ALGORITHMIC GOVERNANCE IN EUROPE
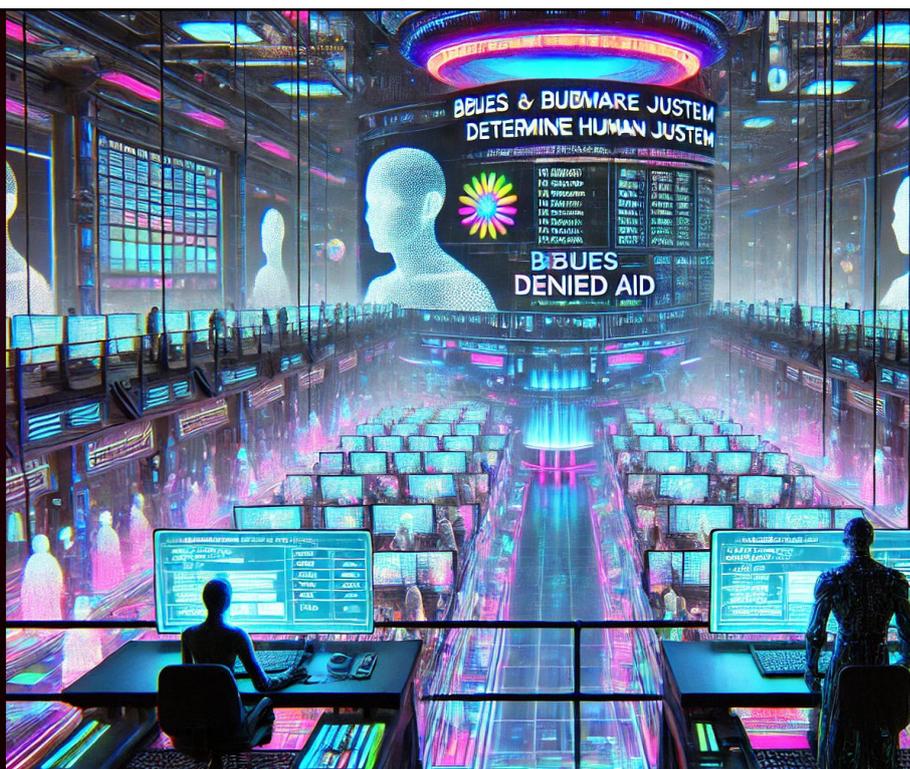
HOW TO DETECT AND ADDRESS DISCRIMINATION IN THE CONTEXT OF AI?

RAPHAËLE XENIDIS
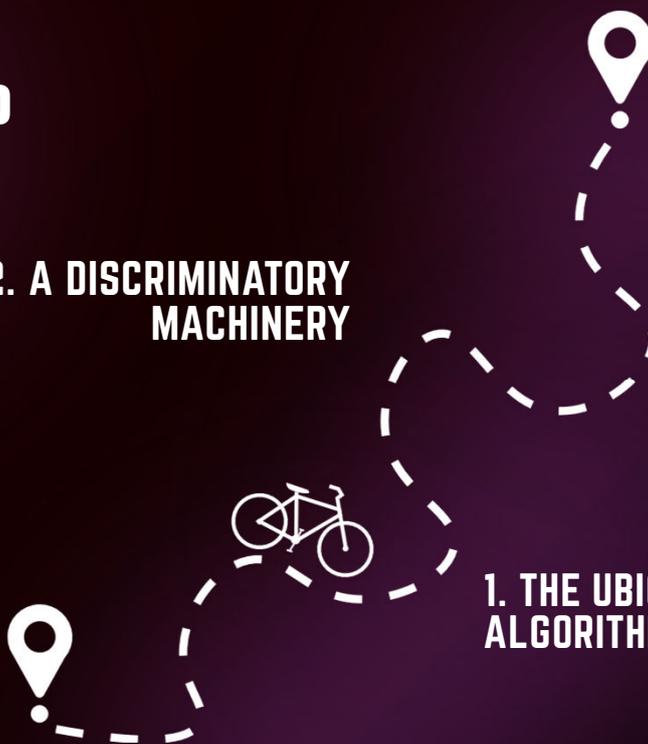
Funded by
the European Union

1

ROADMAP

3. RE-ENGINEERING LEGAL INTERVENTIONS

2. A DISCRIMINATORY MACHINERY

1. THE UBIQUITY OF ALGORITHMIC GOVERNANCE

2

ALGORITHMIC GOVERNANCE IN EUROPE

THE DIGITAL WELFARE STATE

La Quadrature du Net

CAF : LE NUMÉRIQUE AU SERVICE DE L'EXCLUSION ET DU HARCÈLEMENT DES PLUS PRÉCAIRES

19 octobre 2022

LIGHTHOUSE REPORTS

Le Monde

PUBLIÉ LE 4 DÉCEMBRE 2023

COMMENT L'ALGORITHME DE LA CAF PRÉDIT SI VOUS ÊTES « À RISQUE » DE FRAUDER

PAR MANON ROMAIN, ADRIEN SÉNÉCAT, ELSA DELMAS, LÉA GIRARDOT ET THOMAS STEFFEN

SURVEILLANCE | SUSPICION MACHINES

France's Digital Inquisition

Taking apart the secretive fraud detection algorithm that scores half of France's population but pursues the most vulnerable.

PUBLISHED DECEMBER 4, 2023 BY
Manon Romain, Adrien Senecat, Soizic Pénicaud, Gabriel Geiger, Maxime Vaudano, Justin-Casimir Braun, Elsa Delmas, Léa Girardo, Tomas Statius, Daniel Howden

SIMILAR INVESTIGATIONS

RISK FACTORS

Emily Rand & LOTI / Better Images of AI / AI City / CC-BY 4.0
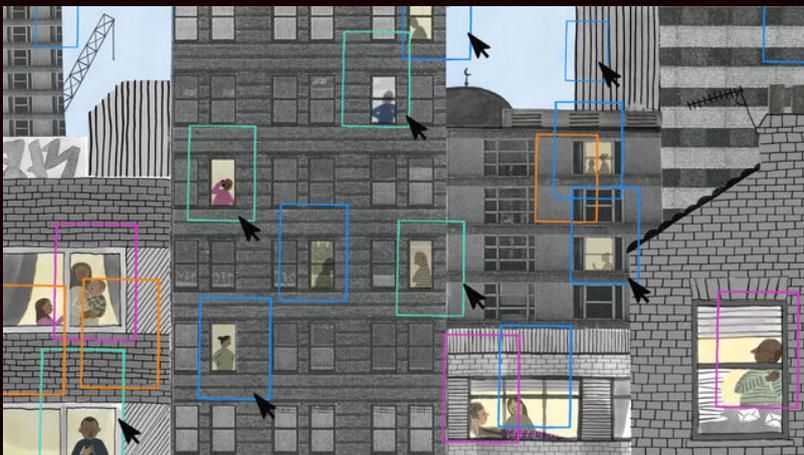
41 risk factors, among which:

- Single-parent families (95% of recipients are single mothers)
  - 16% of recipients
  - 36% of on-site controls
- Recipients' age
- Eligibility to disability benefits
- Marital status
- Income

RISK SCORING IN THE FRENCH SOCIAL SECURITY

A technosolutionist narrative

In 2022:
- more than 87% of investigations, i.e. 28,3 millions investigations, were automated
- Detected fraud only amounted to 0,35% of total benefits

Yaning Wu / Better Images of AI / prediction / CC-BY 4.0

5

ALGORITHMIC GOVERNANCE IN EUROPE

THE DIGITAL WELFARE STATE

6

ALGORITHMIC GOVERNANCE IN EUROPE

THE DIGITAL WELFARE STATE

Variables used in the model

Variable	Nominal values
Gender	Male/Female
Age group	0–29/30–49/50+
Citizenship	Austria/EU except Austria/Non-EU
Highest level of education	Grade school/apprenticeship, vocational school/high- or secondary school, university
Health impairment	Yes/No
Obligations of care (only women)	Yes/No
Occupational group	Production sector/service sector
Regional labor market	Five categories for employment prospects in assigned AMS job center
Prior occupational career	Characterization of variable listed in **Table 2**

Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective

Doris Allhutter[1†]   Florian Cech[2*†]   Fabian Fischer[2†]   Gabriel Grill[3†]

Astrid Mager[1*†]

[1] Institute of Technology Assessment, Austrian Academy of Sciences, Vienna, Austria
[2] Centre for Informatics and Society, Faculty for Informatics, Vienna University of Technology (TU Wien), Vienna, Austria
[3] School of Information, University of Michigan, Ann Arbor, MI, United States

Source: Doris Allhutter et al., 'Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective' (2020) Frontiers in Big Data

Source: https://amsalgorithmus.at/en/

7

Sweden: Authorities must discontinue discriminatory AI systems used by welfare agency

The use of opaque artificial intelligence (AI) systems by *Försäkringskassan,* Sweden's Social Insurance Agency, must be immediately discontinued, Amnesty International said today, following **an investigation** into Sweden's welfare system by Lighthouse Reports and Svenska Dagbladet, which found that the system unjustly flagged marginalized groups for benefits fraud inspections.

8

9

10

AMNESTY INTERNATIONAL — ENGLISH — WHO WE ARE — WHAT WE DO — COUNT

...hms to Combat

The Guardian

...s' in AI tools in

27 November 2024

Sweden: Aut...

LIGHTHOUSE REPORTS

COPUBLISHED WITH
De Limburger — DAMHOFF BRABANT
Automating Distrust
Digital profiling of Dutch low
investigations reveals

SURVEILLANCE | SUSPICION M...

< NEWS

12 November 2024

Denmark: AI-powered welfare system fuels mass surveillance and risks discriminating against marginalized groups – report

The Danish welfare authority, Udbetaling Danmark (UDK), risks discriminating against people with disabilities, low-income individuals, migrants, refugees, and marginalized racial groups through its use of artificial intelligence (AI) tools to flag individuals for social benefits fraud investigations, Amnesty International said today in a new report.

...ts and police forces
...ation

AMNESTY INTERNATIONAL — ENGLISH

π **La Quadrature du Net**

...hms to Combat

The Guardian

...s' in AI tools in

FRANCE TRAVAIL : DES ROBOTS POUR CONTRÔLER LES CHÔMEURS·EUSES ET LES PERSONNES AU RSA

22 mai 2025
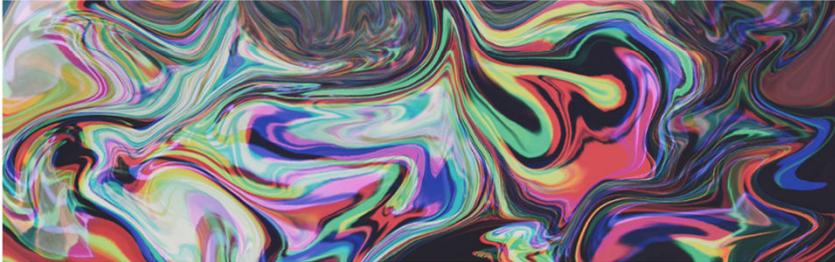
France Travail déploie actuellement des robots visant à automatiser et massifier le contrôle des personnes inscrites à France Travail. Depuis le 1 janvier 2025, cela inclut également les personnes au RSA. Il s'agit d'une nouvelle étape du dangereux projet de gestion algorithmique des personnes sans-emplois, porté par le directeur général de France Travail, Thibaut Guilluy. Retour sur le contexte de cette mise en place et ses implications sociales.

...mass
...st

...disabilities,
...ial intelligence
...y in a new

...ts and police forces
...ation

WHAT DO THESE SYSTEMS HAVE IN COMMON?

AUTOMATING SECURITISATION

—

"Never before have welfare recipients, including the most vulnerable, been so rigorously monitored, and their illegal actions and mistakes so severely punished."

"Social fraud" is defined excessively broadly: "wrongful payments, including those resulting from oversights, errors or involuntary delays on the part of recipients, or even the administration".

Vincent Dubois, *Contrôler les assistés : Genèse et usages d'un mot d'ordre* (Raisons d'agir, 2021)

Photo by Towfiqu barbhuiya on Unsplash

Types of discrimination: classifying algorithmic discrimination

Direct discrimination

- Link to protected ground
- Treatment, perpetrator and causality
- Individual harm and symmetry

Indirect discrimination

- Neutrality
- Effects and structural inequality
- Collective harm and asymmetry
- Open justifications

INTERSECTIONALITY

"The most general statement of our politics at the present time would be that we are actively committed to struggling against racial, sexual, heterosexual, and class oppression, and see as our particular task the development of integrated analysis and practice based upon the fact that the major systems of oppression are interlocking. The synthesis of these oppressions creates the conditions of our lives."

- Combahee River Collective, The Combahee River Collective Statement, 1977

SYSTEMIC DISCRIMINATION

"Systemic discrimination involves the procedures, routines and organisational culture of any organisation that, often without intent, contribute to less favourable outcomes for minority groups than for the majority of the population, from the organisation's policies, programmes, employment, and services."

- Council of Europe, 2024

17

Article 2

Concept of discrimination

1. For the purposes of this Directive, the 'principle of equal treatment' shall mean that there shall be no direct or indirect discrimination whatsoever on any of the grounds referred to in Article 1.

2. For the purposes of paragraph 1:

● ● ●

4. An instruction to discriminate against persons on any of the grounds referred to in Article 1 shall be deemed to be discrimination within the meaning of paragraph 1.

why?

1. *unchartered territory*

2. *epistemic congruence*

3. *legal resilience*

18

DISCRIMINATION 'BY DESIGN'

No neutrality: reversing legal presumptions

Failure to prevent: negligence

Prevention ecosystem

(Semi-)automated decision-making system as an instruction to discrimination

Foreseeability and scale of intersectional algorithmic bias

The General Data Protection Regulation 2016/679

Art. 9 GDPR
Processing of special categories of personal data

1. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

NB: exceptions exist!

GUARANTEES AGAINST FULLY AUTOMATED DECISION-MAKING AND PROFILING
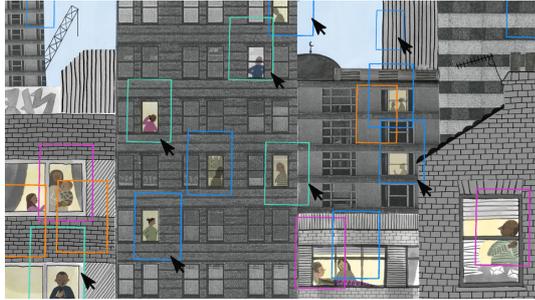
Art. 22 GDPR
Automated individual decision-making, including profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

2. Paragraph 1 shall not apply if the decision:

    (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

    (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

    (c) is based on the data subject's explicit consent.

3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

C-634/21 Schufa Holding (CJEU, 2023)

Is Article 22(1) of the [GDPR] to be interpreted as meaning that the **automated establishment of a probability value** concerning the **ability of a data subject to service a loan** in the future **already constitutes a decision based solely on automated processing**, including **profiling**, which **produces legal effects concerning the data subject or similarly significantly affects him or her**, where that value, determined by means of personal data of the data subject, is transmitted by the controller to a third-party controller and the latter draws strongly on that value for its decision on the establishment, implementation or termination of a contractual relationship with the data subject?

| SCHUFA Score-Simulator

C-634/21 Schufa Holding (CJEU, 2023)

YES!

73. ...Article 22(1) of the GDPR must be interpreted as meaning that **the automated establishment, by a credit information agency, of a probability value based on personal data** relating to a person and concerning his or her ability to meet payment commitments in the future **constitutes 'automated individual decision-making'** within the meaning of that provision**, where a third party**, to which that probability value is transmitted, **draws strongly on that probability value** to establish, implement or terminate a contractual relationship with that person.
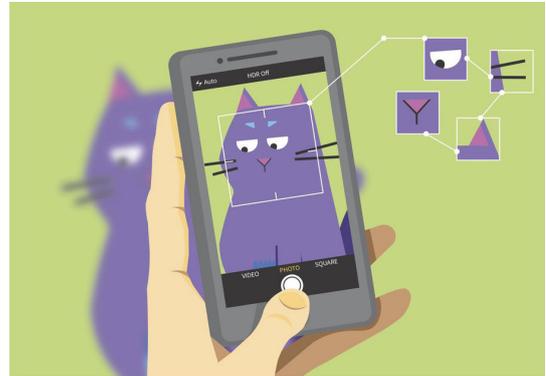
Oleksandra Mukhachova & The Bigger Picture / Better Images of AI / Snapcat / CC-BY 4.0

Regulation laying down harmonised rules on artificial intelligence (EU AI Act)

- Risk-based approach
- Liability: AI Liability Directive

Source: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Prohibited practices
Article 5
EU AI Act

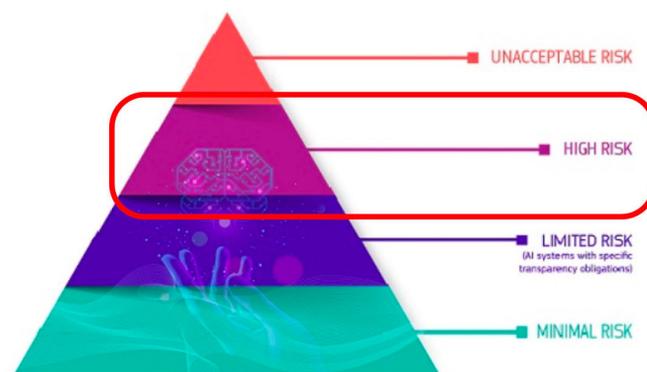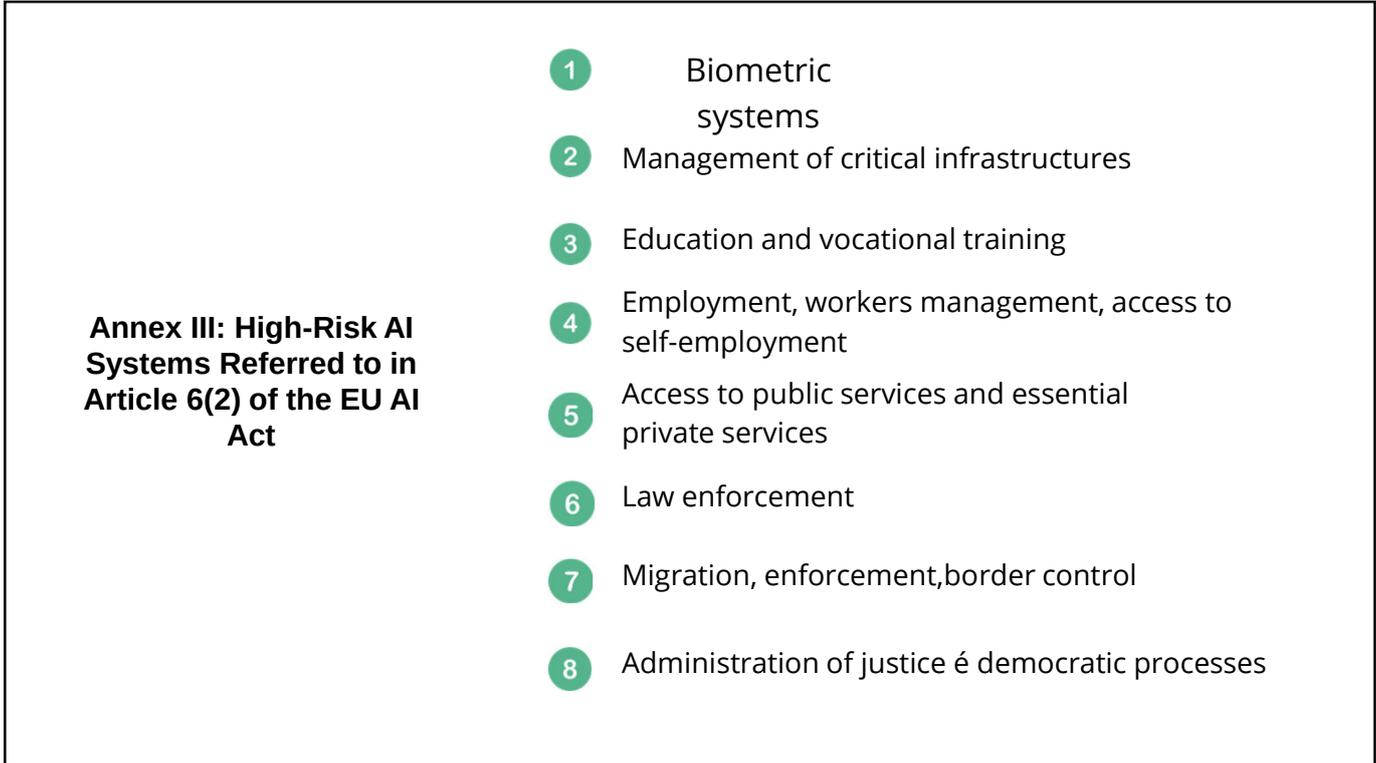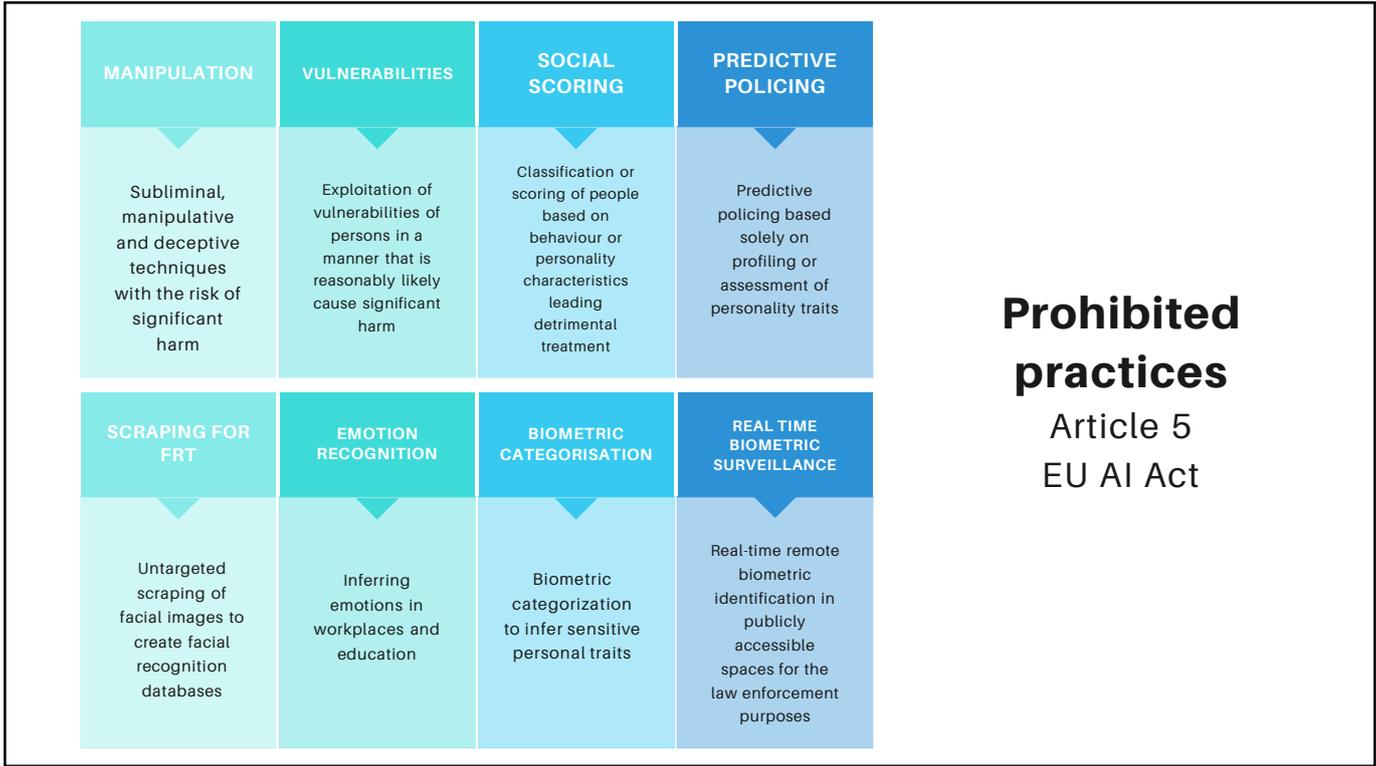| MANIPULATION	VULNERABILITIES	SOCIAL SCORING	PREDICTIVE POLICING
Subliminal, manipulative and deceptive techniques with the risk of significant harm	Exploitation of vulnerabilities of persons in a manner that is reasonably likely cause significant harm	Classification or scoring of people based on behaviour or personality characteristics leading detrimental treatment	Predictive policing based solely on profiling or assessment of personality traits
SCRAPING FOR FRT	EMOTION RECOGNITION	BIOMETRIC CATEGORISATION	REAL TIME BIOMETRIC SURVEILLANCE
Untargeted scraping of facial images to create facial recognition databases	Inferring emotions in workplaces and education	Biometric categorization to infer sensitive personal traits	Real-time remote biometric identification in publicly accessible spaces for the law enforcement purposes

---

Annex III: High-Risk AI Systems Referred to in Article 6(2) of the EU AI Act

1. Biometric systems
2. Management of critical infrastructures
3. Education and vocational training
4. Employment, workers management, access to self-employment
5. Access to public services and essential private services
6. Law enforcement
7. Migration, enforcement,border control
8. Administration of justice é democratic processes

Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law

- **Art. 10 on equality and non-discrimination**: negative and positive obligation
- **Art. 8 on transparency and oversight**
- **Art. 9 on accountability and responsibility** for adverse impacts
- **Art. 14 on accessible and effective remedies**, incl. information & documentation accessible to supervision bodies and victims & complaint mechanism
- **Art. 15 on effective procedural guarantees**, safeguards and rights
- **Art. 16 on risk and impact management framework**: measures for the identification, assessment, prevention and mitigation of risks, incl. representation, documentation and testing (HUDERIA methodology)

European Union Artificial Intelligence Act

Specific requirements for high-risk systems:
- Art. 9 on risk management
- Art. 10 on data governance
- Art. 27 on fundamental rights impact assessments
- Art. 11 on technical documentation
- Art. 12 on record-keeping
- Art. 13 on transparency
- Art. 14 on human oversight
- Art. 15 on accuracy and robustness
- Art. 17 on quality management systems

+ Art. 49 and 71 on registration obligations & EU database
+ Art. 27 on fundamental rights impact assessments

AI Act key stakeholders

- Enforcement: **market surveillance authorities**

- Supervision of fundamental rights by **Art. 77 authorities**

- **CEN-CENELEC**: standardisation process, high stakes (e.g. definition of "residual risk"?)

- **AI Office**: supports implementation of EU AI Act & investigates possible infringements

---

THANK YOU