PPML AND THE AI ACT'S FUNDAMENTAL RIGHTS ASSESSMENT FOR ML SYSTEMS

Prof. Dr. Mireille Hildebrandt
Vrije Universiteit Brussel (Law)
Radboud Universiteit (CS)

What's up?

- PP = preserving what?
- Privacy as the protection of the incomputable self
- Correlatability linkability profiling and manipulation
- FRIA in the GDPR
- FRIA in the proposed EU AI Act
- PPML may solve one problem and create another

What's up?

- PP = preserving what?
- Privacy as the protection of the incomputable self
- Correlatability linkability profiling and manipulation
- FRIA in the GDPR
- FRIA in the proposed EU AI Act
- PPML may solve one problem and create another

PP = preserving what?

- Privacy as privation, e.g. being deprived of the human gaze
  - Isolation, atomism, disconnection, *freedom from*
- Privacy as a provocation, e.g. returning the human gaze
  - Reaching out, relationality, *freedom to*
- Privacy as control over information flows about the self
  - To share or not to share: what - with whom – how – when – how long?
- Privacy: being addressed as an incomputable agent
  - Beyond prediction, similarity, highlighting *in-dividuality* and singularity

What's up?

- PP = preserving what?
- Privacy as the protection of the incomputable self
- Correlatability linkability profiling and manipulation
- FRIA in the GDPR
- FRIA in the proposed EU AI Act
- PPML may solve one problem and create another

Privacy as the protection of the incomputable self

- The self (the I and the me) is indivisible and incomputable
- But it can be made computable
- In different ways – and they make a difference
  - ML system design decisions matter
  - Any quantification assumes qualification (what 'counts as' the same)

- Privacy as the protection of the incomputable self:
  - The reflective shield of Perseus
  - Bouncing the petrifying force of the computational gaze
  - Rejecting computational qualification as this or that type of person

What's up?

- PP = preserving what?
- Privacy as the protection of the incomputable self
- Correlatability linkability profiling and manipulation
- FRIA in the GDPR
- FRIA in the proposed EU AI Act
- PPML may solve one problem and create another

Correlatability, linkability, profiling and manipulation

- HE, MPC and DP:
  - Allow to Ignore individual datapoints while enabling pattern recognition
  - Preserving ‘privacy’ while enabling knowledge acquisition

- They safeguard against the construction of individual profiles
  - Enabling future entrapment by the inferred patterns
  - Due to the application of ‘group’ profiles to ‘matching’ individuals

Correlatability, linkability, profiling and manipulation

- Group profiling based on distributive profiles
  - Allows detailed information about X inferred from a population
  - Infringement of various fundamental rights:
    - Privacy, non-discrimination, effective remedy

- Group profiling based on non-distributive profiles
  - Allows micro-targeting despite incorrect inferences
  - Infringement of various fundamental rights:
    - Non-discrimination, presumption of innocence, effective remedy

What's up?

- PP = preserving what?
- Privacy as the protection of the incomputable self
- Correlatability linkability profiling and manipulation
- FRIA in the GDPR
- FRIA in the proposed EU AI Act
- PPML may solve one problem and create another

FRIA in the art. 24.1 GDPR

- “Taking into account the nature, scope, context and purposes of processing
- as well as the risks of varying likelihood and severity for
  - the rights and freedoms of natural persons,
- the controller shall implement:
  - appropriate technical and organisational measures
  - to ensure and to be able to demonstrate
  - that processing is performed in accordance with this Regulation.
  - Those measures shall be reviewed and updated where necessary.”

FRIA in the art. 25.1 GDPR

- Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing
  - as well as the risks of varying likelihood and severity for
  - rights and freedoms of natural persons posed by the processing,
- the controller shall,
  - both at the time of the determination of the means for processing and at the time of the processing itself,
- implement appropriate technical and organisational measures,
  - such as pseudonymisation,
- which are designed to implement data-protection principles,
  - such as data minimisation,
- in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.

FRIA in the art. 4(5) GDPR

- ‘pseudonymisation’ means:
- the processing of personal data in such a manner that
  - the personal data can no longer be attributed to a specific data subject
  - without the use of additional information,
  - provided that such additional information is kept separately and
  - is subject to technical and organisational measures
  - to ensure that the personal data are not attributed to an identified or identifiable natural person;

FRIA in the art. 25.2 GDPR

- The controller shall implement
  - appropriate technical and organisational measures for ensuring that,
  - by default,
  - only personal data which are necessary for each specific purpose of the processing are processed.
- That obligation applies to:
  - the amount of personal data collected,
  - the extent of their processing,
  - the period of their storage and
  - their accessibility.
- In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.

FRIA in the art. 35.1 GDPR

Where a type of processing in particular using new technologies, and

- taking into account the nature, scope, context and purposes of the processing,
- is likely to result in a high risk
  - to the rights and freedoms of natural persons,
- the controller shall,
  - prior to the processing,
- carry out an assessment of the impact of the envisaged processing operations
- on the protection of personal data.

A single assessment may address a set of similar processing operations that present similar high risks.

What's up?

- PP = preserving what?
- Privacy as the protection of the incomputable self
- Correlatability linkability profiling and manipulation
- FRIA in the GDPR
- FRIA in the proposed EU AI Act
- PPML may solve one problem and create another

FRIA in the proposed AI Act

This proposal imposes some restrictions on

- the freedom to conduct business (Article 16) and
- the freedom of art and science (Article 13)
- to ensure compliance with overriding reasons of public interest such as health, safety, consumer protection and
- the protection of other fundamental rights (‘responsible innovation’)
- when high-risk AI technology is developed and used.

Those restrictions are proportionate and limited to the minimum necessary to prevent and mitigate serious safety risks and likely infringements of fundamental rights.

FRIA in the proposed AI Act

Chapter 1 of Title III sets the classification rules and identifies two main categories of high-risk AI systems:

- AI systems intended to be used as safety component of products that are subject to third party ex-ante conformity assessment;
- other stand-alone AI systems with mainly fundamental rights implications that are explicitly listed in Annex III.

FRIA in the proposed AI Act Recital 28

- The extent of the adverse impact caused by the AI system on the fundamental rights protected by the Charter is of particular relevance when classifying an AI system as high-risk.
- Those rights include the right to human dignity, respect for private and family life, protection of personal data, freedom of expression and information, freedom of assembly and of association, and non-discrimination, consumer protection, workers' rights, rights of persons with disabilities, right to an effective remedy and to a fair trial, right of defence and the presumption of innocence, right to good administration.

ANNEX III HIGH-RISK AI SYSTEMS

1. Biometric identification and categorisation of natural persons:

(a) AI systems intended to be used for the ‘real-time’ and ‘post’ remote biometric identification of natural persons;

ANNEX III HIGH-RISK AI SYSTEMS

2. Management and operation of critical infrastructure:

(a) AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity.

ANNEX III HIGH-RISK AI SYSTEMS

3. Education and vocational training:

(a) AI systems intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions;

(b) AI systems intended to be used for the purpose of assessing students in educational and vocational training institutions and for assessing participants in tests commonly required for admission to educational institutions.

ANNEX III HIGH-RISK AI SYSTEMS

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.

ANNEX III HIGH-RISK AI SYSTEMS

5. Access to and enjoyment of essential private services and public services and benefits:

(a) AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services;

(b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use;

(c) AI systems intended to be used to dispatch, or to establish priority in the dispatching of emergency first response services, including by firefighters and medical aid.

ANNEX III HIGH-RISK AI SYSTEMS

6. Law enforcement:

(a) AI systems intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for potential victims of criminal offences;

(b) AI systems intended to be used by law enforcement authorities as polygraphs and similar tools or to detect the emotional state of a natural person;

(…)

ANNEX III HIGH-RISK AI SYSTEMS

7. Migration, asylum and border control management:

(a) AI systems intended to be used by competent public authorities as polygraphs and similar tools or to detect the emotional state of a natural person;

(b) AI systems intended to be used by competent public authorities to assess a risk, including a security risk, a risk of irregular immigration, or a health risk, posed by a natural person who intends to enter or has entered into the territory of a Member State;

(…)

ANNEX III HIGH-RISK AI SYSTEMS

8. Administration of justice and democratic processes:

(a) AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts.

FRIA in the proposed AI Act
Article 13 Transparency and provision of information to users

1. High-risk AI systems shall be
   - designed and developed in such a way to ensure
   - that their operation is sufficiently transparent
   - to enable users to interpret the system's output and use it appropriately.

FRIA in the proposed AI Act
Article 13 Transparency and provision of information to users

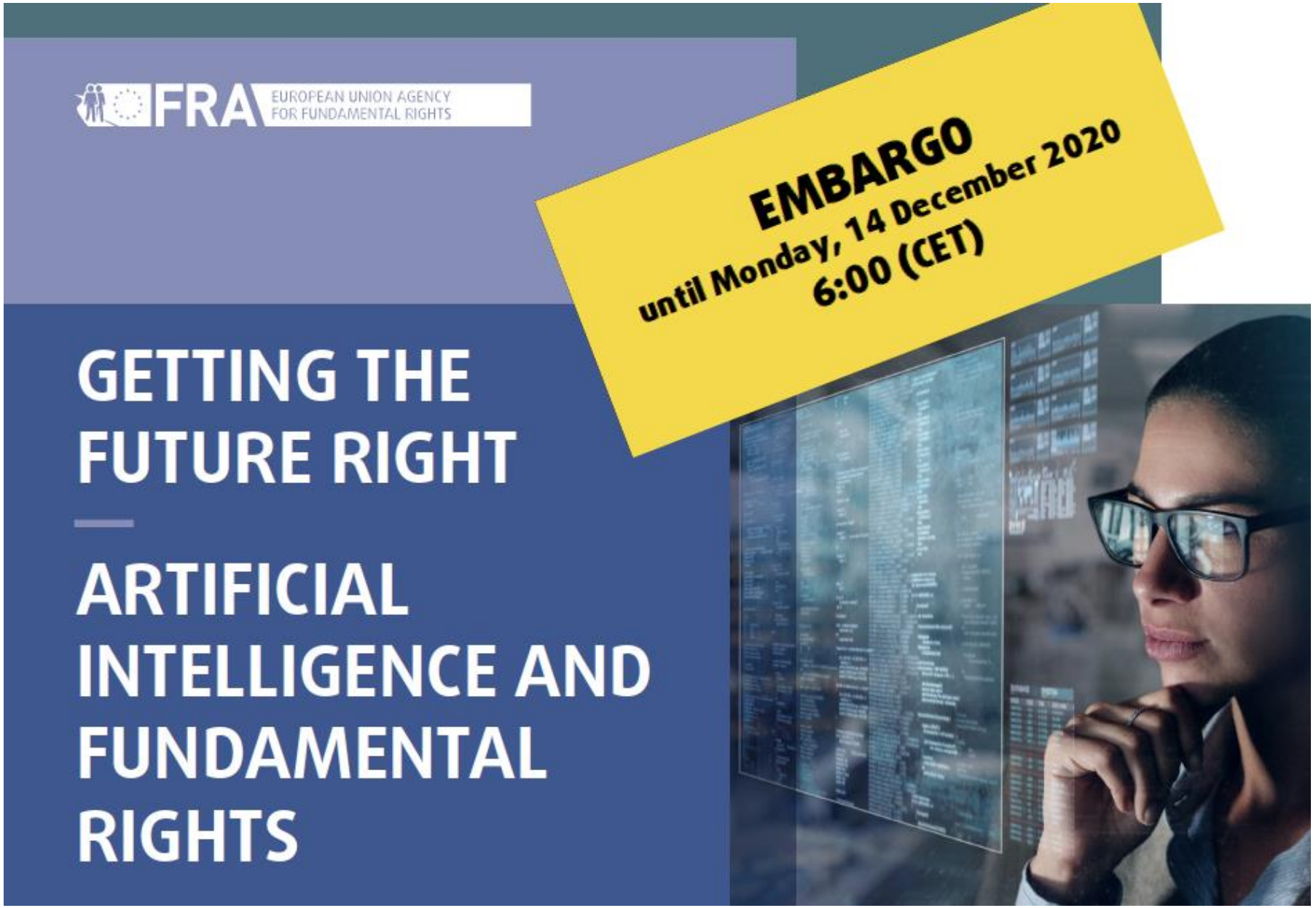
3. The information referred to in paragraph 2 shall specify:

iii. any known or foreseeable circumstance,

- related to the use of the high-risk AI system
- in accordance with its intended purpose or
- under conditions of reasonably foreseeable misuse,
- which may lead to risks to the health and safety or fundamental rights;

FRIA in the proposed AI Act
Article 14 Human oversight

2. Human oversight shall aim at
   - preventing or minimising the risks to health, safety or fundamental rights
   - that may emerge when a high-risk AI system is used
     - in accordance with its intended purpose or
     - under conditions of reasonably foreseeable misuse,
     - in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter.


FRA
EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS
EMBARGO
until Monday, 14 December 2020
6:00 (CET)
GETTING THE FUTURE RIGHT
ARTIFICIAL INTELLIGENCE AND FUNDAMENTAL RIGHTS

An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems ☆

Alessandro Mantelero[#,*], *Maria Samantha Esposito*[#]

Department of Management and Production Engineering, Polytechnic University of Turin, Turin, Italy

ARTICLE INFO

Keywords:
Artificial intelligence
Human rights
Human Rights Impact Assessment
Data protection
AI regulation
Data ethics

ABSTRACT

Different approaches have been adopted in addressing the challenges of Artificial Intelligence (AI), some centred on personal data and others on ethics, respectively narrowing and broadening the scope of AI regulation. This contribution aims to demonstrate that a third way is possible, starting from the acknowledgement of the role that human rights can play in regulating the impact of data-intensive systems.

The focus on human rights is neither a paradigm shift nor a mere theoretical exercise. Through the analysis of more than 700 decisions and documents of the data protection authorities of six countries, we show that human rights already underpin the decisions in the field of data use.

Based on empirical analysis of this evidence, this work presents a methodology and a model for a Human Rights Impact Assessment (HRIA). The methodology and related assessment model are focused on AI applications, whose nature and scale require a proper contextualisation of HRIA methodology. Moreover, the proposed models provide a more measurable approach to risk assessment which is consistent with the regulatory proposals centred on risk thresholds.

The proposed methodology is tested in concrete case-studies to prove its feasibility and effectiveness. The overall goal is to respond to the growing interest in HRIA, moving from a mere theoretical debate to a concrete and context-specific implementation in the field of data-intensive applications based on AI.

What's up?

- PP = preserving what?
- Privacy as the protection of the incomputable self
- Correlatability linkability profiling and manipulation
- FRIA in the GDPR
- FRIA in the proposed EU AI Act
- PPML may solve one problem and create another

PPML may solve one problem and create another

- What if PPML legitimises ML that enables microtargeting
  - That may be discriminatory, manipulative or incorrect?
- How does PPML interact with debiasing or with checking for bias?
- How does PPML interact with verification/validation in terms of other fundamental rights, e.g.
  - presumption of innocence, effective remedy, freedom of information, fair administration?
- Should new methods be developed to check which type of PPML (HE, MPC, DP) best fits with
  - Checking for, prevention of or mitigation of other FR infringements?

PPML may solve one problem and create another

- Why should CS care?
  - Can one say: well just add debiasing, don't manipulate?
  - Can one say: just don't ask me, I just do PPML?

PPML may solve one problem and create another

- Even if a developer or system's architect takes that position
  - Controllers (GDPR) and providers (AI Act) cannot afford it
  - They need to anticipate fundamental rights infringements
  - And avoid or mitigate them
- Raising Question Zero: should we do ML here at all?

THE END