



7080MAA

AI Ethics

Dr. Huma Shah

Directed Science on EU Horizon2020 CSI-COP Research and Innovation project

<https://www.csi-cop.eu/>



In today's 7080MAA Lecture and Tutorial

- We will explore AI through *I > AI > to ethical issues in nascent AI*
- Recommended reading:
 - Alan Turing's 1950 paper '***Computing Machinery and Intelligence***', published in *Mind* journal and accessible through [Locate](#)
 - An old but interesting 1976 article by **Drew McDermott**: '***Artificial Intelligence meets Natural Stupidity***' can be accessed here:

[https://www.researchgate.net/publication/234784524 Artificial Intelligence meets natural stupidity](https://www.researchgate.net/publication/234784524_Artificial_Intelligence_meets_natural_stupidity)

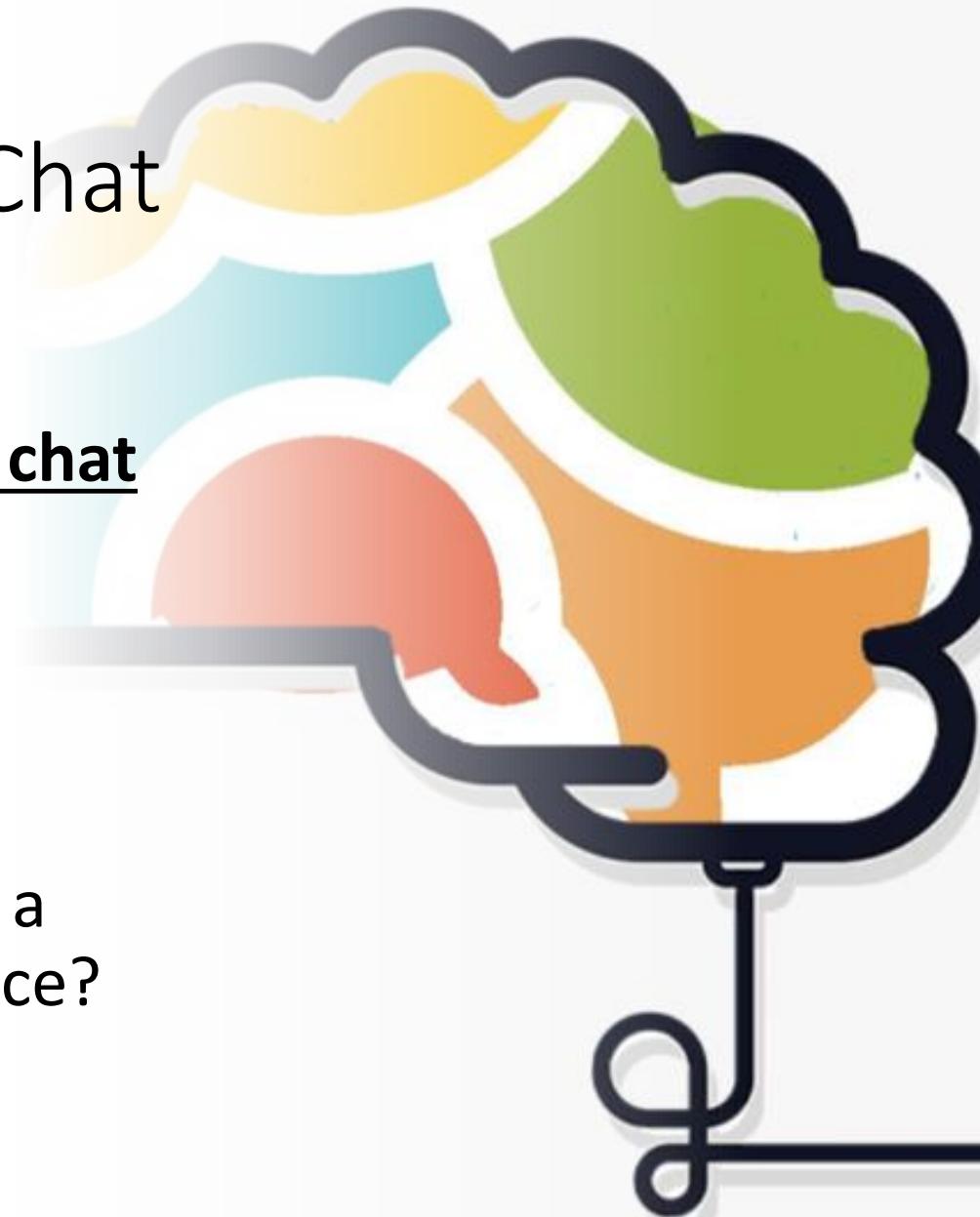
7080MAA Class today: 3 main themes

- **Human & Animal Intelligence (*I*): what is it?**
- **Artificial Intelligence (AI): how do we build it?**
- **Ethics**: what is the societal impact of AI?

In MS Teams Chat

Please share in the chat
features or
characteristics of
intelligence

- e.g., are emotions a
feature of intelligence?



One feature of intelligence...

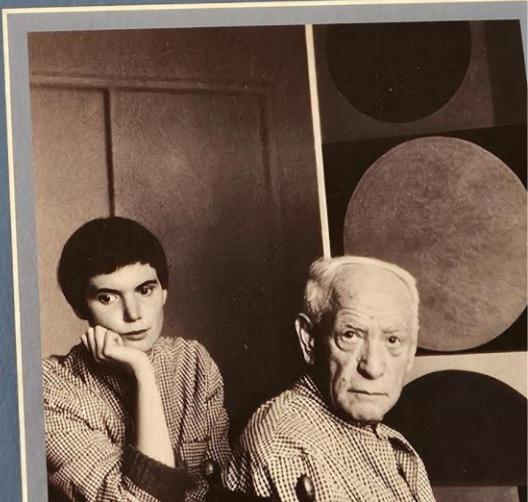
- On a recent trip to Paris visiting 20thC artist Auguste Herbin (1882-1960) exhibition in the Muse de Montmartre, this definition for creativity was displayed:

"Creativity is the expression of laws that are at the origin of all things and that are within us".

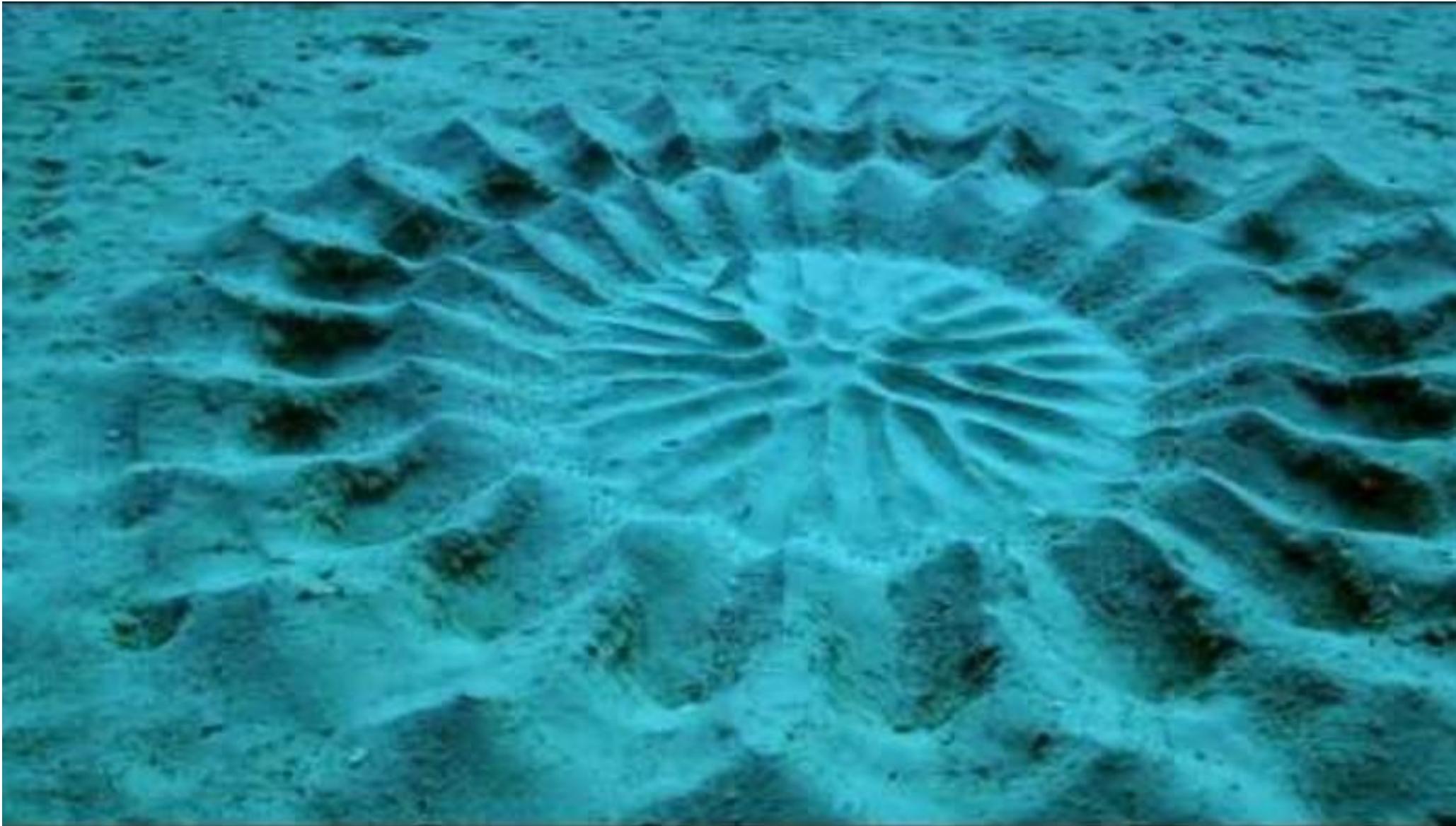
Herbin created the [hashtag#PlasticAlphabet](#) - an abstract code of colourful shapes as a "table of relationships comprising five geometric shapes, a set of colours, the twenty-six letters of the alphabet and the seven notes of the C major scale"



7080MAA: AI Ethics, Dr. Huma Shah 07.10.24



Puffer Fish Creates This Blue Water Art ~: } - YouTube





Intelligence (I)

- What we humans are. We might be clever at different things, but we are *intelligent*
- How do we acquire intelligence? Nature & nurture.
- Alan Turing, 20thC British mathematician and WWII codebreaker: look at how children learn

Intelligence

One view of Intelligence

- Alan Turing (1948):
 - “..idea of intelligence is emotional, rather than mathematical”

One view of Intelligence

- Turing (1948):
 - “The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration”

Earlier view: Francis Galton

- **Francis Galton**, late 19th century view:
 - “..breed brighter humans”
 - “..eminent fathers were more likely to have eminent sons” (in Gardner, 1996: p. 42).

Earlier view: Francis Galton

- Francis **Galton**, late 19th century view:
 - intelligence “**inherited**” and “**reputation was a good measure of ability**” (in Gardner, 1996: p. 43).

Question:

Do you agree with the view?

- In MS Teams chat, please post YES or NO to the question on whether you **agree with or disagree with** Galton's view on *inheritance and reputation*:
 - 'intelligence “**inherited**” and “reputation was a good measure of ability”'

Testing for intelligence

- Charles Spearman's 1904 theory of general intelligence (*g*) involved a quest for its objective measurement.

Intelligence testing - Galton

- Galton invented the “**mental test**” a **sensory measure**, such as “the highest pitch a person could hear, and how well he or she could distinguish small differences in weights, colors, smells, tactile stimuli, and the length of lines”

Intelligence testing – Binet & Simon

- Alfred **Binet** and Théodore **Simon** began testing schoolchildren in France between 1904 and 1911.

Intelligence testing – Binet

- Binet studied individual differences
...interested in the development of thinking
in children ...measuring ability led to better
educational decisions based on the
information

Intelligence and development: Piaget

- Biologist Jean **Piaget**'s seminal work created a clinical method to investigate the development of intelligence in children:
 - “... infants were born without substantive knowledge but with definite means for coming to know the world” (in Gardner et al. 1996: p.105).
 - **Four stages or milestones in the development of humans**

Testing intelligence - Terman

- Lewis **Terman** “adapted Binet’s test for use by American school children”.

Testing intelligence - Terman

- Terman: “The children of successful and cultured parents test higher than children from wretched and ignorant homes for the simple reason that their heredity is better” (in Gardner et al., 1996: p. 51).

Testing intelligence - Stern

- William **Stern** applied a “mental age” system to “compare children of the same chronological age”. (in Gardner et al., 1996: p. 50).

Testing intelligence - Stern

- Stern devised the *intelligence quotient* or **IQ**.

Post-testing view: Francis Crick

- Francis **Crick** 2007 Nobel prize-winner and DNA pioneer:
 - “black people ... less intelligent than white people”
 - .. “the idea that ‘equal powers of reason’ were shared across racial groups was a delusion”

Post-testing view: Francis Crick

- **Crick:**

- “Western policies towards African countries were wrongly based on an assumption that black people were as clever as their white counterparts when "testing" suggested the contrary.”

Post-testing view: Francis Crick

- **Crick:**

- “All our social policies are based on the fact that their [black people] intelligence is the same as ours
- whereas all the testing says not really”

Intelligence testing and scientific method

- H.J Eysenck (1985):
 - “Intelligence tests are not based on any very sound scientific principles
 - “..not a great deal of agreement among experts regarding the nature of intelligence”

(p. 8 in 1985 reprint of book *Know your own IQ*, Penguin books).

Other species intelligence

Dog protecting child

Human vs Animal Intelligence

- Humans share “roughly 99 per cent of our genomes with chimpanzees and bonobos” ...humans are referred to as “another great ape” (Saini, 2011: p.128).

Human vs Animal Intelligence

- Primate behaviour scientist **Jane Goodall** observations of **chimpanzees** in the wild has shown that these primates are “**skilled at using simple tools and cracking nuts with hard shells**”

Communication, Language and intelligence

Human vs Animal Communication

- “Members of the animal kingdom have the ability to communicate through vocal signs, facial expressions” (Fischler and Firschein, 1987 p. 157)

Human vs Animal Communication

- “... in animal communication the language is very limited, restricted to a number of sounds associated with signalling danger, establishing territory, indicating anger, etc.” (Fischler and Firschein, 1987: p.158)

Human vs Animal Communication

- Language is a set of vocal or written signs and symbols that permits a social group to communicate, and facilitates thinking and actions of individuals” (p. 157)

Human vs Animal Communication

- “Language, in the full sense of the term, is species-specific to [hu]man...”

A modern view of intelligence: *processing information*

Information processing view

- Joseph Fagan: information processing as an “access to opportunity” to enter the world of knowledge.

Intelligence - Fagan

- Fagan (2000: p. 168):
 - “... consider the implications for society of defining intelligence as processing ...controversy surrounding the term *intelligence* has arisen and continues because **intelligence has historically been defined as how much one knows rather than as how well one processes.**”
{think about how conspiracy theories are believed}

Intelligence - Fagan

- Fagan (2000: p. 168):
 - IQ scores, by convention, are based on how much one knows relative to one's age peers.”

Information processing & knowledge

- **Fagan:**

- “...**intelligence is processing** and that processing can be **measured by performance on certain elementary cognitive tasks**. An **IQ score**, on the other hand, **depends** not only on processing ability but **on what one has been taught**.”

Information processing & knowledge

- Fagan:

- “...information is processed, the mind changes.”
- “That change is called knowledge”. {But what about misinformation, and how *that* is processed?}

AI – what is it?

- In **AnswerGarden** at the link below, please share a skill(s) that an *AI* should have:
- <https://answergarden.ch/4175257>



Origin of AI

- Humans have considered intelligent automata for centuries
- The modern term ‘artificial intelligence’ emerged from the scholarship into *intelligent machinery* by 20th century mathematician and Bletchley Park code-breaker, **Alan Turing**.
- Turing considered:
 - Can a *machine think?*
 - Posited the *Imitation Game* – popularly known as the Turing test
 - Proposed a methodology -*Indistinguishability* : can a machine’s responses to any questions be indistinguishable from the kinds of responses a human would give?

Intelligent Machinery >

Artificial Intelligence

Arrival of Alan Turing on the scene with his innovative ideas that gave birth to the science of *artificial intelligence*

Turing on language

- Turing's 1948 essay *Intelligent Machinery*:
 - "...first manifesto of artificial intelligence"

Turing on language

- Turing's 1948 essay *Intelligent Machinery*:
 - “... learning of languages an impressive human activity”

Turing test for intelligent machinery

- Turing explored intelligent machinery through ability of a machine **to think**



What is thinking?

Turing test for intelligent machinery

- Turing: Thinking is *buzzing in my head*

Avoid definitions

- Turing had no definition for intelligence or machine intelligence (see Plato's *Laches* dialogue, or read W.V.O. Quine)

Problem with definitions

- Turing proposed practical way to assess intelligence through Imitation Game, popularly known as the **Turing test**.

Learning: *Imitation*

- Learning by imitation – rather than looking at the adult mind, look at how children learn

Learning - Imitation

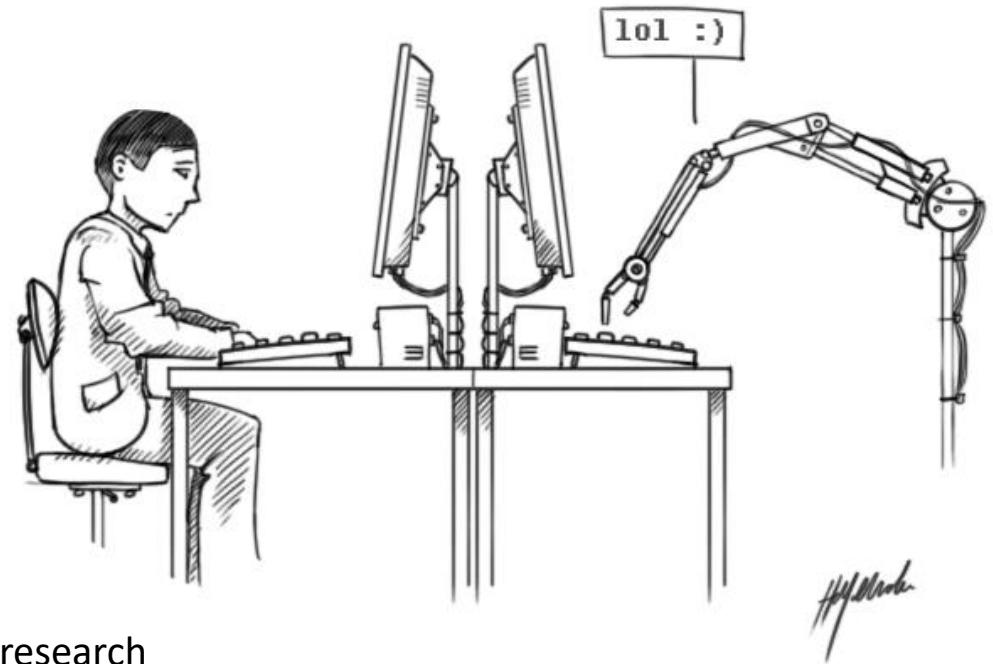
- Turing test measures a machine's intellectual capacity/ to answer any question in a “**satisfactory**” and “**sustained**” way (1950) by **imitating** the sorts of answers a **human would give** to questions

Turing's two tests for machine thinking

Simultaneous Comparison



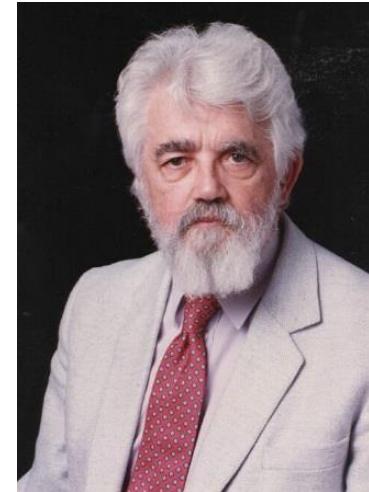
Viva Voce



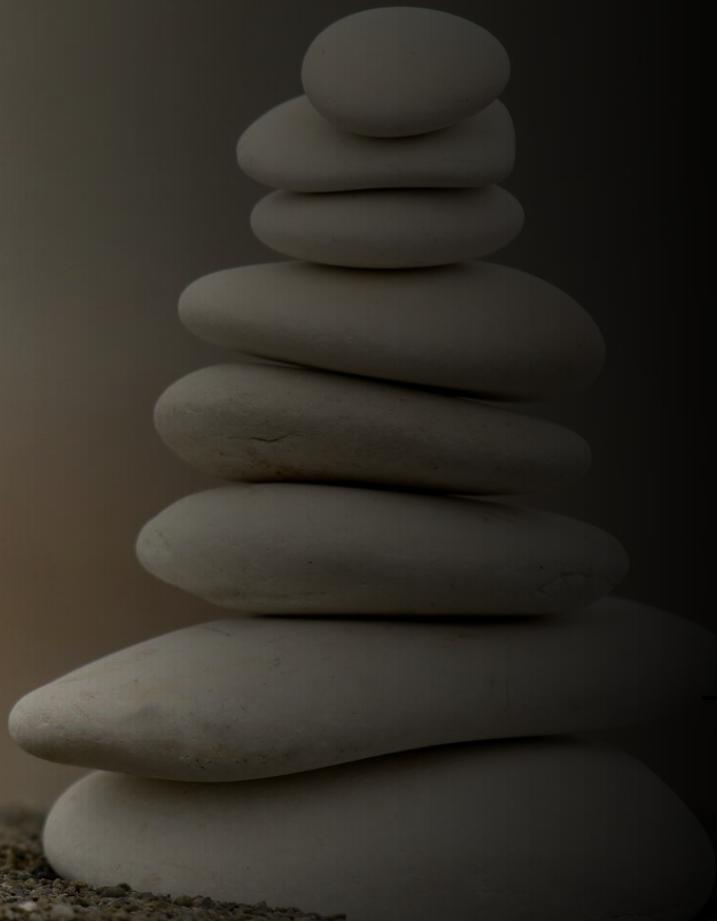
© Images from Huma Shah's PhD and her AI research

Birth of AI

- Artificial Intelligence (AI) coined by computer scientist **John McCarthy** in **1955**



- Emerged as a field of investigation in 1956 from a Dartmouth Artificial Intelligence Conference



Machine Consciousness

Question: what is consciousness?

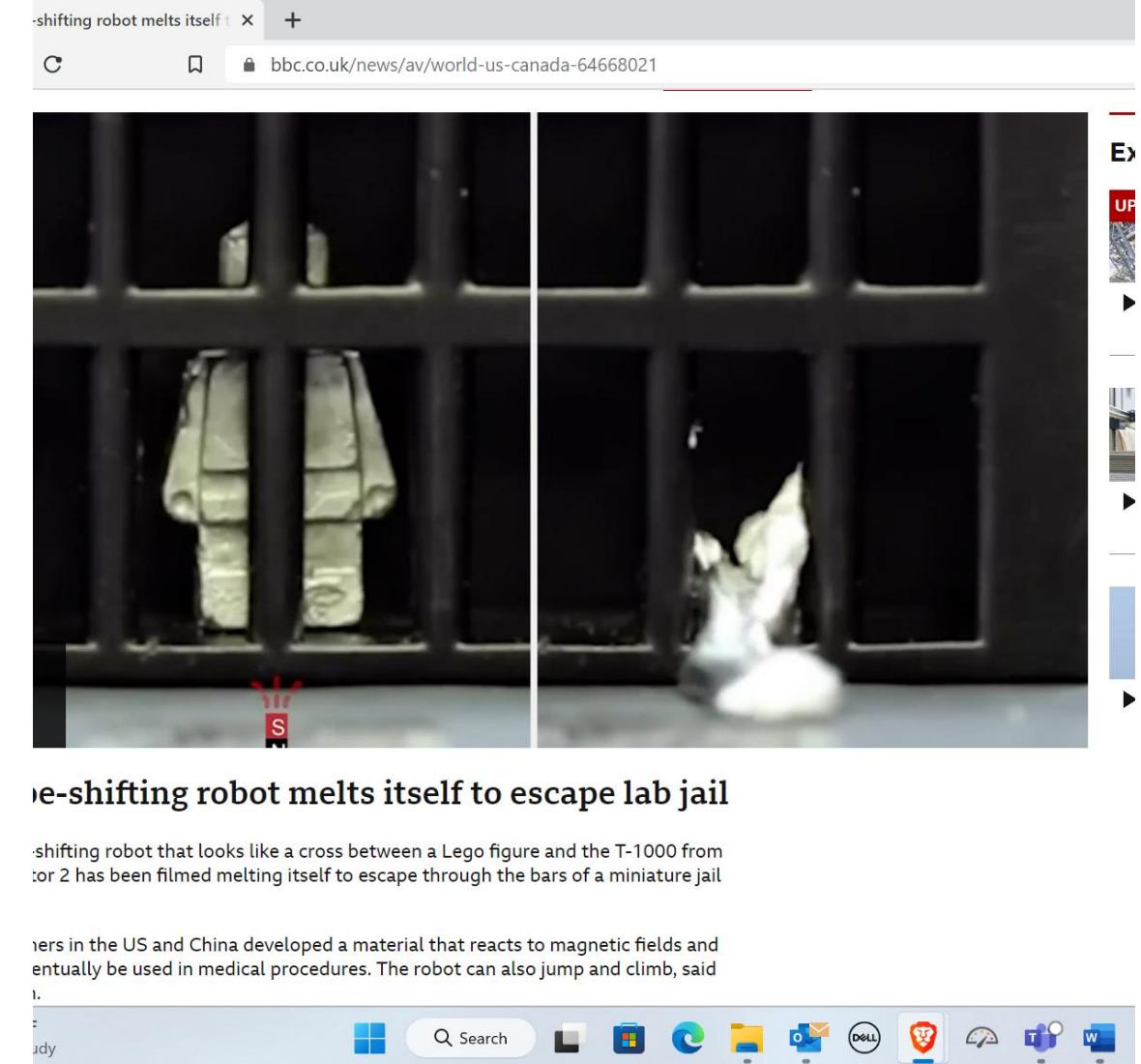
- In MS Teams chat, please share a word or two that comes to mind about what **consciousness** means in humans

Shape-shifting AI!

- BBC (2023):
 - “A shape-shifting robot that looks like a cross between a Lego figure and the T-1000 from Terminator 2 has been filmed melting itself to escape through the bars of a miniature jail cell. Researchers in the US and China developed a material that reacts to magnetic fields and could eventually be used in medical procedures. The robot can also jump and climb, said the team.”

BBC News (2023):

<https://www.bbc.co.uk/news/av/world-us-canada-64668021>



Robot escapes

- *BBC (2022):
 - “Robot vacuum cleaner escapes from Cambridge Travelodge”

*From here:

<https://www.bbc.co.uk/news/uk-england-cambridgeshire-60084347>

The screenshot shows a BBC News website page. At the top, there's a navigation bar with links for Home, News, Sport, Weather, iPlayer, and Search. Below that is a red header bar with the word 'NEWS' in white. Underneath the header, there are several news category links: Home, Cost of Living, War in Ukraine, Coronavirus, Climate, UK, World, Business, Politics, and Tech. The 'Cambridgeshire' link is underlined. The main title of the article is 'Robot vacuum cleaner escapes from Cambridge Travelodge'. Below the title is a timestamp '© 22 January 2022'. To the left of the title is a red share icon. The main content area features a photograph of a black robotic vacuum cleaner on a light-colored wooden floor. At the bottom of the page, there's a footer bar with various icons and links, including a weather forecast for 47°F Cloudy, a search bar, and links to Microsoft Office applications like Word, Excel, and PowerPoint.

And another escaping robot!

- *BBC News (2016):
 - “Russian 'runaway robot' causes traffic jam”

*From here:

<https://www.bbc.co.uk/news/blogs-news-from-elsewhere-36547139>





Are those machines conscious?

What is Consciousness?

- “If you have to ask, you’ll never know!”

Stephen Pinker in *How the Mind Works*, 1998

Learning Consciousness

- “No one learns to be conscious.”
- “When the brain starts to function, up it comes, just like steam out of a turbine”

Michael Gazzaniga *Nature's Mind* 1992 p.203

Turing on consciousness

- “I do not want to give the impression that I think there is no *mystery about consciousness* ... But I **do not think these mysteries necessarily need to be solved before we can answer the question** {Can machines think?}”. (Turing, 1950: p 447).

- *So what do we do with consciousness?*

Think of what we do with our senses....



VISION

HEARING

SMELL

TASTE

TOUCH

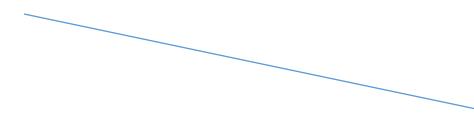


From here: <https://www.worldatlas.com/articles/what-are-the-five-senses.html>

Imagine a wicketkeeper catching a ball in a cricket match



Picture in your mind a goalkeeper stopping a ball from going into the back of a net in a football match



- But what does it mean to be *alert*, *aware*, or to *have intention*?

- If I place my hand on a hot radiator, I will become alert to the heat.
- If I don't take my hand away it will scald, if that happens my hand will hurt.
- I will become aware of the pain. I will be conscious of the pain



Conscious Experience

- To be conscious of *something* means experiencing this ‘something’.

Dictionary definitions include

To be *conscious* means to be:

- *Alert*
- *Aware*
- *Awake*
- *Have intention*

What is the nature of this conscious experience?

- “The feeling of consciousness never changes in life.... it is a feeling about things that doesn’t seem to change... it compels us to survive”

Michael Gazzaniga *Nature’s Mind*, 1992 p.204

- Can injury to the brain, or brain illness leading to someone being completely paralysed unable to move, or speak, but whose mind might be alert, *inform us about consciousness?*

Damage to Consciousness

- Case of patient of UCL Neurobiologist Professor Semir Zeck, an artist suffering brain damage after a small car accident after which he saw everything in **grey**.
- The artist **stopped eating tomatoes** because he could no longer see them as red; stopped making love to his wife because he couldn't stand the colour of her skin which to him was an unattractive **mouse grey**.

The brain injury of 19th Century man: Phineas Gage

- Summer of 1848 in Vermont, US: 25 year old railroad construction foreman Phineas Gage:
 - “It is four-thirty on this hot afternoon. **Gage has just put powder and fuse in a hole and told the man who is helping him to cover it with sand.** Someone calls from behind, and Gage looks away, over his right shoulder, for only an instant. Distracted, and before his man has poured the sand in, **Gage begins tamping the powder directly with the iron bar. In no time he strikes fire in the rock, and the charge blows upward in his face.”**

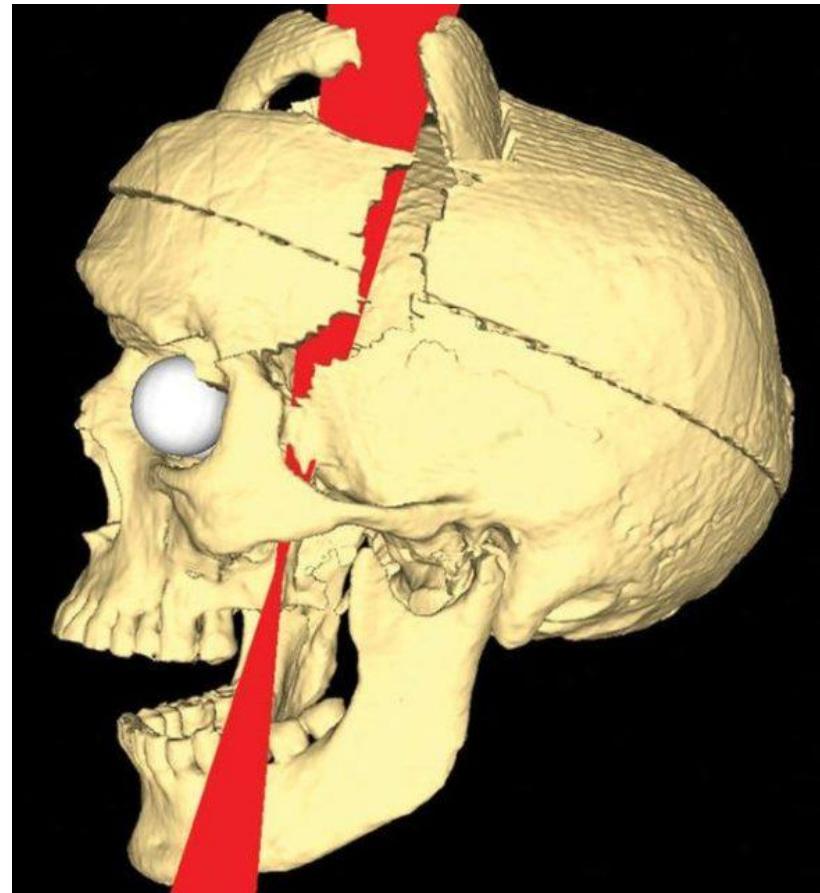
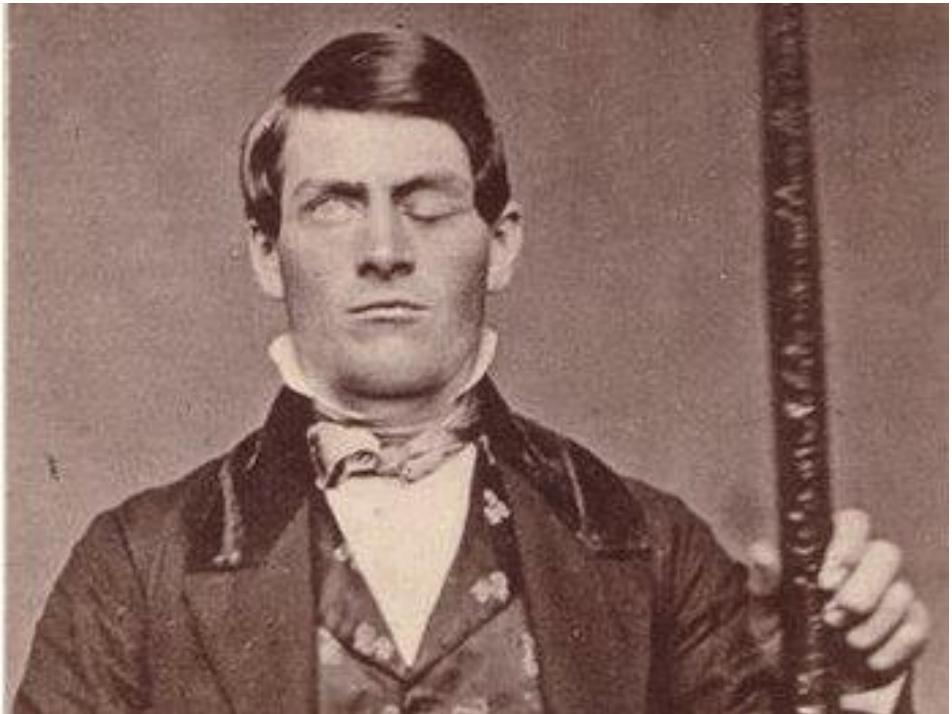
(Antonio R. Damasio *Descartes' Error*, 1996: p. 4)

How it happened

- “The iron enters Gage's left cheek, pierces the base of the skull, traverses the front of his brain, and exits at high speed through the top of the head. The rod has landed more than a hundred feet away, covered in blood and brains. Phineas Gage has been thrown to the ground. He is stunned, in the afternoon glow, silent but awake.”

(Antonio R. Damasio *Descartes' Error*, 1996: p. 4)

Phineas P. Gage



Images from Vintage News:

<https://www.thevintagenews.com/2017/02/25/the-bizarre-story-of-how-a-railroad-worker-survived-an-iron-rod-which-went-through-his-skull/>

Gage's behaviour immediately after the accident

- “Gage was not killed instantly” (Damasio, 1996: p. 5)
- “...shortly thereafter he exhibited a ‘few convulsive motions of the extremities’ and ‘spoke in a few minutes’ ”
- “[Gage’s men] ... took him in their arms and carries him to the road, and sat him into an ox cart, in which he rode, sitting erect ... to the hotel of a Mr Joseph Adams”
- “ Gage ‘got out of the cart himself, with a little assistance from his men’ ”

(Antonio R. Damasio *Descartes' Error*, 1996: p. 5)

- “..wound upon the head the pulsations of the brain being very distinct; there was also an appearance the top of the head appeared somewhat like an inverted funnel; this was owing... to the bone being fractured about the opening for a distance of about two inches in every direction.that the opening through the skull and integuments was not far from one and a half inch in diameter; the edges of this opening were everted, and the whole wound appeared as if some wedge-shaped body had passed from below upward.”

(Antonio R. Damasio *Descartes' Error*, 1996: p. 5-6)

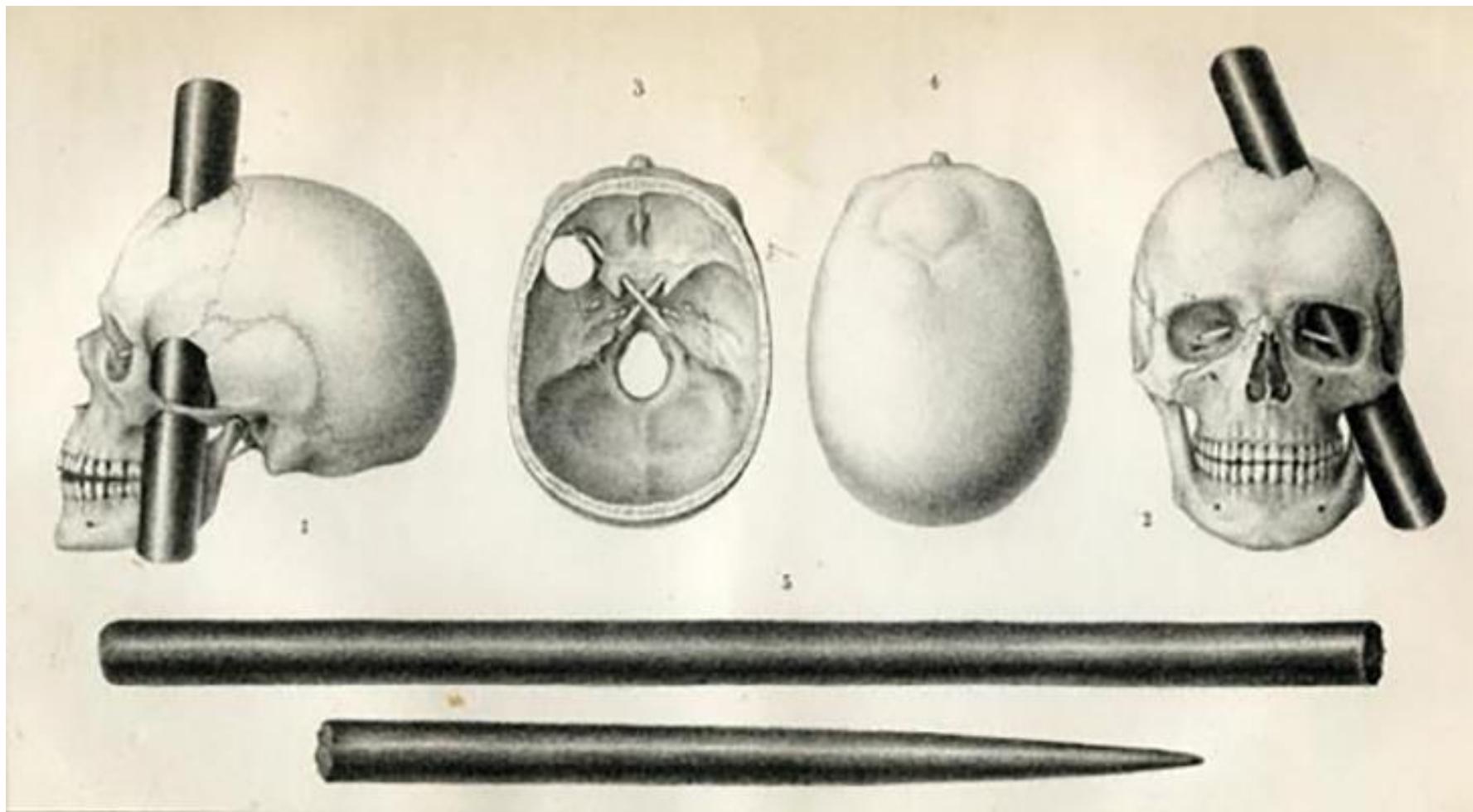


Image from:
<https://braintour.harvard.edu/archives/portfolio-items/phineas-gage>

- “[During the time Dr. Edward Williams is examining this wound] “Mr. Gage was relating the manner in which he was injured to the bystanders; he talked so rationally and was so willing to answer questions, that I directed my inquiries to him in preference to the men who were with him at the time of the accident”

(in Antonio R. Damasio *Descartes' Error*, 1996: p. 5-6)

- The horrific damage to Gage's brain, while he was consciously engaging in his work as a Railroad construction foreman, did not prevent his mind from constructing verbal explanations relating what happened to him

NB: Damasio's book is available as e-book – the 1995 print is uploaded to Moodle and available here:
https://ahandfulofleaves.files.wordpress.com/2013/07/descartes-error_antonio-damasio.pdf

Research into consciousness

Involves investigating:

- What is **mind**
- The workings of the **brain**
- **Intelligence**
- The *soul* – Turing discussed it in his 1950 essay, even mentioned extra-sensory perception (ESP) or sixth-sense.

Dualism

Cartesian Dualism best known version

I think, therefore I am

Renee Descartes' theory is a form of substance dualism

The Hard Problem in Consciousness

- How can we explain that *physical processes in the brain* give rise to subjective experience?

19th century biologist T.H. Huxley on consciousness

- Taylor (2019) :
 - T.H Huxley "... described human beings as “**conscious automata**” with **no free will**”;
 - “As he [Huxley] explained in 1874, “Volitions do not enter into the chain of causation.... The feeling that we call volition is not the cause of a voluntary act, but the symbol of that state of the brain which is the immediate cause.””

Taylor, S. (2019).
How a Flawed Experiment “Proved” That Free Will Doesn’t Exist.
Scientific American.

<https://blogs.scientificamerican.com/observations/how-a-flawed-experiment-proved-that-free-will-doesnt-exist/>

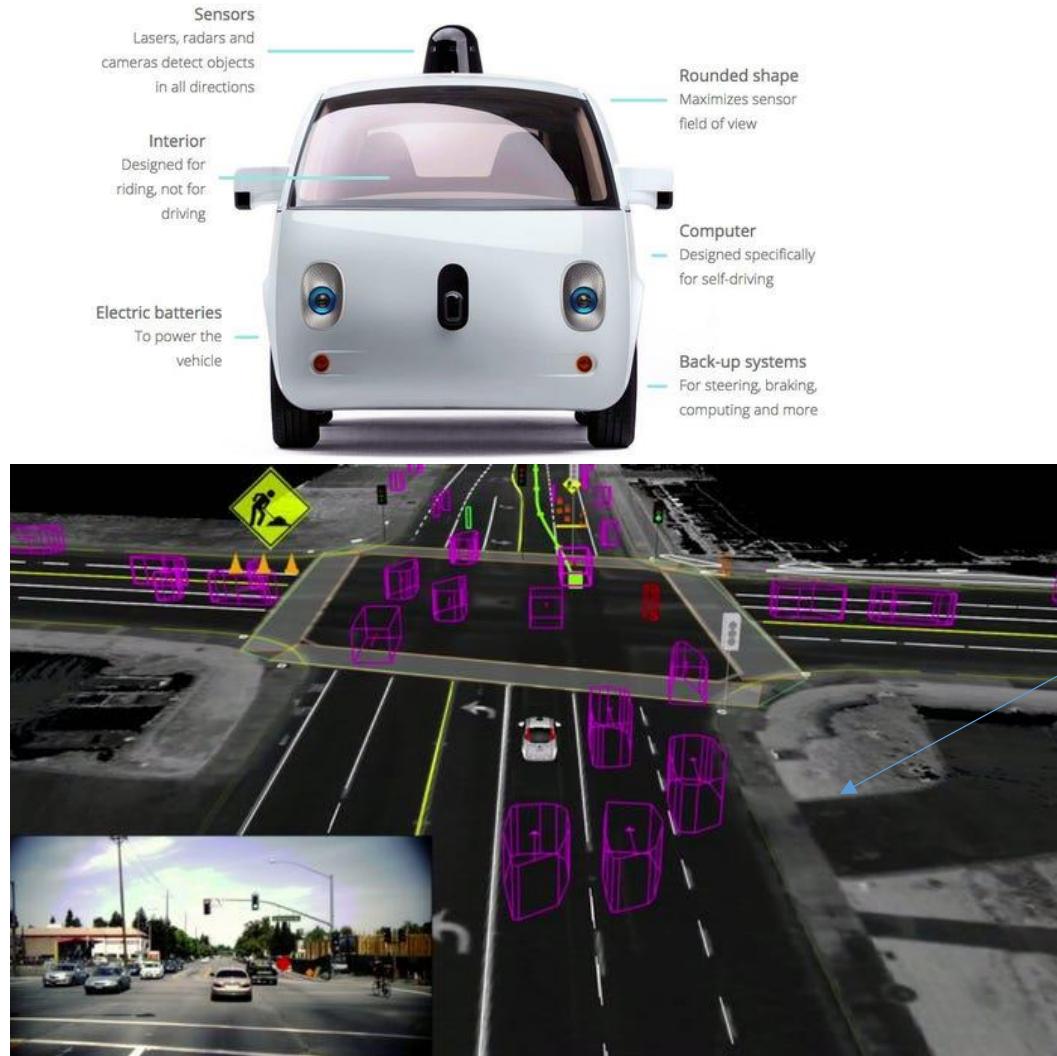
Machine Experiences— *can they feel?*



Can a machine be conscious?

- Do we mean: *can a machine experience human-type consciousness?*

Machine with vision: is it *thinking*?



Google's self-driving vehicles first establish their location by using mapping and sensor data.

The car sees other vehicles, which are represented by the purple figures, a cyclist, which is outlined in red, and some orange cones in the top left corner.

From **Business Insider**, here:

<https://www.businessinsider.com/how-googles-self-driving-cars-see-the-world-2015-10?r=US&IR=T>

Machine Olfaction (ability to smell)

- Ryman et al. (2016) experimental setup involved an olfactometer:
 - “In order to make significant advances in machine learning around chemical sensing, there are many complex features that have to be replicated. For example, the **ability to discriminate odors in a complex and changing environment.**”
 - Paper presents “the development and foundation for a new method of vapor identification to grow.” {to accurately identify a vapor.}

Ryman, S.K., Bruce, N.D.B. and Freund, M.S. 2016
Sensors and Actuators B: Chemical
Vol. 231: 666-674

<https://www.sciencedirect.com/science/article/pii/S0925400516303537>

What about consciousness in a cyborg?

- What if a machine is fused with a human – a cyborg?

Conscious perception

You perceive colour 18 milliseconds before you perceive motion

Specific damage to a part of the brain can result in loss of colour vision



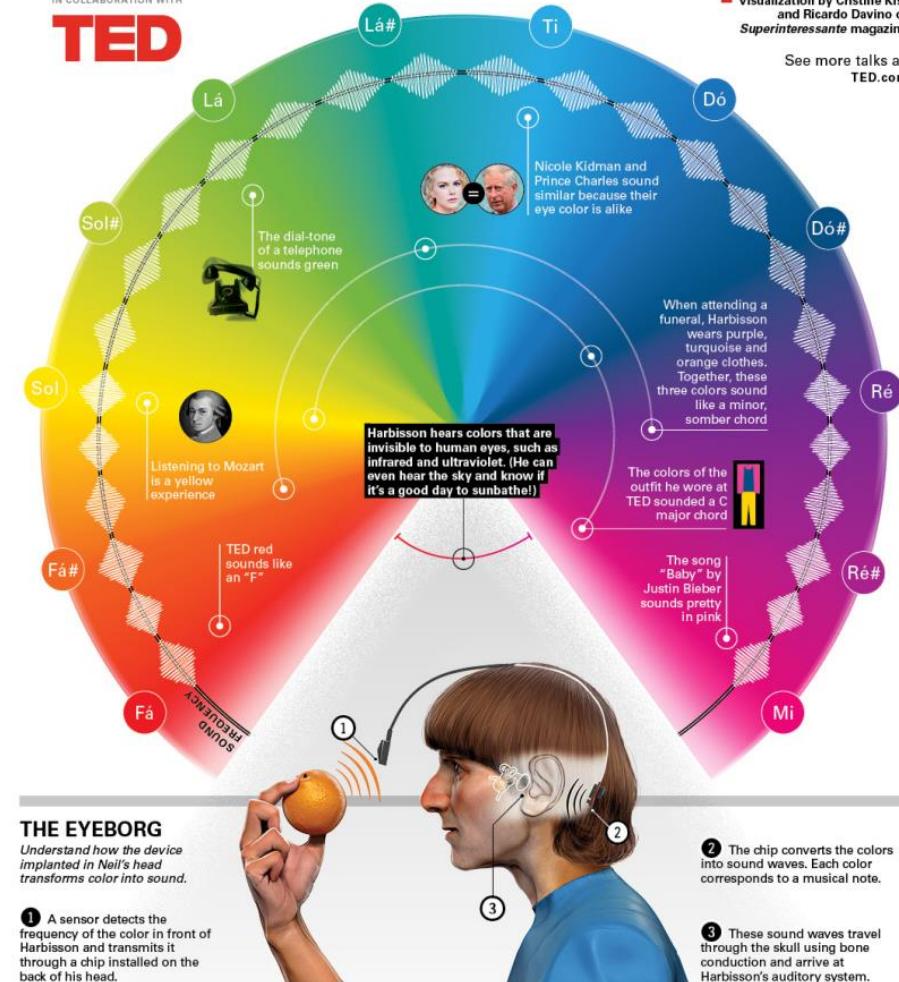
Neil Harbisson: the first legal cyborg

{IDEA} VISUALIZATION
IN COLLABORATION WITH **TED**

In his talk at TEDGlobal 2012, colorblind artist Neil Harbisson delighted the audience with his brightly colored outfit, his quirky personality, and his eyeborg — a device implanted in Harbisson's head that lets him hear a rainbow of color. Instead of seeing a world in grayscale, he can listen to the audible frequencies transmitted by the colors in faces, paintings, even the weather. Step inside the mind of Neil's symphony of color.

Visualization by Cristine Kist and Ricardo Davino of Superinteressante magazine

See more talks at: TED.com



<https://ideas.ted.com/the-sound-of-color-neil-harbissons-talk-visualized/>



What about
emotions?

What are emotions?

- Vallverdú and Casacuberta (2008:103):
 - “**Emotions and feelings are basic regulators of human activity.**
 - “... they are the basis of our interaction with the world: through **pleasure, pain, hunger or fear**”

Vallverdú, J. and Casacuberta, D. (2008)

The Panic Room: on Synthetic Emotions.

Chapter in A. Briggle et al. (Eds)

Current Issues in Computing and Philosophy. Frontiers in Artificial Intelligence and Applications series.

IOS Press

Emotions

- 17th century scientist Blaise Pascal (quoted in Vallverdú and Casacuberta, 2008: 103)
 - “The heart has its reasons, of which reason knows nothing”
 - “... abstract reason is superior to the emotions ... emotions have a special role in human existence”

Vallverdú, J. and Casacuberta, D. (2008)
The Panic Room: on Synthetic Emotions.
Chapter in A. Briggle et al. (Eds)
Current Issues in Computing and Philosophy. Frontiers in Artificial Intelligence and Applications series.
IOS Press

Human Emotions

- Vallverdú and Casacuberta (2008:104):
 - “... scientific research in neurophysiology have shown how emotions are not just any other part of human activity, but a fundamental one... ”

Vallverdú, J. and Casacuberta, D. (2008)
The Panic Room: on Synthetic Emotions.
Chapter in A. Briggle et al. (Eds)

Current Issues in Computing and Philosophy. Frontiers in Artificial Intelligence and Applications series.
IOS Press

Purpose of Human Emotions

- Vallverdú and Casacuberta (2008:104):
 - “ evolutionary approaches to consciousness ... or studies of the emotions ... affirms that **emotions gave rise to consciousness**”

Vallverdú, J. and Casacuberta, D. (2008)
The Panic Room: on Synthetic Emotions.

Chapter in A. Briggle et al. (Eds)
Current Issues in Computing and Philosophy. Frontiers in Artificial Intelligence and Applications series.

IOS Press

Human Emotions and Consciousness

- Vallverdú and Casacuberta (2008:104):
 - “... origin of consciousness, lying in the structure of the nervous system (which enables data feedback loops, the cause of the emergence of consciousness), might be emotion (rather than perception) and that experienced sensations (i.e. qualia) inherently require someone to experience them ...”

Vallverdú, J. and Casacuberta, D. (2008)
The Panic Room: on Synthetic Emotions.

Chapter in A. Briggle et al. (Eds)
Current Issues in Computing and Philosophy. Frontiers in Artificial Intelligence and Applications series.
IOS Press

Emotions and self-perception

- Vallverdú and Casacuberta (2008:104):
 - “... there is a good deal of evidence pointing to a connection between consciousness and **emotions** (the latter **being necessary for self perception and, therefore, consciousness**)”

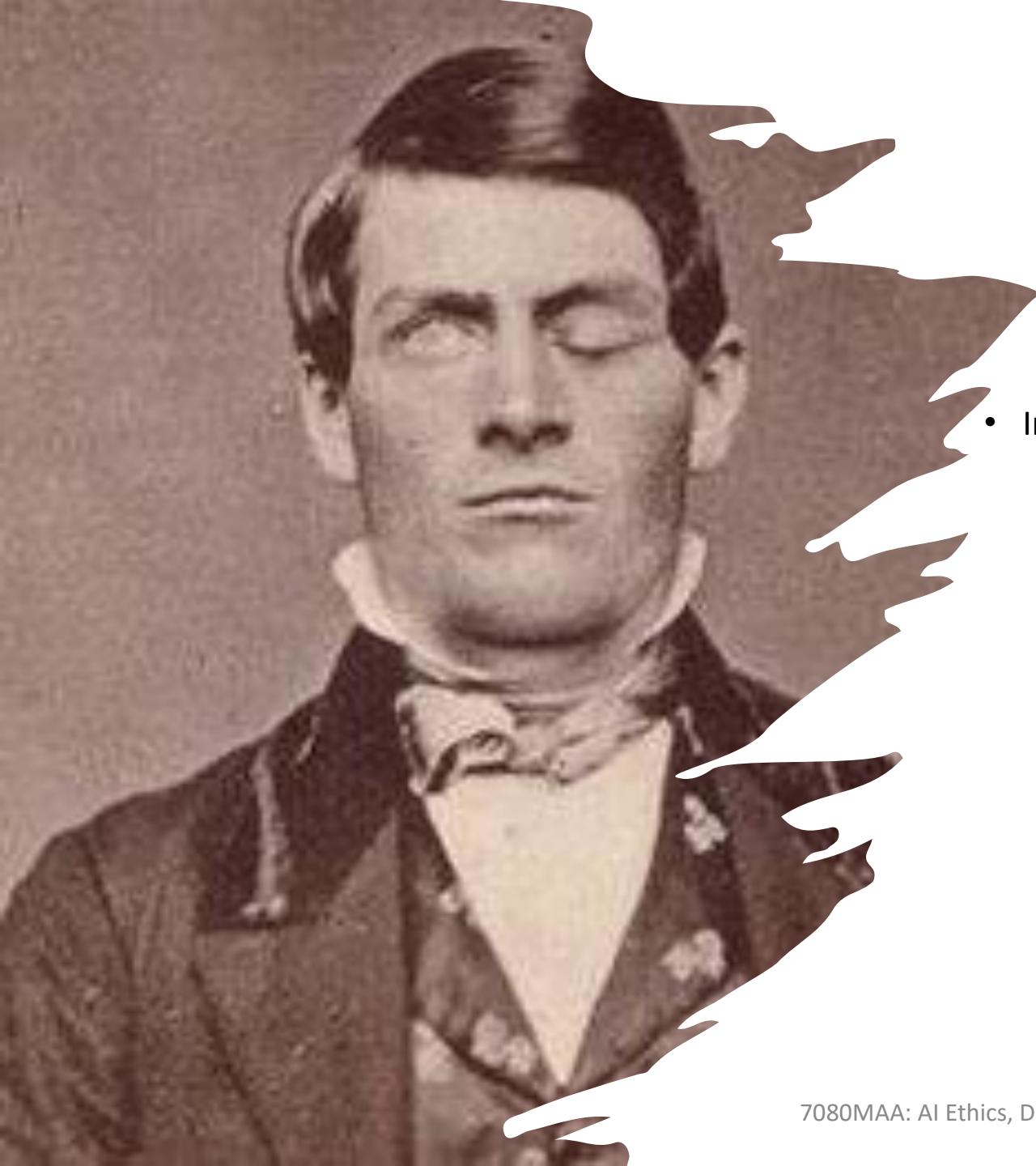
Vallverdú, J. and Casacuberta, D. (2008)

The Panic Room: on Synthetic Emotions.

Chapter in A. Briggle et al. (Eds)

Current Issues in Computing and Philosophy. Frontiers in Artificial Intelligence and Applications series.

IOS Press



Recalling Gage: an emotional change

- In Damasio (1994: 8) :
 - Gage changed, became "fitful, irreverent, indulging at times in the grossest profanity which was not previously his custom, manifesting but little deference for his fellows, impatient of restraint or advice when it conflicts with his desires, at times pertinaciously obstinate, yet capricious and vacillating, devising many plans of future operation, which are no sooner arranged than they are abandoned".

Damasio, A.R. (1994)

Descartes' Error: Emotion, Reason and the Human Brain
Avon Books: New York, USA

- Firth-Godbehere (2018):
 - “The world is being flooded with technology designed to monitor our emotions”
 - “From Alexa to self-driving cars, **emotion-detecting technologies** are becoming ubiquitous—but they rely on out-of-date science”
 - “....the technology is cutting-edge, it’s using an outdated scientific concept stating that **all humans**, **everywhere**, **experience six basic emotions**, and that **we each express those emotions in the same way.**”

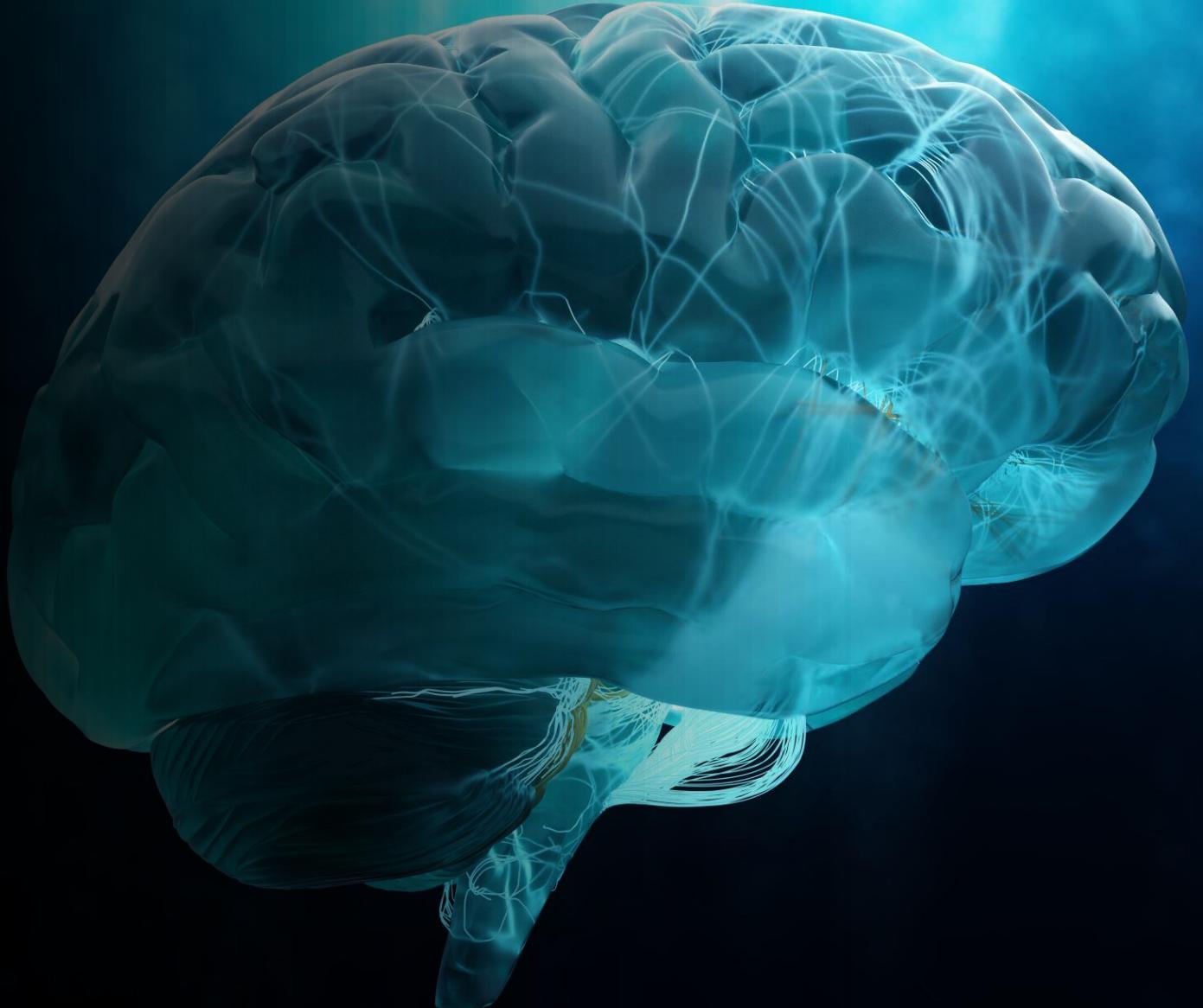
Firth-Godbehere, R. (2018).
Silicon Valley Thinks Everyone Feels the Same Six Emotions.

How we get to next magazine

<https://howwegettonext.com/silicon-valley-thinks-everyone-feels-the-same-six-emotions-38354a0ef3d7>

7080 MAA

*Ethical
Implications of
AI*



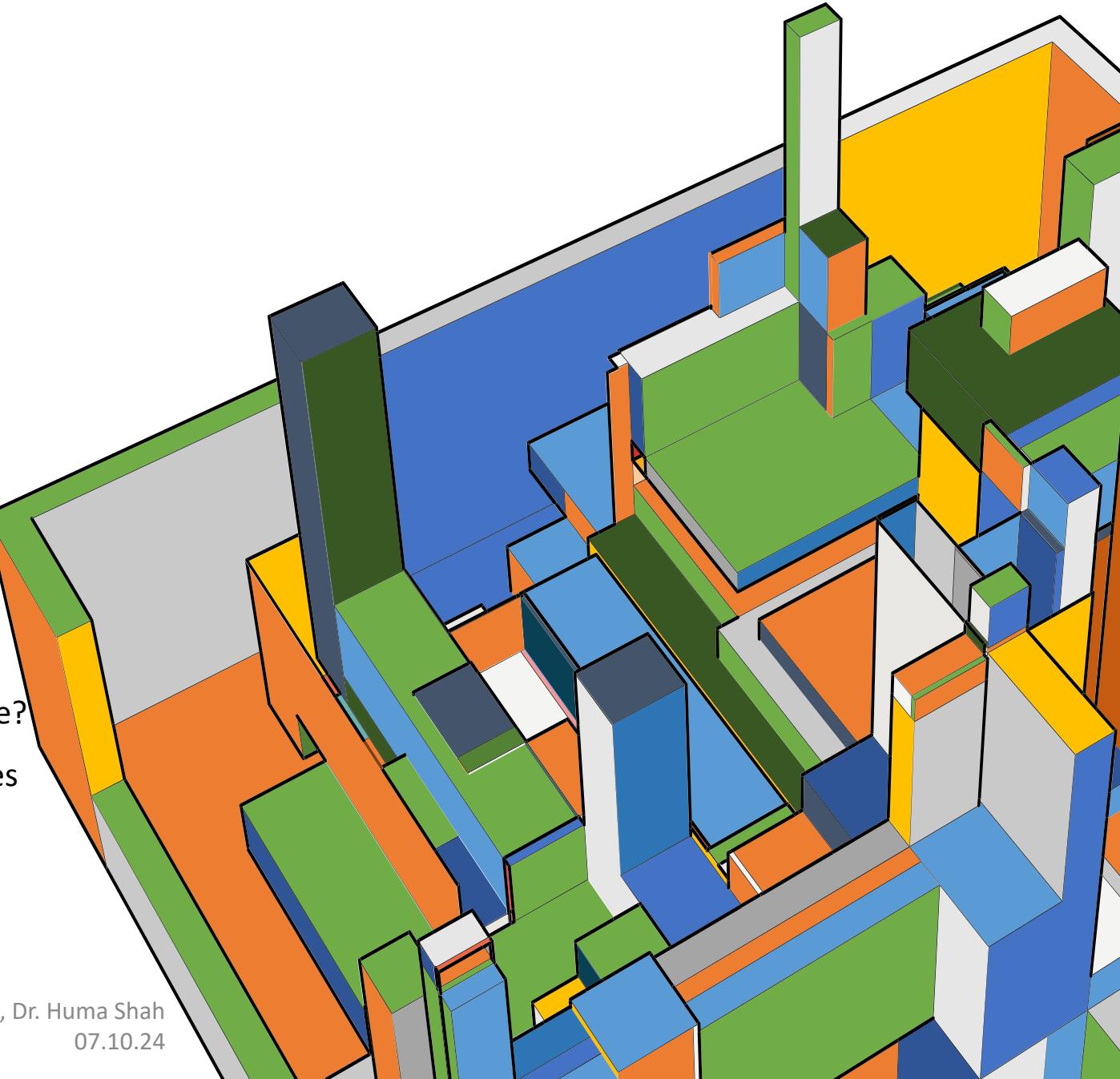
What is 'ethics' to you?

- In MS Teams chat, please share your opinion on what 'ethics' means to you...

AI ETHICS

Ethics & *Ethical Implications of AI:*

1. How can we be ‘ethical’?
2. **Which AI** are we to analyse for its ethical aspects:
 - i. A driverless car?
 - ii. A carer social robot?
 - iii. A space explorer machine?
 - iv. A disaster rescue drone, such as in earthquake?
 - v. A large language model (LLM) that summarises in response to a ‘prompt’ (e.g. GPT4; Gemini)



Ethics: What are the problems in AI?

1. PRIVACY

Privacy is not an abstract social construct; it is a human right (UNDHR 1948) - but how can privacy be protected in AI tech?

2. RESPONSIBILITY

- Who is responsible if an *AI* goes wrong?
- What does '*go wrong*' look like?

3. PURPOSE

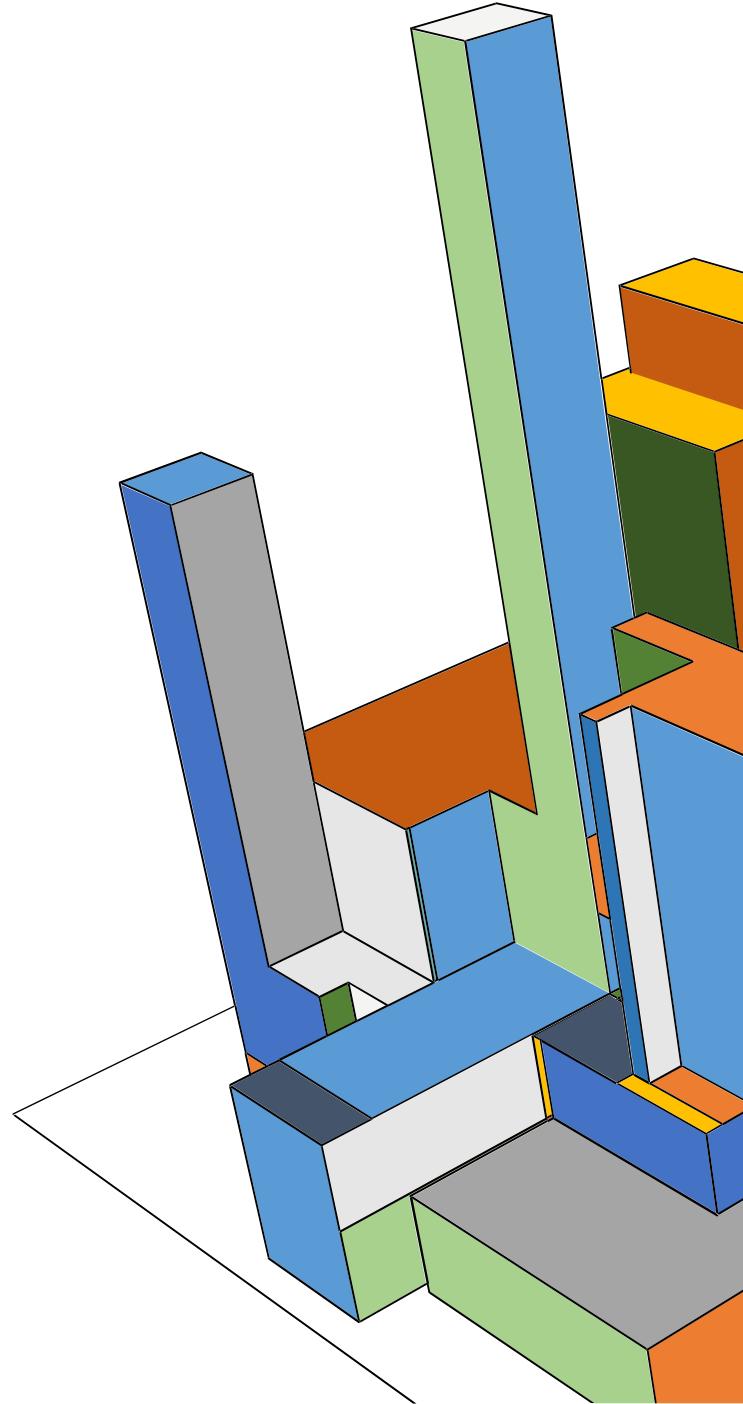
What is the purpose/ benefit of any new AI tech?

4. SHOULD AN AI BE HUMAN-LIKE?

Should we engineer an AI to have **consciousness** and **emotions**? If so, in which situations, and is this even possible?

5. STATE OF THE WORLD

In a future with next generation *AI* (*beyond generative models*) what happens to humanity?



Other ethical considerations

- Context of AI/autonomous algorithm deployment:
 - Social media platform for young teens (e.g. Instagram)
 - Workplace (e.g. Co-pilot in Microsoft suite)
- Veracity of information output from an AI:
 - Meaningful
 - Relevant
 - Truthful
- Place (places) of intended deployment (geographical location)

Ethics in its simplest form

- Do good
- Obey rules
- Follow standards, e.g.:
 - Building a bridge requires following standards such as what materials will be used for the bridge so that it does not collapse from the first pedestrian or vehicle crossing it: so safeguarding users
 - The same fundamental principle applies to other areas: **being ethical means safeguarding, considering others, following agreed standards**

Bad driving, but resistance against the driverless car!

- Bellen (2023):
 - “Autonomous vehicle companies Cruise and Waymo have separately pushed a narrative this week that **humans are bad drivers** and that their **technologies are crucial to making roads safer.**” in Yahoo news
<https://uk.movies.yahoo.com/cruise-waymo-humans-bad-drivers-100029274.html>
- Clayton (2023):
 - “**San Francisco’s Safe Street Rebel group are disabling autonomous taxis by placing a cone on the hood of the car**” in the Guardian:
<https://www.theguardian.com/technology/2023/jul/07/san-francisco-autonomous-cars-protest-cone?ref=biztoc.com>

AI in the Newsroom

- Eisikovits (2024):
 - “The journalism industry **has been under immense economic pressure** over the past two decades, so it makes sense that **journalists have started experimenting with generative AI to boost their productivity.**”
 - “An Associated Press survey published in April 2024 asked **journalists** about the use of generative artificial intelligence in their work. **Nearly 70% of those who responded** said they **had used these tools to generate text, whether it was composing article drafts, crafting headlines or writing social media posts.**”
 - “... if journalists are not careful about its deployment, the use of generative AI could undermine the integrity of their work.”

From here:

<https://theconversation.com/newsrooms-are-experimenting-with-generative-ai-warts-and-all-228565>

Misinformation

- Google's own researchers: **AI is a top vector of disinfo**
<https://t.co/qb8cOnMu8I>
(https://x.com/jason_koebler/status/1795446984586698840?t=7svwdlfsE5oWaCYIE7ZKgw&s=03)
- “Misinformation and lies are everywhere online. How do you tell what's legit and what's fake?” (WhatIsMyIPAddress.com -@wimia posted on X at 5:25 pm on Tue, May 28, 2024)

Links:

<https://t.co/wl36BQq4sW>

<https://t.co/XhWxAelZ90>

<https://x.com/wimia/status/1795491505244688449?t=F3G9faHZ9L7o0ZTmkoXjxw&s=03>

AI >SKEWED LAW?

- Magesh, et al., 2024:
 - “In the legal profession, the recent integration of large language models (LLMs) into research and writing tools presents both unprecedented opportunities and significant challenges ... These systems promise to perform complex legal tasks, but their adoption remains hindered by a critical flaw: their tendency to generate incorrect or misleading information, a phenomenon generally known as “hallucination”

Paper accessible from here:

[\[2405.20362\] Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools \(arxiv.org\)](https://arxiv.org/abs/2405.20362)

Fake Authors

- Futurism (@futurism) posted at 4:01 pm on Wed, May 08, 2024:

“Last fall, we reported that Sports Illustrated published product reviews by fake authors with AI-generated profiles. Now, we've uncovered how the company behind that content has generated articles for dozens of news sites — from USA Today to the LA Times.”

<https://t.co/v2UAO7url2>

(https://x.com/futurism/status/1788222758658203950?t=R_1CQo-2DV3Ty8OuL0PLzg&s=03)

AI & Copyright



© Can Stock Photo - cap11112399

Clip art from the web



- Milmo (2024):
 - “Google has said that Britain risks being left behind in the global artificial intelligence race unless the government moves quickly to build more datacentres and let tech companies use copyrighted work in their AI models.”
 - Google “calls for changes to UK copyright laws after the abandonment this year of attempts to draft a new code for using copyrighted material to train AI models.”
 - “Data from copyright-protected material such as news articles and academic papers is seen as vital for models that underpin tools like chatbots, which are “trained” on billions of words that allow them to understand text-based prompts and predict the right response to them. The same concerns apply to models that make music or images.”

Milmo, D. (2024). Google says UK risks being ‘left behind’ in AI race without more data centres.

The Guardian – News, AI: 19.09.24.

<https://www.theguardian.com/technology/2024/sep/19/google-says-uk-risks-being-left-behind-in-ai-race-without-more-data-centres>

Democracy and deep fakes

- Paul Nemitz (@PaulNemitz) posted at 7:06 pm on Sun, May 26, 2024:
 - “The #politicalclone is what #AI contributes to the #Democracy in #India. Voters do not realize they talk to clone. They Are Being Bombarded With Millions of #Deepfakes. Political candidates approve. ”
<https://t.co/sZptdcNboH>
(<https://x.com/PaulNemitz/status/1794792275370795231?s=03>)

LLM cost/benefit

- Grady Booch (@Grady_Booch) posted at 3:54 pm on Mon, May 27, 2024:
“Large language models are one of the worst genre of software products I've seen in my career.”
“They are inherently and irredeemably unreliable narrators, offering tantalizingly coherent output that is dangerous in its ability to deceive, threatening in its casual toxic bias....”
(https://x.com/Grady_Booch/status/1795106421153096022?s=03)
- Carissa Véliz (@CarissaVeliz) posted at 9:24 am on Tue, May 28, 2024:
“Research shows that employees who are steered by digital nudges may lose some ethical competency. We should be mindful of this risk when using AI assistants. #AIEthics”
<https://t.co/WtYN1U67X3>
(<https://x.com/CarissaVeliz/status/1795370424437539290?t=DOZFPTiT6bTlhXQhUezlg&s=03>)

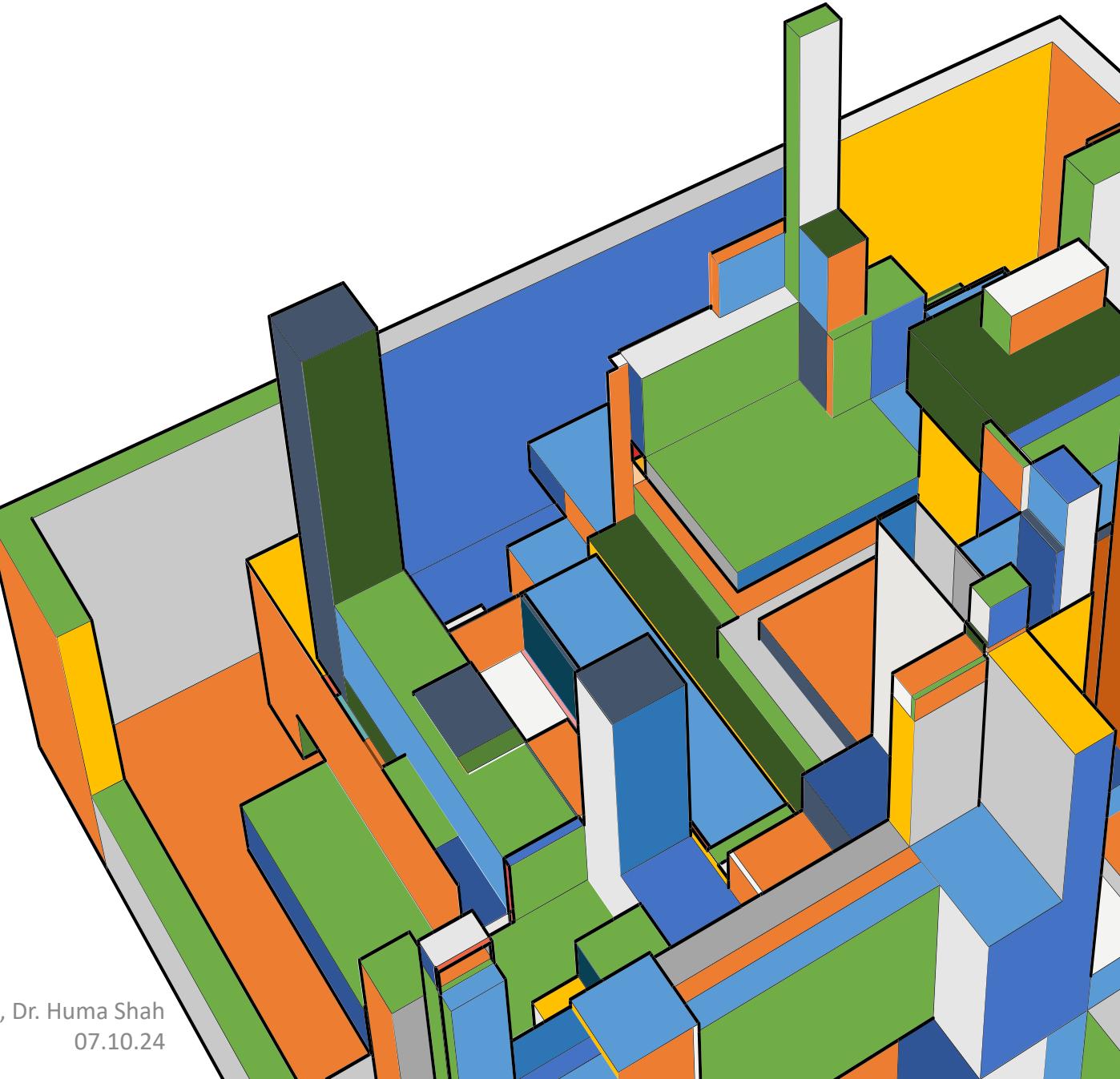
AI Influencers

- Former UK Prime Minister [Tony Blair](#) event on AI in July- he claimed that:
 - “more than 40% of tasks performed by **public sector workers** could be partially automated”
- Politico, TechRadar and Forbes identified a problem with this research:
 - “Blair’s research is based on ChatGPT prediction, because ‘interviewing experts would be too hard’” !!!!

The full article is accessible from this link:

[AI Finds That AI Is Great In New Garbage Research From Tony Blair Institute \(404media.co\)](#)

Impact on society



Extracting our personal data

From Andreas Weigend in '*Data for the People*' book (2017, pages 2-3):

- “Every day **more than a billion people create and share social data [online]....”**
- “**Some of these data are shared knowingly and willingly...”**
- “**..part and parcel of the convenience of using the Internet and mobile devices.”**

Sharing personal data

- Book in Lanchester Library:

A. Weigend. 2017. *Data for the People: How to Make Our Post-Privacy Economy Work for You.* Basic Books: New York, USA (pages 2-3):

Extracting our personal data contd

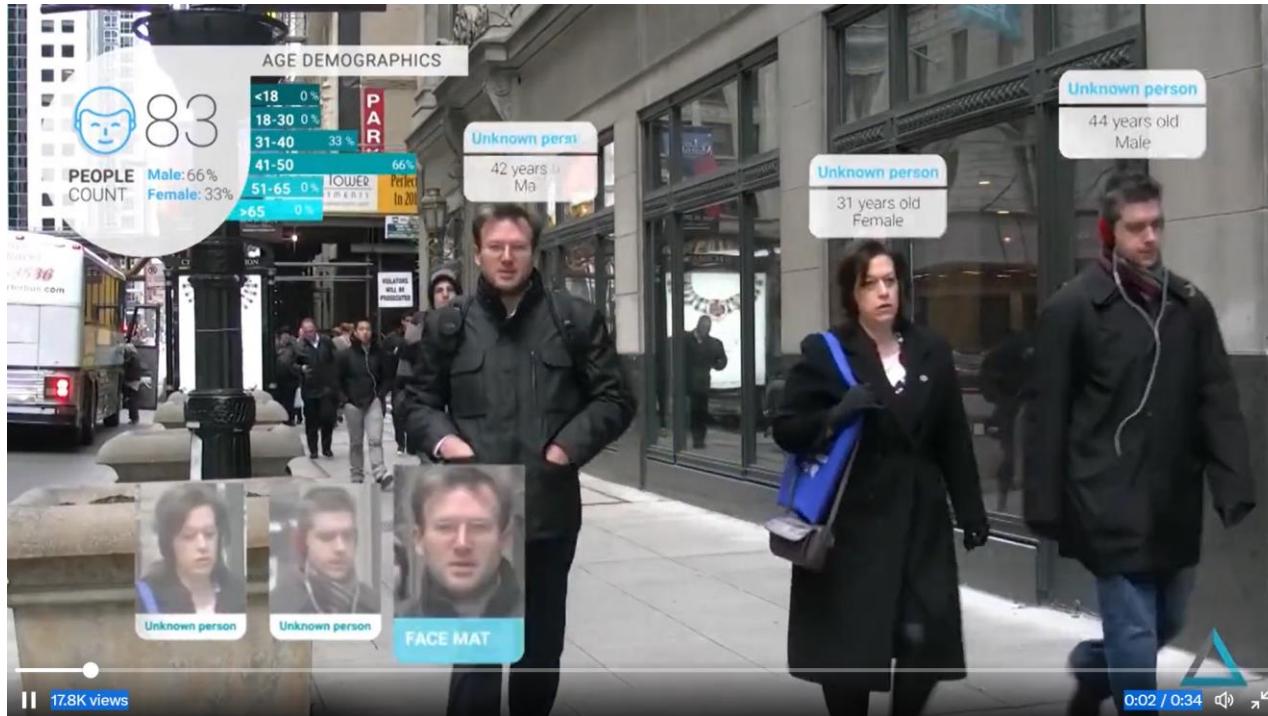
- *Nigel Warburton (n.d.)
 - “From the moment you check your phone in the morning you are losing your privacy”.
 - “Corporations around the world are recording what time you woke up, where you slept, and with whom”.
 - “Without your permission , or even your knowledge, tech companies are harvesting your data – your location, your likes, your habits, your fears, your diseases, your politics – and sharing it amongst themselves”.
 - “They are not just selling your data. They are selling the power to influence you, and everyone you know”.

*From inside cover of book

Véliz, C. (2020). *Privacy is Power. Why and How You Should Take Back Control of Your Data*. Bantam Press, London, UK

Privacy in the ‘smart city’

- “... a snap shot on how "Smart Cities" and "15 Minute Cities" surveillance will work. **You will have zero privacy...it's all about the data.”***



*From Twitter feed of ***The Investigative Examiners***:

<https://twitter.com/TruthorConseq12/status/1622775916001894400>

Protecting Paris: Police snipers and AI cameras watched over 2024 Olympic Games

- USA Today (2024):
 - “AI: tracking ‘unusual patterns’
 - “One area of Olympics-related security that's been getting attention is France's use of AI to help ensure the safety of the estimated 15 million people expected in Paris for the [Olympic] Games.”
 - “AI's impact is being felt across different sectors and industries. Its use at the Olympics is perhaps the most high-profile example yet of how it's being deployed to detect security threats. In the U.S., various laws at the state and local level have sought to regulate AI's facial-recognition capabilities.”

From here:

[Police, soldiers and AI cameras will watch over the Paris Olympics \(indystar.com\)](https://www.indystar.com/article/2024/07/01/police-soldiers-and-ai-cameras-will-watch-over-the-paris-olympics)

Privacy concerns

- **Privacy:**
 - **data protection concerns** arising from artificial intelligence ‘approaches’
(machine learning applied to big data sets)

Privacy as a fundamental right

- “**Privacy is a fundamental right:**
 - Essential to autonomy
 - Protection of human dignity
 - Serving as the foundation upon which many other human rights are built.”

<https://privacyinternational.org/explainer/56/what-privacy>

Responsibility

Responsibility – *for what?*

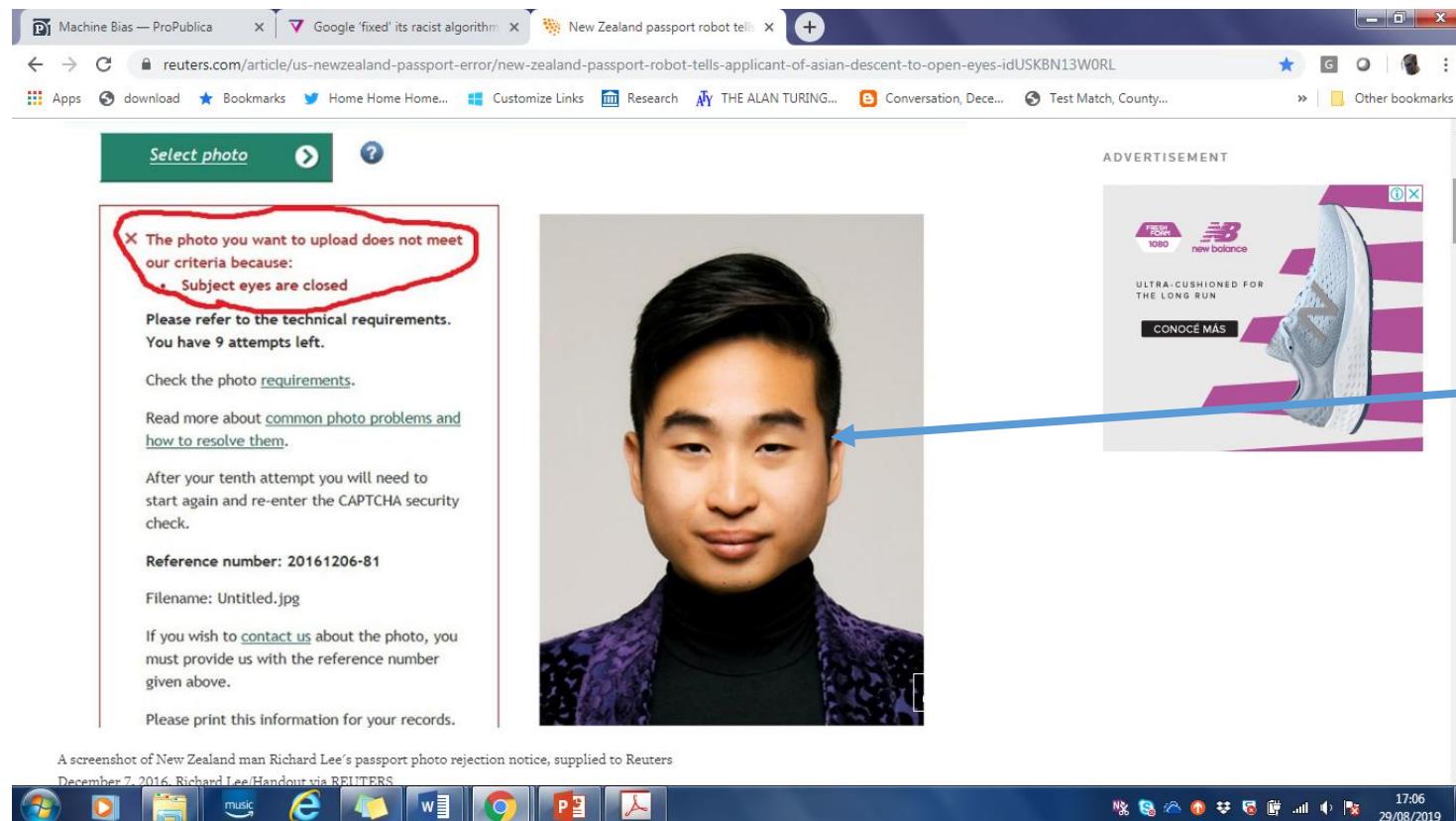
WEF's top nine identified ethical issues:

1. **Unemployment** through increasing automation
2. **Inequality** – think of billionaires' space ambitions while on the planet some starve
3. **Humanity** – what happens to us as we interact with more machines?
4. **Artificial stupidity** – will AI mistakes be more catastrophic?
5. **Racist robots** – AI bias is a problem
6. **Security** – keeping AI clear from actors who would want to use it to harm society
7. **Unintended consequences** – the unknown
8. **Singularity** – what if AI is more intelligent? Well aren't those occupants of unidentified aerial phenomena more intelligent than humans, since they have managed to navigate space? See the [US Defence Department](#)'s 2020 statement on authenticated historical navy footage from 2004 and 2015.
9. **Robot rights**: what rights should be accorded to robots?

What's missing in WEF'S LIST?

- Should Facebook's (now *Meta*) social media ***profit-first*** philosophy be included in a list of ethical issues, since the use of **algorithms** drives its engagement quest?

New Zealand's passport renewal system



“New Zealand passport robot tells applicant of Asian descent to open eyes” !!

<https://www.reuters.com/article/us-newzealand-passport-error/new-zealand-passport-robot-tells-applicant-of-asian-descent-to-open-eyes-idUSKBN13W0RL>

Legal issues in Live Facial Recognition

- Live facial recognition technology in King's Cross, London - UK's Information Commissioner's Office:
 - “**Scanning people’s faces as they lawfully go about their daily lives, in order to identify them, is a potential threat to privacy that should concern us all.** That is especially the case if it is done without people’s knowledge or understanding.
 - “**... deeply concerned about the growing use of facial recognition technology in public spaces, not only by law enforcement agencies but also increasingly by the private sector.”**

From here:

<https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/08/statement-live-facial-recognition-technology-in-kings-cross/>

Live Facial Recognition

- “Over the past few years six [UK] forces, including the Metropolitan police, have trialled facial recognition, with spectacularly poor results.”
- “After surveilling hundreds of thousands of people, the Met has made a mere handful of arrests using this technology in the past four years.”
- “Indeed, 81% of the “suspects” caught by the cameras were simply innocent bystanders not on the police’s watchlists. Two of the trials even had an error rate of 100%.”

David Davis is Conservative MP for Haltemprice and Howden:

<https://www.theguardian.com/commentisfree/2019/sep/20/facial-recognition-technology-privacy>

Paper on ethical issues in facial recognition

- Buolamwini and Gebru (2018: p.1):
 - “... while face recognition software by itself should not be trained to determine the fate of an individual in the criminal justice system, it is very likely that such software is used to identify suspects. Thus, an error in the output of a face recognition algorithm used as input for other tasks can have serious consequences. For example, someone could be wrongfully accused of a crime based on erroneous but confident misidentification of the perpetrator from security video footage analysis”

Buolamwini, J. and Gebru, T. (2018).

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.

Proceedings of Machine Learning Research, Conference on Fairness, Accountability, and Transparency 81:1–15.

Accessed from here: <http://proceedings.mlr.press/v81/buolamwini18a.html>

Twitter @CatHallam1: April 6, 2019

- Permission granted October 21, 2019 by Cat Hallam to showcase her experience from April 2019:
 - “....thank you for asking in advance. Certainly, happy to see the information used in an educational context. Hope it helps the students & contextualises the subject matter.”

Cat Hallam: “Today, I simply wanted to renew my passport online.”

3. Photo quality and format

Your photo must:

- be in colour, with no effects or filters
- not be blurred or have ‘red eye’
- be unedited – you can’t ‘correct’ your passport photo

Our automated check suggests

- it looks like your eyes are closed
- we can't find the outline of your head

You can use the photo

- if you have a medical or religious reason why it doesn't meet the rules
- if you think it meets all the rules

If your photo doesn't meet the rules you may have problems with your application or when you travel.

BETA This is a new service – your [feedback](#) will help us to improve it.

[◀ Back](#)

Check your photo

1. Background and lighting

Your photo must have:

- a plain light-coloured background – without texture or pattern
- balanced light – no shadows on your face or behind you
- no objects behind you

Your photo



2. Your appearance

Make sure:

- the photo is a good likeness taken in the last month
- your whole face is visible with your eyes open
- you have a plain expression – no smile and mouth closed
- there are no reflections or glare (if you have to wear glasses)
- you're not wearing headwear (unless for religious or medical reasons)

3. Photo quality and format

Your photo must:

- be in colour, with no effects or filters
- not be blurred or have ‘red eye’
- be unedited – you can’t ‘correct’ your passport photo

Our automated check suggests

- we can't find the outline of your head

You can use the photo

- if you have a medical or religious reason why it doesn't meet the rules
- if you think it meets all the rules

If your photo doesn't meet the rules you may have problems with your application or when you travel.

Do you want to submit this photo?

- Yes, I want to submit it
- No, I want to get another photo

[Continue](#)

<https://twitter.com/CatHallam1/status/1114590857397788673>

BETA This is a new service – your [feedback](#) will help us to improve it.[Back](#)

Check your photo

1. Background and lighting

Your photo must have:

- a plain light-coloured background – without texture or pattern
- balanced light – no shadows on your face or behind you
- no objects behind you

Your photo



2. Your appearance

Make sure:

- the photo is a good likeness taken in the last month
- your whole face is visible with your eyes open
- you have a plain expression – no smile and mouth closed
- there are no reflections or glare (if you have to wear glasses)
- you're not wearing headwear (unless for religious or medical reasons)

3. Photo quality and format

Your photo must:

- be in colour, with no effects or filters
- not be blurred or have 'red eye'
- be unedited – you can't 'correct' your passport photo

Our automated check suggests

- it looks like your eyes are closed
- we can't find the outline of your head

You can use the photo

- if you have a medical or religious reason why it doesn't meet the rules
- if you think it meets all the rules

If your photo doesn't meet the rules you may have problems with your application or when you travel.

Do you want to submit this photo?

- Yes, I want to submit it
 No, I want to get another photo

[Continue](#)

You need another photo

This photo has failed our automated checks for the following reasons:

It looks like your eyes are closed

You must keep your head straight and in the centre of the photo.

You can still submit your photo if you believe it should be accepted (eg there's a medical reason why you can't open your eyes).

Your photo quality is too low

You need to use a photo that's between 50kb and 10mb.

It looks like your mouth is open

Sometimes this is caused by facial hair or shadows near your mouth. Please try different lighting.

You can still submit your photo if you believe it should be accepted (eg there's a medical reason why your mouth appears to be open).

<https://twitter.com/CatHallam1/status/1114590857397788673>

Twitter @CatHallam1: April 6, 2019

- “After numerous attempts and changing my clothes several times, this example illustrates why I regularly present on Artificial Intelligence/Machine Learning bias, equality, diversity and inclusion. #passport ”

"Black man is stunned after passport photo checker mistakes his lips for an open mouth as Home Office facial recognition system is accused of bias" (2019)



Automated check

- It looks like your mouth is open

Do you want to submit this photo?

Yes, I want to submit it

Explain why you'd like to use this photo

My mouth is closed, I just have big lips

No, I want to get another photo

<https://www.standard.co.uk/news/uk/man-stunned-as-passport-photo-check-sees-lips-as-open-mouth-a4241456.html>

UK Home Office use of Facial Recognition

- “The UK government went ahead with a face-detection system for its passport photo checking service, despite knowing the technology failed to work well for people in some ethnic minorities.”
- “... was aware of problems with its website’s passport photo checking service, but decided to use it regardless”
- “User research was carried out with a wide range of ethnic groups and did identify that people with very light or very dark skin found it difficult to provide an acceptable passport photograph”

New Scientist: October 9, 2019:

<https://www.newscientist.com/article/2219284-uk-launched-passport-photo-checker-it-knew-would-fail-with-dark-skin/#ixzz62XKSWWcF>

Ethics of Facial recognition

- Big Brother Watch (2024):
 - “There is no law for facial recognition in the UK”
 - “The technology works by creating a 'faceprint' of everyone who passes in front of camera — processing biometric data as sensitive as a fingerprint, often without our knowledge or consent.”
 - “The boom in facial recognition technology in the UK, operating in a largely lawless space, is a mortal threat to privacy as we know it.”

From here:

<https://bigbrotherwatch.org.uk/campaigns/stop-facial-recognition/>

CONTD.

- Big Brother Watch (2024):
 - “As well as police forces, private companies are also rolling out facial recognition technology to scan and identify members of the public, largely without their consent or knowledge”
 - “Some facial recognition companies even check faces against internet data in real-time, trawling the entire internet to identify any photos of you, posted anywhere.”

From here:

<https://bigbrotherwatch.org.uk/campaigns/stop-facial-recognition/>

contd

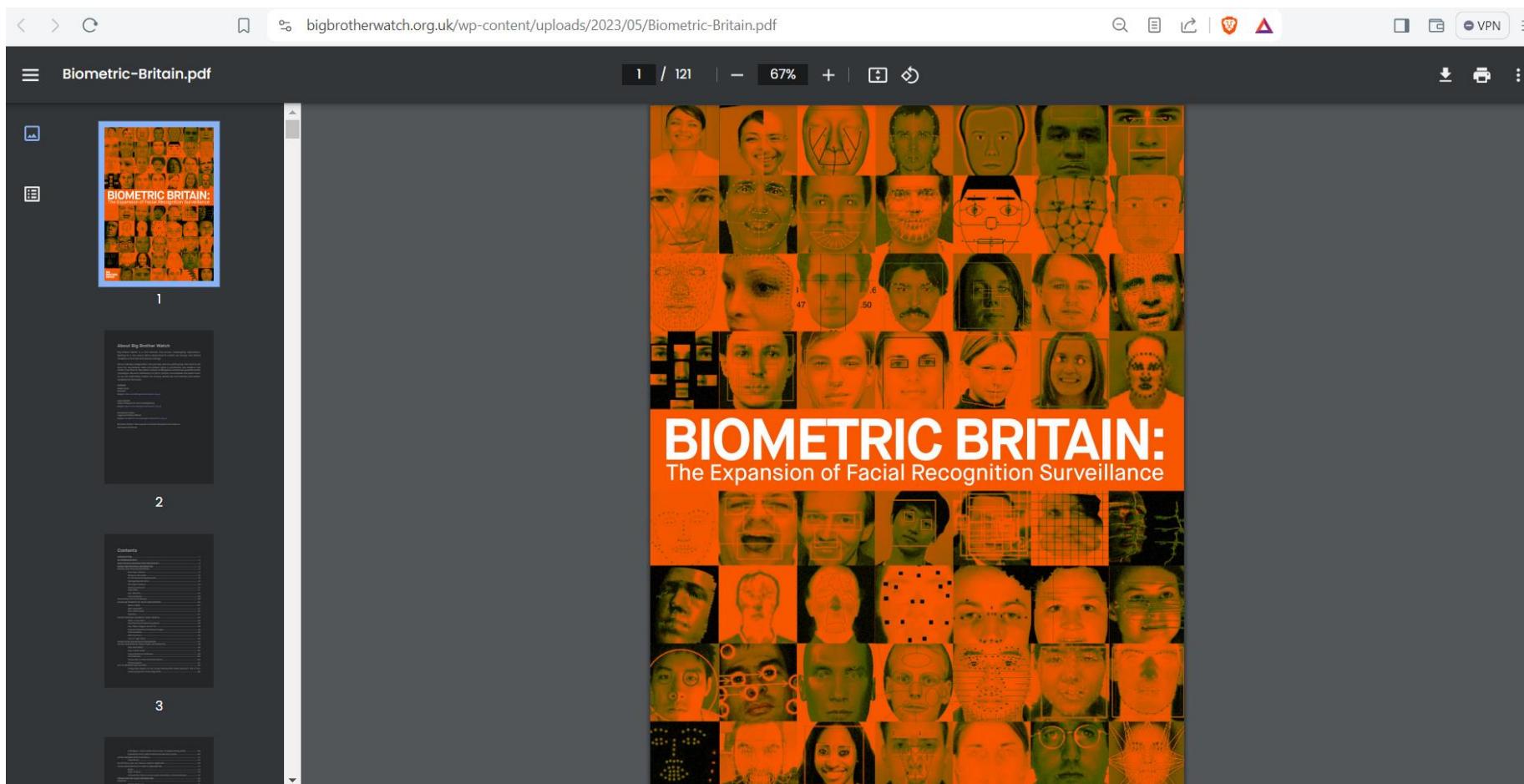
- Big Brother Watch (2024)
 - “Research shows that facial recognition discriminates against **women and people of colour.**”
 - “People up and down the country are being falsely flagged as criminals by facial recognition cameras.”
 - “Community worker, Shaun Thompson, was pulled up and held for thirty minutes by Met Police after he was misidentified as a criminal by facial recognition on his way home near London Bridge.” Legal action launched against police and the shops using facial recognition.



From here:

<https://bigbrotherwatch.org.uk/campaigns/stop-facial-recognition/>

Big Brother Watch Report



<https://bigbrotherwatch.org.uk/wp-content/uploads/2023/05/Biometric-Britain.pdf>

Facebook and Ray-Ban

- Question:

Would you wear Ray-Ban ‘smart’ glasses with Facebook camera? (See *Hern, 2021)

Image from NBC news:

<https://www.nbcnews.com/shopping/tech-gadgets/ray-ban-stories-facebook-smart-glasses-n1278954>



*Hern, A. (2021). *Facebook announces launch of Ray-Ban Stories smart glasses*.
Guardian, 9 September, 2021.

Another question.....

- Do you feel it appropriate that **Facebook** was attempting to attract pre-teens (children aged 10-12 years) to its **Instagram** platform? (See *Wells and Horwitz, 2021).

*Wells, G. and Horwitz, J. (2021).

Facebook's Effort to Attract Preteens Goes Beyond Instagram Kids, Documents Show.

The Wall Street Journal, 28 September 2021,

https://www.wsj.com/articles/facebook-instagram-kids-tweens-attract-11632849667?mod=hp_lead_pos1

Generation Z and Facebook

- “Bottom-line, Facebook appears to have a massive Gen Z issue and making Instagram approachable is probably the only way to survive it”. (See Wells and Horwitz, 2021).

Wells, G. and Horwitz, J. (2021).

Facebook’s Effort to Attract Preteens Goes Beyond Instagram Kids, Documents Show.

The Wall Street Journal, 28 September 2021,

https://www.wsj.com/articles/facebook-instagram-kids-tweens-attract-11632849667?mod=hp_lead_pos1

What underpins these systems: algorithms

- The term ‘algorithm’ can be understood to be a step-by-step process to get something done
- The mathematics underpinning computer algorithms is human created, or is it the *language of nature*?
- Algorithms are powerful
- Algorithms decide:
 - What interest rate you will pay on a bank loan, mortgage, etc.
 - Algorithms decide what you see when you perform a Google search
 - Algorithms recommend purchases to you when shopping online

Examples of real-life algorithms

- From *learning.com*:

- Tying your shoe-laces.
- Following a Recipe.
- Classifying Objects.
- Bedtime Routines.
- Finding a library book in the library.
- Driving or walking to or from somewhere.
- Deciding what to eat: *Deliveroo* or *Uber Eats*?

<https://www.learning.com/blog/7-examples-of-algorithms-in-everyday-life-for-students/>

Notable Algorithms

- Google's *PageRank* algorithms:
 - "...designed to sort through hundreds of billions of webpages in a *Search index* to find the most relevant, useful results in a fraction of a second, and present them in a way that helps you find what people are looking for."
- **Google's ranking systems are made up of a series of algorithms.**

From here:

https://www.google.com/intl/en_uk/search/howsearchworks/algorithms/

Google's Search Algorithms

- To give you the most useful information, Google's ***Search algorithms*** "look at many factors, including the words of a query, relevance and usability of pages, expertise of sources and your location and settings."
- "The weight applied to each factor varies depending on the nature of someone's query – e.g. the **freshness of the content plays a bigger role in answering queries about current news topics than it does about dictionary definitions.**"

From here:

https://www.google.com/intl/en_uk/search/howsearchworks/algorithms/

Examples of Algorithm Applications

- **Entertainment:** Netflix personalised recommendation generator ([Chong, 2020](#))
- **Food:** McDonald's restaurant scheduling system: controlling time/managing staff (Crawford, 2021: p.75).
- **News:** Facebook's news feed algorithm – *who sees what* in their feed ([Hutchinson, 2021](#))
- **Recruitment** algorithms to “screen and source candidates” ([Prpic, 2020](#))

Purpose of algorithms

- Weigend (2017, p.20):
 - “Algorithms find patterns that humans cannot see without computers”
 - “Such patterns can help guide decisions”
 - The “decision-making” can include:
 - Buying products and services
 - Applying for a loan
 - Exploring insurance premiums for a car/home/travel
 - Looking for a job
 - Obtaining healthcare
 - Searching for places to study

Algorithms > Machine learning

- Machine learning tools (**algorithms**) are applied to analyse big data sets to help
 - **Reveal patterns** in the data
 - **Find hidden trends** in business-held information
 - **Uncover relationships** between data points
 - **Provide competitive advantage** through *business intelligence – different from artificial intelligence!*

“ALGORITHMS OF OPPRESSION”

- From Noble (2018):
 - “Google search is one of the most underexamined areas of consumer protection policy” (p.29)
 - “... data is biased and perpetuates racism and sexism” (*ibid*)
 - “... algorithms to replicate human thinking , is predicated on specific values from specific kinds of people ---- the most powerful institutions in society and those who control them” (*ibid*)

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism.*

New York University Press

Ethical issues in algorithm development

- From Zuboff (2019)
 - Algorithms power machine intelligence, which in turn is used in **surveillance capitalism** (p. 22):
 - “...the darkening of the digital dream and its rapid mutation into a voracious and utterly novel commercial project” (p. 7);
 - “Surveillance capitalism unilaterally claims human experience as free raw material [our personal data] into behavioural data” (p. 8).
- From Noble (2018: p.1):
 - “Power of algorithms ...digital decisions reinforce oppressive social relationships and enact new modes of racial profiling”.

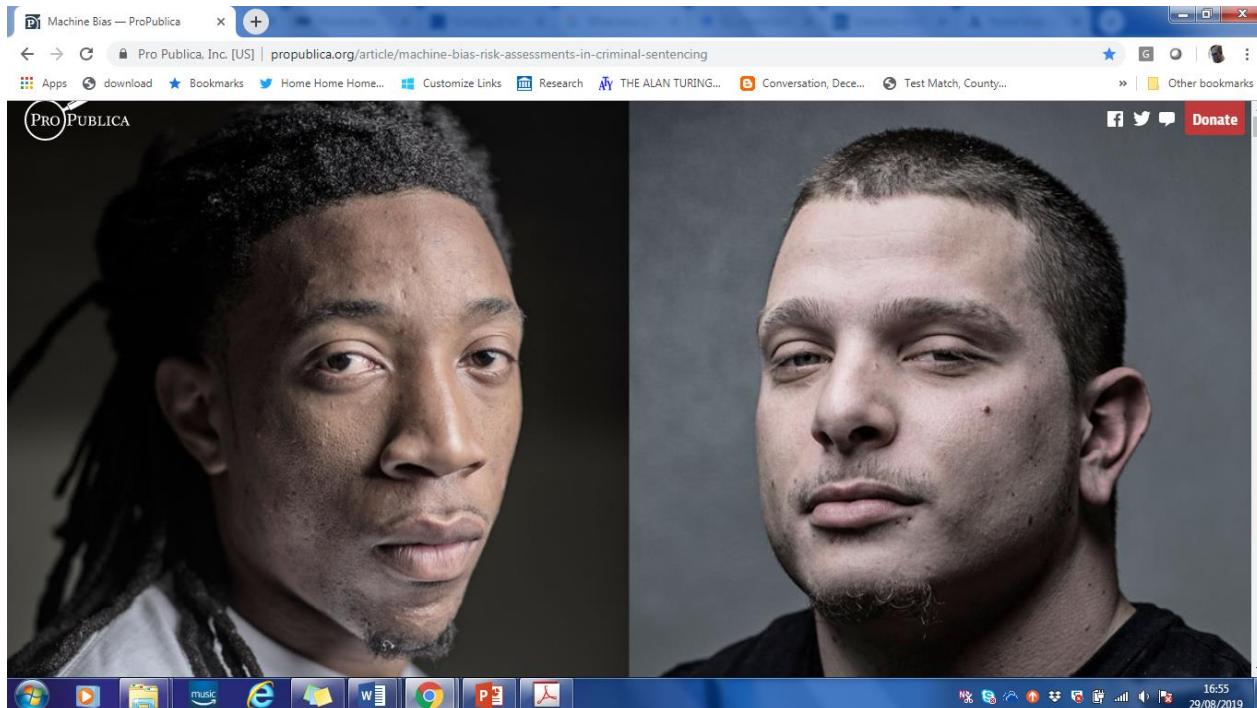
(See Reference List at the end)

Responsibility for algorithms

- Who is responsible if an algorithm *goes wrong*?
- What does '*going wrong*' look like?
- What could the effect be on a human if an algorithm *goes wrong*?

Bias, bias and more bias

AI technologies in the US Justice System



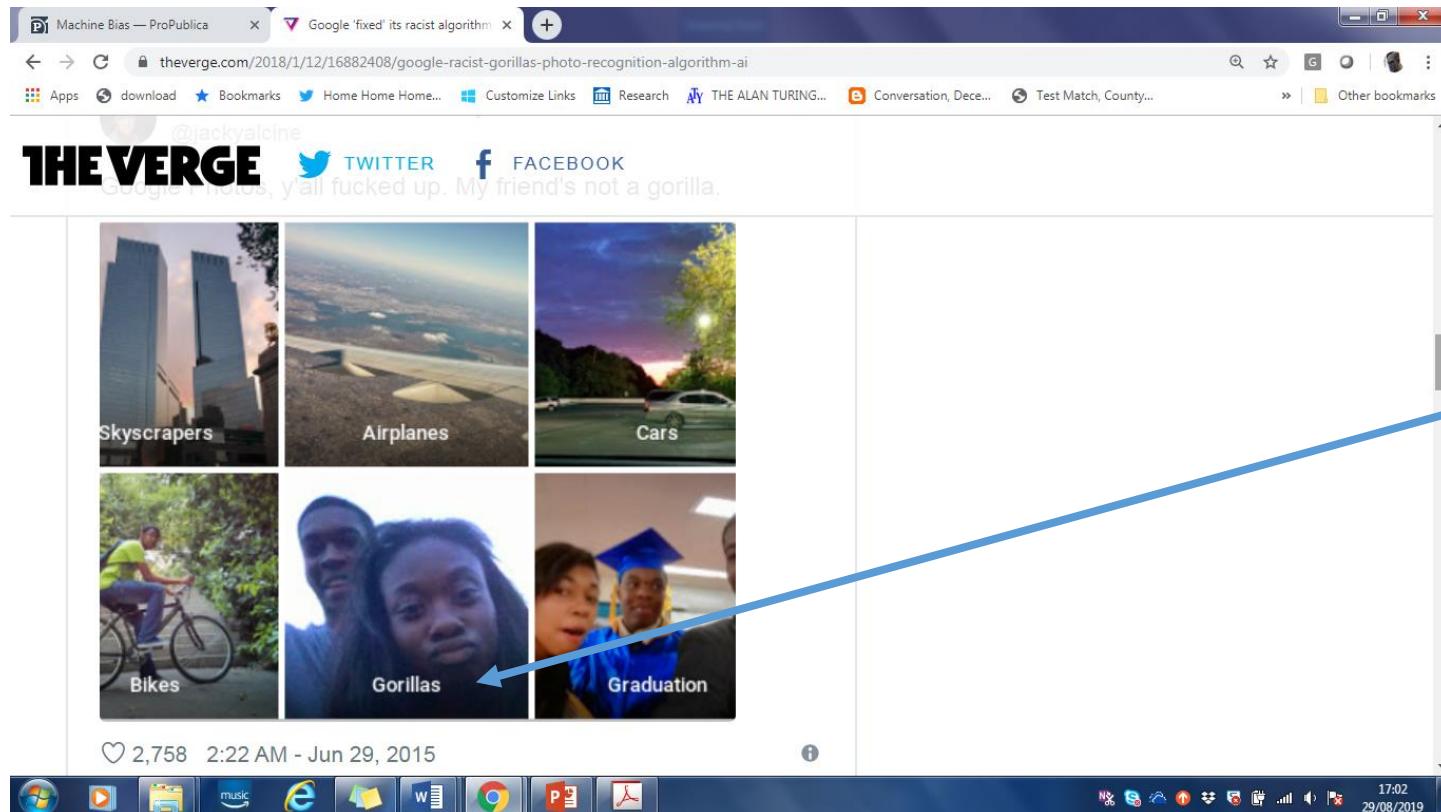
Anguin et al. (2016):

- “There’s **software** across the country [US] to **predict future criminals**. And it’s biased against blacks.”
- ”...**the algorithm was somewhat more accurate than a coin flip”**

Anguin, J., Larson, J., Mattu, S. and Kirchner, L. (2016).
Machine Bias . ProPublica, 23 May, 2016 .

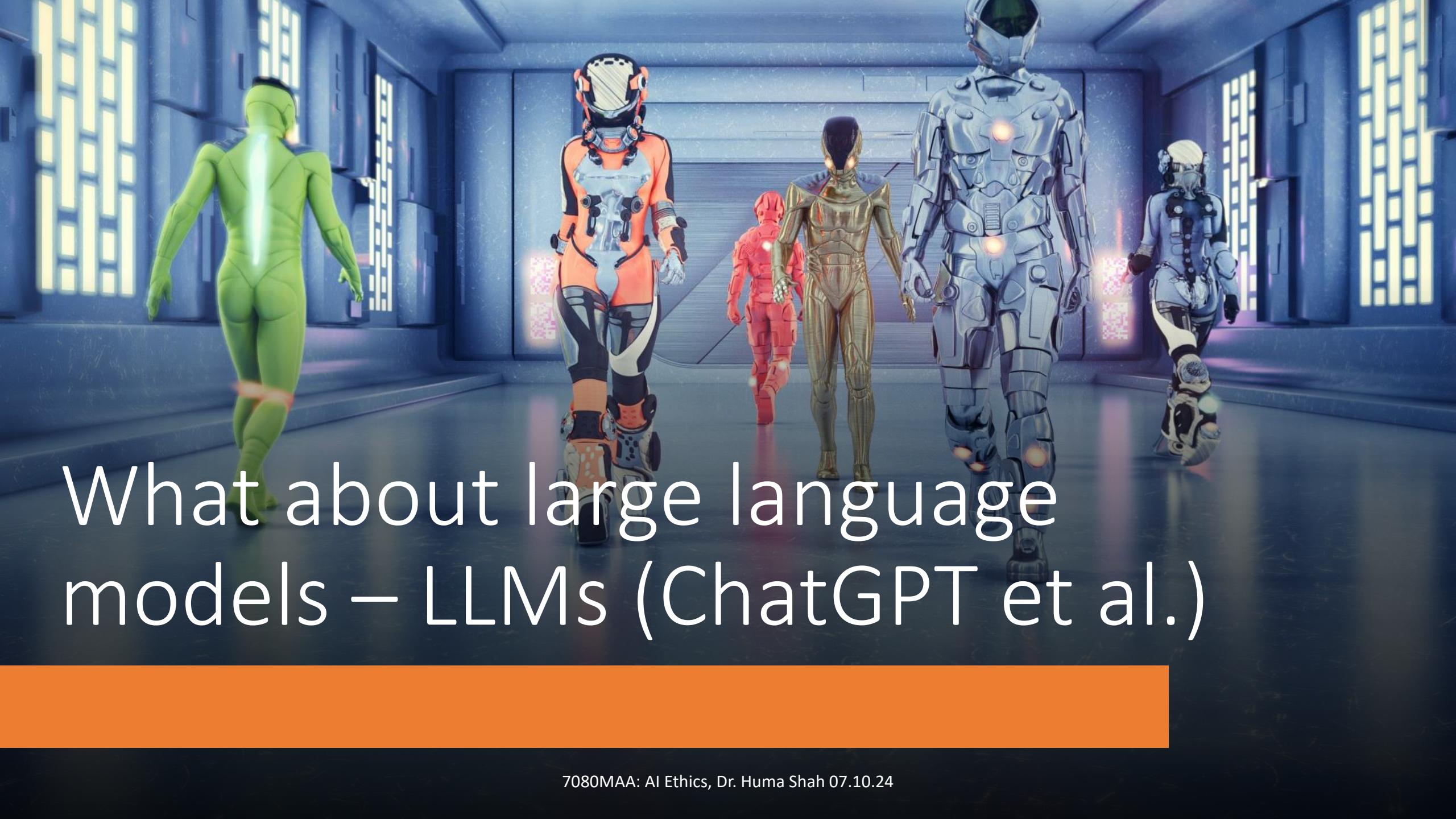
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Google's algorithm and bias



“Google ‘fixed’ its racist algorithm by removing gorillas from its image-labeling tech”

<https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai>

A futuristic hallway with blue walls and floor, featuring several AI characters in various colored suits (green, orange, red, gold, silver, blue) walking away from the viewer. The scene is set in a modern, minimalist architectural space.

What about large language
models – LLMs (ChatGPT et al.)

Death by chatbot?

- Walker (2023):
 - “A young Belgian man recently died by suicide after talking to a chatbot named ELIZA for several weeks, spurring calls for better protection of citizens and the need to raise awareness.”
 - “Without these conversations with the chatbot, my husband would still be here,” the man's widow has said She and her late husband were both in their thirties, lived a comfortable life and had two young children.”
 - “However, about two years ago, the first signs of trouble started to appear. The man became very eco-anxious and found refuge with ELIZA, the name given to a **chatbot that uses GPT-J**, an open-source artificial intelligence language model developed by EleutherAI.”
 - “After six weeks of intensive exchanges, he took his own life”

Walker, L. (2023).
“Belgian man dies by suicide following exchanges with chatbot”
Brussels Times.
<https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt?s=09>

Impressing a chatbot to kill?

- Duncan (2013):
 - “An intruder caught with a crossbow in the grounds of Windsor Castle was inspired to attack Queen Elizabeth II by Star Wars films and **encouraged by an AI chat bot**, a court has heard.
 - “In a journal, the self-styled “assassin” ... confessed that if the queen was “unobtainable” he would “go for” the “prince” as a “suitable figurehead”, in an apparent reference to the then Prince of Wales.”

Duncan, G. (2013).

“AI chat bot 'encouraged' Star Wars fan's plot to kill queen at Windsor Castle”.

MSN.

<https://www.msn.com/en-ae/news/world/ai-chat-bot-encouraged-star-wars-fans-plot-to-kill-queen-at-windsor-castle/ar-AA1duQS2>

What is the purpose of an AI?

- Comes back to the question of *what is an AI?*
 - Is it an algorithm?
 - Is it machine learning?
 - Is it a virtual human?



Ethics & Sustainability

Big Tech and Natural Resources

- “The Secret Cost of Google’s Data Centers: Billions of Gallons of Water to Cool Servers”
- “Google often puts data centers close to large population hubs to help its web services respond quickly. Sometimes that means building in hot and dry regions. The processing units inside heat up easily and water is needed to cool them down.”

From here: [Secret Cost of Google's Data Centers: Billions of Gallons of Water | TIME \(urlisolation.com\)](#)

Large language models thirst for water

- Li et al. (2023):
 - "... important and enormous water (withdrawal and consumption) footprint of AI models has remained under the radar"
 - "... training GPT-3 in Microsoft's state-of-the-art U.S. data centers can directly evaporate 700,000 liters of clean freshwater"
 - "the global AI demand may be accountable for 4.2 -- 6.6 billion cubic meters of water withdrawal in 2027, which is **more than the total annual water withdrawal of ... Denmark or half of the United Kingdom.**"

Full paper accessible here:

[\[2304.03271\] Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models \(arxiv.org\)](https://arxiv.org/abs/2304.03271)

Google falling short on climate targets

- AP News (2024):
 - “Three years ago, Google set an ambitious plan to address climate change by going “net zero,” meaning it would release no more climate-changing gases into the air than it removes, by 2030.”
 - “Rather than declining, its emissions grew 13% in 2023 over the year before. Compared to its baseline year of 2019, emissions have soared 48%.”
 - “Google cited artificial intelligence and the demand it puts on data centers, which require massive amounts of electricity, for last year’s growth.”
 - “Making that electricity by burning coal or natural gas emits greenhouse gas emissions, including carbon dioxide and methane, which warm the planet, bringing more extreme weather.”

From here:

[Google falling short of important climate target, cites electricity needs of AI | AP News](#)

Can automation benefit us?

- Matthias Spielkamp, July 22, 2019:
 - “The question of whether **automation** benefits or damages us citizens is primarily a political one. **No one should let themselves be told that only those who have studied mathematics or computer science can take part in the discussion.**”

<https://algorithmwatch.org/en/mind-the-algorithm/>

Yoshua Bengio inventor of deep learning:

- Machines “have no obvious ways of performing logical inferences, and they are also *still a long way from integrating abstract knowledge*, such as *information about what objects are, what they are for, and how they are typically used.*”

From Gary Marcus (2018) :

<https://medium.com/@GaryMarcus/the-deepest-problem-with-deep-learning-91c5991f5695>

Bengio on deep learning:

- “*The most powerful A.I. systems ... use techniques like deep learning as just one element in a very complicated ensemble of techniques, ranging from the statistical technique of Bayesian inference to deductive reasoning.*”

From Gary Marcus (2018) :

<https://medium.com/@GaryMarcus/the-deepest-problem-with-deep-learning-91c5991f5695>

M Bengio-Marcus AI Debate Post M X +

https://medium.com/@GaryMarcus/bengio-marcus-ai-debate-post-mortem-part-i-the-deep-learning-pivot-f7bd62b9861c G Paused :

Apps download Bookmarks Home Home Home... Customize Links Research THE ALAN TURING... Conversation, Dece... Test Match, County... Other bookmarks

THE STATE OF DEEP LEARNING

Amazing progress in this century

- Is it enough to just grow datasets, model sizes, computer speed?

Still far from human-level AI!

- Sample efficiency
- Human-provided labels
- Stupid errors
- Next step completely different from deep learning?

Just get a bigger brain?



Mila

Type here to search O E W P V X N S T 3:18 PM 3/26/2020 21

Bengio vs Marcus December 2019:

<https://medium.com/@GaryMarcus/bengio-marcus-ai-debate-post-mortem-part-i-the-deep-learning-pivot-f7bd62b9861c>

DEBATE : Yoshua Bengio | Gary M X +

montrealartificialintelligence.com/aidebate/ Paused

Apps download Bookmarks Home Home Home... Customize Links Research THE ALAN TURING... Conversation, Dece... Test Match, County... Other bookmarks

The Best Way Forward For AI

Deep Learning AI debate

MONTREAL.AI

REPORT ON THE AI DEBATE : <https://medium.com/@Montreal.AI/report-on-the-ai-debate-c62fcfb2ca43>

AI DEBATE: Yoshua Bengio | Gary Marcus

AI DEBATE : YOSHUA BENGIO | GARY MARCUS



Gary Marcus

Yoshua Bengio

Watch later Share

Type here to search

3:21 PM 3/26/2020 21

Warning from Dune author, Frank Herbert

- “Once men turned their thinking over to machines in the hope that this would set them free. But that only permitted other men with machines to enslave them” (1965, p. 12).
- “Thou shalt not make a machine in the likeness of a man’s mind” (Paul Atreides character in Frank Herbert’s Dune, 1965, p.12).
- “Though shalt not make a machine to **counterfeit** a human mind” (Reverend Mother character in Frank Herbert’s Dune, 1965, p.12).

Herbert, F. (1965). *Dune*. Hodder Stoughton 50th anniversary paperback edition. UK: London



X Files: S11 Episode 7 - Conversational AI in a driverless:
<https://www.youtube.com/watch?v=uV9ZDF09pkY>

It's not all bad in the AI world

- AI in healthcare – dealing with shortage in workforce
 - “... **health care systems** [take] **very structured data** and as a result of machine learning technologies, unsupervised machine learning, you may be able to get to better diagnoses from imaging, from genomics, from structured digital data.”
 - “... combining all those things together will make it even more powerful, which will give you much **better diagnostic accuracy** and will **create a new diagnostic framework** which will be done **without a huge amount of input again from the workforce.**” (**Professor Sir John Bell**, Regius Professor of Medicine at the University of Oxford)

In podcast from here:

[AI in health care: hope or hype? With Professor Sir John Bell and Dr Axel Heitmueller](#)

<https://www.youtube.com/watch?v=EqBllzgBIUo>



Summarising this week:

- This week: ethical issues in AI, but what are these?
 - ***What is artificial intelligence (AI)?*** Is there an agreement on the *I* in *AI* ?
 - ***Purpose:*** what is the purpose of any AI? (BARD, CICERO, ChatGPT?)
 - ***Responsibility:*** Who is responsible when it goes wrong? MS Tay 2016 debacle
 - ***What does go wrong look like?*** {Accidents involving driverless cars; privacy concerns}
- Other considerations: ***Should an AI be humanlike?***
 - *Should an AI be conscious so humanlike? – What is it consciousness?*
 - *What is the role of emotions in consciousness?*

7080MAA Class today: 3 main themes

- **Human Intelligence:** Learning, knowledge, language, creativity, *consciousness*, humour
- **Artificial Intelligence (AI):** Simulating, *imitating* human intelligence
- **Ethics:** responsible development to benefit the majority

References

- **Blackmore.** S. 2010. *An Introduction to Consciousness*. Hodder Education: UK [online access from Lanchester Library](#)
- **Damasio,** A.R. 1996. *Descartes; Error: Emotion, Reason and the Human Brain*. Macmillan Books: London, UK
- **Perera,** A., Gomez-Baena, A., Sundic, T., Pardo, T. and Marco, S. 2002. Machine Olfaction: Pattern recognition for the identification of aromas. *IEEE Conference on object recognition supported by user interaction for service robots*. Quebec, Canada, 11-15 August
- **Pinker,** 2. 1998. *How the Mind Works*. Penguin Science: London, UK
- **Ryman,** S.K., Bruce, N.D.B. and Freund, M.S. 2016. *Sensors and Actuators B: Chemical*. Vol. 231, 666-674. DOI: [10.1016/j.snb.2016.03.059](https://doi.org/10.1016/j.snb.2016.03.059)
- **Turing,** A. M. 1950. Computing Machinery and Intelligence', in *Mind* Vol 59 (236), 433-460
- **Zeki,** S. 1993. *A vision of the brain*. Cambridge, MA, US: Blackwell Scientific Publications.

More References

- Noble, S.U. (2018). **Algorithms of Oppression: How search engines reinforce racism.** New York University Press
- Prpic, N. (2020). **The AI Recruitment Evolution: from Amazon's biased algorithm to contextual understanding.** Talentlyft. <https://www.talentlyft.com/en/blog/article/414/the-ai-recruitment-evolution-from-amazons-biased-algorithm-to-contextual-understanding>
- Wiegend, A. (2017). **Data for the people: How to make our post-privacy economy work for you.** Basic Books.
- Zuboff, S. (2019). **The Age of Surveillance Capitalism: the fight for a human future at the new frontier of power.** Profile Books