bcs
The Chartered Institute for IT

The Ethics of AI and IT

Chris Rees
President, BCS

Adapted by Mike Pickup FBCS

Agenda **The Ethics of AI and IT**

1. Some definitions
2. Systems of Ethics
3. Why is ethics so important in AI?
4. Bias
5. Transparency and Explainability
6. Correlation vs Causation
7. Harmlessness
8. Responsibility and Liability
9. Sharing the benefits fairly & mitigating negative effects
10. How to be ethical in AI and IT
11. 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". (Henry Sidgwick, 1893)

Ethical Systems

Deontological ethical standards – Focus on Intention and means

- Essential for medical systems!

Teleological ethical standards – Focus on ends and outcomes

- Might be used to justify data collection and experimentation for AVs

Apply the Golden Rule? Does it work with Deep Learning Systems?

Cultural relativism vs. Universalism

- Use of AI for Chinese oppression of Uyghurs in Xinjiang
- ***"Totalitarian determination and modern technology have produced a massive abuse of human rights"*** *– The Economist May 2018*

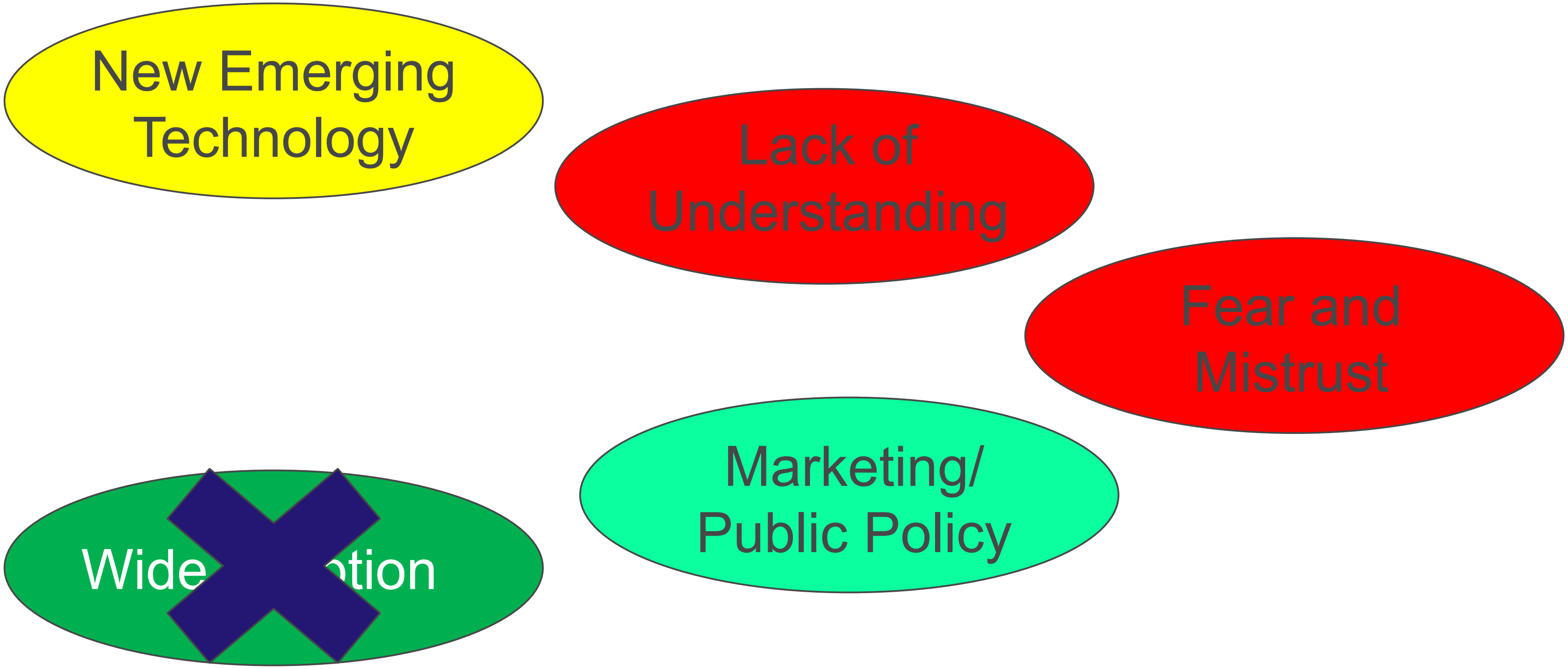
Some More Definitions

Moral Agent – has the ability to judge right from wrong and to act on the basis of reasons. Thus responsible for its actions. Should AIs be moral agents? Could you sue an AI?

Moral Patient – Moral patients are entities towards which moral agents can have moral responsibilities. On this definition, all moral agents are also moral patients, but moral patients need not be moral agents. Human beings are moral patients. Are human embryos, animals, future people moral patients? Are AIs/robots moral patients? Should they be?

A baby is not a moral agent but is a moral patient:
those with moral agency should care about its well-being.

Why is ethics so important?



New Emerging Technology

Lack of Understanding

Fear and Mistrust

Marketing/ Public Policy

Wide tion

Unless AI is perceived to be ethical, this may not happen.

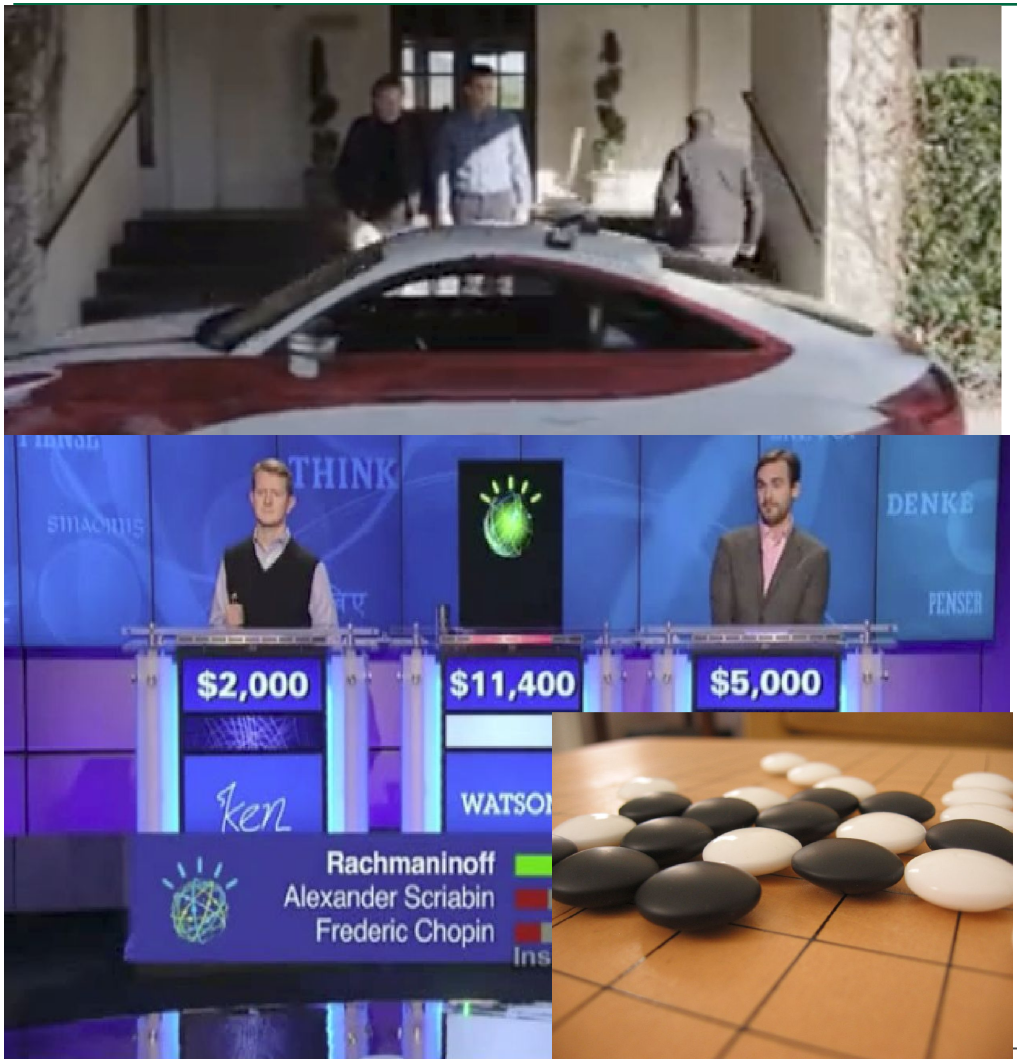
The huge benefits could be lost. Remember GM foods

2017 Eurobarometer

- 84% of respondents agree that robots can do jobs that are too hard/dangerous for people
- 68% agree that robots are a good thing for society because they help people
- **61%** of respondents have a **positive view of robots**
- **88%** of respondents consider robotics a technology that **requires careful management**
- **72%** of respondents think robots **steal people's jobs**

Essential to address ethical (as well as legal, societal, and economic) issues in advance.

The State of the Art



AI is already
“superhuman” at
chess, Go,
speech transcription,
lip reading,
Deception detection
Forging voices,
hand-writing & video
Cancer detection…

This is real intelligence

But Narrow AI, not AGI

Bias in the data - facial recognition software

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

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

She is a Rhodes Scholar, a Fulbright fellow, a Stamps scholar, an Astronaut Scholar and an Anita Borg Institute scholar.

Founder of Algorithmic Justice League

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.

Sources of Bias

Bias: A predisposition, inclination, or leaning for or against one person or group, especially in a way considered to be unfair. (OED)

Bias may be introduced **accidentally** – lack of diversity among developers

deliberately, i.e. intentionally

from **biased data** in the **training dataset**

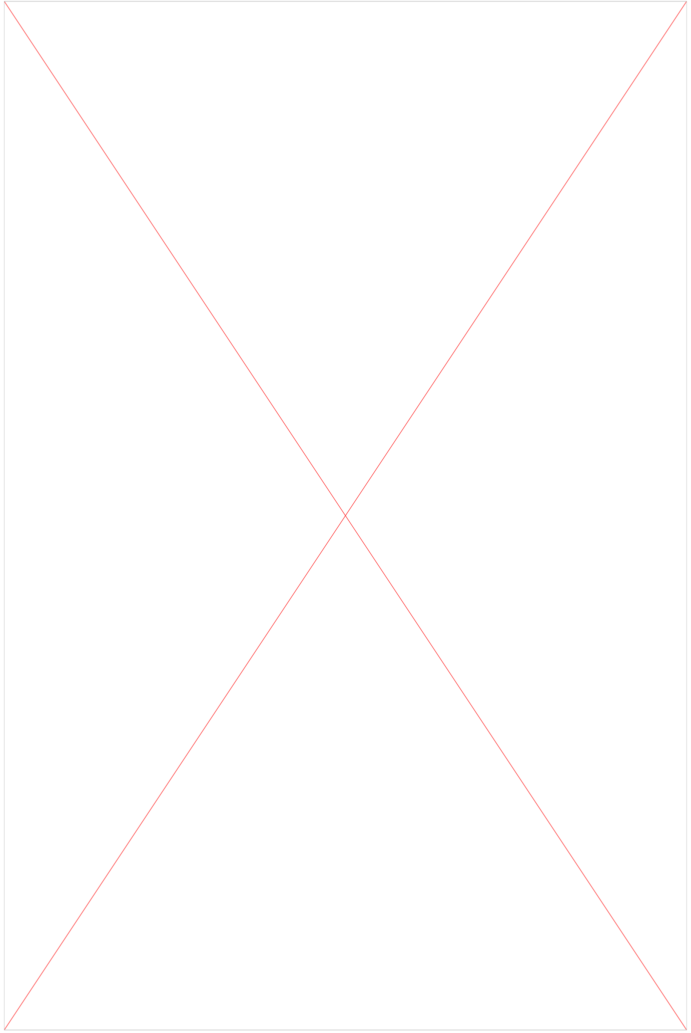
The larger the dataset, the more biases

We all have biases. We are not aware of all of them. Not all are bad.

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
2. **Accidental bias** – Designers and developers are not diverse
3. **Implicit bias** – Datasets on which the AI is trained are biased
4. **Deliberate bias** – Introduced intentionally during development

Deliberately introduced bias



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

There are other examples, including a “right turn"

How should we address machine learning bias?

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

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

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. Is it possible?

Some human decisions are impenetrable or illogical.

Intelligibility: Transparency & Explainability

House of Lords (HoL) 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

Transparency e.g. access to source code

Helpful for experts, regulators, not for the layman

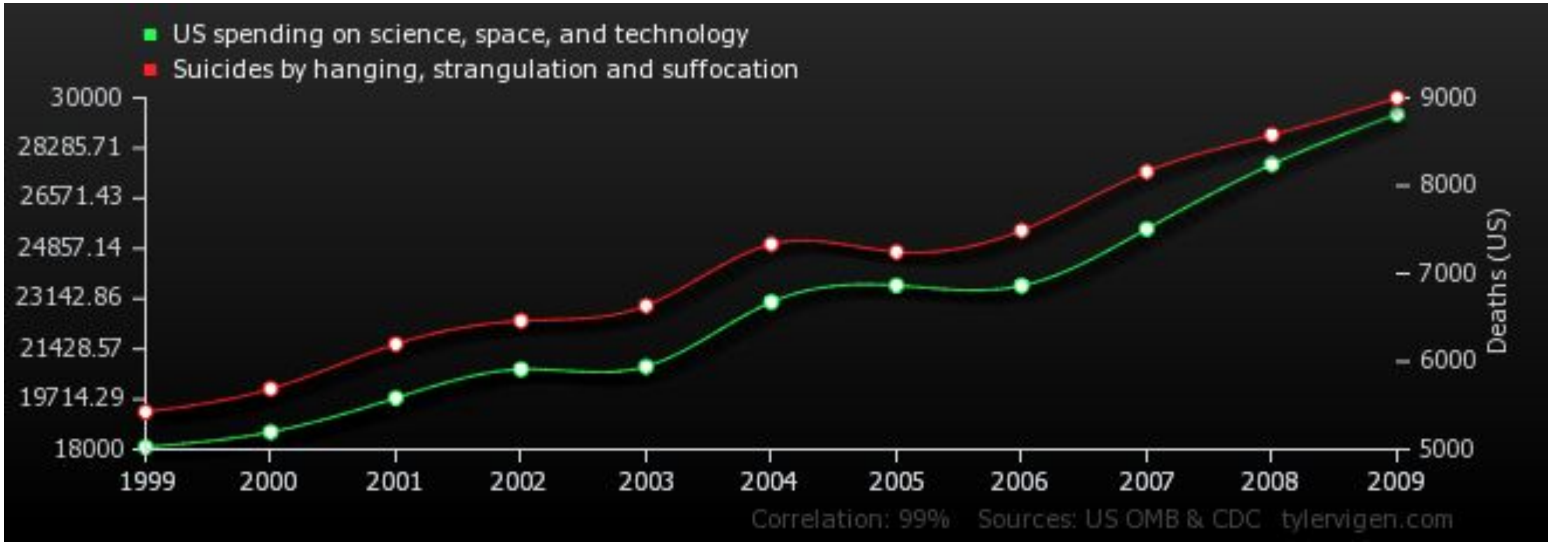
May not explain how a decision was reached

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

Correlation vs Causation

It’s an old problem in statistics – Correlation does not imply causation

US spending on science, space, and technology and suicides correlates with deaths by hanging, strangulation and suffocation



Statisticians have tests to detect the problem, AI systems do not. The problem is exacerbated by AI

Harmlessness – Malicious Use of AI

Like all technologies, AI, ML 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

Isaac Asimov’s 3 Laws of Robotics - 1942

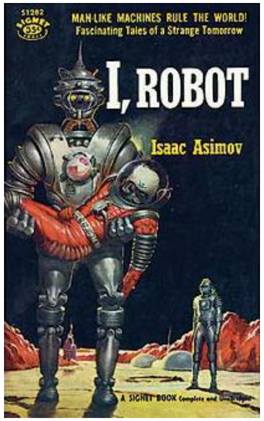
1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

0. A robot may not harm humanity, or, by inaction,

allow humanity to come to harm.


S1282
SIGNET
35c
MAN-LIKE MACHINES RULE THE WORLD!
Fascinating Tales of a Strange Tomorrow
I, ROBOT
Isaac Asimov
A SIGNET BOOK Complete and Unabridged

Harmlessness - LAWS

Lethal Autonomous Weapons Systems (LAWS) pose particular ethical challenges

- If they are not under human control
  - Will they discriminate between combatants and civilians?
  - What if the situation has changed?
  - Could the attack be aborted? Drone attacks often are.
- 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 offense and defence
- 3,100 Google staff recently forced **Google** to withdraw from **Project Maven**
- **Non-state actors** may well be developing them too
- Would a ban work? AI-driven drones can be weaponised

Liability: a legal, societal and ethical issue

When an accident occurs which is the ‘fault’ of the AV, who is liable?

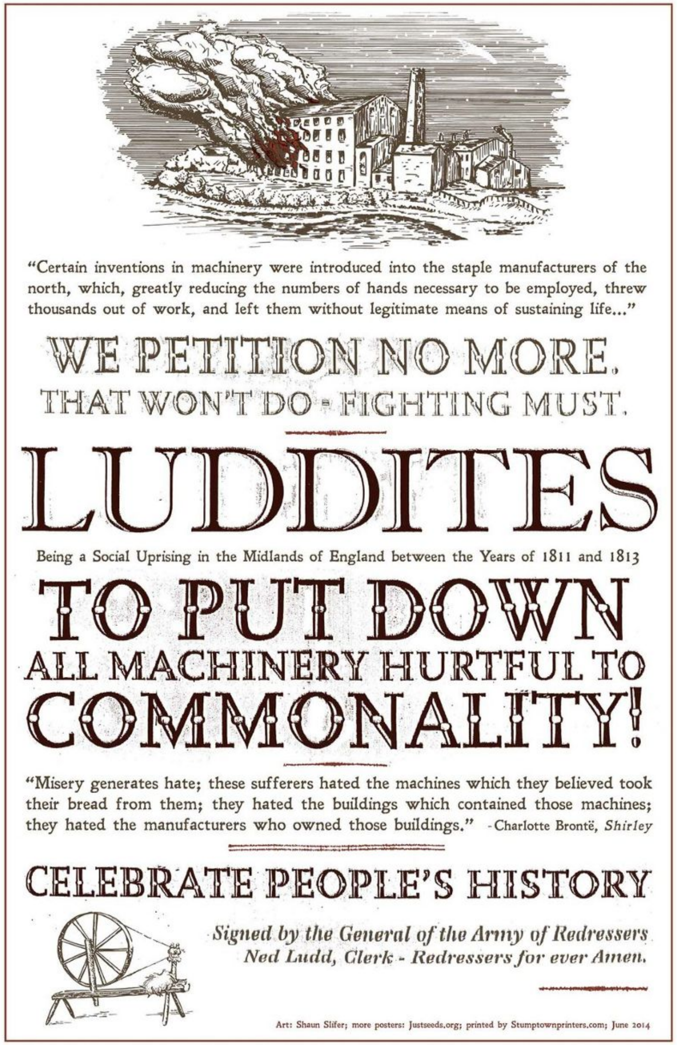
- The AV, an artefact, should not be a moral agent, so Not the AV!
- The ‘driver’ (if there is one)?
- The owner?
- The retailer that sold it to the customer?
- The manufacturer?
- The designer of the AV or of the failing component?
- What if the AV has been hacked?
- What if the owner has failed to 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.

The Fourth Industrial Revolution

We have been here before



"Certain inventions in machinery were introduced into the staple manufacturers of the north, which, greatly reducing the numbers of hands necessary to be employed, threw thousands out of work, and left them without legitimate means of sustaining life..."

WE PETITION NO MORE.
THAT WON'T DO - FIGHTING MUST.

LUDDITES

Being a Social Uprising in the Midlands of England between the Years of 1811 and 1813

TO PUT DOWN
ALL MACHINERY HURTFUL TO
COMMONALITY!

"Misery generates hate; these sufferers hated the machines which they believed took their bread from them; they hated the buildings which contained those machines; they hated the manufacturers who owned those buildings." -Charlotte Brontë, *Shirley*

CELEBRATE PEOPLE'S HISTORY

Signed by the General of the Army of Redressers
Ned Ludd, Clerk - Redressers for ever Amen.

Art: Shaun Slifer; more posters: Justseeds.org; printed by Stumptownprinters.com; June 2014

The Fourth Industrial Revolution

Have we not been here before? The only group that completely lost employment in the Industrial Revolution were the horses. New jobs were created to replace the ones lost.

OECD report shows that worries about “massive technological unemployment” are to some extent overblown. Instead the risks are of “further polarisation of the labour market” between highly paid workers and other jobs that may be “relatively low paid and not particularly interesting. Only 13m jobs in the USA are at risk, not the 47m predicted by Fey and Osborne. *(OECD report on Automation, skills use and Training)*

However 13m is more than all the manufacturing jobs in the USA in 2018!

Andy Haldane

“Previous revolutions had “a wrenching and lengthy impact on the jobs market, on the lives and livelihoods of large swathes of society”

“Jobs were effectively taken by machines of various types, there was a hollowing out of the jobs market, and that left a lot of people for a lengthy period out of work and struggling to make a living.”

“That hollowing out is going to be potentially on a much greater scale in the future, when we have machines both thinking and doing” *Andy Haldane, Bank of England*

Sharing the Benefits and Mitigating the Negative Effects

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

We need to ensure that the benefits do not all accrue to a small elite,
while the risks and dislocation fall upon the many with fewer skills.

SOLUTIONS?

Universal income??

Re-training by companies, Colleges of Further Education, etc

Funded by individuals, companies, the state?

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?

From conception through design to training and use

At every stage we should consider:

Is what we are about to do 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 – "It's what everyone does…"
- Work Pressures: 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 you face an ethical issue?

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?

Write down the essential Facts:
Who is affected?,
The Pressures affecting the situation (e.g. time, money, reputation, job security)
Are there legal issues?
What are the ethical issues?
What are the guiding principles?

What Choices do I have?

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

Make the decision and ask – would I be happy if this appears in the papers?

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

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 AI and 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

If AI is not ethical it may be rejected by the public.