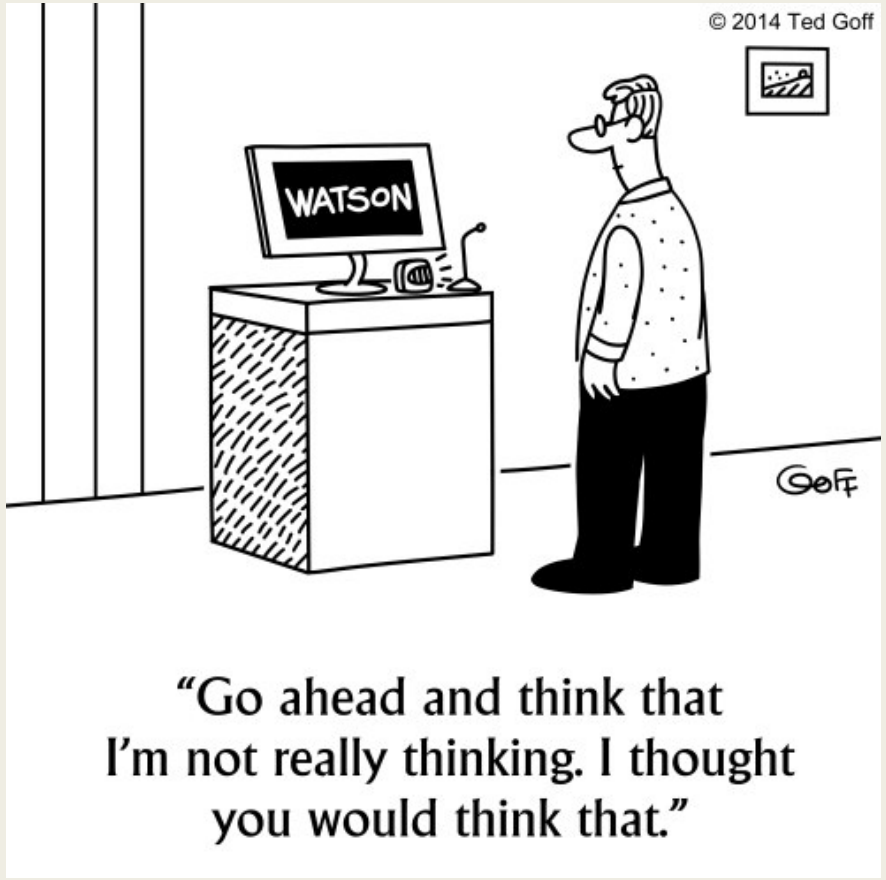
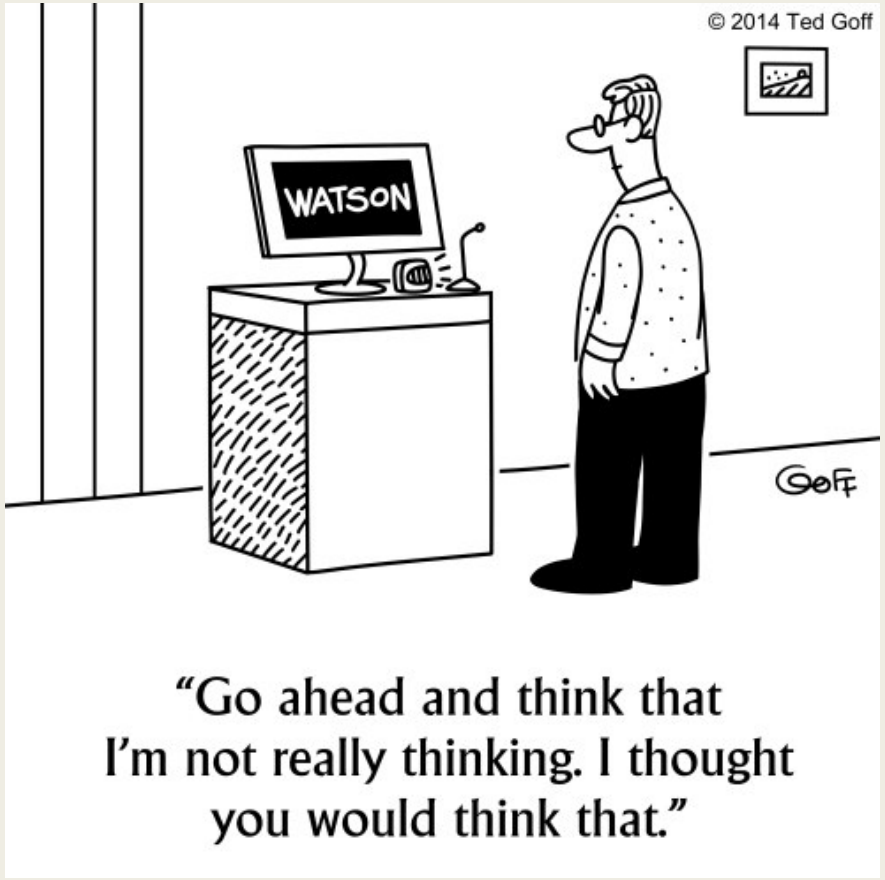
PROFILE TRANSPARENCY AND AUTOMATED DECISIONS UNDER THE GDPR

Mireille Hildebrandt
Science Faculty, Radboud University Nijmegen
Faculty of Law and Criminology, Vrije Universiteit Brussel


© 2014 Ted Goff
WATSON
GOFF
"Go ahead and think that I'm not really thinking. I thought you would think that."

who is speaking?



What's new?

1. Privacy as the protection of the incomputable self
2. Data-driven 'everyware'
3. Resisting computational overdetermination
4. GDPR compliance and new business models
   - *Consent*
   - *Automated decisions and profile transparency*

Privacy as incomputability

- Not everything that can be counted counts, not everything that counts can be counted

- *Anything* is computable - but not *everything*
- We are computable in many ways (though not in any way), depending on who wants to know what

- For the philosophers: think Mead, Plessner, Arendt, Ricoeur, and Herbert Simon
- Think natality and learning, remember *no algorithm can be trained on future data* (Godel, Wolpert)

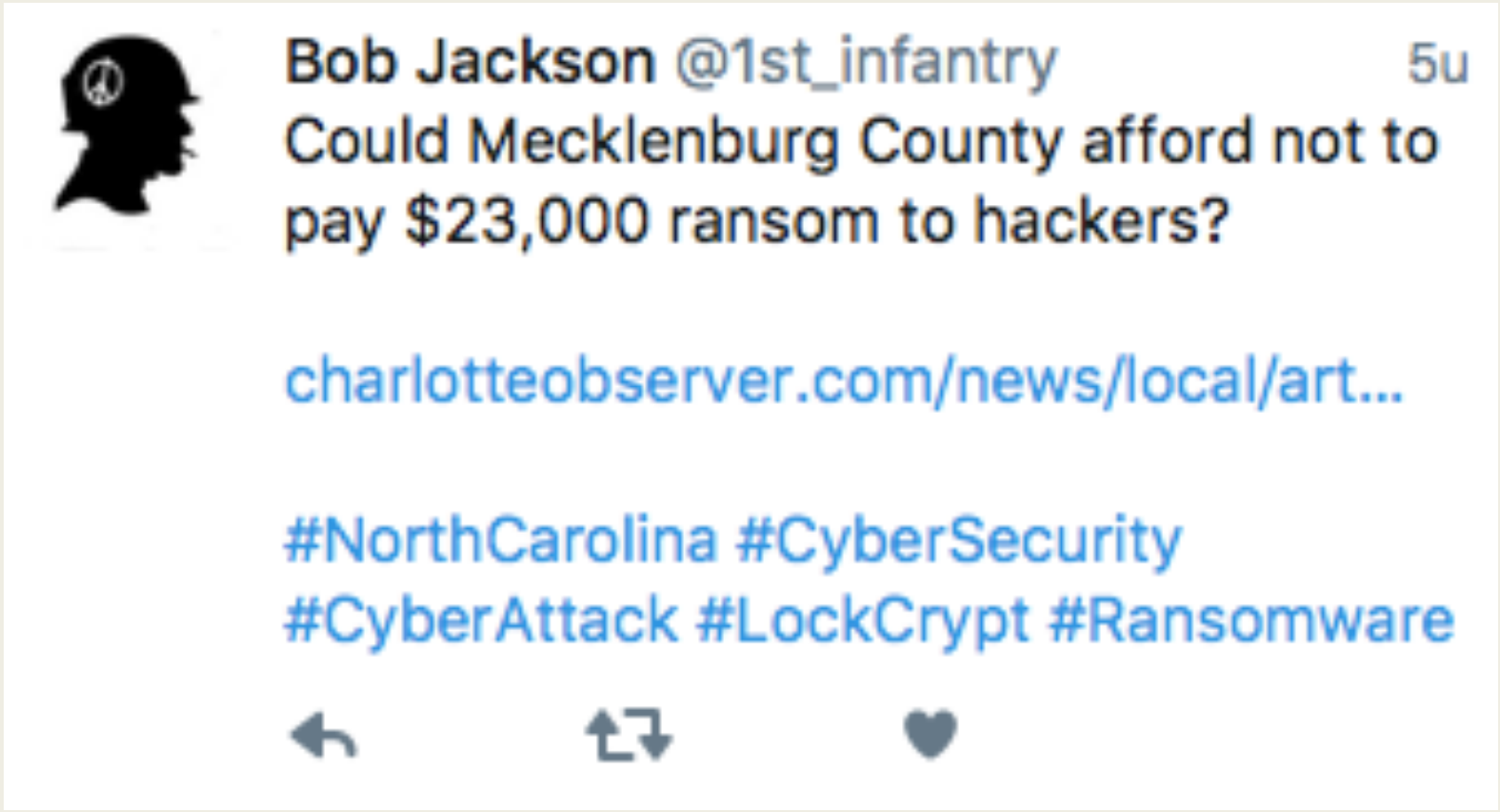
- The idea of the incomputable self asserts that we can be computed in many ways, but never in a final way, *we are fundamentally underdetermined*

Privacy as incomputability

- Data-driven business-models *'capture'* data and reconfigure their environment (us)
- We are *f*udged to produce as many data as possible, to enable *prediction* and to *influence*
- Data-driven business-models result in *specific choice architectures*

- Think e.g. tracking walls, then translate this to IoT applications
  - *connected cars, smart energy grids, smart city infrastructure*

- These choice architectures may *overdetermine* us, based on what choice they (do not) provide
- More generally, applied machine learning may result in *computational overdetermination*


Bob Jackson @1st_infantry 5u
Could Mecklenburg County afford not to pay $23,000 ransom to hackers?
charlotteobserver.com/news/local/art...
#NorthCarolina #CyberSecurity #CyberAttack #LockCrypt #Ransomware

Data-driven ‘everyware’

Tom Mitchell:

“A computer program is said to learn

- from experience E (training set: legal text)
- with respect to some class of tasks T (e.g. prediction of judgements)
- performance measure P (compare to actual outcome, or to prediction by legal experts)

if

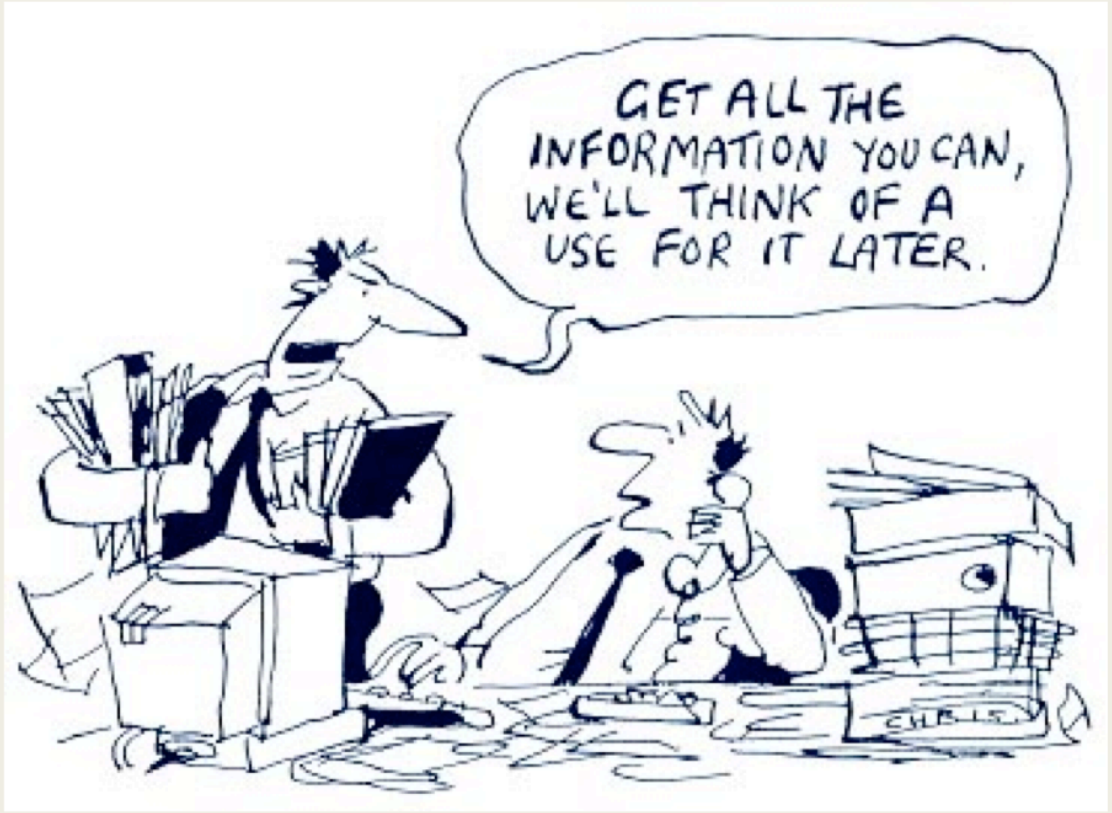
- its performance at tasks in T,
- as measured by P,
- improves with experience E.”

- ML often ‘parasites on’ human domain expertise (notion of ‘ground truth’)

Data-driven ‘everyware’

- The politics are in who get to determine E, T and P
  - The ethics are in how they are determined
    - Law concerns the contestability

Bad research design: ‘low hanging fruit’



Bad research design: 'low hanging fruit'

First Law of Informatics (Van der Lei, 1991, Cabitza et all, 2017):

- *'Data shall be used only for the purpose for which they were collected. And the collateral: If no purpose was defined prior to the collection of data, then the data should not be used.'*

Data-driven ‘everyware’

- Core to ML:
  - *finding the mathematical function that optimizes the relationship between input and output*
  - *based on the assumption that such an ideal target function exists*

  - *E = training data, validation data, test data*
  - *T = predict outcome of case*
  - *P = percentage correct compared to actual cases, or to human prediction*

  - *algorithms cannot be trained on future data*
  - *Wolpert’s NFL theorem: temporality = uncertainty*

RESISTING computational overdetermination

- machine learning is about optimizing the approximation of an ideal mathematical target function that links input data with output data
- since ML cannot train on future data the uncertainty of infinity has the final say (Godel, Wolpert)

- it becomes important to ensure actionable transparency:
  - *contestability and foreseeability*
- ito serve an actionable right to object:
  - *against processing of personal data*
  - *against machine decisions*

AI abolitionism?

GDPR: the big differences

1. Regulation instead of Directive (direct application)
2. Extraterritorial jurisdiction (processing personal data of individuals in the EU)
3. Smart enforcement framework (fines, compensation, injunction, collective action)

GDPR: the big differences

Smart enforcement framework (fines, compensation, injunction, collective action):

77 Right to lodge a complaint with a supervisory authority

78 Right to an effective judicial remedy against a *supervisory authority*

79 Right to an effective judicial remedy against a *controller or processor*

80 Right to mandate a non-for-profit to lodge a complaint and to exercise the right to receive compensation

82 Right to compensation and liability

83 General conditions for imposing administrative fines

- *83.1 effective, proportionate and dissuasive*
- *83.4 maximum 2% global turnover*
- *83.5 maximum 4% global turnover*

GDPR compliance and (new) business models

What's wrong with compliance?

[spoiler: nothing]

Rule of Law (normative infrastructure):

- our rights should not depend on the ethical inclinations of individual companies
- business should be able to act ethically without being pushed out of the market

- e.g. tracking walls don't align with macro-protection against computational overdetermination
- they will be outlawed: back to genuine consent, subscription models and contextual advertising

Consent

Art. 6.1(a)

consent of the data subject for *one or more specific purposes*

Art. 5.1(c):

processing must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed

Consent

Art. 7. 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.

7.2. 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.

Consent

Art. 7.3. 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*.

Consent

Art. 7.4. 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.

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.

NB! Check the negotations on the EP draft ePrivacy Regulation, notably 8.1a ePR

EP draft of the ePR

8.1a. No user shall be denied access to any information society service or functionality, regardless of whether this service is remunerated or not, on grounds that he or she has not given his or her consent under Article 8(1)(b) to the processing of personal information and/or the use of processing or storage capabilities of his or her terminal equipment that is not necessary for the provision of that service or functionality.

Automated decisions and profile transparency

4(4) GDPR ‘profiling’ means:

- any form of automated processing of personal data consisting of
- *the use of personal data to evaluate certain personal aspects relating to a natural person,*
- *in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;*

Automated decisions and profile transparency

- Art. 15.1(h) GDPR provides a legal right to obtain information about:
  - *the existence of automated decision-making,*
  - *including profiling,*
  - *referred to in Article 22(1) and (4) and, at least in those cases,*
  - *meaningful information about the logic involved, as well as*
  - *the significance and the envisaged consequences of such processing for the data subject.*

➢ *Art. 13.1(g) and art. 14.1(h) obligate the controller to provide this information*

➢ Art. 12.1 requires that the information is provided 'in a concise, transparent, intelligible and easily accessible form, using clear and plain language,

Automated decisions and profile transparency

22.1 The data subject shall have the right:

- not to be subject to a decision based solely on automated processing,
- including profiling,
- which produces legal effects concerning him or her or similarly significantly affects him or her.

Automated decisions and profile transparency

Recital (71) adds:

- *(…) such as automatic refusal of an online credit application or e-recruiting practices without any human intervention.*
- *Such processing includes 'profiling' that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person,*
- *in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements,*
- *where it produces legal effects concerning him or her or similarly significantly affects him or her.*

Automated decisions and profile transparency

Art. 29 WP considers that

- ‘if someone routinely applies automatically generated profiles to individuals
- without any actual influence on the result,
- this would still be a decision based solely on automated processing’

[calling this ‘*fabrication of human involvement*’]

Automated decisions and profile transparency

22.2 Paragraph 1 shall not apply if the decision:

a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

c) is based on the data subject's explicit consent.

Automated decisions and profile transparency

22.3 In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall:

- implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests,
- at least the right to obtain human intervention on the part of the controller,
- to express his or her point of view and
- to contest the decision.

Automated decisions and profile transparency

22.4 Decisions referred to in paragraph 2 shall:

- not be based on special categories of personal data referred to in Article 9(1),
- unless point (a) or (g) of Article 9(2) applies and
- suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

Automated decisions and profile transparency

-

Recital (71) adds:

- *In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision (my emphasis).*

Automated decisions and profile transparency

-

Recital (71) adds:

- *'That right should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software. However, the result of those considerations should not be a refusal to provide all information to the data subject.'*

Automated decisions and profile transparency

1. Don’t confuse profile transparency or explanations with legal justification of the decision itself!

   - *justification could be ‘freedom to contract’ [but is not unlimited]*
   - *in case of government agencies, legality principle requires attribution of competences*
   - *in case of criminal charge high standards apply, e.g. presumption of innocence*

Automated decisions and profile transparency

2. Don’t be overly impressed with the claim that there is necessarily a trade-off between predictive accuracy and ‘interpretability’ of the system

   - *this depends on uncontroversial ‘ground truth’, otherwise accuracy cannot be established*

Automated decisions and profile transparency

3. Discriminate between exploratory and confirmatory research:

   - *performance claims should be based on confirmatory research*
   - *this implies research in causalities*
   - *note: the more training data, the more spurious patterns*

- Duncan Watts on ‘Interpretation and Prediction’:

1. Exploratory ML
2. Confirmatory ML

- **Duncan Watts on Interpretation and Prediction:**

1. <u>Exploratory ML</u> 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.*

- **Duncan Watts on Interpretation and Prediction:**

1. **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).***

Automated decisions and profile transparency

4. ’Meaningful information about the logic of processing’:

   - *ex ante (info on research design, on relevant output, e.g. classifiers)?*
     - research design info, think of DPAs
   - *ex post (determination of parameters that informed individual decision)?*
     - LIME, counterfactuals, built in transparency

Automated decisions and profile transparency

Wachter, Mittelstadt and Russell propose 'counterfactual explanations':

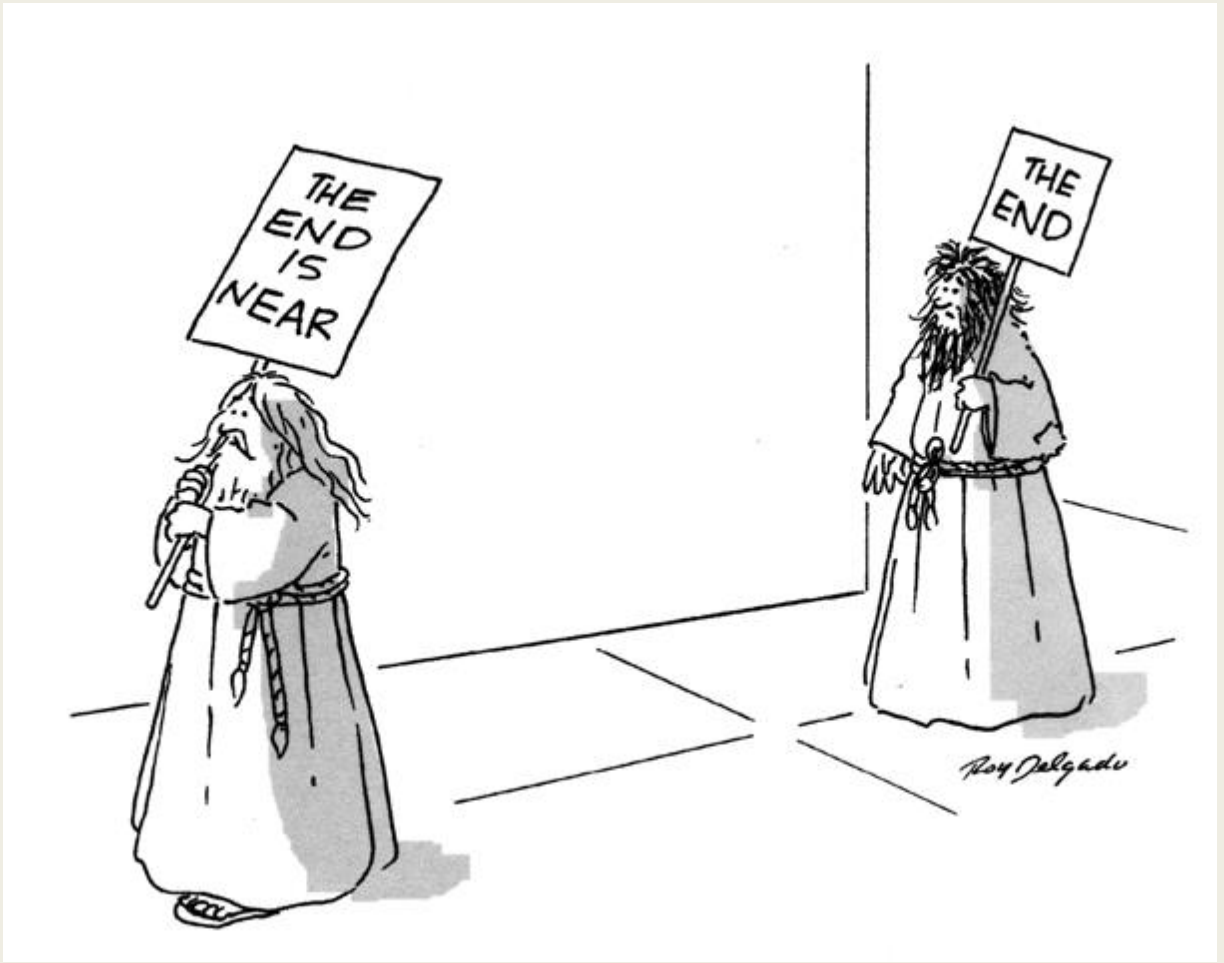
- *'Counterfactuals describe a dependency on the external facts that lead to that decision without the need to convey the internal state or logic of an algorithm'*
- *'What would need to change in order to receive a desired result in the future, based on the current decisionmaking model'*

Automated decisions and profile transparency

- the ‘need to change’ concerns all sides (not just the data subject):
- includes a redistribution of actionability and responsibility amongst developers, profilers and those profiled

Rethinking business models

- computational overdetermination (C) will suffer a setback
- good for the methodological integrity of ML ‘anywares’
- business models based on CO will also suffer a setback
- new compliant, creative, business models will emerge
- both for
  - *those who are actually selling a service or a product*
  - *and for those selling TETs*


THE END IS NEAR
THE END
Roy Delgado