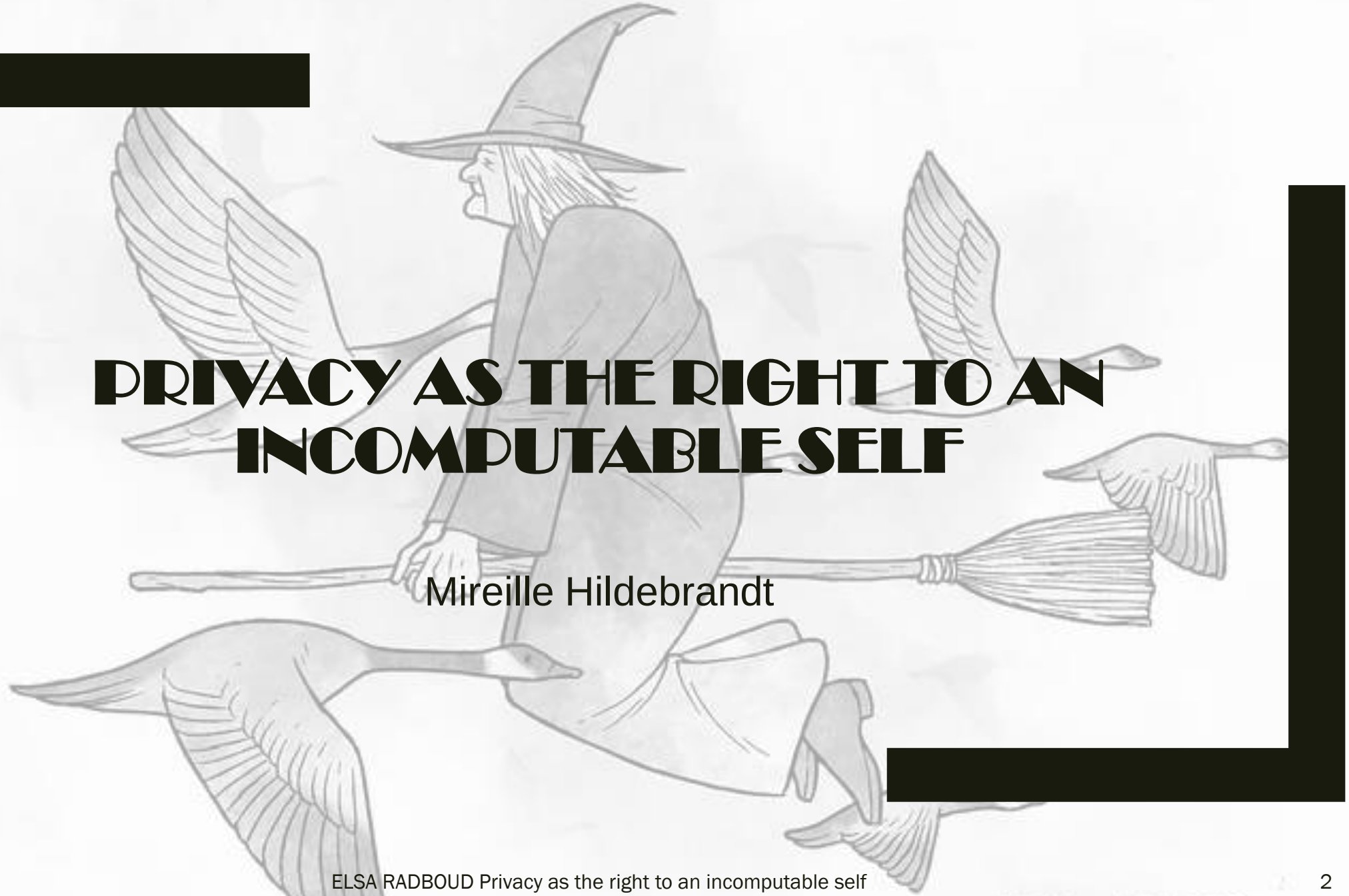
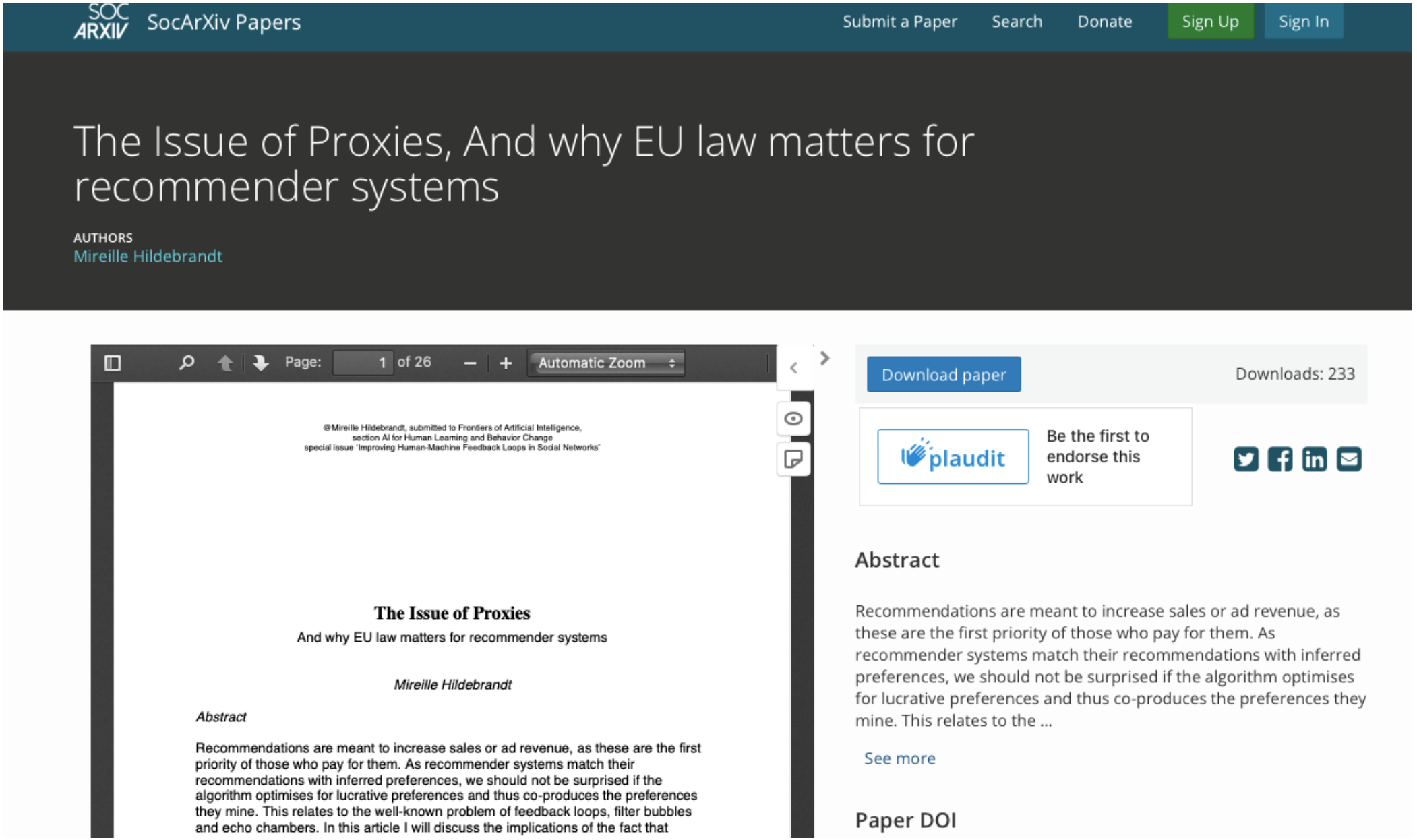
# PRIVACY AS THE RIGHT TO AN INCOMPARABLE SELF

Mireille Hildebrandt



# PRIVACY AS THE RIGHT TO AN INCOMPUTABLE SELF

Mireille Hildebrandt


SocArXiv Papers
Submit a Paper
Search
Donate
Sign Up
Sign In
The Issue of Proxies, And why EU law matters for recommender systems
AUTHORS
Mireille Hildebrandt
Page: 1 of 26
Automatic Zoom
@Mireille Hildebrandt, submitted to Frontiers of Artificial Intelligence, section AI for Human Learning and Behavior Change special issue 'Improving Human-Machine Feedback Loops in Social Networks'
The Issue of Proxies
And why EU law matters for recommender systems
Mireille Hildebrandt
Abstract
Recommendations are meant to increase sales or ad revenue, as these are the first priority of those who pay for them. As recommender systems match their recommendations with inferred preferences, we should not be surprised if the algorithm optimises for lucrative preferences and thus co-produces the preferences they mine. This relates to the well-known problem of feedback loops, filter bubbles and echo chambers. In this article I will discuss the implications of the fact that
Download paper
Downloads: 233
plaudit
Be the first to endorse this work
Abstract
Recommendations are meant to increase sales or ad revenue, as these are the first priority of those who pay for them. As recommender systems match their recommendations with inferred preferences, we should not be surprised if the algorithm optimises for lucrative preferences and thus co-produces the preferences they mine. This relates to the ...
See more
Paper DOI

# What's new?

- GOFP (good old fashioned privacy)
- Privacy, data protection and fundamental rights
- Privacy as the protection of the incomputable self
- GDPR: FRIA and DPbDD
- AI Act: FRIA and LPbD
- LPbD, from Q0 to mitigation

# What's new?

- GOFP (good old fashioned privacy)
- Privacy, data protection and fundamental rights
- Privacy as the protection of the incomputable self
- GDPR: FRIA and DPbDD
- AI Act: FRIA and LPbD
- LPbD, from Q0 to mitigation

GOFP

- Horizontal (direct):
  - *Privacy tort, peer-2-peer (art. 6:162 CC)*
  - *Public disclosure of photographs (Warren and Brandeis in the US)*

- Vertical:
  - *Constitutional right to privacy, international human right to privacy*
  - *Government-subject*
  - *Criminal law investigations, secret services (Klass, Weber and Saravia ECtHR)*

- Horizontal (indirect):
  - *Human right violations by states ignoring their positive obligations*
  - *Government-subject/subject*
  - *This may even involve an obligation to criminalise violations by private parties*

# What's new?

- GOFP (good old fashioned privacy)
- Privacy, data protection and fundamental rights
- Privacy as the protection of the incomputable self
- GDPR: FRIA and DPbDD
- AI Act: FRIA and LPbD
- LPbD, from Q0 to mitigation

# Privacy and data protection

- Privacy is mainly seen as a negative right
  - *Natural persons' freedom from interference by private parties and public authorities*
- Data protection is often seen as a positive right
  - *For controllers: conditional freedom to process personal data*
  - *For data subjects: freedom to control processing of personal data*
- Privacy is more than data protection (e.g. physical integrity, decisional privacy)
- Data Protection is more than privacy (e.g. unlawful processing of personal data may concern non-discrimination or the presumption of innocence rather than privacy)

# Privacy the modern approach: data protection

- Data-driven infrastructure, access to information/housing/insurance/education

- Reasonably foreseeable consequences of the processing of personal data:
  - *Dutch tax office flags low income people for potential fraud*
  - *UK government targets disabled receivers of welfare benefits as potential frauds*

# DWP urged to reveal algorithm that 'targets' disabled for benefit fraud

**Manchester group launches action after people with disabilities report high number of stressful checks for potential scams**


The Department for Work & Pensions has been asked to explain how artificial intelligence is being used to identify potential benefit fraudsters. Photograph: Andy Rain/EPA

Disabled people are being subjected to stressful checks and months of frustrating bureaucracy after being identified as potential benefit fraudsters by an algorithm the government is refusing to disclose, according to a new legal challenge.

# Court clears 39 post office operators convicted due to 'corrupt data'

**Theft, fraud and false accounting convictions quashed after one of England's biggest ever miscarriages of justice**

**Written evidence from PAUL MARSHALL, barrister, of Cornerstone Barristers (PPS0024)**

PAUL MARSHALL, barrister, of Cornerstone Barristers, 2-3 Gray's Inn Square, Gray's Inn, London WC1R 5JH will say:

1. I am a barrister in private practice. I have been invited to make a written statement to the Justice Committee. While it is unusual to do so, I have provided footnotes to avoid cluttering up the text of this statement with references.

2. The occasion, as I understand it, is a request from the Criminal Cases Review Commission following upon the disaster of the flawed Post Office prosecutions of its sub-postmasters and sub-postmistresses (SPMs), from introduction of the Horizon system in 1999 until about 2014, that was exposed by the judgments of Mr Justice Fraser in *Bates v Post Office*.[1]

In particular, *Bates v the Post Office Ltd (No 6: Horizon Issues) (Rev 1)* [2019] EWHC 3408 (QB) https://www.bailii.org/ew/cases/EWHC/QB/2019/3408.html (168 pages, 1,030 paragraphs – the Technical Appendix extends to 452 paragraphs).

# Belastingdienst stuurt brief over fraudesignalering

Nieuwsbericht | 25-03-2021 | 17:57

Ongeveer 240.000 belastingplichtigen krijgen de komende maanden van de Belastingdienst en de Dienst Toeslagen te horen dat hun gegevens in de zogeheten fraudesignaleringsvoorziening (FSV) hebben gestaan. De eerste 2000 krijgen nog deze maand het bericht, waarin de Belastingdienst spijt betuigt voor het gebruik van dit systeem dat niet aan de privacywet AVG voldeed. De overige brieven worden, tot eind juni, in twee zendingen verstuurd.

In de brief legt de Belastingdienst onder meer uit wat FSV was en waarvoor het werd gebruikt. De meeste mensen hebben van deze registratie geen gevolgen ondervonden. Bij een nog onbekend aantal is mogelijk een verzoek om een betalingsregeling of minnelijke schuldsanering automatisch afgewezen omdat zij in FSV stonden. Zij worden in de brief opgeroepen zich te melden. Ook als zij zich niet hebben gemeld en nader onderzoek door de Belastingdienst wijst uit dat zij onterechte nadelige gevolgen hebben ondervonden van vermelding in FSV, krijgen zij later een tweede brief.

Niet iedereen die in het systeem was opgenomen krijgt een brief. De ouders in de hersteloperatie kinderopvangtoeslag krijgen in plaats van een brief bericht van hun persoonlijke zaakbehandelaar. Ook in situaties waarin de (fiscaal)-juridische, toezicht- of opsporingsbelangen zwaar wegen, gaat er geen bericht uit.

**De Belastingdienst heeft jarenlang mensen met een laag inkomen geselecteerd voor extra controle bij aanvragen voor toeslagen voor kinderopvang. Zij kwamen eerder in het vizier bij fraudebestrijding. Hogere inkomens werden bewust ontzien.**

Dit bevestigt de Belastingdienst / Toeslagen in antwoorden op vragen van RTL Nieuws en Trouw. Wie een laag inkomen had, liep een grotere kans eruit gepikt te worden door het zogeheten risicoclassificatiemodel van Toeslagen.

Dit risicoclassificatiemodel (RCM) werd juli vorig jaar uit de lucht gehaald, na een kritisch rapport van consultancybureau KPMG. De Belastingdienst heeft daarna geprobeerd het systeem te verbeteren en weer in de lucht te krijgen.

## Selectiemodel was 'stigmatiserend'

Uit een intern vervolgonderzoek bleek in maart dit jaar dat het model mogelijk 'onrechtmatig' en 'discriminerend' was. Burgers zouden 'gestigmatiseerd' kunnen worden, zo blijkt uit de eigen analyse die aan RTL Nieuws en Trouw is verstrekt. Daarna is besloten het RCM voorgoed niet meer te gebruiken. (De kern van dit rapport lees je hier).

# Gegevensbeschermings-effectbeoordeling (GEB)

## Doelgericht ingrijpen op een aangevraagde toeslag (M1354): Het Risicoclassificatiemodel

**Ministerie van Financiën, directie Toeslagen**

**Utrecht, 11 maart 2021**

- Recht om aan uitsluitend geautomatiseerde besluitvorming te worden onderworpen:
  Het verwerken van persoonsgegevens in het risicoclassificatiemodel kwalificeert als profilering in de zin van artikel 4, aanhef en onder 4, van de AVG. Aanvragen die het risicoclassificatiemodel selecteert worden handmatig beoordeeld door een medewerker van Toeslagen. Er is door deze menselijke tussenkomst geen sprake van geautomatiseerde besluitvorming als bedoeld in artikel 22 van de AVG.

| Nr. | Risico | Oorsprong | Kans[54] (1-5) | Impact[55] (1-5) |
|---|---|---|---|---|
| 1. | Het risico kan ontstaan dat het resultaat onnauwkeurig en/of discriminerend is en dat het model niet meer representatief is. | Regressiemodellen moeten vaak getraind worden met actuele trainingcases, maar dit gebeurt niet altijd. | 3 | 4 – Externe impact<br>Proces: A, B<br><br>Als het model de verkeerde indicatoren gebruikt dan bestaat de kans dat er burgers een onterecht hoge score van het model krijgen, wat een vertraging van het ontvangen van toeslag kan betekenen. Het heeft invloed op de toekenning van de betrokkene. Als er al een toekenning is en de burger stuurt een wijziging in dan wordt de wijziging niet verwerkt en is deze ook niet zichtbaar voor de burger. Pas na handmatige controle en goedkeuring door een medewerker wordt de wijziging doorgevoerd. |

| Nr. | Risico | Oorsprong | Kans[54] (1-5) | Impact[55] (1-5) |
|---|---|---|---|---|
| 1. | Het risico kan ontstaan dat het resultaat onnauwkeurig en/of discriminerend is en dat het model niet meer representatief is. | Regressiemodellen moeten vaak getraind worden met actuele trainingcases, maar dit gebeurt niet altijd. | 3 | 4 – Externe impact<br>Proces: A, B<br><br>Als het model de verkeerde indicatoren gebruikt dan bestaat de kans dat er burgers een onterecht hoge score van het model krijgen, wat een vertraging van het ontvangen van toeslag kan betekenen. Het heeft invloed op de toekenning van de betrokkene. Als er al een toekenning is en de burger stuurt een wijziging in dan wordt de wijziging niet verwerkt en is deze ook niet zichtbaar voor de burger. Pas na handmatige controle en goedkeuring door een medewerker wordt de wijziging doorgevoerd. |

| | | | | |
|---|---|---|---|---|
| 10. | Het risico kan ontstaan dat de output van het model niet meer uitlegbaar is, wat een probleem zal vormen op het moment dat een burger zijn recht op inzage wil uitoefenen. | De logica van het model kan complex zijn. De risicoscore per indicator is niet altijd uitlegbaar. | 2 | 3 Externe impact<br>Proces: A, B, C<br><br>De burgers zouden geen duidelijke informatie over de verwerking krijgen en het zou kunnen dat ze niet begrijpen waarom hun aanvraag als hoog risico is geselecteerd. |

# Fundamental rights

- Is this about privacy? Yes, but about much more
- I advise each of you to download and read this summary report to get a better sense of what concrete risks are spelled out: https://www.rtlnieuws.nl/sites/default/files/content/documents/2021/11/22/GEBmaart2021RCM-risico%27s.pdf

- Note that this is an art. 35 DPIA, to be conducted by the controller
- There is no obligation to make a DPIA public (this one has been made public by RTL4)

# Private interests public goods

- Two questions:
  - *Is privacy a private interest (with which I can trade)?*
  - *Or is it (also) a public good (we don't want folk to trade with)?*

# Private interests public goods

- Questions:
  - *Are the implications of ADM/AI a matter of cost-benefit or of power (im)balances?*
  - *Or do they concern public goods (that are by definition not computable)?*

# What's new?

- GOFP (good old fashioned privacy)
- Privacy, data protection and fundamental rights
- Privacy as the protection of the incomputable self
- GDPR: FRIA and DPbDD
- AI Act: FRIA and LPbD
- LPbD, from Q0 to mitigation

# My 3 cards on the table
## (if you are not at the table, you are on the menu)

1. Things that matter are not computable
2. They can nevertheless be made computable
3. They can be computed in different ways and the difference matters

# What's new?

- GOFP (good old fashioned privacy)
- Privacy, data protection and fundamental rights
- Privacy as the protection of the incomputable self
- GDPR: FRIA and DPbDD
- AI Act: FRIA and LPbD
- LPbD, from Q0 to mitigation

# FRIA 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 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 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 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 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 new?

- GOFP (good old fashioned privacy)
- Privacy, data protection and fundamental rights
- Privacy as the protection of the incomputable self
- GDPR: FRIA and DPbDD
- AI Act: FRIA and LPbD
- LPbD, from Q0 to mitigation

# 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 new?

- GOFP (good old fashioned privacy)
- Privacy, data protection and fundamental rights
- Privacy as the protection of the incomputable self
- GDPR: FRIA and DPbDD
- AI Act: FRIA and LPbD
- LPbD, from Q0 to mitigation

# AI systems under the rule of law?

- Rule of law in a constitutional democracy is a normative undertaking:
  - *It aims to protect*
    - the incomputable nature of human agency
    - a shared world that affords privacy, diversity, inclusion
    - transparency, accountability and contestability of big players
    - by way of a series of institutional checks and balances
    - notably 'effective and practical' fundamental rights
- In the context of pervasive code- and/or data-driven environments this requires
  - *Legal protection by design and default*

# Legal Protection by Design

- From legal protection based on **text-driven ICT**
- To legal protection at the level of relevant **data- and code-driven ICT**

This is NOT about ethics

- We do not want to depend on the ethical inclinations of developers
- Or those who fund them (big tech)

# Legal Protection by Design

This is NOT about ‘legal by design’ (which is an oxymoron)

Legal protection by design:

- integrating the checks and balances of the rule of law
- based on democratic law-making
- combining **legal certainty** with **contestability**

# Legal Protection by Design

- Controllers (GDPR) and providers (AI Act)
- Have a legal obligation to anticipate fundamental rights infringements
- And avoid or mitigate them

- Raising Question Zero: should we do ML here at all?

THE END