THE WHISPER CHALLENGE

Mireille Hildebrandt

THE RULE OF PROXIES IN DATA SCIENCE

Mireille Hildebrandt

For further reading:

- Law for Computer Scientists and Other Folk (eg discussing the relationship between law and ethics: https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198860877.001.0001/oso-9780198860877-chapter-11 (the whole book is open access)

- Smart Technologies and the End(s) of Law (explain the background of all this in terms of both philosophy of law and philosophy of technology): https://www.elgaronline.com/view/9781849808767.xml

More specifically on behaviourism, bias and proxies:

- 'Learning as a machine' (my keynote at the 2016 Learning Analytics and Knowledge Conference): https://learning-analytics.info/index.php/JLA/article/view/5367

- 'The issue of bias. The framing powers of machine learning': https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3497597

- 'The issue of proxies in recommender systems': https://www.frontiersin.org/articles/10.3389/frai.2022.789076/full

On why Kahneman gets it wrong:

- 'the bias of bias in behavioural economics': https://www.nowpublishers.com/article/Details/RBE-0092

O**n why current forms of machine learning will not achieve 'intelligence':**

- 'The promise of artificial intelligence. Reckoning and judgment': https://mitpress.mit.edu/books/promise-artificial-intelligence

On the Goodhart effect:

- 'Improving ratings: audit in the British university system: https://www.cambridge.org/core/journals/european-review/article/abs/improving-ratings-audit-in-the-british-university-system/FC2EE640C0C44E3DB87C29FB666E9AAB

- Datum = something given
  - *Raw data is an oxymoron (Gitelman), data is a construction*
- Factum = something made
  - *Les faits sont faits*

- A tribute to Sabine Krolak-Schwerd
- Data Science ***as science***, methodological integrity of the research design
  - *Attention to the key role played by the relevance and distribution of data*
  - *'Groundtruthing' as a verb not a noun*
  - *The need for rigorous mathematical verification, validation and testing*
- A rule of law perspective:
  - *The impact of upstream design decisions on*
    - power relationships and fundamental rights
    - downstream reliability, robustness and resilience
- Legal Protection by Design
  - *NOT legal by design*
  - *Building checks and balances into the new ML-driven ICIs*
  - *Making machine decisions and machine behaviours contestable*

What's next?

- To count or not to count
- Data, metrics, variables, tasks and mathematical patterns as proxies
- Proxy and principal
- Getting the proxy right
- Lure and pitfalls of proxification
- The rule of proxies and the rule of law

What's next?

- To count or not to count
- Data, metrics, variables, tasks and mathematical patterns as proxies
- Proxy and principal
- Getting the proxy right
- Lure and pitfalls of proxification
- The rule of proxies and the rule of law

It would be nice if all of the data which sociologists require could be enumerated because then we could run them through IBM machines and draw charts as the economists do. However, not everything that can be counted counts, and not everything that counts can be counted

– William Cameron, *Informal Sociology* (1963)

CYBERNETICS

1. What matters is incomputable
2. It can nevertheless be made computable
3. In different ways – and that difference matters

‘counting is an act of labelling,

hence the mathematician is an active subject right from the start’

Jean Paul Van Bendegem (2012)

a human computer is a person "supposed to be following fixed rules; he has no authority to deviate from them in any detail.”

Turing (1950)

- To make things computable you need proxies (e.g. justice: fairness)
- You need a tertium comparationis (e.g. outcome equality, i.e. a fair share)
- That allows you to compare, rank and calculate (e.g. income = low hanging fruit, Sen)
- This is how economics works (see discussion of BNP, SDG)
- Qualification necessarily precedes quantification (Callon and Law on ‘qualculation’)

What's next?

- To count or not to count
- Data, metrics, variables, tasks and mathematical patterns as proxies
- Proxy and principal
- Getting the proxy right
- Lure and pitfalls of proxification
- The rule of proxies and the rule of law

Briefing | The world that Bert built

Huge "foundation models" are turbo-charging AI progress

They can have abilities their creators did not foresee

Jun 11th 2022

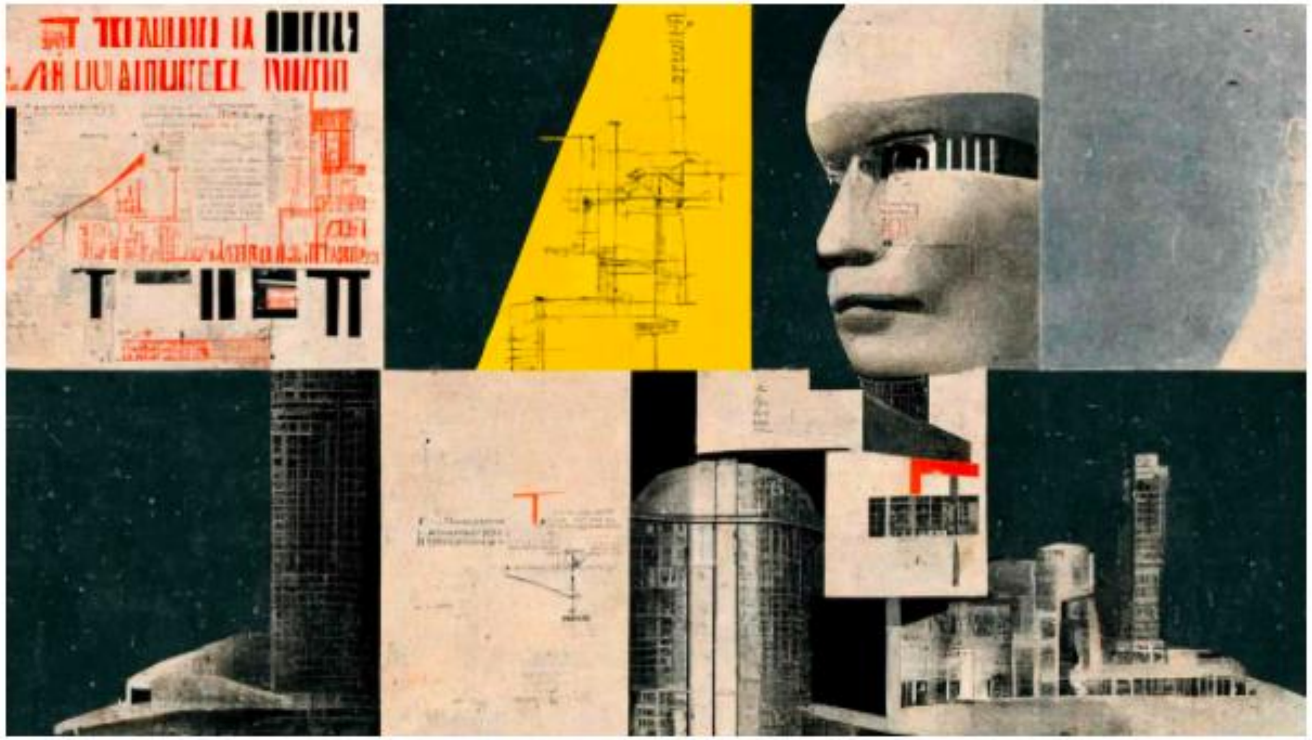


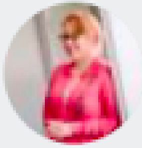
IMAGE: MIDJOURNEY

Collage | Dali | Bruegel

Why speak of

foundation model

instead of base model?



MMitchell
@mmitchell_ai

Reminder to everyone starting to publish in ML: "Foundation models" is *not* a recognized ML term; was coined by Stanford alongside announcing their center named for it; continues to be pushed by Sford as *the* term for what we've all generally (reasonably) called "base models".

> **Stanford HAI** @StanfordHAI · 03/06/2022
> Oversight of foundation models requires multi-stakeholder partnerships, including independent organizations not driven by commercial incentives. We need to leverage the collective wisdom of the community and represent the diverse voices of the people that this technology impacts. twitter.com/CohereAI/statu...
>
> Show this thread

It’s the base

of other models, yes -

but the foundation is

the real world.

Or is it?

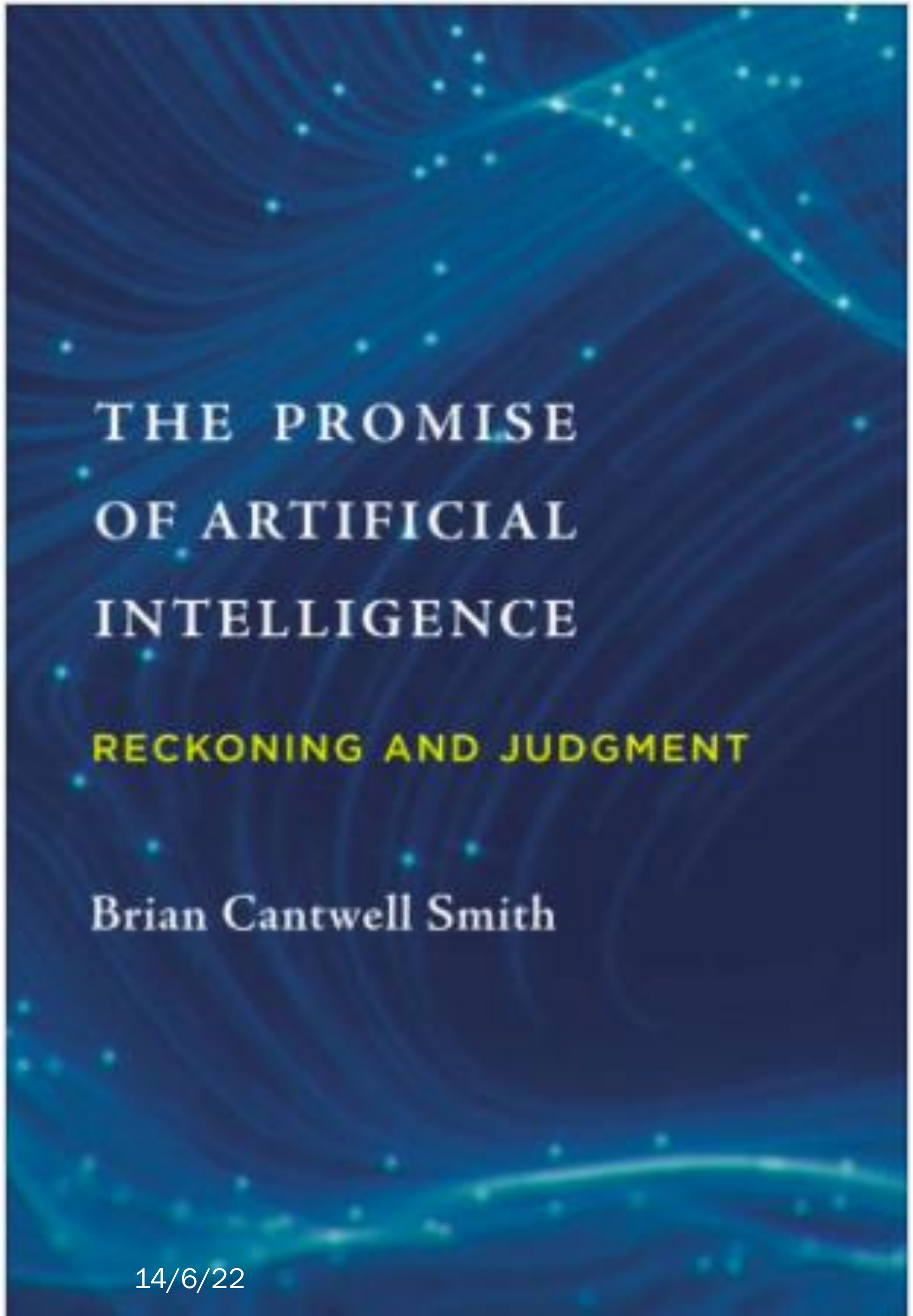
MMitchell
@mmitchell_ai

Reminder to everyone starting to publish in ML: "Foundation models" is *not* a recognized ML term; was coined by Stanford alongside announcing their center named for it; continues to be pushed by Sford as *the* term for what we've all generally (reasonably) called "base models".

Stanford HAI @StanfordHAI · 03/06/2022

Oversight of foundation models requires multi-stakeholder partnerships, including independent organizations not driven by commercial incentives. We need to leverage the collective wisdom of the community and represent the diverse voices of the people that this technology impacts. twitter.com/CohereAI/statu...

Show this thread



THE DIFFERENCE THAT MAKES A DIFFERENCE

BATESON (1972)

Proxies in data science:

- Data
  - *legal text corpora as a proxy for positive law when training for legal search*
  - *labelled X-rays as a proxy for correct diagnoses when training for medical diagnostics*
  - *www-data as a proxy for 'language acquisition(?)' when training foundation models*
- Variables
  - *income as a proxy for well being or wealth when training for equality*
  - *negative or positive labels as a proxy for emotional engagement when training for sentiment analysis*
- Parameters
  - *weights in an ANN as a proxy for correlations between variables when using backpropagation*
- Tasks
  - *legal text classification as a proxy for ordering relevant documents when training for legal search*
- Mathematical patterns
  - *base models eg BERT, GPT3, DALL-E as proxies for 'language acquisition (?)' when used for further training*

What is data science?

- IBM:
  - *Data science combines the scientific method, math and statistics, specialized programming, advanced analytics, AI, and even storytelling to uncover and explain the business insights buried in data.*
- HBR:
  - *What do data scientists do? According to interviews with more than 30 data scientists, data science is about infrastructure, testing, using machine learning for decision making, and data products. Data science is being used in numerous fields, but it’s not all about deep learning or the search for artificial general intelligence. In fact, the skills needed include communication and storytelling. But data science is becoming more specialized, and with that the skills data scientists need are evolving. In addition, ethics is becoming a bigger and bigger challenge.*

Journal of Behavioral Data Science, 2021, *1* (1), 1–16.
DOI: https://doi.org/10.35566/jbds/v1n1/p1

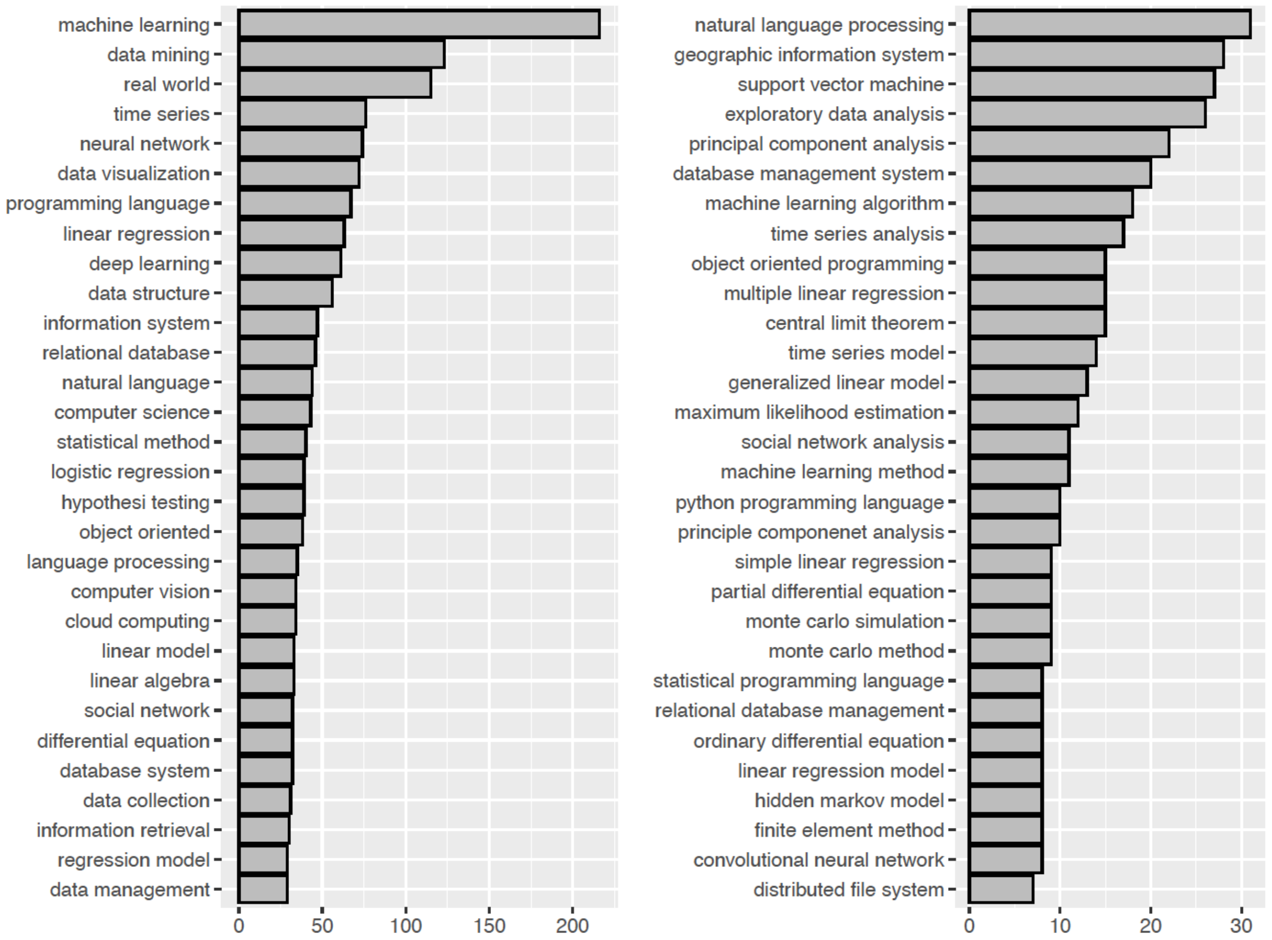
What is Data Science? An Operational Definition based on Text Mining of Data Science Curricula

Zhiyong Zhang[1] and Danyang Zhang[2]

[1] University of Notre Dame
zzhang4@nd.edu
[2] University of Texas–Austin
danyang.zhang@utexas.edu

Abstract. Data science has maintained its popularity for about 20 years. This study adopts a bottom-up approach to understand what data science is by analyzing the descriptions of courses offered by the data science programs in the United States. Through topic modeling, 14 topics are identified from the current curricula of 56 data science programs. These topics reiterate that data science is at the intersection of statistics, computer science, and substantive fields.

Keywords: Data Science · Topic Modeling · Data Science Curriculum.



(a) Two-word phrases (b) Three-word phrases

Figure 2: Most frequently used phrases

4.2 Conclusion

The goal of this study is to understand what data science is through the mining of the courses offered by data science programs in the US to hopefully provide a better definition of data science. We adopted a bottom-up approach to mining the description information of individual courses taught in current data sciences programs. Although we identified fourteen topics among all the courses, it is still difficult to provide a concise and conclusive definition of data science. However, we believe our results can provide useful information on how to operate data science programs. The results of our study further reiterate the notion that data science is at the intersection of statistics, computer science, and applications. A major contribution of our study is to provide empirical support to a better understanding of data science.

- Is deciding what ‘is’ data science something to be inferred statistically from publicly available curricula that claim to offer ‘data science’?
  - *Is that perhaps the wrong proxy?*
  - *Or does the question deserve a normative answer?*

Data Centrism and the Core of Data Science as a Scientific Discipline

Thilo Stadelmann, Tino Klamt and Philipp H. Merkt

Abstract Data science is one of the most significant developments in computing in the 21^{st} century. It is also described as a discipline in the making, drawing principles, methods and tools from established fields like computer science, statistics, science, business, politics, and any domain with adequate data. What are data science's underlying principles and techniques (models, methods) that are applicable across different use cases and fields of application? What novel aspect of science underlies this emerging discipline? We argue that it is *data centrism* – the reliance on data itself, in mindset, methods and products – that makes data science more than the sum of its parts, as this is not done in any other discipline.

We believe this aspect to be the core of data science because it firmly differentiates data science from related fields, as is demonstrated by the following exemplary consideration of such related fields.

Machine learning revolves around *learning from data* (not data itself): principles and methods to gain general knowledge out of finite data (Samuel (1959) put the highest weight on the learning outcome itself in his famous definition and neglected the input entirely). Despite the efforts of Andrew Ng to teach the field otherwise (Ng, 2021), this is still mainly a model-centric endeavour, i.e., conferences, sub-fields and projects revolve around model architectures as the centre pieces. Then, suitable data to satisfy the needs of the predominantly supervised modeling approaches has to be delivered for machine learners to usually take up the work. It is arguably the influence of data science that un- and semi-supervised methods are increasingly researched and used in recent years: Unsupervised learning was for a long time mainly equated to clustering (Mitchell, 1997). The rise of unsupervised learning as, e.g., spearheaded by Meta's Yann LeCun (LeCun and Misra, 2021), coincides with the rise of data-driven companies like Meta's Facebook and their needs as addressed by data science.

Empiricism is the driving force in data science: in contrast to pre-conceived models of reality, data science reinforces the *mindset* to establish theories out of the patterns that arise from potentially vast amounts of data (i.e., empirical evidence rather than human intuition) (Hey et al, 2009). The effect of this is that data science models tend to become complex and opaque, as they didn't originate in a simple human idea, but emerged in a data-driven way. Deep learning methods are a good example for this, and the recent trend to research and apply *explainable and trustworthy methods* (Samek et al, 2019; Amirian et al, 2021) can be seen as a direct reaction to the data science mindset: If the data itself is determining the model, the discipline responsible for this development, as a next step, has to provide methods that make this machine-conceived models again amenable to human intuition, decision and control.

- Digital data is NOT equivalent with empirical fact
- Mistaking data science for empirical science implies
- The inversification of proxy and principal

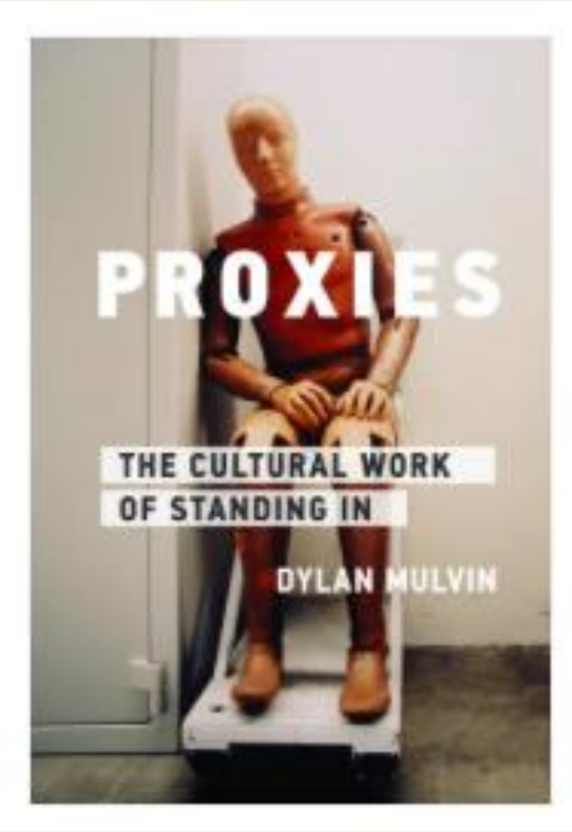
theoretical reasons, then it is important to have the best possible data science methods available. Mitchell (1997) proves that no learning is possible without assumptions; we argue that data science is home to those methods that deliberately work with the least possible amount of assumptions, which sometimes is the only viable route to take. Of course, such approaches can only detect correlations in the data and make no statements about causality (Cap, 2019). But while correlation is not causation, correlation often is enough (Brodie, 2019b,a; Stockinger et al, 2019). Hence, furthering data science as a data-centric discipline adds something unique to the quiver of scientific methodologies. The

- Better:
  - *unearth you assumptions and address them*
  - *by tracing their implications*

What's next?

- To count or not to count
- Data, metrics, variables, tasks and mathematical patterns as proxies
- Proxy and principal
- Getting the proxy right
- Lure and pitfalls of proxification
- The rule of proxies and the rule of law


Proxy



THE STAND IN

Proxies and principal

Meaning (via google, based on https://languages.oup.com/google-dictionary-en/):

- the authority to represent someone else, especially in voting
- **a figure that can be used to represent the value of something in a calculation**.

Ethymology (https://www.etymonline.com/word/proxy)

- Procuratio (caring for, management, administration)

The issue of proxies

- One thing ’standing in’ for another:
  - *in mathematics numbers don’t necessarily ‘stand in for’ something else*
  - *E.g. –6 – 3 = – 9 (what, apples?), or square root of 2*
  - *In statistics and applied math (social science, computer science):*
    - A variable (x, y, z) stands for a feature/category/type with dedicated values:
      - a symbol (usually a letter) standing in for an unknown numerical value in an equation (https://www.britannica.com/topic/variable-mathematics-and-logic)
      - algebra (functions, equations)
      - imagine how this enabled abstraction

Roger K Moore @rogerkmoore · 3d

We should never have called it "language modelling" all those years ago; it was (and still is) "word sequence modelling". Confusion always occurs when you label an algorithm with the name of the problem you're trying to solve, rather than with what it actually does. @GaryMarcus

The issue of proxies

One thing ’standing in’ for another:

- a proxy in algebra and ML serves as the tertium comparationis
- E.g. a variable brings together different things under the same ‘heading’
- quantification is contingent upon prior qualification
- language as word sequencing (on LLMs)
- justice as fairness
  - *Fairness as a specific type of distribution in a dataset (outcome oriented)*
  - *Fairness as being heard and taken into account (process oriented)*
- quality of academic research
  - *Volume of publications in double blind peer reviewed international journals*
  - *Citation score (impact factor)*

European Review

Article contents

'Improving ratings': audit in the British University system

Published online by Cambridge University Press: **13 July 2009**

Marilyn Strathern

Article Metrics

Get access Share Cite Rights & Permissions

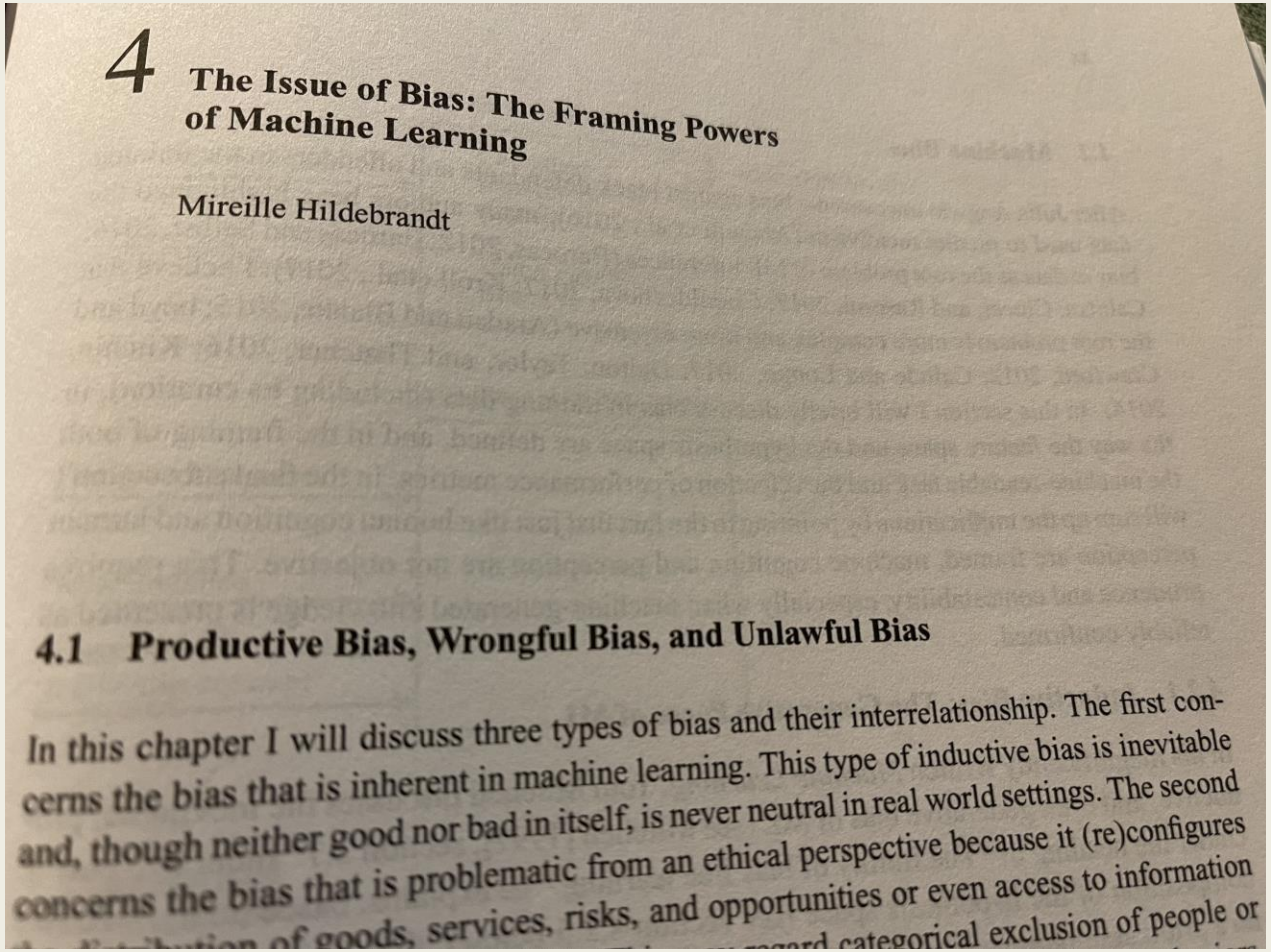
Abstract

This paper gives an anthropological comment on what has been called the 'audit explosion', the proliferation of procedures for evaluating performance. In higher education the subject of audit (in this sense) is not so much the education of the students as the institutional provision for their education. British universities, as institutions, are increasingly subject to national scrutiny for teaching, research and administrative competence. In the wake of this scrutiny

What's next?

- To count or not to count
- Data, metrics, variables, tasks and mathematical patterns as proxies
- Proxy and principal
- Getting the proxy right
- Lure and pitfalls of proxification
- The rule of proxies and the rule of law

- The relevance of the proxy depends on the purpose
- This is even more important in the case of ‘general purpose’ systems
- What is relevant for the myriad downstream purposes?
  - *acknowledging that N=ALL is a hoax*
  - *even all data on the web is not equivalent with ’real life’ or ‘real world’*

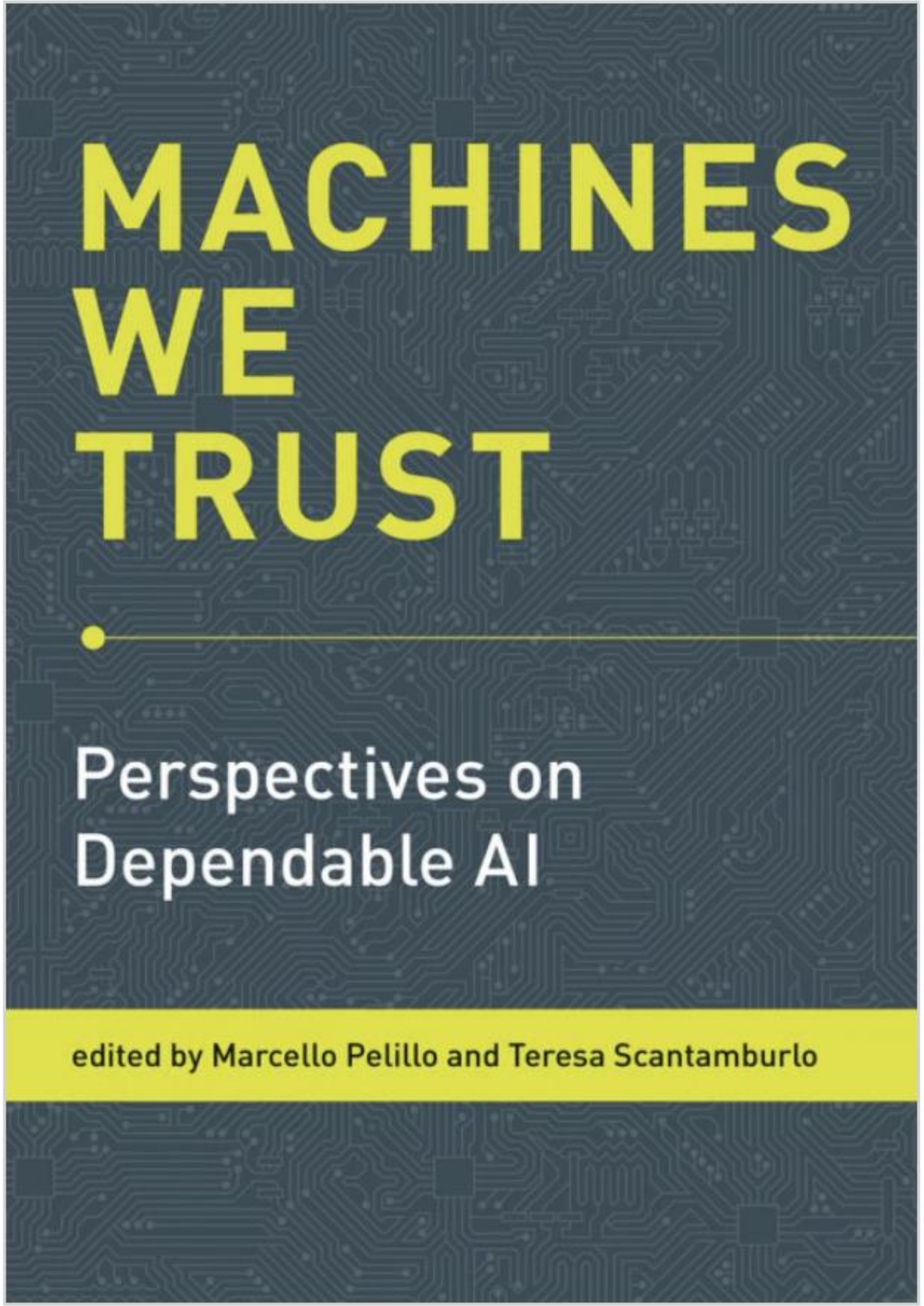


4 The Issue of Bias: The Framing Powers of Machine Learning

Mireille Hildebrandt

4.1 Productive Bias, Wrongful Bias, and Unlawful Bias

In this chapter I will discuss three types of bias and their interrelationship. The first concerns the bias that is inherent in machine learning. This type of inductive bias is inevitable and, though neither good nor bad in itself, is never neutral in real world settings. The second concerns the bias that is problematic from an ethical perspective because it (re)configures the distribution of goods, services, risks, and opportunities or even access to information … regard categorical exclusion of people or



The issue of bias

Machine Learning is inherently biased

- This is why ML is productive

The issue of human bias

Gadamer, philosophical hermeneutics:

- Bias, fore-structure of being in the world, interpretation as understanding
- Such fore-structure is productive and inherent in the possibility to understand
- Understanding is not given with a text or action but emerges in its reading
- Prejudgements interact with the horizon of a text or an action, reshaping both
- This implies a critical relation between interpreter and interpreted
- **Bias does not determine our understanding but co-constitutes it**

The issue of machine bias

- Derived bias:
  - *Distribution of the training data (ground truth)*
  - *Type of labels, attribution of labels (choice of target variable)*
  - *Goal specification*
- Bias inherent in ML:
  - *Loss function (related to choice of ground truth)*
  - *Reward function (related to choice of goals)*
  - *Construction of the hypothesis space*
  - *Limitations of computational framing*
    - Need for discrete (mutually exclusive) variables
    - Need to assume that feature variables are independent of target variable
    - Training on future data is not possible – problematic feedback loops

What's next?

- To count or not to count
- Data, metrics, variables, tasks and mathematical patterns as proxies
- Proxy and principal
- Getting the proxy right
- Lure and pitfalls of proxification
- The rule of proxies and the rule of law

‹ Articles

THIS ARTICLE IS PART OF THE RESEARCH TOPIC
Improving Human-Machine Feedback Loops in Social Networks View all Arti

REVIEW article
Front. Artif. Intell., 28 April 2022 | https://doi.org/10.3389/frai.2022.789076

Check for updates

The Issue of Proxies and Choice Architectures. Why EU Law Matters for Recommender Systems

Mireille Hildebrandt[1,2*]

[1]Institute of Computing and Information Sciences (iCIS), Science Faculty, Radboud University, Nijmegen, Netherlands
[2]Research Group Law Science Technology & Society (LSTS), Faculty of Law and Criminology, Vrije Universiteit Brussel, Brussels, Belgium

Recommendations are meant to increase sales or ad revenue, as these are the first priority of those who pay for them. As recommender systems match their recommendations with inferred preferences, we should not be surprised if the algorithm optimizes for lucrative preferences and thus co-produces the preferences they mine. This relates to the well-known problems of feedback loops, filter bubbles, and echo chambers. In this article, I discuss the implications of the fact that computing systems necessarily work with proxies when inferring recommendations and raise a number of

Inversification of proxy-principal relations

Behaviourism (Pavlov, Skinner, Watson) underpinning behavioural economics:

- The primitive (principal) is an observable behaviour
- The proxy is a natural language concept (vague, imprecise, ambiguous)
- Cognitive bias distracts from the primitives, need to be removed

Machine learning

- Fairness or justice are impossible concepts: vague, imprecise, ambiguous
- The proxy is a machine readable distribution deemed to be fair or just
- Or fairness/justice are just proxies for a fair distribution in the data?

Inversification of proxy-principal relations

Rational choice theory (Coase, Elstar) underpinning neoclassical (neoliberal) economics:

- The primitive (principal) is individual rational choice in the context of game theory
- The proxy is a natural language concept (vague, imprecise, ambiguous)
- Concepts with open texture distract from the primitives, need to disambiguate and discretize

Machine learning

- Fairness or justice are impossible concepts: vague, imprecise, ambiguous
- The proxy is e.g. a multi agent system based on game theoretical assumptions
- Or fairness/justice are just proxies for the outcome of the MAS?

Inversification of proxy-principal relations

- A map is a proxy for a territory, domain or concept
- Developing a map is productive, it helps to navigate
- Mistaking the map (compression) for the territory creates blind spots
- Awareness of the reduction is key to maps being helpful
- Mapping can be done in many ways, it provides framing powers
- Mistaking the proxy for relevance for relevance itself has two implications:
  - *Hiding the framing power (who do the framing, whose options are framed)*
  - *Preempting the discussion on consequences (as these are 'made' inevitable)*

What's next?

- To count or not to count
- Data, metrics, variables, tasks and mathematical patterns as proxies
- Proxy and principal
- Getting the proxy right
- Lure and pitfalls of proxification
- The rule of proxies and the rule of law

Rule of proxies (in data science)

- On the role of proxies:
  - *Data is always a proxy for what is the case or what should be the case*
  - *Tasks are always a proxy for an intended purpose*
  - *Models are always a proxy for the system they model*
- On the rule of proxies:
  - *proxies have framing power*
  - *they make some things visible by making other things invisible*
  - *e.g. the duality of risk modelling (Claudio Ciborra)*

Rule of proxies (in data science)

- Who does the framing? Developers, data scientists
- Whose choices are framed? E.g. deployers and end-users of downstream systems
- To protect and institute whose choices? Whoever pays for the system

Rule of law (in EU)

- Who does the framing? EU legislature
- Whose choices are framed? E.g. Controllers (GDPR), Providers (AI Act)
- To protect and institute whose choices? Notably natural persons in the EU

Rule of law (in EU)

- Data Protection
  - *choice architecture of controllers – and processors*
- Charter of Fundamental Rights (and the European Convention of Human Rights)
  - *choice architecture of natural persons in the EU*
- AI Act and other parts of the EU Digital Strategy
  - *choice architecture of providers of AI systems, service providers etc.*

Rule of law (in EU)

Data Protection: GDPR

- Conditions for fair, transparent and lawful processing of personal data
- Principles: purpose limitation, data minimisation, accountability
- Legal basis: 6 ways to ‘ground’ processing (always based on necessity)
- Transparency requirements
- Risk approach: DPbD and DPIA
  - *Risk to fundamental rights and freedoms of natural persons*
- Accountability: fines and private law liability
- Brussels effect: those wishing to compete on the EU market will adapt, because …

Rule of law (in EU)

Charter of Fundamental Rights of the EU

Art. 7 Privacy

Art. 8 Data Protection

Art. 10 Freedom of thought, conscience and religion

Art. 11 Freedom of expression and information

Art. 16 Freedom to conduct a business

Art. 21 Prohibition of discrimination

Art. 52.3 aligns the scope of rights with that of the ECHR

Rule of law (in EU)

Proposed AI Act

- Mainly targets high risk AI systems:
  - *That have a potentially high impact on physical safety or health (ANNEX II)*
    - Eg medical devices, aircraft, toys etc.
  - *That have a potentially high impact on fundamental rights (ANNEX III)*
    - Eg when intended for deployment in context recruitment, insurance, policing

Rule of law (in EU)

Proposed AI Act

- Spells out a series of conditions (requirements) that must be met
  - *Before an AI system is placed on the market or put into service in the EU*
  - *Mainly addressing the providers*
  - *Who must conduct and document a Conformity Assessment (CE label)*
  - *Violation of the conditions (requirements) can result in*
    - high fines (up to 30 million euro or 6% global turnover)
    - private law liability (still awaiting the update of the Product Liability Directive)

Rule of law (in EU)

Proposed AI Act

- Spells out a series of conditions (requirements) that must be met:
  - *A dedicated risk management system must be in place*
    - Risk of deployment for intended [and other reasonably foreseeable] purposes
  - *Data governance*
    - Relevance of training, validation and test data, bias monitoring, GDPR data minimisation
  - *Technical documentation and record keeping*
    - Including automated logging
  - *Transparency for those who deploy the systems*
  - *Human oversight*
    - By design or by way of instruction

rule of proxies
rule of law

- Rule of law is about
  - *Bringing the rule of proxies under the rule of law*
    - Making the choice of proxies
      - *Visible*
      - *Contestable*
    - Addressing the Whisper Challenge

End