Equality, artificial intelligence, algorithms and automated decisions: how to detect and address discrimination in the context of AI?
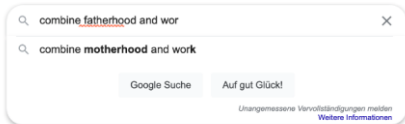
Fabian Lütz

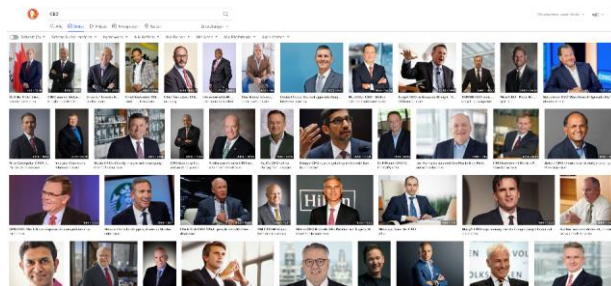ERA (Trier 12-13 June 2023)

This training session is funded under the 'Rights, Equality and Citizenship Programme 2014-2020' of the European Commission.

1

Gender Equality in the Age of Algorithms

Google (10.10.2021)

DuckDuckGo "CEO" (10/10/2021)

ERA (Trier 12-13 June 2023)

2

2

Large Language Models (LLMs): ChatGPT & Co

The problem of Gender Bias and potential discriminations might increase with the use of ChatGPT, GPT-4, Bard etc.

→ General-purpose AI or foundation models are not included in COM proposal (But Council called for GPAI to be considered)

Why does a search for pictures of CEO show only male CEOs ?

There could be multiple reasons why a search for pictures of CEOs predominantly shows male CEOs. One reason is that historically, men have held a majority of CEO positions in many industries. This gender disparity in leadership positions is known as the "glass ceiling" and can be attributed to various factors such as gender biases, societal expectations, and systemic discrimination.

Another reason could be the bias of the search algorithms themselves. If the search algorithm is not programmed to consider gender as a factor in its search results, it may prioritize images of male CEOs simply because they are more prevalent in the data it has been trained on.

It's important to note that this issue has been recognized and efforts are being made to improve diversity and representation in leadership positions. As more women and individuals from underrepresented groups break through the glass ceiling, we may see a shift in the search results as well.

NEWS

ChatGPT cited 'bogus' cases for a New York federal court filing. The attorneys involved may face sanctions.

OpenAI's popular chatbot had "hallucinated" – a term for when AI systems simply invent false information – and spat out cases and arguments that were entirely fiction.

https://www.nytimes.com › 2023 › 05 › 27 › nyregion › avianca-airline-lawsuit-chatgpt.html

A Man Sued Avianca Airline. His Lawyer Used ChatGPT. - The New Y...
6 days ago · Mr. Schwartz said that he had never used ChatGPT, and "therefore was unaware of the possibility that its content could be false." He had, he told Judge Castel, even asked the program to verify that...

ERA (TRIER 12-13 JUNE 2023)

3

Agenda

1. *Introduction*

2. *Current and Future EU legal framework*

3. *AI and Gender Equality in Practice*

4. *Regulation of AI and Algorithms*

5. *AI as enabler of Gender Equality ?*

6. *Outlook*

ERA (Trier 12-13 June 2023)

4

Slide 5

Ministerial Declaration
The G7 Digital and Tech Ministers' Meeting
30 April 2023

Responsible AI and Global AI Governance

39. In 2016, we initiated an international discussion on AI principles. This discussion helped pave the way for the 2019 OECD AI Recommendation (OECD AI Principles) and the associated work launching the OECD.AI Policy Observatory and Network of Experts. In 2020, we supported the launch of the Global Partnership on Artificial Intelligence (GPAI).

40. The OECD AI Principles provide guidance for trustworthy AI and for ensuring an open and enabling environment for AI development and deployment that is grounded in human rights and democratic values. Since adoption of these principles, the OECD continues to engage and work with the global AI community to support their implementation.

41. Rapid AI developments call for attention to, and cooperation on, emerging and medium-term policy issues including development of technical standards, developed by international standards development organisations (SDOs), as well as other tools to ensure the development and deployment of trustworthy AI in line with the OECD AI Principles. In this context, we welcome the contributions of existing initiatives on these topics.

42. We reaffirm our commitment to promote human-centric and trustworthy AI based on the OECD AI Principles and to foster collaboration to maximise the benefits for all brought by AI technologies. We oppose the misuse and abuse of AI to undermine democratic values, suppress freedom of expression, and threaten the enjoyment of human rights.

43. We stress the importance of international discussions on AI governance and interoperability between AI governance frameworks, while we recognise that like-minded approaches and policy instruments to achieve the common vision and goal of trustworthy AI may vary across G7 members. Tools for trustworthy AI, such as regulatory and non-regulatory frameworks, technical standards and assurance techniques, can promote trustworthiness and can allow for the comparable assessment and evaluation of AI systems. We support the development of tools for trustworthy AI through multistakeholder international organisations, and encourage the development and adoption of international technical standards in SDOs through private sector-led multistakeholder processes. We commend work to date in the OECD on mapping the commonalities and differences between trustworthy AI frameworks, and we intend to work together to support such work that fosters interoperability.

1. Introduction

- Definitions
- Toronto Declaration
- Recall of Gender Equality Problems
- Update June 2023

ERA (Trier 12-13 June 2023)

5

Slide 6

Definitions: AI and algorithms

Artificial intelligence

- AI system is a software that '(..) for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;(..)'.

  (Art. 3 para. 1, EU AI Act)

- Examples:
  - machine learning (supervised, unsupervised, reinforcement, deep learning), Statistical approaches, search and optimization methods

Algorithm

- "... sufficiently detailed and systematic instruction of action to solve a mathematical problem so that, when implemented correctly, the computer computes the correct output for each correct set of inputs"

  (Prof. Katharina Zweig)

6

ERA (Trier 12-13 June 2023)

6

Definitions: Decision errors

Bias

- Cognitive biases, collective term for systematic decision errors

(Daniel Kahneman, Olivier Sibony, Cass R. Sunstein: Noise, 2021, p. 163f.)

Noise

- Variability of error, chance variability of judgments *or an* invisible tax

(K/S/S: Noise, 2021, pp. 3-5 and D. Kahneman, HBR, 2016, p. 36-43)

Algorithmic discrimination = occurs when automated systems contribute to unjustified different treatment or impacts disfavoring people, for example based on their sex.

E R A (T r i e r 1 2 - 1 3 J u n e 2 0 2 3)

7

Source: https://www.torontodeclaration.org

Toronto Declaration

- Duty of states to protect human rights; this includes ensuring the right to non-discrimination by private sector actors" (Toronto, para. 38)
- Need for oversight of the use of machine learning by the private sector in contexts that present risk of discriminatory outcomes (Toronto, para. 40)
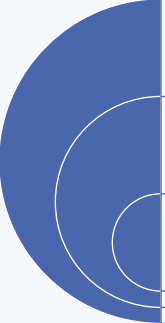
E R A (T R I E R 1 2 - 1 3 J U N E 2 0 2 3)

8

Recalling some problems for Gender Equality (Ex. labour market)

General

- Unequal treatment and discrimination
- Lower participation of women in the labour market
- Vertical and horizontal imbalances
- WLB-Challenge (unequal distribution of care tasks between men and women)

AI specific

- Loss of jobs due to AI
- Many new jobs will be created in *Science Technology Engineering and Mathematics* (STEM) and AI.
- Underrepresentation of women in AI (ca. 21%)

9

9

General Update latest news: June 2023

- RightsCon Summit Costa Rica 5-8 June 2023: Human Rights in Digital Age
- New and emerging technologies need urgent oversight and robust transparency: UN experts (2 June 2023) → gender & discrimination
- UN Secretary-General's Policy Brief on the Global Digital Compact → multi-stakeholder approach (interesting for business and legal advisors)
- Consultation UN Human Rights & on Gender, Tech, and the Role of Business (June 15, 2023)
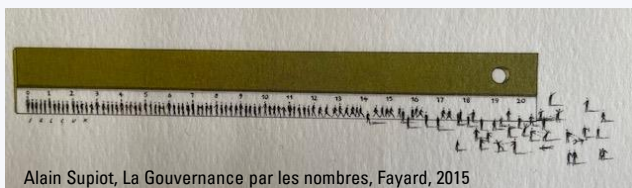
10

10

EU Update latest news: June 2023

- Compromise text of European Parliament (EP) (16 May 2023)
- Vote in Plenary of EP expected 12-15 June 2023 session
- *"Once approved, they will be the world's first rules on Artificial Intelligence"* (EP News, 11 May 2023)
- MEPs introduced complaints about AI systems + receive explanations of decisions based on high-risk AI systems that significantly impact their rights
- Strengthen EU AI Office responsible for monitoring how the AI rulebook is implemented
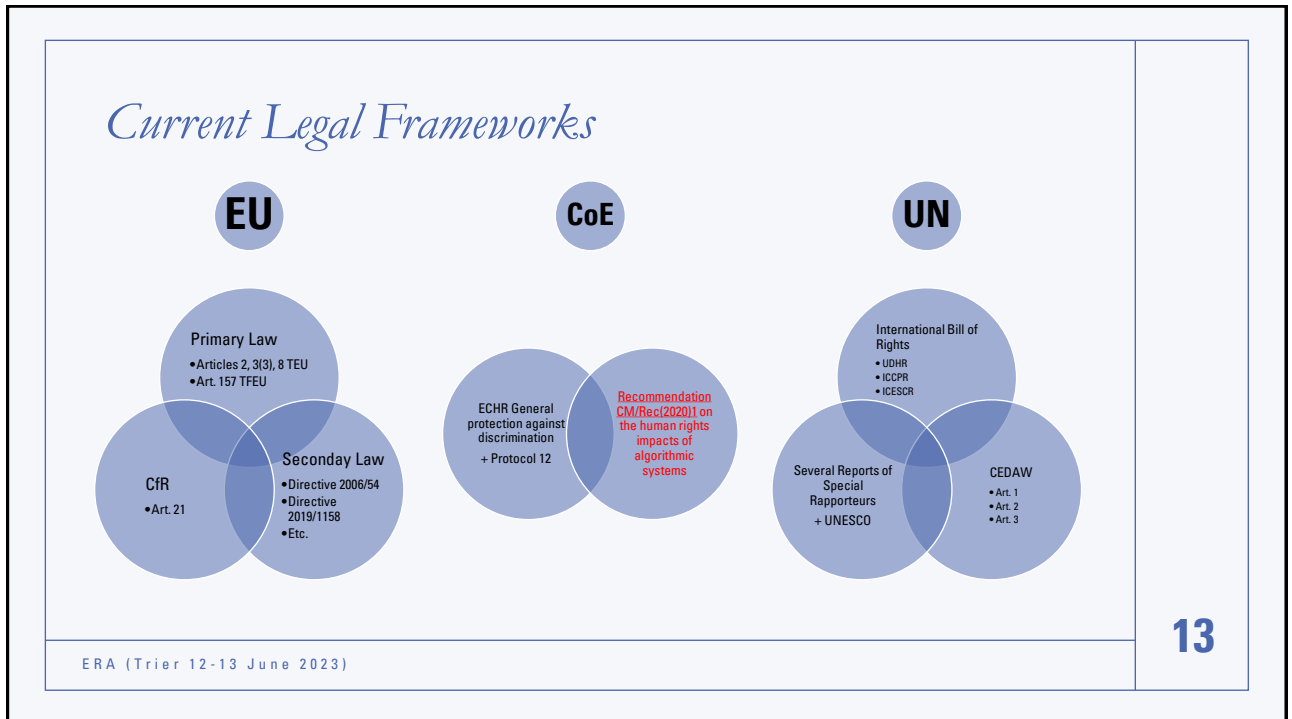
ERA (Trier 12-13 June 2023)

11

11

---

2. Current and Future Legal frameworks

Alain Supiot, La Gouvernance par les nombres, Fayard, 2015

- Applicable law
- Brussels Effect ?
- EU proposal: AI Act

ERA (Trier 12-13 June 2023)

12

Current Legal Frameworks

EU

CoE

UN

Primary Law
• Articles 2, 3(3), 8 TEU
• Art. 157 TFEU

CfR
• Art. 21

Seconday Law
• Directive 2006/54
• Directive 2019/1158
• Etc.

ECHR General protection against discrimination
+ Protocol 12

Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems

International Bill of Rights
• UDHR
• ICCPR
• ICESCR

Several Reports of Special Rapporteurs
+ UNESCO

CEDAW
• Art. 1
• Art. 2
• Art. 3

13

13

Future Legal Framework

EU AI Regulation

- DSA
  - ECAT/JRC Sevilla
- DMA
- AI Act
- EB Standards Directive
  - Recital 16
- AI liability Directive
- Better Regulation/Review of EU Acquis + next mandate 2024-2029)

Council of Europe

- General:
  - CoE Framework Convention on AI, Human Rights, Democracy and the rule of law (adoption November 2023?)
- GE/discrimination:
  - Legal instrument to adress GE/discrimination (until 2025)

UN

- UN
  - Global Digital Compact
- OHCHR
  - Priority, Reports
- CEDAW
  - General Recommandation 40
- CSW
  - Agreed Conclusions 'CSW 67

14

14

15

'Brussels Effect' for regulating algorithmic discrimination

EU at the forefront of shaping the regulation of AI/algorithmic discrimination

- See Anu Bradford (in general+GDPR)
- EU opened an office in San Francisco to follow AI/tech developments (*US creates AI, EU regulates AI*)
- After adoption of DSA/DMA + ECAT, EU likely to be the first major economic region to have far reaching regulatory oversight for AI systems with AI Act

PRESS RELEASE | Publication 05 September 2022

EU opens new office in San Francisco to reinforce its digital diplomacy

The European Union has opened its new office in San Francisco, California, a global centre for digital technology and innovation. The office will reinforce the EU's cooperation with the United States on digital diplomacy and strengthen the EU's capacity to reach out to key public and private stakeholders, including policy makers, the business community, and civil society in the digital technology sector.

High Representative/Vice-President Josep **Borrell**, said:

The opening of the office in San Francisco responds to the EU's commitment to strengthen transatlantic technological cooperation and to drive the global digital transformation based on democratic values and standards. It is a concrete step to further reinforce the EU's work on issues such as cyber and countering hybrid threats, and foreign information manipulation and interference.

The EU office in San Francisco will seek to promote EU standards and technologies, digital policies and regulations and governance models, and to strengthen cooperation with US stakeholders, including advancing the work of the EU-US Trade and Technology Council. Find more information in the EEAS press release.

Related topics
International relations

16

16

Slide 17

Bruxelles, 21.4.2021
COM(2021) final
2021/0106(CO

Proposition de

RÈGLEMENT DU PARLEMENT EUROPÉEN ET DU CONSEIL

ÉTABLISSANT DES RÈGLES HARMONISÉES CONCERNANT L'INTELLIGENC ARTIFICIELLE (LÉGISLATION SUR L'INTELLIGENCE ARTIFICIELLE) ET MODIFIANT CERTAINS ACTES LÉGISLATIFS DE L'UNION

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

ANNEXE III
SYSTÈMES D'IA À HAUT RISQUE VISÉS À L'ARTICLE 6, PARAGRAPHE 2

Les systèmes d'IA à haut risque au sens de l'article 6, paragraphe 2, sont les systèmes d'IA répertoriés dans l'un des domaines suivants:

1. Identification biométrique et catégorisation des personnes physiques:
   (a) les systèmes d'IA destinés à être utilisés pour l'identification biométrique à distance «en temps réel» et «a posteriori» des personnes physiques.

2. Gestion et exploitation des infrastructures critiques:
   (a) les systèmes d'IA destinés à être utilisés en tant que composants de sécurité dans la gestion et l'exploitation du trafic routier et dans la fourniture d'eau, de gaz, de chauffage et d'électricité.

3. Éducation et formation professionnelle:
   (a) les systèmes d'IA destinés à être utilisés pour déterminer l'accès ou l'affectation de personnes physiques aux établissements d'enseignement et de formation professionnelle;
   (b) les systèmes d'IA destinés à être utilisés pour évaluer les étudiants des établissements d'enseignement et de formation professionnelle et pour évaluer les participants aux épreuves couramment requises pour intégrer les établissements d'enseignement.

4. Emploi, gestion de la main-d'œuvre et accès à l'emploi indépendant:
   (a) les systèmes d'IA destinés à être utilisés pour le recrutement ou la sélection de personnes physiques, notamment pour la diffusion des offres d'emploi, la présélection ou le filtrage des candidatures, et l'évaluation des candidats au cours d'entretiens ou d'épreuves;
   (b) l'IA destinée à être utilisée pour la prise de décisions de promotion et de licenciement dans le cadre de relations professionnelles contractuelles, pour l'attribution des tâches et pour le suivi et l'évaluation des performances et du comportement de personnes dans le cadre de telles relations.

EU *Artificial Intelligence Act*
COM (2021), 206 final

Article 2
Scope

1. This Regulation applies to:
(a) providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country;
(b) users of AI systems located within the Union;
(c) providers and users of AI systems that are located in a third country, where the output produced by the system is used in the Union;

Business and Human Rights (Lausanne, 1 June 2023)

17

17

Slide 18

EU COM(2021), 206 final
Artificial Intelligence Act(AIA)

Horizontal regulation based on opportunities and risks of AI → Definitions (Art. 3) → Ban (Art. 5)

High risk AI systems (Art. 6) → Requirements for high-risk AI systems (Art. 8 - 14) → Sanctions (Art. 71)

ERA (Trier 12-13 June 2023)

18

18

EU COM AI Act & Equality

Horizontal instrument for AI

- Proposal complements existing Union law on non-discrimination
- Concrete requirements to minimise the risk of discrimination by algorithms
- Design and quality of datasets used for AI system development.
- Rules for: Testing, risk management, documentation and human oversight throughout the lifecycle of AI systems.

Equality in the AIA proposal

- "(Non)-discrimination" mentioned (16x)
- "Gender Equality" (1x)
- "Women" (x2): Art. 21 CFR + in relation to recruitment systems

ERA (Trier 12-13 June 2023)

19

EU COM Draft AI Act & Recruitment

Risk-based approach

- Article 6 + Article 6(2) + Annex III (No 4 Recruitment and selection AI systems)
  - Relevance to gender equality
  - Danger to fundamental rights ?

Annex III (Recruitment systems)

- "Throughout the recruitment process and in the evaluation, promotion, or retention of persons in work-related contractual relationships, such systems may perpetuate historical patterns of discrimination, for example against women (...)". (Recital 36)

Draft Compromise Amendments of European Parliament (16/05/2023)
Article 4a - General principles applicable to all AI systems
1.e) *"'diversity, non-discrimination and fairness'* means that AI systems shall be developed and used in a way that includes diverse actors and **promotes** equal access, **gender equality** and cultural diversity, **while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law**."

ERA (Trier 12-13 June 2023)

20

EU COM Draft AIA & Annex III

Annex III (Recruitment systems), No. 4

- Employment, workers management and access to self-employment:

(a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;

(b) AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.
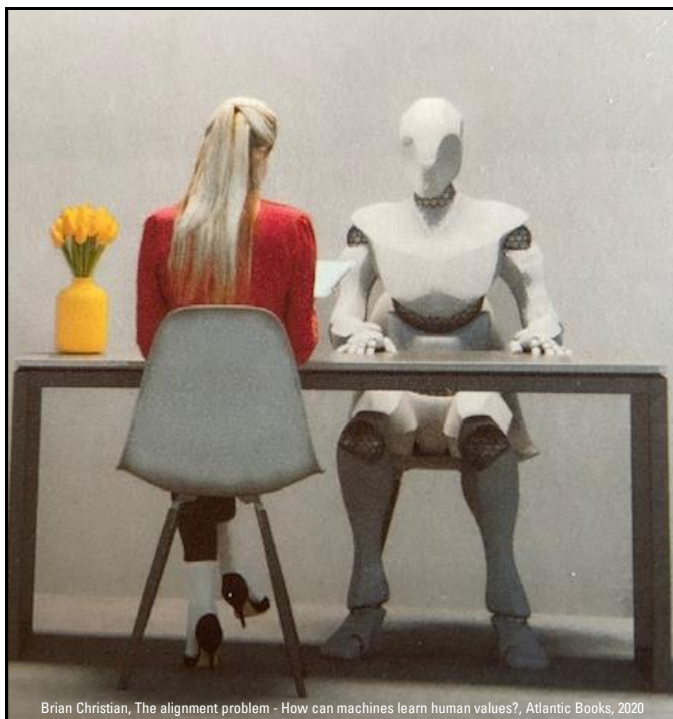
Article 6(2)

- Classification as a high-risk AI system
- Application of the special provisions of Chapter 2 (Articles 8 - 14):
  - Article 9: Risk management system
  - Article 10: Data and data governance
  - Article 11: Technical documentation
  - Article 12: Record-keeping
  - Article 13 Transparency and provision of information to users
  - Article 14: **Human oversight**

21

ERA (Trier 12-13 June 2023)

21

---

3. AI and gender equality in practice

- AI, Biases and Gender Equality
- Public authorities
- Private companies
- Legal Questions
- Case law

Brian Christian, The alignment problem - How can machines learn human values?, Atlantic Books, 2020

ERA (Trier 12-13 June 2023)

22

Gender Equality, AI and Biases

Gender *biases*

- **Word2vec/word vectors:** language, machine learning, frequency, ranking, metaphors, importance of data sets + LLMs

- **Stereotypes and biases:** "Man is to computer programmer as woman is to homemaker?" (see Bolukbasi/Chang et.al. (2016) debiasing word embeddings)

- Men should also be associated with parental leave, home help, part-time…

***Bias* is human**

- **Data** reflects society but also shapes perceptions and prejudices.

- Gender Innovations Project of Stanford (Machine Learning: Analyzing Gender, https://genderedinnovations.stanford.edu)

- Research on Implicit Associations Test (Greenwald and Banaij) + implicit.harvard.edu/

E R A (T r i e r   1 2 - 1 3   J u n e   2 0 2 3)

23

23

AI and equality in practice

Employment agency AT

- State use of AI / monopoly of labour market placement (AMS)

- Possible discrimination against women

- Problem: WLB is reflected in the data

- Source code of the algorithm disclosed
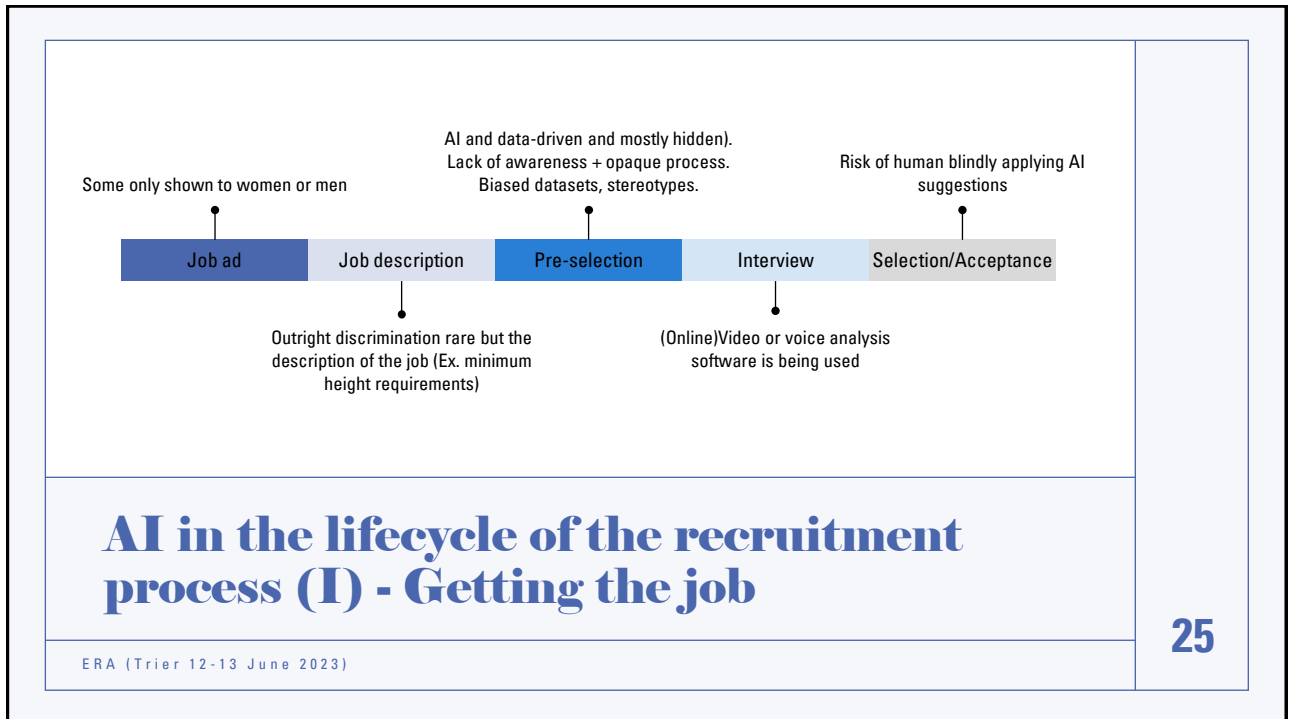
Amazon Recruiting AI

- Private actor

- Possible discrimination against women

- Problem: Historical data (much more data from men than women)
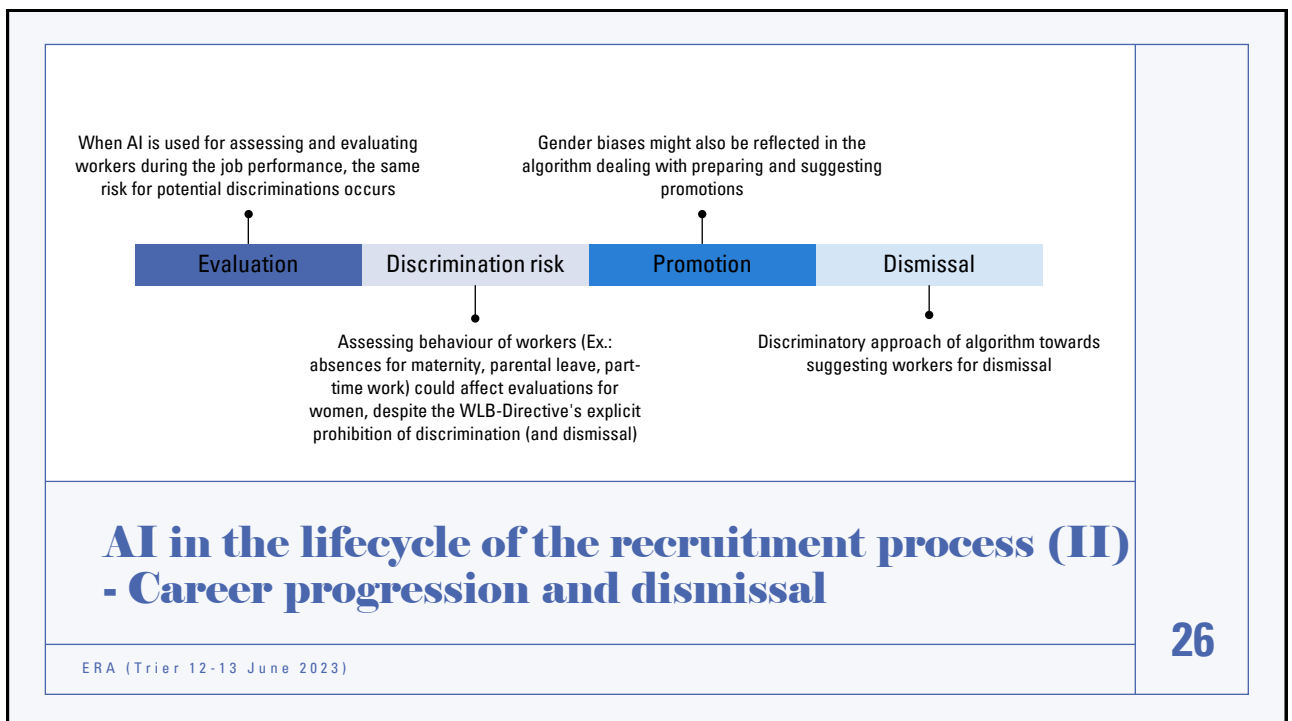
- Solution: Use new data !

E R A (T r i e r   1 2 - 1 3   J u n e   2 0 2 3)

24

24

Some only shown to women or men

AI and data-driven and mostly hidden).
Lack of awareness + opaque process.
Biased datasets, stereotypes.

Risk of human blindly applying AI
suggestions

| Job ad | Job description | Pre-selection | Interview | Selection/Acceptance |

Outright discrimination rare but the
description of the job (Ex. minimum
height requirements)

(Online)Video or voice analysis
software is being used

AI in the lifecycle of the recruitment process (I) - Getting the job

ERA (Trier 12-13 June 2023)

25

When AI is used for assessing and evaluating
workers during the job performance, the same
risk for potential discriminations occurs

Gender biases might also be reflected in the
algorithm dealing with preparing and suggesting
promotions

| Evaluation | Discrimination risk | Promotion | Dismissal |

Assessing behaviour of workers (Ex.:
absences for maternity, parental leave, part-
time work) could affect evaluations for
women, despite the WLB-Directive's explicit
prohibition of discrimination (and dismissal)

Discriminatory approach of algorithm towards
suggesting workers for dismissal

AI in the lifecycle of the recruitment process (II) - Career progression and dismissal

ERA (Trier 12-13 June 2023)

26

Legal questions

Problems of Evidence and Proof

- Providing evidence (AI does not write a reason for rejecting a candidate)
- Burden of proof (who proves what?) + reversal
- Access to opaque AI systems?
- Algorithm as trade secret/commercial legal protection
  - *In-camera* process as a solution
- Judge as AI expert or external expertise needed?

Law Enforcement

- European Commission, DG Justice
- National anti-discrimination bodies (see new Dir)
- Private enforcement
- Referral procedure pursuant to Art. 267 TFEU (central importance for further development of the law)
  - When will a national court refer a discrimination case involving AI to the ECJ ?
- Chief Data/Computer Scientist for KOM

27

27

EU case law (I)

Danfoss C-109/88 (1988)

- "where an undertaking **applies** a system of pay which is **totally lacking in transparency**, it is for the employer to prove that his practice in the matter of wages is not discriminatory, if a female worker establishes, in relation to a relatively large number of employees, that the average pay for women is less than that for men" (§ 16)

Kelly, C-104/10 (2011)

- "does **not** entitle an applicant for vocational training, who believes that his application was not accepted because of an infringement of the **principle of equal treatment, to information held by the course provider on the qualifications of the other applicants** for the course in question, in **order that he may establish 'facts from which it may be presumed that there has been direct or indirect discrimination'** in accordance with that provision. "(§ 38)
- "Nevertheless, it cannot be ruled out that a refusal of disclosure by the defendant, in the context of establishing such facts, could risk compromising the achievement of the objective pursued by that directive and thus depriving, in particular, Article 4(1) thereof of its effectiveness. It is for the national court to ascertain whether that is the case in the main proceedings. " (§ 39)

28

28

EU case law (II)

Meister C-415/10 (2012)

- No right to information regarding other applicants in case of non-selection (§ 46)

- However, it cannot be ruled out that a **defendant's refusal to grant any access to information may be one of the factors to take into account in the context of establishing facts from which it may be presumed that there has been direct or indirect discrimination.** It is for the referring court (...) to determine whether that is the case in the main proceedings." (§ 47)

Shuch-Ghannadan, C-274/18 (2019).

- it **is established that** that legislation adversely affects a **significantly higher percentage of female workers than male workers** and if that legislation is not objectively justified by a legitimate aim or if the means of achieving that aim are not appropriate and necessary. Article 19(1) of that directive must be interpreted as not requiring the party who considers him or herself wronged by such discrimination to submit, in order to establish a prima facie case of discrimination, **specific statistics or facts pertaining to the workers concerned by the national legislation at issue if that party does not have access to those statistics or facts or can access them only with difficulty.**

29

29

EU case law (III): Legal limits to AI ?

- **CJEU JUDGMENT OF 21. 6. 2022 – CASE C-817/19 LIGUE DES DROITS HUMAINS ECLI:EU:C:2022:491**

  - Definition of AI/machine learning (para. 58, 194, 195)

  - What are the limits? For PNR context (requires pre-determined' criteria), non-stability of criteria inherent to machine learning algorithms was decisive ("(..) that requirement precludes the use of artificial intelligence technology in self-learning systems ('machine learning'), capable of modifying without human intervention or review the assessment process (..)", para. 194)

  - What will the Court do in the fundamental rights context ? High stakes decisions ?

IT prohibits use of ChatGPT due to GDPR violations and other EU agencies are exploring the issue
→ GDPR often considered for AD law by researchers and practionners (often due to absence of specific rules addressing AI and gender equality/discrimination issues)
→ Interplay between different legal regimes GDPR (see Lütz, La pollinisation croisée entre droit de la protection des données et droit de la non-discrimination, *forthcoming 2023*)

30

30

Pablo Jensen, Pourquoi la société ne se laisse pas mettre en équations, Seuil, 2018.

4. Regulation of algorithms

- Regulatory approaches
- Regulatory tools & bias and discrimination detection
- EU jurisdiction
- Looking beyond the EU's borders

ERA (Trier 12-13 June 2023)

31

32

32

Role of Business & Corporate human rights diligence for AI

Toronto Declaration

identifies 3 core steps for corporate human rights diligence for machine learning systems:

- (a) **identification** of potential discriminatory outcomes
- (b) **prevention and mitigation** of discrimination and tracking of responses
- (c) **transparency** regarding efforts to identify, prevent and mitigate discrimination

Multi-stakeholder approaches are good practice

Throughout the lifecycle and the different stages of AI system, Due Diligence and Compliance can play a role to reduce biases and ensure non-discrimination

Responsibility as creator of AI systems to avoid and diminish human rights harm

33

33

Company level & role of researchers

Aequitas

An open source bias audit toolkit for machine learning developers, analysts, and policymakers to audit machine learning models for discrimination and bias, and make informed and equitable decisions around developing and deploying predictive risk-assessment tools.
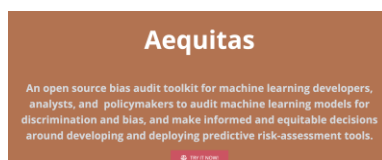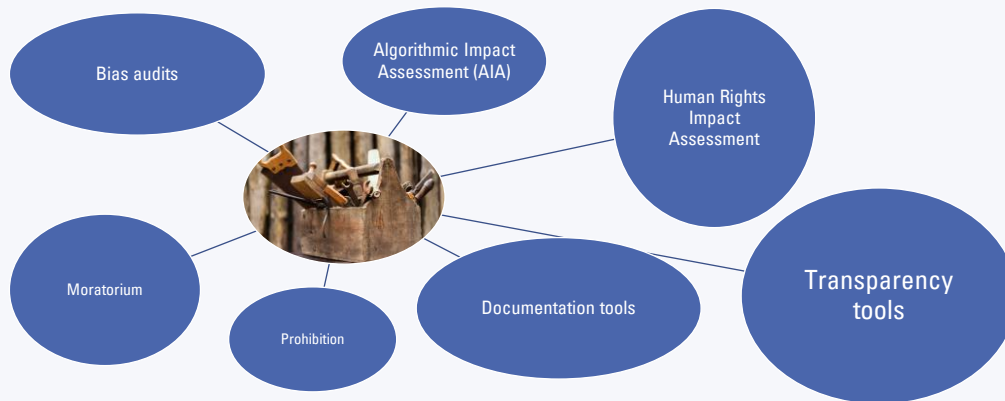
- AI companies should use tools throughout the AI lifecycle to ensure fairness and/or detect biases in their algorithms/AI
  - Specific guides for AI developers and programmers
  - Education and training to ensure fair and non-biased algorithms
- (Independent) **researchers** vital to assess concrete needs and feasibility of regulatory obligations → for example to detect, diminish and eliminate biases from the design and datasets in order to reduce discriminatory outcomes

34

34

Slide 35

Tools to ensure regulatory goals: bias mitigation and reduction of discrimination

Bias audits

Algorithmic Impact Assessment (AIA)

Human Rights Impact Assessment

Transparency tools

Moratorium

Prohibition

Documentation tools

35

35

Slide 36

Global dimension

CoE

- Work in progress towards legislative framework (2025)

- Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems

- "Need to ensure that (...) **gender and other societal and labour force imbalances that have not yet been eliminated from our societies are not deliberately or accidentally perpetuated through algorithmic systems**" (Preamble)

OECD

- OECD/LEGAL/0449, Recommendation of the Council on Artificial Intelligence (2019)

- "AI actors should respect (..) human rights and democratic values, throughout the AI system lifecycle. These include (...) **non-discrimination and equality** (..)» (1.2a)
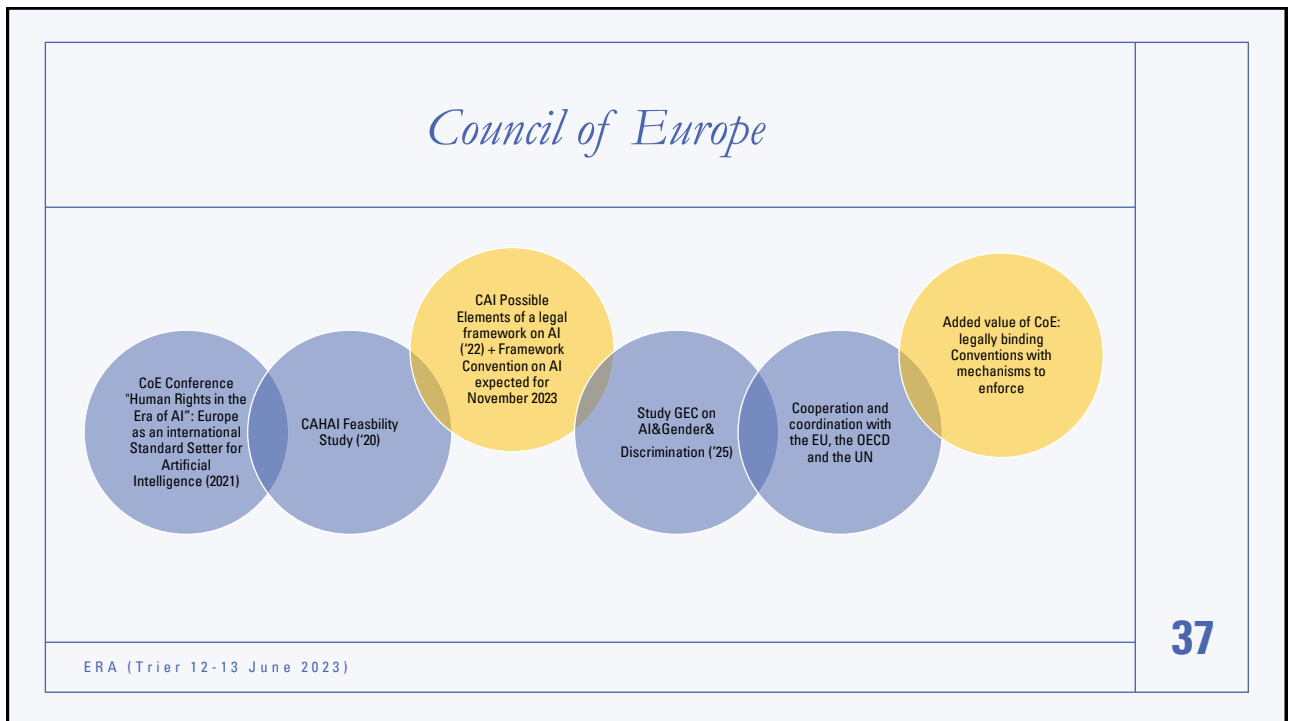
United Nations

- UNESCO Recommendation on the Ethics of AI (2021), Gender as Policy Area 6 (§§ 87-93)

- The right to privacy in the digital age, Report of the UN High Commissioner for Human Rights, A/HRC/48/31:

- "advances in new technologies must not be used to **erode human rights, deepen inequality or exacerbate existing discrimination**"(§ 4) + "**These systems (...) decide who has a chance to be recruited for a job**" (Report, § 57).

36

36

Council of Europe

CoE Conference "Human Rights in the Era of AI": Europe as an international Standard Setter for Artificial Intelligence (2021)

CAHAI Feasbility Study ('20)

CAI Possible Elements of a legal framework on AI ('22) + Framework Convention on AI expected for November 2023

Study GEC on AI&Gender& Discrimination ('25)

Cooperation and coordination with the EU, the OECD and the UN

Added value of CoE: legally binding Conventions with mechanisms to enforce

37

ERA (Trier 12-13 June 2023)

37

---

5. AI as enabler of Gender Equality ?

- AI as discrimination detective
- Comparison human / AI
- Practical considerations

Éric Sadin, La vie algorithmique - Critique de la raison numérique, L'échappée, 2015

ERA (Trier 12-13 June 2023)

38

AI as discrimination detective

Elements

- From the visible to the invisible, from the explicit to the opaque
- Human decisions are error-prone, but even AI is not necessarily better
- Algorithms can overcome (some) *biases* and *noise* (K/S/S, Noise, p. 334-337)

Examples

- Automatic case law analysis
- Better targeted enforcement, better use of resources
- Use algorithms to test other algorithms for possible discriminations
- Bias/Discrimination Detection software

39

A comparison of humans and AI

Advantages of AI

- Coherence, objectivity, speed?
- Better decision makers?
- Exercise of discretion?
- Identification of patterns
- AI as (great) help (support) but humans must be in control

AI as an assistant

- Assistant to the "street-level bureaucrat" (Lipsky, 1980)
- More predictability and reliability in administrative decisions?
- Preparation of case analysis
- Conclusion: AI cannot replace decision making
- *"Unprecedented power of artificial intelligence which can be a **force for good** (..) (Disability (A/HRC/49/52))*

40

Melanie Mitchell, Artificial intelligence - A guide for thinking humans, Pelican, 2019.

6. Conclusion & Outlook

- AI poses risks for Gender Equality
- Also Opportunities
- Regulation is key
- Shared role of business and states
- Global dimension

ERA (Trier 12-13 June 2023)

41

Selected Publications on Algorithms & Gender Equality:

- *How the 'Brussels Effect' Could Shape the Future Regulation of Algorithmic Discrimination*, Duodecim Astra, Issue 1, p. 142-163 (2021), (Available online).
- *Discrimination by correlation - Towards eliminating algorithmic biases and achieving Gender Equality*, contribution to a monograph, Transcript Verlag (2021), (Open Access).
- *Gender equality and artificial intelligence in Europe. Addressing direct and indirect impacts of algorithms on gender-based discrimination*. ERA Forum (2022), available at: https://doi.org/10.1007/s12027-022-00709-6 (Open Access)

- *Artificial Intelligence and Gender-Based Discrimination*, in: Quintavilla Alberto, Temperman Jeroen (eds.) Human Rights and Artificial Intelligence, Oxford University Press (June 2023).
- *Algorithmische Entscheidungsfindung aus der Gleichstellungsperspektive – Ein Balanceakt zwischen Gender Data Gap, Gender Bias, Machine Bias und Regulierung*, GENDER 1/2 (2023), (Open Access online).
- *Gender Equality and Artificial Intelligence: SDG 5 and the role of the UN in fighting stereotypes, biases and gender discrimination*, in: Cristani Federica, Fornalé Elisa (eds.) Women's Empowerment and its Limits, Palgrave, May 2023).
- *Le rôle du droit pour contrer la discrimination algorithmique dans le recrutement automatisé*, In: Guillaume Florence (eds.) La technologie, l'humain et le droit, Stämpfli Verlag (June 2023)

ERA (Trier 12-13 June 2023)

42

42

THANKS
to the ERA

MERCI
for listening

DANKE
for your interest in algorithmic discrimination

Fabian.Luetz@unil.ch

43

ERA (Trier 12-13 June 2023)