The Ethics of IT and AI:

When we can do it, should we do it?

Chris Rees
President, BCS

Edinburgh Branch, 2nd May 2018

Agenda

1. Some definitions
2. Making IT Good for Society
3. When should we think about ethics?
4. The obstacles to ethical action
5. What to do about it
6. The Ethics of AI
7. Conclusions

Some Definitions

Ethics - Of or relating to **moral principles**, esp. as forming a system... (OED)

Moral - Of or relating to **human character or behaviour considered as good or bad**; of or relating to the **distinction between right and wrong**, or **good and evil**, in relation to the actions, desires, or character of responsible human beings; ***ethical***. (OED)

"***Ethics*** is the 'study of **what is right or what ought to be**, so far as this depends upon the voluntary action of individuals; assuming that whatever we judge to be '**good**', we implicitly judge to be something which we '**ought**' to bring into existence". (Sidgwick, 1893)

Making IT Good for Society

How can we do that if we do not always act ethically?

When should we consider whether what we are about to do is ethical?

When we conceive

specify

design

develop

test

implement

use a system or application

Including selecting datasets, training AI's, etc.

At every stage we should consider whether what we are about to do is ethical

It's easier if ethical risks are identified as early as possible –

Ethical by Design

Is it easy to be ethical?

Most people are ethical and want to act ethically – though not all!

In many cases, there is no ethical issue

But if there is, there can be significant obstacles:

- The desire to conform – others do it, don’t step out of line…“It’s what everyone does…”
- The need to make your budget, achieve your target, deliver to the deadline…
- You are too junior to rock the boat
- You have been told to do it
  - – Refusing an instruction may put your career on the line

What should you do?

If there may be an ethical issue in what you are about to do, what should you do?

Fall still, and then ask yourself the question:

Is what I am about to do completely ethical?

*If the answer is **no**, fall still again, and ask yourself:*

What should I do?

There will not be a standard answer, but the answer – or answers – may well arise, together with the strength to act on them.

You may need help and guidance

BCS may be able to help with a code of ethics, guidance, methods, training

Part 2 - **The Ethics of AI**

1. Some more definitions
2. Bias in the data
3. Transparency and Explainability
4. Correlation vs Causation
5. Harmlessness
6. Liability
7. Sharing the benefits fairly & mitigating negative effects

The State of the Art

AI is already “superhuman” at chess, Go, speech transcription, lip reading, deception detection from posture, forging voices, hand-writing & video.

This is real intelligence

But narrow, not general AI

Some more definitions

Bias: expectation derived from experience of regularities in the world

Stereotype: bias based on regularities we do not wish to persist

Prejudice: acting on stereotypes.

For Instance:

Bias: Knowing what *programmer* means, including that most are male

Stereotype: Knowing that most programmers are male

Prejudice: Hiring only male programmers

Bias in the data - facial recognition software

Google's image recognition technology confused people with dark skins and gorillas.

This bias has history:

- White balance feature of digital cameras was made for white people.
- Digital cameras could be differently tuned but they were built to replicate film.

Research by Joy Buolamwini, researcher at the M.I.T. Media Lab
Unrecognised by the algorithm until she put on a white mask

Gender was misidentified in **up to 1 percent of lighter-skinned males** in a set of 385 photos

Gender was misidentified in **up to 7 percent of lighter-skinned females** in a set of 296 photos

Gender was misidentified in **up to 12 percent of darker-skinned males** in a set of 318 photos

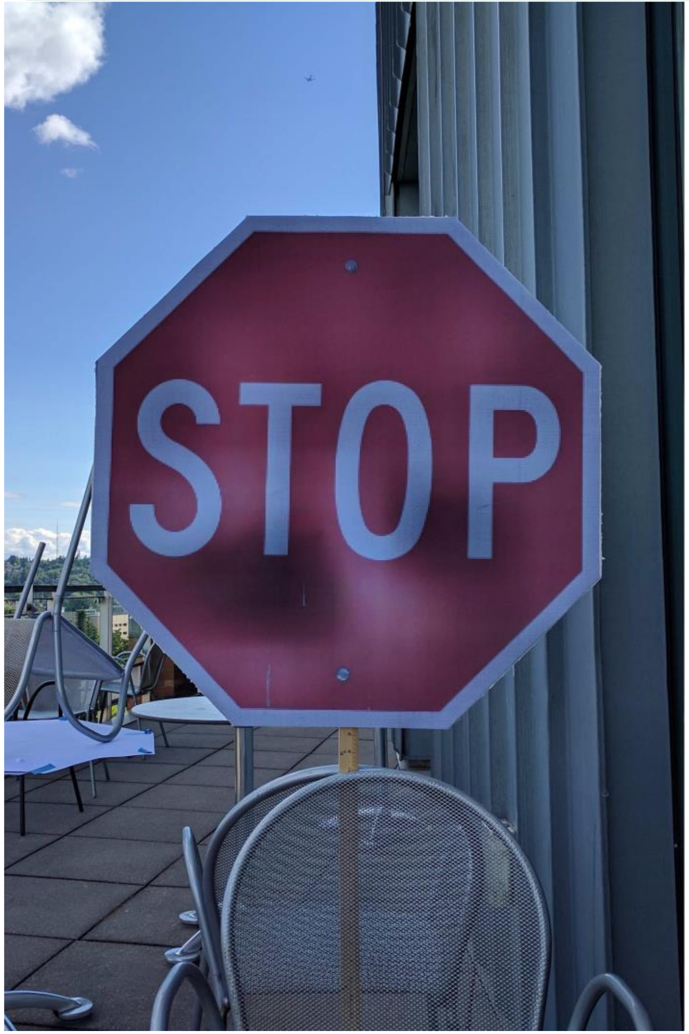
Gender was misidentified in **35 percent of darker-skinned females** in a set of 271 photos.

Why is there bias in AI and datasets that train AI?

1. Because we have biases and stereotypes
   - – e.g. Programmers are male
   - – Datasets contain these biases
   - – Some biases can be accurate
   - – Stereotypes are culturally determined

   There is no algorithmic way to discriminate stereotype from bias
2. Designers and developers are not diverse – **Accidental bias**
3. Datasets on which the AI is trained are biased – **Implicit bias**
4. Introduced intentionally in the development process – **Deliberate bias**

Deliberately introduced bias



STOP sign that Machine Learning, trained on a standard database of road signs, recognises as a 45MPH speed limit.

There are other examples, including a “right turn": useful if you lived on a rat run.

How should we address machine learning bias?

Bias is inevitable because we are biased

We address it in the same way as we address our own biases:

We have to recognise it, compensate for it
and eliminate prejudiced biases

Implicit bias – compensate with design, architecture

Accidental bias – As above but also diversify work force,
then test, log, iterate, improve

Deliberate bias – audits, regulation

Transparency & Explainability:
The Black Box problem

An AI system recently cracked a German Enigma code in 13 minutes

- It did not “know” what it had done or how it had done it
- **Its designers did not know how it had done it**

A common characteristic of deep learning systems

IBM and others argue that Black Box systems should not be marketed at all

Should we distinguish between safety- or life-critical systems and others?

e.g. medical diagnosis vs. language translation

This may limit what can be achieved with AI

Some human decisions are impenetrable or illogical.

Why impose higher standards on AI?

Intelligibility: Transparency & Explainability

House of Lords Report identified domains requiring intelligibility such as

- Judicial & legal affairs
- Some financial products & services (e.g. personal loans, insurance)
- Autonomous vehicles
- Weapons systems

The HoL report distinguishes between Technical Transparency & Explainability

Technical Transparency – e.g. access to source code

Helpful for experts, regulators perhaps, not for the layman

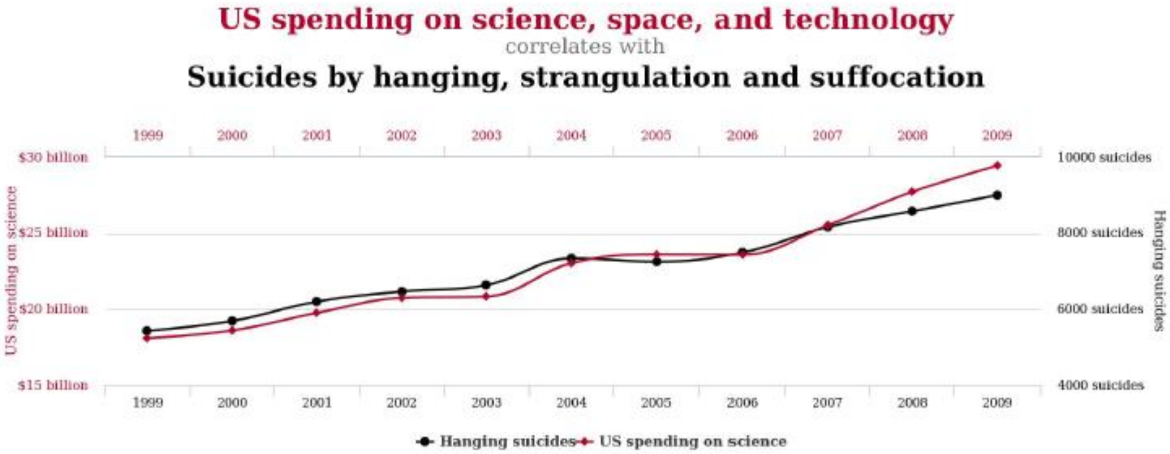
May not explain how a decision was reached

Explainability - AI is developed so that it can explain the information and logic it used to reach a decision

Development of Explanation systems (NB GDPR Art. 22)

Correlation vs Causation

- It’s an old problem in statistics – Correlation does not imply causation but may seem to



- Statisticians have tests to detect the problem, AI systems do not
- Exacerbated by AI

Harmlessness

- Like all technology, ML, AI and Robotics are ethically neutral
- By the same token they are dual use – can be used for good and ill
- Major risks of malicious use:

  spear phishing, hacking, impersonation, poisoning data, use of drones, political influence…

  – Expansion of existing threats

  – Introduction of new threats

  – Change in character of threats, by AI and to AI
- All obviously unethical – but they impose an obligation to develop defence and counter-measures

Harmlessness - LAWS

- Lethal Autonomous Weapons Systems (LAWS) pose particular ethical challenges
  - E.g. Will they discriminate between combatants and civilians
  - Should they be banned like chemical and biological weapons?
    - The International Committee for Robot Arms Control (ICRAC) and The Campaign to Stop Killer Robots think so
  - All major powers are developing them, for offensive and defensive applications
  - Non-state actors may well be developing them too

Would a ban work?

Liability: a legal, societal and ethical issue

Self-driving cars/autonomous vehicles (AVs) will not be on the road for consumers until the liability issues have been resolved:

When an accident occurs which is the ‘fault’ of the AV, who is liable (Not the AV!):

- The ‘driver’ (if there is one)?
- The owner?
- The company that sold it?
- The manufacturer?
- The designer of the AV or the failing component – e.g. the AI control system?
- What if the AV has been hacked?
- What if the owner has failed to install updates issued by the manufacturer?

It would be unethical if the allocation of responsibility were unfair

This requires legislation based on full consultation between parliament, the industry, insurers and the public.

Sharing the Benefits and Mitigating the Negative Effects

The rapid spread of AI and robotics will have a material impact on employment, as the machines perform human functions faster, cheaper, better than humans

This is not a new phenomenon – e.g. in the Industrial Revolution

The tech industry argues that it will create new, skilled, well-paid jobs

BUT

There were several recessions and massive unemployment in the Industrial Revolution

In the **long term** there will be new jobs but in the **short term** there will be dislocation

There will be new jobs and job functions, but will they suit those put out of work?

The new job functions may be performed more effectively by machines than people

SOLUTIONS?

Universal income??

Re-training by companies, Colleges of Further Education, etc. (eg AT&T)

Funded by individuals, companies, the state. (cf Thames bridges)

Conclusions

If IT is to be Good for Society, it must be ethical

We have a duty and a need to act ethically in the development and use of IT

In AI we need to guard against bias in the training data, eliminate stereotypes & prejudices

Critical AI systems must be explainable

Guard against confusing correlation with causation

Protect society from malicious AI and LAWS

We need a public conversation about liability

We must find a way to share the benefits and mitigate the negative effects – Training is key

It is our responsibility to set an example