Council of Europe study on the impact of artificial intelligence, its potential for promoting equality, including gender equality, and the risks to non-discrimination

Ivana Bartoletti and Raphaële Xenidis

Ivana Bartoletti is Global Chief Privacy Officer at Wipro, Visiting Policy Fellow at the Oxford Internet Institute, University of Oxford and co-founded the Women Leading in AI Network.

Raphaële Xenidis is Lecturer in EU law, University of Edinburgh, School of Law and Marie Curie Fellow, iCourts, University of Copenhagen.

The structure of the report

- Setting the scene and establishing the key points (what makes AI different)
- The good and bad: AI as an opportunity rifled with risks
- The origins of bias and socio-technical components of AI system
- The law: where we are, intersection between disciplines and where AI falls between the cracks
- Rebalancing the burden of the proof to recognise existing power (and information asymmetries)
- Recommendations

Premise

- Opportunities of AI, algorithmic decision making and automated systems. However...
- *without dedicated effort, the use of algorithmic technologies perpetuates and amplifies societal inequalities and harmful stereotypes*.
- Is there a definition of AI? CoE defines “as a ‘blanket term’ for various computer applications based on different techniques, which exhibit capabilities commonly and currently associated with human intelligence. But there is no set definition – and this is problematic.
- What makes AI different? Key point to unlock to define measures and adequate policies.

What is unique and different about AI?

AI Risks are dynamic

- Algorithms learn from new input data
- A model tha.t was low-risk yesterday may be high-risk today, including in whether or not it is fair.

AI systems are complex

- Fairness, for example, has different lenses, not just technical or legal;
- AI systems may be complex to interpret

AI operates in an evolving legal landscape

Regulation around AI is evolving at the intersection between privacy, consumer, data protection, competition and human rights law.

Technology teams lack diversity

Lack of diversity impacts on the abitlity to identify potential bias at both design and implementation stage.

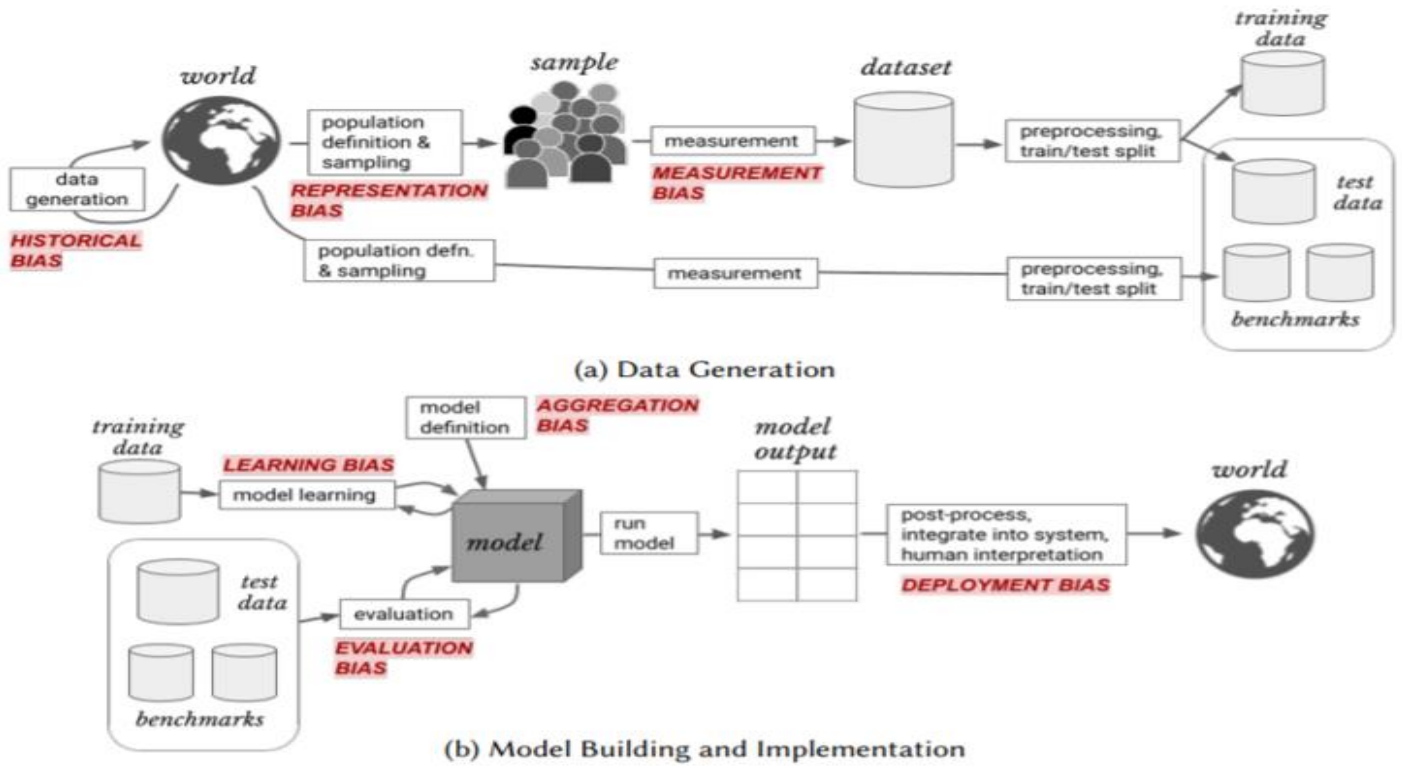
AI systems as socio – technical tools

Overview of bias: it is not just about the data

Types of bias: historical bias, representation bias, learning bias measurement bias, aggregation bias, evaluation bias, and deployment bias.

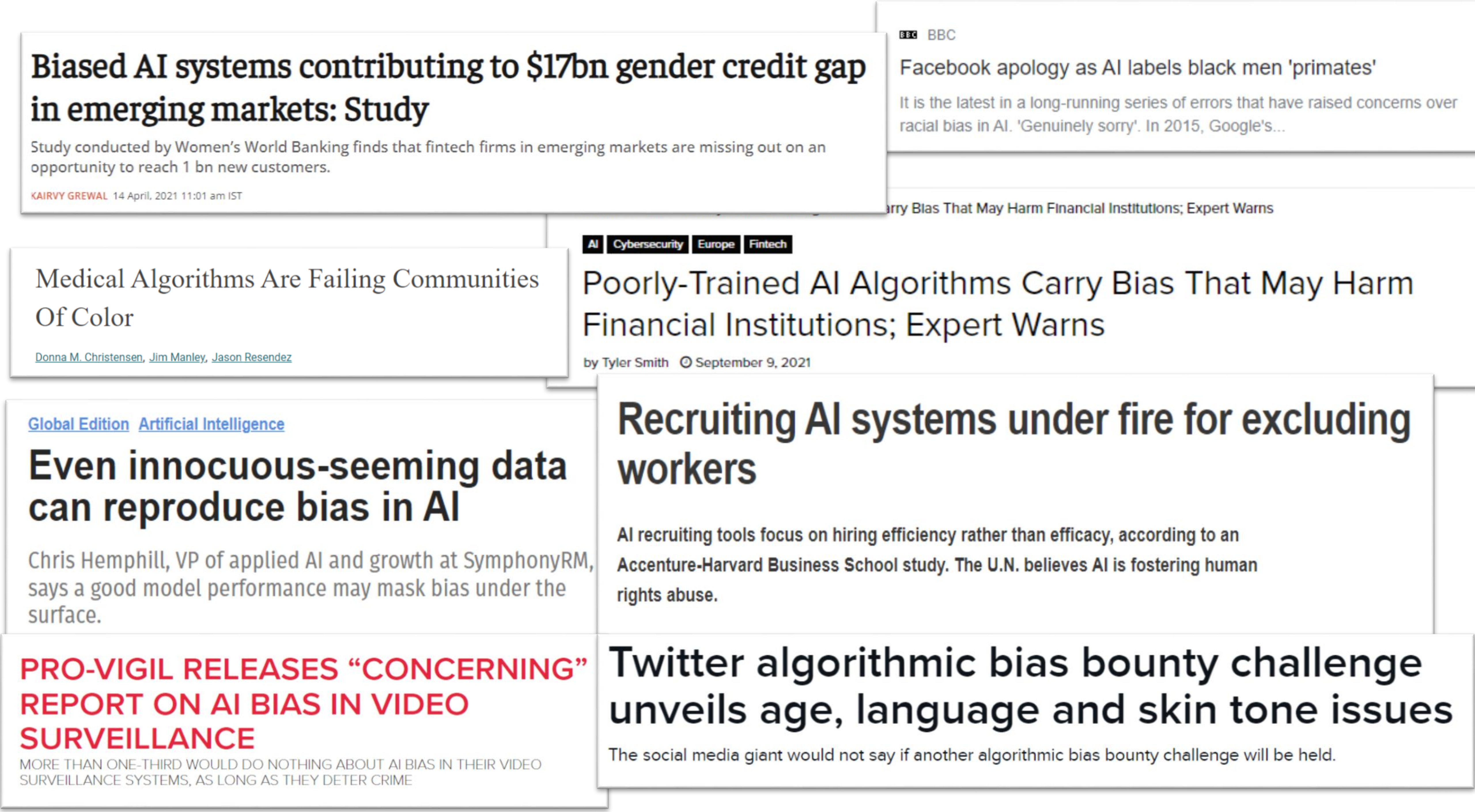
In a nutshell, this means that bias can emerge at any point of the AI lifecycle.

A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle



(a) Data Generation

(b) Model Building and Implementation



Biased AI systems contributing to $17bn gender credit gap in emerging markets: Study

Study conducted by Women's World Banking finds that fintech firms in emerging markets are missing out on an opportunity to reach 1 bn new customers.

KAIRVY GREWAL 14 April, 2021 11:01 am IST

BBC

Facebook apology as AI labels black men 'primates'

It is the latest in a long-running series of errors that have raised concerns over racial bias in AI. 'Genuinely sorry'. In 2015, Google's...

rry Bias That May Harm Financial Institutions; Expert Warns

Medical Algorithms Are Failing Communities Of Color

Donna M. Christensen, Jim Manley, Jason Resendez

AI Cybersecurity Europe Fintech

Poorly-Trained AI Algorithms Carry Bias That May Harm Financial Institutions; Expert Warns

by Tyler Smith September 9, 2021

Global Edition Artificial Intelligence

Even innocuous-seeming data can reproduce bias in AI

Chris Hemphill, VP of applied AI and growth at SymphonyRM, says a good model performance may mask bias under the surface.

Recruiting AI systems under fire for excluding workers

AI recruiting tools focus on hiring efficiency rather than efficacy, according to an Accenture-Harvard Business School study. The U.N. believes AI is fostering human rights abuse.

PRO-VIGIL RELEASES "CONCERNING" REPORT ON AI BIAS IN VIDEO SURVEILLANCE

MORE THAN ONE-THIRD WOULD DO NOTHING ABOUT AI BIAS IN THEIR VIDEO SURVEILLANCE SYSTEMS, AS LONG AS THEY DETER CRIME

Twitter algorithmic bias bounty challenge unveils age, language and skin tone issues

The social media giant would not say if another algorithmic bias bounty challenge will be held.

The role of diversity in AI

- Spotting bias and solutions impacting the more vulnerable
- Innovation requires deliberation
- Any algorithm built by a majority group is at risk of failing to embed perspectives of marginalised minority groups, resulting in algorithms that only work for the majority.
- Addressing diversity should be viewed as mission critical
- State members have reported awareness as well as initiatives being undertaken.

©marketoonist.co

The law: limits, intersections and where does algorithmic discrimination fall into the cracks?

The ECHR equality framework: what relevance for existing legal and policy instruments?

- **Art 14 ECHR**: protected criteria include ‘sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status’
- And **Art 1 Protocol 12** ECHR**, Istanbul Convention**, Framework Convention for the Protection of **National Minorities**, European Charter for **Regional or Minority Languages**
- Policy instruments incl. Recommendation on ‘Preventing and Combating **Sexism**’, Council of Europe **Gender Equality Strategy** 2018-2023, Recommendation on ‘Combating **hate speech**’
- **Case law** of the European Court of Human Rights, in particular on ‘**positive obligations**’

= Legal and policy framework **relevant** to addressing **algorithmic discrimination** in its various dimensions including algorithmic exclusion, violence, stereotyping and disadvantage across many different areas

Main features:

- **Open-ended** list of protected criteria
- **Direct** vs. **indirect** discrimination
- Open regime of **justifications**: existence of a legitimate aim and proportionality between the means employed and the aim sought
- **Shift of the burden of proof** onto the defendant when prima facie case shown
- Importance of **public**/private divide
- **Failure** to act by the state = **discrimination**

The limits of the equality law framework: where does algorithmic discrimination falls into the cracks?

1) **Mismatch** between existing legal **concepts** and the **forms** of algorithmic discrimination:

- ➢ Difficulty in / adequacy of distinguishing direct vs indirect algorithmic discrimination due to role of proxies (NB: less consequences than under EU legal framework because unified justification regime under ECHR)

2) **Procedural issues** linked to evidence, justification and responsibility:

- **Proof**: information asymmetries between algorithmic subjects and decision-makers → obstacles to bringing pima facie evidence of algorithmic discrimination → lacking evidence can amount to access to justice issue
- **Proportionality**: intelligibility of technical trade offs (‘fairness metrics’) in judicial review process? Role of algorithmic opacity and trade secrets in shielding algorithmic decision-making from judicial reviewability?
- **Responsibility** and **liability**: who should be held liable for algorithmic discrimination in the absence of legal personhood of AI systems? Allocation among providers and users of ADM systems?

3) Challenges linked to the protection of **specific characteristics** by the law:

- **Proxy discrimination**: what are the limits of the scope of protection of Art 14?
- **New algorithmic groups**: deprived of social salience, dynamically evolving
- **Intersectional** discrimination: big data and fine-grained clustering

Ways forward: some propositions for a human rights based approach to algorithmic discrimination

- Shifting the regulatory paradigm:
    - Working around legal presumptions to reflect the pervasiveness of algorithmic bias
    - Establishing ex ante accountability obligations and preventive safeguards
    - Adjusting rules on the burden of proof to reflect new power asymmetries
    - Centering negligence in reflections on liability

- Putting positive action at the centre and in a holistic manner:
    - Identifying and addressing new and structural algorithmic vulnerabilities
    - Addressing the lack of diversity, equal representation and equal participation in educational and professional fields related to the AI industry
    - Utilising positive obligations as a legal basis to mainstream equality-related concerns in the development of ADM systems
    - Thinking about the strategic use of quota and other positive action measures

Ways forward: some propositions for a human rights based approach to algorithmic discrimination

- Introducing **preventive obligations** in the form of **human rights impact assessments** *ex ante*, *ex post* and throughout the AI lifecycle, third-party **certification** mechanisms, **audits**

- Setting up **transparency** and **explainability** obligations to reduce power asymmetries and facilitate access to justice

- **Public supervision**, monitoring, information dissemination and awareness-raising: empowering NHRIs, equality bodies and DPAs to **monitor**, **test prevent** and **address** algorithmic discrimination in dialogue with providers and users

- Democratic **participation** in standards-setting and public **consultations** with CSOs with a legitimate interest