

# Why and how algorithms may be treated as, or in, archival records

IRFD Network presentation, Aalborg University

Herbjørn Andresen, 18.01.19

# Concept of records, introductory emphasis

- Schellenberg: *Byproducts of administrative activity*
- Yeo (2007): *Records may also contain representations of ... [these 'alsos' are normally] subordinate to the representation of the activity itself. They may be present to [...] or **to supply or explain** its context.*
- No intention of contradicting the prevalence of *activity representation*
- Still, in some cases, understanding *content representations* is necessary
  - This is often trivial – many records explain themselves quite well
  - But some aspects of explanation might be a bit more troublesome

# Transparency and accountability

- The evidential value of the record, as we perceive it, is primarily an effect of its fixed relationship to an activity
  - If I was denied a visa, for instance, a short prose explaining why, for me to understand, will add to the transparency – not to the evidential value
  - Understandable content can not be overlooked in the quest for transparency
- What about an autonomous weapon system, picking its own targets?
  - The evidence – in the records and on the ground – is the targets it picked
  - Transparency involves an explanation: What information or behaviour it responded to, what reasoning was involved, what probability thresholds, etc.

## (Small) shift: Transparency and Explanations in GDPR

- The right to an explanation: Its *basis* is GDPR Art. 22
  - a right not to be subject to automated decisions
  - Often + rightly criticized for vagueness and ambiguity
- Right to explanation: Articles 13(2)(f), 14(2)(g) and 15(1)(h):
  - *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.*

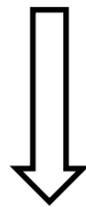
# The 'right to explanation' is debated

- Selbst and Powles (2017) conclude that GDPR introduces an enhanced right to explanation
- Disputed by Wachter, Mittelstadt and Floridi (2017)
  - Elaborates on a difference between *ex ante* and *ex post* explanations
  - *Ex ante* can merely explain the 'functioning of the system'
  - One can only explain what occurred *ex post* (i.e. capture records)
- Mendoza and Bygrave (2017), moderate support for the existence of a right to explanation, mainly as a consequence of a right to contest automated decisions

# Explanations internal or external to the records

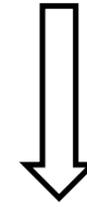
*A tempting, but probably too simple, illustration*

*Ex ante*



External

*Ex post*



Internal

# ***Ex ante* explanations, external to the record**

- Deterministic automation, a certain input yields a certain record
  - A recipe for the record system
  - The explanation can predict the outcome, in lieu of actual records
- Non-deterministic automation, predicts different possible outcomes
  - for instance profiling of travel documents and various traits of traveling behaviour in order to pick which passengers the customs' officers should expose to a more thorough luggage check
  - An *ex ante* explanation can provide transparency on criteria and probabilities – but is not capable of reconstructing an actual decision for a specific point in time

# ***Ex post* explanations, internal to the record**

- ‘Internal to’ the record is not an unequivocal term...
  - A strong sense of the term: ‘The Archive inside the document’
    - Concept study in a Norwegian municipality, embedding archival traces in pdfs using xmp
  - Probably more common: Information that forms a record, as a compound of various elements, within a defined records system
- If an automated decision is to be explained, *ex post*, the record system could (and should) include a design that captures the information used and deduced, the reasoning, weights and probabilities etc.

## Is all good, then?

- One kind of situation is still not covered: Sometimes you *can* predict an outcome, without being able to explain it
  - Cf. Schum (2001/1994, p. 198)
  - Applies to some machine learning algorithms
- An *ex ante* explanation is called for
  - but what could an explanation of a ‘non-understood prediction’ consist of?
- One thinkable way out: ‘reification’ of the most troublesome of algorithms
  - Assigning an identifier for reference (e.g. a patent no.) for *ex post* documentation
  - If successful, how useful would it be?

# Conclusions

1. Transparency and a right to explanation is one (and definitely not the only) example of external circumstances that impact the theoretical and conceptual frames of our merry trade
2. This was a short peek under the blanket, to confirm 'yes, there is a monster under here' – and it might be a bit of overthinking
3. Ideas and advice are welcome