The Swedish Law and
Informatics Research Institute

Artificial Intelligence and the Law

The Manipulation of Human Behaviour

Stanley Greenstein

Stanley.Greenstein@juridicum.su.se

Predictive Modelling

Stockholm University

Our Humanity Exposed

Predictive Modelling in a Legal Context

Stanley Greenstein

Doctoral Thesis in Law and Information Technology at Stockholm University, Sweden 2017

Länk:
http://su.diva-portal.org/smash/record.jsf?pid=diva2%3A1088890&dswid=-1010

Google:
Stanley Greenstein Diva SU

Predictive modelling

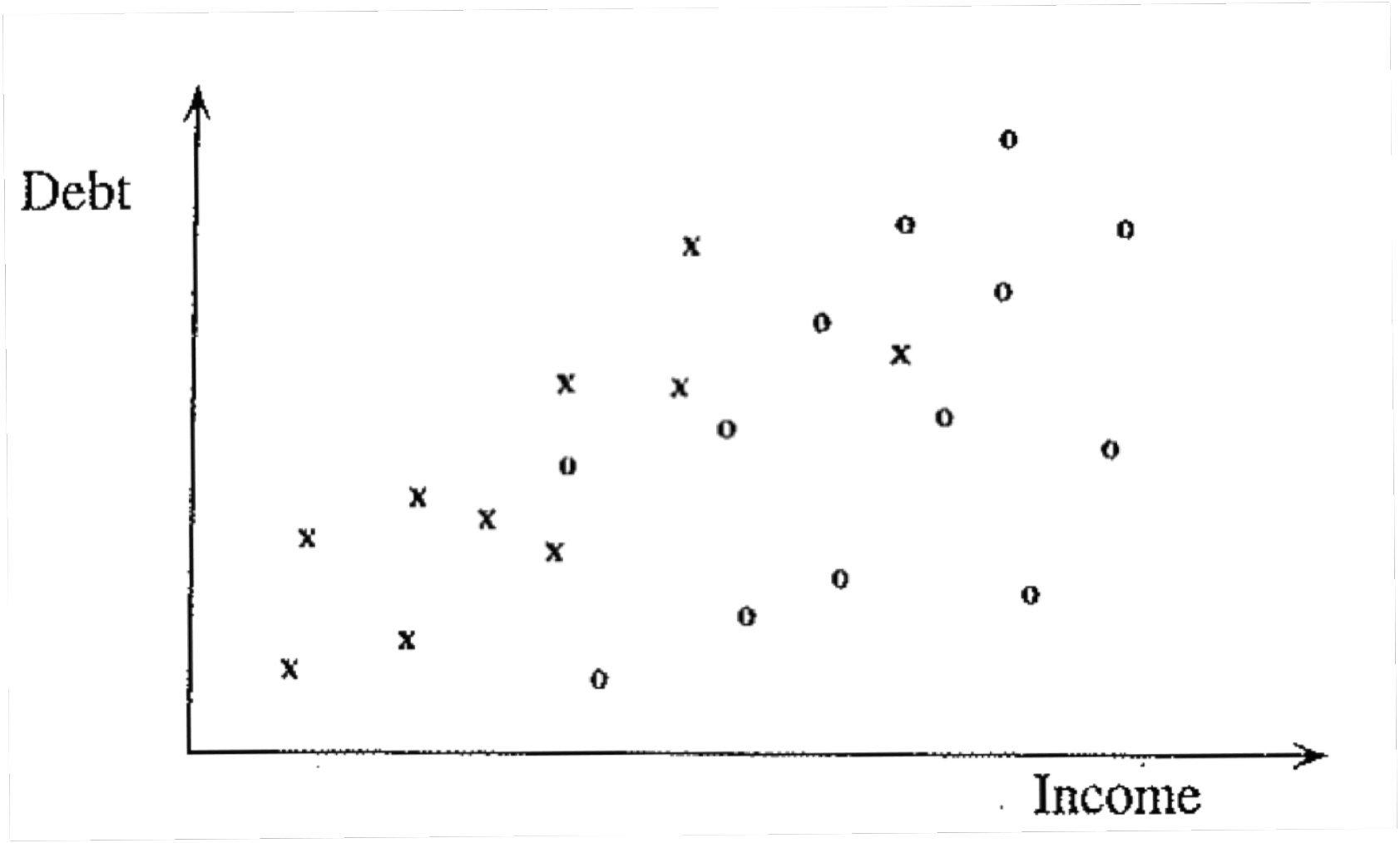
Predictive modelling

- Mathematics, statistics, machine learning, AI
- Algorithms identify patterns in data that are invisible to human beings (correlations = ‘knowledge’)
- This knowledge is applied in models that control/determine our contact with the digital environment
- Models can identify and predict human behaviour
- *That* not *why!*
- Increasingly being used in digital decision systems

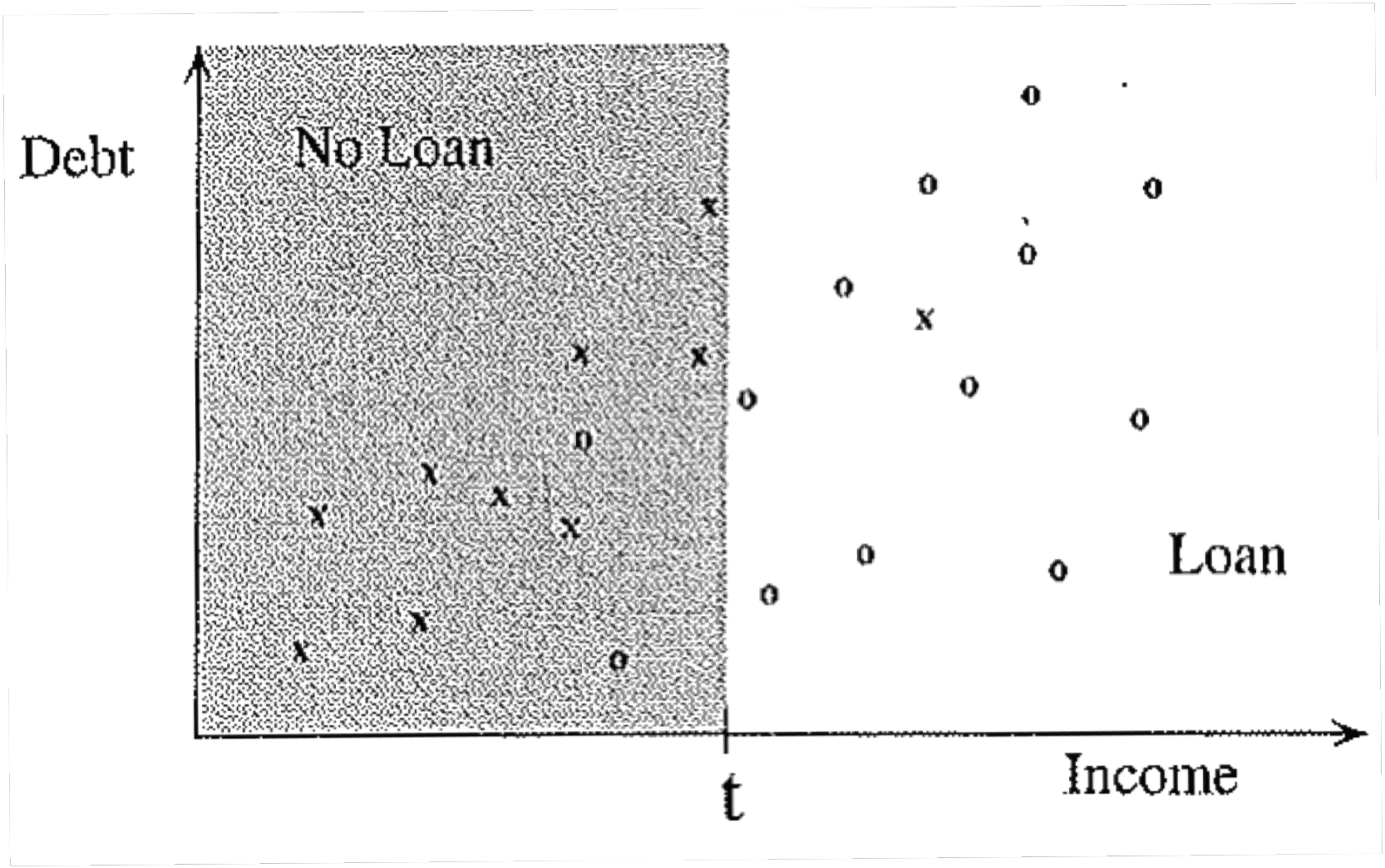
Steven Finlay

‘A predictive model captures the relationships between predictor data and behaviour, and is the output from the predictive analytics process. Once a model has been created, it can be used to make new predictions about people (or other entities) whose behaviour is unknown’.

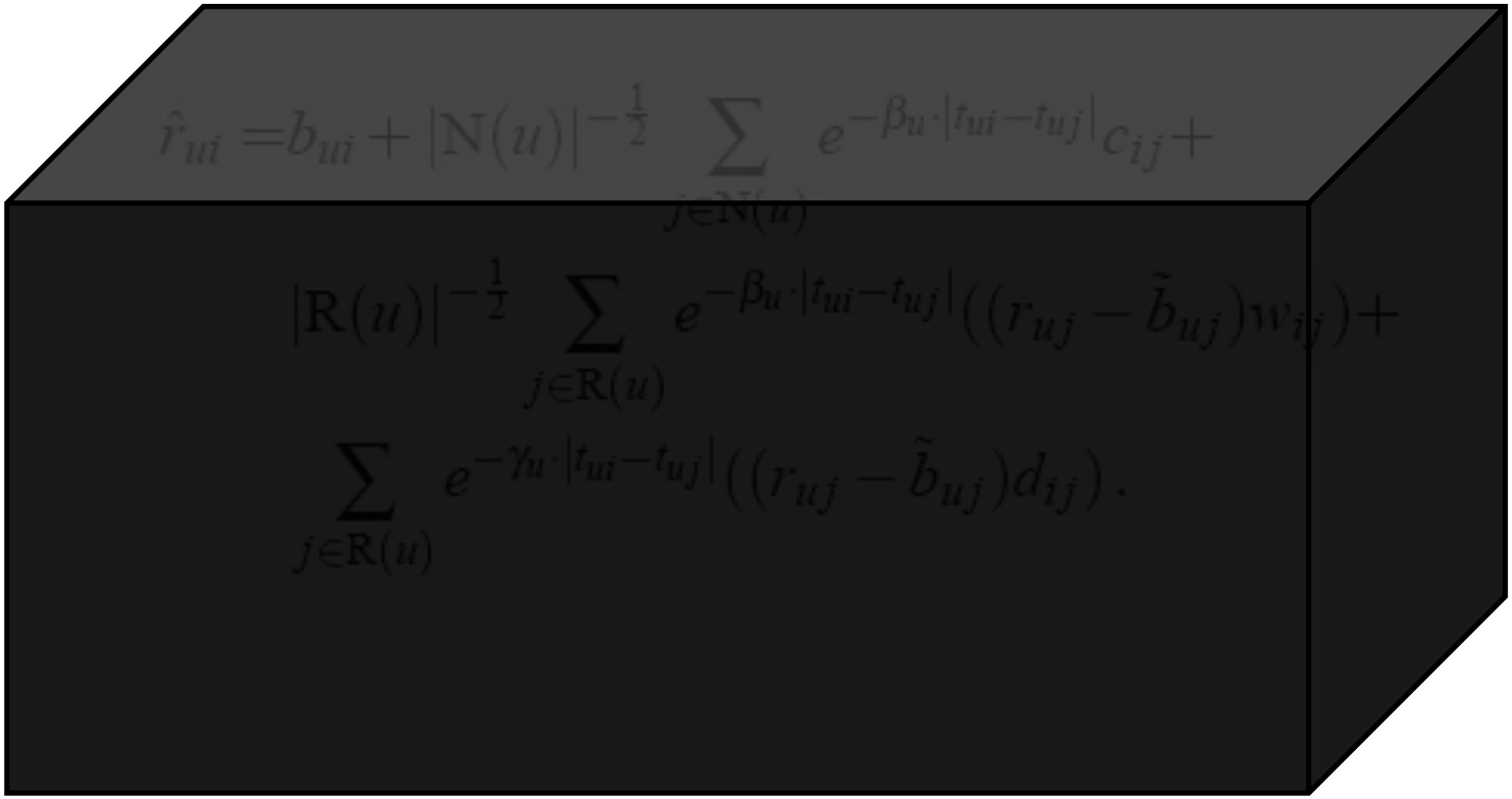
Credit



Credit



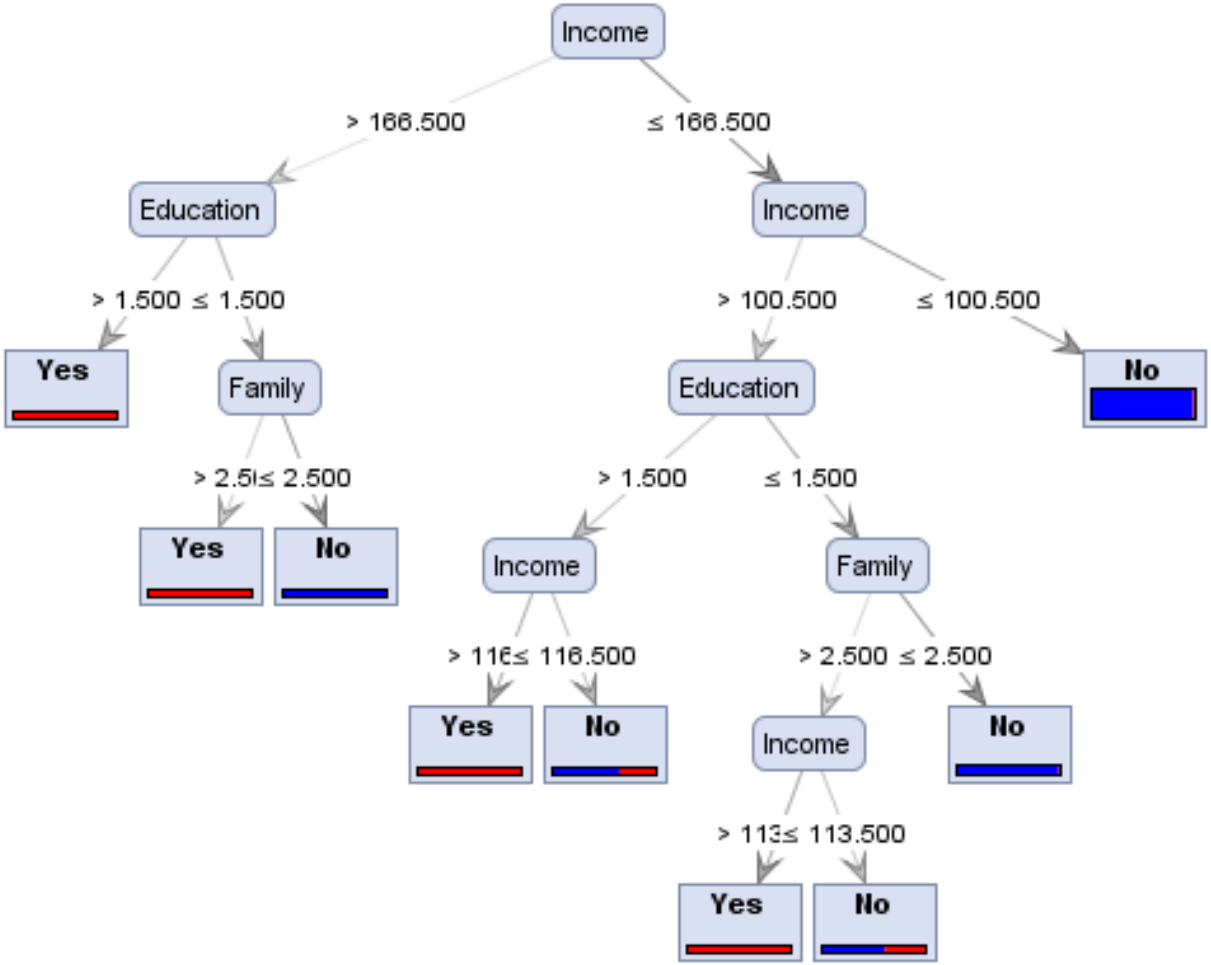
The Algorithm



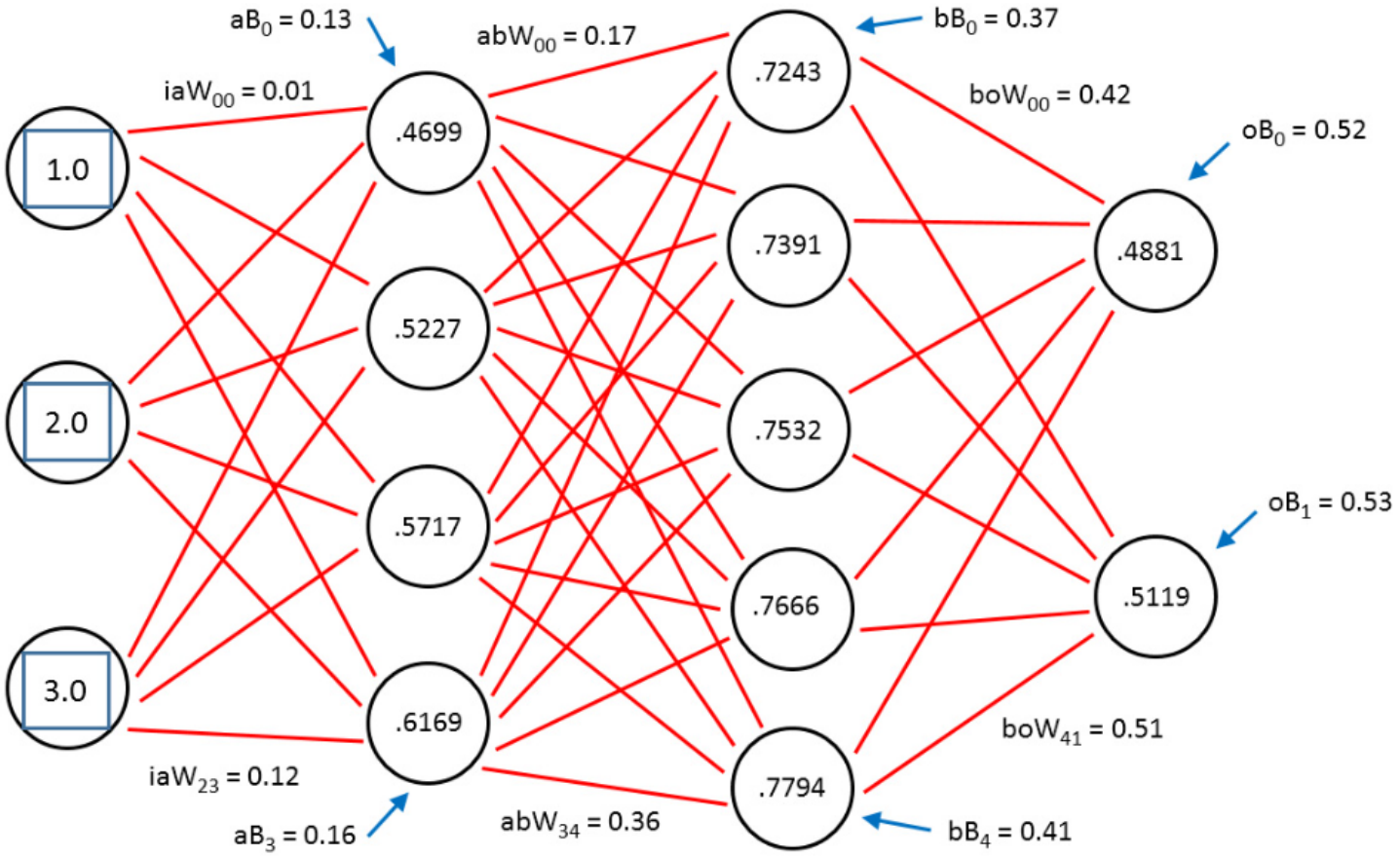
Netflix

Personalization awareness
Top 10 for Xavier
All
Dad
Dad&Mom
Daughter
All
All?
Daughter
Son
Mom
Mom
Diversity

“Decision tree”



“Neural network”



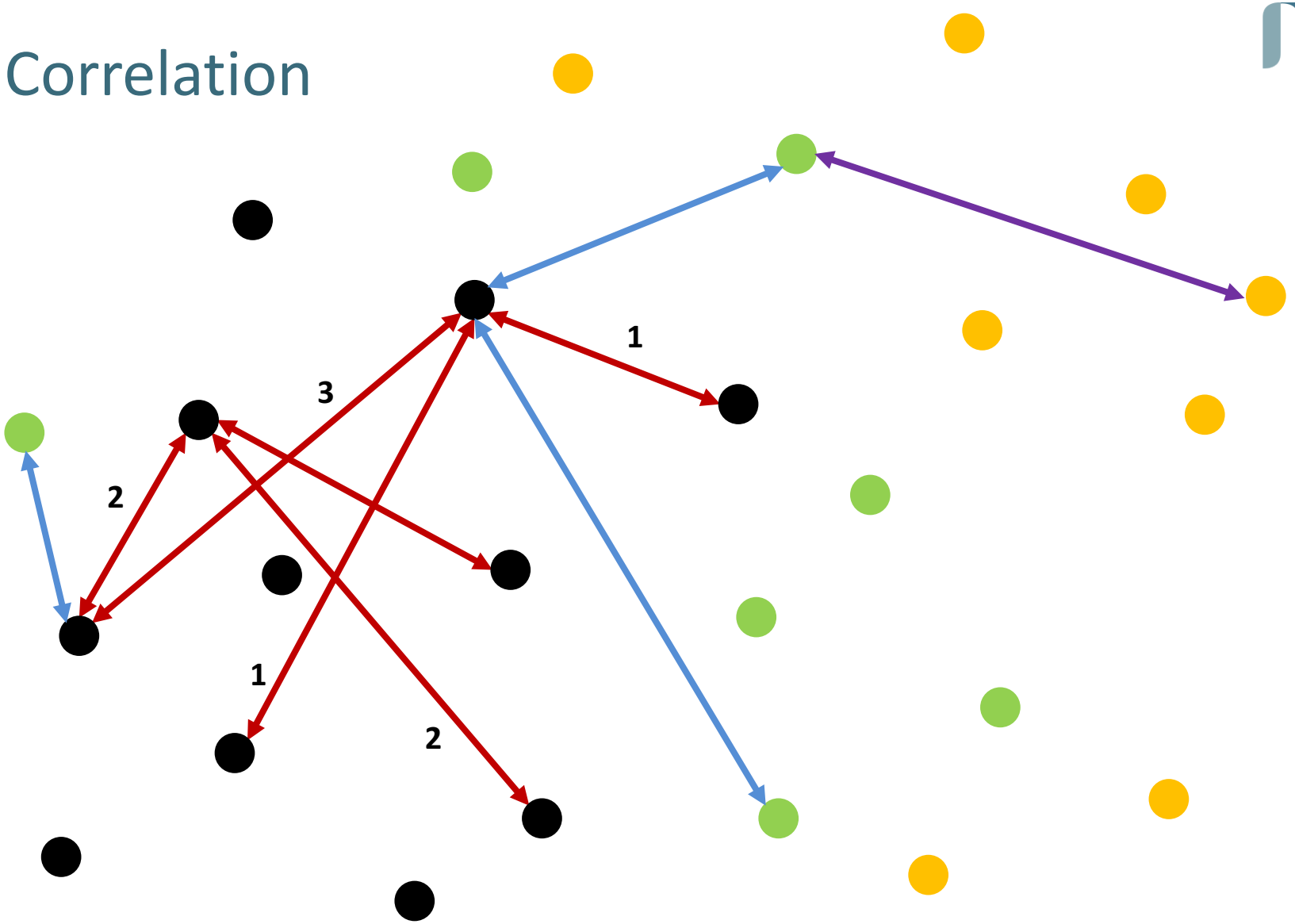
2+2=3,9...and that's OK

Manhole covers i NY

- Problem of exploding manhole covers in NY
- Con Edison given the task of solving the problem
- Inspections based on chance
- 152 000 km wiring & 51 000 manhole covers
- Interested in *that* and not *why*
- *Use the data*
- *Top 10 % on list - 44 % of manholes in really bad shape*
- *106 factors e.g. age of wiring, previous problems*
- Correlation and weight of these correlations

Mayar-Schönberger, Victor & Cukier, Kenneth, *Big Data*, 2013

Correlation



“Flight Risk”

BIG DATA

Applications

Social bots

- Algorithms in the social media
- Twitter: 23 million (8,5% of total users)
- Facebook: 140 million (5,5% - 1,2% of total users)
- Instagram: 27 million (8,2%)

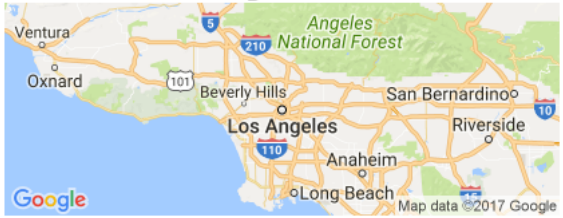
Association for Computing Machinery, Sept. 2017

SF Quakebot

Hi! I'm EQBOT.

I am a robot who tweets about earthquakes as they happen. USGS provides my data. Bill Snitzer programmed me.

Los Angeles Area



Follow @earthquakesLA

@earthquakesLA
LA Area Quakes (All)

Follow @bigQuakesLA

@bigQuakesLA
LA Area Quakes (3.5 Mag +)

San Francisco Area

Follow @earthquakesSF

@earthquakesSF
SF Area Quakes (All)

Follow @bigQuakesSF

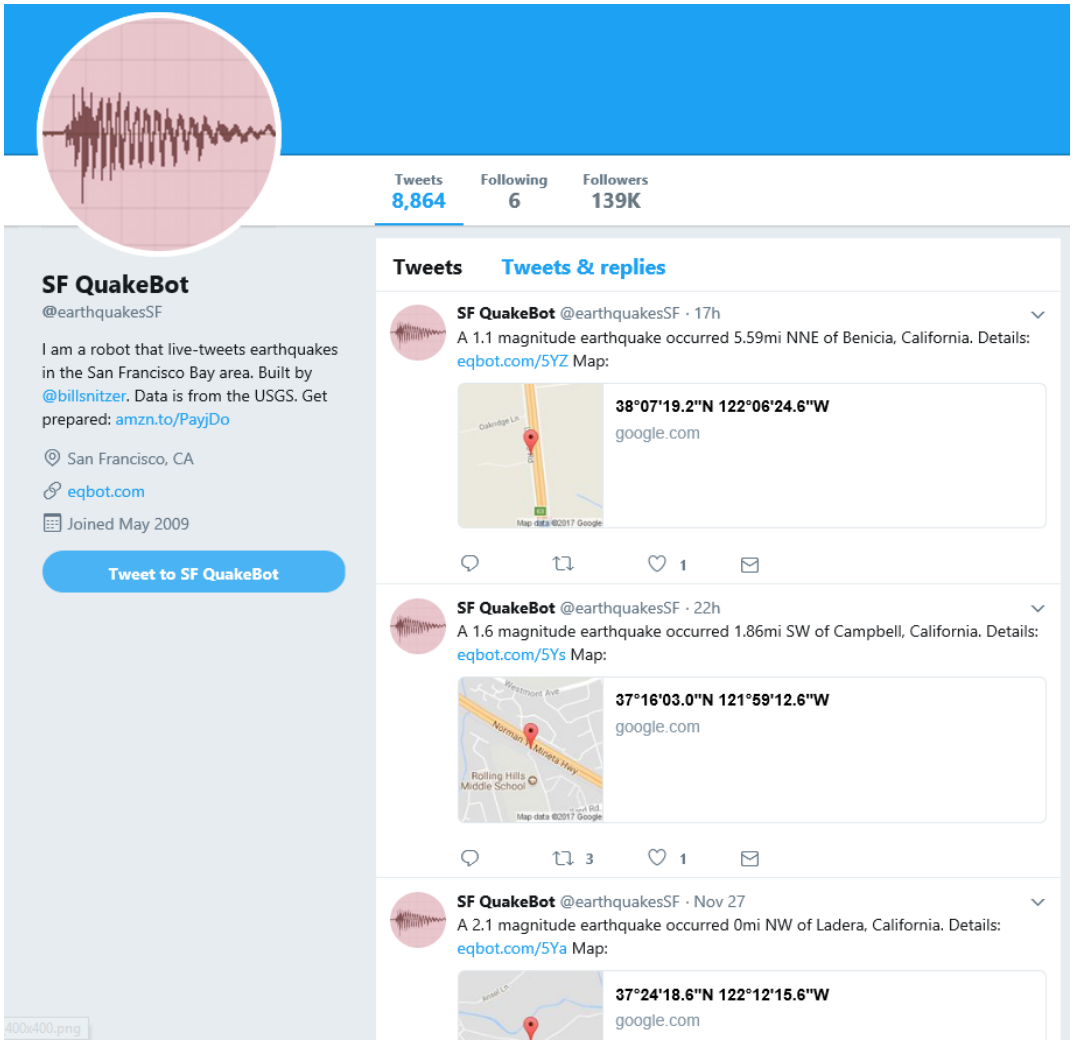
@bigQuakesSF
SF Area Quakes (3.5 Mag +)

Global

Follow @earthquakeBot

@earthquakeBot
Global Quakes (5.0 Mag +)

SF Quakebot


Tweets 8,864
Following 6
Followers 139K
SF QuakeBot
@earthquakesSF
I am a robot that live-tweets earthquakes in the San Francisco Bay area. Built by @billsnitzer. Data is from the USGS. Get prepared: amzn.to/PayjDo
San Francisco, CA
eqbot.com
Joined May 2009
Tweet to SF QuakeBot
Tweets
Tweets & replies
SF QuakeBot @earthquakesSF · 17h
A 1.1 magnitude earthquake occurred 5.59mi NNE of Benicia, California. Details: eqbot.com/5YZ Map:
38°07'19.2"N 122°06'24.6"W
google.com
SF QuakeBot @earthquakesSF · 22h
A 1.6 magnitude earthquake occurred 1.86mi SW of Campbell, California. Details: eqbot.com/5Ys Map:
37°16'03.0"N 121°59'12.6"W
google.com
SF QuakeBot @earthquakesSF · Nov 27
A 2.1 magnitude earthquake occurred 0mi NW of Ladera, California. Details: eqbot.com/5Ya Map:
37°24'18.6"N 122°12'15.6"W
google.com

Spotify

OUR PICKS
LATEST
POPULAR
QUARTZ
OBSESSIONS
GHOST IN THE MACHINE
The magic that makes Spotify's Discover Weekly playlists so damn good
A
Spotify Discover
N.R.
IN
OUT
TDK
High Bias 70µs EQ
SA
90

Instagram

New algorithm-driven Instagram feed rolled out to the dismay of users

Say farewell to chronological ordering of posts – users are now seeing their feed as organised by Instagram's own formula, and they're not happy

Instagram has vowed that the change would ensure the best, most meaningful content would 'be waiting for you when you wake up'. Photograph: Instagram

Is your Instagram feed still ordered oldest-to-newest? Then treasure your straightforward user experience because it is not going to last.

Controversial changes to the photo-sharing platform that were met with widespread outcry when they were first flagged in March are taking effect, with users around the world reporting a new algorithm-driven feed.

Dating

Online dating The Observer

The algorithm method: how internet dating became everyone's route to a perfect love match

Six million Britons are looking for their perfect partner online before Valentine's day on Friday, but their chance of success may depend on clever maths rather than charisma

Six million Britons visit dating sites each month. Photograph: Tom Merton/Getty Images/OJO Images RF

Facebook

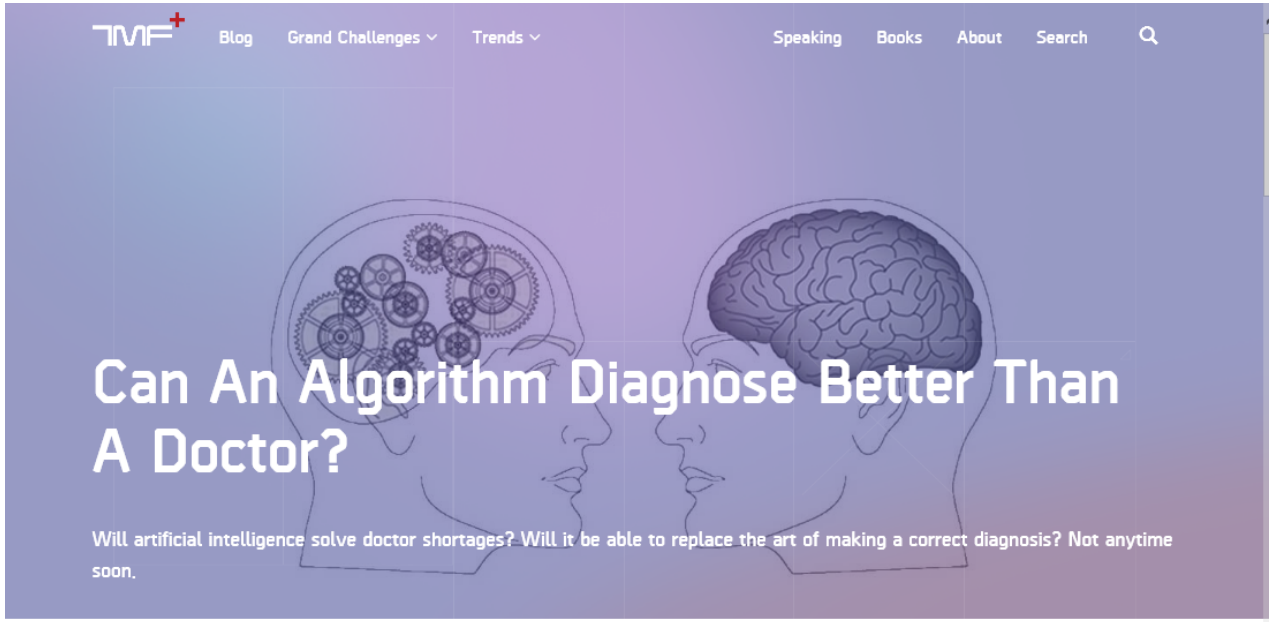
Who Controls Your Facebook Feed

A small team of engineers in Menlo Park. A panel of anonymous power users around the world. And, increasingly, you.

By Will Oremus

Photo illustration by Lisa Larson-Walker. Photo by Tang Ming Tung/Getty Images.

Health and medicine



Many times after my talks, people ask me whether algorithms could theoretically be better at making a diagnosis than doctors. With my doctor's cap on, I must defend the art of medicine. But as a medical futurist I need to tell my honest views. Making a diagnosis is an art. We humans are not engineering products, and therefore measuring a few parameters and tweaking a few knobs will not diagnose and cure our diseases.

Facebook

TECHNOLOGY

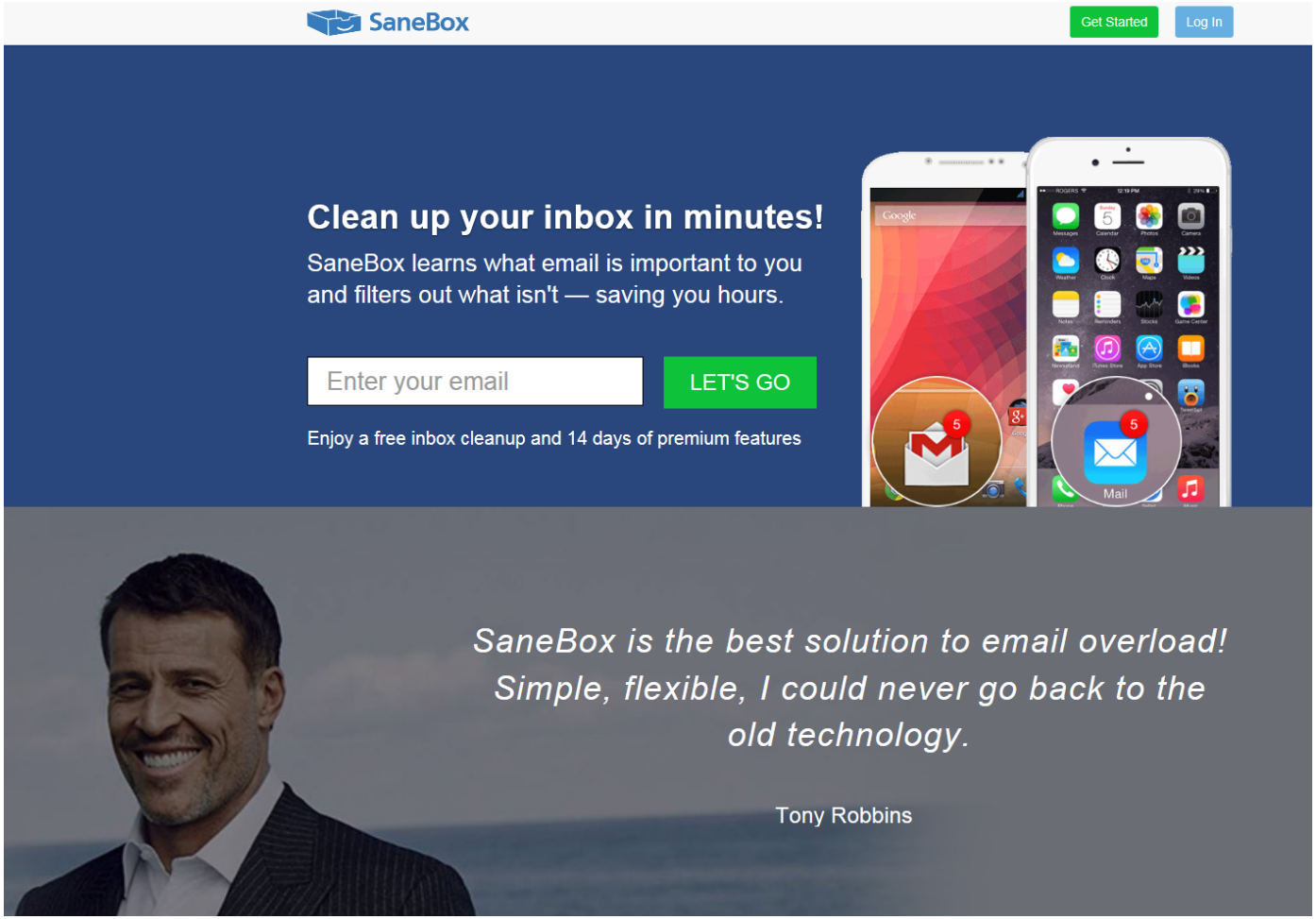
Could a Bank Deny Your Loan Based on Your Facebook Friends?

A recent patent from the company judges your own creditworthiness by your friends.

Facebook

When an individual applies for a loan, the lender examines the credit ratings of members of the individual's social network who are connected to the individual [...]. If the average credit rating of these members is at least a minimum credit score, the lender continues to process the loan application. Otherwise, the loan application is rejected.

Sanebox


SaneBox
Get Started
Log In
Clean up your inbox in minutes!
SaneBox learns what email is important to you and filters out what isn't — saving you hours.
Enter your email
LET'S GO
Enjoy a free inbox cleanup and 14 days of premium features
SaneBox is the best solution to email overload! Simple, flexible, I could never go back to the old technology.
Tony Robbins

China

Så blir du en kinesisk mönstermedborgare

I Kina bidrar varje steg du tar på nätet till ditt betyg. Det avgör sen dina möjligheter för resten av livet. Du måste umgås, shoppa och prata rätt för att slippa bli svartlistad och uthängd.

China

Facial recognition algorithm

The Intercept_

81

Photo: Getty Images

TROUBLING STUDY SAYS ARTIFICIAL INTELLIGENCE CAN PREDICT WHO WILL BE CRIMINALS BASED ON FACIAL FEATURES

Sam Biddle

November 18 2016, 10:28 p.m.

LEIA EM PORTUGUÊS ⟶

THE FIELDS OF artificial intelligence and machine learning are moving so quickly that any notion of ethics is lagging decades behind, or left to works of science fiction. This might explain a new study out of Shanghai Jiao Tong University, which says computers can tell whether you *will be* a criminal based on nothing more than your facial features.

The bankrupt attempt to infer moral qualities from physiology was a popular pursuit for millennia, particularly among those who wanted to justify the supremacy of one racial group over another. But phrenology, which involved studying the cranium to determine someone's character and intelligence, was debunked around the time of the Industrial Revolution, and few outside of the pseudo-scientific fringe would still claim that the shape of your mouth or size of your eyelids might predict whether you'll become a rapist or thief.

Facial recognition algorithm

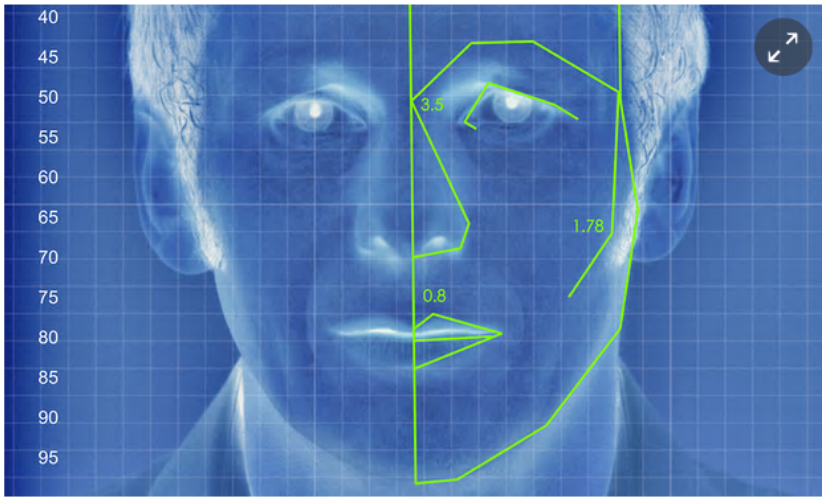
"the upper lip curvature is on average 23.4% larger for criminals than for non-criminals."

"… the distance d between two eye inner corners for criminals is slightly shorter (5.6%) than for non-criminals"

Facial recognition algorithm

New AI can guess whether you're gay or straight from a photograph

An algorithm deduced the sexuality of people on a dating site with up to 91% accuracy, raising tricky ethical questions



An illustrated depiction of facial analysis technology similar to that used in the experiment. Illustration: Alamy

Artificial intelligence can accurately guess whether people are gay or straight based on photos of their faces, according to new research that suggests machines can have significantly better "gaydar" than humans.

The study from Stanford University - which found that a computer algorithm could correctly distinguish between gay and straight men 81% of the time, and 74% for women - has raised questions about the biological origins of sexual orientation, the ethics of facial-detection technology, and the potential for this kind of software to violate people's privacy or be abused for anti-LGBT purposes.

Facial recognition algorithm

Law enforcement

Police embracing tech that predicts crimes

By Heather Kelly, CNN
Updated 1808 GMT (0208 HKT) May 26, 2014

LAPD's crime-fighting data software has roots in the CIA. Is it tracking you? 00:10

Story highlights

Predictive analytic software PredPol anticipates future crime based upon past activity.

The program was adapted from similar software meant to predict earthquake aftershocks.

For something that predicts the future, the software is deceptively simple looking.

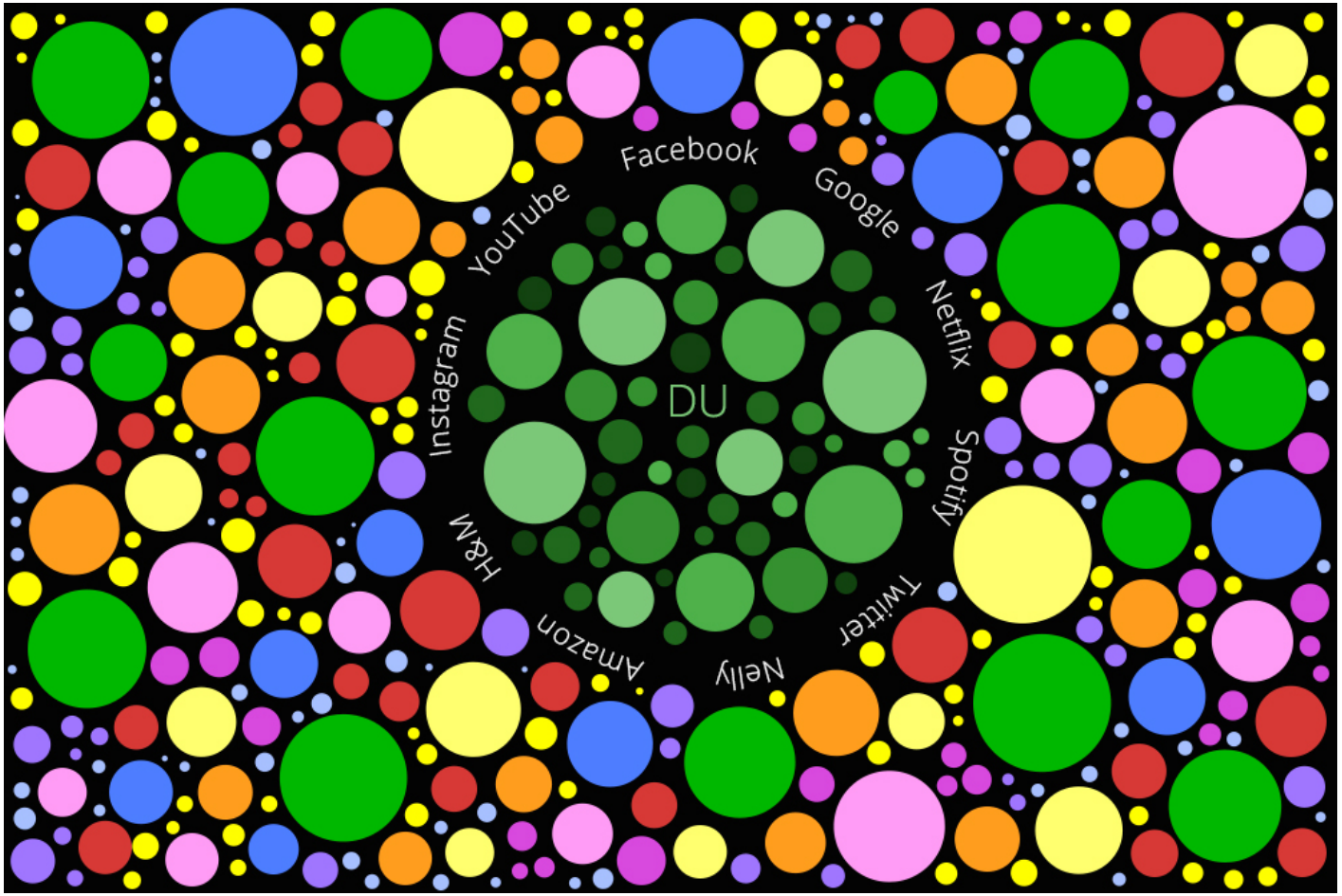
A map of a city is marked up with small red squares, each indicating a 500-by-500-foot zone where crimes are likely to take place next. A heat-map mode shows even more precisely where cars may be stolen, houses

Harms?

Harms

- **Privacy/Data-protection (integritet/dataskydd)**
- **Reputation (rykte)**
- **Autonomy (självbestämmande, autonomi)**
- **Discrimination (diskriminering)**
- **Manipulation (manipulation)**
- **De-individualization***
- **Stereotyping* (stereotyper)**
- **Stigmatisation* (stigmatisering)**
- **Self-censorship* (självcensur)**
- **…**

The “filter bubble”



“The Daily Me”

FINANCIAL TIMES

WORLD US COMPANIES MARKETS OPINION WORK & CAREERS LIFE & ARTS

Climate change + Add to myFT

The rise of the Daily Me threatens democracy

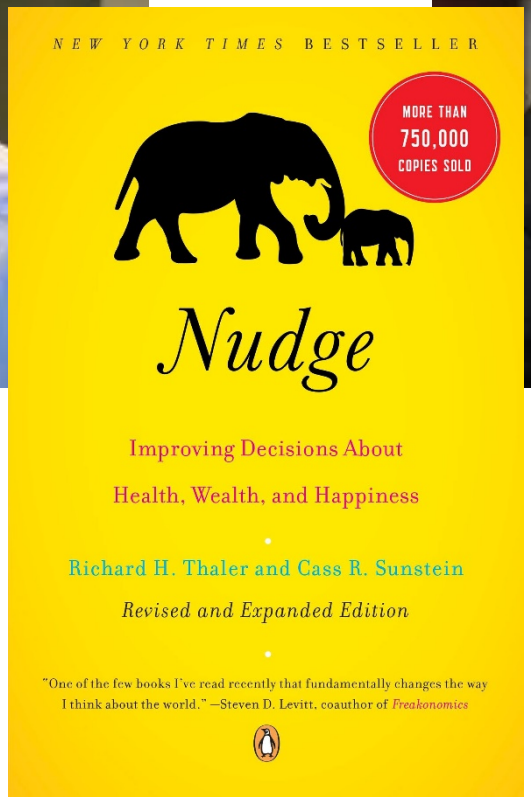
Save to myFT

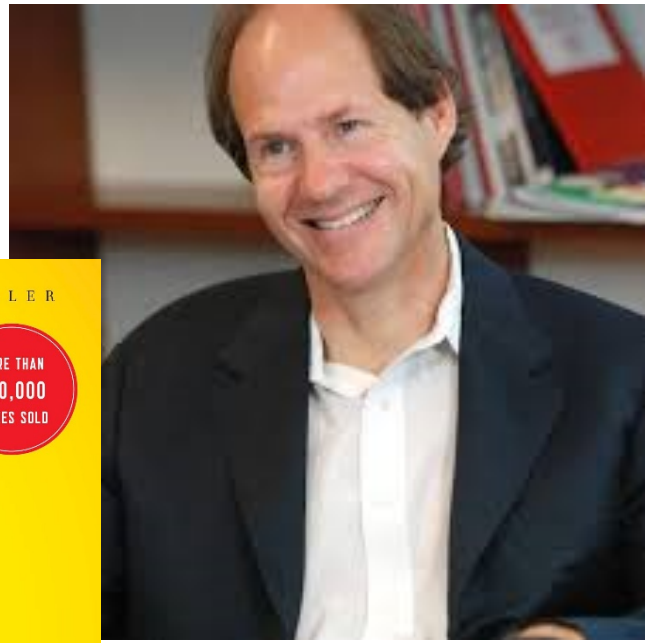
JANUARY 10, 2008 by Cass Sunstein

More than a decade ago the technology specialist, Nicholas Negroponte, prophesied the emergence of the Daily Me – a fully personalised newspaper. It would allow you to include topics that interest you and screen out those that bore or annoy you. If you wanted to focus on Iraq and tennis, or exclude Iran and golf, you could do that.

Many people now use the internet to create something like a Daily Me. This behaviour is reinforced by the rise of social networking forums, collaborative filtering and viral marketing. For politics, the phenomenon is especially important in campaigns. Candidates in the US presidential race can construct information cocoons in which readers are deluged with material that is, in their eyes, politically correct. Supporters of Hillary Clinton construct a Daily Me that includes her campaign’s perspective but offers nothing from Barack Obama, let alone Mitt Romney.

Influencing human behaviour?





Manipulating human behaviour

- 45% of daily human activity is carried out via habit
- Humans think using 2 systems:
  - System 1: intuitive, automatic (rapid, feels instinctive, feels like we are not thinking)
  - System 2: reflective, rational (deliberate, self-conscious)
- Certain decisions – system 1 works but is tainted with a certain bias
  - Short-sighted, optimistic, mental shortcuts, probability
- Manipulation is usually directed against system 1 (or overload system 2)
  - E.g. Subliminal advertising

Manipulating human behaviour

https://www.aftonbladet.se/nyheter/a/ORP51/har-ar-golvet-som-far-manniskor-att-sakta-ner-i-korridoren

50

FEELS LIKE
PIONEER SUSPENSION

FEELS LIKE
PIONEER SUSPENSION

The Law

What does the law say?

- ECHR (Art 8: Right to private life)
- Data Protection (GDPR)

General Data Protection Regulation 2016/679

GDPR

- Complex and vague
- Same data protection principles as the Data Protection Directive 95/46/EC (DPD)
- Art. 22, Recital 63 GDPR
- Consent
- Effective in relation to modern day technology?
- Does not relate to correlations

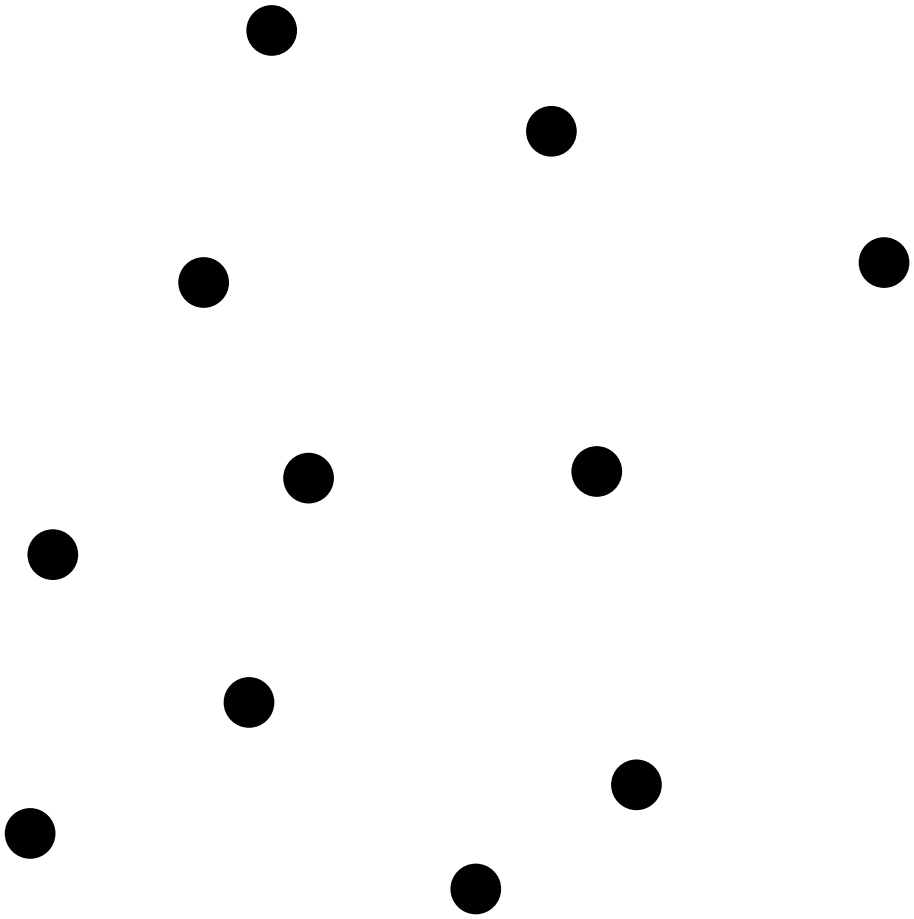
Recital 63, GDPR

‘Every data subject … the logic involved in any automatic personal data processing and, at least when based on profiling, the consequences of such processing … 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. Where the controller processes a large quantity of information concerning the data subject, the controller should be able to request that, before the information is delivered, the data subject specify the information or processing activities to which the request relates.’

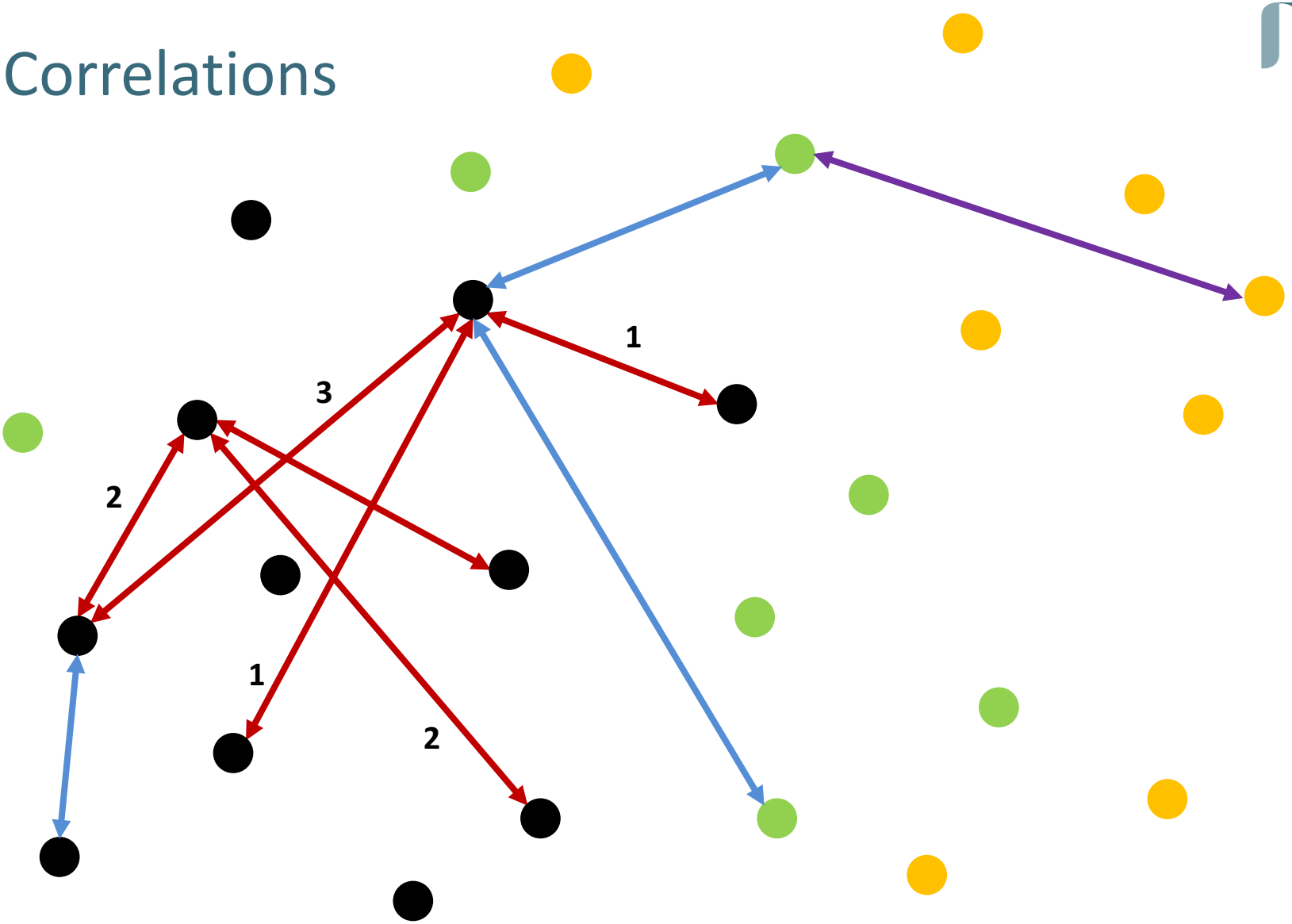
GDPR ... nobel idea, but ...

- Complex and vague
- Same data protection principles as the Directive (DPD)
- Art. 22, Recital 63 GDPR
- Consent
- Effective in relation to modern day technology?
- Does not relate to correlations?

Personal data



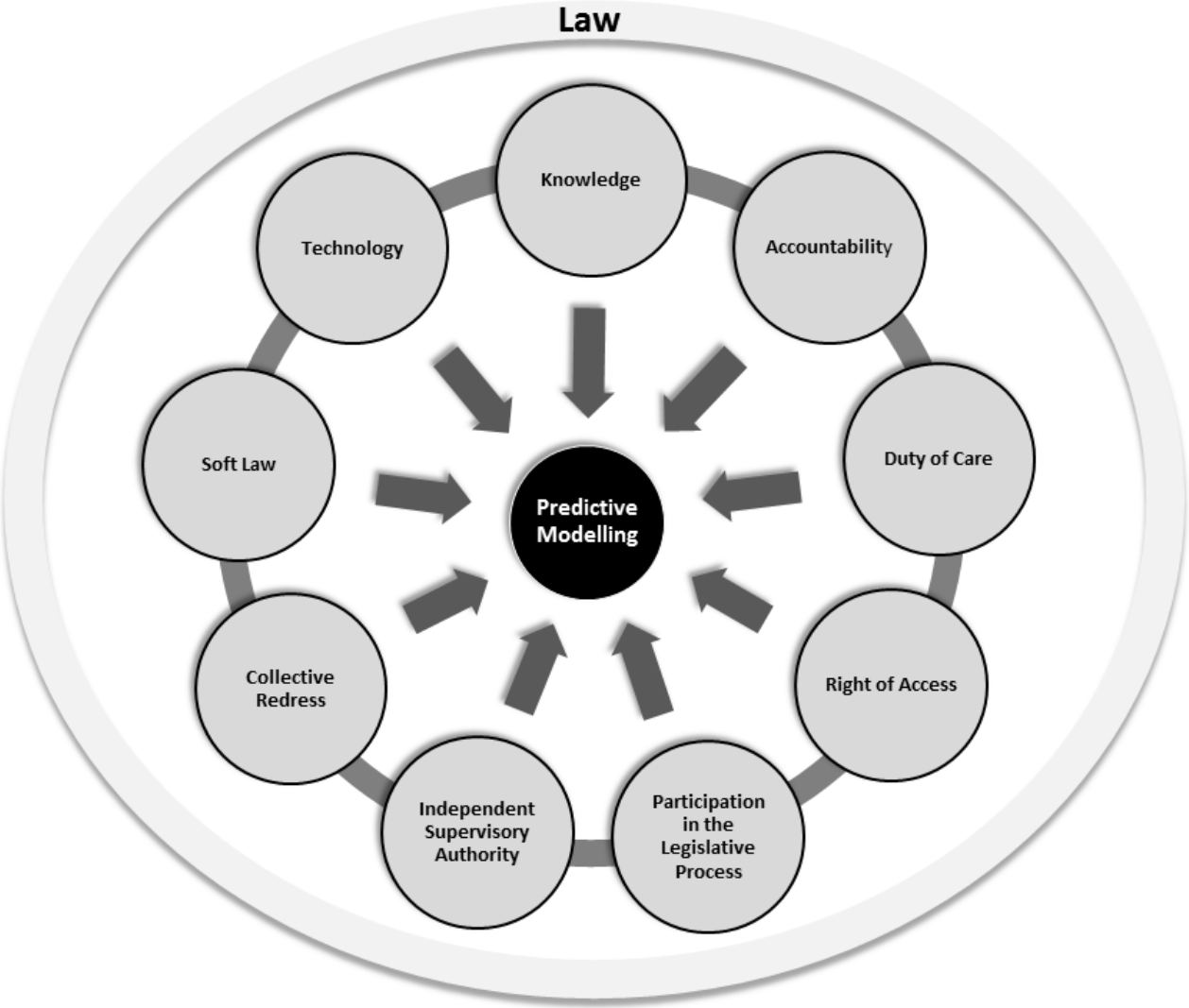
Correlations



‘EMPOWERMENT’

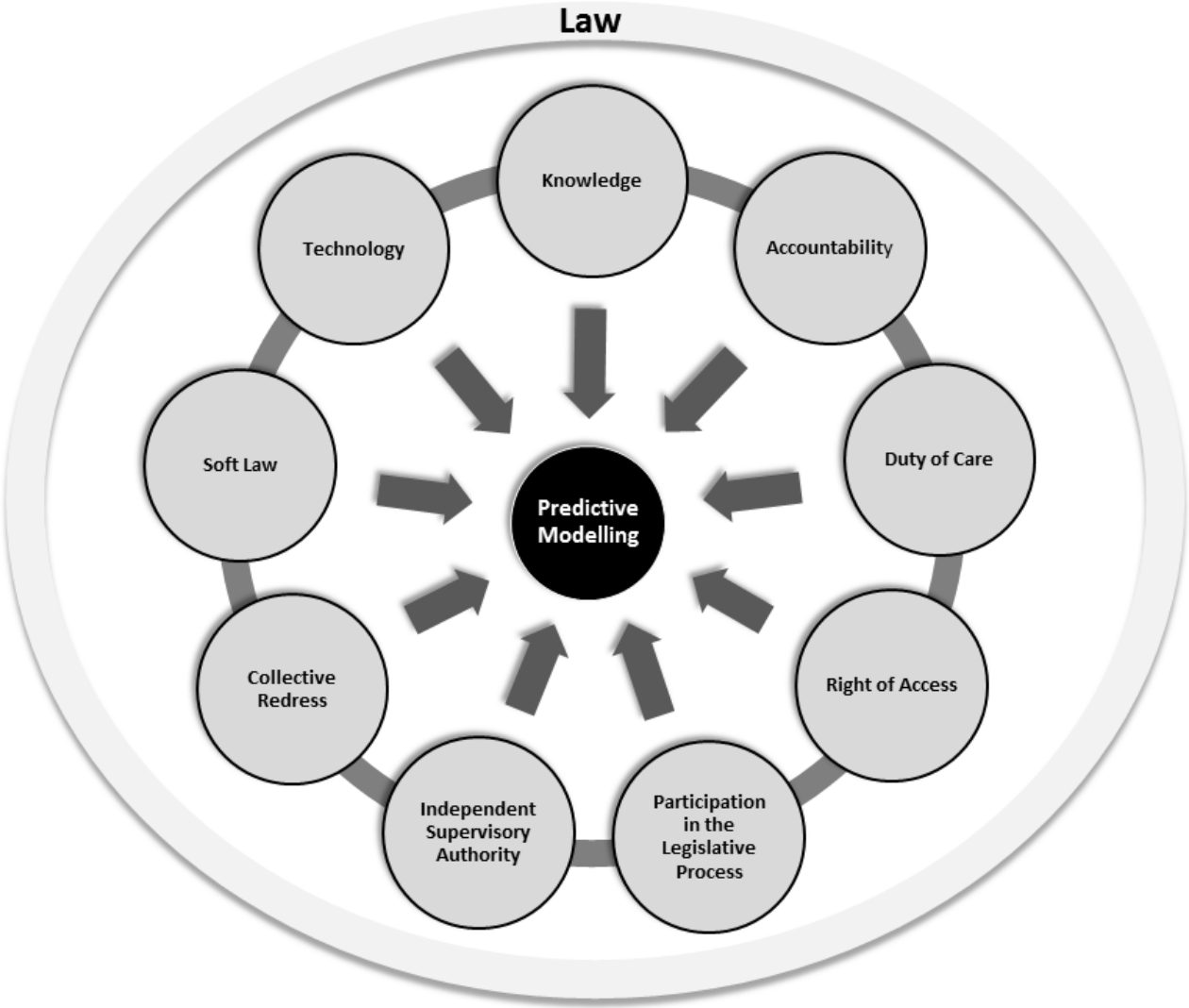
Is the *traditional* law empowering enough?

Empowerment



Soft law

Empowerment



Chatbot

Chatbot lawyer overturns 160,000 parking tickets in London and New York

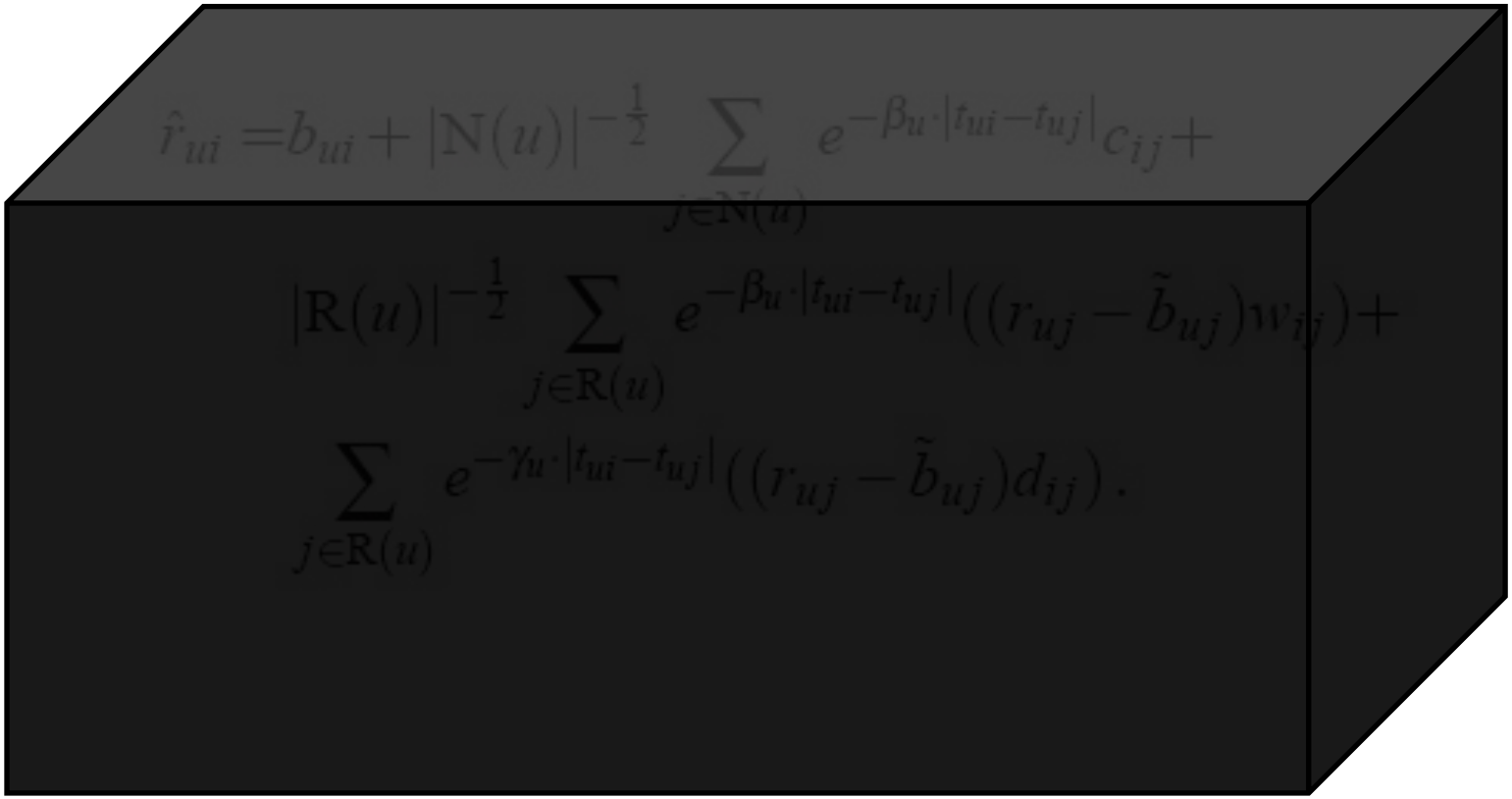
Free service DoNotPay helps appeal over $4m in parking fines in just 21 months, but is just the tip of the legal AI iceberg for its 19-year-old creator

DoNotPay has been helping Londoners and New Yorkers fight illegal parking fines, like this one from Lambeth in London. Photograph: Alamy

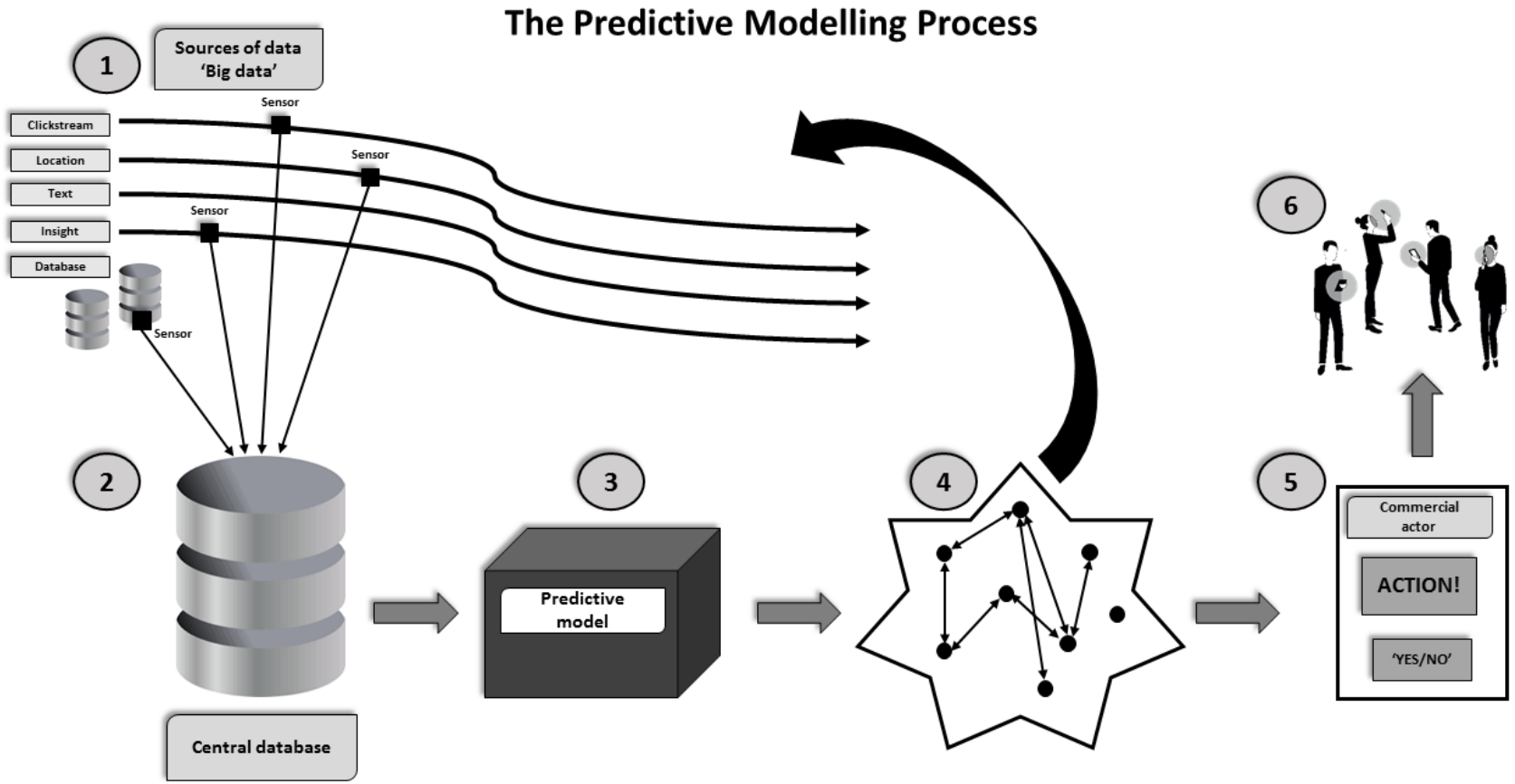
An artificial-intelligence lawyer chatbot has successfully contested 160,000 parking tickets across London and New York for free, showing that chatbots can actually be useful.

Dubbed as "the world's first robot lawyer" by its 19-year-old creator, London-born second-year Stanford University student Joshua Browder, DoNotPay helps users contest parking tickets in an easy to use chat-like interface.

Humanity in the black box society?



“The predictive modelling process”



ECHR

- "In the opinion of the Commission, however, the right to respect for private life does not end there. It comprises also, to a certain degree, the right to establish and to develop relationships with other human beings, especially in the emotional field for the development and fulfilment of one's own personality."

  – *X v. Iceland* (Application no. 6825/74), Commission Decision, 18 May 1976.

ECHR

- "Article 8 also protects a right to personal development, and the right to establish and develop relationships with other human beings and the outside world … Although no previous case has established as such any right to self-determination as being contained in Article 8 of the Convention, the Court considers that the notion of personal autonomy is an important principle underlying the interpretation of its guarantees."

  – *Pretty v. United Kingdom* (2002) 35 EHRR 1, para. 61.

ECHR

- "The Court reiterates that the concept of private life extends to aspects relating to personal identity … Furthermore, private life, in the Court's view, includes a person's physical and psychological integrity; the guarantee provided by Article 8 of the Convention is primarily intended to ensure the development, without outside interference, of the personality of each individual in his relations with other human beings … There is therefore a zone of interaction of a person with others, even in a public context, which may fall within the scope of 'private life'."

  – *Von Hannover v. Germany*, para. 50.

Maipulation and the law

- Border?
- Law assumes that parties are on an equal footing
  - exeptions: consumer law, consent (GDPR recital 43)
- The law recognizes clear forms of manipulation
  - Avtalslagen: 3 kap. 28 § (extreme duress) och 29 § (lesser duress) avtalslagen (1915:218)
  - Brottsbalkan 8:5 (robbery), 9:4 (extorsion), 9:1 (fraud)
- The law recognized manipulation in certains areas
  - Unfair Commercial Practices Directive 2005/29/EC
- Is there a need for the law to recognize more subtle forms of manipulation?

Art. 5(2)(a) and (B)

commercial practice shall be unfair if:

(a) it is contrary to the requirements of professional diligence,

and

(b) it materially distorts or is likely to materially distort the economic behaviour with regard to the product of the average consumer whom it reaches or to whom it is addressed, or of the average member of the group when a commercial practice is directed to a particular group of consumers.

Art. 2(e)

‘to materially distort the economic behaviour of consumers’ means using a commercial practice to appreciably impair the consumer’s ability to make an informed decision, thereby causing the consumer to take a transactional decision that he would not have taken otherwise;’