

RIGHT AND WRONG, IN HUMANITY AND AI.



Good and evil, black and white, vice and virtue. - Since birth, we wire our brains to differentiate easily from these opposing concepts, especially when it comes to decision-making. We understand that hurting others is bad, whilst help and support are obviously linked to goodness.

But how excellent is this personal moral compass we carry anyway?

We are social creatures with a basic sense of guilt.

Various scientific research (mainly psychological and neuroscientific) tell us that morality, our mental ability to tell right from wrong, could be a product of evolution. Apart from the ongoing debates whether particular aspects of behaviour are a product of inherited (i.e., genetic) or acquired (i.e., learned) influences, here's what we're looking at:

Morality helps us live in large social groups.

It improves our capacity to get along and interact with others and without evolution or natural selection, we wouldn't really have morality.

“I fully subscribe to the judgment of those writers who maintain that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important.”

- Charles Darwin, in *The Descent of Man*.

And how do our brains form these thinking patterns?

Scientists have scanned the brains of people while they were looking at different situations, in order to find areas involved with moral thinking.

The scenes varied from someone hurting another; to a viewer having to decide whether to save five (fictional) people by letting someone else die.

The result?

As expected, there wasn't just a single "moral area" that turned on during the experiments, but several. By working together, they make up our conscience, and are referred to as the "moral network."

According to Fiery Cushman of Harvard University in Cambridge, this network is actually made up of three smaller networks.

"One brain network helps us understand other people. Another allows us to care about them. The last one helps us make decisions based on our understanding and caring."

“From a very young age, children have a basic sense of guilt, they know when they have hurt someone. They also know that it’s important for them to make things right again.”

- Amrisha Vaish at the University of Virginia School of Medicine in Charlottesville.

As we grow older and the complexity of our decisions expands accordingly, we get a bit acquainted with grey areas.

The digitalised conscious and unconscious bias.

There is a lot of talk these days within the entrepreneurial ecosystem about conscious and unconscious bias, and how they affect the workplace. From selecting CVs to interviews, and even promotions, recognising our own judgements can be difficult, especially when they’re hidden or unintentional. Of course, that influences how businesses operate.

When adding the element of advanced technologies such as AI in the mix, defining fairness in mathematical terms can bring its own set of challenges.

Algorithms are created by people and use our data. Everything that exists in our real world, is imitated by AI. And that’s quite the burden and responsibility to be held by such budding technologies.

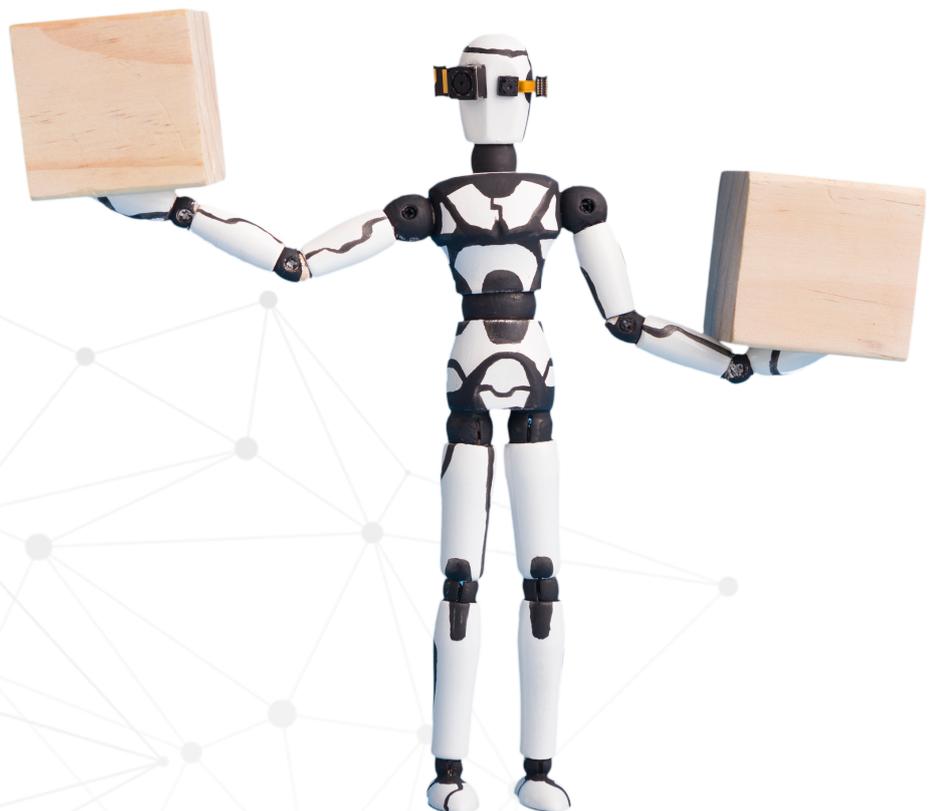
What are some problems AI is facing when it comes to biases?

Our hidden biases in data get embedded in AI engines.

The data itself rather than the algorithm is most often the main source of the issue. It can contain human decisions that don't reflect reality or perhaps match existing prejudices.

If we feed an algorithm on the data from recent books and articles, the word nurse is more likely to refer to a woman, whereas the word programmer is more likely to refer to a man. A simple **google image** search can support this, you can try it out as an experiment.

A lot of features simply can't be quantified in data.



We try to build AI that evaluates complicated qualities of data, which can lead to us settling for easy measurable shortcuts.

But not everything can be described with numbers, and our emotional and creative traits can attest to that.

I.e. When using AI to grade writing in standardised tests in order to save human graders time.

Sentence length and structure can be important features in writing essays, but they don't take into account creativity and clarity within good literature.

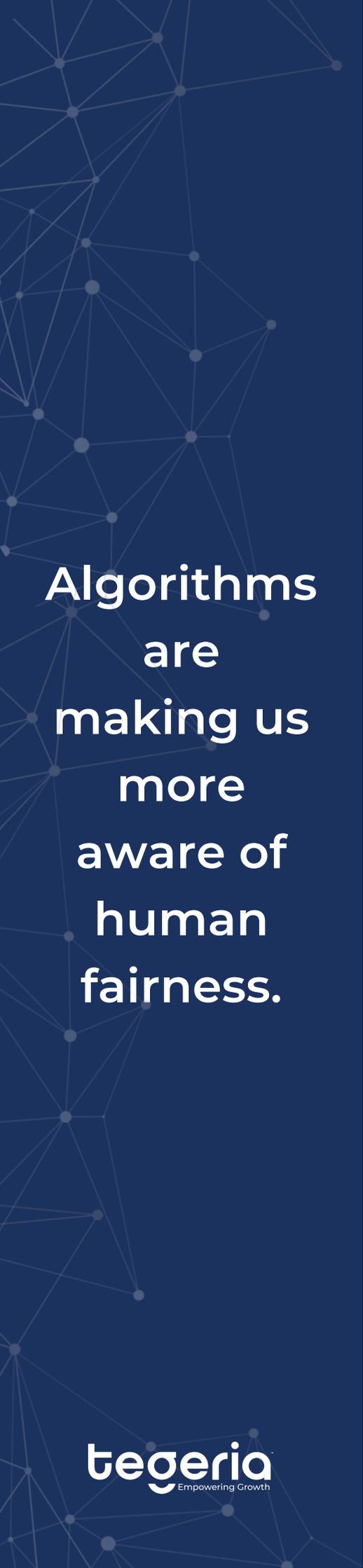
Some MIT students shared the same opinion, so they built an NLP to create essays that made no sense to test one of the algorithms.

They were all graded highly. Kind of too gracious for some nonsense don't you think?!

The positive feedback loop, which amplifies what has happened in the past.

When dealing with heavily biased data due to our socio-political problematic past (and present, for that matter), it can lead to a loop of prediction.

In criminal justice models, that looks as oversampling certain neighbourhoods because they are overpoliced. Thus resulting in recording more crime there, which results in more police (and so on, and so forth). PredPool, which helps law enforcement predict and prevent crime, can be a good example of that.



**Algorithms
are
making us
more
aware of
human
fairness.**

"It's the first time in our history as humankind that we are talking openly about unconscious biases and how to handle them so we don't discriminate against anyone through technology we produce or merely human interaction, especially in the workplace.

Gaining that kind of emotional detachment and distancing ourselves from the decision-making process is making us more logical and fair about how such decisions are made by our algorithms and hence solutions. It's easier to criticise and alter a machine than ourselves depending on such results, so it will make the measurements and return of any investment and solution (Return on Innovation - the new ROI) more tangible and realistic.

It's reverse engineering of the results we see from our solutions to change our very own ideas and perceptions to make them more sustainable and humanly fair. On the other hand, it's through the power and scalability of AI that we will be able to identify hidden problems, defects, and gaps in both our current situations (datasets) and what we are missing to get to the right outcome."

-Diana Xhumari, CEO of Tegeria.

In our daily life, we might often have “*Eureka*” moments. They come to us as shock waves, with no warning for explosion. And it’s the same kind of “aha” moment that makes us reflect when looking at an algorithmic hidden bias. We understand better. AI technology is still in its early stages of development, despite the amazing advancements and promises from businesses and use cases.

Just like a newborn, it’s fragile.

Our innovative infant needs constant collaboration and communication from all of its carers. We’re definitely far from Blade Runner or Ex Machina;

BUT

As a society, we are now building new AI that detects hidden biases, holding companies accountable to fairer outcomes and having more discussions that contemplate the different definitions of justice. All of these efforts contribute to building a better, digital future, that is also mindful of its societal impacts.

Reference:

<https://kids.frontiersin.org/article/10.3389/frym.2016.00003>

<https://www.sciencenewsforstudents.org/article/what-part-us-knows-right-wrong>

<https://unleashpotential.ie/conscious-and-unconscious-bias/>

<https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happens-and-why-its-so-hard-to-fix/>

AI Crash Course by: https://www.youtube.com/watch?v=GvYYFloV0aA&list=PL8dPuuaLjXtO65LeD2p4_Sb5XQ51par_b

<https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans>