Council of Europe

Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks to non-discrimination

Camille Gangloff and Charlotte Gilmartin
Co-Secretaries to the Expert Committee (GEC/ADI-AI)

Study and slides by

Ivana Bartoletti, 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, and

Raphaële Xenidis, Assistant Professor in European law, Sciences Po, Paris.

Structure of the study

- Setting the scene and establishing the key points (what makes AI-driven discrimination different)

- The good and bad: AI as an opportunity which also poses 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

- Recommendations

Premise

- Opportunities of AI, algorithmic decision making and automated systems. However...

without dedicated effort, the use of AI systems perpetuates and amplifies societal inequalities and harmful stereotypes.
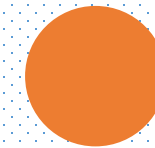
- What makes AI systems different? Key point to unlock to define measures and adequate policies.
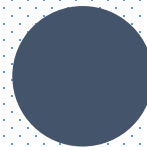
What is unique and different about AI?
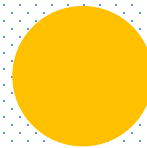
AI Risks are dynamic

- Algorithms learn from new input data

- A model that was low-risk yesterday may be high-risk today, including in whether or not it is fair.
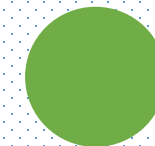
AI systems are complex

- 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 ability 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.

Table and definitions below from: A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle[22]
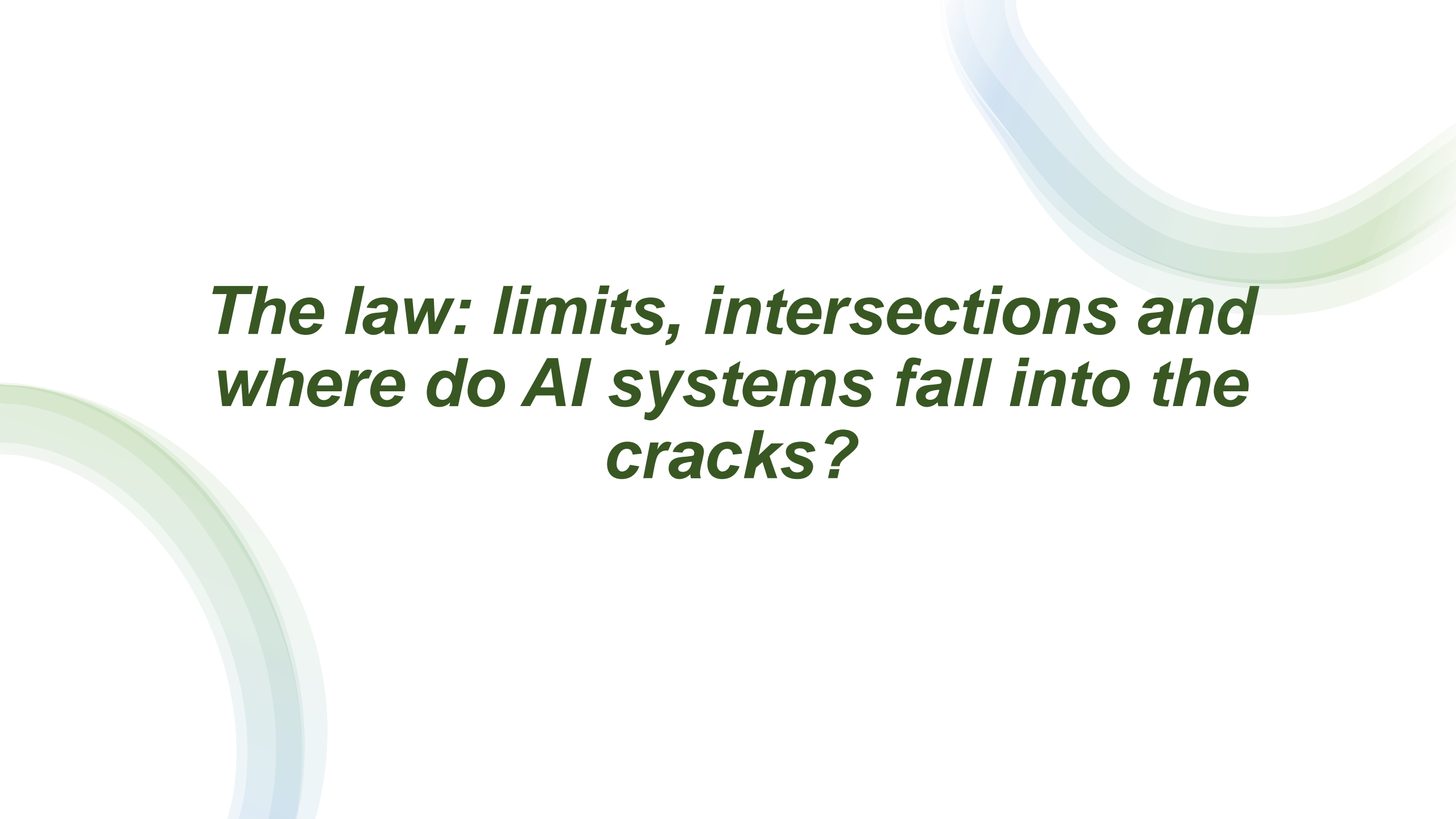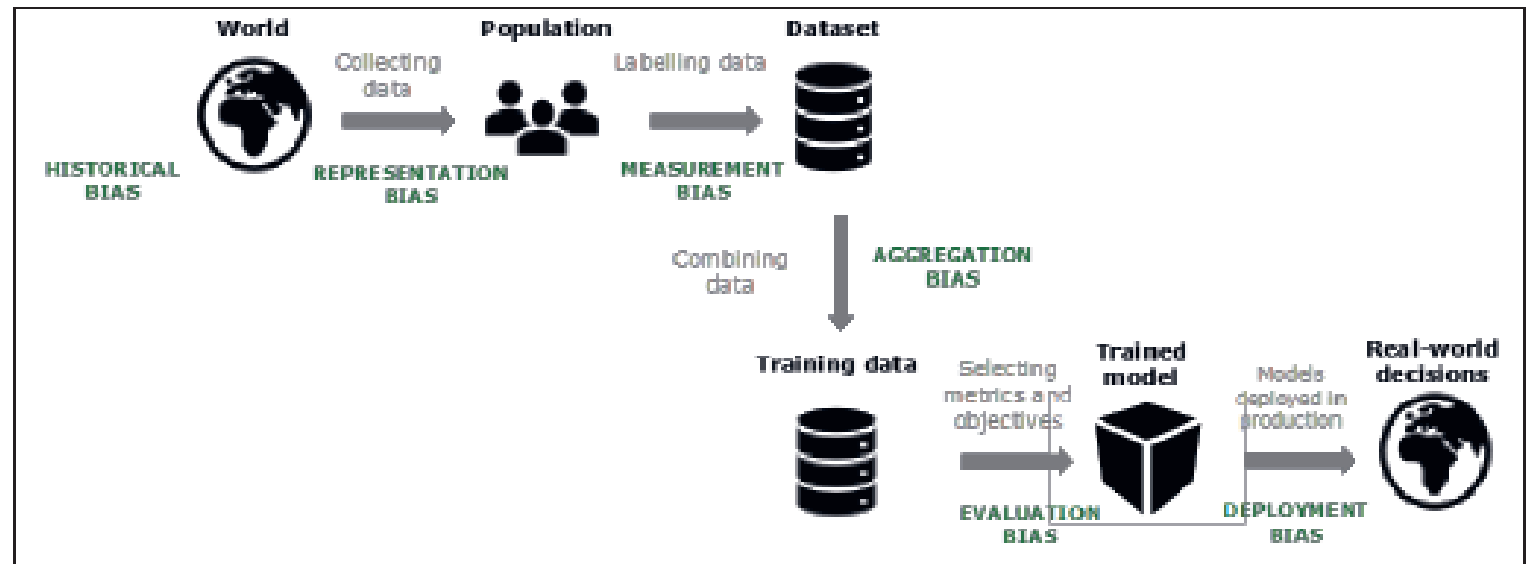
22. Source: Harini Suresh and John Guttag. 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '21), October 5–9, 2021, --, NY, USA.* ACM, New York, NY, USA 9 Pages. https://doi.org/10.1145/3465416.3483305

The role of diversity in AI

- Spotting bias and solutions impacting the more vulnerable
- Innovation requires deliberation, non-bias training and participation
- Any system built by a majority group is at risk of failing to embed perspectives of marginalised minority groups, resulting in systems 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.

WHY DON'T WE HAVE ANY FRESH IDEAS AROUND HERE?

@marketoonist.co

The law: limits, intersections and where do AI systems 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 Strategies** 2018-2023 and 2024-2029, Recommendation on 'Combating **hate speech**'

- **Case law** of the European Court of Human Rights, in particular on '**positive obligations**'

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**

Ways forward: some propositions for a human rights based approach

Shifting the regulatory paradigm

Working around legal presumptions to reflect the pervasiveness of 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 vulnerabilities

Addressing the lack of diversity, equal representation and equal participation in educational and professional fields related to the AI industry

Using positive obligations as a legal basis to mainstream equality-related concerns in the development of AI 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

We want to hear from you!

- What initiatives is your organisation taking to promote equality in the use of AI systems?

- What challenges have been faced?

- What do you think is missing in the current legal/regulatory landscape?