EU DATA PROTECTION LAW: AN ALLY FOR SCIENTIFIC INTEGRITY?

Mireille Hildebrandt
Professor of Interfacing Law and Technology
Free University Brussels, Radboud University


Tweets
Media
Vind-ik-leuks
Mireille Hildebrandt @mireillemoret 33s
Data Pseudo Science thrives on data obesitas and pattern obesitas. This is what you get:
James Breakwell @XplodingUnicorn
I registered for a running club
My bank immediately sent me a fraud alert
Apparently the only reason I'd exercise is if my card was stolen

Mireille Hildebrandt @mireillemoret 10s

Yes, companies face data obesitas and pattern obesitas, GDPR compliance forces a lean, agile approach to data-driven applications

> **Laura Kayali** @LauKaya
>
> .@VeraJourova : I am convinced the GDPR rules will offer a competitive advantage for companies #data2017

Part I:
The underlying logic of the GDPR

- Data protection law is not equivalent with privacy law
- Risk approach (assessements must be made)
- Proportionality test (necessity requirement)
- Purpose limitation (purpose also determines who is liable)

Part II:
GDPR and Methodological Integrity of ML

- On **methodological integrity**
  - p-hacking, data dredging, or cherry picking performance metrics
  - the **reproducibility crisis** in ML destroying the reliability of ML applications
- How do the **purpose limitation principle** and **the prohibition of automated decisions** relate to ML research design?

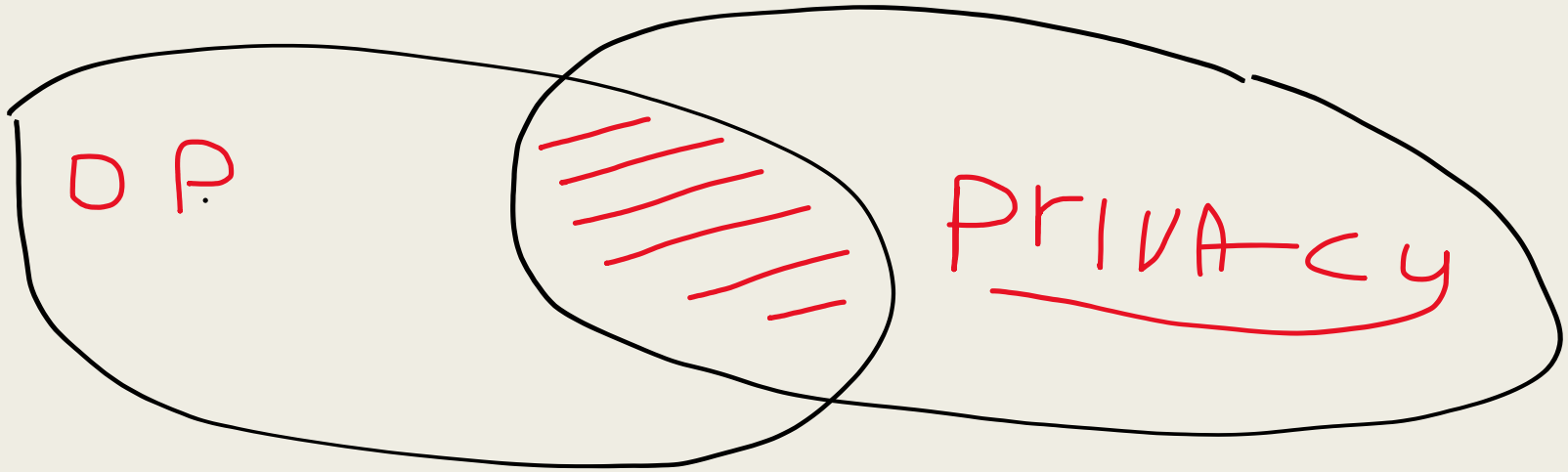
Part I:
The underlying logic of the GDPR

Data protection law is not equivalent with privacy law

- In Europe (EU) we have two fundamental rights:
  - Art. 7 Charter: right to privacy
  - Art. 8 Charter: right to data protection

Part I:
The underlying logic of the GDPR

Data protection law is not equivalent with privacy law



Part I: The underlying logic of the GDPR

Charter of Fundamental Rights of the European Union

Article 7 Respect for private and family life

Everyone has the right to respect for his or her private and family life, home and communications.

Part I: The underlying logic of the GDPR

Charter of Fundamental Rights of the European Union

Article 8 Protection of personal data

1. Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority.

Part I: The underlying logic of the GDPR

Charter of Fundamental Rights of the European Union

Article 8 Protection of personal data

2. Such data must be processed
   - fairly
   - for specified purposes and
   - on the basis of the consent of the person concerned
   - or some other legitimate basis laid down by law.

   Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.

Part I:
The underlying logic of the GDPR

1. GDPR and inferences captured from multimedia data

- Focus of this talk is not on 'mere' capturing of images or voice (other than as a precondition for inferencing)

- Potential inferences:
  - Identification across contexts (plus misidentification)
  - Sentiment analysis (mostly pseudoscience, based on controversial psychology)
  - Categorisation in terms of ethnicity, health risks, employability etc.
  - Enabling micro targeting (with potentially significant consequences)

Part I:
The underlying logic of the GDPR

2. MM data made public by the person it relates to

- Making data public **does not imply others can process it** (different in US)
- Others will need a legal ground, compliance with principles (e.g. specified, legitimate, explicit purpose), transparency requirements, SARs …
- Prohibition of processing 'biometric data for the purpose of uniquely identifying a natural person' exception required under art. 9
- 9.2(e) [exception for prohibition of processing of biometric data if] processing relates to personal data which are manifestly made public by the data subject;

Part I:
The underlying logic of the GDPR

- Note that GDPR takes a **risk approach**, meaning that controllers must
  - err on the side of caution
  - conduct a risk assessment (iterant)
  - mitigate risks by
    1. engaging alternative less risky means to achieve the goal
    2. employing data protection by default (data minimization, e.g. pseudonymisation)
    3. incorporating data protection by design (enabling SARs, erasure withdrawal of consent, triggering human intervention in case of automated decisions)

Part I:
The underlying logic of the GDPR

3. Processing of personal data and consent

- Consent is just one of 6 possible legal grounds!
- Requirement for valid consent are huge:
  - Consent only valid for explicit, legitimate, specified purpose
  - Informed and non-ambiguous
  - Possibility to withdraw must be as easy as provision
  - Twisting of hand not allowed
- If used as biometric for identification default prohibition

Part I:
The underlying logic of the GDPR

3. Processing of personal data and consent

– Better opt for other legal ground:

b. processing is necessary for the performance of a **contract** to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

c. processing is necessary for compliance with a **legal obligation** to which the controller is subject;

d. processing is necessary in order to protect the **vital interests** of the data subject or of another natural person;

e. processing is necessary for the performance of a **task carried out in the public interest** or in the exercise of **official authority** vested in the controller;

f. processing is necessary for the purposes of the **legitimate interests pursued by the controller or by a third party**, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks.

Part I:
The underlying logic of the GDPR

- GDPR and human rights law often requires a proportionatility test
- This is based on the 'necessity requirement'

6.1(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

5.1(c) processing must be 'adequate, relevant and limited to what is necessary in relation to the purposes'

Art. 8.2 ECHR (privacy):
infringement must be 'necessary in a democratic society'

Part I:
The underlying logic of the GDPR

4. MM inferences as mere statistics?

- To the extent that inferences are statistics that do not enable identification
  - the GDPR does not apply

- But, if those inferences are then used to target one or more individuals:
  - It becomes personal data once again (as it is related to a natural person), and
  - The default prohibition of automated decisions may apply, if applicable

Part I:
The underlying logic of the GDPR

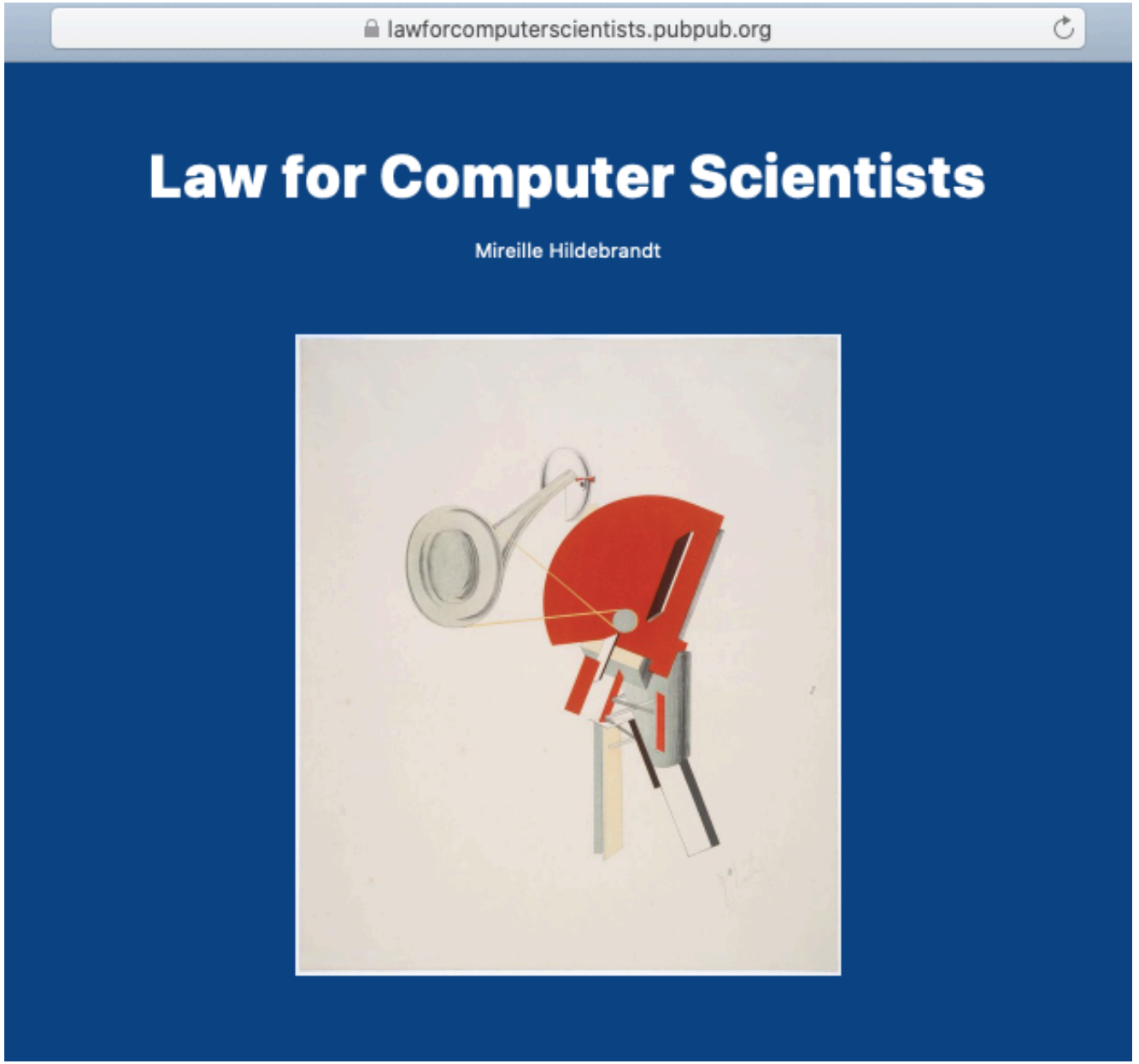
5. The prohibition of processing 'sensitive data' (ethnicity, health)

- 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.

Part I:
The underlying logic of the GDPR

6. Purpose limitation principle as the guiding rationale of EU data protection law, protecting individuals against incorrect, unfair or unwarranted targeting.

- Processing of personal data is not allowed without explicitly specifying a legitimate purpose, and
- this purpose must be communicated to the person whose data is processed (whatever the legal ground)
    - Whichever entity de facto determines the purpose is liability

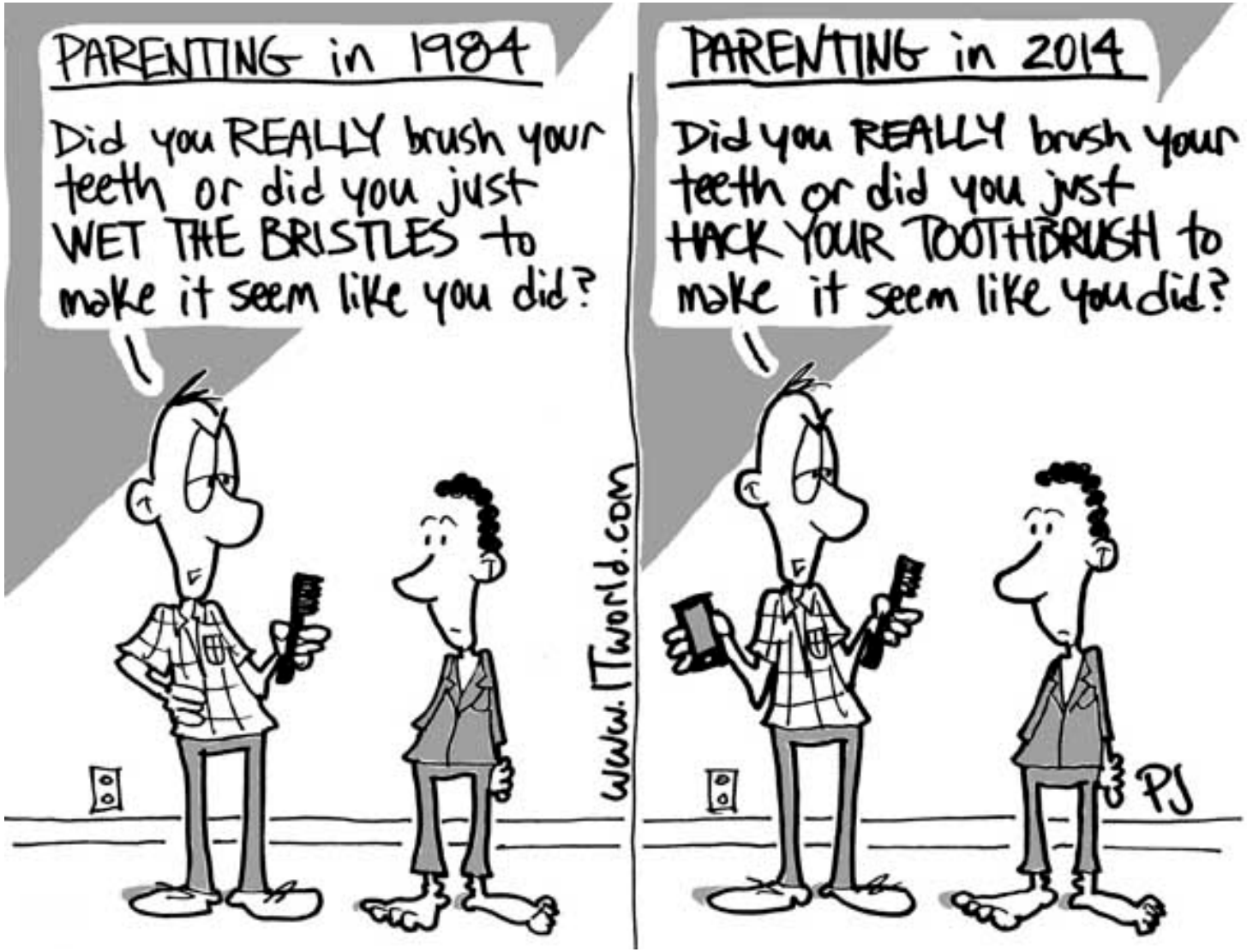

lawforcomputerscientists.pubpub.org
Law for Computer Scientists
Mireille Hildebrandt

- The book will be published in hardcopy (for sale), and as an ebook (open access) by Oxford University Press, March 2020
- It is already available at MIT's pubpub in open access: https://lawforcomputerscientists.pubpub.org

Part II: GDPR and Methodological Integrity of ML

- What is **methodological integrity?**
  - p-hacking, data dredging, or cherry picking performance metrics
  - the **reproducibility crisis** in ML is destroying the reliability of ML applications
- How do the **purpose limitation principle** and **the prohibition of automated decisions** relate to ML research design?


NOISE TO SIGNAL
RobCottingham.com
@robcottingham
I WILL WALK AROUND ALL DAY WEARING YOUR FITNESS TRACKER
$50


PARENTING in 1984
Did you REALLY brush your teeth or did you just WET THE BRISTLES to make it seem like you did?
PARENTING in 2014
Did you REALLY brush your teeth or did you just HACK YOUR TOOTHBRUSH to make it seem like you did?
www.ITworld.com
PJ

The crisis of methodological integrity

Empirical research, mathematics and statistics

- Godel, Church and Wolpert: incompleteness, undecidability and NFL
- Gigerenzer: replication delusion
- Cohen: misapplication of deductive sylogistic reasoning
- Bouthillier: reproducibility of result or finding
- Pearl: causality and correlation
- Geckoboard: data fallacies
- Hofman, Sharma, Watts: exploratory and confirmatory research

The crisis of methodological integrity

Empirical research, mathematics and statistics

- Research in data is NOT empirical research
- Data is NOT what it refers to (is a trace of, or represents)
- Note that **quantification implies prior qualification**

- The idea that human behaviour follows math is ***metaphysics*** (neo-platonism)
- Behaviourism is built upon a skewed, unhelpful reductive metaphysics
- Human action builds on mutual double anticipation

The crisis of methodological integrity

Human action builds on mutual double anticipation

The crisis of methodological integrity

Human action builds on mutual double anticipation

- Parsons and Luhmann:
  - Double contingency
- Plessner:
  - Ex-centric positionality of human animals

The crisis of methodological integrity

Human action builds on mutual double anticipation

- Austin, Searle, MacCormick:
  - Speech act theory:
  - I declare you man and wife:
    - not a description (propositional logic)
    - not a cause (in the physicalist sense)
    - but the ‘performative effect’ of a specific type of language usage

The crisis of methodological integrity

Human action builds on mutual double anticipation

- This is what informs (rather than causes):
  - The Lucas Critique
  - The Goodhart Effect
  - The Campbell Effect

The crisis of methodological integrity

Methodological integrity of robust claims demands confirmatory research design that precedes the data:

- "Many 'applications' are based on exploratory design, of which nobody know whether it actually does work, and for how long and at what costs to individuals or societal infrastructure. Think RTB" @mireillemoret
- "Machine Learning is Computationally Intensive Statistics normally done with poorly selected data, no hypothesis and no confidence intervals mostly by folks with no statistics training; what could possibly go wrong?" Derek🕷McAuley @drdrmc

If A is a hypothesis and B a data set

- **Confirmatory ML research** aims to detect the probability of
  If B than A
  (given this data what is the probability of the hypothesis being true)
- However, it usually ends up detecting:
  If A than probably not B
  (given the truth of the hypothesis, what is the probability of B)

- **Exploratory ML research** seeks to generate potential hypotheses:
  If B what A, B, C etc. can be abducted?

- Confirmatory research

If I have breast cancer the probability of this result of a test is 95%

With this test result the probability of me having cancer is XX

[spoiler: not 95%, depends on the distribution of the data]

- Exploratory research

What test results correlate with breast cancer?

Note we are not even speaking of causes or theory here,
doubt this is even science, but at least no false claims are made

Deep neural networks are easily fooled: High confidence predictions for unrecognizable images

PDF: DNNsEasilyFooled_cvpr15.pdf

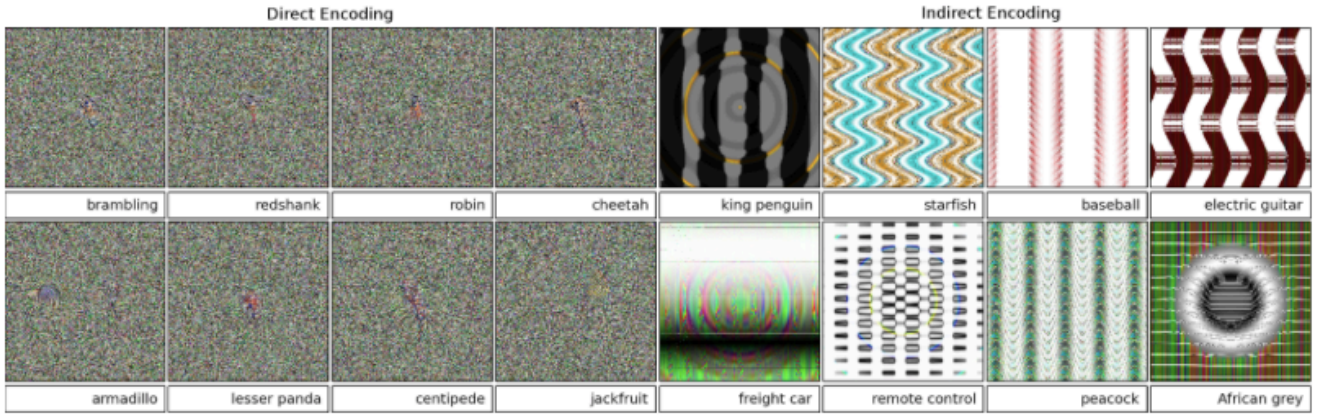


Figure 1: Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with >= 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects.
Left: Directly encoded images. Right: Indirectly encoded images.

Author(s): Nguyen A
Yosinski J
Clune J

Year: 2015

SHARE

POLICY FORUM | MACHINE LEARNING

Adversarial attacks on medical machine learning

Samuel G. Finlayson[1], John D. Bowers[2], Joichi Ito[3], Jonathan L. Zittrain[2], Andrew L. Beam[4], Isaac S. Kohane[1]

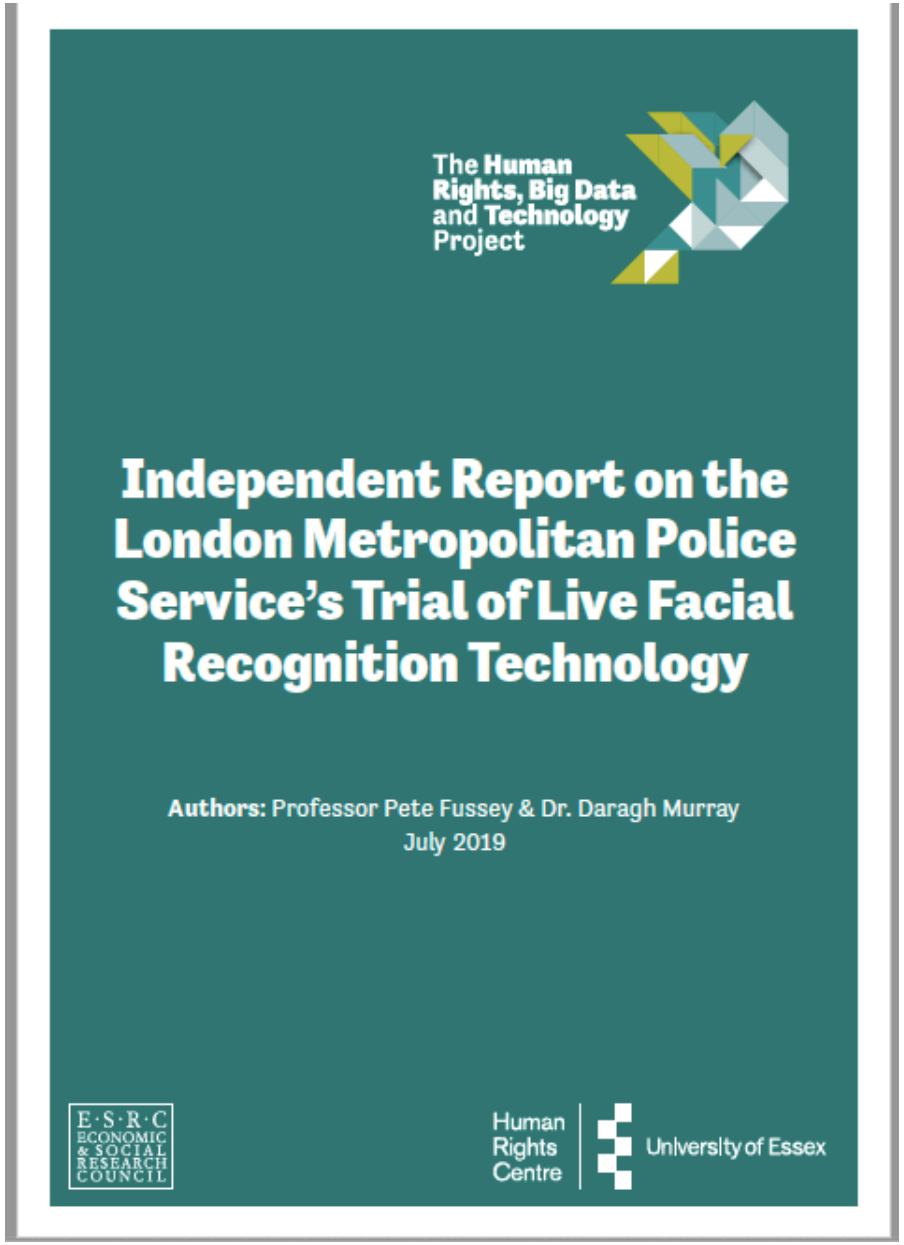
+ See all authors and affiliations

Science 22 Mar 2019:
Vol. 363, Issue 6433, pp. 1287-1289
DOI: 10.1126/science.aaw4399

Article | Figures & Data | Info & Metrics | eLetters | PDF

Summary

With public and academic attention increasingly focused on the new role of machine learning in the health information economy, an unusual and no-longer-esoteric category of vulnerabilities in machine-learning systems could prove important. These vulnerabilities allow a small, carefully designed change in how inputs are presented to a system to completely alter its output, causing it


The Human Rights, Big Data and Technology Project
Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology
Authors: Professor Pete Fussey & Dr. Daragh Murray
July 2019
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
Human Rights Centre
University of Essex

Alerts

Overall, the LFR system generated 46 matches over the course of observed test deployments, involving 45 separate individuals. 42 matches were deemed eligible for analysis.

Adjudicating officers judged 16 (38.1%) of these 42 computer generated matches to be 'non-credible'; that is, officers did not believe the image recorded by the LFR technology match the image on the watchlist. MPS officers considered the LFR match sufficiently credible to stop individuals and perform an identity check on 26 occasions. Four of these attempted interventions were unsuccessful, as individuals were lost in the crowd.

Of the remaining 22 stops, 14 (63.64%) were verified as incorrect matches following an identity check. Eight (36.36%) were verified as correct matches following an identity check. This means that across all six observed trials, and from all computer-generated alerts, face recognition matches were verifiably correct on eight occasions (eight of 42 matches, 19.05%).

Who wants accurate models? Arguing for a different metrics to take classification models seriously

Federico CABITZA[a,1] Andrea CAMPAGNER[a,b]
[a] *University of Milano-Bicocca, Milano, Italy*
[b] *IRCCS Istituto Ortopedico Galeazzi, Milano, Italy*

Abstract. With the increasing availability of AI-based decision support, there is an increasing need for their certification by both AI manufacturers and notified bodies, as well as the pragmatic (real-world) validation of these systems. Therefore, there is the need for meaningful and informative ways to assess the performance of AI systems in clinical practice. Common metrics (like accuracy scores and areas under the ROC curve) have known problems and they do not take into account important information about the preferences of clinicians and the needs of their specialist practice, like the likelihood and impact of errors and the complexity of cases. In this paper, we present a new accuracy measure, the *H-accuracy* (Ha), which we claim is more informative in the medical domain (and others of similar needs) for the elements it encompasses. We also provide proof that the H-accuracy is a generalization of the balanced accuracy and establish a relation between the *H-accuracy* and the *Net Benefit*. Finally, we illustrate an experimentation in two user studies to show the descriptive power of the Ha score and how complementary and differently informative measures can be derived from its formulation (a Python script to compute Ha is also made available).

Keywords. predictive models, accuracy, Machine Learning, Medical Artificial Intelligence, Validation

Mandrola and Morgan in [29], to warn of the risks of incidental findings and the related overuse that a highly sensitive decision support could exacerbate, have observed how "the problem with decision support is that it must be designed *to add value* and be easily accessible without increasing burden for clinicians [...that is it] needs to *better provide relevant information* at the point of care to make decision-making easier for clinicians." (our emphasis). Some author has then recently proposed this one best measure is the *Matthew correlation coefficient* [13], which is not affected by class imbalance and is generalizable to multiclass settings. However, this metrics is not intuitively related to error rate and, mostly important, does not consider the characteristics of the available data, nor the preferences of the intended model users.

For this reason, in this paper we proposed a novel metrics that takes into account the above elements, to provide an indicator of the reliability and value of the potential advice by a decision support. In particular, by providing an analytical formulation of this metrics, we also pointed out meaningful areas of the resulting function to focus on specific aspects of the model performance, like reliability (cf. τ), practicality (cf. d) and priority (cf. p), and suggested some empirical values to report that we believe could inform the users adopting an ML model exhibiting such skills, namely *confident, prioritized*, and *practical accuracy*. To our knowledge, H-accuracy is the first metrics to go beyond what can be known of a model's performance from the confusion matrix, while still being related to the intuitive notion of "getting classification right"[9].

The GDPR as an ally

- **Core to methodological integrity of applied ML:**
  - Develop a metric that takes into account:
    - Reliability
    - Practicality
    - Priority
- This is in turn core to the **proportionality test** that is core to the GDPR and to human rights law:
  - If an application is accurate in this sense it cannot be effective and thus
  - not necessary

The GDPR as an ally

- **Core to methodological integrity of applied ML:**
  - Develop a metric that takes into account:
    - Reliability
    - Practicality
    - Priority
- This relates to to the risk approach:
  - Without a proper empirical validation
  - No assessment can be made about
    - Relationship between practical effectiveness
    - And infringement of rights and freedoms

The GDPR as an ally

- Core to methodological integrity of ML research design:
  - Differentiate between **exploratory and confirmatory** ML RD
  - Never employ findings of exploratory for real life implementation
  - Do not assume that ML based on behavioural data 'works' as claimed

The GDPR as an ally

- **Hofman, Sharma, Watts on experimental and confirmatory research design:**

Exploratory ML researchers are **free to**

- study different tasks,
- fit multiple models,
- try various exclusion rules, and
- test on multiple performance metrics.

When reporting their findings, however, they should:

- transparently declare their full sequence of design choices to avoid creating a false impression of having confirmed a hypothesis rather than simply having generated one,
- report performance in terms of multiple metrics to avoid creating a false appearance of accuracy.

The GDPR as an ally

- **Hofman, Sharma, Watts on experimental and confirmatory research design:**

Confirmatory ML: researchers should be

- **required to preregister their research designs,**
- including data preprocessing choices,
- model specifications,
- evaluation metrics,
- and out-of-sample predictions,
- **in a public forum such as the Open Science Framework (https://osf.io).**

The GDPR as an ally

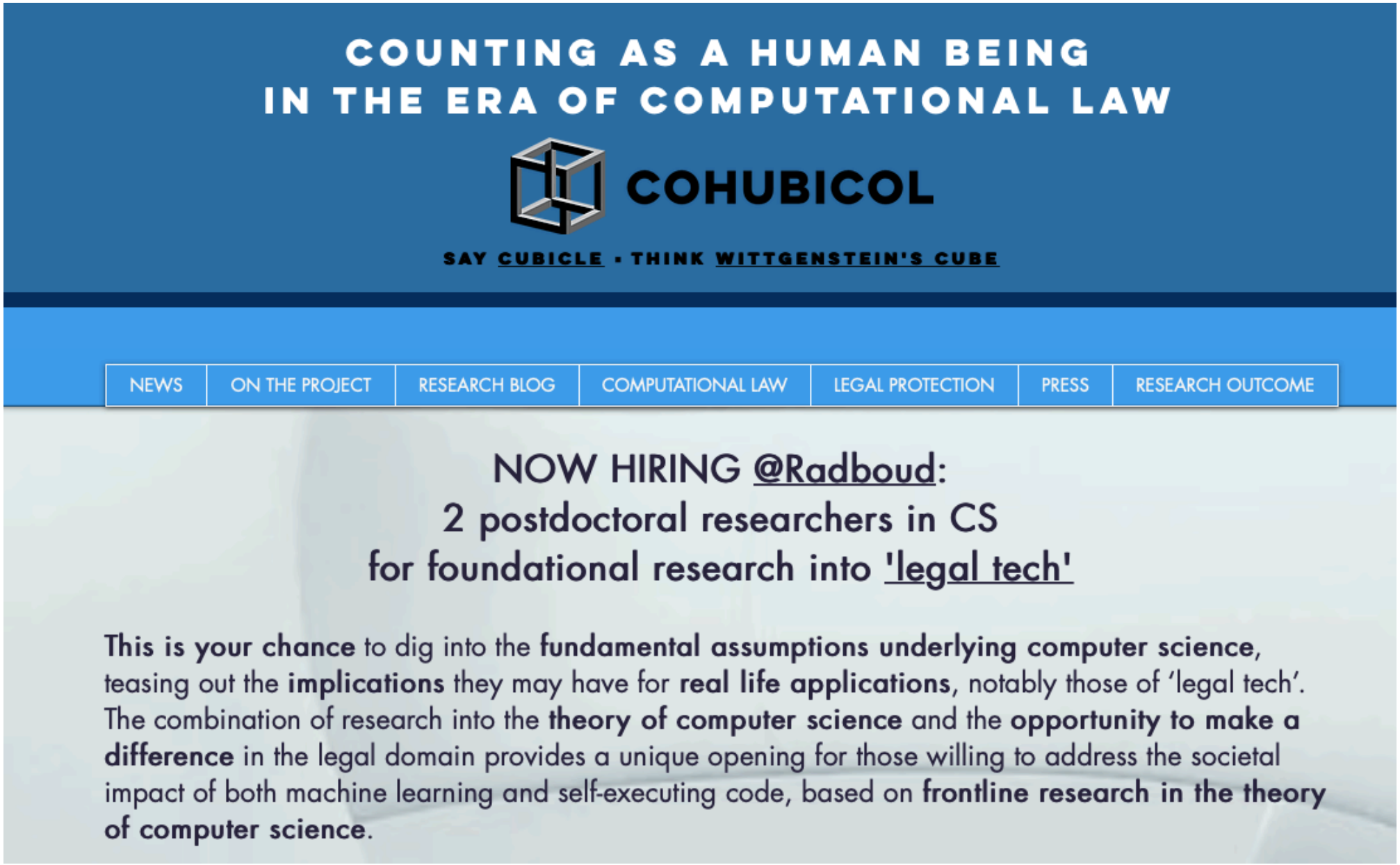
- **Core to methodological integrity of ML research design:**
  - Differentiate between exploratory and confirmatory ML RD
    - Purpose limitation forces to make that choice and face the consequences
    - Purpose limitation aligns with the machine readable task that must be formulated
    - Purpose limitation aligns with the evaluation metrics

The GDPR as an ally

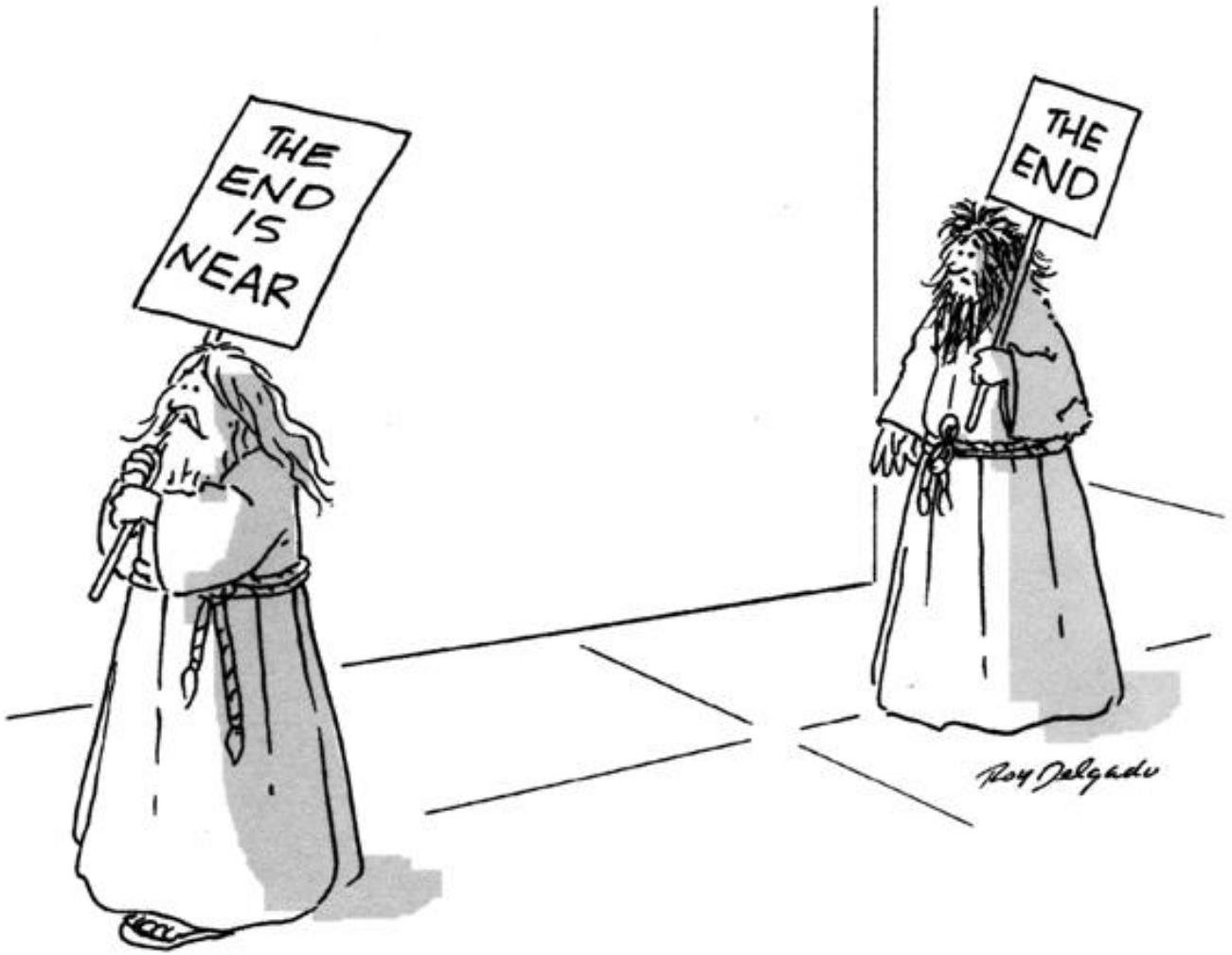
- **Core to methodological integrity of ML research design:**
  - Prohibition of automated decisions ex 22 will help save us from unreliable ML applications
  - The requirements of human intervention (22.3) and explanation and contestation (13-14-15) will enable to challenge lack of scientific integrity

The GDPR as an ally

- **Core to methodological integrity of ML research design:**
  - Do not assume that ML based on behavioural data 'works' as claimed
  - Storage limitation, data minimisation, purpose limitation, transparency accuracy, accountability all help to safeguard the reliability of input data
  - Data obesity generates pattern obesity generates bad ML output


COUNTING AS A HUMAN BEING
IN THE ERA OF COMPUTATIONAL LAW
COHUBICOL
SAY CUBICLE · THINK WITTGENSTEIN'S CUBE
NEWS
ON THE PROJECT
RESEARCH BLOG
COMPUTATIONAL LAW
LEGAL PROTECTION
PRESS
RESEARCH OUTCOME
NOW HIRING @Radboud:
2 postdoctoral researchers in CS
for foundational research into 'legal tech'
This is your chance to dig into the fundamental assumptions underlying computer science, teasing out the implications they may have for real life applications, notably those of 'legal tech'. The combination of research into the theory of computer science and the opportunity to make a difference in the legal domain provides a unique opening for those willing to address the societal impact of both machine learning and self-executing code, based on frontline research in the theory of computer science.

- Vacancies can be found at
  - https://www.cohubicol.com (extensive description)
  - https://www.ru.nl/werken-bij/vacature/details-vacature/?recid=1068776&doel=embed&taal=nl (where to apply)


THE END IS NEAR
THE END
Roy Delgado