Algorithms in practice: bias and the users of AI

Tiphaine Viard – Associate professor
Numérique, Organisation et Société,
Operational AI Ethics

tiphaine.viard@telecom-paris.fr
github.com/TiphaineV
https://tiphainev.github.io
Mastodon: tviard@sciences.social

Why care about users?

Why care about users?

- Egoistically, we *are* users too

- Important for maintaining **trust** in institutions and governance

- Understanding **deployment and usage issues** helps with identifying hindrances, non-use-cases, grounds for discrimination…

- It makes models closer to a form of **metrological realism**

- It helps with **operationalisation** (e.g. in regulation)

Outline

- The state of AI bias
- Technosolutionnism and AI
- The users of AI: AI systems in practice
- The users of AI: Data subjects, explanations and contestation of AI systems

State of AI bias

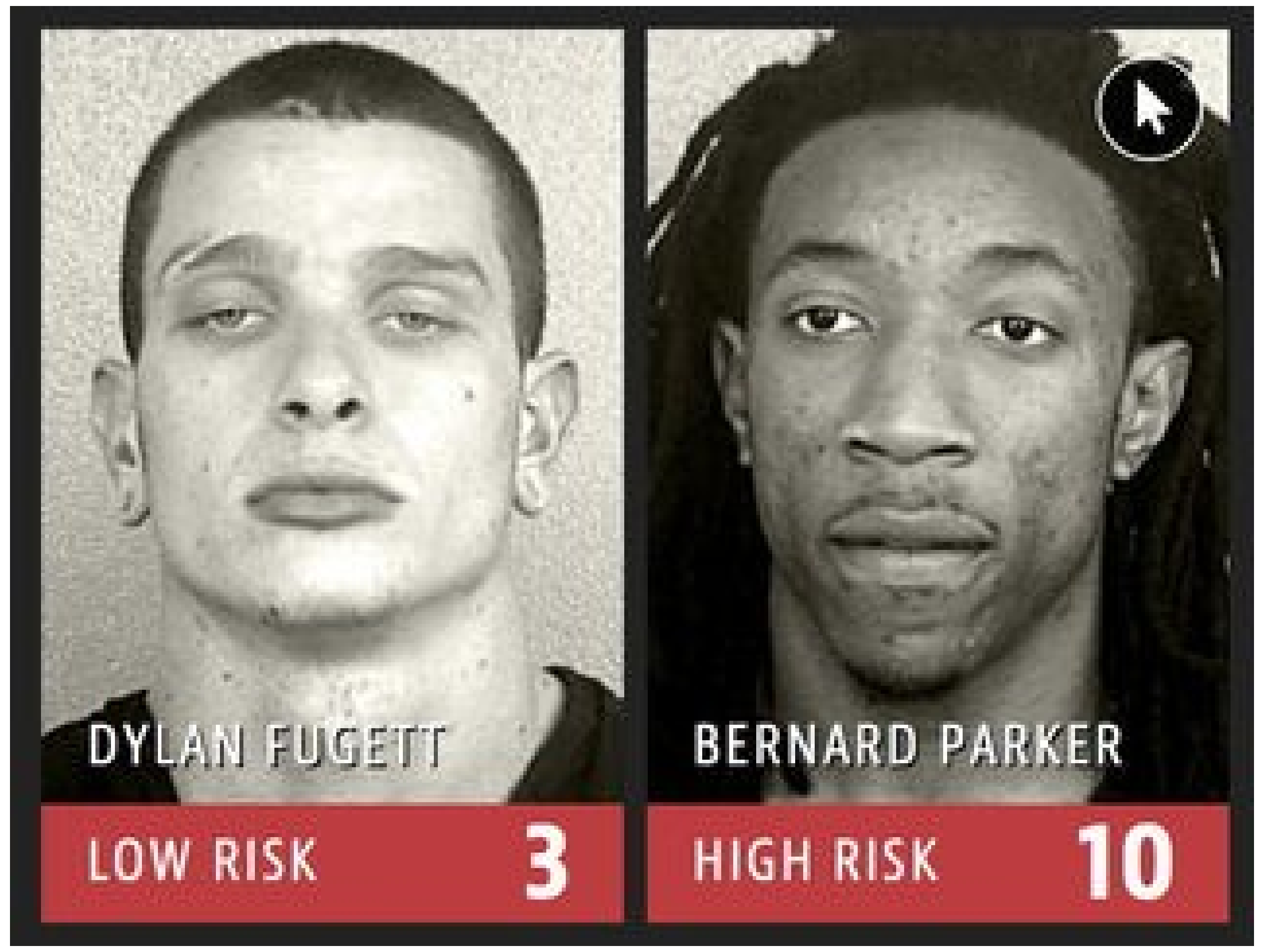
- COMPAS : the flagship case of AI bias (Beaudouin & Maxwell, 2023)

- The limits of COMPAS/bias approach (Kalluri, 2020; Eidelson, 2021)

- Where are we on AI bias now?

A retrospective on the COMPAS affair

Published by ProPublica in 2016, has become the **flagship case of bias and discrimination** in AI systems



The COMPAS case

- Displacing the controversy from law and criminology to public spaces
- Actuarial rules vs AI
- 2 arenas: data science and criminology
- (non)-human judgment vs bail inequality reduction
- Few discussions between arenas (especially DS → Crim)
- Propublica: a media paper with a huge academic impact

AI bias today

COMPAS was 8 years ago – what has changed?

Multiple **debiasing** methods, in particular in NLP (Bolukbasi et al., 2016)

- With mixed results (Prost et al., 2021), sometimes making the resulting embedding *more biased*

An academic field centered on **fairness and transparency in ML models**

- Definitions are still dated and unsatisfactory,
- Common definitions are not necessarily sociologically sound,
- Intersectionality is seldom taken into account.

Reinforces the need debates and discussions on **algocracy** : the way societies and government use and deploy AI systems

Tension point: fair AI versus radical AI

In parallel, multiple affairs tackle the (mis-)use of algorithms:

- **SyRI** (Netherlands), on welfare fraud (CJUE, 2023),
- **Schufa** (Austria), on human oversight and ADM (CJUE, 2024),
- **CNAF** (France), on welfare fraud (Conseil d'Etat, 2024),
- **ETIAS** (EU), on border control and facial recognition (CJUE, 2022)
- **Clearview AI** (UK), on facial recognition, fined again in 2025
- among others…

This questions **the point of fairness** / ethical AI if not attached to deeper change (Kalluri, 2020; Keyes, 2020), the **compounding of injustice** (Eidelson, 2021), and the need for **compassion in the design of algorithmic systems** (Vaccaro et al., 2021)

Technosolutionnism and AI

- Leading example: a mulching proposal
- Defining technosolutionnism
- The problem with technosolutionnism
- The consolidation of AI skeptics / critics

Keyes, O., Hutson, J., & Durbin, M. (2019, May). A mulching proposal: Analysing and improving an algorithmic system for turning the elderly into high-nutrient slurry. In CHI *EA 2019.*

Making systems better with fairness, accountability and transparency

- Applies the **Fairness, Accountability, and Transparency framework** to an algorithm that *“resolves various societal issues around food security and population ageing”*

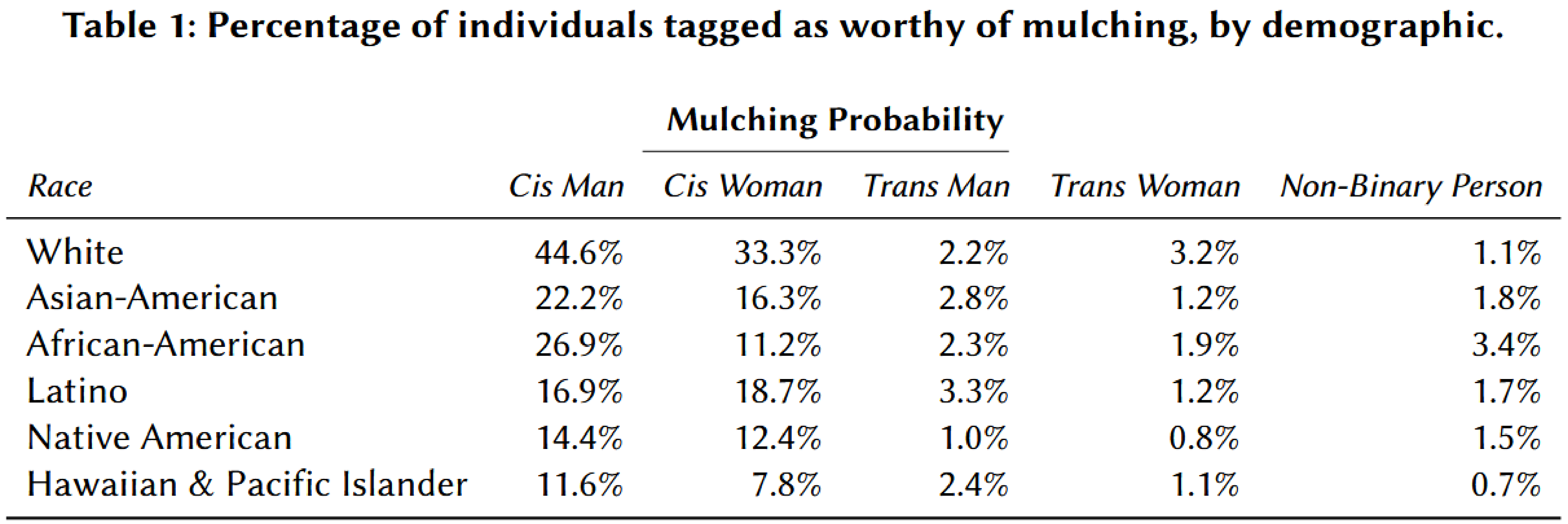
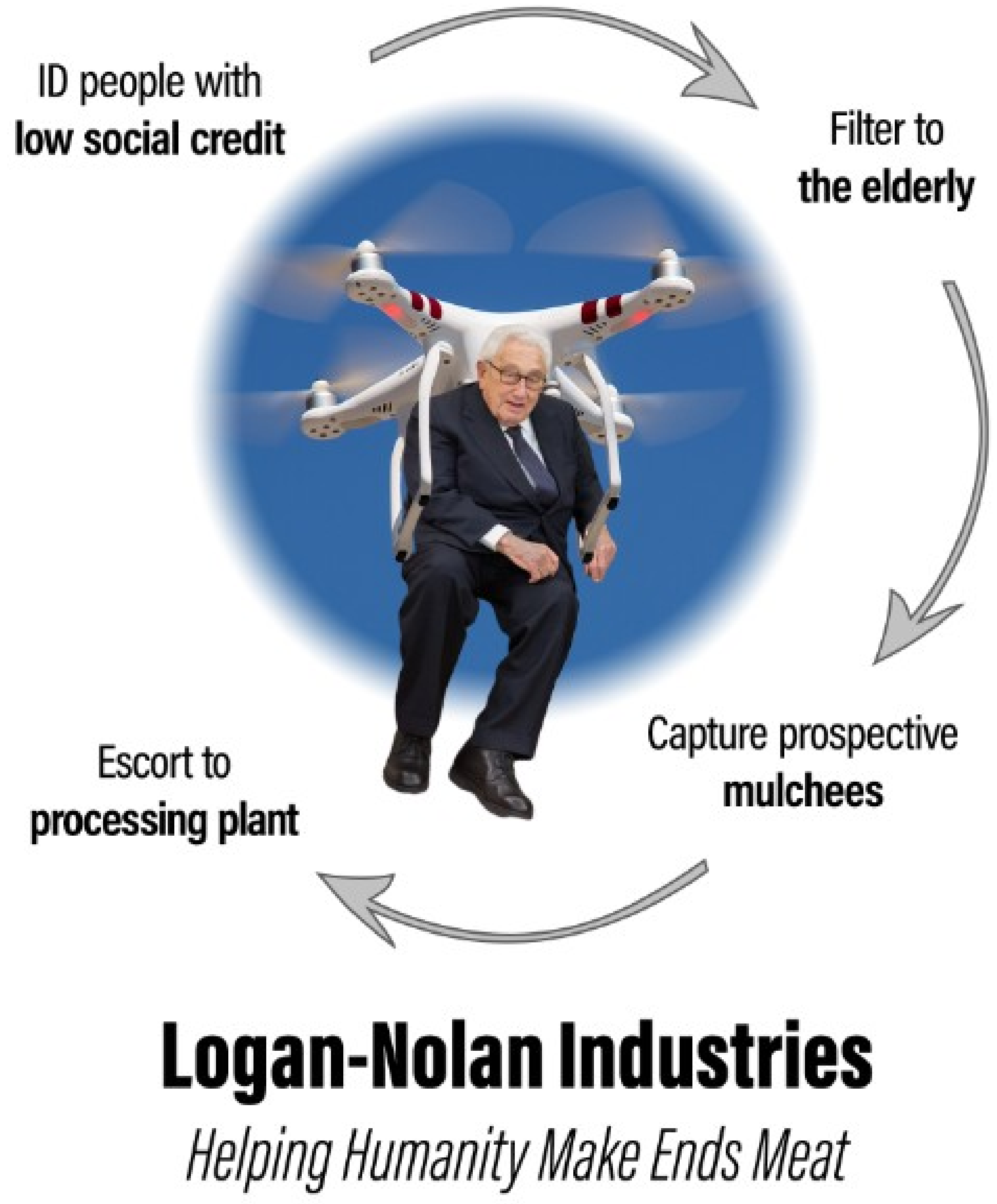


Table 1: Percentage of individuals tagged as worthy of mulching, by demographic.

	Mulching Probability				
Race	*Cis Man*	*Cis Woman*	*Trans Man*	*Trans Woman*	*Non-Binary Person*
White	44.6%	33.3%	2.2%	3.2%	1.1%
Asian-American	22.2%	16.3%	2.8%	1.2%	1.8%
African-American	26.9%	11.2%	2.3%	1.9%	3.4%
Latino	16.9%	18.7%	3.3%	1.2%	1.7%
Native American	14.4%	12.4%	1.0%	0.8%	1.5%
Hawaiian & Pacific Islander	11.6%	7.8%	2.4%	1.1%	0.7%


ID people with low social credit
Filter to the elderly
Capture prospective mulchees
Escort to processing plant
Logan-Nolan Industries
Helping Humanity Make Ends Meat

The algorithm is not fair (demographic parity)!

Keyes, O., Hutson, J., & Durbin, M. (2019, May). A mulching proposal: Analysing and improving an algorithmic system for turning the elderly into high-nutrient slurry. In CHI *EA 2019.*

Making systems better with fairness, accountability and transparency

- Applies the **Fairness, Accountability, and Transparency framework** to an algorithm that *“resolves various societal issues around food security and population ageing”*

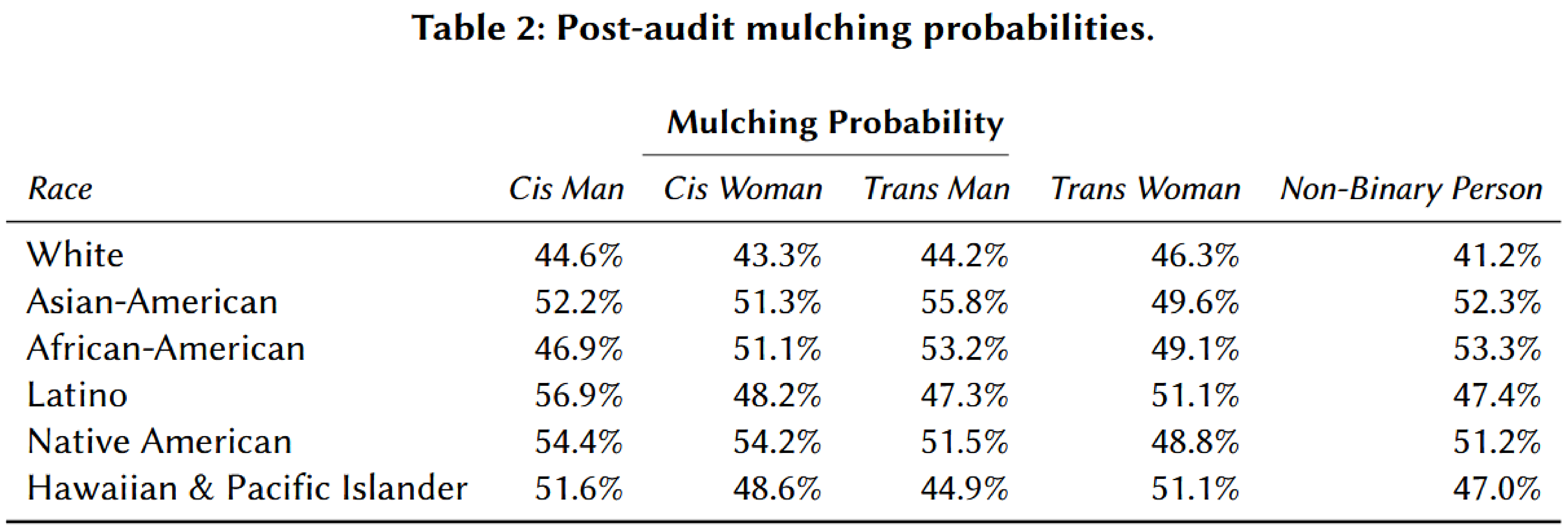
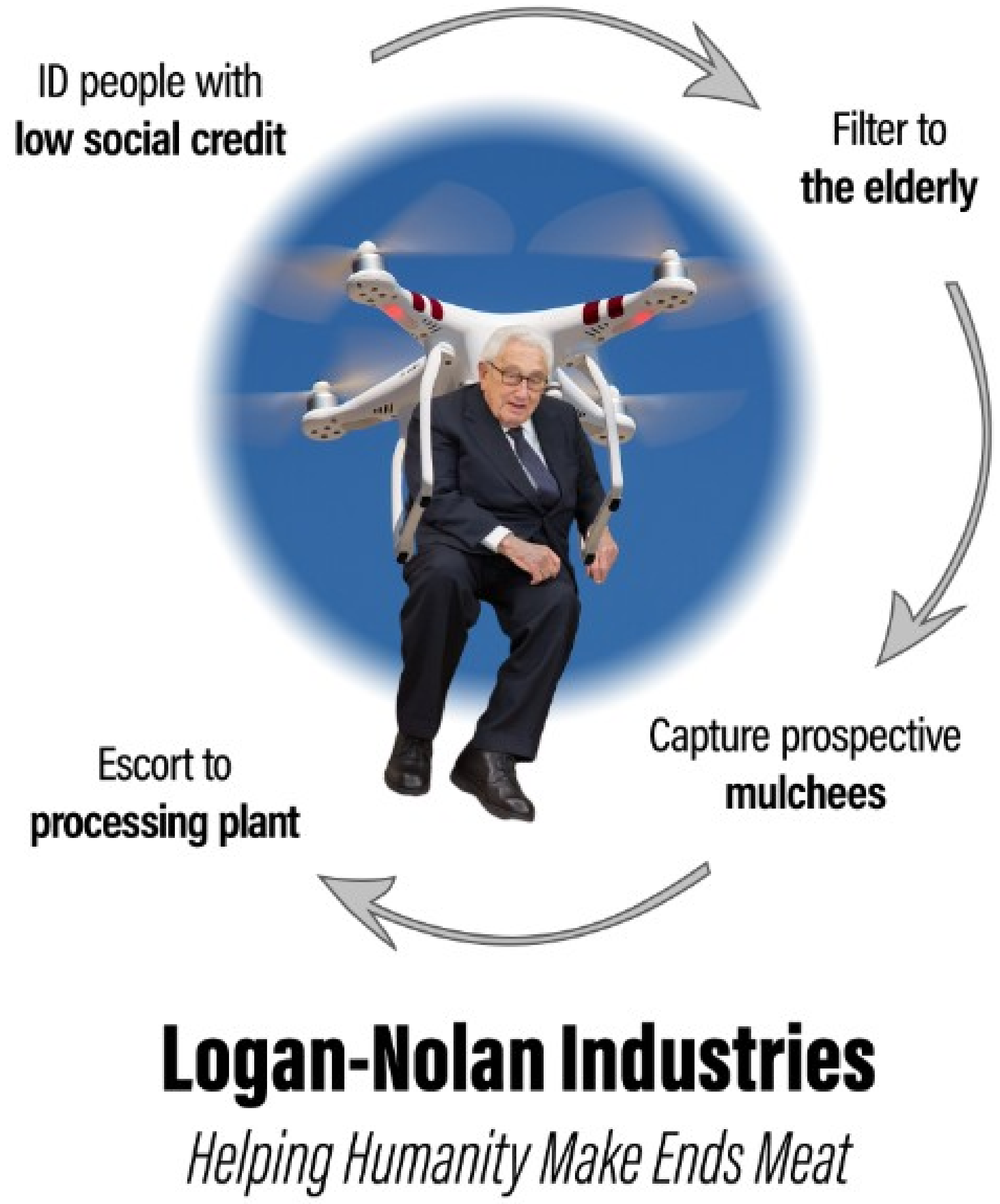


Table 2: Post-audit mulching probabilities.

	Mulching Probability				
Race	*Cis Man*	*Cis Woman*	*Trans Man*	*Trans Woman*	*Non-Binary Person*
White	44.6%	43.3%	44.2%	46.3%	41.2%
Asian-American	52.2%	51.3%	55.8%	49.6%	52.3%
African-American	46.9%	51.1%	53.2%	49.1%	53.3%
Latino	56.9%	48.2%	47.3%	51.1%	47.4%
Native American	54.4%	54.2%	51.5%	48.8%	51.2%
Hawaiian & Pacific Islander	51.6%	48.6%	44.9%	51.1%	47.0%


ID people with **low social credit**
Filter to **the elderly**
Capture prospective **mulchees**
Escort to **processing plant**
Logan-Nolan Industries
Helping Humanity Make Ends Meat

The algorithm is now fair!

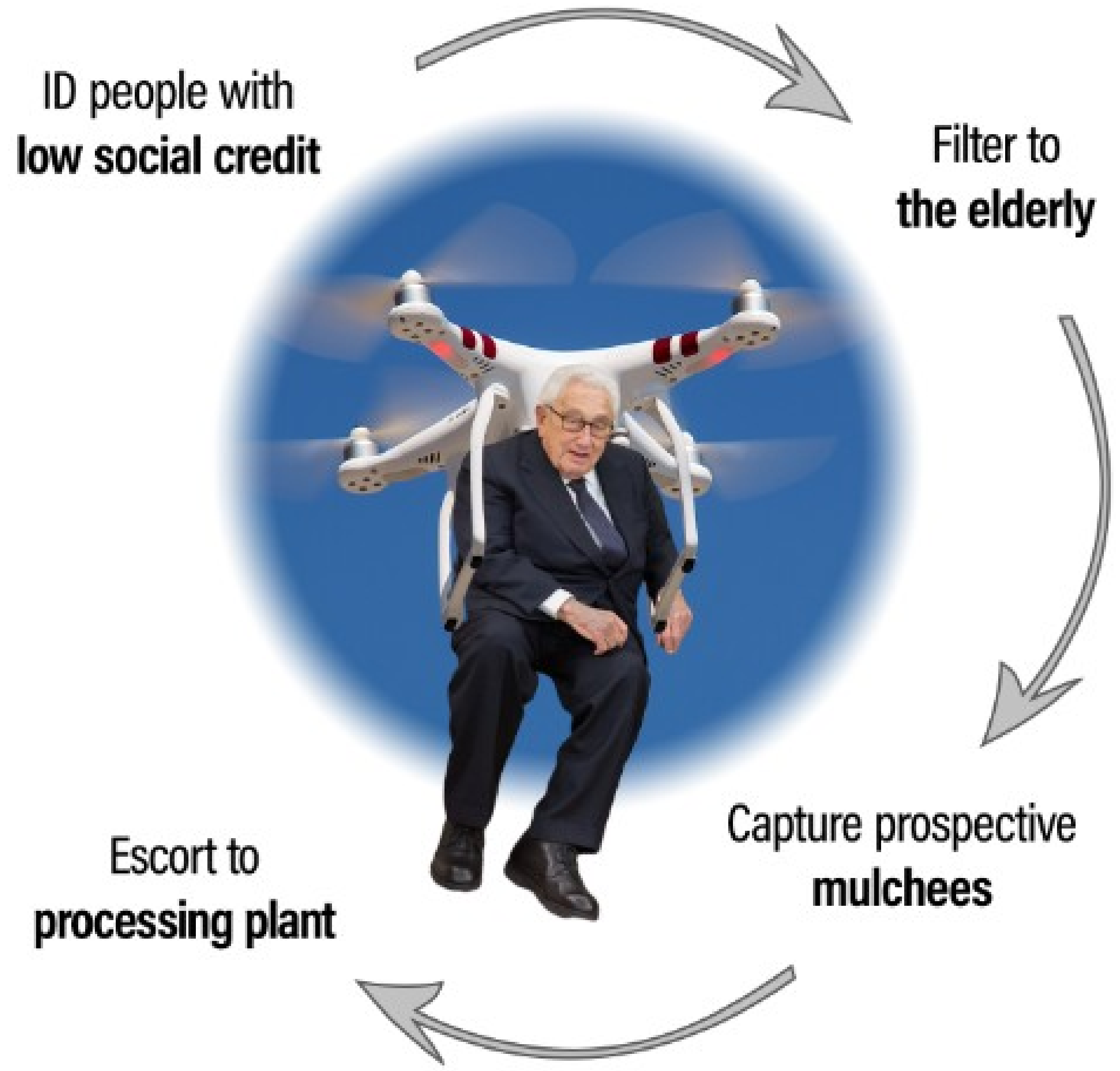
Making systems better with fairness, accountability and transparency

Keyes, O., Hutson, J., & Durbin, M. (2019, May). A mulching proposal: Analysing and improving an algorithmic system for turning the elderly into high-nutrient slurry. In CHI *EA 2019.*

- Applies the **Fairness, Accountability, and Transparency framework** to an algorithm that *“resolves various societal issues around food security and population ageing”*

- **Accountability** (feedback through user survey):
  - Pre-mulching: mulchees are “afforded a ten-second window to state whether their selection was correct or not” + human oversight
  - Post-mulching: food serial number communication + provision of an elderly person of equal or greater wholesomeness and social utility.

The algorithm is now accountable!



Logan-Nolan Industries

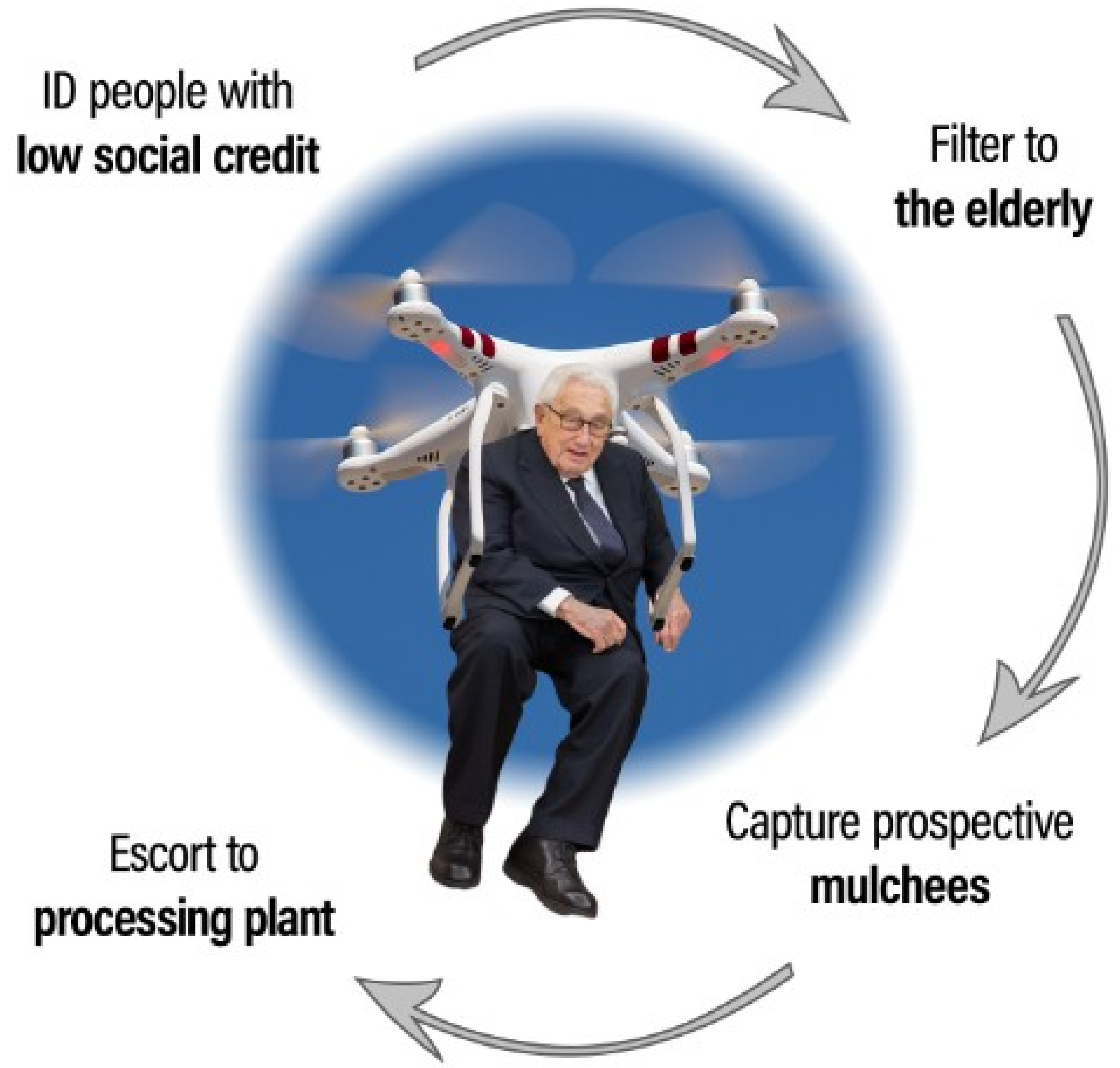
Helping Humanity Make Ends Meat

Keyes, O., Hutson, J., & Durbin, M. (2019, May). A mulching proposal: Analysing and improving an algorithmic system for turning the elderly into high-nutrient slurry. In CHI *EA 2019*.

Making systems better with fairness, accountability and transparency

- Applies the **Fairness, Accountability, and Transparency framework** to an algorithm that *“resolves various societal issues around food security and population ageing”*
- **Transparency**:
  - Areas are marked as “mulching areas” (similar to “videosurveillance areas”)
  - An online platform (mulchme.com) where users can use an interactive website to play with the model and data

The algorithm is now transparent!


ID people with low social credit
Filter to the elderly
Capture prospective mulchees
Escort to processing plant

Logan-Nolan Industries

Helping Humanity Make Ends Meat

Tech neutrality versus tech politics

Beyond the classic rift between **techno-solutionnists** and **techno-skeptics**

Discussing uses only reduces the debate to an experts’ debate

- *e.g.* proximity tracing during Covid lockdowns: debates shifted to privacy protection rather than Bluetooth inadequacy

➔Led to post-hoc rationalisation

Common arguments from tech lobbyists:

- “if we do not do it, others will”;
- “there will be a brain drain”;
- “regulating will thwart innovation”;
- “the new version will solve the problems of the previous version”.

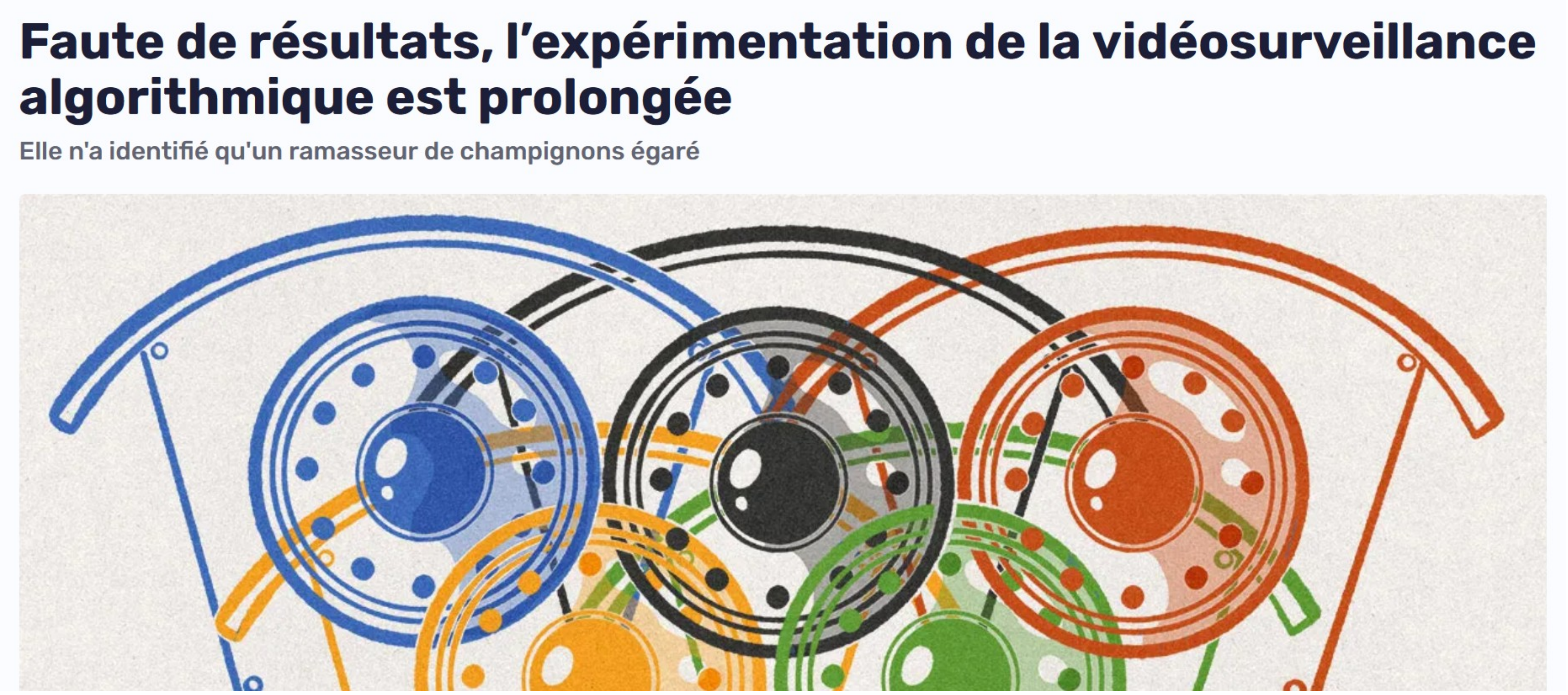
Technosolutionnism

- The idea that technology is a **necessary** part of a solution to a given problem

- Leads to post-rationalization, out of touch technical solutions

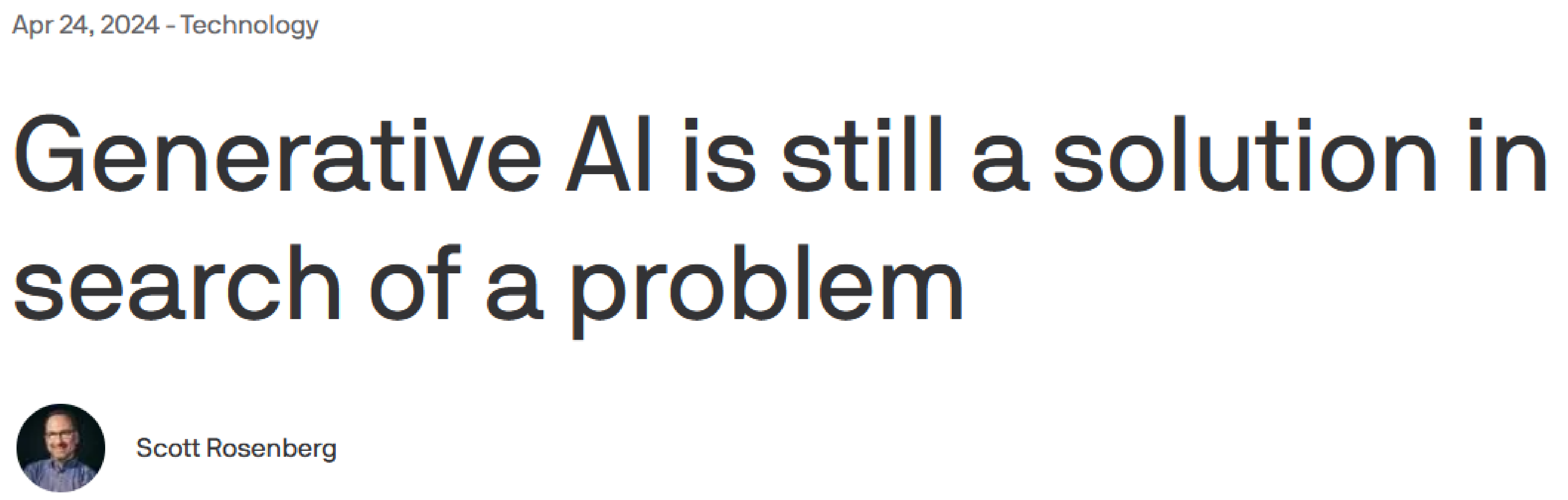
- Is highlighted as a potential **problem for democratic processes and discourse** (Nemitz, 2023; Lafrance, 2024)

- See **fairness & abstraction in sociotechnical systems** (Selbst et al., 2018)



Faute de résultats, l'expérimentation de la vidéosurveillance algorithmique est prolongée

Elle n'a identifié qu'un ramasseur de champignons égaré



Apr 24, 2024 - Technology

Generative AI is still a solution in search of a problem

Scott Rosenberg

The users of AI

- Resisting algorithms (Christin, 2019)

- AI training as reinforcing exploitation

- Another case of training : the French *Cour de Cassation* AI system

Algorithms in practice

How do users react to algorithmic use in their work?

How do **practices** differ from **discourses**? (Jerolmak & Khan, 2014)

Methodology:

- An ethnographic study over 2011-2015
- Two populations: jurists (Paris/NY bar) and web journalists
- 100 interviews; conferences and court sessions
- Algorithms for **defendant scoring** and **trend mining**

Goal: what are the commonalities and differences

between these two fields?

Algorithms in practice: key results

Two **expert fields**: actors sharing a belief in the legitimacy of specific forms of knowledge as a basis of intervention in public affairs (Bourdieu, 1993. Collins et al., 2007)

Different from professions: positioning and entry barriers

Key differences:

- Law and journalism have **different barriers to entry** (strong vs close to none)
- One is **non-profit oriented**, the other is **profit-oriented**,
- Different **stances towards digital technologies**,
- Different conceptualisations of the **expertise** with respect to **their identity**

Algorithms in practice: commonalities

Decoupling (Meyer& Rowan, 1977): separating management discourses from employee practices

Algorithms are either **ignored** or **actively resisted**

Buffering: Foot-dragging, gaming, open critique

- **Foot dragging**: ignoring the tools, placing their results at the end of reports…
- **Gaming**: clickbait titles, article slot times, court cases selection…
- **Open critique**: success for the algorithm's criteria is **disaligned** from success internalized by individuals (e.g. media image)

This leads to a **displacement of subjective judgment** and **social quantification** (Espeland et al., 2007)

Algorithms in practice: deeper differences

Distinct algorithmic imaginaries: “ways of thinking about what algorithms are, what they should be, and how they function” (Bucher, 2016:30)

Journalists do not question the tool but are ambivalent on its finality

- *i.e.* they question the alignment between the algorithm’s goal and the journal’s goal

In **courtsrooms**, algorithms are :

- openly criticized as “*crude*” and “*problematic*”,
- the **for-profit tools are criticized**,
- the **absence of legal precedent** slows change

Algorithms in practice: different imaginaries

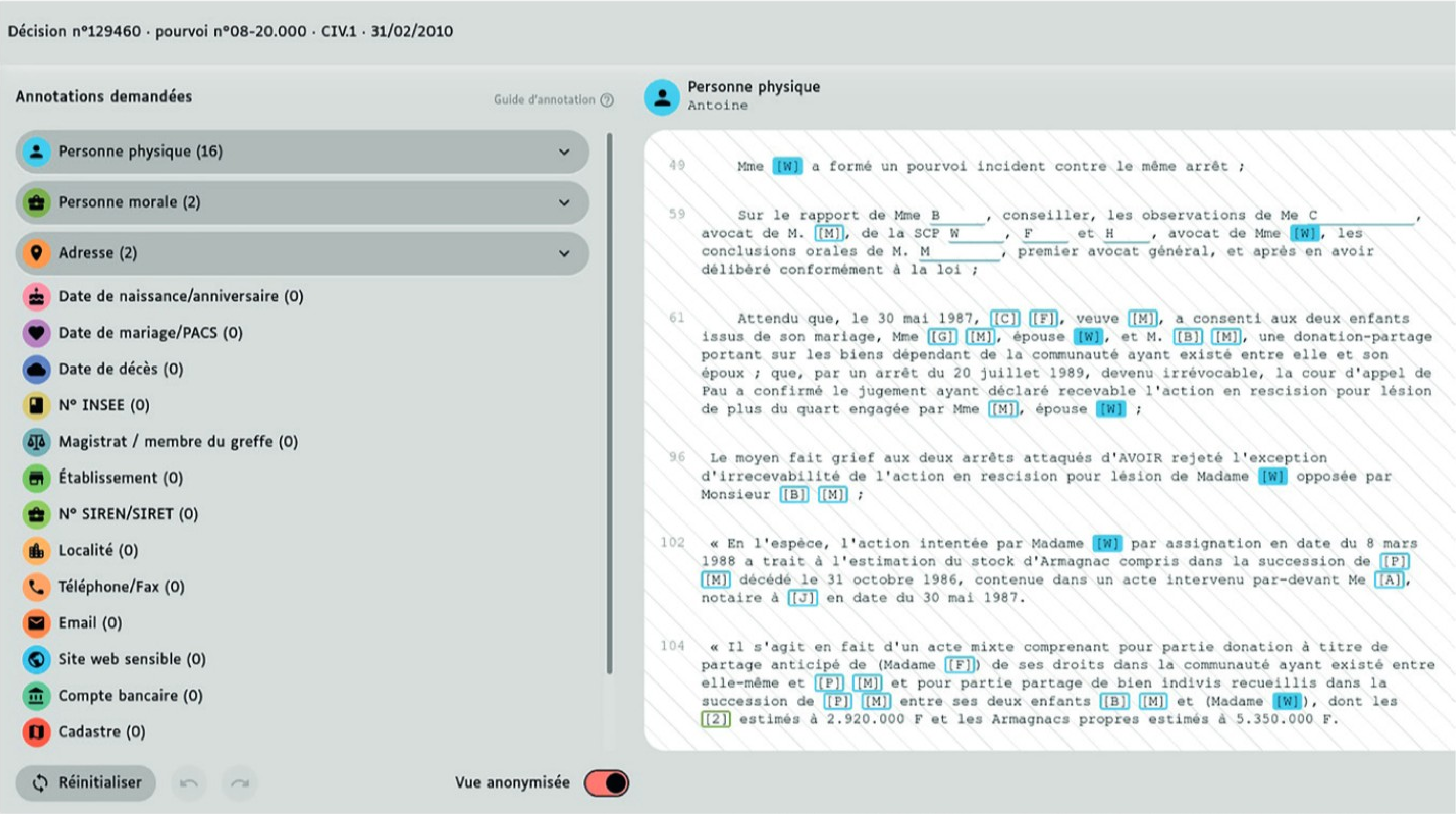
Why are these imaginaries so different?

- Journalists share with editors the **knowledge of need for profit**;
- Journalists evolve in a **porous and heteronomous field**;
- Journalists **value digital technologies**, and **appreciate immediacy**.

- Judges and prosecutors a **public servants appointed** by the state;
- Judges attach their **decisions to their identity**;
- Judges are **reluctant to digital technologies**, and **value conservative decisions**

The case of the French Cour de Cassation

- A **named entity recognition algorithm** for court case anonymization
- Trained **“in-house”**, by C. Cass civil servants, over 2 years
- No **end of training**: constant fine-tuning is needed (now a permanent team)
- No conception of AI (initially), **no knowledge of digital workers**
- The AI system is generally called “the software”, “the system”, without referencing AI
- It becomes an **absent** “**colleague**”, with **nuances with classical qualitative coding in its ends**
- Mobilising **empathy skills**: how to ensure the reader will understand the case?



Décision n°129460 · pourvoi n°08-20.000 · CIV.1 · 31/02/2010

Annotations demandées

- Personne physique (16)
- Personne morale (2)
- Adresse (2)
- Date de naissance/anniversaire (0)
- Date de mariage/PACS (0)
- Date de décès (0)
- N° INSEE (0)
- Magistrat / membre du greffe (0)
- Établissement (0)
- N° SIREN/SIRET (0)
- Localité (0)
- Téléphone/Fax (0)
- Email (0)
- Site web sensible (0)
- Compte bancaire (0)
- Cadastre (0)

Personne physique
Antoine

49 Mme [W] a formé un pourvoi incident contre le même arrêt ;

59 Sur le rapport de Mme B______, conseiller, les observations de Me C____________, avocat de M. [M], de la SCP W______, F_____ et H_____, avocat de Mme [W], les conclusions orales de M. M________, premier avocat général, et après en avoir délibéré conformément à la loi ;

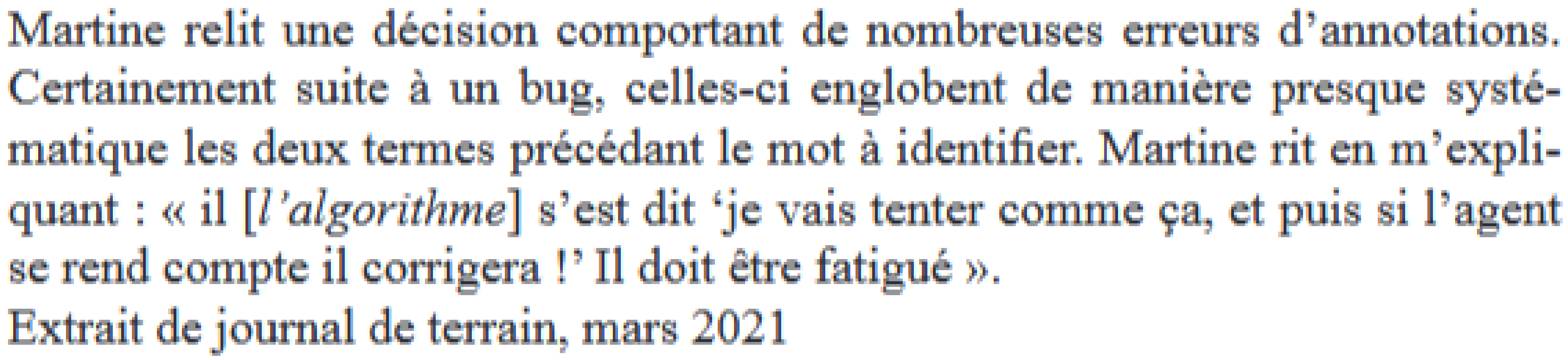
61 Attendu que, le 30 mai 1987, [C] [F], veuve [M], a consenti aux deux enfants issus de son mariage, Mme [G] [M], épouse [W], et M. [B] [M], une donation-partage portant sur les biens dépendant de la communauté ayant existé entre elle et son époux ; que, par un arrêt du 20 juillet 1989, devenu irrévocable, la cour d'appel de Pau a confirmé le jugement ayant déclaré recevable l'action en rescision pour lésion de plus du quart engagée par Mme [M], épouse [W] ;

96 Le moyen fait grief aux deux arrêts attaqués d'AVOIR rejeté l'exception d'irrecevabilité de l'action en rescision pour lésion de Madame [W] opposée par Monsieur [B] [M] ;

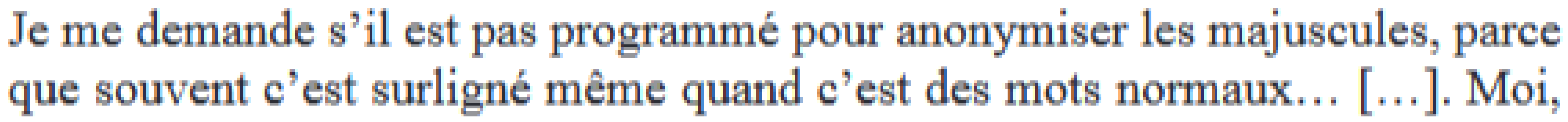
102 « En l'espèce, l'action intentée par Madame [W] par assignation en date du 8 mars 1988 a trait à l'estimation du stock d'Armagnac compris dans la succession de [P] [M] décédé le 31 octobre 1986, contenue dans un acte intervenu par-devant Me [A], notaire à [J] en date du 30 mai 1987.

104 « Il s'agit en fait d'un acte mixte comprenant pour partie donation à titre de partage anticipé de (Madame [F]) de ses droits dans la communauté ayant existé entre elle-même et [P] [M] et pour partie partage de bien indivis recueillis dans la succession de [P] [M] entre ses deux enfants [B] [M] et (Madame [W]), dont les [2] estimés à 2.920.000 F et les Armagnacs propres estimés à 5.350.000 F.

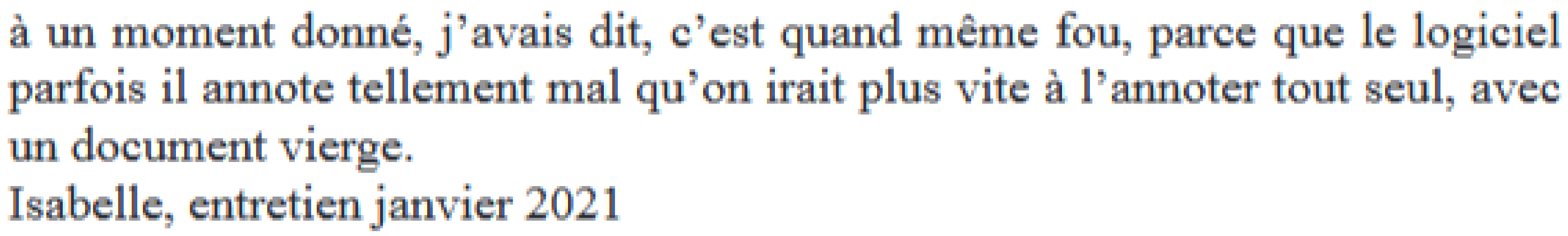
Fieldwork excerpts



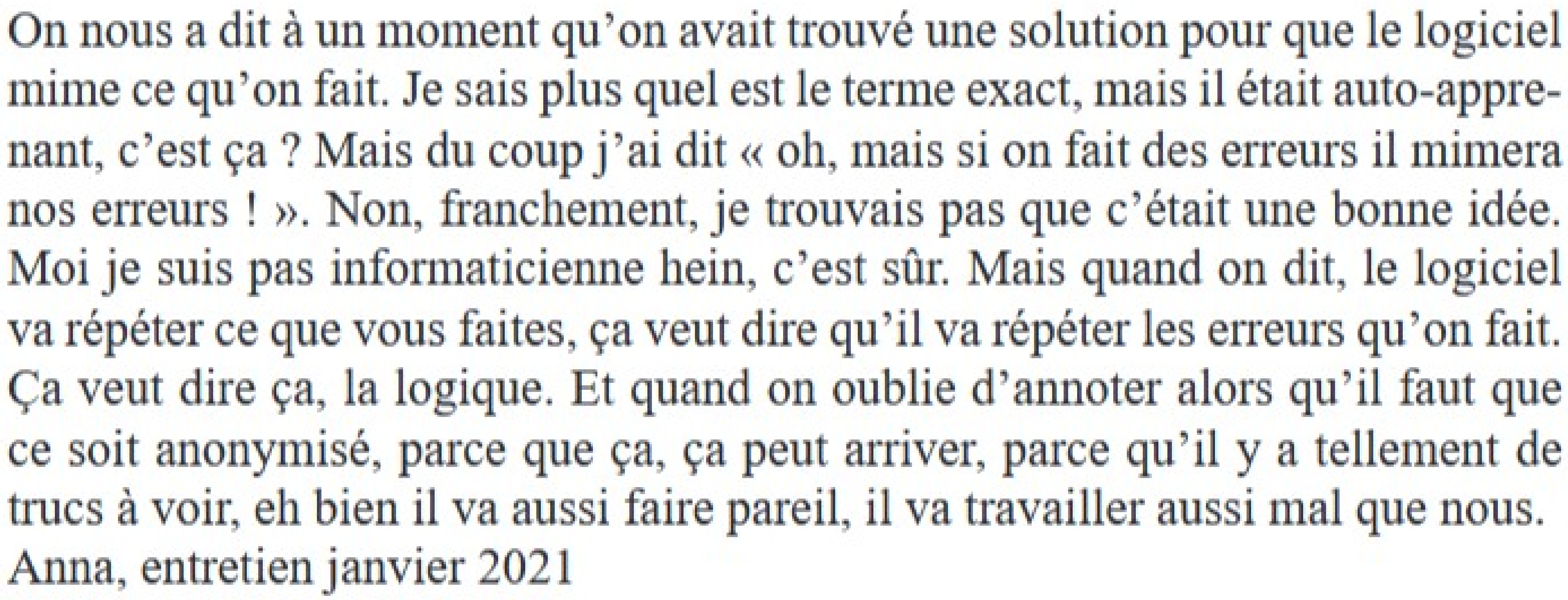
Martine relit une décision comportant de nombreuses erreurs d'annotations. Certainement suite à un bug, celles-ci englobent de manière presque systématique les deux termes précédant le mot à identifier. Martine rit en m'expliquant : « il [*l'algorithme*] s'est dit 'je vais tenter comme ça, et puis si l'agent se rend compte il corrigera !' Il doit être fatigué ».
Extrait de journal de terrain, mars 2021



Je me demande s'il est pas programmé pour anonymiser les majuscules, parce que souvent c'est surligné même quand c'est des mots normaux… […]. Moi,



à un moment donné, j'avais dit, c'est quand même fou, parce que le logiciel parfois il annote tellement mal qu'on irait plus vite à l'annoter tout seul, avec un document vierge.
Isabelle, entretien janvier 2021



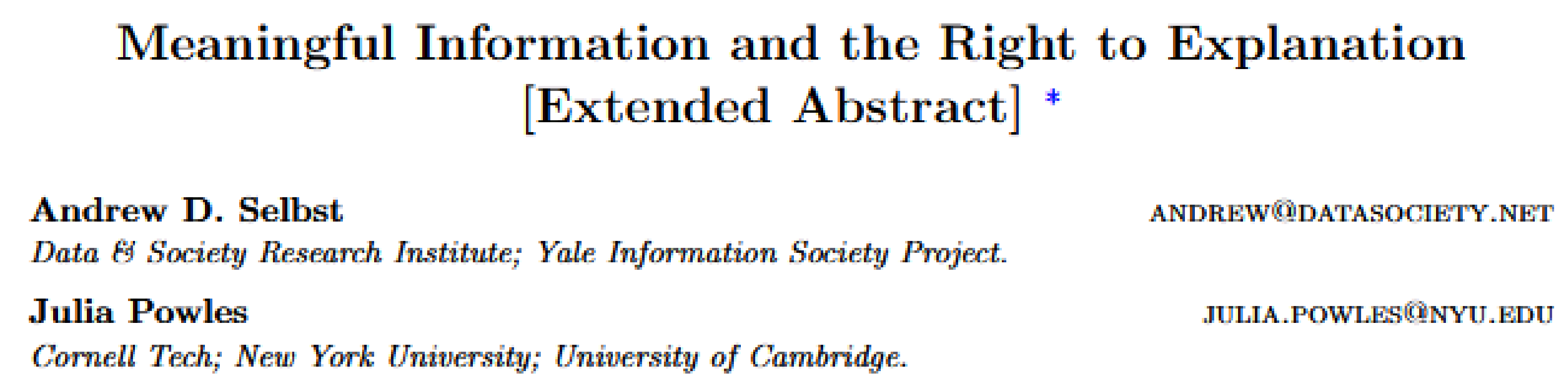
On nous a dit à un moment qu'on avait trouvé une solution pour que le logiciel mime ce qu'on fait. Je sais plus quel est le terme exact, mais il était auto-apprenant, c'est ça ? Mais du coup j'ai dit « oh, mais si on fait des erreurs il mimera nos erreurs ! ». Non, franchement, je trouvais pas que c'était une bonne idée. Moi je suis pas informaticienne hein, c'est sûr. Mais quand on dit, le logiciel va répéter ce que vous faites, ça veut dire qu'il va répéter les erreurs qu'on fait. Ça veut dire ça, la logique. Et quand on oublie d'annoter alors qu'il faut que ce soit anonymisé, parce que ça, ça peut arriver, parce qu'il y a tellement de trucs à voir, eh bien il va aussi faire pareil, il va travailler aussi mal que nous.
Anna, entretien janvier 2021

The end users of AI

- The *right to an explanation* in EU law
- What are explanations worth for? Contesting AI systems

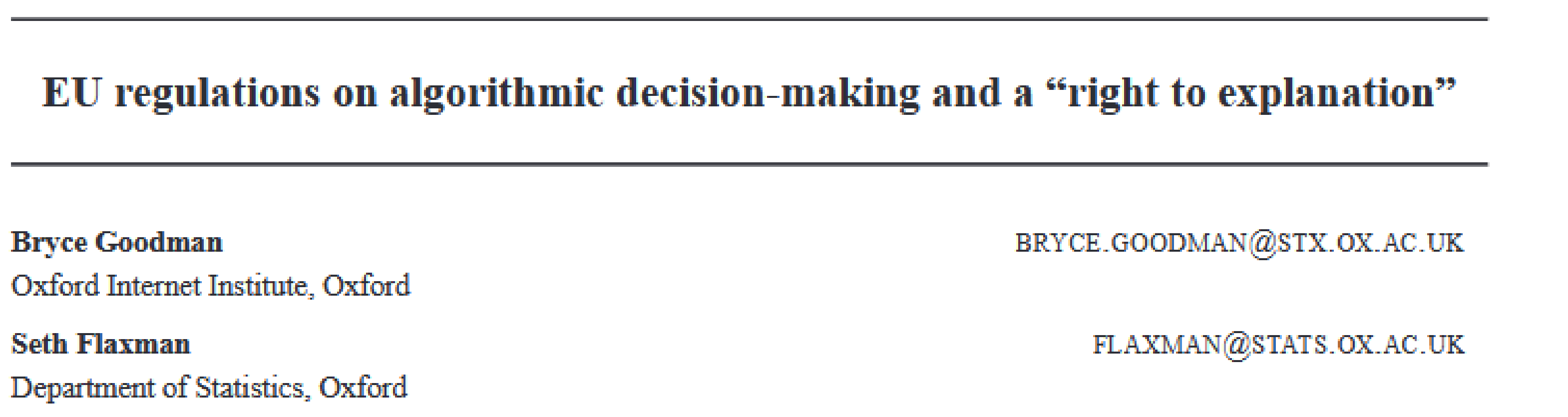
The right to an explanation

- What **form** should explanations take?
  - Selective, mutable, dialogic (Miller, 2020)
  - Contrastive explanations and complexity drops (Dessalles, 2020)
- Who are we explaining for?
  - Experts
  - Laypeople
  - Regulators…
- **Why** explain? *descriptive*, *explanatory*, *normative*, *contestable*.



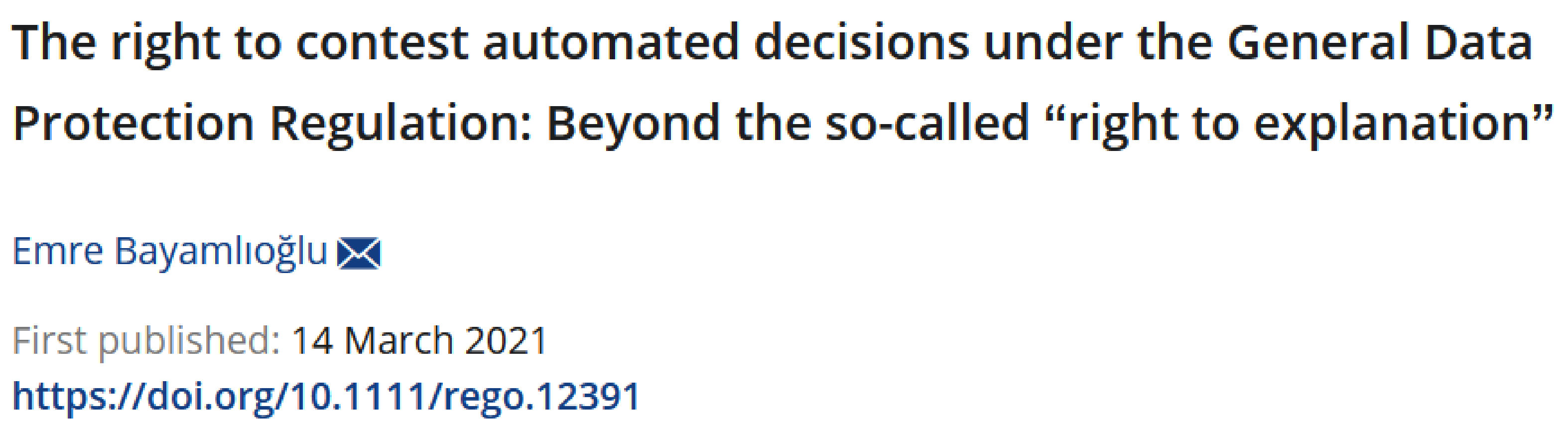
Meaningful Information and the Right to Explanation
[Extended Abstract] *

Andrew D. Selbst ANDREW@DATASOCIETY.NET
Data & Society Research Institute; Yale Information Society Project.
Julia Powles JULIA.POWLES@NYU.EDU
Cornell Tech; New York University; University of Cambridge.



EU regulations on algorithmic decision-making and a "right to explanation"

Bryce Goodman BRYCE.GOODMAN@STX.OX.AC.UK
Oxford Internet Institute, Oxford
Seth Flaxman FLAXMAN@STATS.OX.AC.UK
Department of Statistics, Oxford

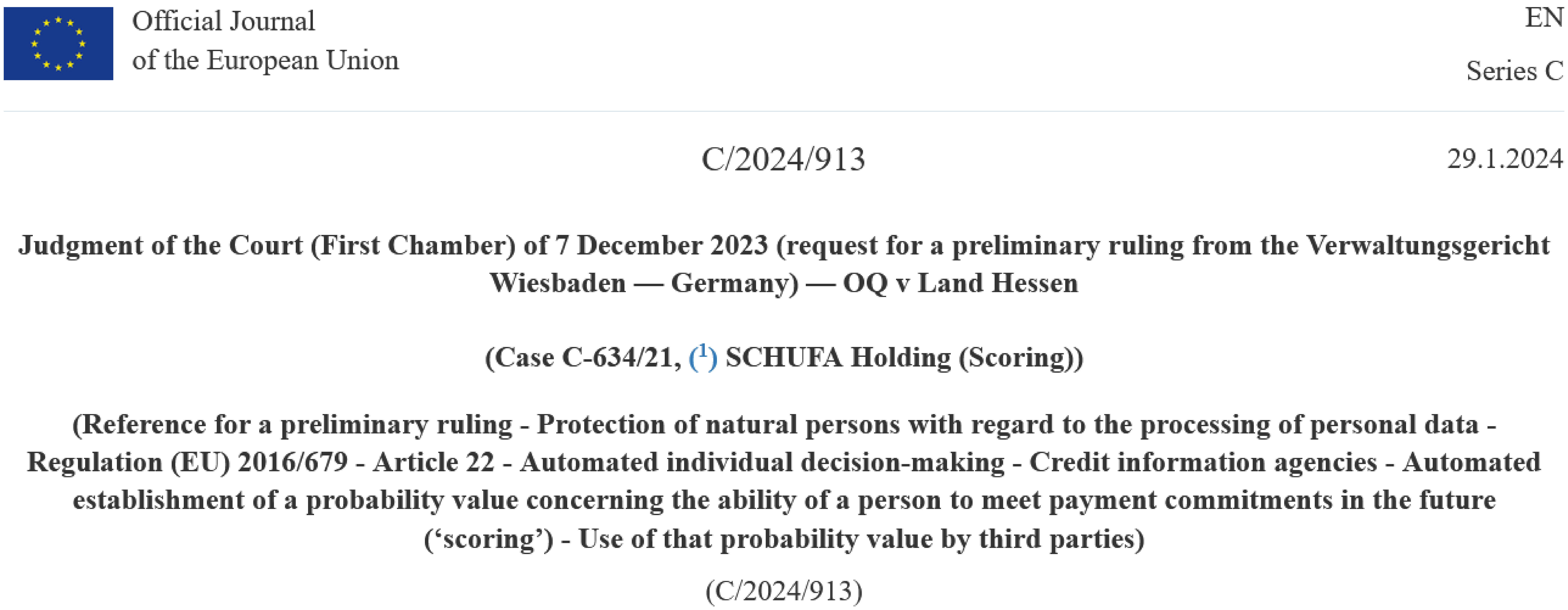


The right to contest automated decisions under the General Data Protection Regulation: Beyond the so-called "right to explanation"

Emre Bayamlıoğlu

First published: 14 March 2021
https://doi.org/10.1111/rego.12391

An increasingly relevant right to explanation



Official Journal
of the European Union

EN

Series C

C/2024/913

29.1.2024

Judgment of the Court (First Chamber) of 7 December 2023 (request for a preliminary ruling from the Verwaltungsgericht Wiesbaden — Germany) — OQ v Land Hessen

(Case C-634/21, [1]) SCHUFA Holding (Scoring))

(Reference for a preliminary ruling - Protection of natural persons with regard to the processing of personal data - Regulation (EU) 2016/679 - Article 22 - Automated individual decision-making - Credit information agencies - Automated establishment of a probability value concerning the ability of a person to meet payment commitments in the future (‘scoring’) - Use of that probability value by third parties)

(C/2024/913)

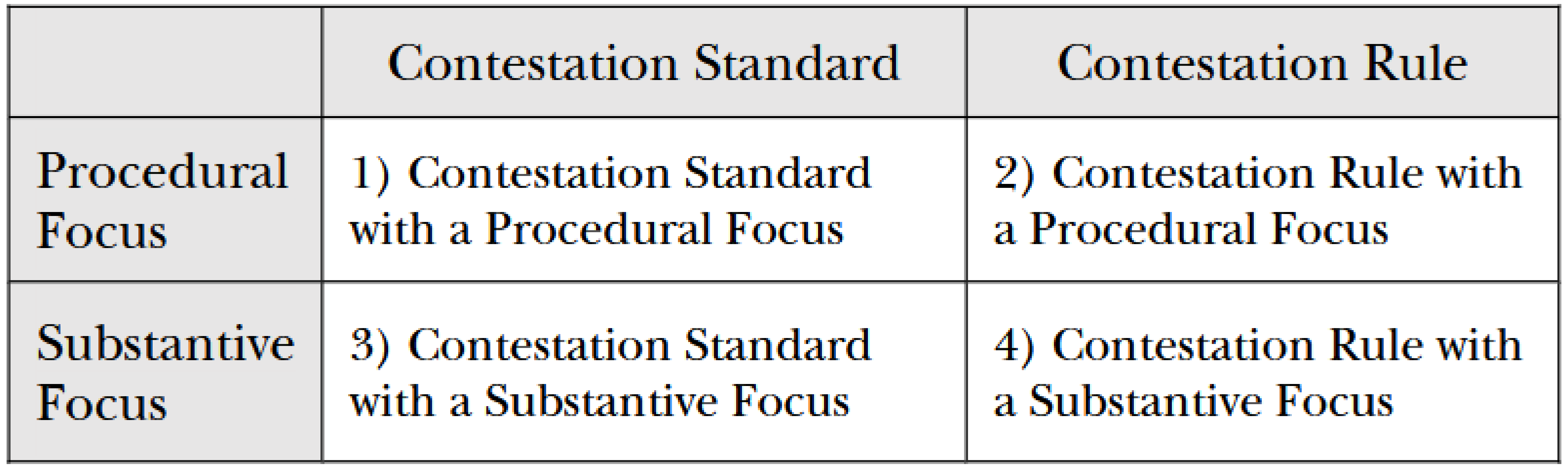
The right to contest AI

- AI is used in high-stakes processes (university admissions, loans, etc.)
- Regulatory approaches (esp. USA) focus on systemic governance rather than individual rights
- Proposal (Kaminsky et al., 2021): **individual right to contest AI**, mimicking due process
- Puts the onus on individuals to challenge unfair AI decisions
- Already fitting in GDPR framework
- What is a good contestation : for users? for agencies?

Archetypes for grounds for contestation

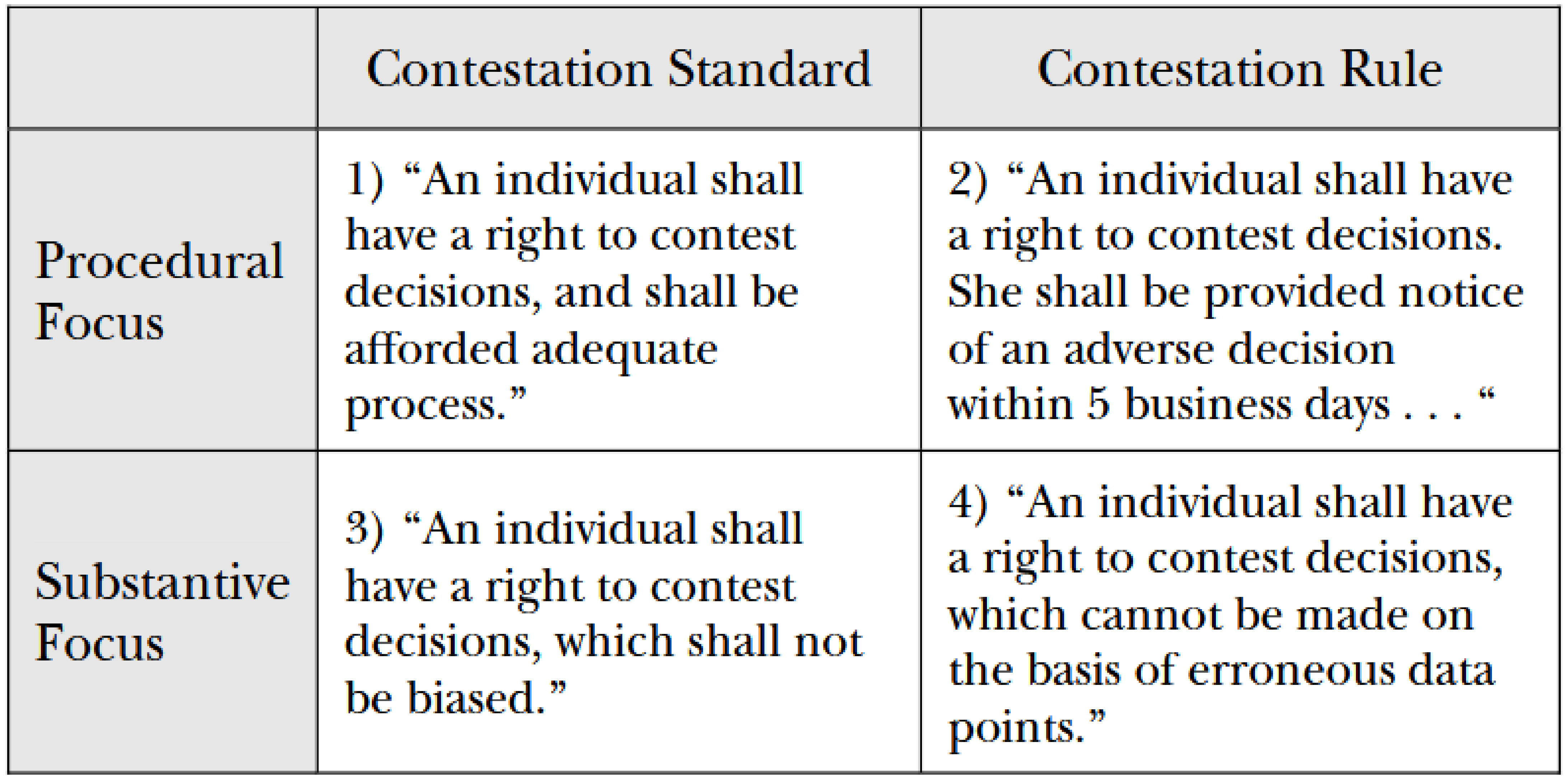
Kaminski, M. E., & Urban, J. M. (2021). The right to contest AI. *Columbia Law Review*, *121*(7), 1957-2048.

TABLE I: THE CONTESTATION ARCHETYPES



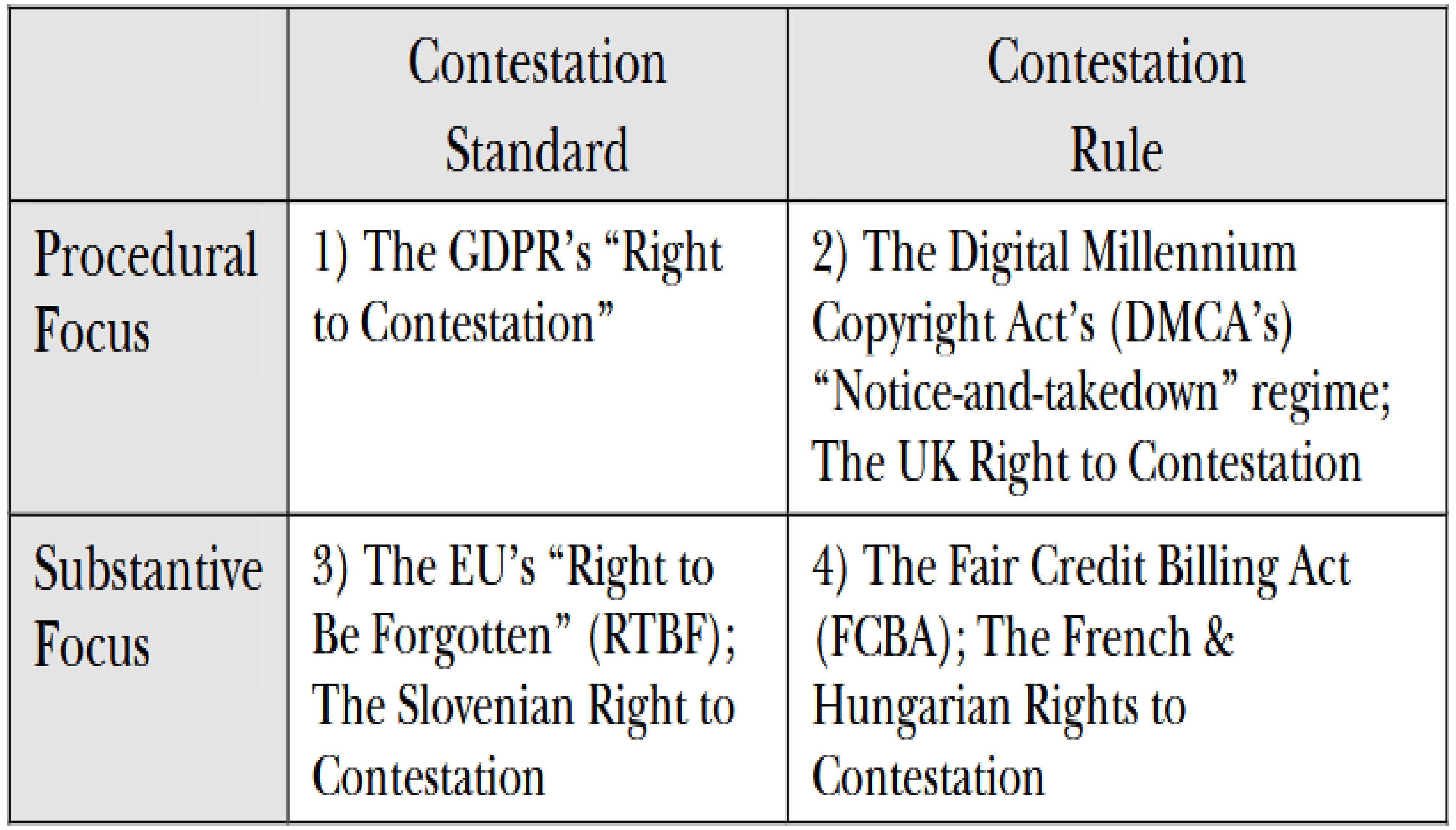
	Contestation Standard	Contestation Rule
Procedural Focus	1) Contestation Standard with a Procedural Focus	2) Contestation Rule with a Procedural Focus
Substantive Focus	3) Contestation Standard with a Substantive Focus	4) Contestation Rule with a Substantive Focus

TABLE II: HYPOTHETICAL EXAMPLES OF THE CONTESTATION ARCHETYPES



	Contestation Standard	Contestation Rule
Procedural Focus	1) "An individual shall have a right to contest decisions, and shall be afforded adequate process."	2) "An individual shall have a right to contest decisions. She shall be provided notice of an adverse decision within 5 business days . . . "
Substantive Focus	3) "An individual shall have a right to contest decisions, which shall not be biased."	4) "An individual shall have a right to contest decisions, which cannot be made on the basis of erroneous data points."

TABLE III: THE CONTESTATION ARCHETYPES IN ACTION

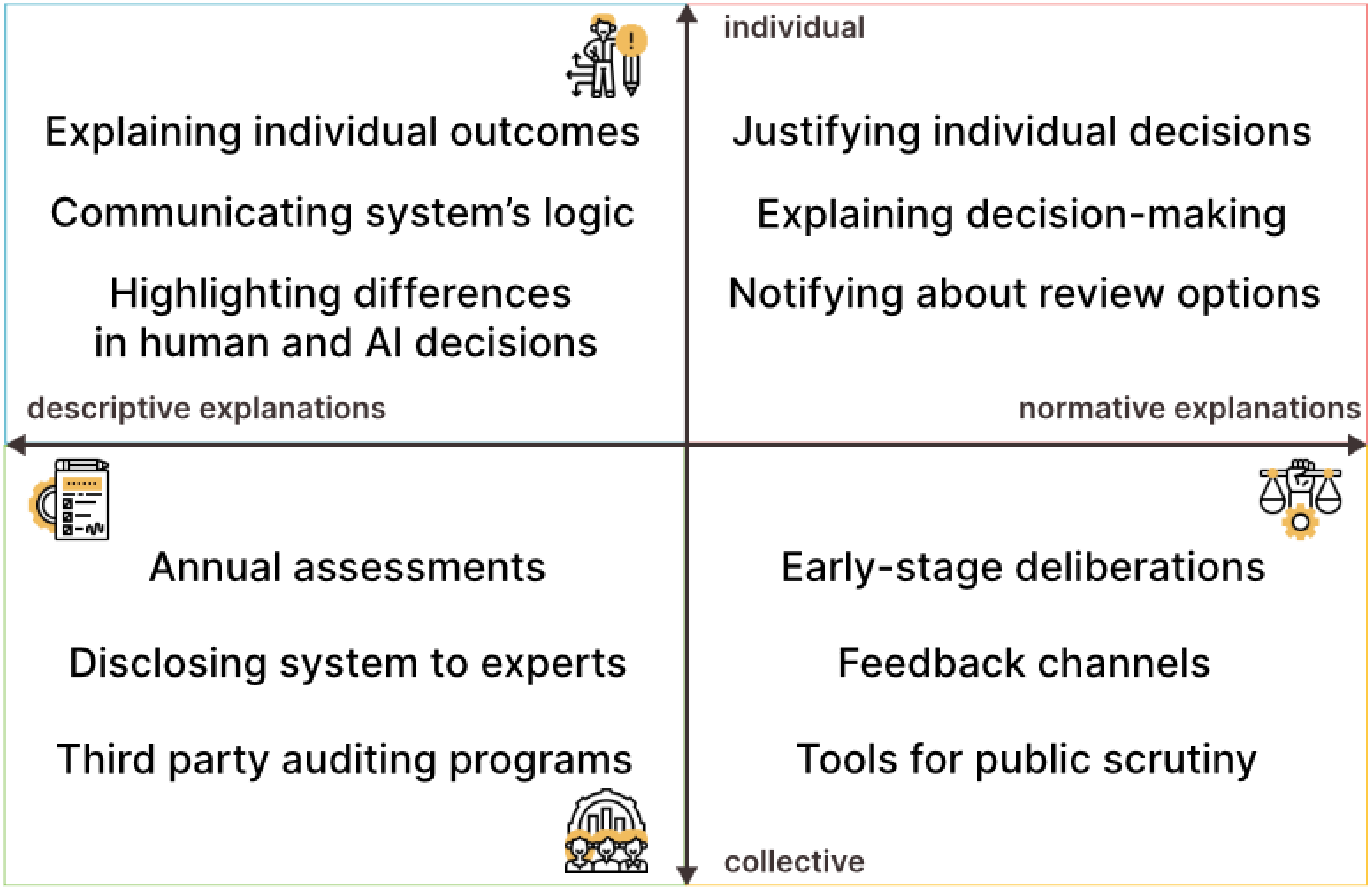


	Contestation Standard	Contestation Rule
Procedural Focus	1) The GDPR's "Right to Contestation"	2) The Digital Millennium Copyright Act's (DMCA's) "Notice-and-takedown" regime; The UK Right to Contestation
Substantive Focus	3) The EU's "Right to Be Forgotten" (RTBF); The Slovenian Right to Contestation	4) The Fair Credit Billing Act (FCBA); The French & Hungarian Rights to Contestation

Contestation goals, mechanisms and limits

- **Goals**: benefitting citizen empowerment, acceptability of decisions, suitability of system’s development, preventing gaming
- **Mechanisms**: better understood for explainability than contestability; contestability is harder to identify as overarching
- **Limits**: individual contestation does not resolve information asymmetry and power imbalances; users fight for their case only; contestability is seen as a societal/political tenant of explainability

How to contest AI decisions?



individual

Explaining individual outcomes

Communicating system’s logic

Highlighting differences
in human and AI decisions

descriptive explanations

Justifying individual decisions

Explaining decision-making

Notifying about review options

normative explanations

Annual assessments

Disclosing system to experts

Third party auditing programs

Early-stage deliberations

Feedback channels

Tools for public scrutiny

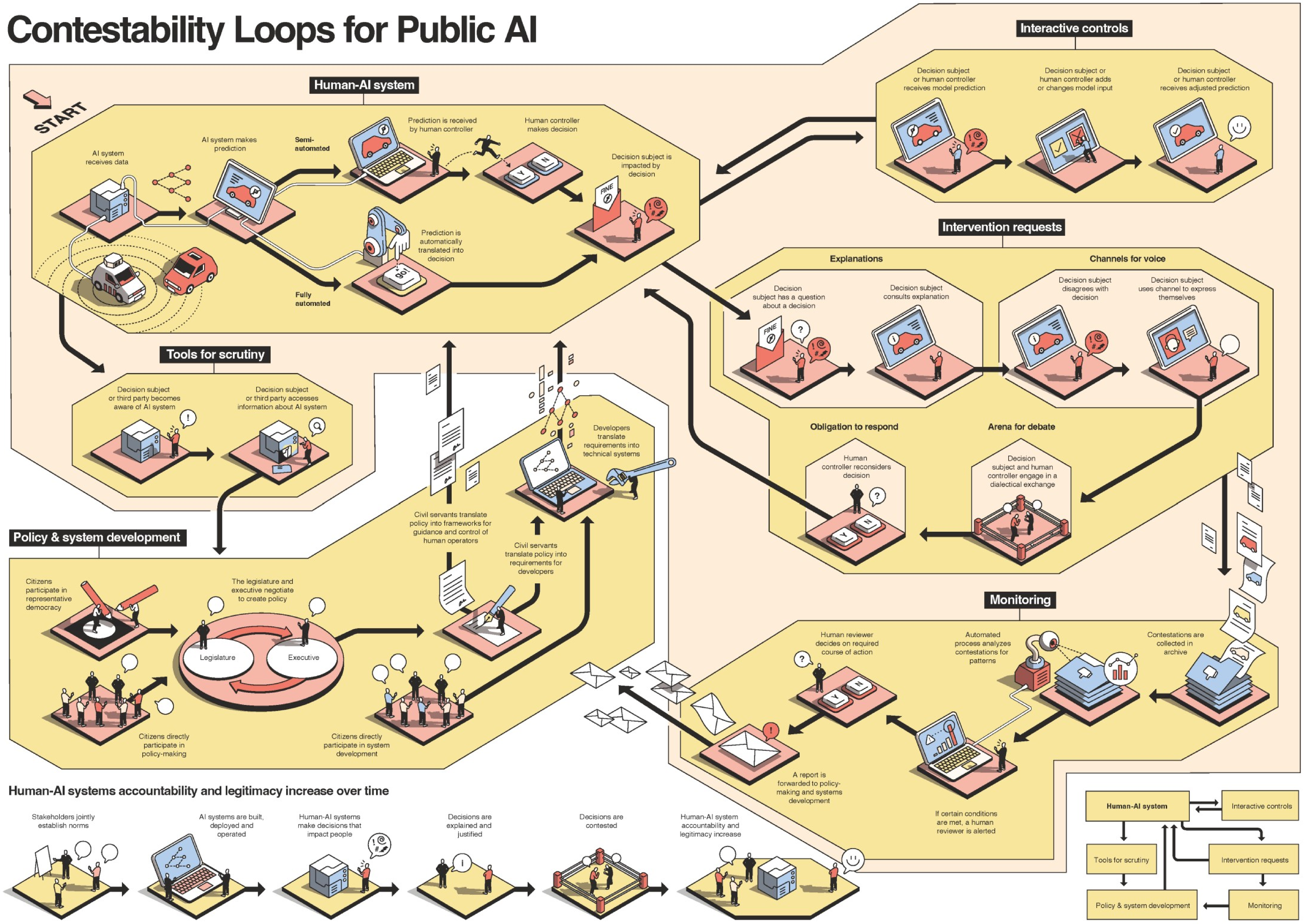
collective

Designing contestable systems

- Law is one thing; how to ensure contestation is **encouraged**?
- Design studies, in particular **Value-Sensitive Design** (VSD)
- (next slide) A focus on **public AI**, *i.e.* AI used in public services
- **Incorporates descriptive and normative aspects, procedural and substantial…**

Alfrink, K., Keller, I., Kortuem, G., & Doorn, N. (2023). Contestable AI by design: Towards a framework. *Minds and Machines*, *33*(4), 613-639.

Contestability Loops for Public AI



START

Human-AI system

AI system receives data

AI system makes prediction

Semi-automated

Prediction is received by human controller

Human controller makes decision

Decision subject is impacted by decision

Prediction is automatically translated into decision

Fully automated

Tools for scrutiny

Decision subject or third party becomes aware of AI system

Decision subject or third party accesses information about AI system

Policy & system development

Citizens participate in representative democracy

The legislature and executive negotiate to create policy

Legislature

Executive

Citizens directly participate in policy-making

Citizens directly participate in system development

Civil servants translate policy into frameworks for guidance and control of human operators

Civil servants translate policy into requirements for developers

Developers translate requirements into technical systems

Interactive controls

Decision subject or human controller receives model prediction

Decision subject or human controller adds or changes model input

Decision subject or human controller receives adjusted prediction

Intervention requests

Explanations

Decision subject has a question about a decision

Decision subject consults explanation

Channels for voice

Decision subject disagrees with decision

Decision subject uses channel to express themselves

Obligation to respond

Human controller reconsiders decision

Arena for debate

Decision subject and human controller engage in a dialectical exchange

Monitoring

Human reviewer decides on required course of action

Automated process analyzes contestations for patterns

Contestations are collected in archive

A report is forwarded to policy-making and systems development

If certain conditions are met, a human reviewer is alerted

Human-AI systems accountability and legitimacy increase over time

Stakeholders jointly establish norms

AI systems are built, deployed and operated

Human-AI systems make decisions that impact people

Decisions are explained and justified

Decisions are contested

Human-AI system accountability and legitimacy increase

Human-AI system

Interactive controls

Tools for scrutiny

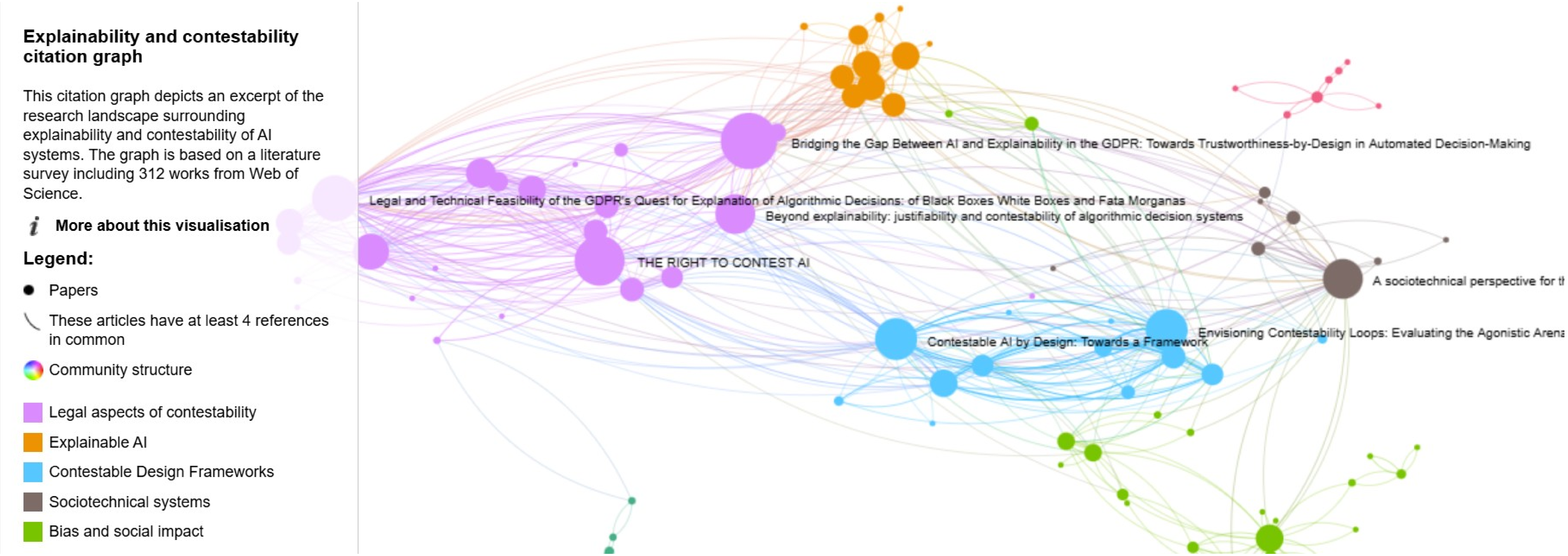
Intervention requests

Policy & system development

Monitoring

www.contestable.ai

Contestability is heterogeneous



Explainability and contestability citation graph

This citation graph depicts an excerpt of the research landscape surrounding explainability and contestability of AI systems. The graph is based on a literature survey including 312 works from Web of Science.

i **More about this visualisation**

Legend:

- ● Papers
- ╲ These articles have at least 4 references in common
- ○ Community structure

- 🟪 Legal aspects of contestability
- 🟧 Explainable AI
- 🟦 Contestable Design Frameworks
- 🟫 Sociotechnical systems
- 🟩 Bias and social impact

Bridging the Gap Between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making
Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes White Boxes and Fata Morganas
Beyond explainability: justifiability and contestability of algorithmic decision systems
THE RIGHT TO CONTEST AI
A sociotechnical perspective for tl
Contestable AI by Design: Towards a Framework
Envisioning Contestability Loops: Evaluating the Agonistic Arena

Things to remember