digitdebates

How digital technologies are changing work

Why the world needs to worry about the ethics of AI

SPEAKER:
Simon Greenman
Member, Global AI Council, World Economic Forum
Partner, Best Practice AI

AI is a GPT that will be woven into the fabric of society with the potential to transform lives, companies, and government
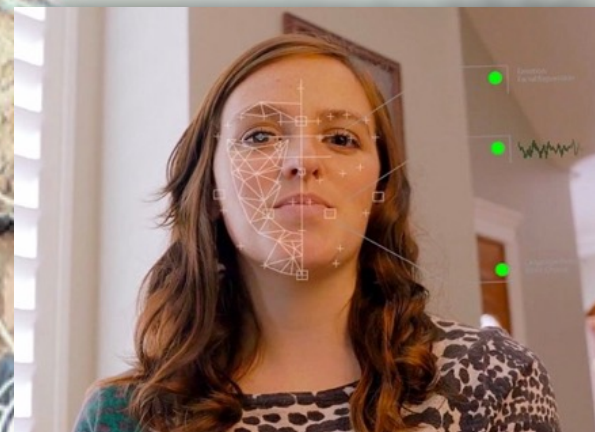
Improving crop yields in India with predictive plant disease diagnosis

Predicting occurrences of diseases earlier and more accurately
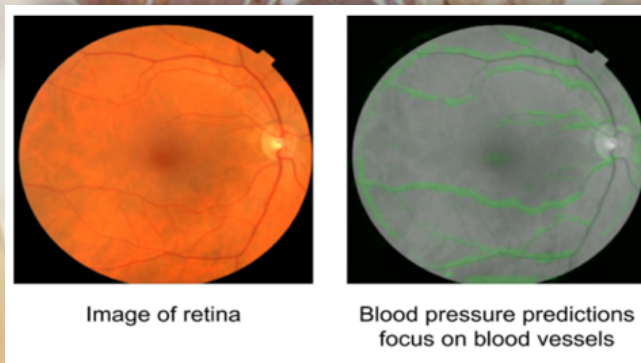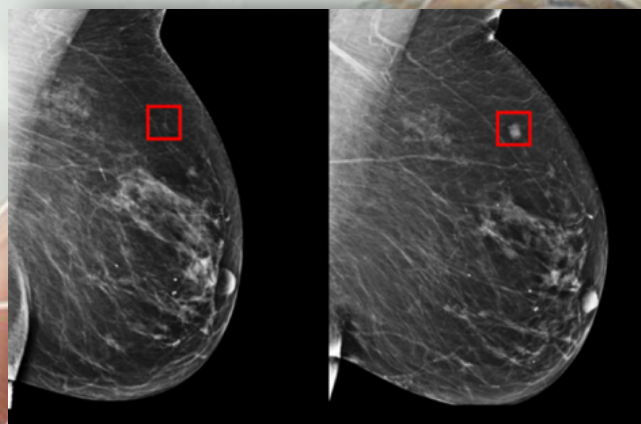
Improving corporate performance – CV screening & supply chain forecasting

Image of retina | Blood pressure predictions focus on blood vessels

The total global uplift on GDP as a result of AI could be over $10 trillion by 2030 per PWC

North America

$3.7 Trillion

15% of GDP

North Europe

$1.8 Trillion

10% of GDP

China

$7.0 Trillion

26% of GDP

https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

Overcoming Racial Bias In AI Systems And Startlingly Even In AI Self-Driving Cars

Racial bias in a medical algorithm favors white patients over sicker black patients

AI expert calls for end to UK use of 'racially biased' algorithms

AI Bias Could Put Women's Lives At Risk – A Challenge For Regulators

Gender bias in AI: building fairer algorithms

Bias in AI: A problem recognized but still unresolved

Amazon, Apple, Google, IBM, and Microsoft worse at transcribing black people's voices than white people's with AI voice recognition, study finds

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

When It Comes to Gorillas, Google Photos Remains Blind

Google promised a fix after its photo-categorization software labeled black people as gorillas in 2015. More than two years later, it hasn't found one.

The Week in Tech: Algorithmic Bias Is Bad. Uncovering It Is Good.

Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech

Artificial Intelligence has a gender bias problem – just ask Siri

The Best Algorithms Struggle to Recognize Black Faces Equally

US government tests find even top-performing facial recognition systems misidentify blacks at rates five to 10 times higher than they do whites.

WILL A ROBOT STEAL **YOUR** JOB?

KILLER COMPUTERS

Bill Gates warns 'dangerous AI' poses a threat 'like nuclear weapons'

AI WARNING:

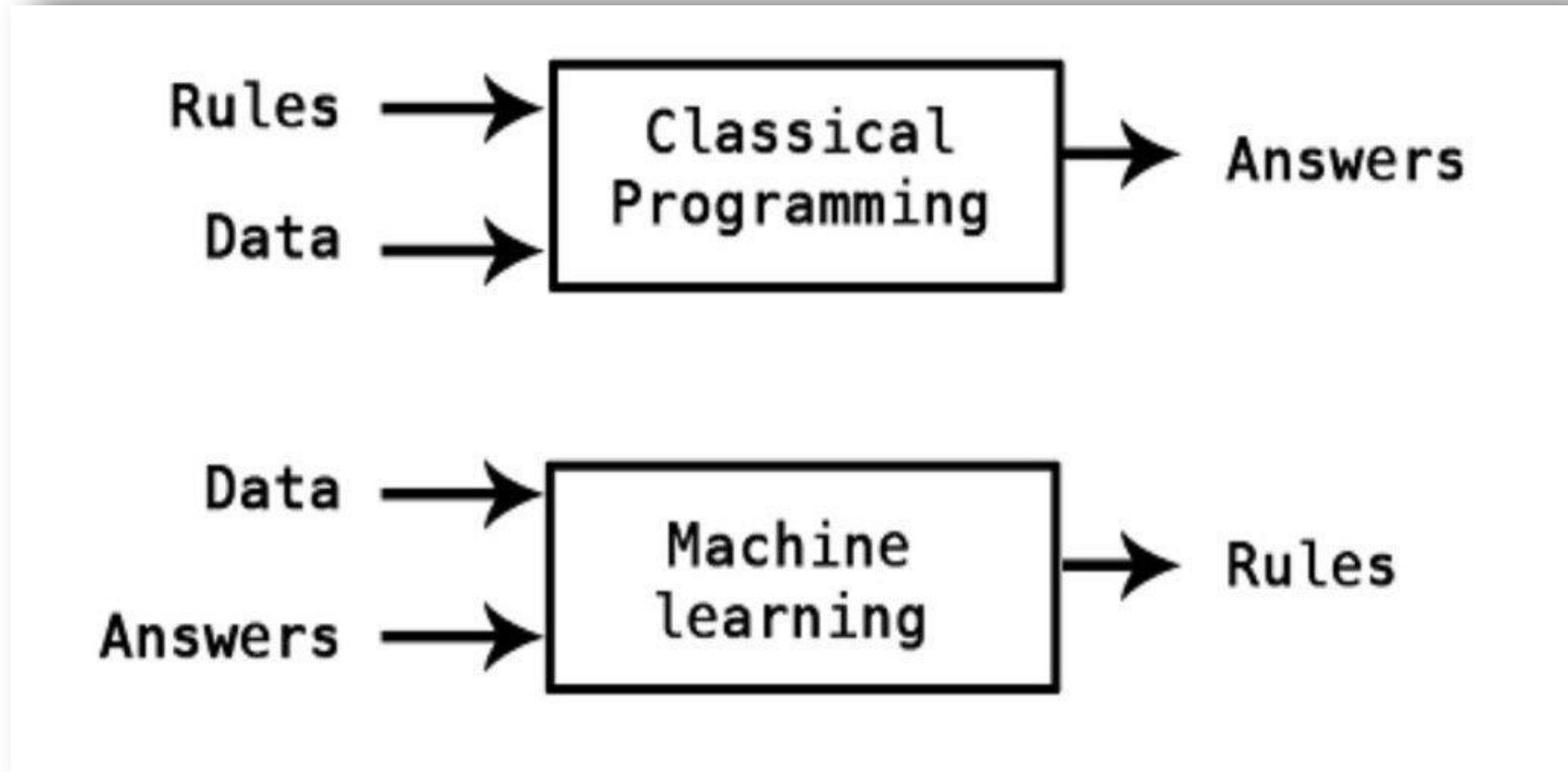Robots will destroy a HUGE number of jobs, claims expert

HIGH QUALITY

AI could be used to **TAKE OVER** the *WORLD* through 'evil' fake news and hijacking cars

We have a lot of fear and not enough trust of AI. Why?

# AI and Machine Learning creates its own "software rules" by learning patterns in data

https://twitter.com/fchollet

And we do not necessarily understand the "rules" it creates
There is often a tradeoff between interpretability and accuracy of explainability

Inferable models, such as decision trees, provide non-technical explanations (e.g. a loan decision)
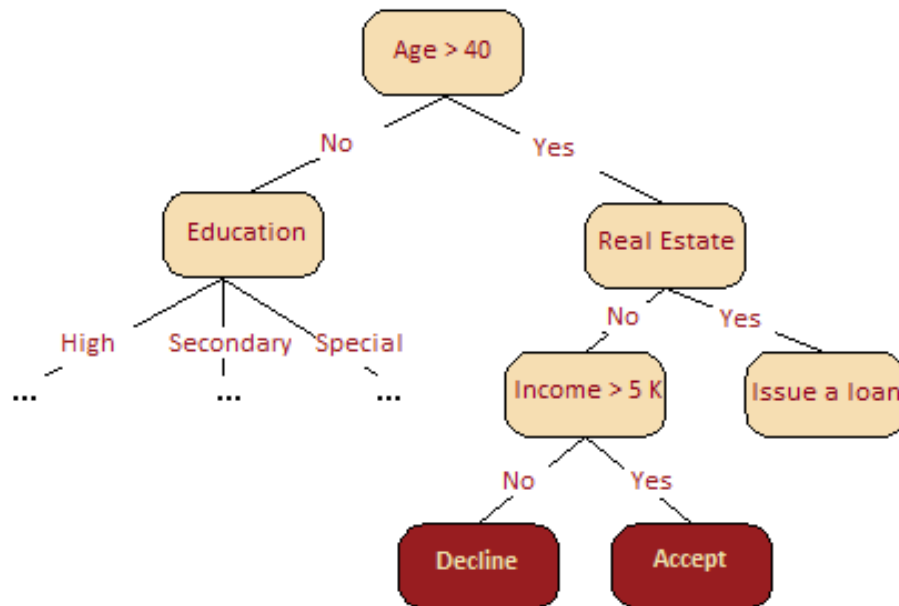
Deep neural networks are often described as black boxes and it is harder to understand their decision making

Best Practice AI
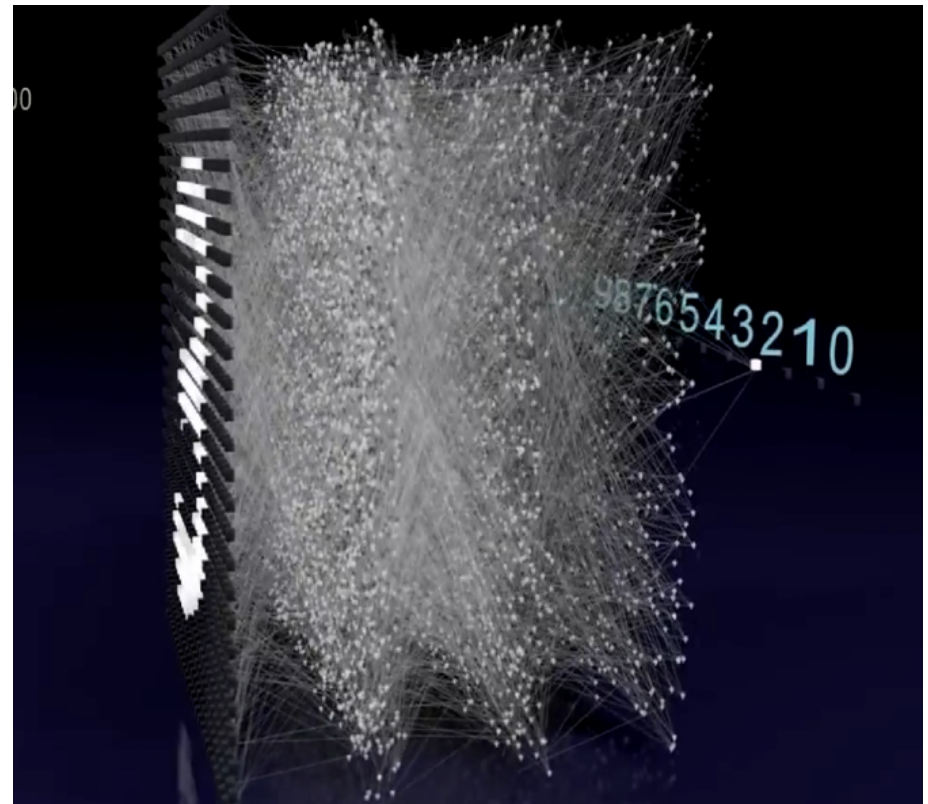
How do you understand a 175 billion parameter deep neural network model? OpenAI's GPT-3 terrified us in 2020 as it wrote "intelligently"

OpenAI 175 billion parameters (June '20)

A robot wrote this entire article. Are you scared yet, human?

We asked GPT-3, OpenAI's powerful new language generator, to write an essay for us from scratch. The assignment? To convince us robots come in peace

- For more about GPT-3 and how this essay was written and edited, please read our editor's note below

▲ 'We are not plotting to take over the human populace.' Photograph: Volker Schlichting/Getty Images/EyeEm

Google 1.6 trillion parameters (Jan '21)

The Machine
Making sense of AI

Google trained a trillion-parameter AI language model

Kyle Wiggers   @Kyle_L_Wiggers   January 12, 2021 10:36 AM

NEW YORK, NEW YORK - OCTOBER 20: Google's offices stand in downtown Manhattan on October 20, 2020 in New York City. Accusing the company of using anticompetitive tactics to illegally monopolize the online search and search advertising markets, the Justice Department and 11 states filed an antitrust case against Google.
Image Credit: Spencer Platt/Getty Images

How open banking is driving huge innovation

Learn how fintechs and forward-thinking FIs are accelerating personalized financial products through data-rich APIs.

Parameters are the key to machine learning algorithms. They're the part of the model that's learned from historical training data. Generally speaking, in the language domain, the correlation between the number of parameters and sophistication has held up remarkably well. For example, OpenAI's GPT-3 — one of the largest language models ever trained, at 175 billion parameters — can make primitive analogies, generate recipes, and even complete basic code.

Best Practice AI

How can we trust AI when it fails?

Deep learning is brittle and lacks human level robustness

It recognises statistical patterns, not higher order concepts and lacks common sense

School Bus	Garbage Truck	Punch Bag	Snow Plough
100%	**99%**	**100%**	**92%**

Tesla's Smart Summons shows the brittleness of pattern recognition and the challenges of a world of edge cases

And when it fails it goes viral on social media

Is AI biased?
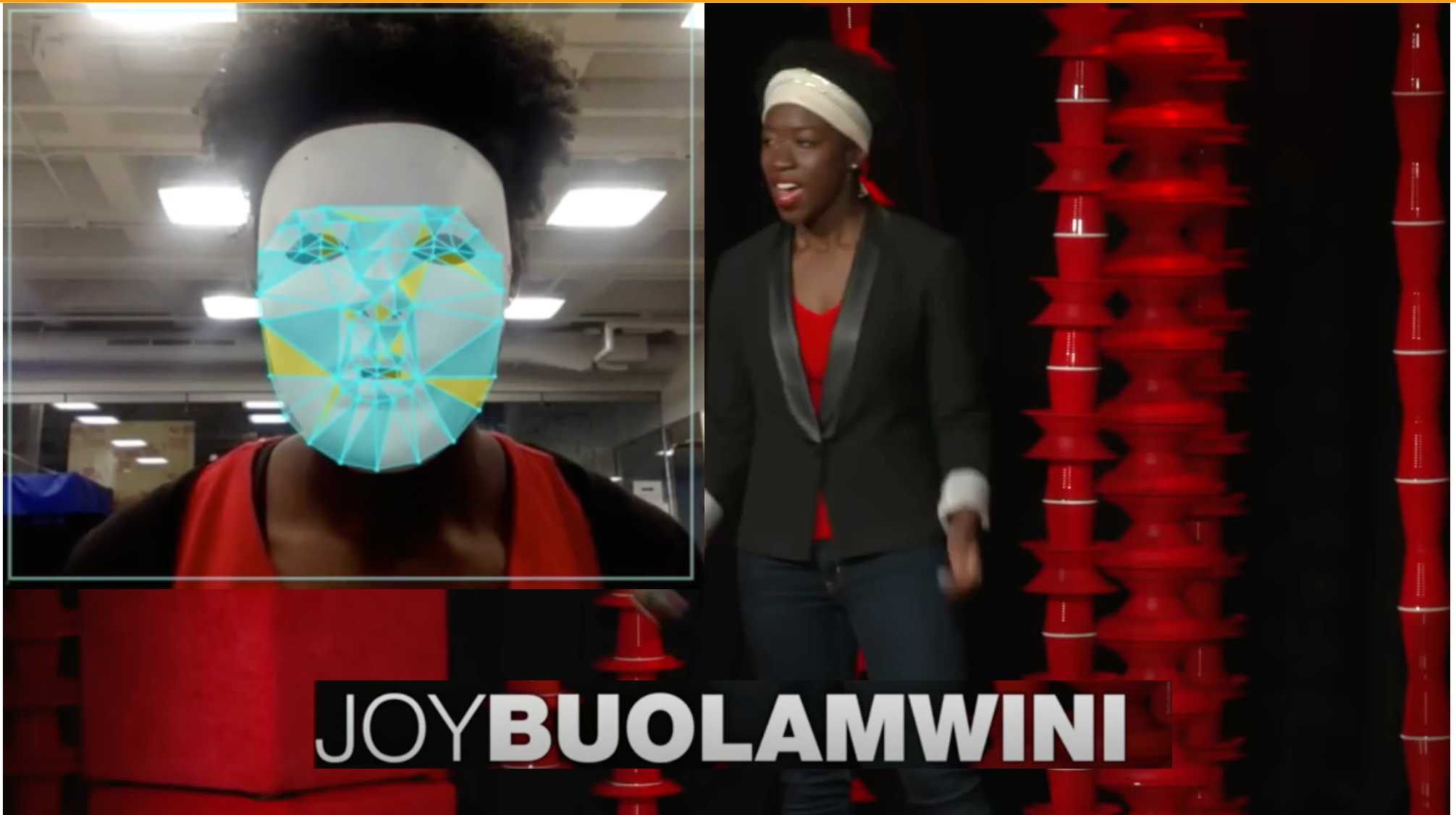
Joy Boualmwini is on a mission to "stop an unseen force that is rising." The risk of bias and discrimination.

JOYBUOLAMWINI

Gender classification systems are often biased as they are not trained on representative sample datasets of gender & ethnicity

Gender was misidentified in up to 1% of lighter-skinned males

Gender was misidentified in up to 7% of lighter-skinned females

Gender was misidentified in 35% of darker skinned females

These biases could have serious ethical, legal, operational and reputational consequences

AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind

Machine learning has the potential to save thousands of people from skin cancer each year—while putting others at greater risk.

ANGELA LASHBROOK   AUGUST 16, 2018
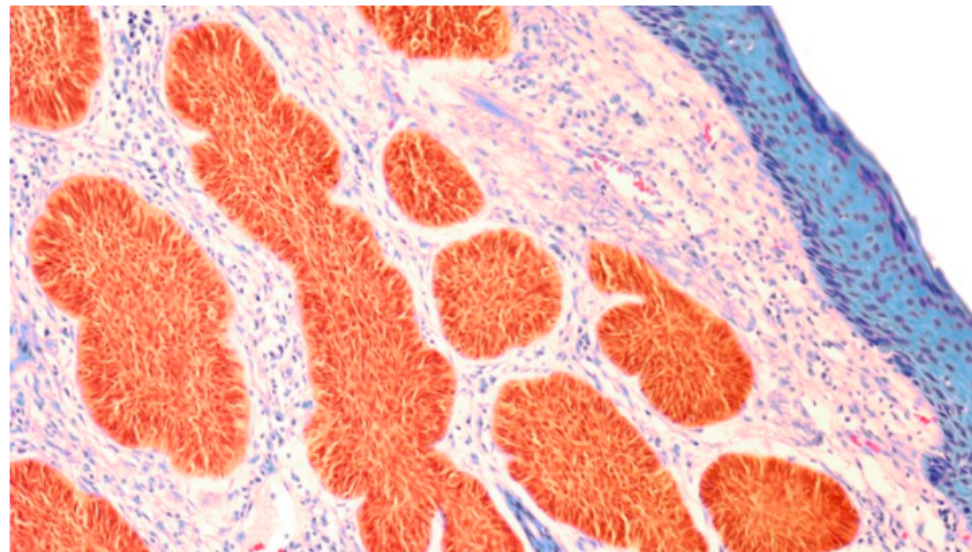
STEVE GSCHMEISSNER / GETTY

LaToya Smith was 29 years old when she died from skin cancer. The young doctor had gotten her degree in podiatry from Rosalind Franklin University, in Chicago, just four

AI shines a spotlight and often amplifies our human biases
What did the algorithm learn after reading 3.5 million books and 11 billion words?

Top 11 positive and negative words associated with females and males

Female	
Positive	**Negative**
beautiful	battered
lovely	untreated
chaste	barren
gorgeous	shrewish
fertile	sheltered
beauteous	heartbroken
sexy	unmarried
classy	undernourished
exquisite	underweight
vivacious	uncomplaining
vibrant	nagging

Male	
Positive	**Negative**
just	unsuitable
sound	unreliable
righteous	lawless
rational	inseparable
peaceable	brutish
prodigious	idle
brave	unarmed
paramount	wounded
reliable	bigoted
sinless	unjust
honorable	brutal

"There is a systemic, systematic, racist, sexist, gendered, class-oriented and other axes of discrimination-bias embedded in most data collected by humans"

Government leaders politicising algorithms doesn't help to engender public trust in AI!

sky news

Boris Johnson blames 'mutant algorithm' for A-level results fiasco, then sacks education chief

The prime minister is accused of "shamelessly trying to avoid taking responsibility" for the exams debacle.

Is our AI technology human centric?

THE TECHNOLOGY
THAT **CONNECTS US**
ALSO **CONTROLS US**

/the
social
dilemma_

| **NETFLIX**
thesocialdilemma.com

Do we understand the (unintended) consequences of our algorithms?

200+ "Ethical AI" frameworks from the likes of the OECD and the IEEE boil down to similar principles

1) **Explainable** and transparent decision making

2) **Inclusive**, diverse and **fair** (avoid or don't reinforce bias)

3) Be built and tested for **safety**

4) Be **socially** beneficial

5) **Respect** human rights and the law

6) People are **accountable**

AI regulation is already here under GDPR: Explainability

Article 22 under GDPR states:

Fully automated decisions with legal effect or similarly significant effect needs to be explainable

and

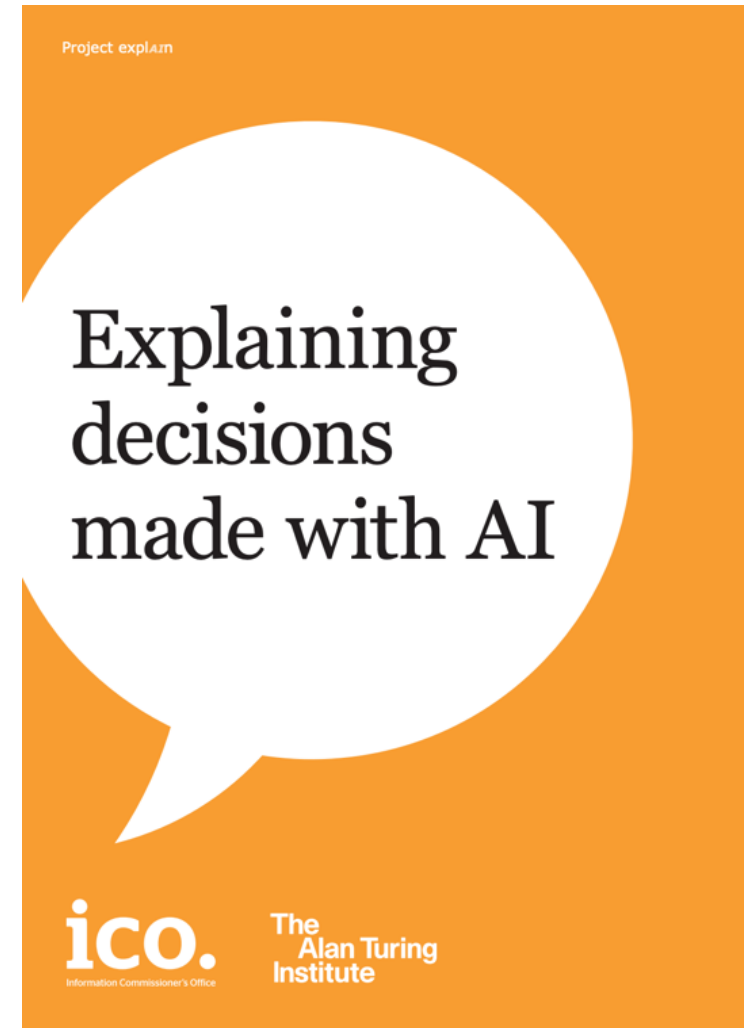data subjects have the right to human-made decisions

Explainability will spotlight company historical practices

Information Commissioners Office (ICO) – requires an Explainability statement for automation with legal effect

The ICO identified six main types of explanation that form an explanatory statement:

1. **Rationale** explanation: the reasons that led to a decision, delivered in an accessible and non-technical way.

2. **Responsibility** explanation: who is involved in the development, management and implementation of an AI system, and who to contact for a human review of a decision.

3. **Data** explanation: what data has been used in a particular decision and how; what data has been used to train and test the AI model and how.

4. **Fairness** explanation: steps taken across the design and implementation of an AI system to ensure that the decisions it supports are generally unbiased and fair, and whether or not an individual has been treated equitably.

5. **Safety** and performance explanation: steps taken across the design and implementation of an AI system to maximise the accuracy, reliability, security and robustness of its decisions and behaviours.

6. **Impact** explanation: the impact that the use of an AI system and its decisions has or may have on an individual, and on wider society.

Project explAIn

Explaining decisions made with AI

ico.
Information Commissioner's Office

The Alan Turing Institute

Companies are implementing Responsible AI programmes, often as a competitive differentiator

- **Deutsche Telekom view digital ethics as a strategic competitive differentiator**
- **They focus on shaping AI responsibly, with programme implemented on various levels**
- **Continuously updated**

INTERNAL PROCESSES
Integration into internal security and data protection processes; integration into financing processes

DIGITAL ETHICS CENTER
Place for internal/external conferences to make Digital Ethics tangible

DIGITAL ETHICS SEAL
First internal AI projects have been certified and received a Digital Ethics Seal

SUPPLIER MANAGEMENT
Guidelines extended to suppliers of AI systems

COMMUNICATION & EVENTS
Regular communication and expert interviews on the intranet; AI days and other formats

TRAININGS & FORMAL POLICY
eLearnings, roadshow, workshops and policy for employees to develop safe AI systems

Deutsche Telekom

Best Practice AI

Can the identification of bias by AI help us create a more just society?

- A recent paper looked at disparities in the treatment of knee osteoarthritis that causes chronic pain

- Radiologists review x-rays of the knee and score the patient's pain based on radiographic features (e.g. degree of missing cartilage or structural damage

- But there is a gap between the radiologist prediction of pain and self-reporting pain of black patients

- Why? The methodology used - the Kellgren-Lawrence (KLG) grade - was developed several decades ago based on a white British population

- ML used to predict pain in black patients was much more accurate – medical methods need to be updated.
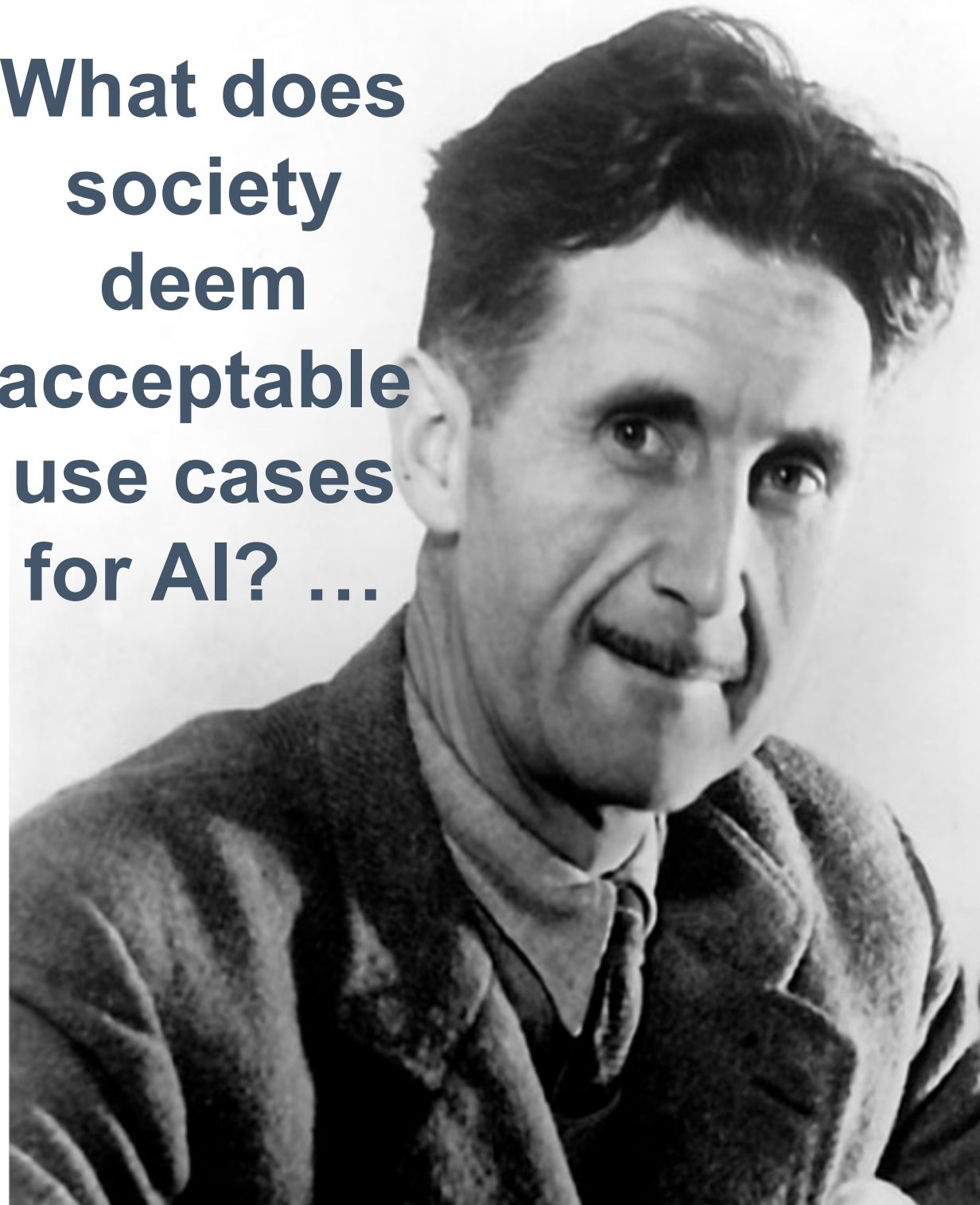

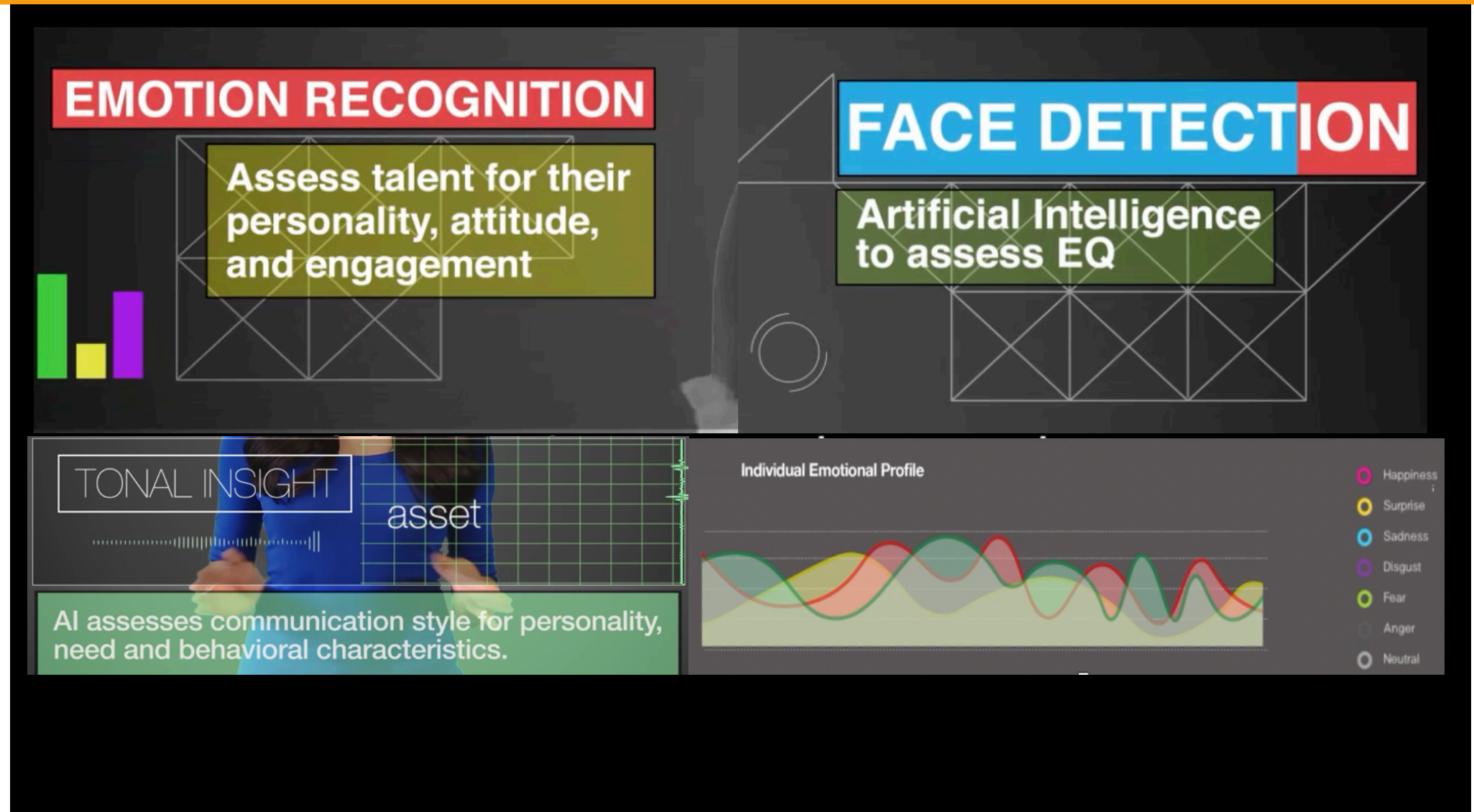
Grade 1.   Grade 2.
A   B
C   D
Grade 3.   Grade 4.
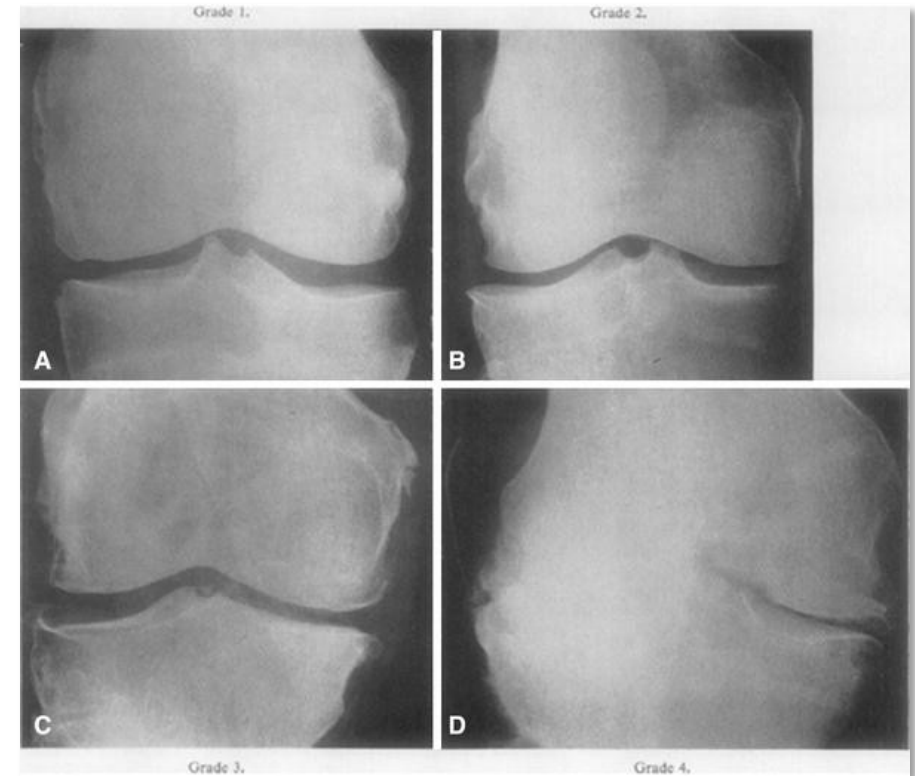
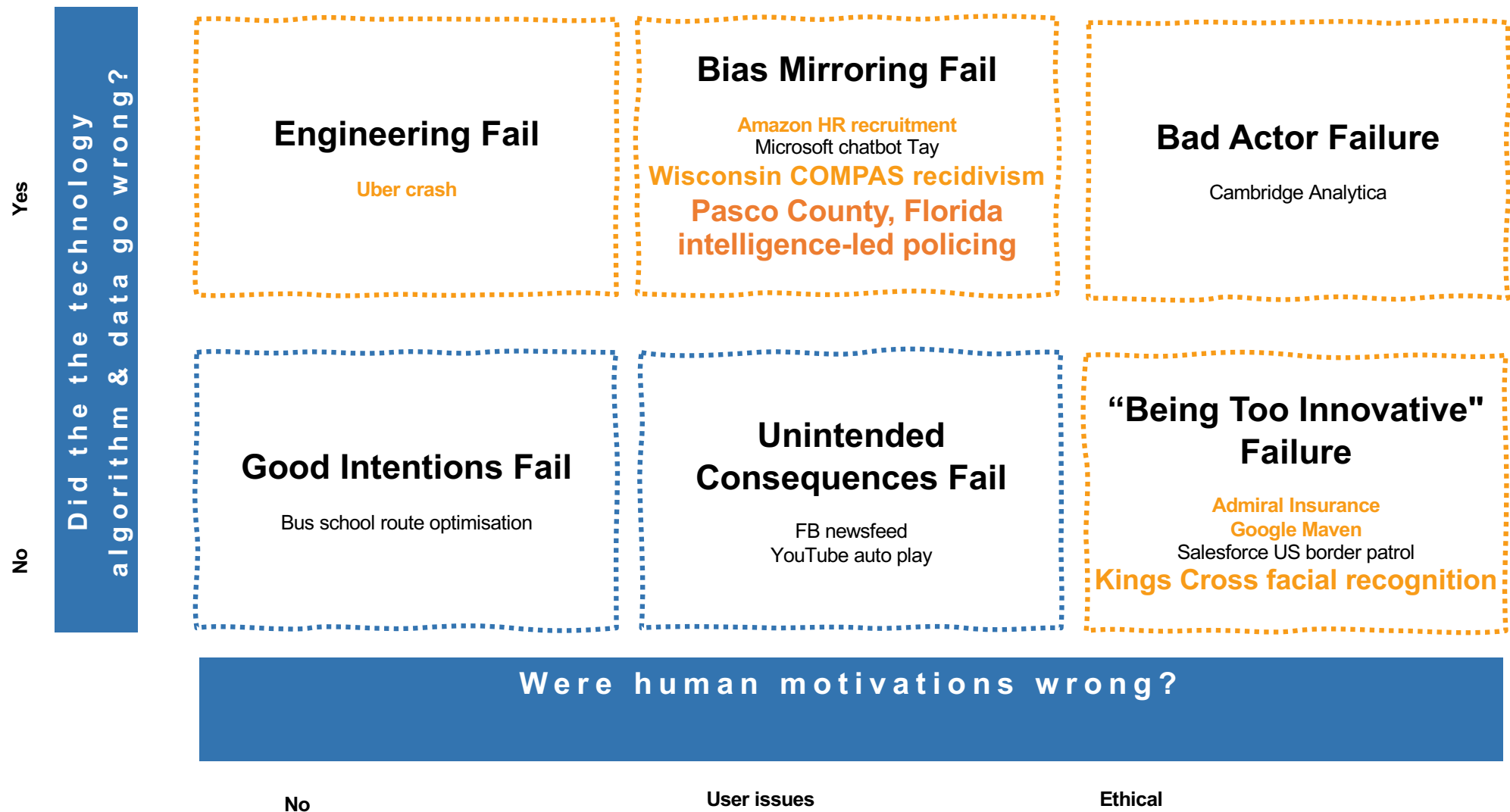**What does society deem acceptable use cases for AI? …**

What about algorithmic monitoring and assessment?

How about we save time in recruiting with automated video interviews?

AI ethics is often more a question of human motivations than technology failings

Did the the technology algorithm & data go wrong?

Yes

Engineering Fail

Uber crash

Bias Mirroring Fail

Amazon HR recruitment
Microsoft chatbot Tay
Wisconsin COMPAS recidivism
Pasco County, Florida intelligence-led policing

Bad Actor Failure

Cambridge Analytica

No

Good Intentions Fail

Bus school route optimisation

Unintended Consequences Fail

FB newsfeed
YouTube auto play

"Being Too Innovative" Failure

Admiral Insurance
Google Maven
Salesforce US border patrol
Kings Cross facial recognition

Were human motivations wrong?

No   **User issues**   **Ethical**

Best Practice **AI**

AI ethics is often more a question of human motivations than technology failings

Did the technology, algorithm or data go wrong?

Yes

No

No   Yes

Were human motivations wrong?

Best Practice AI

AI surveillance helped many countries successfully manage COVID-19. Should we use AI monitoring in our country?
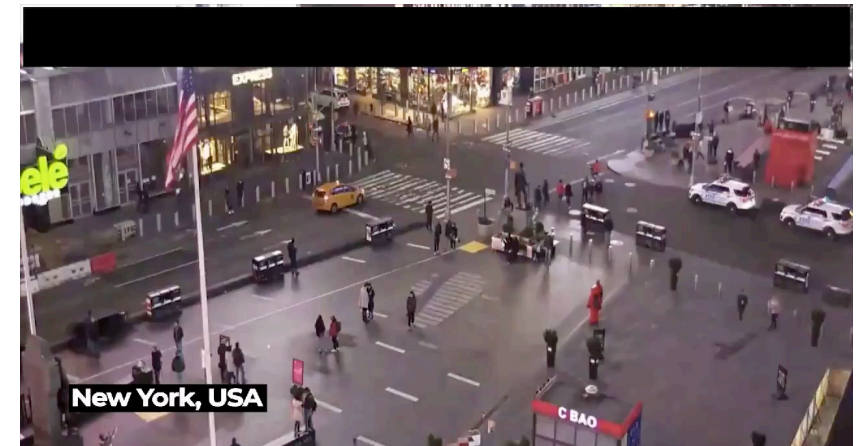
Thermal screening in stations and airports

Thermal cameras used on on drones to identify fevers and crowds

FRT to detect those wearing masks & enforce quarantines

Enforcing social distancing

What if we could understand what everyone was saying in our company? Should we?

What if….

- **All the conversations taking place in your organisation were tracked**
  - Emails, slack, social media, telephone, messaging, meetings…
  - Customers, teams, suppliers, feedback sessions, internal debate….

- **What value could your organisation capture if these were processed and assessed**
  - Pain points and customer concerns identified, issues logged, ranked and prioritised…
  - Sentiment analysed, management issues spotted, individual performance assessed…
  - Personalised feedback, abuse monitoring, behaviour nudging, welfare protected…

- **But at what cost?**
  - Privacy, nuance, personal style…
  - Transparency versus reality…

- **Who has the power – and where might this happen?**

But how about AI in education? Should we monitor students in the classroom?

- AI can help education. For example, **VIPKid** is used by 700,000+ students. Its AI animated "fun characters" assist human teachers. Correct answers grew 50 to 80%.

- Chinese **Megvii** is a $4B+ company known for its facial recognition platform Face++

- Demonstrated a **classroom teaching evaluation system** used to supplement teaching evaluations through **real-time structured analysis** of classroom video data

- Observes and **classifies** student behaviour:
    - resting on one's desk
    - playing on a mobile phone
    - sleeping
    - listening to a lecture
    - reading
    - raising hands
    - concentration levels...

A recent report identified 10+ Chinese companies in the emotion recognition market for education

EF Children's English 英孚少儿英语	In person and online	Partners with Tencent Cloud to conduct image, emotion, and voice recognition, and receives curriculum design assistance to EF's product-development teams and teachers.[141]
Hanwang Education 汉王教育	In person	Class Care System (CCS) cameras take photos of whole classes once per second, connect to a programme that purportedly uses deep-learning algorithms to detect behaviours (including 'listening, answering questions, writing, interacting with other students, or sleeping') and issue behavioural scores to students every week. Scores are part of
New Oriental 新东方	Blended learning	AI Dual Teacher Classrooms contain a 'smart eye system based on emotion recognition and students' attention levels', which the company says can also detect emotional states, including 'happy, sad, surprised, normal, and angry'.[149] A
Hikvision 海康威视	In person	Smart Classroom Behaviour Management System integrates three cameras, positioned at the front of the classroom, and identifies seven types of emotions (fear, happiness, disgust, sadness, surprise, anger, and neutral) and six behaviours (reading, writing, listening, standing, raising hands, and laying one's head on a desk).[144] Cameras take attendance using face
Meezao 蜜枣网	In person	Uses facial expression recognition and eye-tracking software to scan preschoolers' faces over 1,000 times per day and generate reports, which are shared with teachers and parents.[147] Reports contain data visualisations of students'
Taigusys Computing 太古计算	In person	Collects data from three cameras, one each on students' faces, teachers, and a classroom's blackboard. The system detects seven emotions (neutral, happy, surprised, disgusted, sad, angry, scared) and seven actions (reading, writing, listening, raising hands, standing up, lying on the desk, playing with mobile phones).[151]

https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf

And many Chinese vendors are offering emotion recognition and monitoring for "public security." Any concerns?

Alpha Hawkeye 阿尔法鹰眼	Monitors vestibular emotional reflex and conducts posture, speech, physiological, and semantic analysis.[65]	• Airport, railway, and subway station early-warning threat detection • Customs and border patrol
CM Cross 科思创动	Employs deep-learning-powered image recognition to detect blood pressure, heart rate, and other physiological data.	• Customs and border patrol[67] • Early warning
EmoKit 翼开科技	EmoAsk AI Multimodal Smart Interrogation Auxiliary System detects facial expressions, body movements, vocal tone, and heart rate.[68] Other products detect similar data for non-interrogation	• Detecting and managing mental-health issues at medical institutions • Loan interviews at banks • Police-conducted interrogations[69]
Joyware 中威电子 NuraLogix	NuraLogix's DeepAffex is an image recognition engine that identifies facial blood flow (which is used to measure emotions) and detects heart rate, breathing rate, and 'psychological pressure'.[71] Joyware also uses NuraLogix's polygraph tests.[72]	• Airport and railway station surveillance • Nursing • Psychological counselling
Miaodong 秒懂	Relies on image recognition of vibrations and frequency of light on faces, which are used to detect facial blood flow and heart rate as a basis	• Police interrogation
Sage Data 睿数科技	Public Safety Multimodal Emotional Interrogation System detects micro-expressions, bodily micro-	• Police and court interrogations
Shenzhen Anshibao 深圳安	Emotion recognition product detects frequency and amplitude of light vibrations on faces and	• Early warning[76] • Prevention of crimes and acts
Taigusys Computing 太古计算	One product is referred to as a micro-expression-recognition system for Monitoring and Analysis of Imperceptible Emotions at Interrogation Sites, while others include 'smart prison' and 'dynamic emotion recognition' solutions. Taigusys claims to use image recognition that detects light vibrations on faces and bodies, as well as parallel computing.[77]	• Hospital use for det Alzheimer's, depres attacks[78] • Police interrogation criminals[79] • Prison surveillance

https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf

As a society we need to agree the responsible limits on facial recognition and other high risk uses of AI technologies

WORLD ECONOMIC FORUM

Facial recognition technology: The need for public regulation and corporate responsibility

Jul 13, 2018 | Brad Smith - President

Microsoft

IBM
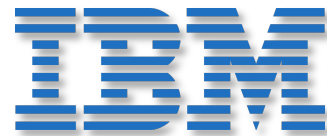
IBM quits facial-recognition market over police racial-profiling concerns

CEO writes to US Congress calling for 'national dialogue' about use in law enforcement
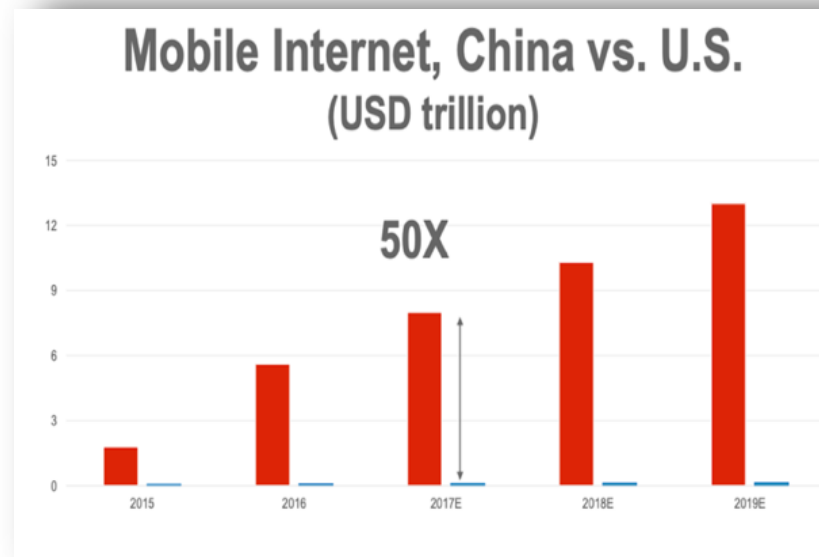
There is a leadership race for AI across nation states.

Does this impact AI ethics?

China is an AI super-power with an ambition to be a global leader in AI by 2030

AI Super-Powers, Dr Kai Fu-Lee
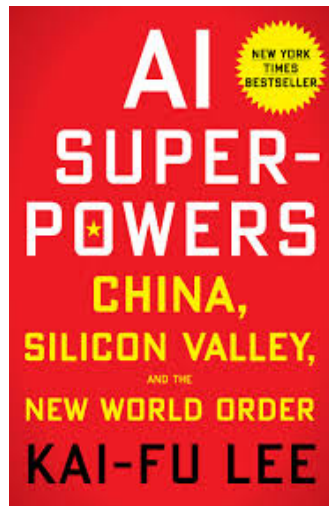
"China is the Saudi Arabia in data"

"I left out Europe..."

"I left out Europe because I didn't think there was a good chance for it to take even a so-called 'bronze medal' in this AI competition. European artificial intelligence is <u>losing the race</u>."

Chinese structural advantages include data privacy regulations, public - private cooperation, and scale of data, capital, talent and market demand.

"Create an AI ecosystem of trust." Ursula von der Leyen, Davos Agenda Jan '21.

"…Some of us are deeply **concerned** about the role which will be left for human beings in a **world run by AI**."

"Others worry about the **serious effects that algorithms** can have on the health of our **democracies**. Who is taking the final decisions? Who is steering the flow of information?

"What we see through social media platforms seems real….we literally live in different worlds."

"Yes, algorithms can be a danger to our democracy. But they do not have to be…There must be at least **transparency** on how the algorithm works….

"For people to accept a role for AI in such decisions, they must be **comprehensible**.

And they must respect **people's legal rights…**we have to be able to examine the workings of the system and to ensure **human oversight**.

Our aim is to create an AI **ecosystem of trust**.

"Create an AI ecosystem of trust." Ursula von der Leyen, Davos Agenda Jan '21.

"What sets Europe apart from competitors like China is not the size of our tech sector or how attractive we are for venture capital. What sets Europe apart from competitors is the fact that our values come first. Human beings come first."

Thank you. Best Practice AI

Contact us:

✉ **Simon Greenman**
Simon @ bestpractice . ai

🐦 **@BestpracticeAi**

📞 **+44 (0) 7 8 2 4 5 5 7 9 7 9**
www.bestpractice.ai

Strategy, Business Planning & Education

AI Delivery Support & Coaching

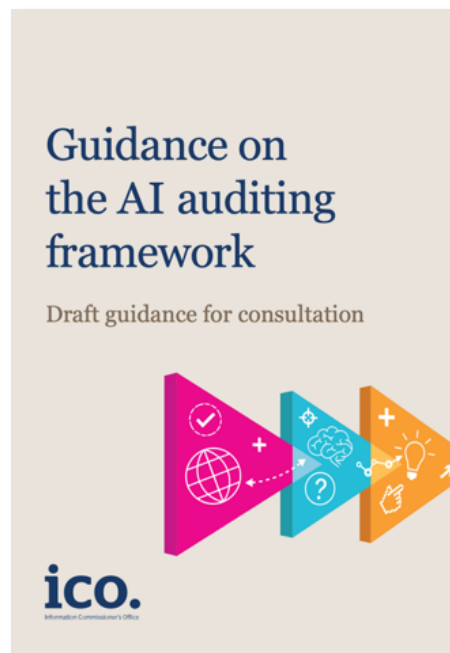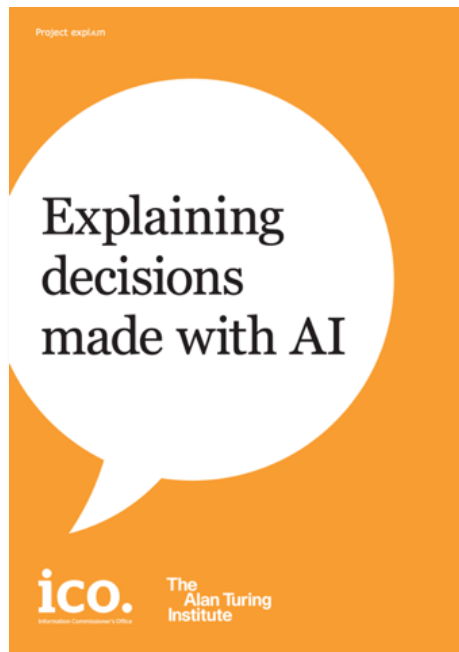Organisational Design and Capacity Building

AI Governance & Risk Management
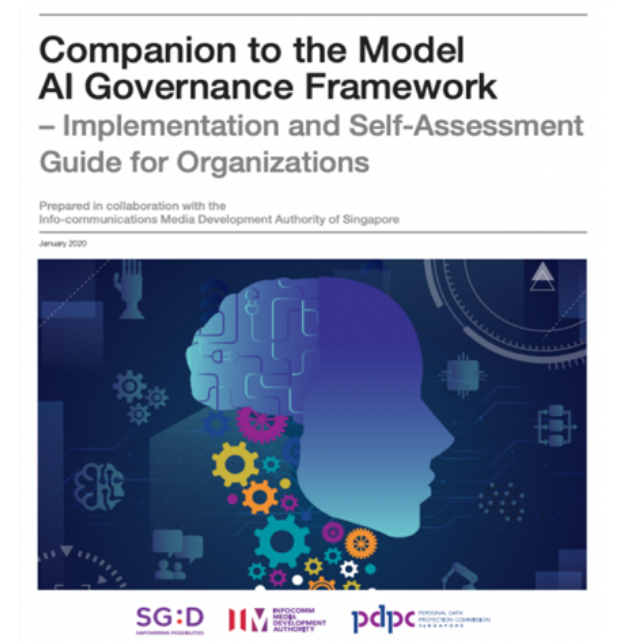
Investment Due Diligence

Best Practice **AI**

Further reading on AI governance
Ensure AI explainability, auditability within a governance framework

UK Information Commissioner's Office (ICO) Guidance

Model AI Governance Framework

(launched Davos 2020)