

ARTIFICIAL INTELLIGENCE

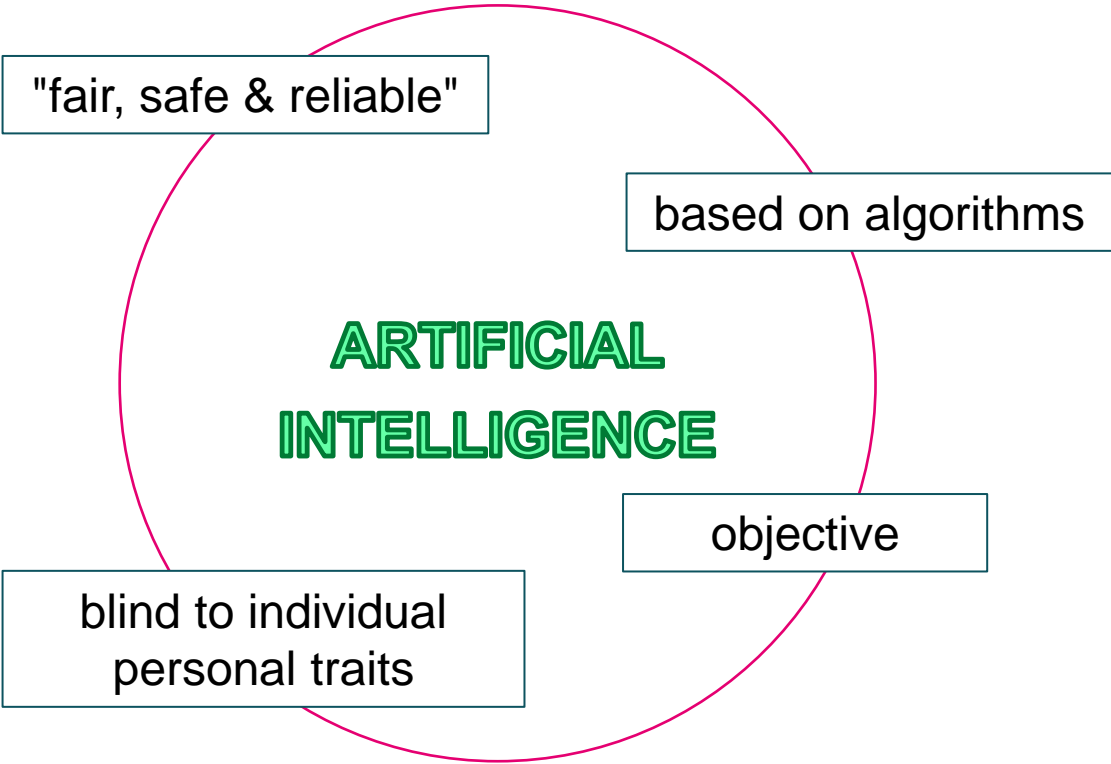
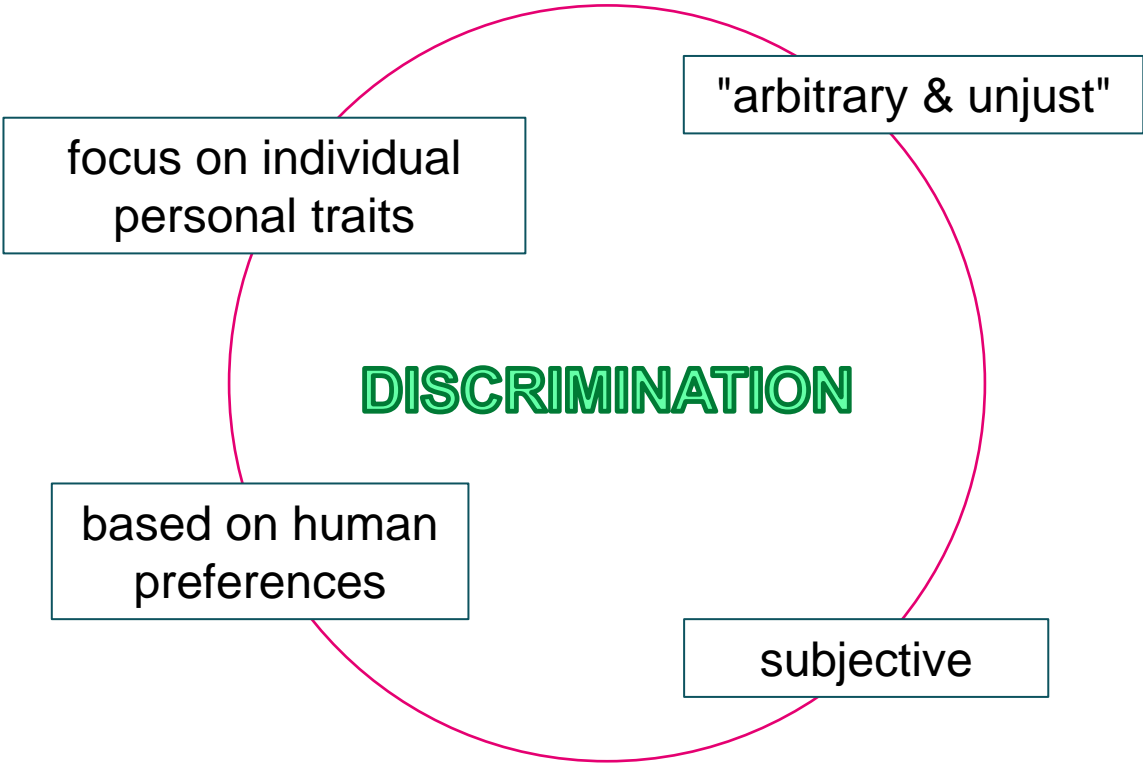
BIAS & DISCRIMINATION

AGENDA

- ❖ A LACK OF OVERLAP?
- ❖ A TASTE OF BIAS
- ❖ NON-DISCRIMINATION LAW IN THE EU
- ❖ AI TOOLS IN HR & (KNOWN) DISCRIMINATION
- ❖ RECOMMENDATIONS

a lack of overlap?

at first glance



AGENDA

❖ THE OVERLAP

❖ A TASTE OF BIAS

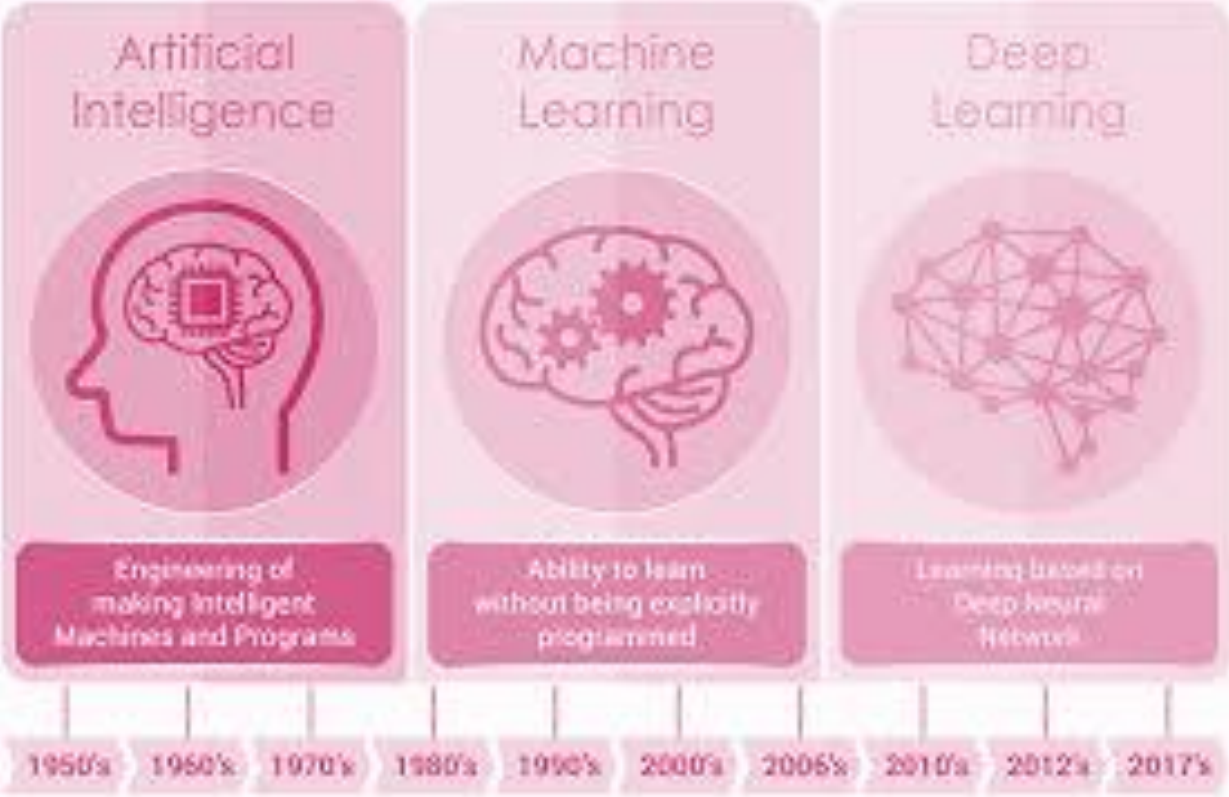
❖ NON-DISCRIMINATION LAW IN THE EU

❖ AI TOOLS IN HR & (KNOWN) DISCRIMINATION

❖ RECOMMENDATIONS

a taste of bias

how ai learns



a taste of bias

how ai learns

Rule-based learning



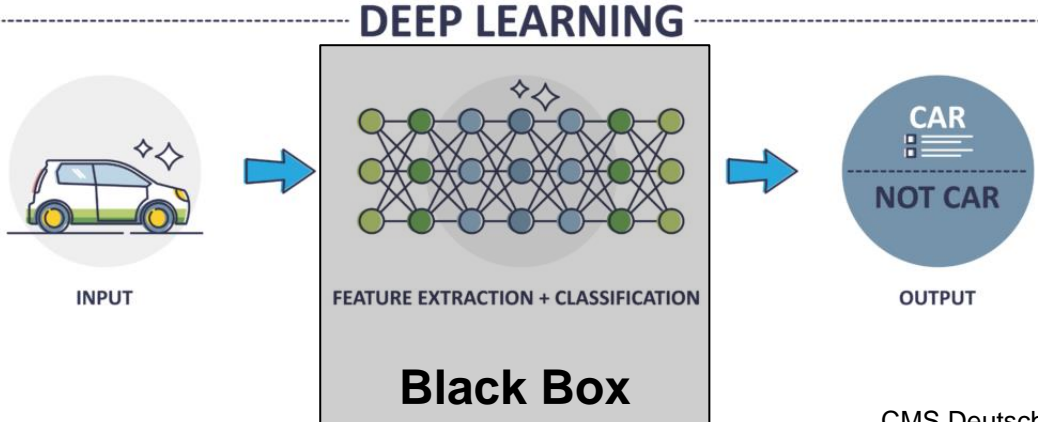
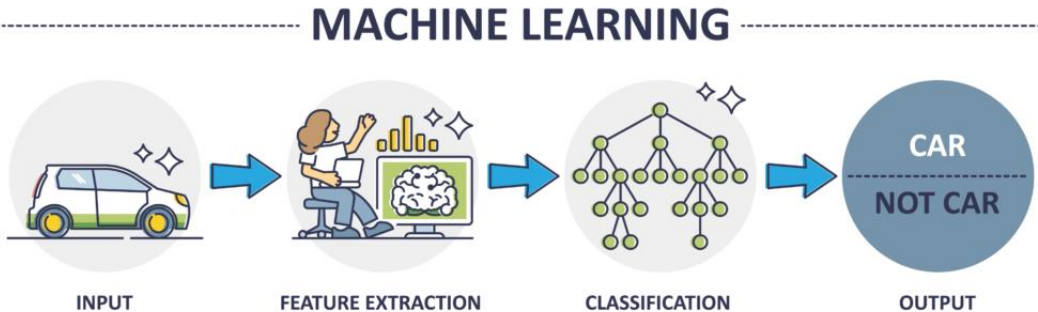
4 wheels = car



2 wheels = bike



default = human?



a taste of bias

how ai learns

Journalistic investigative research has shown that Kenyan workers had to watch violence and racism for ChatGPT to learn:



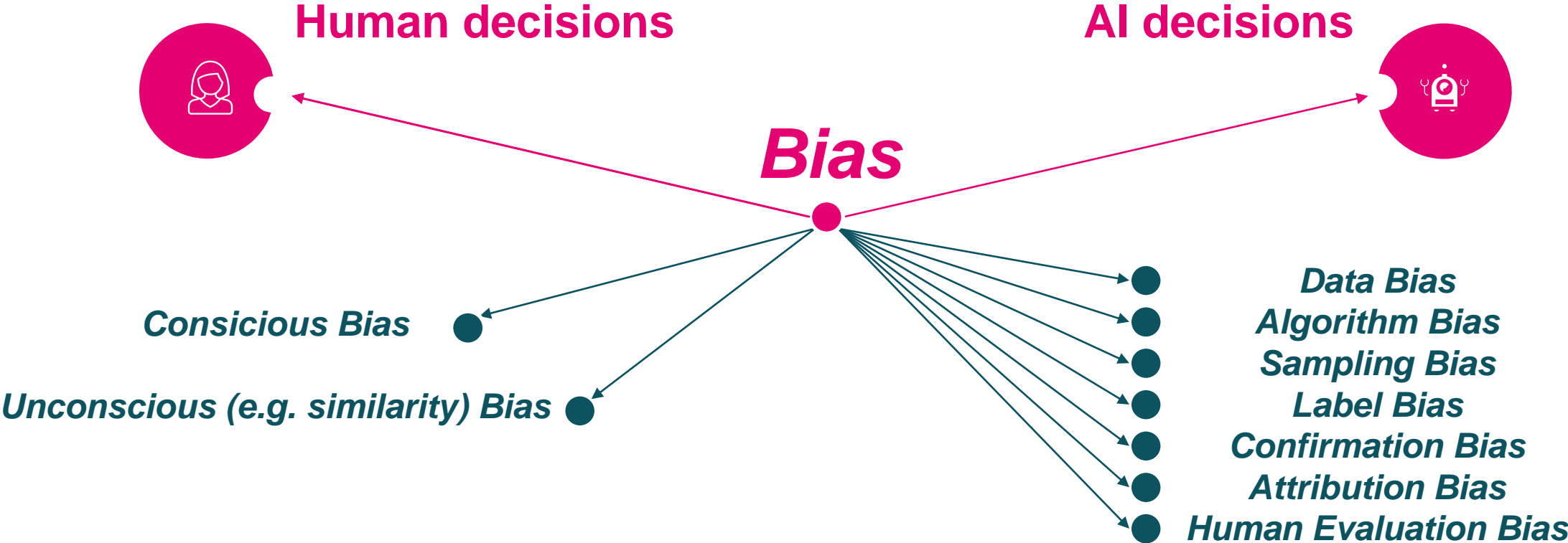
OpenAI commissioned a company in Kenya to have their employees read and evaluate thousands of text excerpts with brutal descriptions of violence and abuse. Thousands of text clippings were sent to the company, which is based in San Francisco but employs click workers in a satellite office in Kenya who annotate artificial intelligence training data.

The goal: to generate training data for ChatGPT so that the model can better recognize problematic content.

Not only were they poorly paid - they received between 1.32 and two dollars per hour - but they also suffered from the psychological consequences of the traumatizing work.

a taste of bias

your ideal decision maker?



a taste of bias

arbitrary correlation

AI promises to discover patterns
that are not visible for humans by
means of Data Mining



Suspicious (criminal) behaviour

Social Scoring

Individual Online Pricing



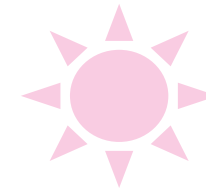
AI can see that

a high demand for sunscreen

often correlates with

a low demand for coal

**But: AI is not able to explain the
causality between findings**



AGENDA

❖ THE OVERLAP

❖ A TASTE OF BIAS

❖ NON-DISCRIMINATION LAW IN THE EU

❖ AI TOOLS IN HR & (KNOWN) DISCRIMINATION

❖ RECOMMENDATIONS

non-discrimination law in the EU

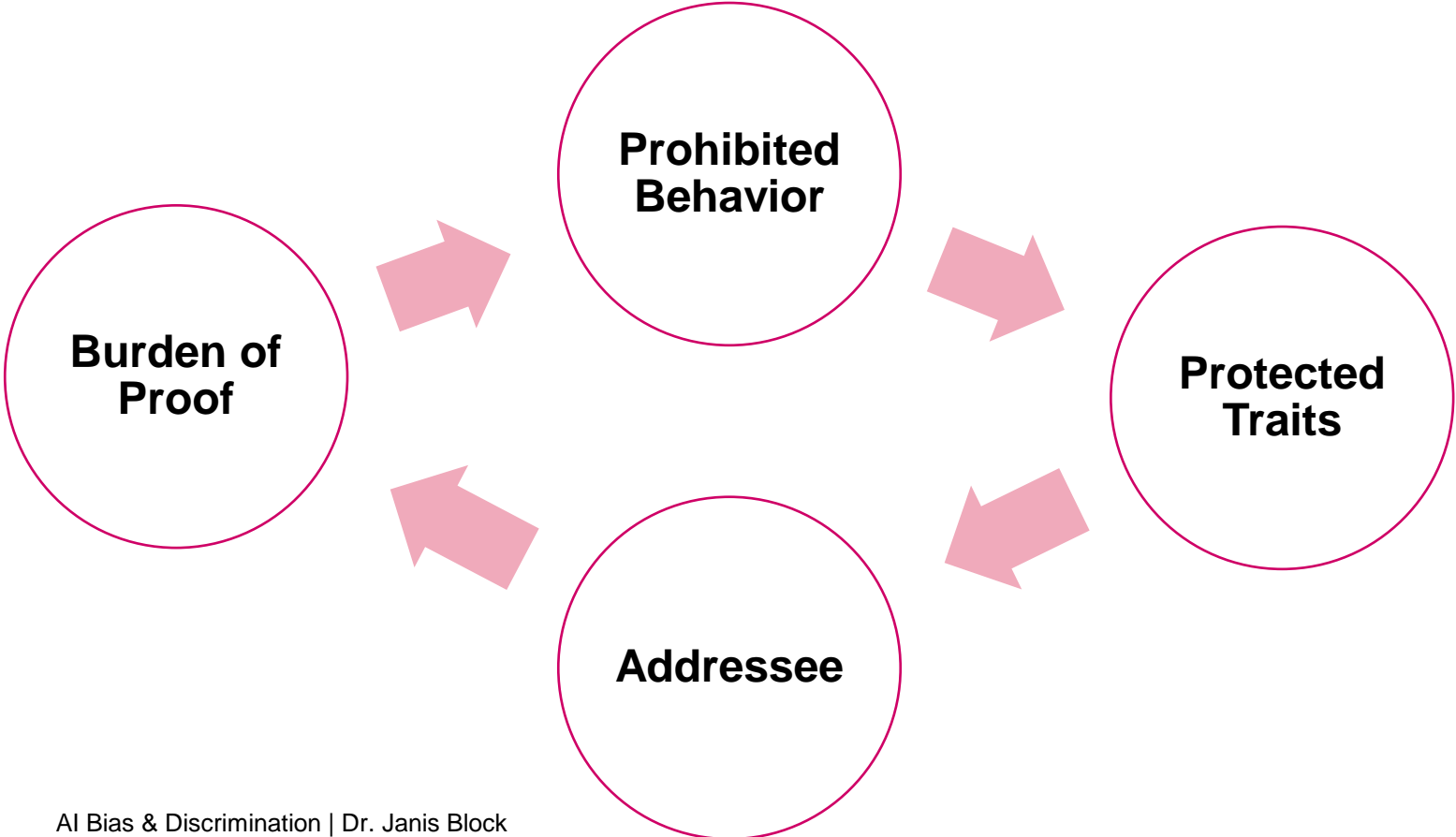
eu directives - old news?



- 2000/43** implementing the principle of equal treatment between persons irrespective of **racial or ethnic origin**
- 2000/78** establishing a **general framework** for equal treatment in **employment and occupation**
- 2002/73** implementation of the principle of equal treatment for **men and women** as regards access to **employment, vocational training and promotion, and working conditions**
- 2004/113** implementing the principle of equal treatment between **men and women** in the **access to and supply of goods and services**

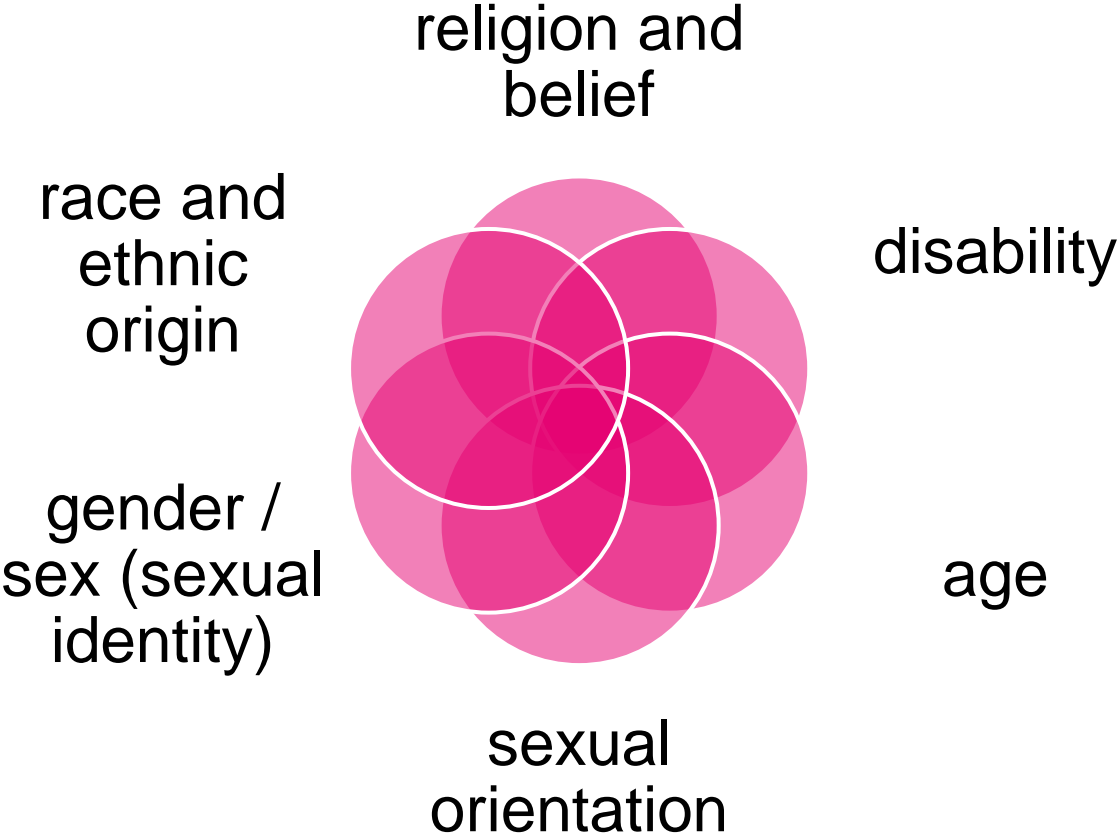
non-discrimination law in the EU

basics



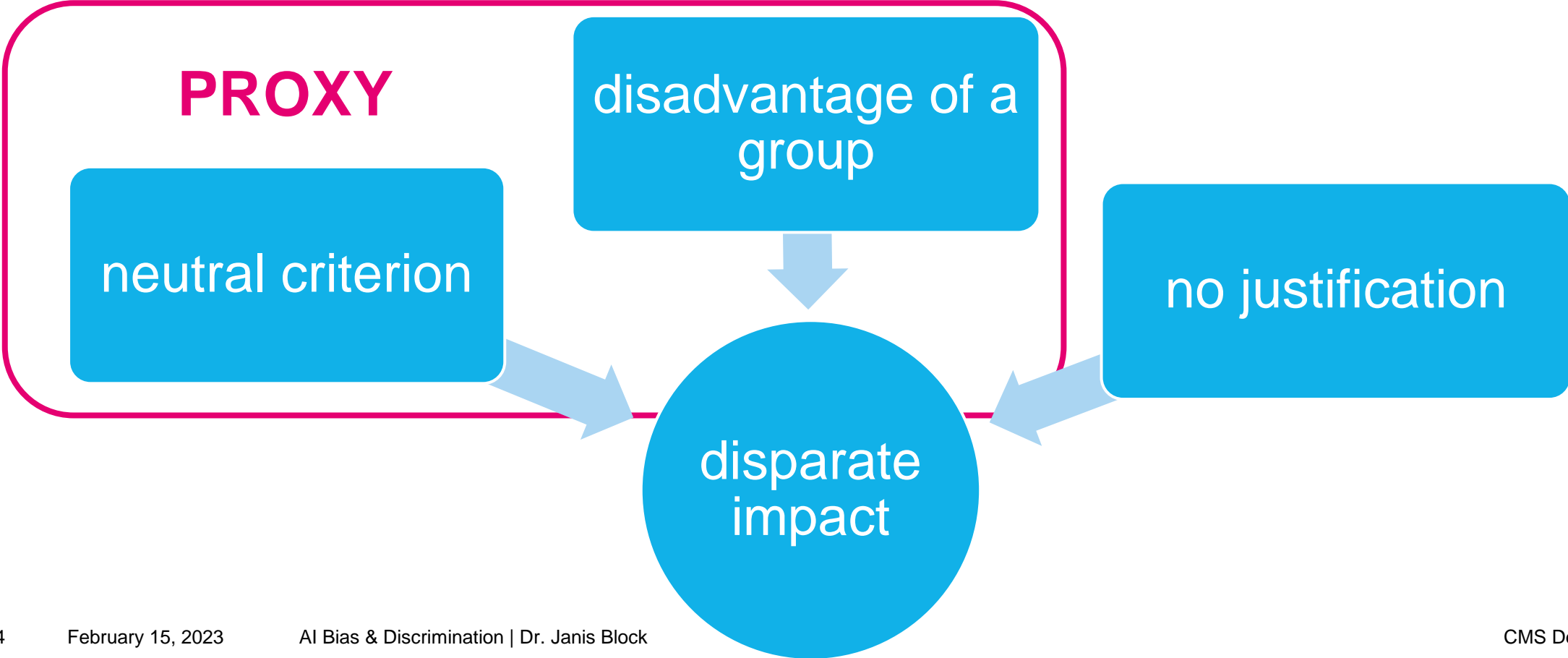
non-discrimination law in the EU

protected traits



non-discrimination law in the EU

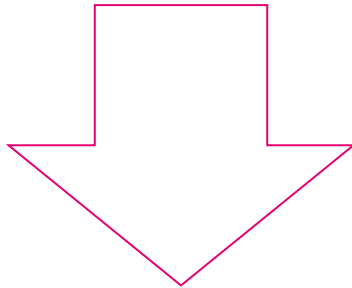
indirect discrimination



non-discrimination law in the EU

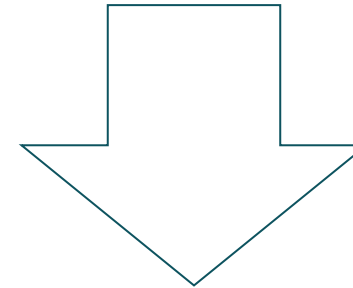
addressee

With regards to all protected traits



Employers

With regards to racial & ethnic origin and sex/gender



everyone who provides access to and supply of goods and services which are available to the public, including housing

non-discrimination law in the EU

burden of proof



prima facie proof: establish facts from which it may be presumed that there has been direct or indirect discrimination

proof that there has been no breach of the principle of equal treatment

non-discrimination law in the EU

arbitrary correlation vs. indirect discrimination

Discrimination requires a causal connection between the unequal treatment and a protected trait. However, it is not necessary that the relevant trait is the exclusive or even only a significant motive for the discriminating party's actions.

AI does not need to directly work with protected traits

Statistical data can indicate that there is a causal connection, not only a correlation

BUT: AI output alone is not enough, as the linking proxy is missing

non-discrimination law in the EU

artificial intelligence vs. human accountability

Employer

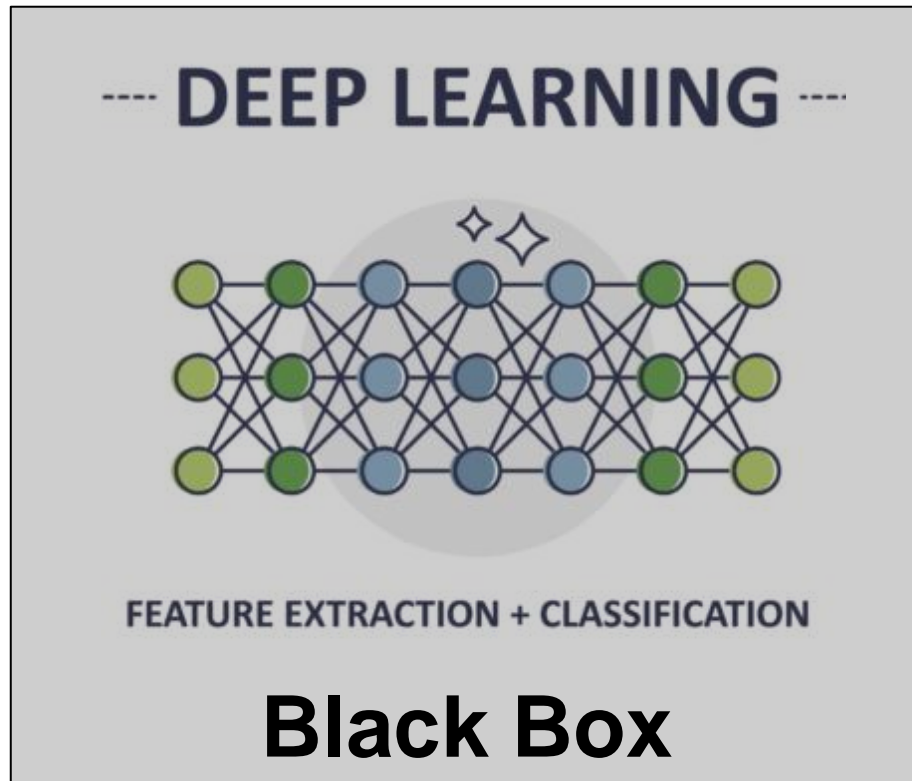


Employer

non-discrimination law is technology neutral

non-discrimination law in the EU

black box vs. burden of proof



prima facie proof: establish facts from which it may be presumed that there has been direct or indirect discrimination

non-discrimination law in the EU

solutions de lege lata

AS OF TODAY, IT IS TECHNICALLY AND LEGALLY ALMOST IMPOSSIBLE TO PROVE AI BASED DISCRIMINATION DUE TO LACK OF KNOWLEDGE THAT AI IS BEING USED IN THE FIRST PLACE AND/OR LACK OF COMPREHENSION, HOW AI TOOK A DECISION

AGENDA

❖ THE OVERLAP

❖ A TASTE OF BIAS

❖ NON-DISCRIMINATION LAW IN THE EU

❖ AI TOOLS IN HR & (KNOWN) DISCRIMINATION

❖ RECOMMENDATIONS

ai tools in hr & (known) discrimination

use of ai in the workplace

Selection and planning decisions

Recruiting, compensation, promotion, people analytics, vacation, shift schedules, workforce planning



AI supports supervisor

Instructions by AI

Assistance systems, "pick-by-voice" order picking



AI replaces supervisor

AI as a workforce

Cobot, chatbot, human-machine interaction



AI as a colleague

ai tools in hr & (known) discrimination

use of ai in the workplace

Hiring process

People Analytics, assessment of candidates/selection of applicants, AI assistance in interview process, designing efficient hiring process

Remuneration decisions

Recommendations for performance-based bonuses, payments for achieved results

Workforce planning

Decisions on requests and overall planning

ai tools in hr & (known) discrimination

use of ai in the workplace

Appraisal decisions

Performance appraisal, risk assessments, CPD recommendations

Compliance

Monitoring, investigation and assessment of breaches of duty

Predictive Analytics

Analysis of data on individual employees for predictions

ai tools in hr & (known) discrimination

overview of service providers for ai in hr



retorio

video & audio analysis



SAP & Nestlé

employee chatbot
Interface HR data base



zortify

matching & AI based
decision making



HireVue

U.S. market leader
audio analysis



Moveworks

virtual assistant for
employees



100 Worte

job postings

fountain

automated hiring



IBM Watson Talent

multifunctional solution,
chatbot, automation of
HR processes, hiring

ai tools in hr & (known) discrimination

examples of discrimination – speech analytics

- Used in the hiring process to **find suitable applicants**
- Interview by AI system to **determine personality traits**
- **Prosodic** and **linguistic** analysis

ISSUES

- Downgrading of applicants with slow speaking pace (disability) or use of language (ethnic origin)

ai tools in hr & (known) discrimination

examples of discrimination – rankings and ratings

ISSUES

- Search algorithm variables inferred by labeling profile images and queries
- Women receive significantly fewer reviews
- POC suppliers receive worse ratings than people of Caucasian or Asian American descent

- Suppliers can **offer small services on marketplace**
- Analysis of approx. 3.700 profiles, based on tasks suppliers are willing to complete and **ratings/reviews** they have received from customers
- **No self-reporting on protected traits**, i.e. gender & race

ai tools in hr & (known) discrimination

examples of discrimination – recruiting

- Search for top-talent based on **successfully submitted CV's** over 10 year period
- Search and evaluate applicants' CVs by **analyzing word patterns**
- Candidates received **scores ranking from 1-5 stars**

ISSUES

- Most applications in the reference period came from male candidates
- Technology favored candidates who described themselves with language more commonly associated with men
- System taught itself that male applicants were preferable

ai tools in hr & (known) discrimination

examples of discrimination – jobs ads

Age discrimination in job ads

- Companies personalized audience of their job ads = "Microtargeting"
- Older Facebook users did not see job ads
- Class-action lawsuit in 2017 and settlement in 2019

Gender discrimination in job ads

- Lawsuit in 2018 against Facebook and 10 employing companies
- "Lookalike audience" service
- Settlement in 2019

ai tools in hr & (known) discrimination

examples of discrimination – outside of hr

Predictive Policing

- To increase efficiency of **police patrolling**, algorithms propose routes or areas where there is a high probability that crimes will be committed
- Higher presence of police will detect more criminal activities
- Algorithms is reinforced to believe that this is a social hotspot

Criminal Profiling

- To predict the **recidivism probability** of known offenders, courts use algorithms to determine the appropriate penalty
- 137 variables are evaluated, among others how many "friends/acquaintances" have already been arrested, reprimandments at school or how old the accused was when his parents separated
- For offenders of colour, the prognosis is much worse

AGENDA

❖ THE OVERLAP

❖ A TASTE OF BIAS

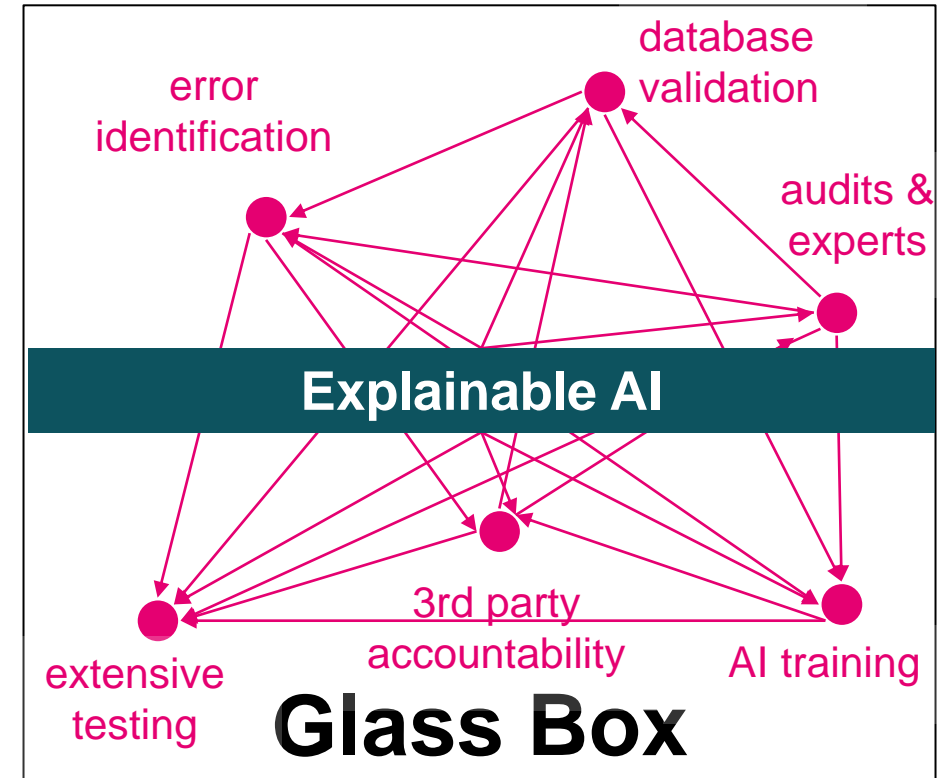
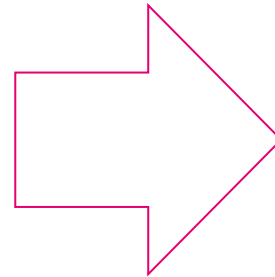
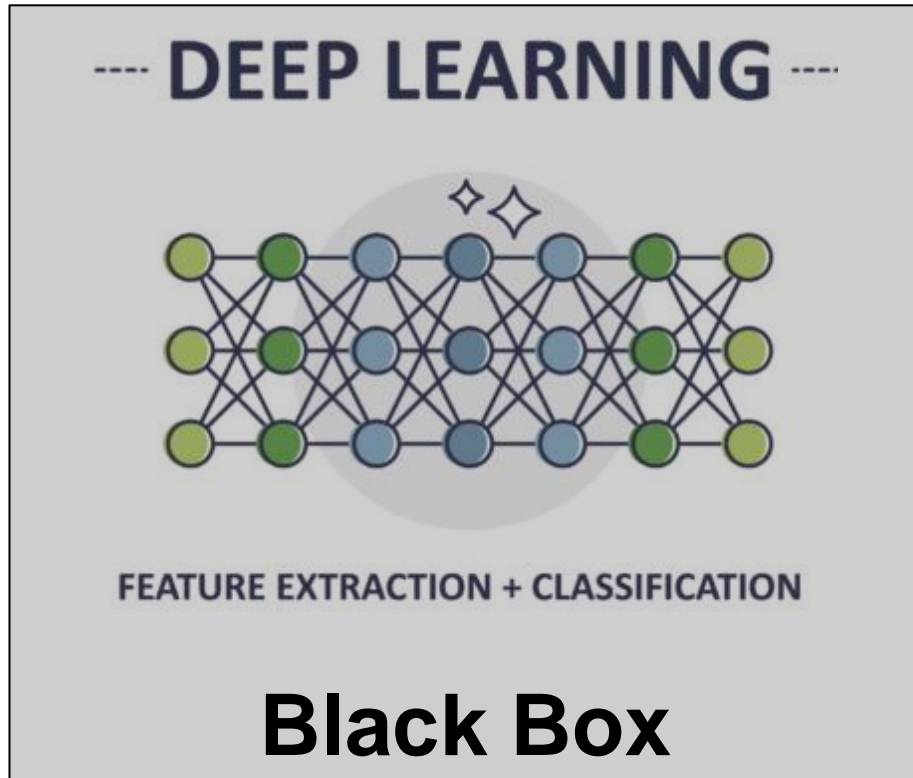
❖ NON-DISCRIMINATION LAW IN THE EU

❖ AI TOOLS IN HR & (KNOWN) DISCRIMINATION

❖ RECOMMENDATIONS

recommendations

what needs to be done



recommendations

what needs to be done

- Any predictions new to be "well calibrated": probability needs to proportionally reflect group representation and weight of single features in used data
- Apply "balance for the positive class" and "balance for the negative class" as correctives to reflect the statistical likelihood that a proxy is actually (not) represented in a group

Thanks!



Dr. Janis Block

Counsel, Attorney at Law

Labor & Employment Law

T +49 160 94627913

E janis.block@cms-hs.com

[in](#) [Dr. Janis Block | LinkedIn](#)