UCL

Fair Trade Algorithms?

A provenance-y provocation

Michael Veale @mikarv

Supporting Algorithm Accountability using Provenance

King's College London

12 July 2018

Department of Science, Technology,
Engineering & Public Policy
University College London

aah! lgorithms

FRANK PASQUALE

THE BLACK BOX SOCIETY

The Secret Algorithms

AE Systems' Taranis drone has autonomous elements, but relies on humans for combat decisions.

Ethics of artificial intelligence

Four leading researchers share their concerns and solutions for reducing societal risks from intelligent machines.

make all targeting dec
Existing AI and r
can provide physical p
motor control, navigat
decision-making and
They just need to be
ple, the technology alr
self-driving cars, toge
like tactical control le
DQN system, could s
and-destroy missions.
Two US Defense
Projects Agency (D
foreshadow planned
Lightweight Autonom
rative Operations in
(CODE). The FLA pr
rotorcraft to manoe
speed in urban areas
CODE aims to devel
mous aerial vehicles
of a strike mission —
engage, assess" in situ
signal-jamming make
a human commander

28 MAY 20

© 2015

rights reserved

Heuristics of the algorithm: Big Data, user interpretation and institutional translation

Göran Bolin and Jonas Andersson Schwarz

Abstract
Intelligence on mass media audiences was founded on representative statistical sam
market departments of media corporations. The techniques for aggregating user
ubiquitous personal media (e.g. laptops, smartphones, credit cards/swipe cards

Synthese (2009) 169:593–613
DOI 10.1007/s11229-008-9438-z

The philosophy of simulation: hot new issues or same old stew?

Roman Frigg · Julian Reiss

Received: 5 December 2007 / Accepted: 17 October 2008 / Published online: 4 December 2008
© Springer Science+Business Media B.V. 2008

Abstract Computer simulations are an exciting tool that plays important roles in many scientific disciplines. This has attracted the attention of a number of philosophers of science. The main tenor in this literature is that computer simulations not only constitute interesting and powerful new *science*, but that they also raise a host of new *philosophical* issues. The protagonists in this debate claim no less than that simulations call into question our philosophical understanding of scientific ontology, the epistemology and semantics of models and theories, and the relation between experimentation and theorising, and submit that simulations demand a fundamentally new philosophy of science in many respects. The aim of this paper is to critically evaluate

The Relevance of Algorithms

Tarleton Gillespie

forthcoming, in *Media Technologies*, ed. Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot. Cambridge, MA: MIT Press.

Algorithms play an increasingly important role in selecting what information is considered most relevant to us, a crucial feature of our participation in public life. Search engines help us navigate massive databases of information, or the entire web. Recommendation algorithms map our

Equality of Opportunity in Supervised Learning

Moritz Hard Eric Price Nathan Srebro

October 11, 2016

Abs

We propose a criterion fo imination nst a speci sitive ate in su- ised learning, where th is to pr me target on av features. ming data about the pre target, and otected are avail- able, we show how to optima st any learne s to remov imination according to our definition. ramework a mproves incentives by sh e cost of

Article

Certifying and removing disparate impact

Michael Feldman Haverford College
Sorelle A. Friedler Haverford College
John Moeller University of Utah
Carlos Scheidegger University of Arizona
*Suresh Venkatasubramanian** University of Utah

Abstract
What does it mean for an algorithm to be biased? In U.S. law, unintentional bias is encoded via *disparate impact*, which occurs when a selection process has widely different outcomes for different

Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining

Sara Hajian
Eurecat
Barcelona, Spain
sara.hajian@eurecat.org

Francesco Bonchi
ISI Foundation
Turin, Italy
francesco.bonchi@isi.it

Carlos Castillo
Eurecat
Barcelona, Spain
chato@acm.org

ABSTRACT
lgorithms and decision making based on Big Data have be- me pervasive in all aspects of our daily lives lives (offline nd online), as they have become essential tools in personal nance, health care, hiring, housing, education, and poli- es. It is therefore of societal and ethical importance to ask hether these algorithms can be discriminative on grounds uch as gender, ethnicity, or health status. It turns out that he answer is positive: for instance, recent studies in the con- ext of online advertising show that ads for high-income jobs re presented to men much more often than to women [5]; nd ads for arrest records are significantly more likely to ow up on searches for distinctively black names [16]. This *algorithmic bias* exists even when there is no discrim- nation intention in the developer of the algorithm. Some-

1. INTRODUCTION
At the beginning of 2014, as an answer to the growing concerns about the role played by data mining algorithms in decision-making, USA President Obama called for a review of big data collecting and analysing practices. The resulting report[1] concluded that "big data technologies can cause societal harms beyond damages to privacy." In particular, it expressed concerns about the possibility that decisions informed by big data could have discriminatory effects, even in the absence of discriminatory intent, further imposing less favorable treatment to already disadvantaged groups.
In the data mining community, the effort to design discrimination-conscious methods has developed two groups of solutions: (1) techniques for *discrimination discovery from databases* [13] and (2) discrimination prevention by

BIG DATA'S DISPARATE IMPACT

*Solon Barocas**
Andrew D. Selbst†

Big data claims to be neutral. It isn't.
Advocates of algorithmic techniques like data mining argue that they eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data mining can inherit the prejudices of prior decision-makers or reflect the widespread biases that persist in society at large. Often, the "patterns" it discovers are simply preexisting societal patterns of inequality and exclusion. Unthinking reliance on data mining can deny members of vulnerable groups full participation in society. Worse still, because the resulting discrimination is almost always an unintentional emergent property of the algorithm's use rather than a conscious choice by its programmers, it can be unusually hard to identify the source of the problem or to explain it to a court.
This Article examines these concerns through the lens of American anti-discrimination law—more particularly, through Title VII's prohibition

Profiling the European Citizen

Mireille Hildebrandt • Serge Gutwirth
Editors

Cross-Disciplinary Perspectives

INTERNET POLICY REVIEW
Journal on internet regulation
Volume 2 | Issue 3

Governance by algorithms

Francesca Musiani
MINES ParisTech , France , francesca.musiani@mines-paristech.fr

Published on 09 Aug 2013 | DOI: 10.14763/2013.3.188

Abstract: Algorithms are increasingly often cited as one of the fundamental shaping devices of our daily, immersed-in-information existence. Their importance is acknowledged, their performance scrutinised in numerous contexts. Yet, a lot of what constitutes 'algorithms' beyond their broad definition as "encoded procedures for transforming input data into a desired output, based on specified calculations" (Gillespie, 2013) is often taken for granted. This article seeks to

Article

Science, Technology, & Human Values
1-25
© The Author(s) 2015
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0162243915606523
sthv.sagepub.com
SAGE

Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness

Mike Ananny[1]

Abstract
rt of un ding the meaning wer of algo means asking at new ds they might ma hical frame and how they

Article

chnology, & Huma
© The Author
Reprints and pe
sagepub.com/journalsPermis
DOI: 10.1177/016224391
sthv.sage

aring ccount-able Witness to the Ethical Algorithmic System

Daniel Neyland[1]

Abstract
This paper explores how accountability might make otherwise ob and inaccessible algorithms available for governance. The potential i

Original Article

Science, Technology, & Human Values
1-16
© The Author(s) 2015
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0162243915589635
sthv.sagepub.com
SAGE

Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics

Kate Crawford[1,2,3]

bstract
is how politi y ap al s.
ns of the
es of agon alism, thi epicts alg public life
stinct scene ler to ask estion, wha f politics de
nstantiate? A s are wo hin highly c d online spa
ublic discou h as Yo and Facebo ere incomp
erspectives Yet also re designe duce clear

Article

Want to be on the top? Algorithmic power and the threat of invisibility Facebook

Taina Bucher
University of Oslo, Norway

Abstract
This article explores the new modalities of a focus on the social networking site Faceb Panopticism – that is, the architectural str

Science, Technology, & Human Values
1-33
© The Author(s) 2015
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0162243915587360
sthv.sagepub.com
SAGE

Article

Algorithms, Governance, and Governmentality: On Governing Academic Writing

Lucas D. Introna[1]

Abstract
Algorithms, or rather algorithmic actions, are seen as problematic because they are inscrutable, automatic, and subsumed in the flow of daily practices. Yet, they are also seen to be playing an important role in organizing opportunities, enacting certain categories, and doing what David Lyon calls "social sorting." Thus, there is a general concern that this increasingly

Governing Algorithms: A Provocatio

Solon Barocas sib237@nyu.edu
Sophie Hood sah7@nyu.edu
Malte mcz3
New York University
March 29, 2013

Paper prepared for the "Governing Algorithms" conference, May 16-17, 2013, at New York University

1. Al ave develo to somewhat of a modern myth. They "compet n 2011), "determine how a billion plus people get where they're g eady written symphonies as moving as those composed by Beet 2012) ered action on the Internet but also contain aspects of f action ology 2011), and "free us from sorting throug irrelevant ow do they do all this, and more? What exac

Technol (2011) 13:251–260
07/s10676-010-9234-7

Is there an ethics of algorithms?

Felicitas Kraemer · Kees van Overveld · Martin Peterson

Published online: 3 July 2010

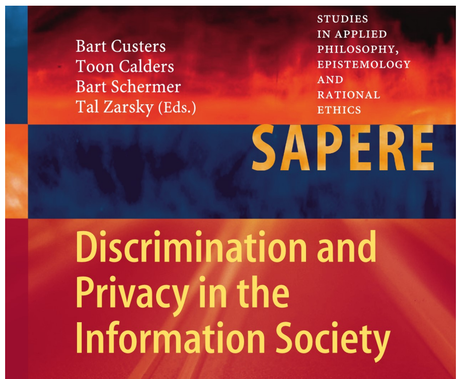
Abstract We argue that some algorithms are value-laden, and that two or more persons who accept different value-judgments may have a rational reason to design such algorithms differently. We exemplify our claim by dis-

detail how to solve a problem. Both c use algorithms for solving a wide However, in this paper we shall be c with algorithms implemented in com

COMPUTING AND ACCOUNTABILITY

Helen Nissenbaum

A teacher stands before her sixth-grade class demanding to know who shot the spitball in her ear. She threatens punishment for the whole class if someone does not step forward. Eyes are cast downward and nervous giggles are suppressed as a boy in the back row slowly raises his hand.
The boy in the back row has answered for his actions. We do not know whether he shot at the teacher intentionally or merely missed his true target, whether he acted alone or under goading from classmates, or even whether the spitball was in protest for an unreasonable action taken by the teacher. While all of these factors are relevant to determining a just response to the boy's action, the boy, in accepting responsibility for his action, has fulfilled the valuable social obligation of accountability.
In an increasingly computerized society, where computing, and its broad application, brings dramatic changes to our way of life, and exposes us to harms and risks, accountability is very important. A community (a society or professional community) that insists on accountability, in which agents are expected to answer for their work, signals esteem for high-quality work, and encourages diligent, responsible practices. Furthermore, where lines of accountability are maintained, they provide the foundations for just punishment as well as compensation for victims. By contrast, the absence of accountability means that no one answers for harms and risks. Insofar as they are regretted, they are seen as unfortunate accidents—consequences of a brave new technology. As with accidents due to natural disasters such as hurricanes and earthquakes, we sympathize with the victims' losses, but do not demand accountability.
This article maintains that accountability is systematically undermined in our computerized society—which, given the value of accountability to society, is a disturbing loss. While this systematic erosion of accountability is not an inevitable consequence of computerization, it is the inevitable consequence of several factors working in unison—an overly narrow conceptual understanding



ALGORITHMIC ACCOUNTABILITY

Journalistic investigation of computational power structures

Nicholas Diakopoulos

Every day automated algorithms make decisions that can amplify the power of businesses and

An A–Z of **Algorithmic Harms**

- **Accountability**: ‘It was the algorithm, not me!’
- **Competition**: Centralisation of power
- **Control**: Using data to make and sell models without permission
- **Dignity**: Kafkaesque ‘computer says no’
- **Discrimination**: Machines re-producing and reinforcing biases
- **Manipulation:** Microtargeting, censorship and democratic effects
- **Oversight**: Intransparency preventing scrutiny
- **Privacy**: Inference of especially sensitive data
- **Safety**: Critical systems (e.g. cars, security)

See generally Lilian Edwards and Michael Veale (2017) Slave to the Algorithm? Why a ‘Right to an Explanation’ is Probably Not the Remedy You are Looking For. 16 Duke Law & Technology Review 18.

An A–Z of Algorithmic Harms

- **General Data Protection Regulation**
  - Automated decisions, art 22
  - Information rights, arts 12–15
  - ‘Supercomplaintant’ system, art 80
  - Lawful basis for processing, art 6
  - Special category data, art 9
- **Equality Act 2010 (UK)**
- **Draft ePrivacy Regulation**
- **Competition Law, Labour Law, Product Liability [...]**

See generally Lilian Edwards and Michael Veale (2017) Slave to the Algorithm? Why a ‘Right to an Explanation’ is Probably Not the Remedy You are Looking For. 16 Duke Law & Technology Review 18.

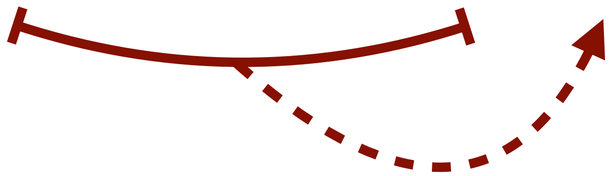

MACHINE
LEARNING
acer
acer
acer

Two core machine learning tasks

Supervised learning

labelled data

age	income	education	repaid loan?
			Y
			N
			Y
			Y

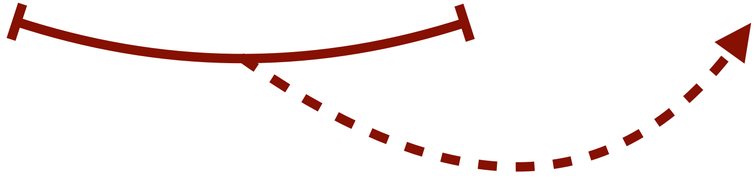


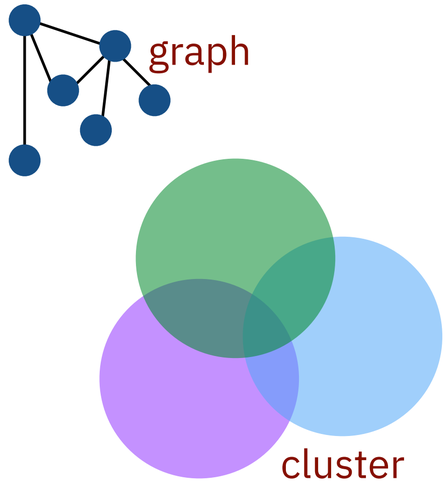
when “correct” values are *possible*

Unsupervised learning

unlabelled data

age	income	education

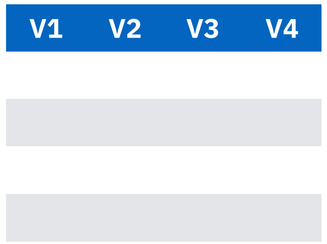




when “correct” values are *less possible*

Reinforcement learning is a further type, but we won’t go into that here as it has very few deployed uses.

Machine learning as a *process*



set aside

test data

evaluate with unseen test data

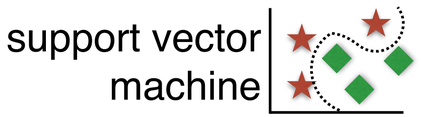
new data

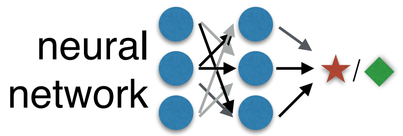
data → preprocess → training data → learning algorithm → trained model → predict

choose model type
tune its ‘hyperparameters’

by ‘feature engineering’ the data to make variables that you think will be predictively useful

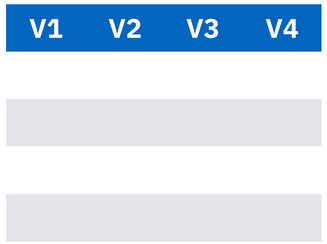
model types





given a model structure, a learning algorithm estimates its parameters

Machine learning as a *process*



set aside

test data

evaluate with unseen test data

new data

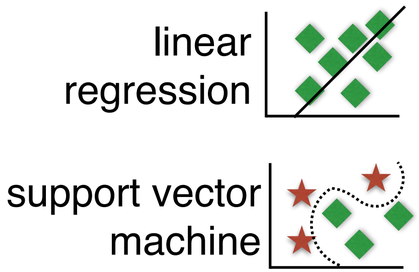
data → preprocess → training data → learning algorithm → trained model → predict

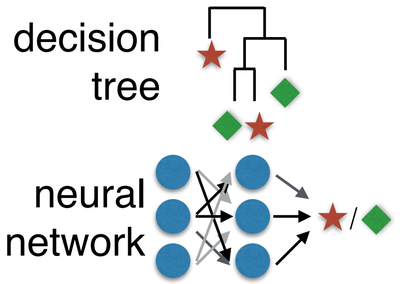
choose model type
tune its ‘hyperparameters’

by ‘feature engineering’ the data to make variables that you think will be predictively useful

Not necessarily from the same jurisdiction!

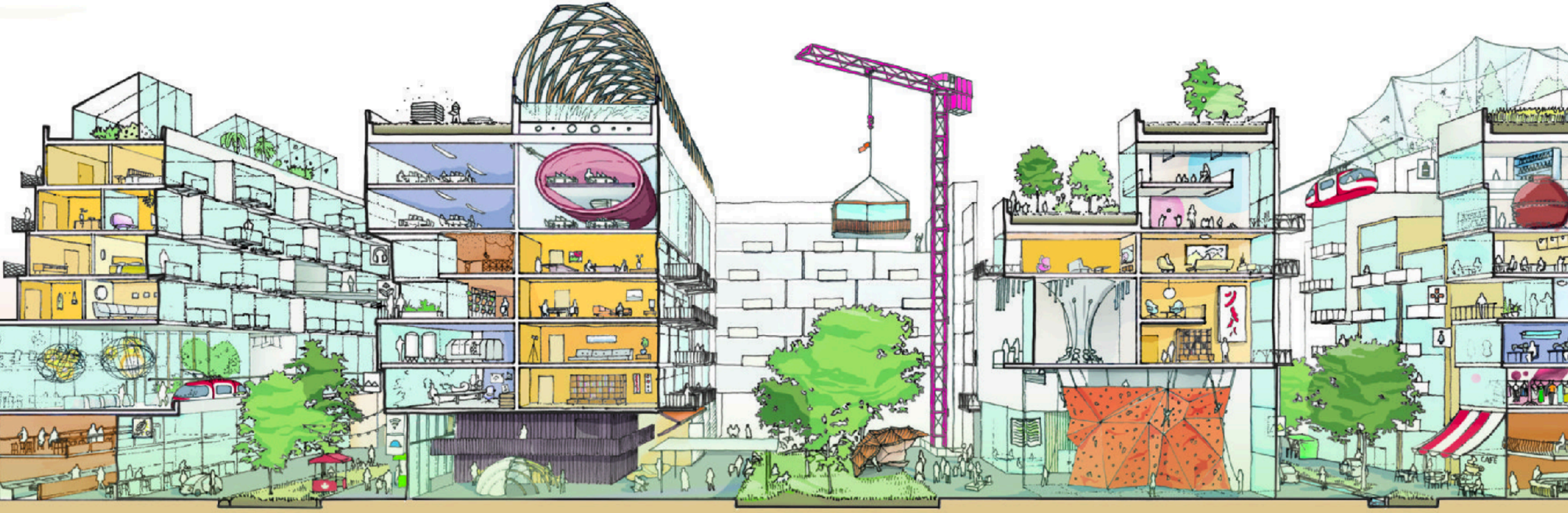
model types





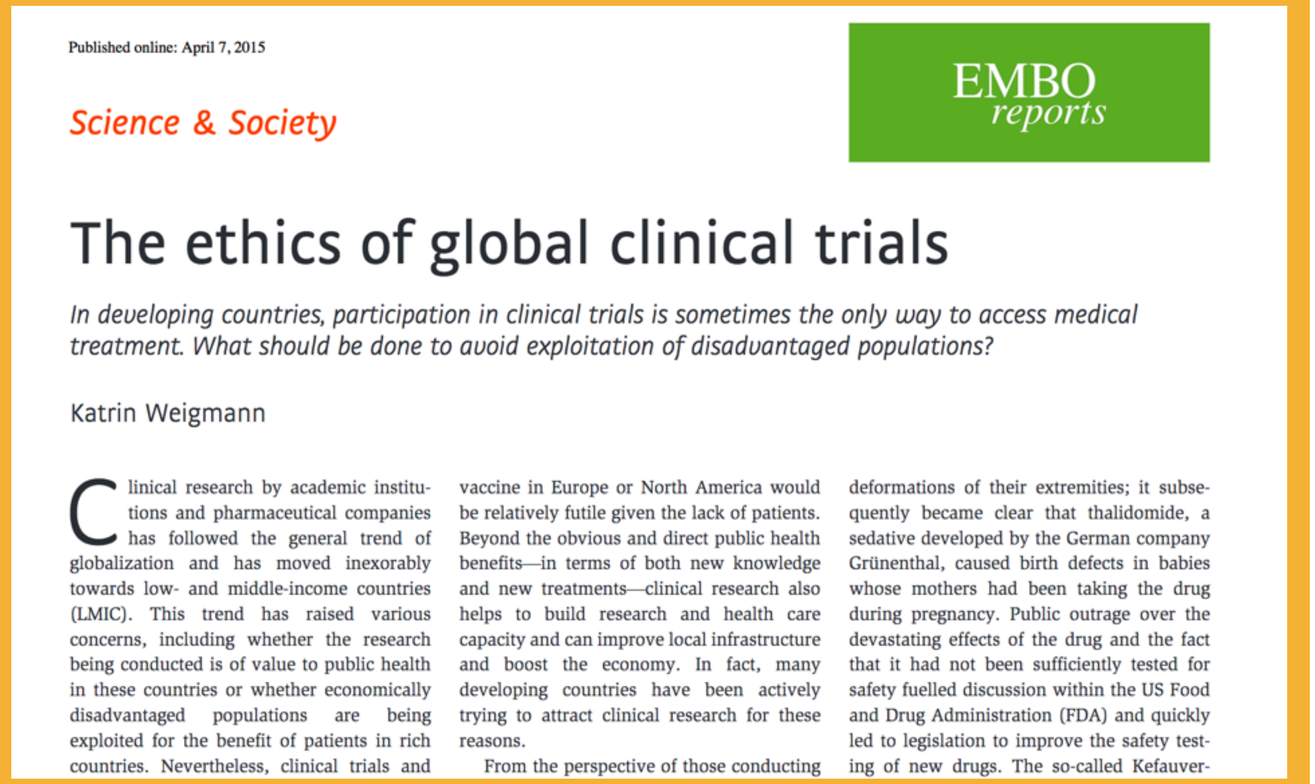
given a model structure, a learning algorithm estimates its parameters

Look to the Future



Google's Guinea-Pig City

Will Toronto turn its residents into Alphabet's experiment? The answer has implications for cities everywhere.



Published online: April 7, 2015

Science & Society

EMBO reports

The ethics of global clinical trials

In developing countries, participation in clinical trials is sometimes the only way to access medical treatment. What should be done to avoid exploitation of disadvantaged populations?

Katrin Weigmann

Clinical research by academic institutions and pharmaceutical companies has followed the general trend of globalization and has moved inexorably towards low- and middle-income countries (LMIC). This trend has raised various concerns, including whether the research being conducted is of value to public health in these countries or whether economically disadvantaged populations are being exploited for the benefit of patients in rich countries. Nevertheless, clinical trials and vaccine in Europe or North America would be relatively futile given the lack of patients. Beyond the obvious and direct public health benefits—in terms of both new knowledge and new treatments—clinical research also helps to build research and health care capacity and can improve local infrastructure and boost the economy. In fact, many developing countries have been actively trying to attract clinical research for these reasons.

From the perspective of those conducting deformations of their extremities; it subsequently became clear that thalidomide, a sedative developed by the German company Grünenthal, caused birth defects in babies whose mothers had been taking the drug during pregnancy. Public outrage over the devastating effects of the drug and the fact that it had not been sufficiently tested for safety fuelled discussion within the US Food and Drug Administration (FDA) and quickly led to legislation to improve the safety testing of new drugs. The so-called Kefauver-

Where knowledge-generation is invasive where is it created?

scientificamerican.com

Most Experimental Drugs are Tested Offshore–Raising Concerns about Data

Rebecca Robbins,STAT

13-17 minutes

The clinical trial for a herpes vaccine flouted just about every norm in the book: American patients were flown in to the Caribbean island of St. Kitts for experimental injections. Local authorities didn't give permission. Nor did the Food and Drug Administration. Nor did a safety panel.

That's why the trial — run by a startup that has since received funding from billionaire investor Peter Thiel — prompted widespread alarm and censure when it was reported last week by Kaiser Health News.

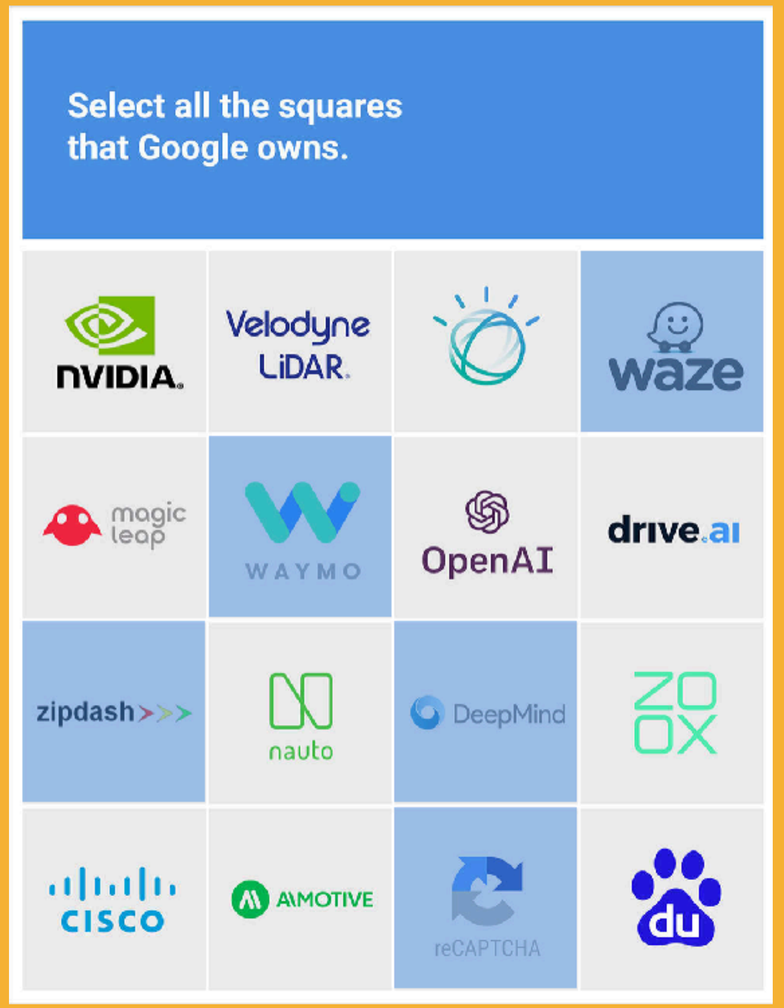
But in some respects, the herpes vaccine trial isn't all that unusual. Nearly all drug makers seeking U.S. approval today rely in part on overseas locations and populations to test their drugs, the result of a decades-long push by industry to try to cut costs and speed recruitment of patients. In fact, a STAT analysis found that 90 percent of new drugs approved this year were tested at least in part outside the U.S. and Canada.

41

TECHNOLOGY NEWS 30 January 2017

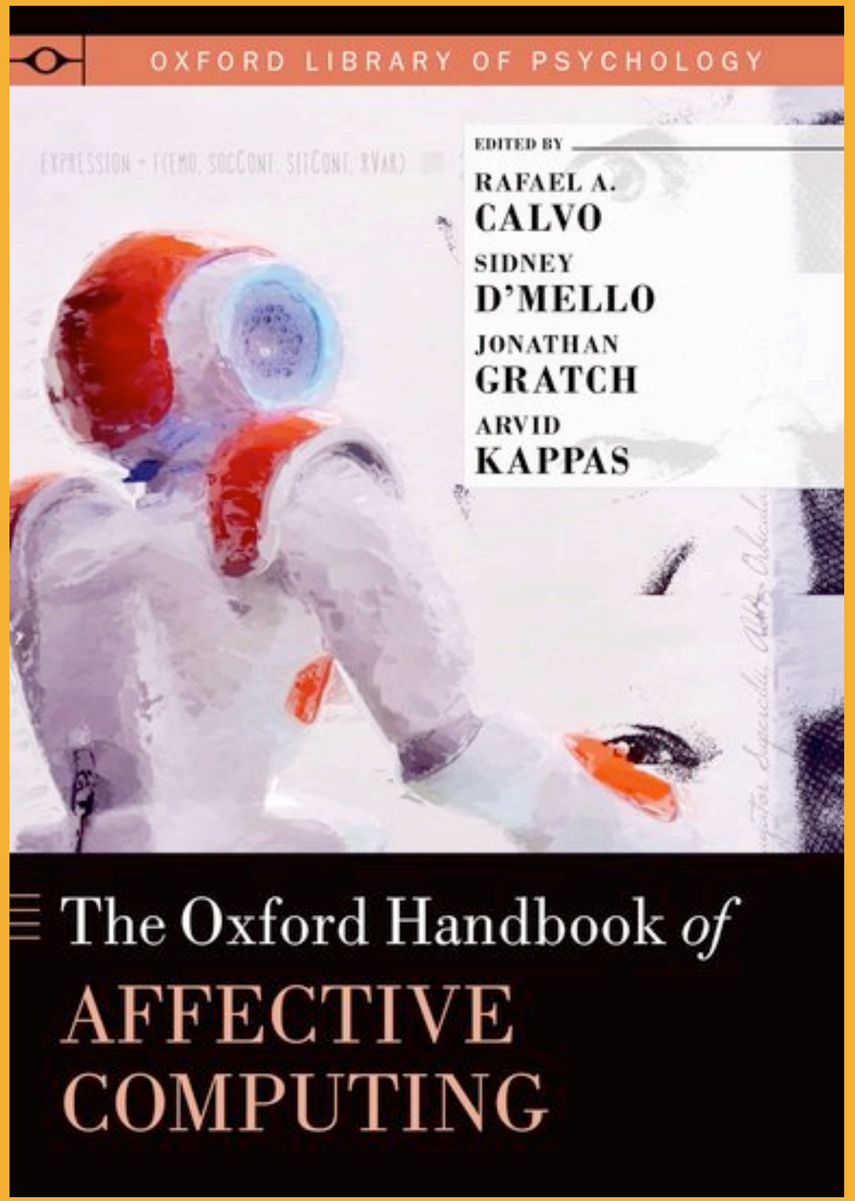
AI tracks your every move and tells your boss if you're slacking

(re)CAPTCHA:
“Completely Automated Public Turing test to tell Computers and Humans Apart”



@mikarv


AFFECTIVE COMPUTING
Rosalind Picard


OXFORD LIBRARY OF PSYCHOLOGY
EDITED BY
RAFAEL A.
CALVO
SIDNEY
D'MELLO
JONATHAN
GRATCH
ARVID
KAPPAS
The Oxford Handbook of
AFFECTIVE
COMPUTING

@mikarv

HELP OUR AI TO LEARN EMPATHY

Which image makes you feel more empathic towards victims of Syria crisis?
Help our AI to learn empathy: select which images are more likely to inspire empathy and help us train our algorithm to get better.

CLICK HERE TO TAKE THE SURVEY

(Survey images might contain sensitive and graphic content.)

Progress: 46/50

HELP OUR AI TO LEARN EMPATHY!

CLICK ON WHICH IMAGE MAKES YOU FEEL MORE EMPATHETIC TOWARDS VICTIMS OF THE CRISIS IN SYRIA.

Progress: 14/50

P OUR AI TO LEARN EMPATHY!

L MORE EMPATHETIC TOWARDS VICTIMS OF THE CRISIS IN SYR

WIRED

ADRIAN CHEN BUSINESS 10.23.14 6:30 AM

THE LABORERS WHO KEEP DICK PICS AND BEHEADINGS OUT OF YOUR FACEBOOK FEED

A contractor at the Manila office of TaskUs, a firm that provides content moderation services to U.S. tech companies. MOISES SAMAN/MAGNUM

@mikarv

Compouding the issue: The rise of **model trading**

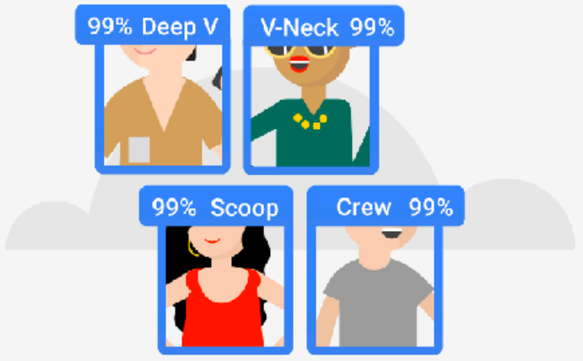
- Data protection has long (and in theory) limited the wholesale trade of personal data: under the GDPR, moreso.
- Companies move to trading models rather than data.
- **Packaged models** vs **API access**
- Supported by specialised hardware and edge computing [e.g. CoreML]

For an extended account, see Michael Veale, Reuben Binns and Lilian Edwards (forthcoming) *Algorithms that Remember: Model Inversion Attacks and Data Protection Law.* Philosophical Transactions of the Royal Society A.

Train Custom Machine Learning Models

Cloud AutoML is a suite of Machine Learning products that enables developers with limited machine learning expertise to train high quality models by leveraging Google's state of the art transfer learning, and Neural Architecture Search technology.

AutoML Vision is the first product to be released. It is a simple, secure and flexible ML service that lets you train custom vision models for your own use cases. Soon, Cloud AutoML will release other services for all other major fields of AI.



Easily train custom vision models

With Cloud AutoML, you can bring your training data to create your own custom vision model with minimum Machine Learning skills required. Start with as little as a few dozen photographic samples and Cloud AutoML will do the rest.

TensorFlow ™ Install Develop API r1.8 Deploy Extend Community Versions TFRC Searc

Develop

GET STARTED PROGRAMMER'S GUIDE TUTORIALS PERFORMANCE MOBILE HUB JAVASCRIPT

Overview

TensorFlow Lite

Introduction to TensorFlow Lite

Developer Guide

Introduction to TensorFlow Lite

TensorFlow Lite is TensorFlow's lightweight solution for mobile and embedded devices. It enables on-device machine learning inference with low latency and a small binary size. TensorFlow Lite also supports hardware acceleration with the Android Neural Networks API.

Developer Discover Design Develop Distri

iOS Overview iPhone X iPad Machine

Build more intelligent apps with machine learning.

Take advantage of Core ML a new foundational machine learning framework used across Apple products, including Siri, Camera, and QuickType. Core ML delivers blazingly fast performance with easy integration of machine learning models enabling you to build apps with intelligent new features using just a few lines of code.

Article 13

Information to be provided where personal data are collected from the data subject

(at point of collection)

Article 14

Information to be provided where personal data have not been obtained from the data subject

(if you can’t contact them, publish online)

Article 15

Right of access by the data subject

(upon email)

	Article 13 **Information to be provided where personal data are collected from the data subject**	*Article 14* **Information to be provided where personal data have not been obtained from the data subject**	*Article 15* **Right of access by the data subject**
Identity and contact details of the data controller	X	X	
Purposes of processing	X	X	X
Legal basis	X	X	
Categories of personal data concerned	X	X	X
Recipients or categories of recipients of the personal data	X	X	X
From which source the personal data originate		X	X
Storage limitation: period and criteria before deletion	X	X	X
Legitimate interests used, where applicable	X	X	X
Meaningful logic about significant automated decision-making	X	X	X
A copy of personal data			X

Data subject

Data controller 1

Data controller 2

Data controller 3

Strong transparency
around provenance

Weak transparency
around provenance

UBER Help

We're here to help

Search rider help

REQUESTING DATA FROM UBER >

To request access to your data, a copy of your data, or a correction to your data, please submit your request in writing.

IF YOU RESIDE IN THE U.S.:
Uber Technologies, Inc.
Attn.: Legal
1455 Market Street, Suite 400
San Francisco, CA 94103

IF YOU RESIDE OUTSIDE THE U.S.:
Uber B.V.
Attn.: Legal
Meester Treublaan 7
1097 DP Amsterdam
Netherlands

On 25th May

~~£10 fee~~

~~No obligation to request by snail mail~~

~~Can receive electronically in “commonly used format”~~

@mikarv

Shattering one-way mirrors – data subject access rights in practice

Jef Ausloos* and Pierre Dewitte*

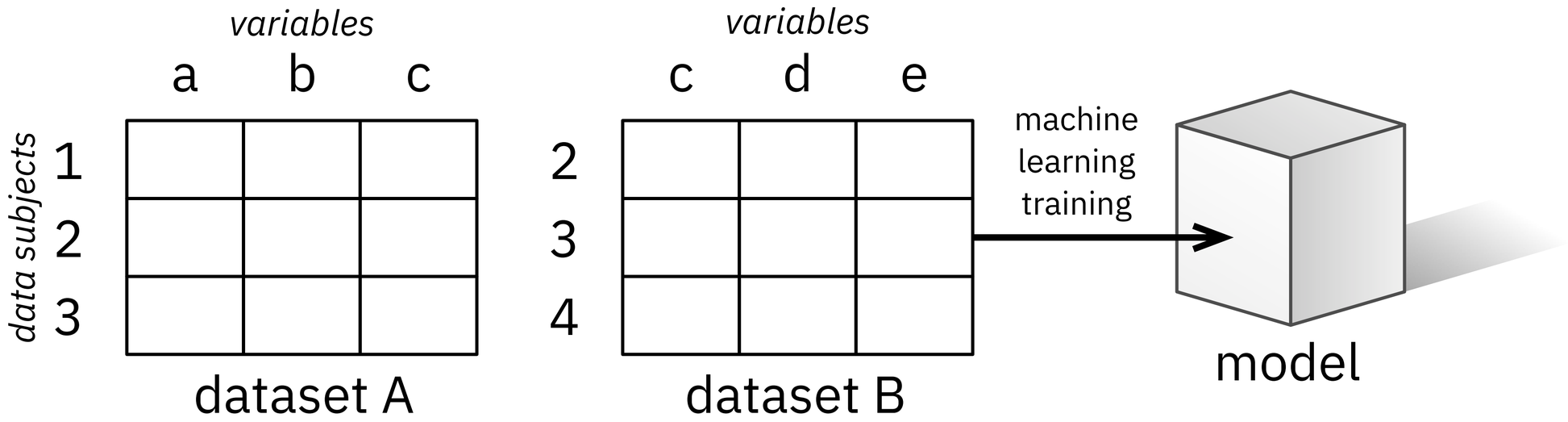
'This type of legislation is the reason we incorporated ***** in the US and not in Belgium. In reality, real users never ask for this type of information. They just delete their account. Our work is to [. . .] in the most trustworthy way. We have now deleted your account and have no data on file anymore, apart from this email in a separate customer support system. We have hereby fulfilled your request. And for all clarity: we treat real users and their privacy with the utmost respect. But we don't spend expensive resources to respond to frivolous requests'.

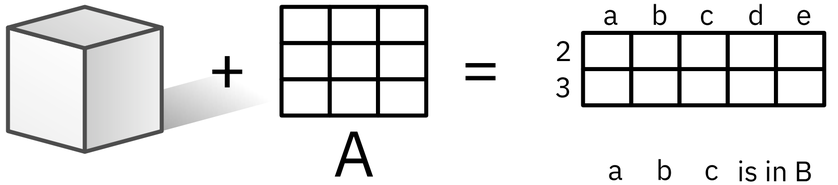
> 'Good morning. ***** being a masculine name, "Dear Sir" will suffice. We really don't have time for this; please look at our privacy policy, all your questions are answered. If you wish to erase your data, you are perfectly entitled to'.

When questioned about the progress of the access request, that same provider replied:

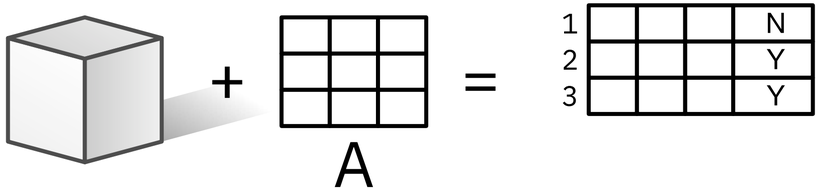
> 'I can't manage to motivate the developers' (translated from French).

Applies to data
What about systems?





: model inversion



: membership inference

Veale, Binns and Edwards (forthcoming) Algorithms that Remember: Model Inversion Attacks and Data Protection Law

@mikarv

Extraterritorial Escapades

- EU data protection law is the strongest, and has severe extraterritorial effects: arguably the Brussels Effect on steroids (albeit enforcement Qs)
- Many countries (e.g. the US) lack an omnibus governance framework for data (although global data privacy laws are more numerous by the month* and there is promise in the modernised Council of Europe Convention 108)
- **Despite this, common types of corporate surveillance and data processing legally permitted in many jurisdictions would be illegal in Europe and some other areas.**

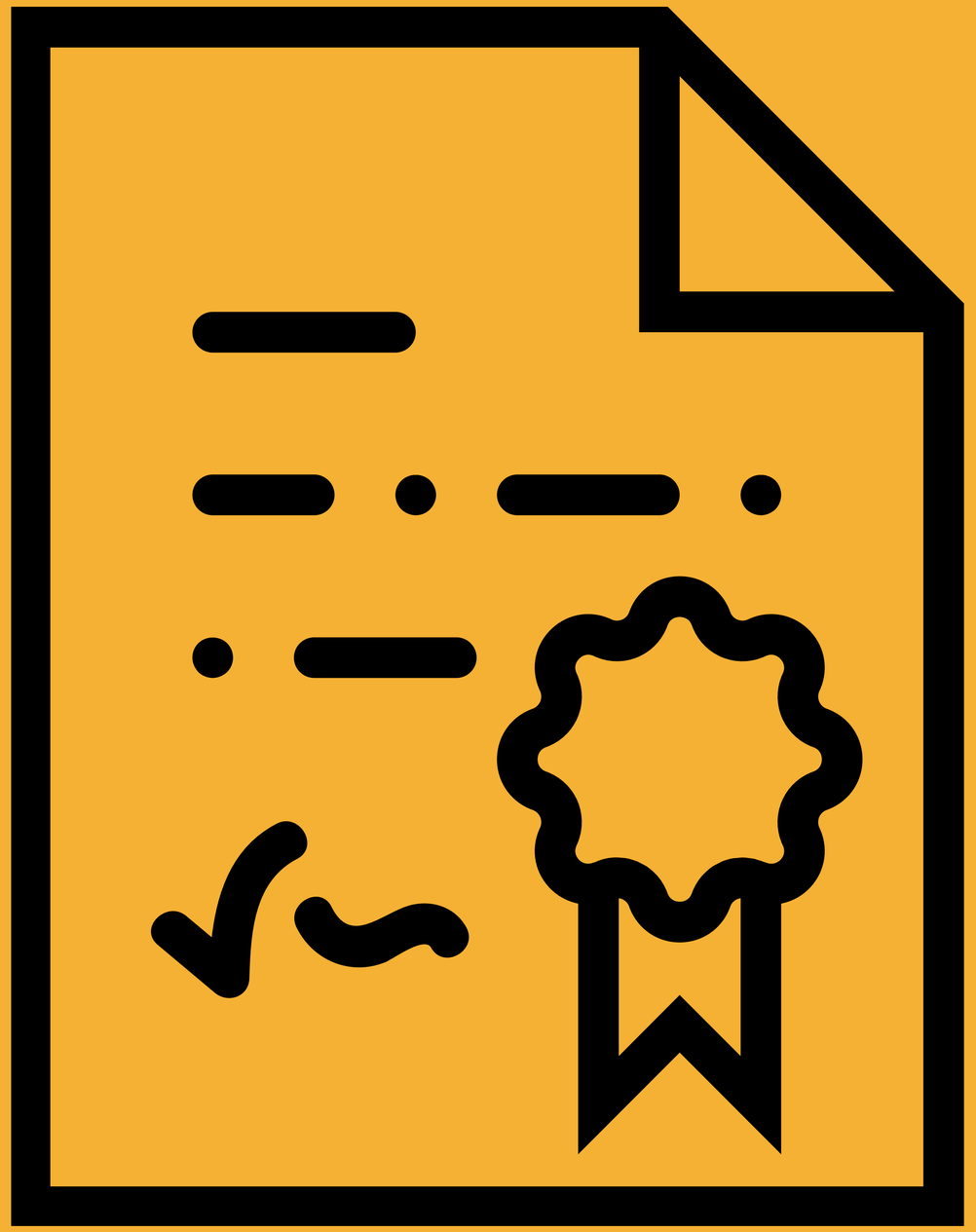
*See the body of work in the area by Graham Greenleaf (available on SSRN).
See also, somewhat surprisingly, the California Consumer Privacy Act 2018.

‘Credence’ attributes

McCluskey, J., 2000. A game theoretic approach to organic foods: An analysis of asymmetric information and policy. Agricultural and Resource Economics Review, 29(1), pp.1–9.

Stop algorithms at the border?

- Can we stop unethical algorithms?
- Other than the challenge of knowing which they are (to follow), legal barriers.
- Repeated WTO cases (e.g. *Dolphin Tuna*, *Shrimp and Turtle*) which see considering credence attributes as a form of protectionism. Not directly appliable but similar concerns with privacy and GATS on services.



Certification for privacy: certify the organisation?

- Turbulent history of privacy certification and seals in international data privacy law: applied to organisations, not datasets.
- Long thought inadequate by scholars, *Safe Harbour* struck down by CJEU.
- Self-certification with audit.

Article 42

Certification

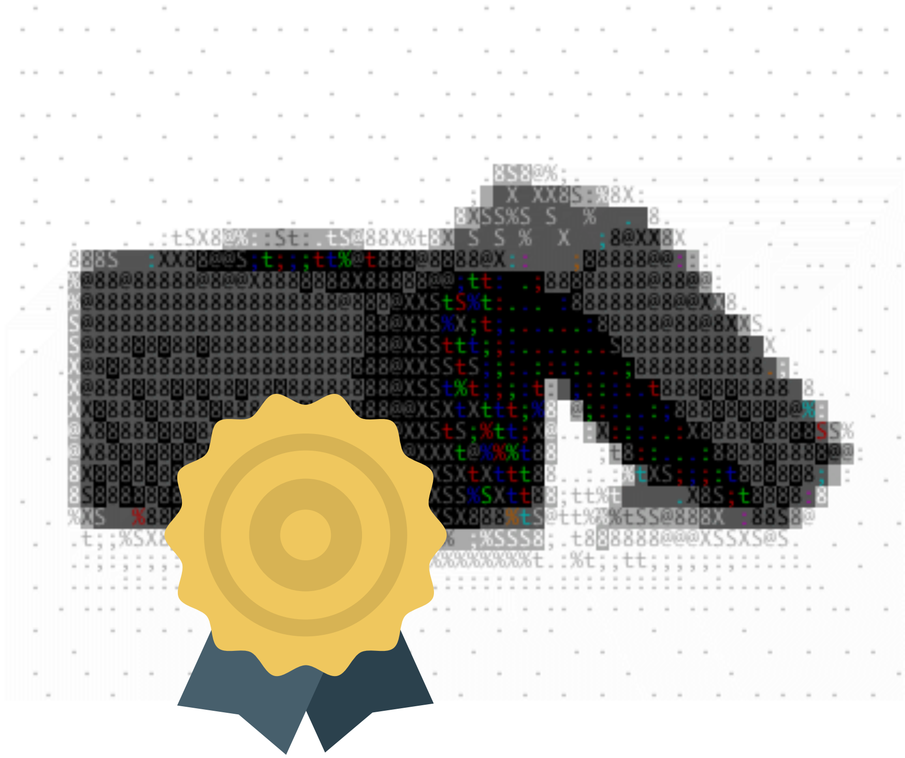
1. The Member States, the supervisory authorities, the Board and the Commission shall encourage, in particular at Union level, the establishment of data protection certification mechanisms and of data protection seals and marks, for the purpose of demonstrating compliance with this Regulation of processing operations by controllers and processors.

[…]

4. A certification pursuant to this Article does not reduce the responsibility of the controller or the processor for compliance with this Regulation […]

Certification for machine learning: certify the model?

- Datasets: can watermark and register them, also difficult to hide the source of high dimensional data. Need a ledger and infrastructure, but if the logic is that ‘only certified data can be used’, possible.
- Models: Harder, as transfer learning and ever changing systems.



Provenance-y tools for models rather than data

- ‘Which data were used to train this model’: possibility to verify using *membership inference attacks*.
- Normal toolkit of e.g. hashing possible for single models, but becomes harder when transfer learning is involved
- A wide array of sophisticated approaches to covertly integrate insights from unethical data into models could be envisaged, such as the use of synthetic data.

Certification for machine learning: certify markets?

- Closer to **chain of custody** certification in commodities.
- Can mandate record keeping, transmission of meta-data with certain models.
- Requirement for e.g. cheaper business insurance.
- Not applying substantive requirements on distant jurisdictions but encouraging transparency, responsible practices, risk management, indirectly.

Governance challenges for ML credence certification

- Certification does not work in a legal vacuum, yet many countries have an effective privacy vacuum, even if on the statute books.
- Who is the consumer? B2B for business risk, but where will pressure come from.
- Technological aspects of ML systems differ from commodities:
  - Can augment, change them, trade them on.
  - Can copy them, access them remotely, deliver as good *or* service
  - Easy to train, and data/labelling can be cheap, while even DPAs cannot enforce existing law.

Concluding remarks

- Certification and standards has been a large topic in governance
- Often focussed on certain social and environmental issues, or technical standards.
- ML brings new challenges from new areas of regulation, but also new technological quirks that may help or hinder these areas.
- Thinking about ‘how certification might work’ in this area highlights the challenges posed by features of the technologies and issues itself to the necessary form of (even potentially) effective governance.

questions?
tweet tweet: @mikarv
papers on algorithmic explanations,
empirical work on public sector ML, and more!
http://michae.lv For early draft of this paper,
e: m.veale@ucl.ac.uk

EPSRC
Engineering and Physical Sciences Research Council

UCL ENGINEERING
Change the world