



INTERPETATION & PREDICTION MAGICAL THINKING IN ML

mireille hildebrandt

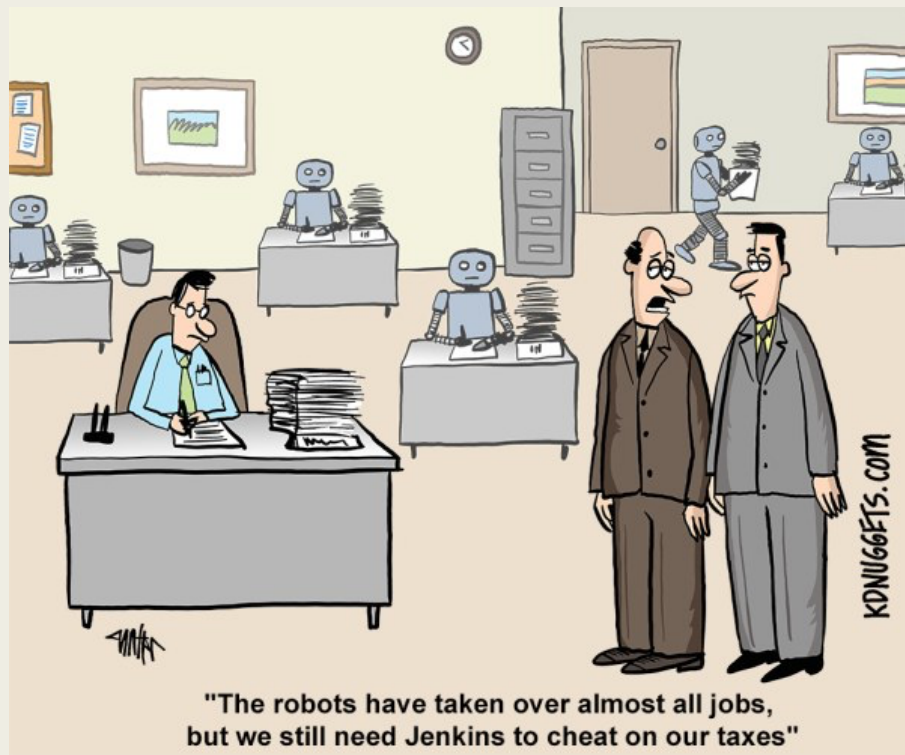
oracular AI



‘Like any oracle, we describe it as any decision aid capable of providing its users with very accurate responses (ie, no more than one error out of 20) and little or no explanation for that advice’ (Cabitza)

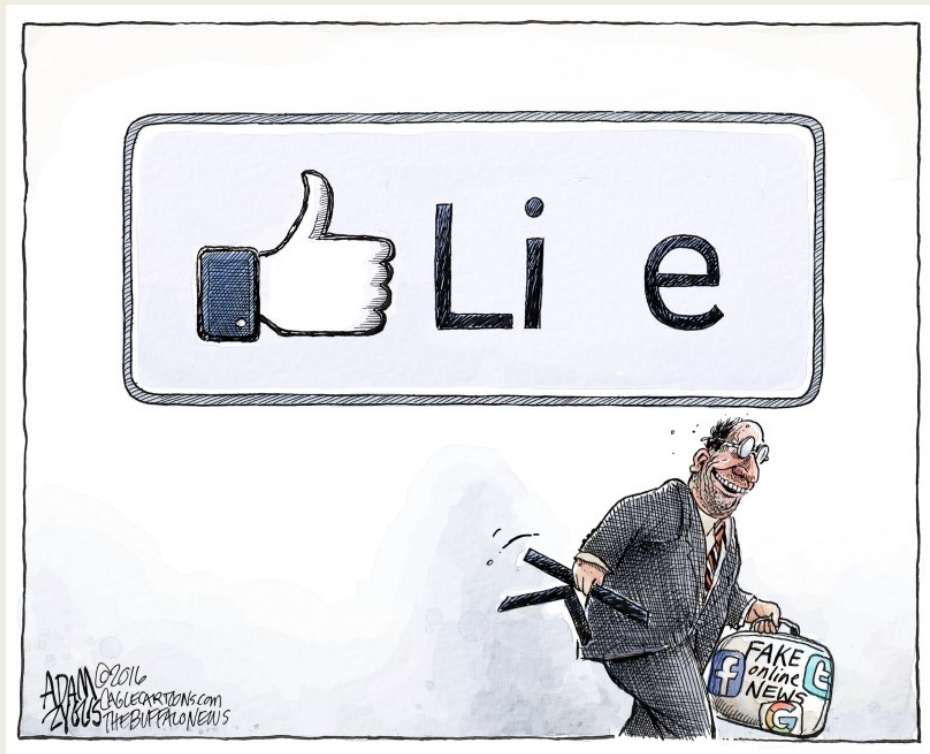
‘interpretability problem’ in machine learning

oracular AI



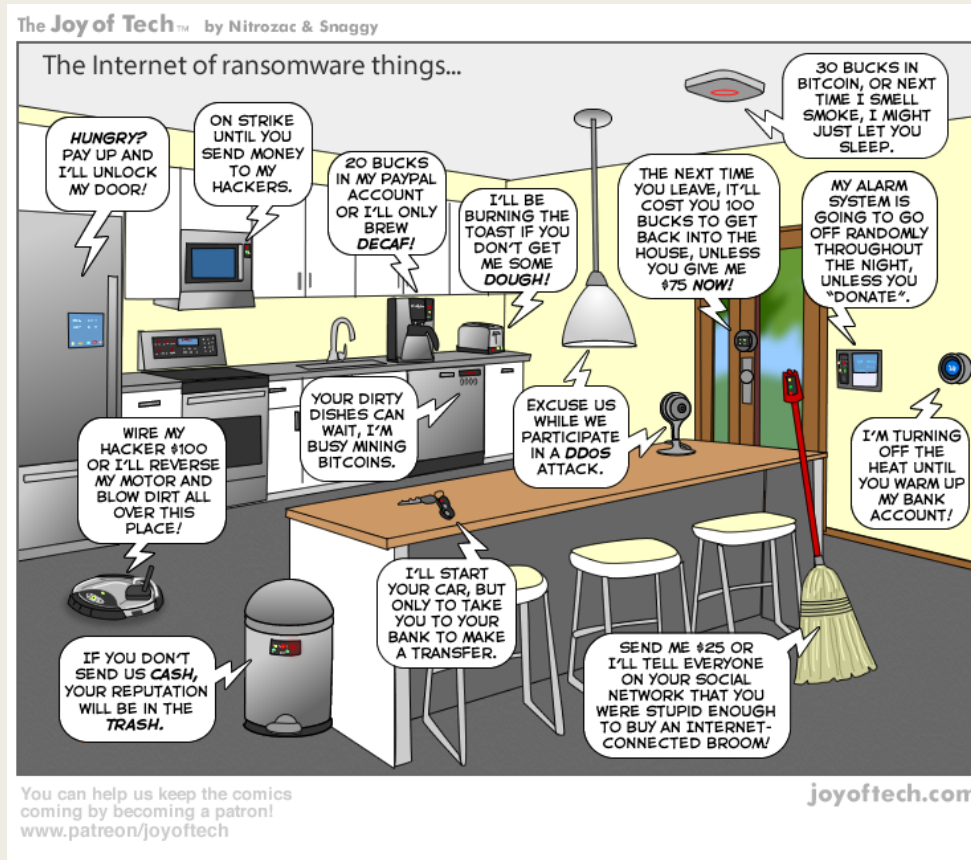
'prediction problem' of real life?

machinic neo-platinism



‘For the data scientist, computation plays the role of the intermediary between the imperfect world of data and the *pure function* that relates the features to the target’ (mcquillan)

machinic neo-platinism



machinic decision making's
real life effects in the IoT?



**“Go ahead and think that
I’m not really thinking. I thought
you would think that.”**

who is speaking?



“Go ahead and think that I’m not really thinking. I thought you would think that.”

What's Next?

1. Two magics of the web
2. Magical thinking
3. The future of the future
4. What machine learning does
5. Agonistic machine learning
6. From magic to rigorous speculation

I. two magics of the web

- Sir Tim Berners Lee (keynote WWW Conference 2007):
 - *Magic = things we don't understand, yet*
- 1. **when** does the microscopic interaction between two people on the Web scale to produce a new macroscopic phenomenon?
- 2. **what** 'creative magic' is required to identify new microscopic designs that could have positive macroscopic effects?

O'Hara, Contractor, Hall, Hendler, and Shadbolt (2012)

**'Web Science: Understanding the Emergence of
Macro-Level Features on the World Wide Web'**

Volume 4, Issue 2-3 Foundations and Trends® in Web Science

I. two magics of the web

- A Manifesto for Web Science @10:
 - *a sociotechnical perspective, including its mathematical properties, engineering principles and social impacts*
 - the notion of prosumer (naïve?)
 - personalization, social networks (echo chambers?)
 - linking documents, data, behaviours, devices, people (privacy: right not to be linked?)
 - who creates, who controls access to data, behaviours & inferences (platform economy)?
 - Web Observatory, Social Machines, Code as Law (code & architecture under the Rule of Law)?
 - issues of emergence, complexity, scale

I. two magics of the web

From interpretation to prediction:

- “The aim will be not **just** to understand the emergent phenomena, but rather to **develop means, methods and tools for controlling large-scale phenomena**, at least partially.”

O'Hara, Contractor, Hall, Hendler, and Shadbolt (2012), p. 6

I. two magics of the web

- **prediction**, control, influence – causality & correlation (*prediction = present futures*)
 - **interpretation**, meaning, understanding – performativity (*interpretation = anticipation*)
-
- **prediction is about the present**
 - **interpretation is about the future**

I. two magics of the web

- “The problem is sharpened by the desideratum that ‘programming the social computer’ must be achievable from within the social computer - the methods of Web Science should:
 - *output **policy for governments and large organisations**, but will also*
 - *democratise control by allowing people to develop social machines to achieve **their own smaller-scale, local, idiosyncratic purposes**”.*

O'Hara, Contractor, Hall, Hendler, and Shadbolt (2012), p.6

I. two magics of the web

- “The problem is sharpened by the desideratum that ‘programming the social computer’ must be achievable from within the social computer - the methods of Web Science should:
 - *output **policy for governments and large organisations, that incentivizes to***
 - *democratise control by allowing people to develop social machines to achieve **their own smaller-scale, local, idiosyncratic purposes**”.*

O'Hara, Contractor, Hall, Hendler, and Shadbolt (2012), p.6

2. magical thinking

- 'Machinic neo-platonism':

"I'm also curious about whether there is a fundamental mathematical law underlying human social relationships that governs the balance of who and what we all care about," Zuckerberg added.

"I bet there is."

Marc Zuckerberg, responding to Stephen Hawking (2015)

2. magical thinking

Magical thinking (wikipedia) =

a term used in anthropology and psychology, denoting the *fallacious* attribution of causal relationships between actions and events.

In anthropology: the attribution of causality between entities grouped with one another (coincidence) or similar to one another.

In psychology: the belief that one's thoughts by themselves can bring about effects in the world or that thinking something corresponds with doing it.

In both cases, the belief can cause a person to experience fear, seemingly not rationally justifiable to *an observer outside the belief system*, of performing certain acts or having certain thoughts because of an assumed correlation between doing so and threatening calamities.

2. magical thinking

Magical thinking (in the political economy of the onlife world) =

a mistaken belief in causal influences and/or predictive accuracy, based on illusionary wishful (or fearful) thinking about the nature of our world

getting rid of magical thinking requires facing

- the temporality of human interaction, and
- the inherent uncertainty this implies, and
- the possibilities this creates

2. magical thinking

is the future like the past?

is this a question of causal (laws of nature) or logical necessity (math)?

how do our present futures co-determine the future present (performativity)?

3. the future of the future

- Since the present futures co-determine the future present,
 - predictions basically *enlarge the probability space* we face;
 - they do not reduce but *expand both uncertainty and possibility*.
 - The question is about the *distribution* of the uncertainty and the possibility:
 - *who gets how much of what?*

Hildebrandt (2017)

3. the future of the future

- Esposito, Luhmann, Koselleck (in my words):
 - *the present futures* co-determine *the future present*
 - *there is only one future present, we better be careful*
 - *the more we predict, the less predictable our world*

3. the future of the future

- **double contingency** is core to prediction and interpretation
 - *life builds on affordances, organisms anticipate behaviours of their environment, this relates to **causality** as the nexus of interpretation and prediction*
 - *human beings anticipate how they are being anticipated, grammatical first person assumes that one can decentre by taking an observer's position, **understanding***
 - *interpretation in speech act sense (**anticipation, performativity**)*
 - *prediction as in **performance metric** or part of **performativity of human action***

4. what machine learning does

- Take an ‘algorithm that uses EHR data to predict which emergency department patients are at high risk for stroke.’
 - *It would learn to diagnose stroke by churning through large sets of routinely collected data.*
 - *Critically, all these data are the product of human decisions: a patient’s decision to seek care, a doctor’s decision to order a test, a diagnostician’s decision to call the condition a stroke.*
- Thus, rather than predicting the biologic phenomenon of cerebral ischemia, the algorithm would predict the chain of human decisions leading to the coding of stroke.’

Ziad Obermeyer, M.D., and Thomas H. Lee, M.D.

Lost in Thought — The Limits of the Human Mind and the Future of Medicine, NEJM.org (2017)

4. what machine learning does

On magical thinking about reinforcement learning (**ML echo chambers**):

- “The papers in conferences fall into two categories.
 1. *One is mathematical results showing that yet another slight variation of a technique is optimal under some carefully constrained definition of optimality.*
 2. *A second type of paper takes a well know learning algorithm, and some new problem area, designs the mapping from the problem to a data representation, and show the results of how well that problem area can be learned.”*

Rodney Brooks, Machine Learning Explained (2017)

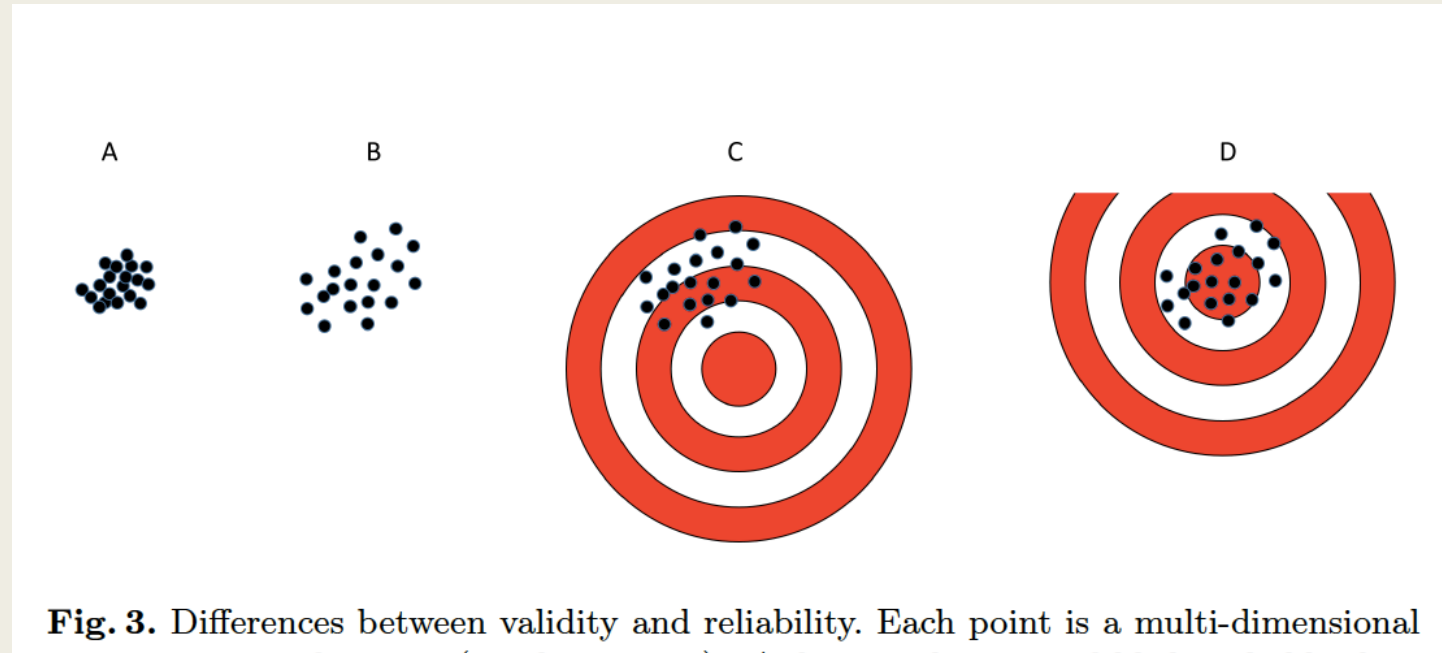
4. what machine learning does

On magical thinking about deep learning:

- **'The bottom line:** Other scientists at the conference said back-propagation still has a core role in AI's future. But Hinton said that, to push materially ahead, entirely new methods will probably have to be invented. "Max Planck said, 'Science progresses one funeral at a time.' The future depends on some graduate student who is deeply suspicious of everything I have said."
- **How it works:** In back propagation, labels or "weights" are used to represent a photo or voice within a brain-like neural layer. The weights are then adjusted and readjusted, layer by layer, until the network can perform an intelligent function with the fewest possible errors.
- But Hinton suggested that, to get to where neural networks are able to become intelligent on their own, what is known as "unsupervised learning," **"I suspect that means getting rid of back-propagation."**
- **"I don't think it's how the brain works," he said. "We clearly don't need all the labeled data."**

Steve Levine, AXIOS.com, September 2015

4. what machine learning does



A giant with feet of clay. On the validity of the data that feed machine learning in medicine, Cabitza (2017)

4. what machine learning does

McNamara Fallacy:

1. The first step is to measure whatever can be easily measured. This is OK as far as it goes.
2. The second step is to disregard that which can't be easily measured or to give it an arbitrary quantitative value. This is artificial and misleading.
3. The third step is to presume that what can't be measured easily really isn't important. This is blindness.
4. The fourth step is to say that what can't be easily measured really doesn't exist. This is suicide.

Daniel Yankelovich "Corporate Priorities: A continuing study of the new demands on business." (1972)

4. what machine learning does

McNamara Fallacy:

1. The first step is to **datafy** whatever can be easily **datafied**. This is OK as far as it goes.
2. The second step is to disregard that which can't be easily **datafied** or to give it an arbitrary quantitative value. This is artificial and misleading.
3. The third step is to presume that what can't be **datafied** easily really isn't important. This is blindness.
4. The fourth step is to say that what can't be easily **datafied** really doesn't exist. This is suicide.

Daniel Yankelovich "Corporate Priorities: A continuing study of the new demands on business." (1972)

5. agonistic machine learning

- Agonistic democracy theory (Chantal Mouffe)
 - *not antagonistic but agonistic*
 - *turning enemies in adversaries*
- Agonistic constructive technology assessment (Arie Rip)
 - *making sure domain experts and those who suffer the consequences are heard*
 - *robust knowledge requires contestation*
- Contestation as the heart of the Rule of Law
 - *without testability no contestability*
 - *restricting the use of OAI (oracular AI)*
 - *purpose specification and the first law of informatics*

5. agonistic machine learning

- the 'infamous' $n=all$ mantra ignores methodological integrity:
 1. *not all ML methods optimize better if more data are available*
 2. *bias in training or input set is more likely than not, and often productive*
 3. *Wolpert's theorem: no free lunch [bias is inevitable]*
 4. *first law of informatics: only train on data collected for the relevant purpose*

5. agonistic machine learning

- Duncan Watts on 'Interpretation and Prediction':

1. Exploratory ML
2. Confirmatory ML

5. agonistic machine learning

■ Duncan Watts on Interpretation and Prediction:

1. 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.*

5. agonistic machine learning

■ 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>).*

6. from magical thinking to rigorous speculation

1. Rethink interpretation and prediction in terms of their interaction
 2. Speculation implies taking seriously the uncertainty that is generated by prediction
 3. Rigorous speculation implies
 - *agonistic machine learning: getting ML out of its echo chambers*
 - *agonistic interaction between ML developers and those subject to its 'jurisdictions'*
-
4. Privacy as the protection of our incomputable artificial nature
 - a. *the right **not to be** linked or profiled*
 - b. *the right **to be** linked, profiled **in a variety of ways***
 - c. *the right **to contest** any particular way of being linked or profiled*



On unpredictability as an effect of prediction:

- M Hildebrandt, 'New Animism in Policing: Re-Animating the Rule of Law?', in *The SAGE Handbook of Global Policing*, ed. Ben Bradford et al. (S.I.: SAGE Publications Ltd, 2016), 406–28.

On novel interaction between humans & machines:

Mireille Hildebrandt,

Smart Technologies and the End(s) of Law.

Novel Entanglements of Law and Technology

(Cheltenham: Edward Elgar, 2015)

