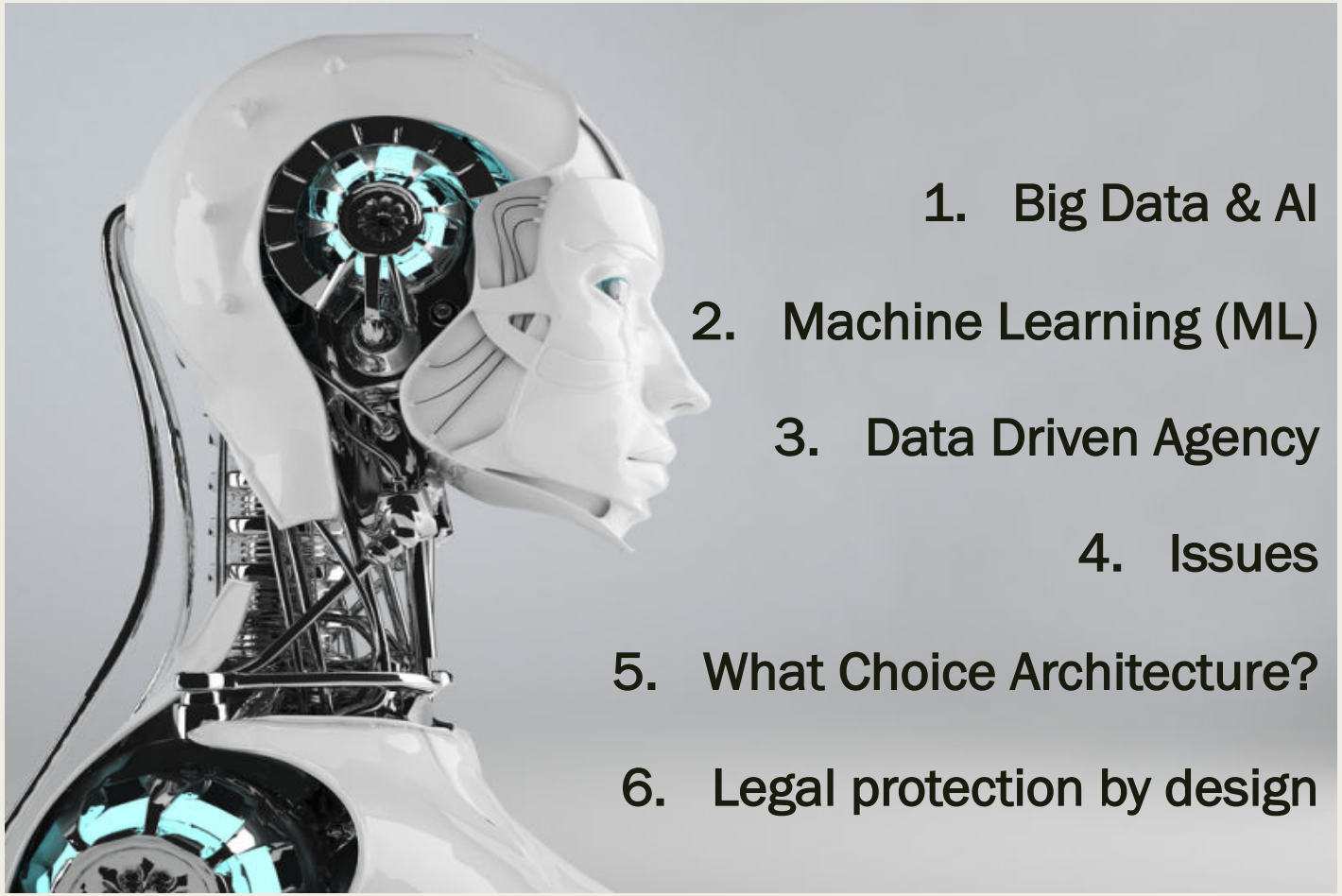
DATA-DRIVEN INTELLIGENCE

Prof. dr. Mireille Hildebrandt
Interfacing Law & Technology Vrije Universiteit Brussel
Smart Environments, Data Protection & the Rule of Law Radboud University

DATA-DRIVEN INTELLIGENCE

Prof. dr. Mireille Hildebrandt
Interfacing Law & Technology Vrije Universiteit Brussel
Smart Environments, Data Protection & the Rule of Law Radboud University

what's next?



1. Big Data & AI
2. Machine Learning (ML)
3. Data Driven Agency
4. Issues
5. What Choice Architecture?
6. Legal protection by design

what's next?

1. Big Data & AI

Fewer businesses investing in big data

By Sead Fadilpašić | Published 2 days ago

8 Comments Like 12 Share 15 G+1 2 Tweet

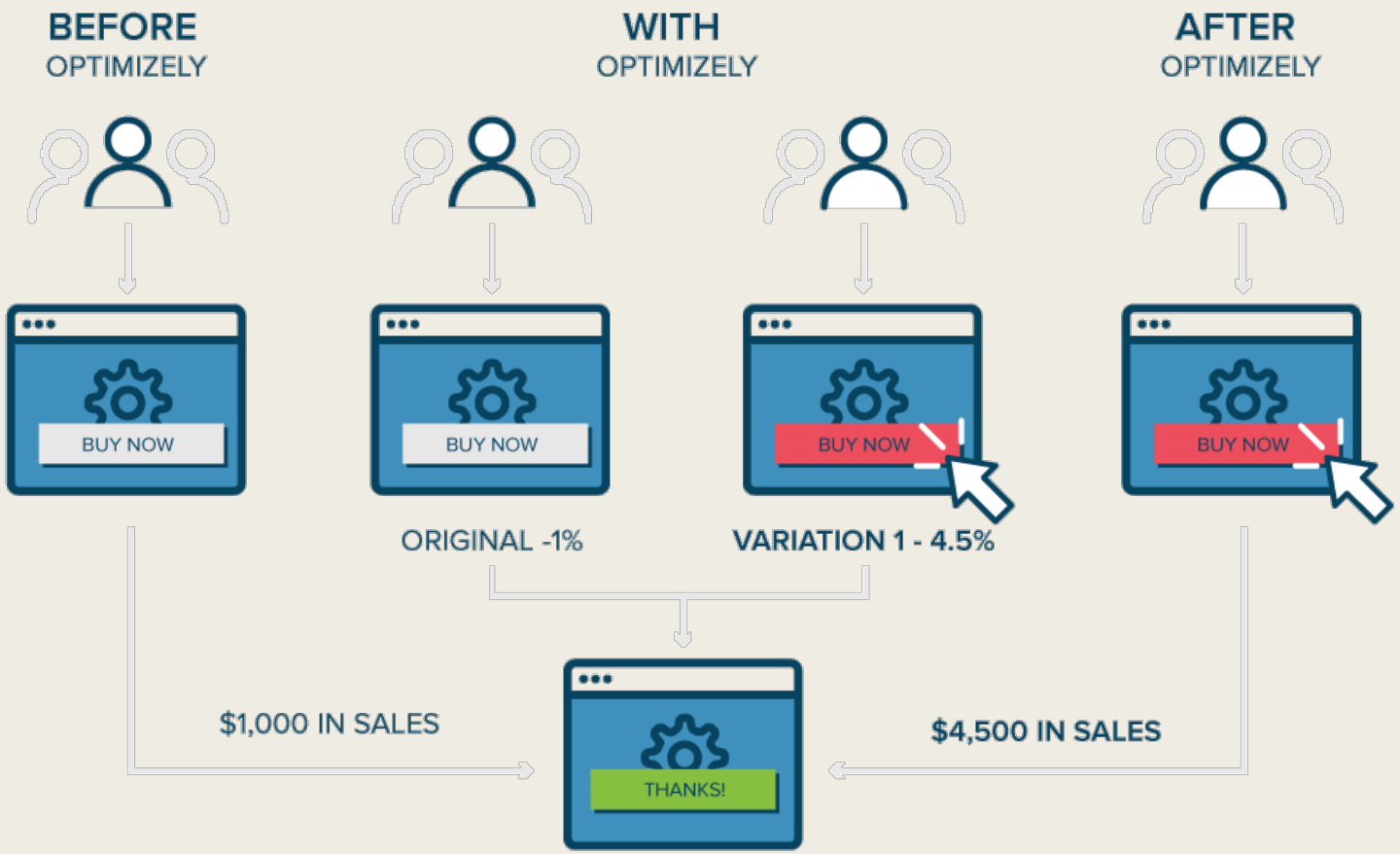
Market analyst company Gartner has issued a report that says that investment in big data is up, but fewer companies are actually planning on investing in this field. While 48 percent of companies have invested in big data in 2016, up three percent compared to the year before, the percentage those who plan on investing within the next two years is down from 31 to 25 per cent.

big data

- **big data** as a condition to conduct reliable **AI** operations
  - *big data: **volume, velocity, variety** makes data unfit for regular analytics & retrieval*

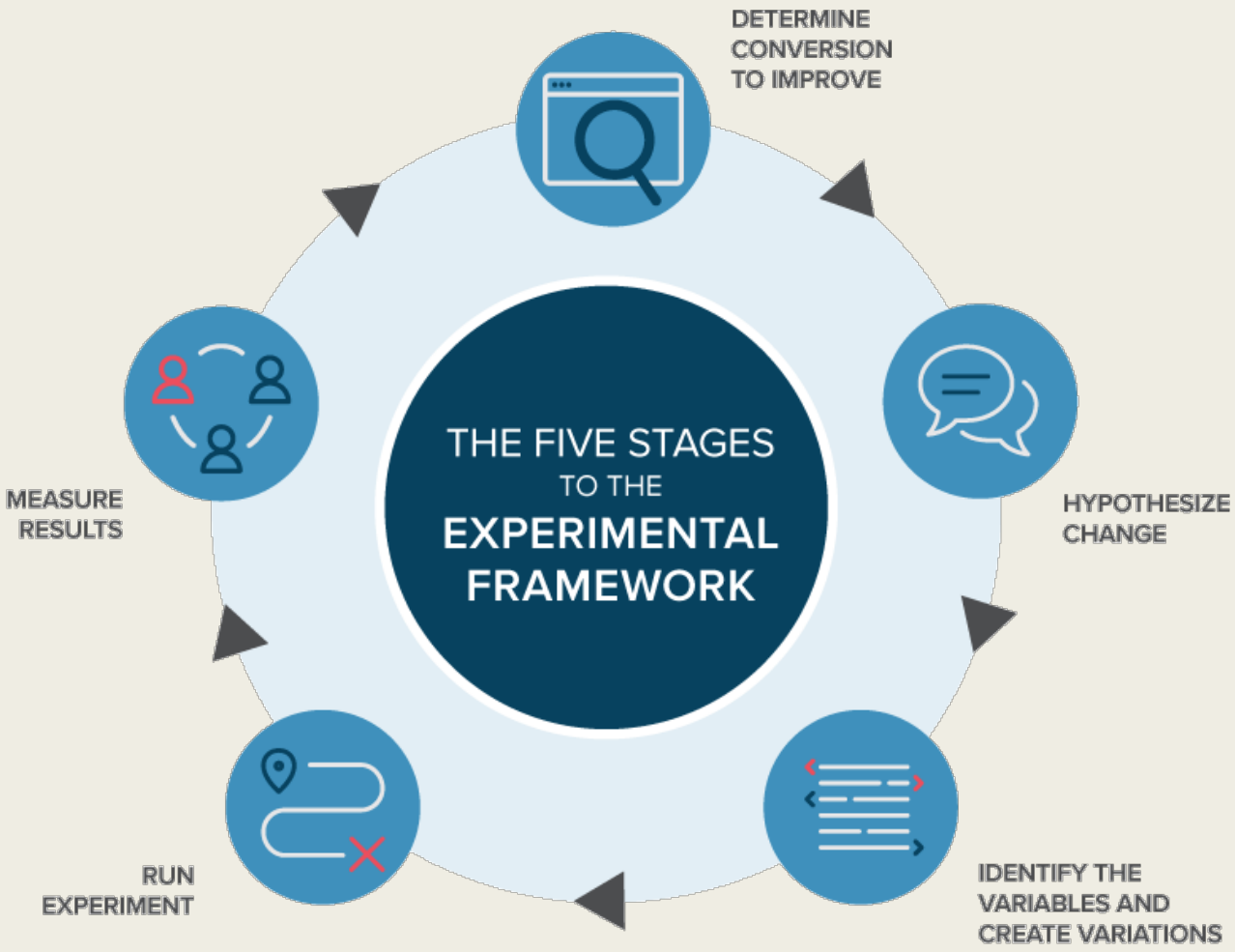
- **AI** as a condition to 'read' and comprehend **big data**
  - *AI now stands for **data-driven intelligence**, depends on relevant training data*

AB testing



- https://www.optimizely.com/ab-testing/

AB testing



- https://www.optimizely.com/ab-testing/

behavioural big data

- AB testing:
  - *experimental design*
  - *unaware guinea pigs*
  - *enables nudging & machine learning*
  - *manipulation? manipulability*
  - *example: medical data & life style data*

big data

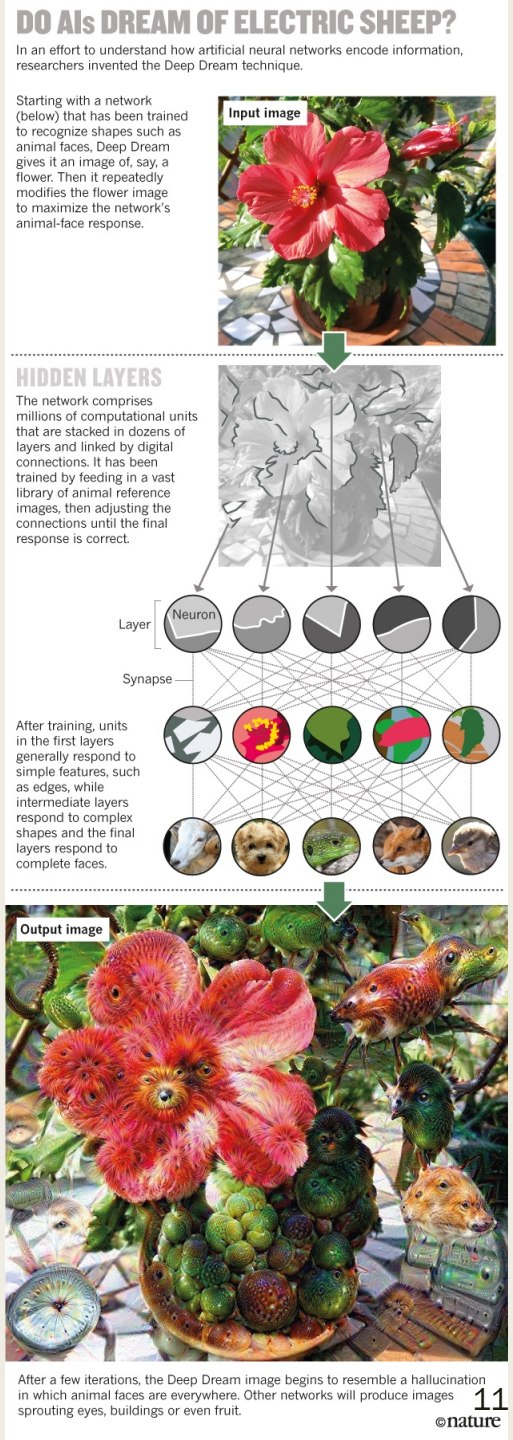
- **data obesitas:** lots of data, but often incorrect, incomplete, irrelevant (low hanging fruit)
  - *any personal data stored presents security and other **risks** (need for DPIA, DPbD)*
  - ***purpose limitation** is crucial: **select before you collect** (and while, and after)*

- **pattern obesitas:** trained algorithms can see patterns anywhere, added value?
  - *training set and algorithms **necessarily** contain bias, this **may** be problematic (need for DPIA, DPbD)*
  - ***purpose limitation** is crucial: to prevent spurious correlations, to **test relevance***

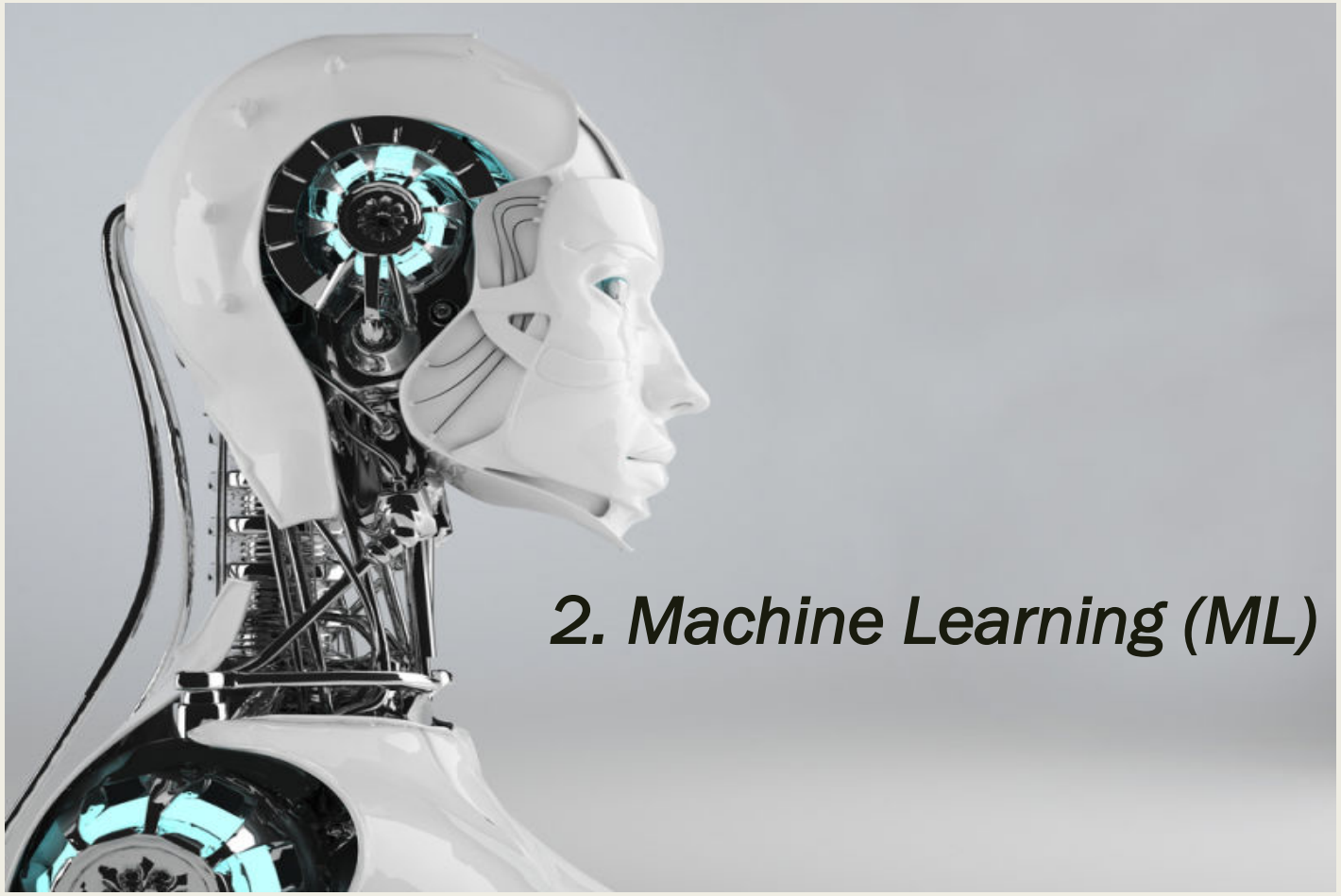
bias
optimisation
spurious correlations

- 2. have a network trained to recognize animal faces
- 1. present it with a picture of a flower
- 2. run the algorithms
- 3. check the output (see what it sees)

http://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731



what's next?



machine learning (ML)

"we say that a machine learns:

- with respect to a particular task **T**,
- performance metric **P**, and
- type of experience **E**,

if

- the system reliably improves its performance **P**
- at task **T**,
- following experience **E**."

(Tom Mitchell)

http://www.cs.cmu.edu/~tom/mlbook.html

machine learning (ML)

vocabulary when speaking of learning algorithms:

- supervised
- reinforcement
- unsupervised

supervised and reinforcement

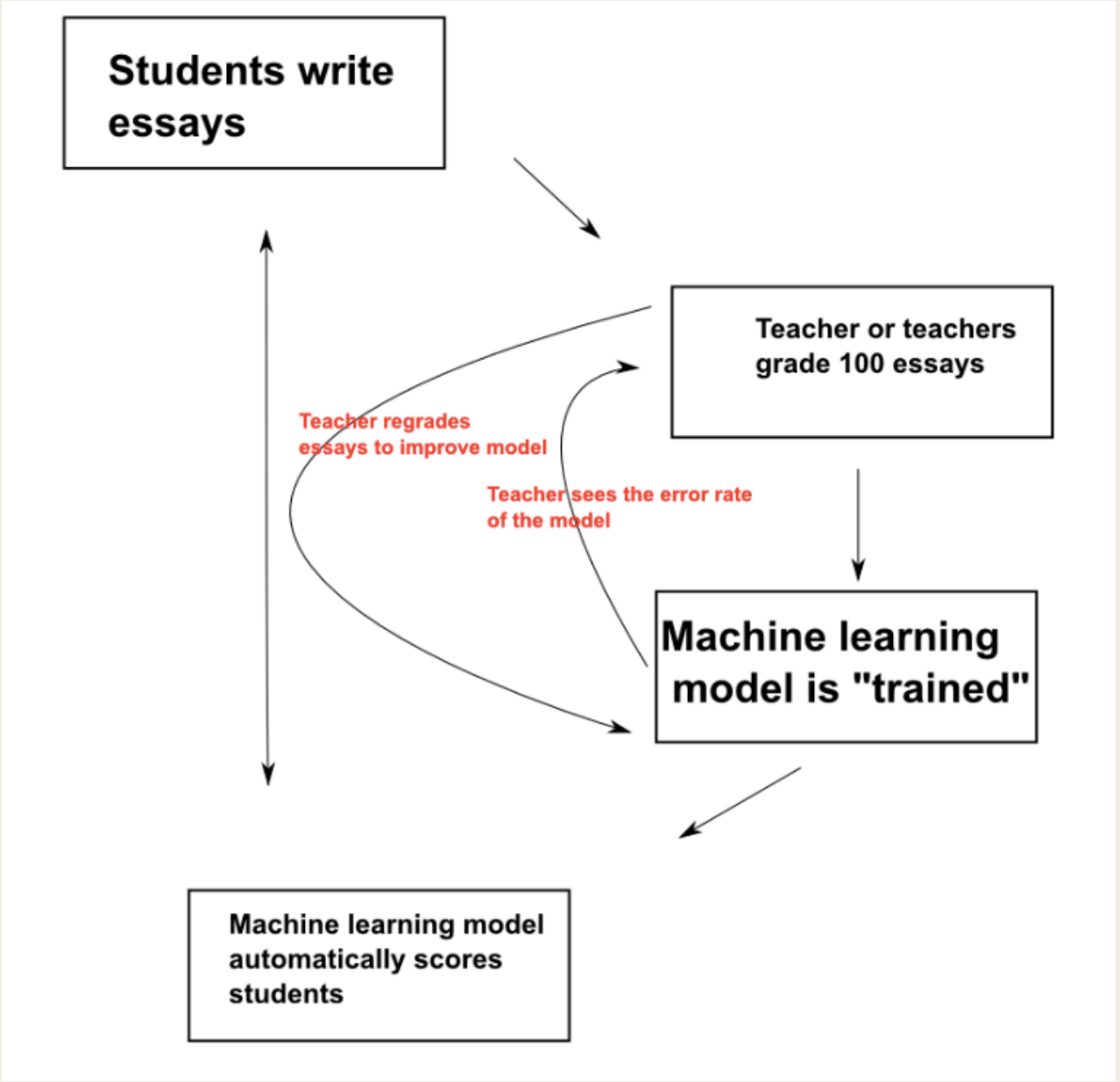
Essay-Grading Software Offers Professors a Break

By JOHN MARKOFF APRIL 4, 2013

EdX, a nonprofit enterprise founded by Harvard and the Massachusetts Institute of Technology, will release automated software that uses artificial intelligence to grade student essays and short written answers. Gretchen Ertl for The New York Times

AES for MOOCs

- automated essay scoring (AES) for
- edX (MOOC founded by Harvard & MIT)
- governed by colleges and universities, open source and non-profit
- http://www.vikparuchuri.com/blog/on-the-automated-scoring-of-essays/


Students write essays
Teacher or teachers grade 100 essays
Teacher regrades essays to improve model
Teacher sees the error rate of the model
Machine learning model is "trained"
Machine learning model automatically scores students

machine learning (ML)

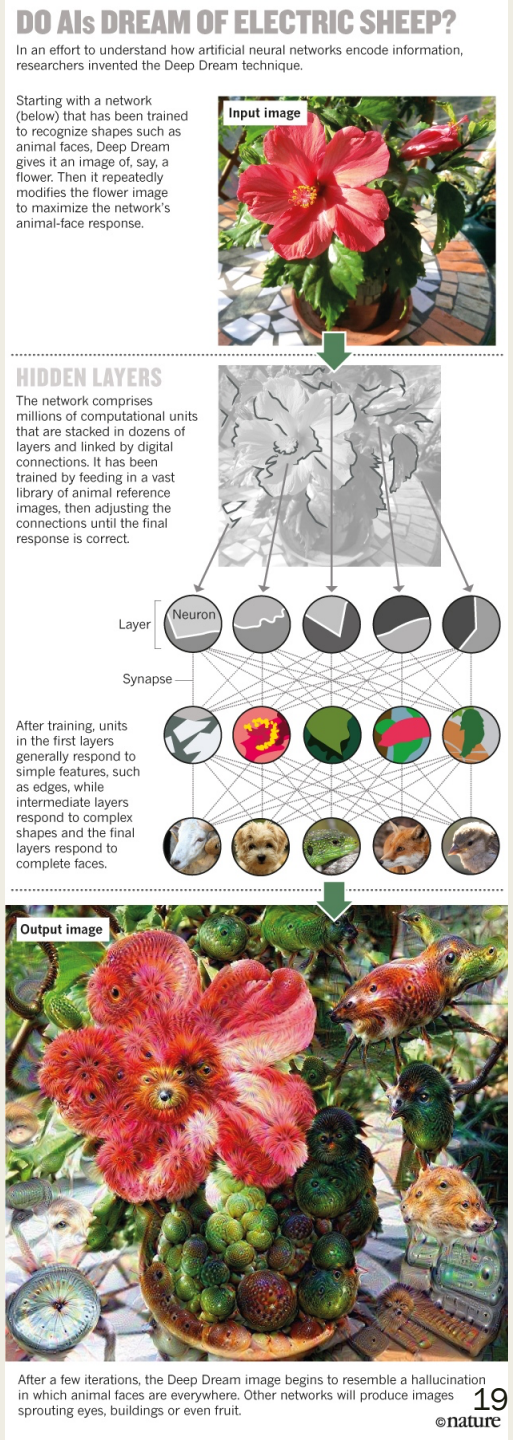
vocabulary when speaking of learning algorithms:

- unsupervised, deep learning, layered neural networks
  - *intuition*
  - *continuous pervasive AB testing*

bias
optimisation
spurious correlations

- 2. have a network trained to recognize animal faces
- 1. present it with a picture of a flower
- 2. run the algorithms
- 3. check the output (see what it sees)

http://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731



machine learning (ML)

vocabulary when speaking of learning algorithms:

- bias, optimization, training sets (synthetic data)
- simulation, multi-agent systems
- *trade-off* around the training set (volume, access, relevance)
- *trade-off* around algorithms (accuracy, speed, overfitting)
- David Wolpert's no free lunch theorem

what's next?

social robotics: uncanny valley

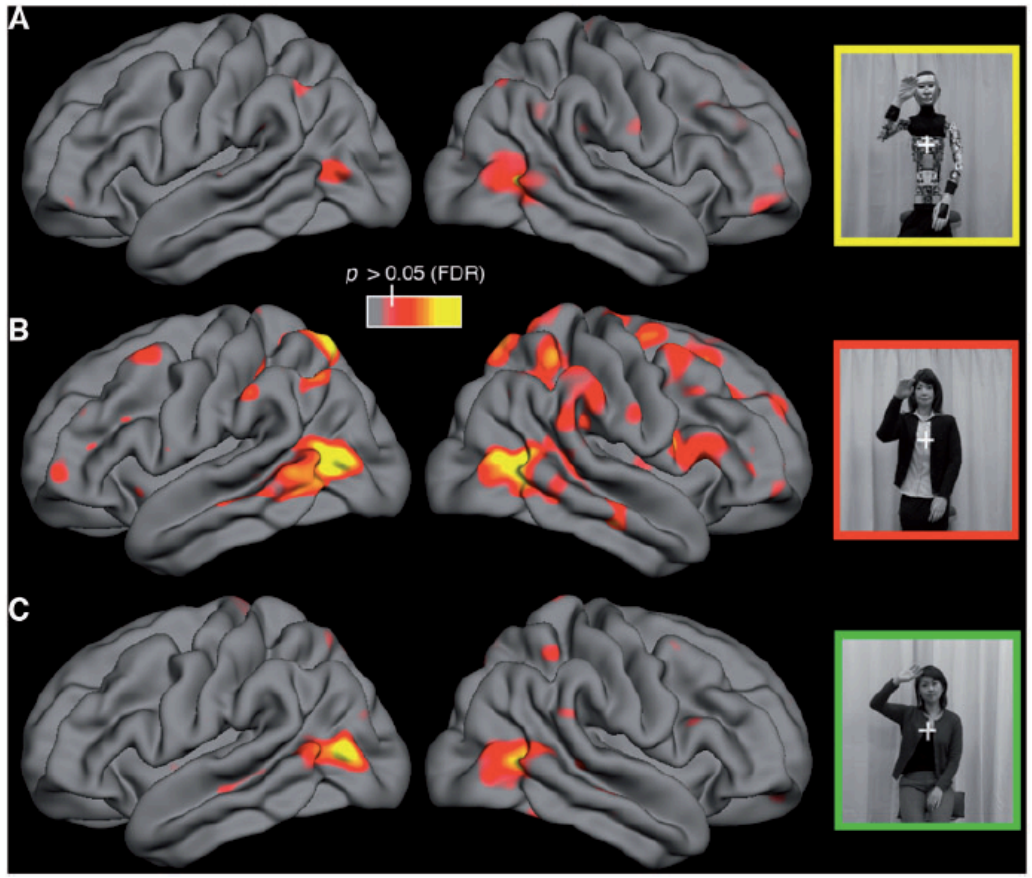


Fig. 2 Repetition suppression. Whole-brain repetition suppression effect for (**A**) Robot, (**B**) Android and (**C**) Human conditions rendered on the lateral views of the cortical of each hemisphere.

- http://scan.oxfordjournals.org/content/early/2011/04/22/scan.nsr025.full.pdf+html

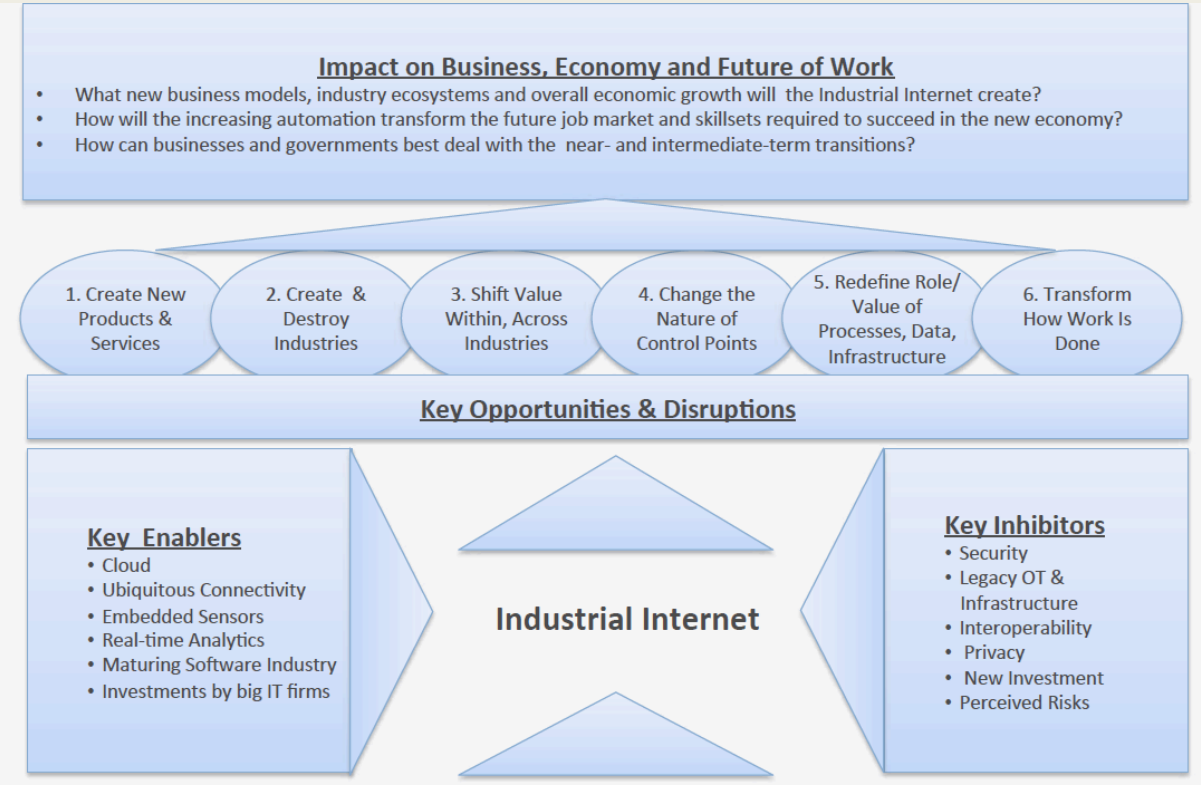
cloud robotics

- from ‘stand alone’ robotics to cloud robotics, eg. http://rapyuta-robotics.com

> The theory is that the advancement and learning of one individual robot will benefit all the rest. Faced with a newly laundered towel for the first time, a robot could query the Rapyuta database and instantly know it wasn't a T-shirt and needed folding differently - after first learning how to do the ironing, naturally.

cyberphysical infrastructures

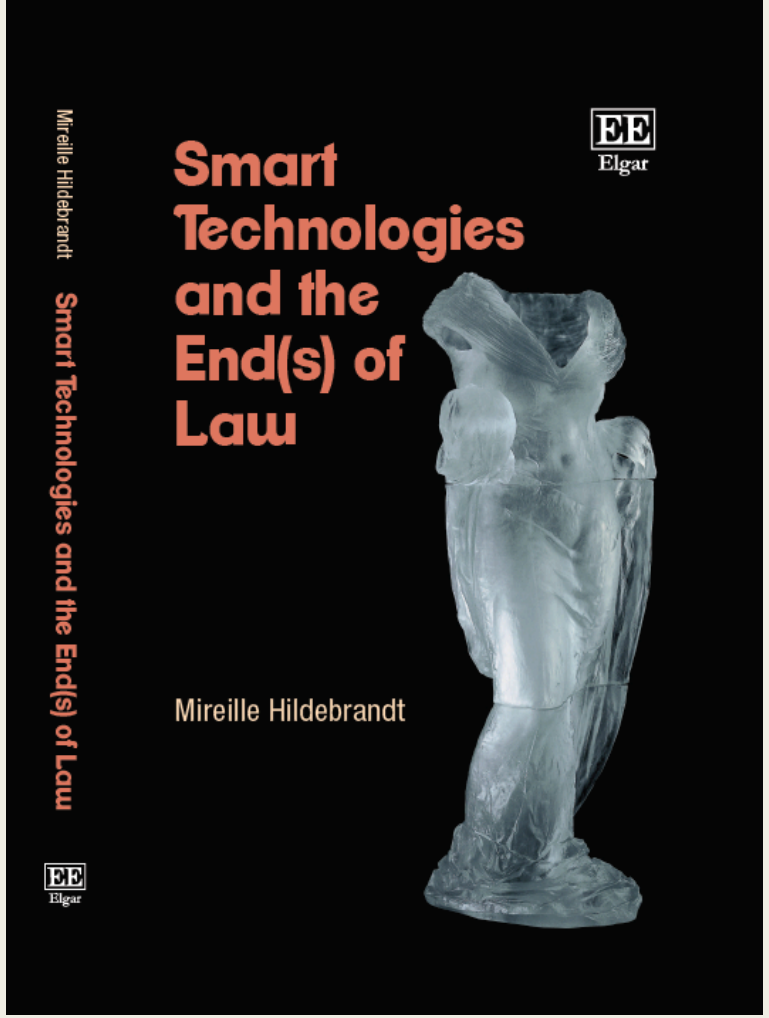
- WEForum: http://www3.weforum.org/docs/WEFUSA_IndustrialInternet_Report2015.pdf



Why speak of agency?

- data driven applications *perceive* their environment and *act* on it
- they *adapt* their own behaviour in view of their *perceived impact*

- they are ***not human agents***, cannot give reasons for their actions, however
- ***such agents foresee us whereas we cannot foresee them***
- ***indeed very often they are distributed we cannot even identify them***



Why speak of agency?

- http://www.e-elgar.com/shop/smart-technologies-and-the-end-s-of-law

what's next?

4. Issues

privacy and autonomy

- *AIs present us with a specific choice architecture:*
  - *pre-emption of our intent*
  - *playing with our autonomy*
  - *routinely making us subject to decisions of data-driven agents*
  - *this choice architecture may generate manipulability*

non-discrimination

- **three types of *bias:***
  - *bias inherent in any action-perception-system (APS)*
  - *bias that some would qualify as unfair*
  - *bias that discriminates on the basis of prohibited legal grounds*

due process & presumption of innocence

- in the case of automated decisions taken by AI systems we need:

1. *to know **that** ML or other algorithms determined the decision*
2. *to know **which data points** inform the decision and how they are weighted*
3. *which are the **envisaged consequences** of the employment of the algorithms*

the opacity argument in ML:

1. *intentional corporate or governmental self-protection and concealment*
   - *trade secrets, IP rights, public security*
2. *current education invests in writing and reading natural language, not in code or ML*
   - *monopoly of the new clerks, the end of democracy*
3. *mismatch between mathematical optimization in high-dimensionality of ML and human semantics*
   - *when it comes to law and justice we cannot settle for 'computer says no'*

- *Jenna Burrell, How the machine 'thinks': Understanding opacity in machine learning algorithms', in* ***Big Data & Society****, January-June 2016, 1-12*

Call for Papers

3rd Workshop on Fairness, Accountability, and Transparency in Machine Learning

Co-located with the Data Transparency Lab 2016

November 18, New York, NY

http://fatml.org/

Submission Deadline EXTENDED: September 16, 2016

OVERVIEW

This workshop aims to bring together a growing community of researchers and practitioners concerned with fairness, accountability, and transparency in machine learning. The past few years have seen growing recognition that machine learning raises novel challenges for ensuring non-discrimination, due process, and understandability in decision-making. In

FAT ML: Fairness

- Can we develop new computational techniques for discrimination-aware data mining?
- How should we handle, for example, bias in training data sets?
- How should we formalize fairness?
- What does it mean for an algorithm to be fair?
- Should we look only to the law for definitions of fairness?
- Are legal definitions sufficient?
- Who decides what counts as fair when fairness becomes a machine learning objective?
- Are there any dangers in turning questions of fairness into computational problems?

FAT ML: Accountability

- What would human review entail if models were available for direct inspection?
- Are there practical methods to test existing algorithms for compliance with a policy?
- Can we prove that an algorithm behaves in some way without having to reveal the algorithm? Can we achieve accountability without transparency?
- How can we conduct reliable empirical black-box testing and/or reverse engineer algorithms to test for ethically salient differential treatment?
- What are the societal implications of autonomous experimentation? How can we manage the risks that such experimentation might pose to users?

FAT ML: Transparency

- How can we develop interpretable machine learning methods that provide ways to manage the complexity of a model and/or generate meaningful explanations?
- Can we use adversarial conditions to learn about the inner workings of inscrutable algorithms? Can we learn from the ways they fail on edge cases?
- How can we use game theory and machine learning to build fully transparent, but robust models using signals that people would face severe costs in trying to manipulate?

Nature Editorial
22 september 2016

THIS WEEK

EDITORIALS

TRIALS US moves to force greater release of clinical results **p.450**

WORLD VIEW Take your vicar to the lab to build understanding **p.451**

DEATH RATTLE Lost genes can explain divergent venoms **p.453**

Algorithm and blues

Powerful computer programs are helping to make decisions that affect all of our lives. To avoid bias and discrimination, greater transparency and accountability are vital.

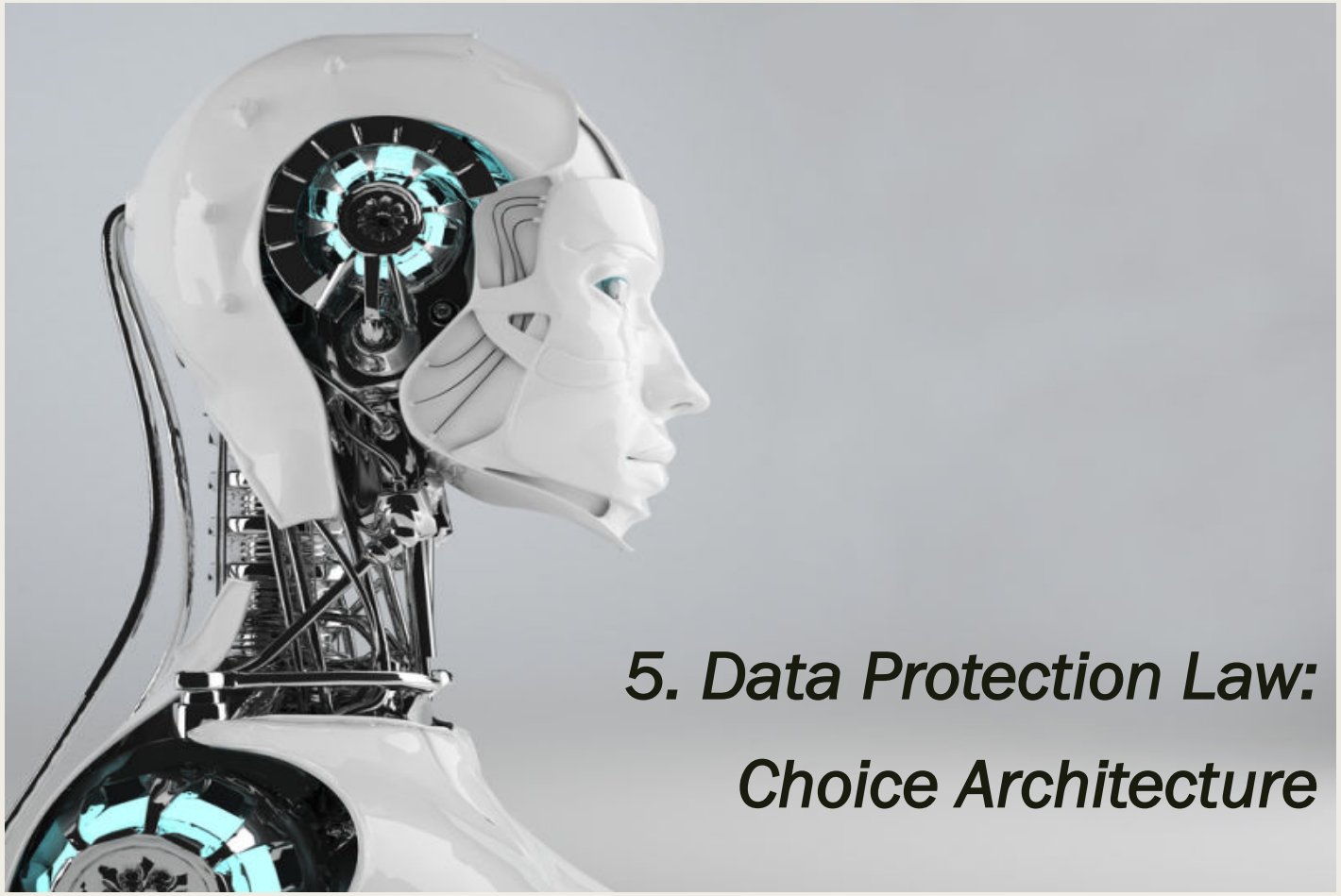
From time to time, scientific equations appear in the media and claim to distil the perfect way to make a cup of tea or identify the most miserable day of the year. Harmless nonsense? Not according to the critics who line up on social media and blogs to complain about the pseudoscience and the commercial interests of those often

turn, could exacerbate unemployment in these areas and generate a vicious circle. Algorithms using crime and other data are also susceptible to self-fulfilling prophecies that discriminate against poorer or minority areas. A big problem is that people usually have no way of knowing what their profiles are based on — or that they exist at all.

Nature editorial 22 september 2016

- "To avoid bias and improve transparency, algorithm designers must make data sources and profiles public."
- "People should have the right to see their own data, how profiles are derived and have the right to *challenge* them."
- "Some proposed remedies are technical, such as developing new computational techniques that better address and correct discrimination both in training data sets and in the algorithms — a sort of affirmative algorithmic action."

what's next?



5. Data Protection Law:

Choice Architecture

Choice Architecture

- nudge theory, cognitive psychology, behavioural economics
- what options does an environment give its inhabitants?

- what options does a data-driven environment give its ‘users’?
- architecture is politics

Choice Architecture

- who/what is using whom/what:
  - *individuals using the web, their smart car or home, mobile apps*
  - *service providers & app developers using behavioural data to improve their service & business model*

Choice Architecture

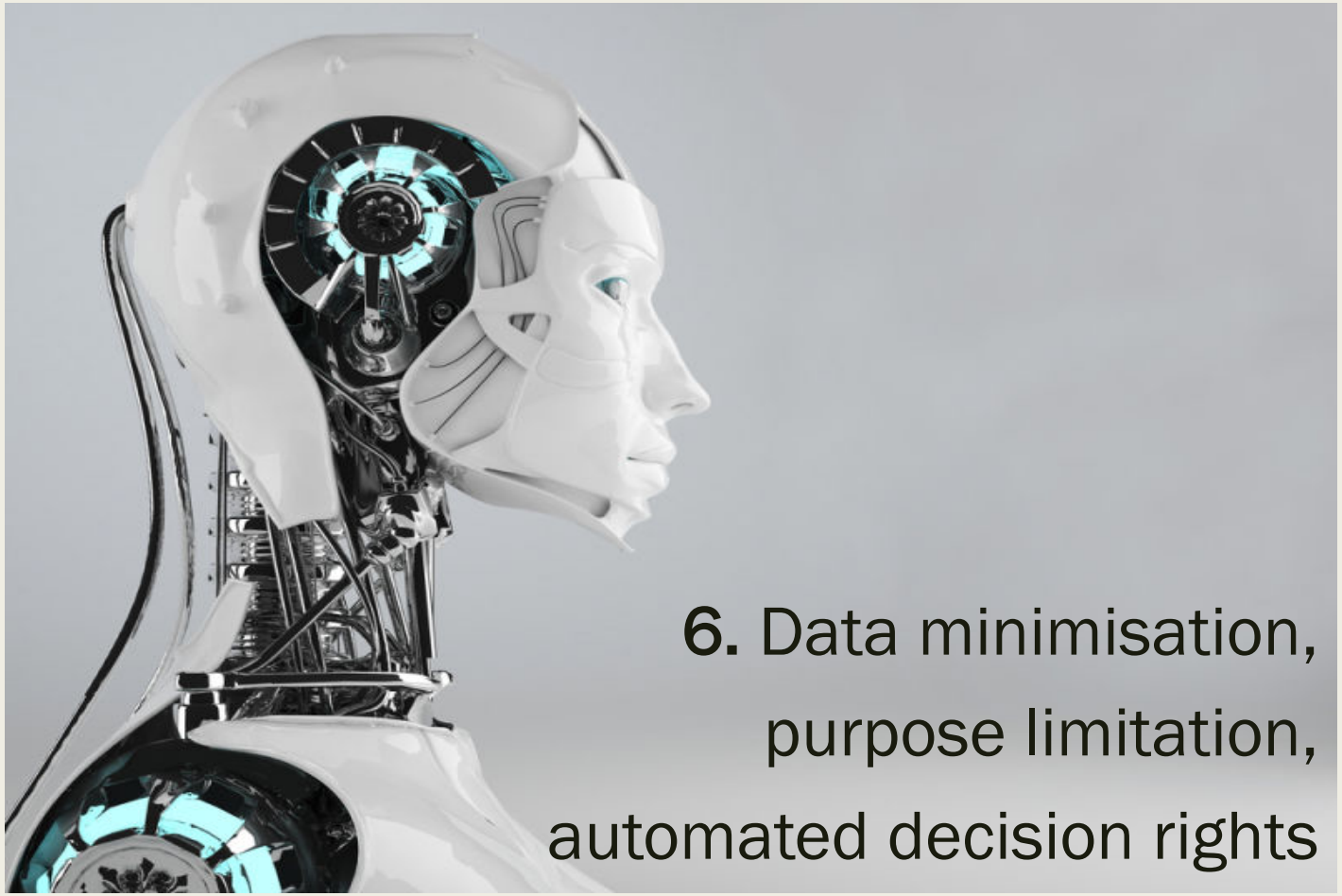
- AB testing, ML and other types of AI configure the choice architectures for their 'users'
  - *whether, and if so what level of service they can choose (consumer goods, credit, insurance)*
  - *to what education they have access, what employment opportunities they will obtain*
  - *what sentence or parole they qualify for; what level of monitoring they 'require'*

Data Protection law *as* Choice Architecture

- how does DP law constrain and reconfigure AI choice architectures?

1. what choice architecture does DP law provide data subjects?
2. what choice architecture does DP law provide data controllers?

what's next?



6. Data minimisation, purpose limitation, automated decision rights

data minimisation

= a choice architecture for data controllers:

- think ‘training sets’: select before you collect
- think of how to avoid ‘low hanging fruit’
- think of how to ensure accuracy, relevance, pertinence
- data minimisation, if done well, should avoid both data and pattern obesitas
  - *detect productive bias, while detecting unfair or prohibited bias*
  - *make data sets available for inspection and contestation*

purpose limitation

= a choice architecture for data controllers

- think ‘training sets’: select before you collect (and while you collect and after)
- think of how to avoid ‘low hanging fruit’ (GIGA)
- think of how to ensure accuracy, relevance, pertinence (depending on purpose)
  - *purpose specification, if done well, should avoid both data and pattern obesitas*
  - *purpose should direct the development and employment of data-driven applications*
  - *experimentation can be a purpose, but not in itself*
- the choice of algorithms should be informed by the purpose

automated decision rights

- current choice architecture of AI:
- ML, IoT, AI is meant to pre-empt our intent
- to run smoothly under the radar of everyday life
- it is all about continuous surreptitious automated decisions

automated decision rights

= choice architecture for data subjects (EU legislation)

1. the right not to be subject to automated decisions that have a significant impact
2. the right to a notification, an explanation and anticipation if exception applies

automated decision rights

= choice architecture for data subjects:

1. the right not to be subject to automated decisions that have a significant impact, unless
a. *necessary for contract*
b. *authorised by EU or MS law*
c. *explicit consent*

under a and c: right to human intervention, possibility to contest

prohibition to make such decisions based on sensitive data

automated decision rights

= choice architecture for data subjects:

2. the right to a notification, an explanation and anticipation if exception applies

- *existence of decisions based on profiling*
- *meaningful explanation of the logic involved*
- *significance and envisaged consequences of such processing*

DP & Privacy law: Choice Architecture

- individual citizens need:
  - *the capability to reinvent themselves,*
  - *segregate their data-driven audiences,*
  - *have their human dignity respected by the data-driven infrastructures*
  - *make sure their robotic social companions don't tell on them beyond what is necessary*
  - *the capability to detect and contest bias in their data-driven environments*

DP & Privacy law: Choice Architecture

- the architects of our new data-driven world need:
  - *integrity of method: rigorously sound and contestable methodologies (bias)*
  - *accountabiity: (con)testability of both data sets and algorithms*
  - *fairness: testing bias in the training set, testing bias in the learning algorithm*
  - *privacy & data protection: reduce manipulability, go for participation and respect*

what's next?

'by design' paradigm

- *architecture is politics*
- translate fairness, methodological integrity, fundamental rights into the architecture
- Data Protection by Default: engineer data minimisation as a requirement
- Data Protection by Design: engineer state of the art DP tools as a requirement


THE END IS NEAR
THE END
Roy Delgado