
- Requires a **balanced approach**, integrating **human oversight** with AI-driven decision-making.





- Organizational Information
- Introduction to Bias in AI
- Introduction to Fairness in AI
- Introduction to Privacy in AI
- Investigate Research Topics
- Presentation Topics

# Understanding Fairness in AI

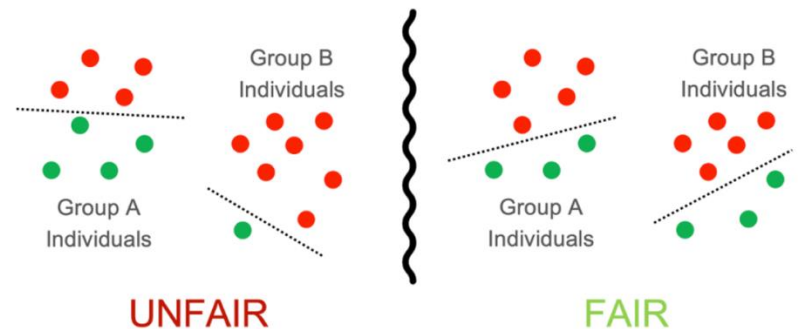
- **Description:**  
Fairness represents a **social and ethical objective**—to ensure AI systems treat individuals and groups equitably.
- **Core Principles:**
  - **Non-discrimination:** No group should be unfairly advantaged or disadvantaged.
  - **Accountability:** AI processes and outcomes should be explainable and justifiable.
  - **Inclusivity:** Systems must account for diverse demographics and contexts.
- **Key Point:**  
Fairness requires a **conscious, value-driven design** — not just the removal of bias.
- “Bias is a symptom; fairness is the cure. Achieving fairness demands both technical precision and ethical responsibility.”



Dimension	Bias	Fairness
<b>Nature</b>	Unintentional deviation from true or expected value	Intentional ethical goal to prevent discrimination
<b>Focus</b>	Technical accuracy	Social justice and equity
<b>Type</b>	Can be positive or negative	Concerned with preventing negative bias
<b>Source</b>	Data imbalance, design flaws, historical context	Ethical and governance frameworks
<b>Goal</b>	Minimize systematic deviation	Promote equitable outcomes
<b>Responsibility</b>	Data scientists, engineers	Entire AI ecosystem – developers, regulators, society

# AI Fairness Definitions: Group Fairness

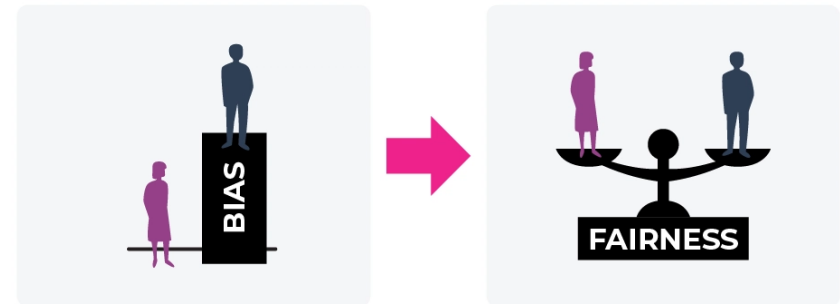
- **Description:**
  - Ensures that **different demographic groups are treated equally or proportionally** in AI systems.
  - Focuses on equality of outcomes across sensitive attributes such as gender, race, or age.
- Can be subdivided into:
  - **Demographic Parity** - Positive and negative outcomes are equally distributed across groups.
  - **Disparate Mistreatment** - Fairness defined by equal **misclassification rates**.
  - **Equal Opportunity** - True positive and false positive rates are equal across demographic groups.
- **Example:**  
Ensuring an AI loan approval system approves and rejects applications at equal rates for men and women with similar credit profiles.



# AI Fairness Definitions: Individual Fairness

- **Description:**
  - Ensures that **similar individuals are treated similarly**, regardless of group membership.
  - Based on **similarity-based** or **distance-based** measures between individuals.
  - Focuses on consistency and equal treatment at the individual level.
- **Example:**

Two applicants with similar qualifications and experience should receive **similar hiring scores**, regardless of gender or ethnicity.



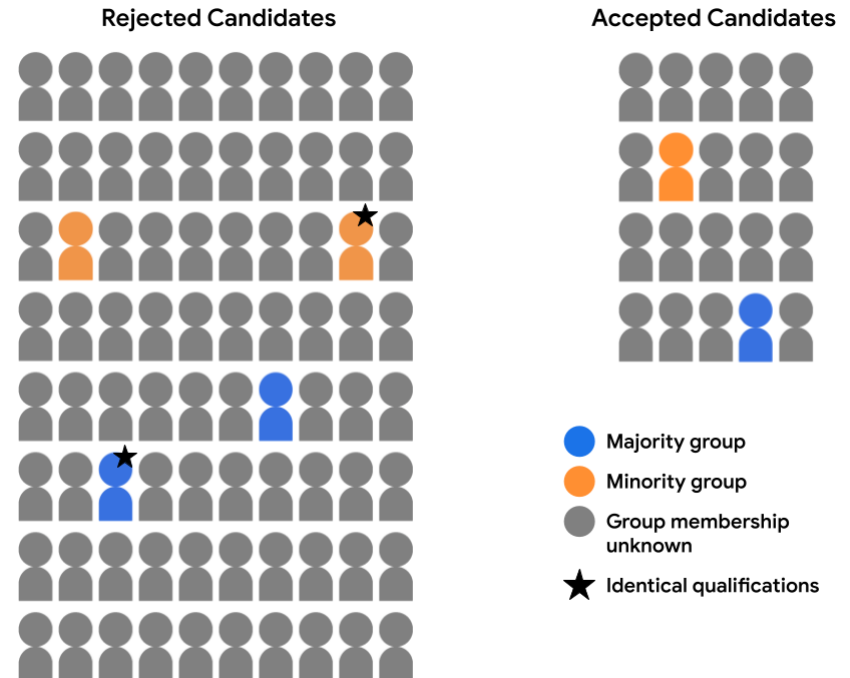
# AI Fairness Definitions: Counterfactual Fairness

- **Description:**

- Aims to ensure fairness even in **hypothetical (counterfactual) scenarios**.
- An AI system is counterfactually fair if it would make the **same decision** for an individual even if their protected attributes (e.g., gender or race) were different.

- **Example:**

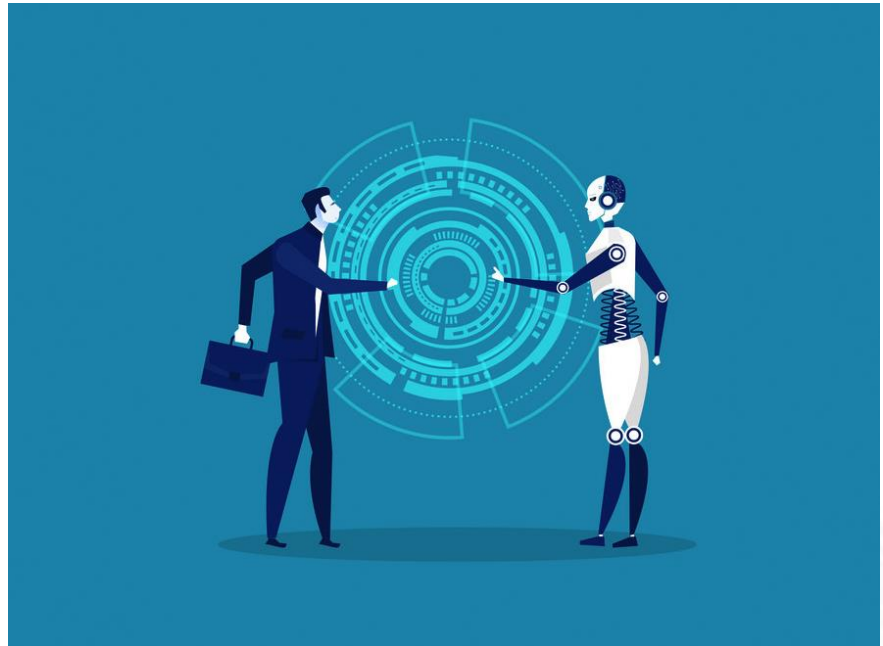
A university admission model should make the **same decision** for a student if their gender were switched from female to male, assuming all other qualifications remain the same.



- **Description:**
  - Focuses on the **fairness and transparency of the decision-making process**, not just the outcomes.
  - Emphasizes explainability, accountability, and visibility into AI decisions.
- **Example:**  
Implementing a **transparent decision-making process** that allows stakeholders to understand how the AI reached its conclusions.



- **Description:**
  - Ensures that AI systems **do not perpetuate historical inequalities or causal dependencies** related to sensitive attributes.
  - Fairness is achieved by identifying and removing biased causal pathways from input features to predictions.
- **Example:**  
Developing an AI recruitment model that removes the **causal influence of gender** on hiring outcomes, focusing instead on objective skill-related variables.



- Organizational Information
- Introduction to Bias in AI
- Introduction to Fairness in AI
- Introduction to Privacy in AI
- Investigate Research Topics
- Presentation Topics

- **Key Idea:** AI systems depend on vast amounts of personal and behavioural data. Without strong privacy safeguards, these systems risk exposing, misusing, or unfairly profiling individuals.
- **Points:**
  - **Privacy** = the right to control personal data and how it's used.
  - **AI systems** process sensitive information (e.g., biometric, health, or financial data).
  - Weak privacy protection can lead to **identity exposure, surveillance, and loss of autonomy.**
  - **Fairness and privacy** are not opposing goals – both protect human dignity and trust in AI.
- **Example:**  
Facial recognition tools store images without consent → privacy violation and risk of **discriminatory misuse.**



# The Fairness-Privacy Trade-off

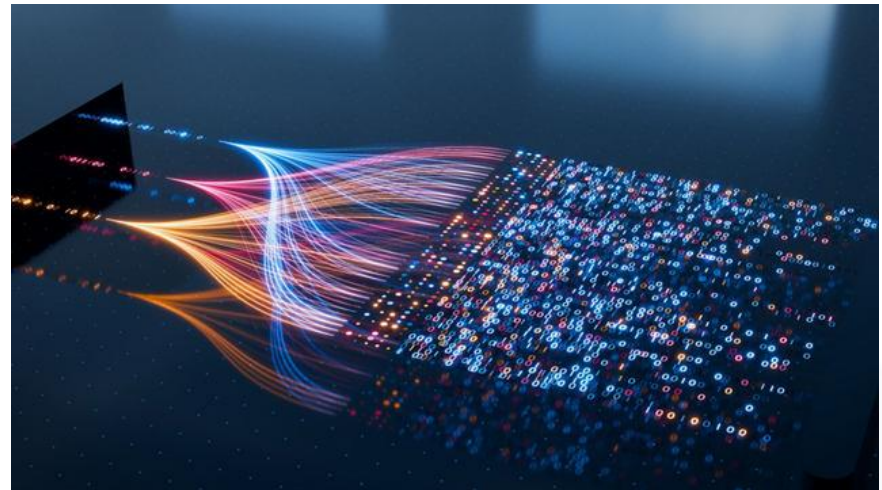
- Efforts to improve fairness sometimes conflict with privacy – achieving both requires balance.

Fairness Goal	Privacy Concern
Collecting more demographic data (e.g., gender, race) to test fairness	Increases risk of <b>re-identification</b> and <b>privacy breaches</b>
Using differential privacy (adding noise to data)	May reduce <b>accuracy</b> and mask unfair patterns
Fairness audits using sensitive data	Requires <b>explicit consent</b> and <b>data protection measures</b>

- Achieving fairness often requires using sensitive data responsibly – not avoiding it entirely.

# Common Privacy Issues in AI

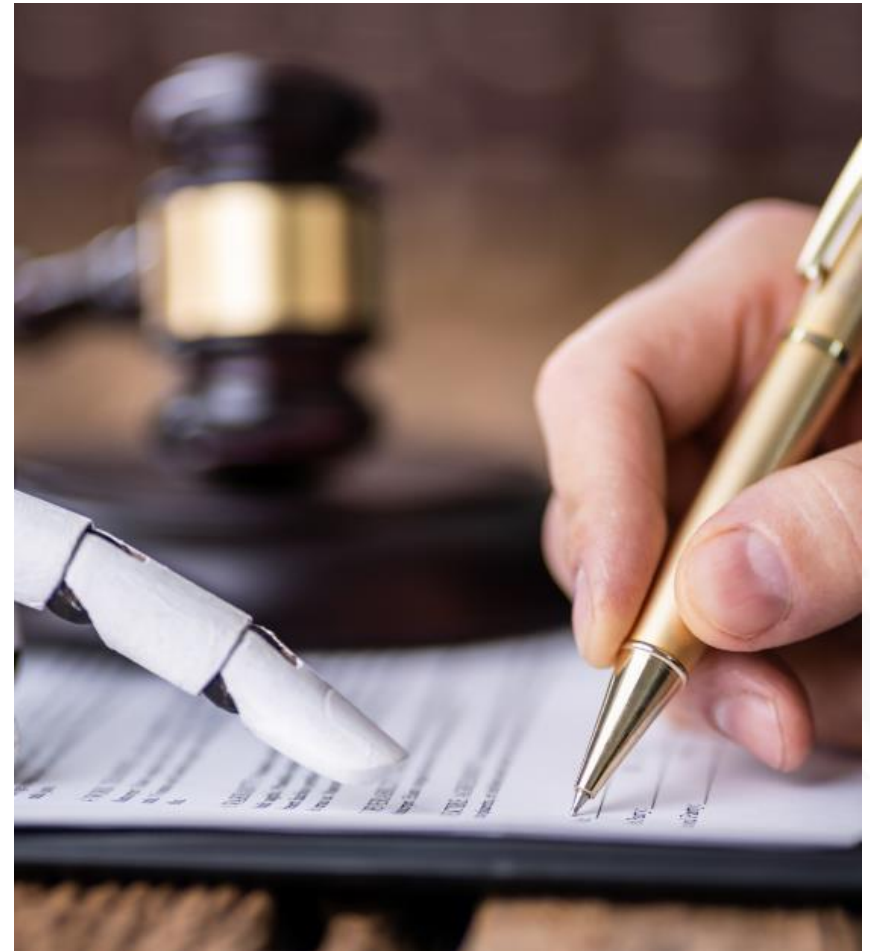
- **Data Collection Without Consent**
  - Many AI datasets are scraped or compiled without individuals' informed permission.
  - Example: Face recognition datasets from social media platforms.
- **Re-identification Risks**
  - Even anonymized data can be **re-linked** to individuals using external information.
- **Data Overreach and Surveillance**
  - Continuous monitoring in workplaces, schools, or smart cities blurs ethical boundaries.
- **Lack of Transparency**
  - Individuals often don't know how their data is used, shared, or monetized.



- **Key Principle:**  
Privacy protection and fairness assurance should operate as **complementary pillars** of trustworthy AI.
- **Best Practices:**
  - **Data Minimization:** Collect only what is necessary.
  - **Differential Privacy:** Add statistical noise to preserve anonymity.
  - **Federated Learning:** Train models without centralizing personal data.
  - **Transparency Reports:** Disclose how data are processed, stored, and deleted.
  - **Ethical Audits:** Regularly evaluate systems for privacy leaks and fairness outcomes.
- **Takeaway Message:**
  - “Fair AI is not only unbiased – it must also be *private, transparent, and accountable.*”

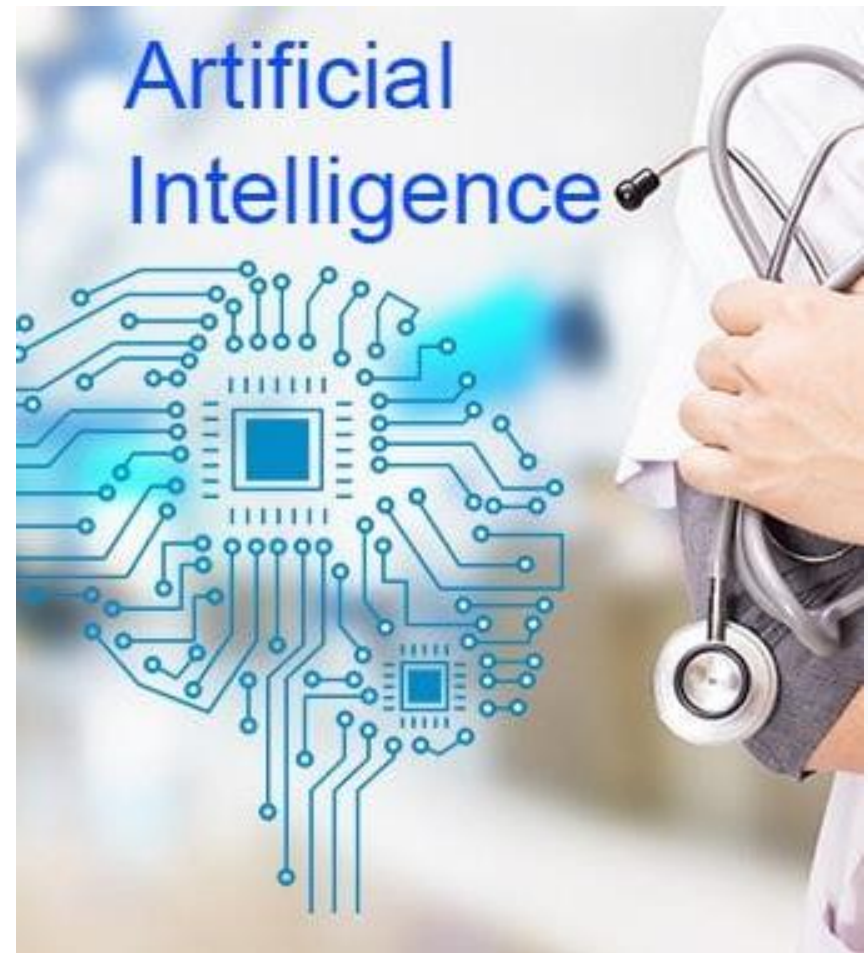
## *COMPAS Algorithm (USA)*

- Predictive policing and recidivism assessment systems **overestimated the risk of reoffending for African-American defendants.**
- Root Cause: **Historical and data bias** in criminal records used for model training.
- **Impact:**
  - Reinforced **racial discrimination** in sentencing and bail decisions.
  - Led to **wrongful risk assessments** and public mistrust in algorithmic justice.



## *Clinical Risk Prediction Models*

- A hospital algorithm used healthcare costs as a proxy for health needs, **underestimating risks for Black patients** who historically receive less care.
- Root Cause: **Measurement bias** and **proxy variable misrepresentation**.
- **Impact:**
  - **Unequal access** to critical care and chronic disease management.
  - **Reinforcement of structural inequities** in medical data.



## *Amazon Recruitment AI*

- The hiring algorithm **penalized résumés containing the word “women’s”**, trained on data from male-dominated tech roles.
- Root Cause: **Sampling bias and historical gender imbalance.**
- **Impact:**
  - **Discriminatory hiring outcomes** disadvantaging women.
  - **Reduced workplace diversity** and violation of fairness standards.



## *AI-Driven Credit Scoring Systems*

- Credit and lending models **systematically downgraded applicants** from minority and low-income backgrounds.
- Root Cause: **Reporting bias** and **group attribution bias** in financial history data.
- **Impact:**
  - **Exclusion from credit markets** for vulnerable groups.
  - **Widening economic inequality** and systemic discrimination.



# Mitigation Strategy: Pre-processing (Data Level Bias Mitigation)

- Approach: Identify and address bias before training the model (e.g., oversampling, undersampling, synthetic data generation).
- Examples:
  - Oversampling darker-skinned individuals in a facial recognition dataset.
  - Data augmentation to boost representation of under-represented groups.
  - Adversarial debiasing to improve model resilience
- Limitations & Challenges:
  - Time-consuming process.
  - May not fully correct bias if original data is heavily skewed.
- Ethical Considerations:
  - **Risk of over- or under-representing** groups, possibly creating new biases.
  - **Privacy concerns** in collecting or augmenting data for vulnerable groups



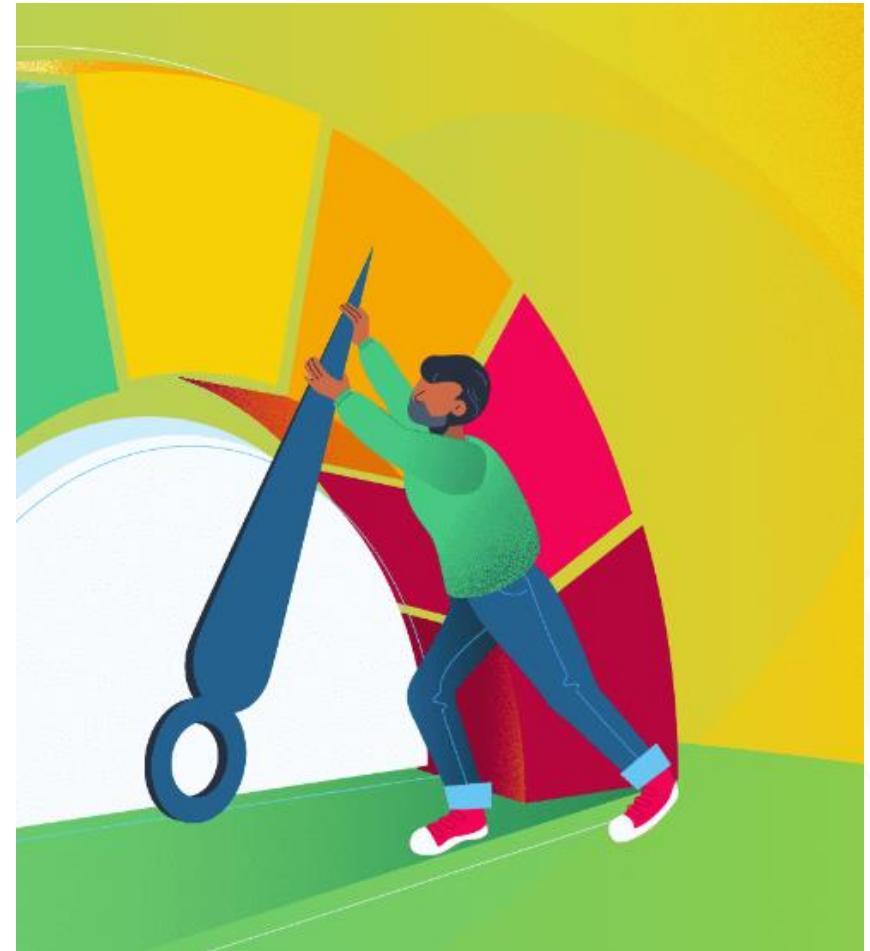
# Mitigation Strategy: Model Selection (Algorithmic Fairness Method )

- **Approach:**
  - Focus on **selecting models** that explicitly prioritize fairness.
  - Methods include **group fairness, individual fairness, regularization, and ensemble approaches** that combine models to minimize bias.
- **Examples:**
  - Classifiers achieving **demographic parity**
  - Model selection based on **group or individual fairness**
  - Regularization to **penalize discriminatory predictions**
  - Ensemble models to **reduce aggregate bias**
- **Limitations & Challenges:**
  - Lack of consensus on what defines “fairness”
  - Trade-offs between **fairness, accuracy, and efficiency**
- **Ethical Considerations:**
  - Choosing fairness criteria may reinforce stereotypes
  - Balancing fairness objectives with performance demands is ethically complex



# Mitigation Strategy: Post-Processing (Decision-Level)

- **Description:**
  - Adjust model **outputs after prediction** to remove or reduce bias.
  - Commonly used to achieve **equalized odds**—ensuring false positives/negatives are equally distributed across groups.
- **Limitations & Challenges:**
  - Computationally complex; may require **large additional datasets**
  - Adjustments can inadvertently affect prediction accuracy
- **Ethical Considerations:**
  - **Trade-offs** between different fairness goals (e.g., equalized odds vs. predictive parity)
  - Potential for **unintended outcome shifts** that disadvantage some groups



- Organizational Information
- Introduction to Bias in AI
- Introduction to Fairness in AI
- Introduction to Privacy in AI
- Investigate Research Topics
- Presentation Topics

- ⌘ Surveys scholarly sources on a specific topic
- ⌘ Provides an overview of current knowledge
- ⌘ Points out gaps in existing research
- ⌘ Appears as part of a dissertation or on its own

# Purpose of the literature review

Demonstrate familiarity with the topic and scholarly context

Develop a theoretical framework and methodology

Position your approach in relation to other researchers

Show how your research fits in

# Conducting a literature review: 5 steps

- Step 1: Search for relevant literature
- Step 2: Evaluate & Select sources
- Step 3: Identify themes, debates and gaps
- Step 4: Outline your structure
- Step 5: Write

## AI in decision making

- Bias & Fairness issues in AI
  - Bias & Fairness in automated hiring system

**How does algorithmic bias in AI-driven hiring systems affect the employment opportunities of women and minority applicants?**

# Identifying keywords:

Key Concept	Synonyms / Related Terms (for database searching)
<b>Algorithmic Bias</b>	data bias, model bias, machine learning bias, automated decision-making bias, discrimination in AI
<b>AI-Driven Hiring Systems</b>	automated recruitment, algorithmic hiring, HR analytics, AI recruitment tools, algorithmic screening, resume-filtering algorithms
<b>Employment Opportunities</b>	hiring outcomes, job selection process, candidate evaluation, recruitment fairness, labor market access
<b>Women and Minority Applicants</b>	gender bias, racial bias, protected groups, diversity in hiring, underrepresented groups, demographic disparities

- ⌘ Use boolean operators (**AND, OR, NOT**)
- ⌘ Read abstracts
- ⌘ Check bibliographies for more sources
- ⌘ Note recurring citations

- ⌘ Our university's library catalogue
- ⌘ [Google Scholar](#)
- ⌘ [JSTOR](#)
- ⌘ [EBSCO](#)
- ⌘ [IEEEExplore](#)
- ⌘ [Springerlink](#)

## Step 2: Evaluate and select sources

What question is addressed?

What are the key concepts?

What are the key theories and methods?

What are the results and conclusions?

How does it relate to other studies?

What are the key insights and arguments?

What are the strengths and weaknesses of the research?

& Quotes

& Summaries of key points

& Source information:

- Author name
- Title & journal name
- Year of publication
- Page numbers

What to look for:

- & Trends in the literature over time
- & Key themes
- & Debates and disagreements
- & Pivotal publications
- & Research gaps

## Examples of trends and gaps

- ✓ **Most research focuses on technical bias detection and fairness metrics**
    - e.g., accuracy comparisons, statistical parity, equalized odds
  - ✓ **Growing interest in fairness-aware machine learning in recruitment**
    - Increased use of HR analytics and automated screening tools
  - ✗ **Less research on *real-world employment outcomes* for affected groups**
    - Few studies examine how AI decisions actually impact women and minorities in hiring practice
  - ✗ **Limited integration of *privacy concerns* with fairness in hiring algorithms**
    - Especially when demographic data is required to assess bias
- ➔ **This opens a meaningful research gap:**
- *Investigating how algorithmic hiring systems shape actual employment experiences and opportunities for women and minority applicants.*

## Common structures

- Chronological: Organize by time
- Thematic: Organize by theme
- Methodological: Organize by methodology
- Theoretical: Organize by theoretical approach

## Format of a literature review

1. Introduction establishing purpose
2. Body analysing the literature
3. Conclusion summarizing key findings

### **Stand-alone literature review:**

- Provide background on the topic
- Describe the objectives of the literature review

### **Dissertation, thesis, or research paper:**

- Reiterate the central problem
- Briefly summarize the scholarly context

- May be divided into sections
- Analyze and interpret
- Critically evaluate
- Synthesize different sources
- Use well-structured paragraphs
- Cite your sources

### **Stand-alone literature review:**

- Discuss the overall implications
- Make suggestions for future research

### **Dissertation, thesis, or research paper:**

- Show how the literature review has informed your approach
- State what gaps your research will address

- Organizational Information
- Introduction to Bias in AI
- Introduction to Fairness in AI
- Introduction to Privacy in AI
- Investigate Research Topics
- Presentation Topics

#	Topic Title	Presentation Focus
1	<b>Sampling Bias in AI Models</b>	When datasets are not representative (e.g., facial recognition misclassification).
2	<b>Algorithmic Bias from Historical Patterns</b>	How models reproduce past inequalities (e.g., Amazon Hiring AI).
3	<b>Confirmation Bias in AI System Design</b>	How developer assumptions influence outcomes.
4	<b>Measurement Bias in Data Collection</b>	Inaccurate data inputs leading to unfair predictions (e.g., Healthcare risk scoring).
5	<b>Generative Bias in Language and Image Models</b>	Cultural dominance and missing perspectives in generative AI.
6	<b>Reporting Bias in User-Generated Data</b>	Imbalance in review/sentiment datasets affecting trend analysis.
7	<b>Automation Bias in Human-AI Decision Making</b>	Over-reliance on AI outputs in high-stakes decisions.
8	<b>Group Attribution Bias in Classification Systems</b>	Stereotype-based predictions affecting individuals (e.g., hiring/credit scoring).

#	Topic Title	Presentation Focus
9	Group Fairness (Demographic Parity and Equalized Odds)	Equality of outcomes across demographic groups.
10	Individual Fairness (Similarity-Based Fairness)	Ensuring similar individuals are treated similarly.
11	Counterfactual Fairness	Testing whether the decision would change if a protected attribute changed.
12	Procedural Fairness & Transparency	Ensuring fairness in the <i>decision-making process</i> itself.
13	Causal Fairness	Removing biased causal influences in prediction pathways.
14	Intersectional Fairness	Protecting individuals with <i>multiple</i> marginalized identities.

#	Topic Title	Presentation Focus
15	<b>Data Collection Without Consent &amp; Data Ownership</b>	Ethical limits of scraping and dataset creation.
16	<b>Re-identification Risks in “Anonymized” Data</b>	Why anonymization is often insufficient.
17	<b>AI Surveillance in Workplaces, Schools, and Cities</b>	Autonomy, monitoring, and power asymmetries.
18	<b>Fairness-Privacy Trade-off and Responsible Data Governance</b>	How fairness goals may require private demographic data, and how to balance.

- Choose any 3 topics from above and mail until 04.11.2025 with one research questions dedicated to each topic.
- I will choose and assign the topic to you.
- Your task is to write a stand-alone literature review.

- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15.
- Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1), eaao5580.
- Fabris, A., Baranowska, N., Dennis, M. J., Graus, D., Hacker, P., Saldivar, J., Zuiderveen Borgesius, F., & Biega, A. J. (2025). Fairness and bias in algorithmic hiring: A multidisciplinary survey. *ACM Transactions on Intelligent Systems and Technology*, 16(1), 1–54.
- Fazelpour, S., & Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, 16(3), e12760.
- Fioretto, F., Tran, C., Van Hentenryck, P., & Zhu, K. (2022). Differential privacy and fairness in decisions and learning tasks: A survey. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, 5470–5477.
- Oduro, S., Moss, E., & Metcalf, J. (2022). Obligations to assess: Recent trends in AI accountability regulations. *Patterns*, 3(11), 100608.
- Singhal, A., Neveditsin, N., Tanveer, H., & Mago, V. (2024). Toward fairness, accountability, transparency, and ethics in AI for social media and health care: A scoping review. *JMIR Medical Informatics*, 12, e50048.
- Verma, S., & Rubin, J. (2018). Fairness definitions explained. In S. Nefti-Meziani & J. Muteba (Eds.), *ACM Computing Surveys*, 15(3), 1–25.
- Ferrara, E., Romero, S., & Carvalho, G. (2023). Fairness, bias, and transparency in AI systems: A survey. *arXiv preprint arXiv:2304.05881*.
- Innis, H. A. (1949). The bias of communication. *The Canadian Journal of Economics and Political Science*, 15(4), 457–476.
- Hrachovec, C. (2009). Binding time: Harold Innis and the balance of new media. *Conference paper on media theory*.
- Ugwudike, P. (2022). Predictive algorithms in justice systems and the limits of tech-reformism. *International Journal for Crime, Justice and Social Democracy*, 11(1), 85–99.