



# **SAVING ML FROM CRAPPY APPLICATIONS**

Mireille Hildebrandt  
Research Professor Vrije Universiteit Brussel  
Radboud University, Nijmegen

# ML Applications:

- RTB, AB testing
- Prediction of energy usage behaviours
- Medical decisions to hospitalize, to conduct surgery
- Prediction of judgements
- Access to education (e-learning)
- Access to employment (recruiting, gig-economy)
- Predictive policing

# What's next?

- Bias is the heart of ML
- GDPR, reseach and purpose limitation
- The First Law of Informatics
- The methodological integrity of ML research design
- Preregistration of MLRD (the only business case for blockchain)

# Bias is the heart of ML

## Bias as a

'fundamental property of inductive inference: a learner that makes no **a priori assumptions** regarding the identity of the target concept has no rational basis for classifying any unseen instances.'

'Thus, we define the inductive bias of a learner as the set of **additional assumptions B** sufficient to justify its inductive inferences as deductive inferences.'

Tom Mitchell (1997), 47, 43

Mitchell, Thomas (1997) *Machine Learning*, New York: McGraw-Hill Education. <http://www.cs.cmu.edu/~tom/mlbook.html>

# Bias is the heart of ML

**Homo economicus** remained the unquestioned ideal, while **Homer Simpson** was teasingly proposed as a more apt description of Homo sapiens, given his bouts of ignorance, laziness, and brilliant incompetence (Thaler and Sunstein, 2008).

The field would have progressed in an entirely different direction had it followed **Herbert Simon's** original vision of behavioral economics, in which psychology comprised more than mental quirks and theory more than expected utility maximization.

Few economists appear to be aware that the bias message is **not representative** of psychology or cognitive science in general.

Gerd Gigerenzer (2018), 304-5

Gigerenzer, Gerd (2018) 'The Bias Bias in Behavioral Economics'. *Review of Behavioral Economics* 5 (3-4): 303-36.

<https://doi.org/10.1561/105.00000092>

# Bias is the heart of ML

The problem is not bias, but **what bias & how it was constructed:**

- What data, what cleansing, what assumptions?
- What feature space, what hypothesis space, what task?
- What performance metrics? What out of sample testing?
- If not interpretable, how do we know what the system gets wrong?

# Charter of fundamental rights of the EU

## **Article 7 Respect for private and family life**

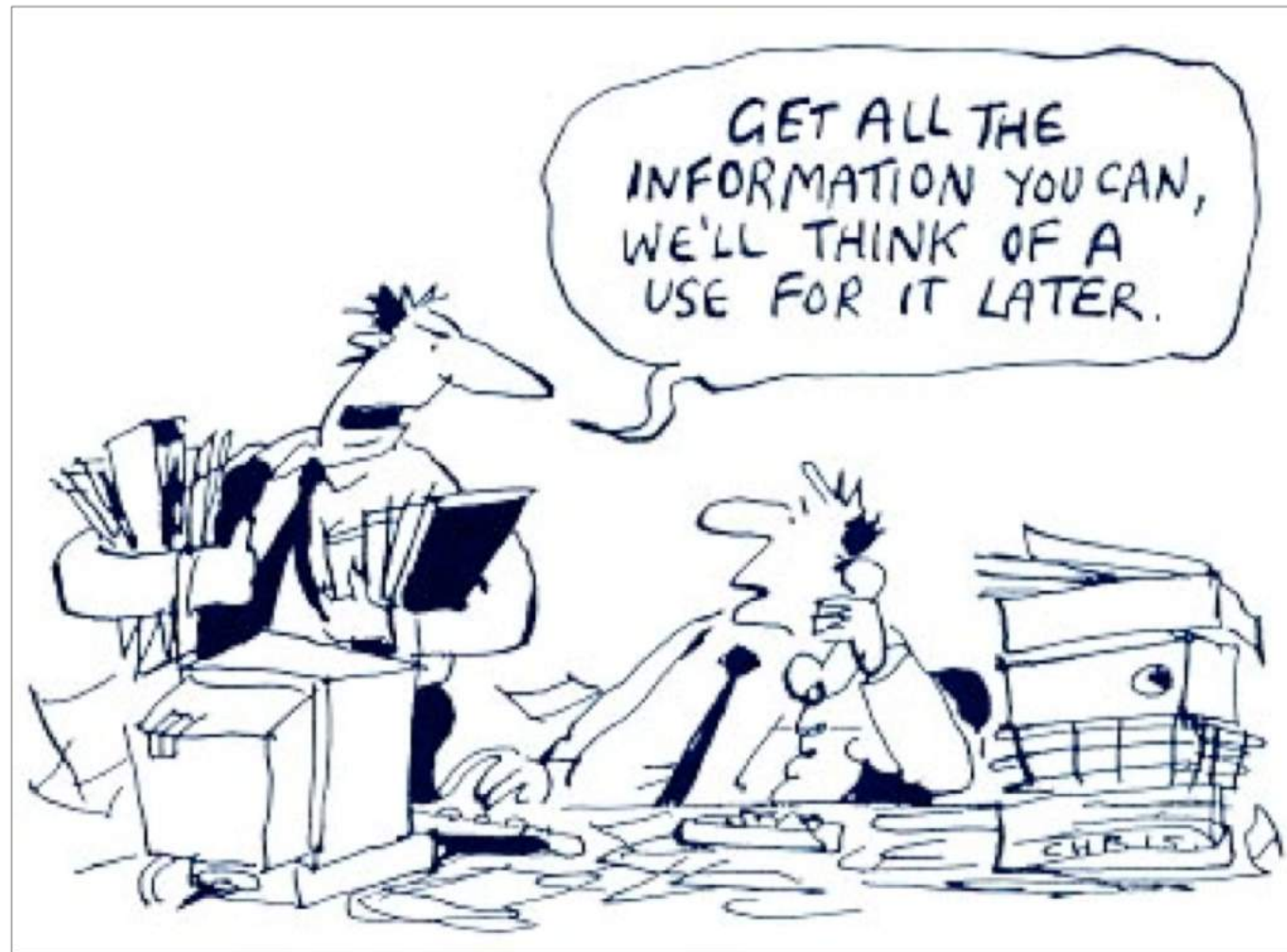
Everyone has the right to respect for his or her private and family life, home and communications

# Charter of fundamental rights of the EU

## 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.





# Charter of fundamental rights of the EU

## Article 16 Freedom to conduct a business

The freedom to conduct a business **in accordance with Community law and national laws** and practices is recognised.

# GDPR objectives

## Article 1 Subject-matter and objectives

1. This Regulation lays down rules relating to the **protection of natural persons** with regard to the processing of personal data and rules relating to the **free movement of personal data**.
2. This Regulation protects **fundamental rights and freedoms of natural persons** and in particular their right to the protection of personal data.
3. The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data.

# GDPR principles

## Article 5 Principles relating to processing of personal data

1. Personal data shall be:

- (a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('**lawfulness, fairness and transparency**');
- (b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('**purpose limitation**');
- (c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('**data minimisation**');

# GDPR purpose limitation

## Article 5 Principles relating to processing of personal data

1. Personal data shall be:

(b) collected for **specified, explicit and legitimate purposes** and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes (**'purpose limitation'**);

# GDPR purpose limitation

## Article 5 Principles relating to processing of personal data

1. Personal data shall be:

(b) collected for specified, explicit and legitimate purposes and **not further processed in a manner that is incompatible with those purposes**; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes (**'purpose limitation'**);

# GDPR purpose limitation

## Article 5 Principles relating to processing of personal data

1. Personal data shall be:

(b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; **further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');**

# GDPR research

## Recital (33) [MH: accepting leeway, but requiring granularity]

It is often **not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection**. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose.



# GDPR research

## Article 89 Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes

1. Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. **Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner**

# GDPR research

## Article 89 Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes

2. Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law **may provide for derogations** from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.
3. Where personal data are processed for archiving purposes in the public interest, Union or Member State law **may provide for derogations** from the rights referred to in Articles 15, 16, 18, 19, 20 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.

# GDPR research

## **Article 89 Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes**

4. Where processing referred to in paragraphs 2 and 3 serves at the same time **another purpose, the derogations shall apply only to processing for the purposes referred to in those paragraphs.**

# GDPR principles

## Article 5 Principles relating to processing of personal data

1. Personal data shall be:

- (a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
- (b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');
- (c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');

# GDPR principles

## Article 5 Principles relating to processing of personal data

1. Personal data shall be:

- d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the **purposes** for which they are processed, are erased or rectified without delay ('accuracy');
- e) kept in a form which permits identification of data subjects for no longer than is necessary for the **purposes** for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
- f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or **unlawful processing** and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

# GDPR liability

## Article 5 Principles relating to processing of personal data

2. The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability').

# GDPR legal basis

## Article 6 Lawfulness of processing

1. Processing shall be lawful only if and to the extent that at least one of the following applies:

- a. the data subject has given **consent** to the processing of his or her personal data for one or more specific purposes;
- 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;

# GDPR legal basis

## Article 6 Lawfulness of processing

1. Processing shall be lawful only if and to the extent that at least one of the following applies:

- 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.



# GDPR consent

## Article 4 Definitions

11. 'consent' of the data subject means any **freely given, specific, informed and unambiguous indication** of the data subject's wishes by which he or she, by a **statement or by a clear affirmative action**, signifies agreement to the processing of personal data relating to him or her;

# GDPR consent

## Article 7 Conditions for consent

1. Where processing is based on consent, the controller shall be able to **demonstrate** that the data subject has consented to processing of his or her personal data.
2. If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is **clearly distinguishable from the other matters**, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.

# GDPR consent

## Article 7 Conditions for consent

3. The data subject shall have **the right to withdraw his or her consent at any time**. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. **It shall be as easy to withdraw as to give consent**.
4. When assessing whether consent is **freely given**, utmost account shall be taken of whether, *inter alia*, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

# GDPR consent

## Recital (43)

(...)

Consent is presumed not to be freely given if (...):

- the performance of a contract, including the provision of a service, is dependent on the consent despite such consent not being necessary for such performance.

The book OUP 2020:



© Lippincott Williams & Wilkins 1990-1991  
The Annals of the New York Academy of Sciences 1992  
<http://www.interscience.wiley.com/jpages/0021-8758>

# **Law for Computer Scientists**

**And other Folks**

**Mireille Hildebrandt**

# First Law of Informatics

Lei, J. van der (1991)

'Use and Abuse of Computer-Stored Medical Records'

*Methods of Information in Medicine* 30 (2): 79-80

# First Law of Informatics

Van der Lei (medical informatics):

**1. data shall be used only for the purpose for which they were collected**

*collateral:*

**2. if no purpose was defined, they should not be used**

# First Law of Informatics

Van der Lei (1991):

Medical records are used as a data source for purposes ranging

- from billing the patient to
- performing epidemiological studies, and
- from performing quality control
- to defending oneself against legal claims.



# First Law of Informatics

Van der Lei (1991):

The quality of the data in medical records has often been lamented:

- the **reliability of clinical data** has been questioned, and
- the tension **between reimbursement schemes and coding schema**

# First Law of Informatics

Van der Lei (1991):

- Burnum, for example, in a paper entitled 'The misinformation era: The fall of the medical record', states: with the advent of the informational era in medicine, we are pouring out a torrent of medical record misinformation.
- He argues that, although paper medical records have long been faulty, computer based medical records will contain even more distorted and misleading information.
- He concludes: all medical record information should be regarded as suspect; much of it is fiction.
- Or, as one of my colleagues recently stated: Medical informatics may prove to be the final nail in the coffin of the medical record.

# First Law of Informatics

Van der Lei (1991):

- Data as such do not exist in nature; data are man-made artefacts.
- Real world entities are observed and described by humans.
- When usage of the medical record changes, the recorded data change accordingly.
- Data are collected with a purpose in mind; that purpose has a direct influence on what data are recorded.



# First Law of Informatics

[Photo taken from a ppt by Van der Lei]

This is what I call:

- **Data obesity**

Resulting in:

- Buffer overflow

- **Pattern obesity**



Mireille Hildebrandt @mireillemoret 10s

Yes, companies face data obesity and pattern obesity, 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

# First Law of Informatics

Caruana, Rich, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad (2015)

'Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission'

In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1721–1730. KDD '15. New York, NY, USA: ACM. <https://doi.org/10.1145/2783258.2788613>

# First Law of Informatics

In machine learning often a **tradeoff** must be made between accuracy and intelligibility [MH: ???]

In the pneumonia risk prediction case study,

- the intelligible model uncovers surprising patterns in the data that previously had prevented complex learned models from being fielded in this domain,
- but because it is intelligible and modular allows these patterns to be recognized and removed.

[MH: accuracy on the patient data, yes, accuracy in real life, no]

# First Law of Informatics

On one of the pneumonia datasets, the rule-based system learned the rule "HasAsthama(x)  $\rightarrow$  LowerRisk(x)", i.e.,

- that patients with pneumonia who have a history of asthma have lower risk of dying from pneumonia than the general population.



# First Law of Informatics

Needless to say, this rule is counterintuitive.

**But it reflected a true pattern in the training data:** patients with a history of asthma who pre-sented with pneumonia usually were admitted not only to the hospital but directly to the ICU (Intensive Care Unit).

The good news is that the aggressive care received by asthmatic pneumonia patients was so effective that it lowered their risk of dying from pneumonia compared to the general population.

# First Law of Informatics

The bad news is that because the prognosis for these patients is better than average,

- models trained on the data incorrectly learn that asthma lowers risk,
- when in fact asthmatics have much higher risk (if not hospitalized).

# First Law of Informatics

## Purposes:

- to predict risk prior to hospitalization so that **a more informed decision** about hospitalization could to be made.
- (ultimate goal) **to reduce healthcare cost by reducing hospital admissions**, while maintaining (or even improving) outcomes by more accurately identifying patients that need hospitalization.

# First Law of Informatics

- The rule-based system was intelligible and modular, making it easy to **recognize and remove dangerous rules** like the asthma rule.
- The decision was made to not use the neural nets not because the asthma problem could not be solved, but because the lack of intelligibility made it **difficult to know what other problems might also need fixing**.

# First Law of Informatics

We now have a number of new learning methods that are very accurate, but unfortunately also relatively unintelligible such as boosted trees, random forests, bagged trees, kernelized- SVMs, neural nets, deep neural nets, and ensembles of these methods.

- applying any of these methods to mission-critical problems such as healthcare remains problematic, in part because usually it is not ethical to modify (or randomize) the care delivered to patients to collect data sets that will not suffer from the kinds of bias described above.
- **learning must be done with the data that is available, not the data one would want.** But it is critical that models trained on real-world data be validated prior to use lest some patients be put at risk, which makes using the most accurate learning methods challenging.

# First Law of Informatics

The main contributions of this paper are that it:

- shows that GA2Ms yield competitive accuracy on real problems;
- demonstrates that the **learned models are intelligible**;
- demonstrates that the predictions made by the model for **individual cases (patients) also are intelligible**, and
- demonstrates how, because the models are modular, they can be **edited by experts**.

# Methodological Integrity

## MLRD

Hofman, Jake M., Amit Sharma, and Duncan J. Watts (2017)

'Prediction and Explanation in Social Systems'

*Science* 355 (6324): 486–88. <https://doi.org/10.1126/science.aal3856>

# Methodological Integrity

## Exploratory MLRD

In **exploratory analyses**, 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.
- Relatedly, they should **report performance in terms of multiple metrics** to avoid creating a false appearance of accuracy.



# Methodological Integrity

## Exploratory MLRD

In cases where data are abundant, moreover, researchers can increase the validity of exploratory research by using a three-way split of their data into

- a **training set** used to fit models,
- a **validation set** used to select any free parameters that control model capacity and to compare different models, and
- a **test set** that is used only once to quote final performance

# Methodological Integrity

## Confirmatory MLRD

Last, having generated a **firm hypothesis** through exploratory research, researchers may then choose to engage in **confirmatory research**, which allows them to make stronger claims.

To qualify research as confirmatory, however, 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>).

Although strict adherence to these guidelines may not always be possible, following them **would dramatically improve the reliability and robustness of results**, as well as facilitating comparisons across studies.

# Methodological Integrity MLRD & p-hacking

Berman, Ron, Leonid Pekelis, Aisling Scott, and Christophe Van den Bulte (2018)  
'P-Hacking and False Discovery in A/B Testing'

SSRN Scholarly Paper ID 3204791. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=3204791>

# Methodological Integrity

## MLRD & p-hacking

Specifically, about 57% of experimenters p-hack when the experiment reaches 90% confidence.

Furthermore, approximately 70% of the effects are truly null, and p-hacking increases the false discovery rate (FDR) from 33% to 42% among experiments p-hacked at 90% confidence.

Assuming that false discoveries cause experimenters to stop exploring for more effective treatments, we estimate the expected cost of a false discovery to be a loss of 1.95% in lift, which corresponds to the 76th percentile of observed lifts.



**Ronny Kohavi** @ronnyk · 16 Dec 2018

Replying to @JohnHolbein1

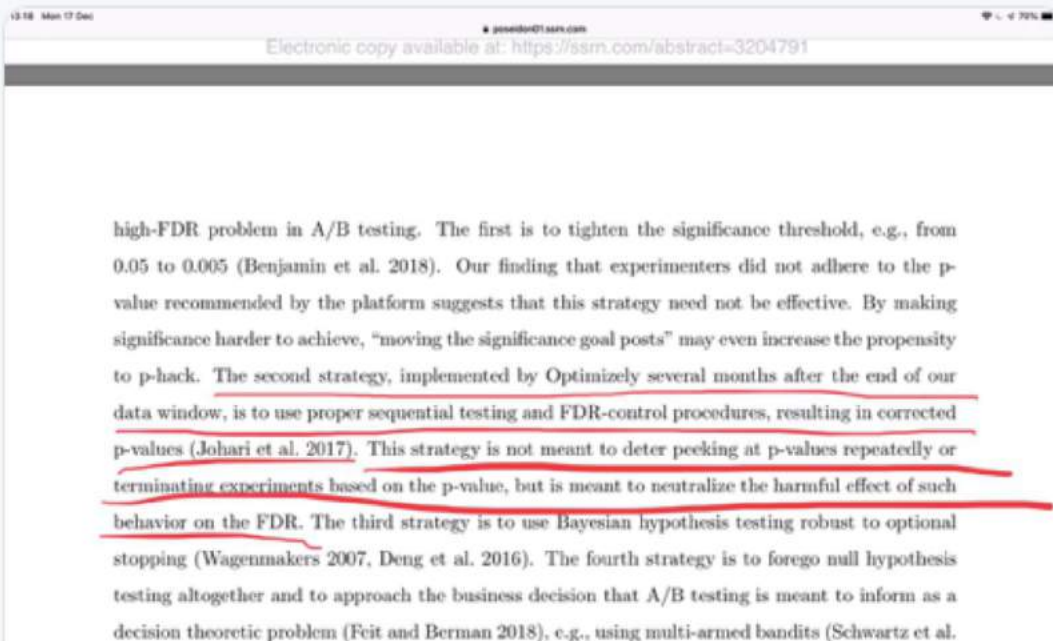
Sensationalized claims. The paper shows that @Optimizely in 2014 encouraged early stopping ("Our data consists of 2,101 experiments run on the Optimizely A/B testing platform in 2014"). I pointed this out in 2014 review: [amazon.com/gp/customer-re...](https://amazon.com/gp/customer-re...)  
Optimizely addressed it  
@koomen

1 6 31



**Henrik Zaunbrecher** @HWFencebreaker · 17 Dec 2018

The paper addresses that Optimizely changed procedures in the conclusion section



■ Check:  
<https://twitter.com/johnholbein1/status/1074091992836009984>



**Saoirse Like Inairse** @HeadAsploding · 16 Dec 2018

Replying to @JohnHolbein1

@EdMillerPoker Should we not do that? I tell you, I gotta plead ignorance on this thing, because if anyone had said anything to me at all ... that that sort of thing is frowned upon... you know, cause I've worked in a lot of offices, and I tell you, people do that all the time.

2 8



# Preregistration of MLRD [the only business case for blockchain]

ML is based on the idea that intelligence concerns the ability to learn from experience, rather than the ability to apply ready-made knowledge. In that sense it favours inductive rather than deductive inferences. In the domain of artificial intelligence many voices now warn against overestimating the effectiveness of inductive learning, without however disqualifying its potential achievements (Brooks 2017, 2018; Marcus 2018).

## The Mechanics of ML

It is interesting to note that human intelligence thrives on what Peirce called abductive inferences (Peirce and Turrisi 1997, 241-56), which are neither inductive nor deductive. Abductive inferencing basically entails an informed guess as to the explanation of a set of observations. Building on Peirce, scientific research can be framed as starting with an abduction based on observation, generating an explanation (theory) from which a hypothesis (prediction) is deduced about subsequent observations, after which the prediction can be inductively tested against new observations. Building on Popper's theory of falsification,<sup>1</sup> hypotheses should be developed in a way that enables the rejection of the explanation – not merely its verification. A theory that explains why all swans are white should not just be verified by detecting ever more white swans, but tested against its potential falsification by searching for black swans.

INING FROM P-HACKING

# Preregistration of MLRD

I propose that whoever puts an ML application on the market should:

- pre-register the research design that was used to develop the application (including subsequent updates, on a blockchain).

This will contribute to the contestability of claims regarding the safety, security, and reliability of such applications,

- while also enabling the contestability of decisions based on such applications in terms of potential violations of fundamental rights such as privacy, data protection, freedom of expression, presumption of innocence and non-discrimination.

If such preregistration were to become a requirement, e.g. in an updated **Machinery Directive**, it would also be a good example of 'legal protection by design'.



# Preregistration of MLRD

A data scientist's goal is one of translation: from data to knowledge to action. After defining a hypothesis but before making a decision, a critical step in the process is transforming data into evidence and assessing the quality of that evidence. Data does not speak for itself. The same observation in disparate contexts will support a disparate set of conclusions. Thus formalized inductive logic, including statistical inference and machine learning, require quantification of both the data and the context. Probabilities and probabilistic reasoning are used almost exclusively to quantify evidence as these frameworks combine observation with context into question-agnostic, cross-disciplinary metrics (1-in 100 chance, 95% confident, 99% accurate etc.).

# Preregistration of MLRD

If not already there, automated intelligence and machine learning will develop a reproducibility crisis of its own.

- Early research wins and models announced with much publicity will not generalize and eventually fail.
- Businesses will perceive their data science teams as underperforming, or not worth the investment.
- Practitioners and strategic leaders would benefit from understanding the limits of inference.
- Models built based on strong theoretical foundations (existing knowledge, context), based on rules that have already shown substantial predictive value, will outperform models developed largely by inference, based on excessive search and selection.

# GDPR: competitive advantage

Practical and effective enforcement of GDPR (art. 78-83):

- Data minimization (in function of necessity)
- Purpose limitation (in function of necessity)
  - Will aid the methodological integrity of ML, and
  - prevent crappy applications from being exploited
  - May transform current business models that favour sloppy MLRD

**This will give EU a competitive advantage, and protect the reputation of ML**

