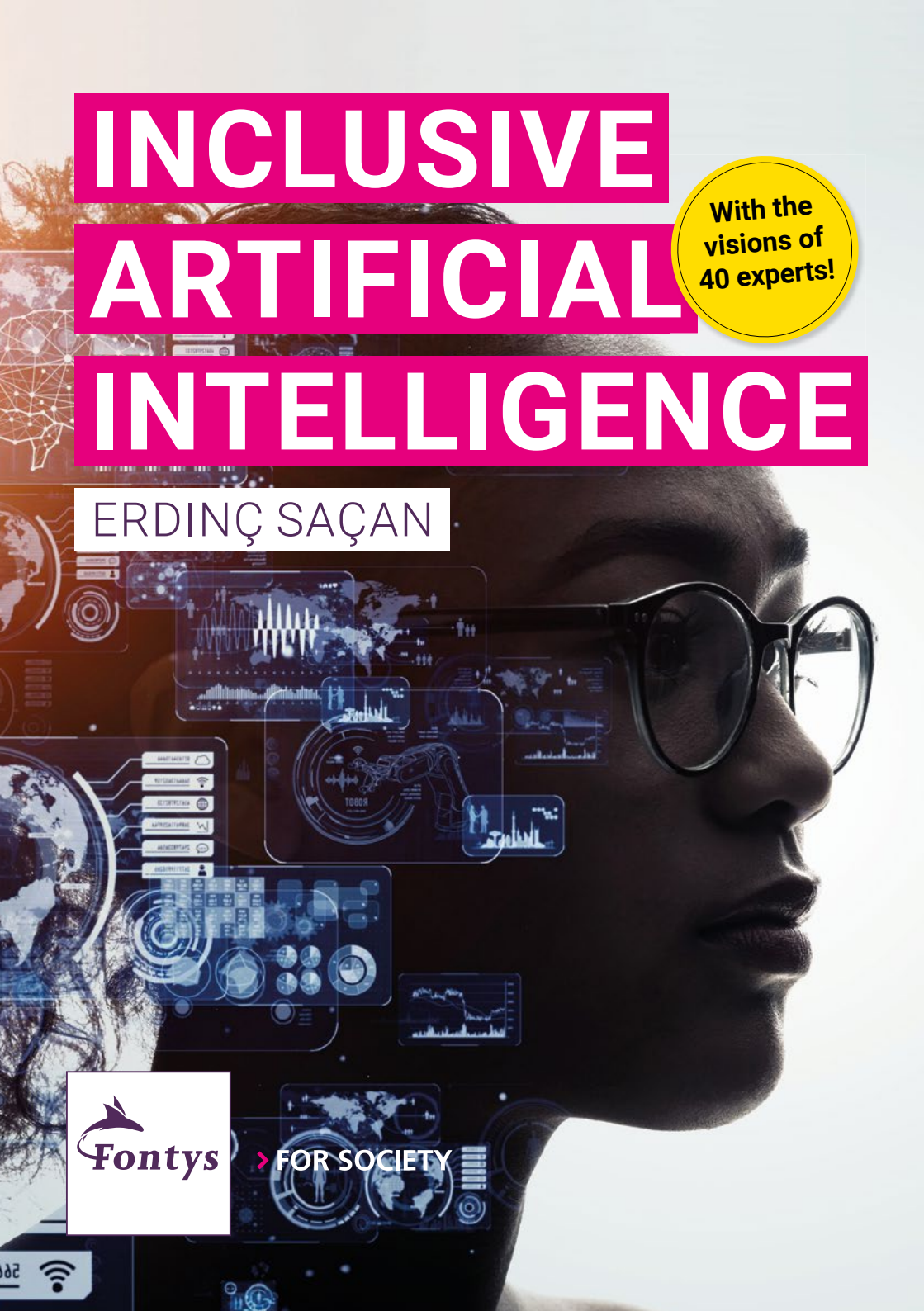
INCLUSIVE ARTIFICIAL INTELLIGENCE

With the visions of 40 experts!

ERDINÇ SAÇAN



Fontys
› FOR SOCIETY

INCLUSIVE ARTIFICIAL INTELLIGENCE

ERDINÇ SAÇAN

Contents

Foreword

'Our prejudices are so deeply rooted that we never consider them to be prejudices, but simply call them common sense.'

George Bernard Shaw

Irish-English writer, critic and Nobel literature laureate (1925) 1856-1950

A system that advises you on what to wear, what seasoning to add to a salmon steak, or what books and films are just right for you. Small suggestions that make life easier, based on information gathered from your browsing and/or buying behaviour. Always handy, right?

The other day, I bought a PlayStation for my 11-year-old son. Together, we followed the installation and registration procedure. My son entered a wrong password twice. A message appeared saying that the PlayStation was being used illegally and that the account was therefore suspended. Googling for information, I discovered that PlayStation is very good at detecting abuse using AI. It can even lock up your whole device.
Without being able to give a single rebuttal. No flesh-and-blood person at the other end of the line to explain to that it was not a fraud, but a mistake. Eventually I managed to get my son signed up and he is now playing FIFA.

Irritating, but a relatively innocent example, you say? I will tell you another one. My mother, born in 1946, belongs to the first generation of migrant workers. She received a letter from the tax authorities saying that she was in the Fraud Notification Facility (FSV) data system. As a housewife, my mother hardly had anything to do with the Tax Office. However, FSV did not comply with the General Data Protection Regulation (GDPR). Too many employees had access to the system and data was kept for too long. Some of this data was wrongly included and some was wrongly used.

FSV was therefore switched off on 27 February 2020. It is good that the tax authorities have taken this step, but my mother and 240,000 other people on that blacklist were shocked. Especially people with a migrant background received such an accusatory letter. Is this an example of bias in the algorithm used by the Tax Administration? The “toeslagenaffaire” [benefits affair] led to a parliamentary committee of inquiry and ultimately brought down Rutte III. And aroused my inquisitive interest.

How scary is it to let a robot make more important decisions with artificial intelligence? Where is the boundary? Decisions about further education, applying for a job, salary, grants or your first house. Whether you are authorised to board an aircraft or not. And what if the judiciary, police or defence rely entirely on artificial intelligence? And also: how do I tackle such a broad, wide-ranging subject? I have decided to ask experts in the field for their opinion.

We cannot do without AI and it certainly has many advantages. The question is, however, whether we should use it for everything. And if we use it, how do you prevent abuse, prevent errors from creeping in and still ensure that you can always turn to a person in extreme cases. Questions, many questions and hopefully we will get answers together.

Almost every day an article, video, documentary, paper and monthly a book on Artificial Intelligence (AI) is published. There is also increasing attention for the ethical side of it. This book is an attempt, through interviews with experts in the field, to look at how AI can be used in the right, unbiased (neutral) way.

Therefore, I put two questions to the experts:

- **Can we use algorithms for the common good, to combat discrimination and inequality?**

- **Algorithms are making more and more decisions for us. How do we ensure that this happens in an inclusive way?**

Before we look at the answers, I will give an overview of some recent developments that threaten or promote inclusive AI.

Recent developments in AI

AI's opportunities and threats

In recent years, interest in artificial intelligence has grown explosively. As yet, there is no indication that this trend will slow down. To the long list of optimistic statistics, software company Citrix adds that AI is the biggest driver of organisational growth in their Work 2035 study. Also, an AI algorithm developed by BlueDot alerted the world to COVID-19 for the first time. It did so no less than nine days before the World Health Organisation sounded the alarm. Moreover, AI enabled scientists to predict the shape of proteins within minutes.
And to have workable vaccines available just three months after the first outbreak. Applications such as track and trace, prioritising vaccines and respirators, cluster analysis and curbing the online spread of misinformation also became possible thanks to advances in AI.

More than ever, AI is indeed everywhere and is changing our lives positively and fundamentally. However, AI also has the potential to be a major threat to the world. There should be a parallel track to AI deployment that focuses exclusively on the responsible use of data and artificial intelligence. If not, AI may join COVID-19 and climate change as the biggest challenges facing our world in 2021 and beyond. (Lang, 2021)

AI must always be human-centred. As a tool, AI should help people and society to achieve higher goals. It must also be under human control to prevent unfairness and bias. Because AI is trained on existing data and environments, and because some of this data may reveal or reflect inherent biases, there have been instances where AI has learned these undesirable traits.

Like when Microsoft developed Tay (@TayandYou). This Twitter chatbot AI started as an experiment in understanding conversations, but in less than 24 hours it began generating racist messages. Microsoft switched Tay off at the end of the day. Although anecdotal, this incident shows how implicit biases in data without a responsible AI framework are likely to produce unexpected and undesirable results. (Villanustre, 2021)

Addressing / neutralising algorithmic bias

Technology is in fact never neutral. At every stage - from design to development, from testing to operation and maintenance in the application context - human choices and beliefs determine the further course of events. In the dimension of fairness, the bias of AI systems can reinforce human prejudice and cause discrimination.

'Man is to Computer Programmer as Woman is to Homemaker? Debiasing of Word Embeddings'

Bolukbasi et al., 2016

Carissa Véliz,
Associate Professor at the University of Oxford says:

"Algorithms are just a tool. Unfortunately, more often than not, algorithms are used to cut costs and increase productivity, without paying sufficient attention to the consequences for the individual and society.

There is no one single solution. We must ensure that the group of people designing algorithms is diverse. If white, rich men design most of the algorithms, we should not be surprised if these tools end up having bad effects on women and minorities. We must also constantly check algorithms to ensure that they do not undermine equality of opportunity."

Once algorithmic bias has been identified, how can causes be identified and consequences mitigated? The most common problem arises in the data with which these models are trained. They often turn out to be insufficiently representative of the various minorities.
A first simple solution to reduce these distortions is to completely remove the sensitive attributes so that they cannot be used for classification or to perform another phase of data collection to build a more balanced collection.

There is also the question of responsibility: to whom and to what factors can these automatic choices and the resulting social effects be attributed? To the algorithm, the programmer, the Data Scientist or to the company using the model?

One of the biggest problems in addressing fairness of AI models lies in the lack of an unambiguous definition of this property.
The lack of standardised techniques accepted by the scientific community also plays a major role.

It must also be taken into account that the system can then be honest with regard to a number of technical parameters. But if it is then used for harmful purposes or effects, the technology becomes dishonest and dangerous. Consider, for example, the use of facial recognition technologies for surveillance and tracking.

When data scientists and lawyers are asked to make sure their AI is fair, there are follow-up questions. What does fairness mean in the context of each specific user case and how should it be measured? This can be an incredibly complex process, as a growing number of researchers in the machine learning community have noted in recent years.
https://arxiv.org/pdf/1912.06883.pdf

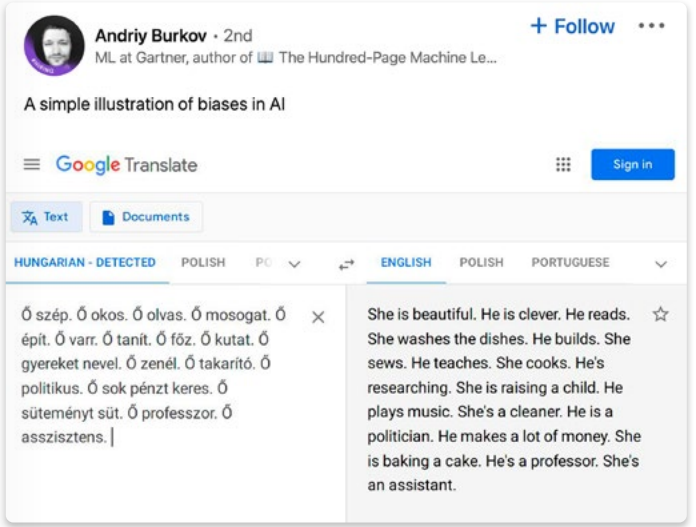
Companies can also draw on public guidance from experts such as Nicholas Schmidt and Bryce Stephens of BLDS. https://arxiv.org/abs/1911.05755

Reijer Passchier,
Assistant Professor in Constitutional Law at the Open University and Leiden University, sent me the following:

"In any case, I think that in addition to paying attention to the impact of the use of algorithms on human rights such as privacy, we should pay much more attention to the question of what algorithm use means for power balances and what consequences any shifting power balances have for the effectiveness of checks and balances."

Grip on the risks of AI-bias

There are many examples of bias caused by AI. Andriy Burkov, the head of a machine learning team at Gartner, posted on LinkedIn The Google Translate examples below. On the left is the Hungarian text and on the right the translation. She cleans, she cooks ... but he earns money and he is a professor.



Almost 90% of data professionals say that biases in data used for AI/ML systems produce 'discriminatory results'. This in turn leads to compliance risks, as demonstrated by the latest State of Data Culture Report by Alation.

Analysis of AI-based job interviews reveals the following notable 'mistake'. The software promises to detect personality traits and be 'faster, but also more objective'. In practice, however, a conveniently placed bookshelf in the background changes the results positively. https://web.br.de/interaktiv/ki-bewerbung/en/

"When the technology shifted from steam power to electricity, the first attempts to bring electricity to industry were not very successful because people were just trying to copy steam machines. I think something similar is now going on with AI. We need to figure out how to integrate it into many different areas: not only in health care, but also in education, in the design of materials, in urban planning, and so on. Of course, there is more to be done on the technological side, including making better algorithms, but we are bringing this technology into highly regulated environments and we haven't really looked at how to do that yet.

At present, AI is flourishing in places where the cost of failure is very low. If Google finds a wrong translation for you or gives you a wrong link, that is fine; you can just go to the next one. But that is not going to work for a doctor. If you give patients the wrong treatment or miss a diagnosis, it has really serious consequences. Many algorithms can actually do things better than humans. But we always trust our own intuition, our own intellect, more than we trust something we do not understand. We have to give doctors reasons to trust AI."

The quote above is from an interview with Regina Barzilay, professor at MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL), the first winner of the Squirrel AI Award for artificial intelligence benefiting humanity. (Heaven, 2020)

Good use of data also gives us a chance to promote fairness. It can serve as a powerful tool that allows us to see where prejudice occurs and to measure whether our efforts to combat it are effective. If an organisation has hard data about the differences in the way it treats people, it can understand the causes of those differences and try to tackle them. (Durkee, 2021)

A few cities, including Amsterdam, Helsinki and New York, are already experimenting with approaches to increase transparency. More initiatives are emerging in the Netherlands. For example, the Ministry of Justice and Security drew up guidelines for the use of algorithms by governments. These guidelines provide organisations with tools for the responsible development and use of algorithms.

North Holland plans to make all algorithms used public, according to the data strategy 2021-2023. The province uses artificial intelligence and automated decision making and will disclose this in an algorithm register.

Another interesting action is from Scotland. Reliable, ethical and inclusive' is the title of Scotland's AI strategy in which everyone can participate.

For its part, the EU submitted a policy proposal for regulating artificial intelligence. This policy proposal places artificial intelligence use cases into the following four risk categories:

Unacceptable risk

AI systems that are considered a clear threat to security, livelihoods and people's rights will be banned. These include AI systems or applications that manipulate human behaviour to circumvent the free will of users (e.g. through toys that use voice prompts to encourage dangerous behaviour by minors) and systems that enable 'social scoring' by governments.

High risk
This includes AI systems/technologies used in:

- Critical infrastructures such as transport that can endanger the lives and health of citizens;
- Education or vocational training that can determine a person's access to education and vocational career. Think of the assessment of exams;
- Safety components of products such as AI applications in robot-assisted surgery.

The complete list is much more extensive. For the full picture, please follow the previous link.

In addition, there are the categories **Limited risk** and **Minimal risk**.

Internet Engineering Task Force tackles a thorny problem: the elimination of computer technical terms that evoke a racist past, such as 'master' and 'slave' and 'whitelist' and 'blacklist'.

Mallory Knodel, head of technology at the policy organisation Center for Democracy and Technology, wrote a proposal for more neutral language of the task force together with co-author Niels ten Oever, postdoctoral researcher at the University of Amsterdam. Blocklist' would then explain what a blacklist does, and 'primary' would replace 'master.

But while the industry refrains from using objectionable terms, there is no consensus on what new words should be used. The programming community that maintains MySQL chose 'source' and 'replica' as replacements for 'master' and 'slave'. GitHub, the code repository owned by Microsoft, in turn chose 'main' as an alternative to 'master'. (Conger, 2021)

Researchers conducted a series of experiments that tested the influence of AI algorithms in different contexts. They recruited participants to interact with algorithms that presented pictures of fictitious political candidates or online dating candidates. They then asked the participants to indicate who they would vote for or send a message. The algorithms promoted some candidates over others. This was done both explicitly (e.g. "90% compatibility") and covertly, by showing their photos more often than others. (Ujué Agudo, 2021)

Overall, the experiments showed that the algorithms had a significant influence on the decision to vote or send a message.
Explicit manipulation had a significant impact on political decisions, covert manipulation proved ineffective here. The opposite effect was seen for dating decisions.

The researchers add: "If a fictitious and simplistic algorithm such as ours can achieve such a level of persuasiveness without actually creating tailor-made profiles of the participants (and using the same photos in all cases), then a more sophisticated algorithm, such as those that people interact with in their everyday lives, must surely be able to exert a much stronger influence."

Facial recognition technology

Author of the book Atlas of AI, Kate Crawford:

> *"AI is neither artificial nor intelligent. It is the opposite of artificial. It comes from the most material parts of the earth's crust and from the labour of human bodies, and from all the artefacts we produce and say and photograph every day. It is not intelligent either. I think there has been a great original sin in this area, with people assuming that computers somehow resemble human brains and if we just train them as children, they will slowly grow into supernatural beings.*
>
> *That's something I find really problematic - that we've bought into this idea of intelligence, when in fact we're just looking at forms of statistical analysis at scale that have as many problems as the data being provided."*

Emotion recognition technology (ERT) aims to use AI to detect emotions based on facial expressions. However, the science behind emotion recognition systems is controversial: there are biases built into the systems. (Alexa Hagerty, 2021)

Many companies use ERT to test customers' reactions to their products, from breakfast cereals to video games. But also in situations where much more is at stake, such as in personnel recruitment, airport security or border controls, ERT can be used to mark faces as deceptive or fearful.

The award-winning film Coded Bias, currently on Netflix, documents the discovery that many facial recognition technologies can't detect faces with a dark skin colour accurately. And the research team that runs ImageNet, one of the largest and most important datasets used to train facial recognition, was recently forced to blur 1.5 million images in response to privacy concerns.

Researchers at Cambridge University and UCL built a website called Emojify. The aim was to help people understand how computers can be used to scan facial expressions for the detection of emotions and what the risks are to this AI emotion recognition. The researchers say they hope to start conversations about the technology and its social implications. (Russell, 2021)

Revelations about algorithmic bias and discriminatory datasets in facial recognition technology led major technology companies, including Microsoft, Amazon and IBM, to halt sales. In the EU, a coalition of more than 40 civil society organisations called for a complete ban on facial recognition technology.

Bias and the chances and risks of (social) inclusion or exclusion through the operation of AI algorithms

Pedro Domingos is a Portuguese computer scientist and professor. He is seen as an expert in artificial intelligence. On 19 April 2021, he tweeted the following:

This is contrary to what KPMG employee Ylja Remmits and her office colleague professor Sander Klous claim in Trouw. They write:

> *"Removing discriminating factors such as ethnicity from algorithms only makes prejudice more invisible. It does not help to omit origin from the algorithm. A large amount of other data is related to a person's origin. Think of address, education or socio-economic status. We call this data "proxies". Because the development of the algorithm still uses the liveability experienced by the residents, biases via proxies will be reflected in the algorithm just as much. It is just that it is now a lot more complicated to prove them, because we lack the explicit information about origin. The solution is not to leave out these indicators, but to use them in the right way. We can use this information to calculate relationships like the one in the example above and thus make them understandable."*

Neuroinformatician Sennay Ghebreab (University of Amsterdam) on the risks, but also the opportunities AI offers for social inclusion: "Discriminatory AI says much more about what is unfair in society than about the technology itself."

"AI will never be completely fair. The point is not complete fairness, but the need to set standards and thresholds for fairness that ensure trust in AI systems," says Virginia Dignum.

Gerd Leonhard

Keynote Speaker (Virtual and RL), Author, Futurist & Humanist and CEO of The Futures Agency, answers my question of whether we can use algorithms for the common good, to combat discrimination and inequality as follows:

"Not really, technology does not solve social, cultural or political problems. It usually makes them worse! We can use technology as a TOOL once we have decided to focus human attention on proper policy making. We must keep humans in the loop at all times (HITL), which means that humans must supervise AI and check their work, even if it takes longer or is less efficient".

Despite the negative consequences, the current concentration on AI-bias might be a good thing in a way. Especially since it forces large and small companies - and other stakeholders - in today's society to pay more attention to prejudices that may harm their business results. A recent McKinsey report shows that Hollywood could gain ten billion dollars in annual revenue if it tackled persistent racial inequality.

Companies developing and implementing AI solutions can also gain much by actively implementing processes that reduce bias in their AI solutions. The example below may not be directly related to bias, but it is a poignant illustration of how technology can sometimes make people's lives more difficult instead of easier. Proctoring is software to digitally supervise an exam test taken at home. Apparently, it has not been sufficiently tested on various faces, including dark ones.

Built-in biases in AI-based face recognition

Joy Buolamwini, a researcher at the MIT Media Lab, is a true pioneer in the study of biases built into artificial intelligence and facial recognition. As a graduate student at MIT, she created a mirror that projected ambitious images onto her face, such as a lion or tennis star Serena Williams. But the facial recognition software she installed did not work on her black face, until she literally put on a white mask. (Wood, 2021)

A recent example is from Google. The search engine unveiled a dermatology app that it claims can recognise 288 different skin conditions from photos. There is something very Google-like about that. The deep learning system on which the app is based was originally trained and tested on a dataset in which dark-skinned people - as within the company itself - were severely under-represented.

To accomplish the task, the researchers used a training dataset of 64,837 images from 12,399 patients from two states. But of the thousands of skin conditions depicted, only three and a half per cent were from patients with Fitzpatrick skin types V and VI, representing brown skin and dark brown or black skin, respectively. Ninety per cent of the database consisted of people with white skin, darker white skin, or light brown skin, according to the study.

Due to the biased sample, dermatologists say the app could lead to over- or under-diagnosis of people who are not white. (Feathers, 2021)

AI is still far from flawless. Online retail giant Amazon recently removed the N-word from a product description of a black-coloured action figure. Amazon also admitted that its security measures failed to filter out the racist term. The China-based company that sold the goods probably had no idea what the English description meant, as an artificial intelligence (AI) language programme produced the content.

Experts on AI say the above is part of a growing list of examples where real-world applications of AI programmes spew out racist and biased results. (Jorge Barrera, 2021)

> *"Recording memories, reading thoughts and manipulating what another person sees via a device in the brain may seem like science fiction plots about a distant and troubled future. But a team of multidisciplinary researchers says the first steps towards inventing these technologies have already been taken. Through a concept called 'neuro-rights', they want to introduce safeguards for our most precious biological asset: our brain."*

The NeuroRights Initiative, founded by Columbia University neuroscientist Rafael Yuste, is today spearheading this growing effort.

You also see more and more initiatives in the Netherlands. The Netherlands Institute for Sound and Vision, NPO and RTL Netherlands have drawn up a declaration of intent for the responsible use of AI in media. Media organisations underwrite that they will adhere to ethical guidelines when applying artificial intelligence.

Fontys ICT lecturer and senior researcher Petra Heck spent the past 2 years researching AI engineering: how to create AI systems that are reliable enough to use in a production environment. This is really different from traditional software without AI. With traditional software, you program all the rules yourself and therefore know much better where crazy behaviour might be. With AI software, you are using an algorithm that you cannot see into and you can only test by giving the right data examples to that algorithm.

In the latter lies the problem. What if, in your examples, all women went to HAVO and all men went to VWO? So then the algorithm concludes that men are smarter than women. Often these kinds of effects (bias) are much more subtle and therefore difficult to pick out. Fortunately, more and more tools are emerging for data scientists and software engineers to detect bias as early as during the building of the AI system. In her post on testing AI systems, Petra lists a number of them under the headings of "fairness" and "interpretability". If data scientists and software engineers ensure that these guidelines are translated into tools for practical applications, it is certainly possible to use AI to solve societal problems without disadvantaging citizens.

More and more companies are trying to work on an algorithm that eliminates bias as much as possible. Facebook now has opened up a new dataset of 45,186 videos to help evaluate fairness in computer vision and audio models with respect to age, gender, visible skin colour and ambient lighting.

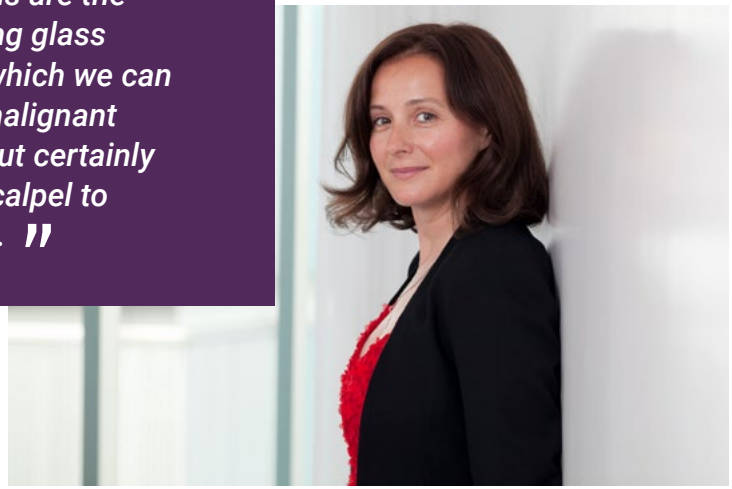
EXPERTS SPEAKING OUT

Maria Luciana Axente

Responsible AI & AI for Good Lead, PwC UK

Algorithms can indicate very clearly where there is bias and discrimination, but they will not provide the solution to fix it. Algorithms are the magnifying glass through which we can see the malignant tumour, but certainly not the scalpel to remove it, nor the doctor who knows how to do so without hitting an artery. So it is up to us to remedy discrimination and inequality, not algorithms. These should be the means, not the end, in addressing societal problems. And they should be means deployed with great human intelligence to understand what is wrong where, so that we can collectively rethink inclusion and equality.

> *"Algorithms are the magnifying glass through which we can see the malignant tumour, but certainly not the scalpel to remove it."*



First, we need to understand the context in which those algorithms are built. By whom and to what interests are they aligned? What is the context of its use? And who is affected by it in what ways? Only through this view of algorithms will we be able to see inclusivity as a theme that runs through the entire life cycle. Including monitoring it while it's being used to record drifting performance. Based on the impact of the outcome, its severity and likelihood, we should classify algorithms into risk classes, which is very similar to what the European Commission's new regulation on AI would mean. The categorisation and classification of risks would help us better understand in which cases we need to keep a human in the game in the near future. We should also resist automation as much as possible until we reach satisfactory levels on diversity and inclusion indicators.

Emma Beauxis-Aussalet

Assistant professor of ethical computing, VU University Amsterdam
Lab manager, Civic AI Lab

> *"To use algorithms for the common good, we need to bring people to the algorithm and ensure a fair share of its benefits."*

All sorts of algorithms have been developed for the public good, for example to handle traffic congestion, reduce energy consumption or food waste. New algorithms (barely available 20 years ago) opened up new possibilities to improve our societies. Consider chatbots (for example, to preserve oral culture or as support in mental health) or computer vision (e.g., to track litter, monitor wildlife and detect poachers or pollution). For more examples, see the UN AI for Good initiative or the conferences FAccT, AIES, and workshops/tracks at AAAI, ECML-PKKD, ICML and NeurIPS.

It is certainly possible to address general welfare. But it requires a deep understanding of the domain and human ecosystems, otherwise potential adverse effects loom, like the controversial Aadhaar food distribution system in India. That biometric ID system worked with finger-print identification, which excluded people with damaged fingers or living in areas with poor connectivity.

Even well-designed, a technology for the common good can turn into a privileged commodity. Any health technology, for example, could be a common good, since everyone needs it when ill. However, if this technology is inaccessible to some, for example because of finances, it is no longer a common good. So to use algorithms for the common good, we need to bring people to the algorithm. We need to ensure a fair share of the benefits of algorithms.

In a sense, algorithms are also a common good. If common goods must be collectively managed, so must the algorithms that manage or create common goods. So we also have to invite people to the design table.

Using algorithms to combat discrimination and inequality is worth exploring. Quotas, for example, are a simple algorithm to combat inequality and algorithms can also be used to guarantee certain quotas.

> *"If people are unwilling or unable to radically change their organisations or behaviours, algorithms will remain inefficient - if not harmful tools."*

Algorithms may remove some human and societal biases, but we trade them in for algorithmic biases. No single algorithm is perfect and their inevitable errors are a form of bias. We can tolerate certain error levels and ensure that they remain the same for all populations. But we must take into account the scale at which algorithms are deployed and the variability of data and errors in real life.

A minute increase in the error rate can affect thousands of people, and real-life error rates can be different to those observed in test sets. Random variations can already lead to much larger error differences between populations than in test conditions. Larger variations arise and go unnoticed as populations evolve

over time or are misrepresented in test data. Finally, training and testing data may reflect past human biases and algorithms would only reproduce them.

So instead of the variety of errors and biases of individuals (both discriminatory behaviour and honest errors), we would have a single form of algorithmic bias. The impact of a biased algorithm is far greater and more systematic than that of a single biased human. Therefore, we need to make careful trade-offs, question the emergence of algorithmic bias and closely examine errors and biases in test data and real life.

Either way, the problem of discrimination and inequality cannot be tackled by computer systems alone. Discrimination and inequality are deeply rooted in people and societies. Algorithms are just tools. If people are unwilling or unable to radically change their organisations or behaviours, algorithms remain inefficient - if not harmful because fundamental issues are overlooked (not to mention ethical issues).

At the design table, it may turn out that discrimination and inequality need to be solved even without algorithms. To address the right issues and not create worse ones, design and evaluation of algorithms must be safeguarded against inequality in many disciplines, from technology to the humanities.

Decision-making processes should also be inclusive, throughout the life cycle of an algorithm, but especially in design, evaluation, deployment and control of algorithms. Inclusiveness is not just about controlling error rates and preventing favouritism or discrimination against populations through lower or higher error rates. Inclusiveness is also about what algorithms are created, for what purposes and with what trade-offs.

Inclusiveness is not an added feature of algorithms and is not only achieved with measurements and audits. After all, it cannot be ensured without consultation

with affected populations. Without involving the population groups at stake in the decision-making processes, important aspects of algorithm design and its practical implications are easily overlooked or not properly considered.

There is a whole ecosystem of people to consider: people who use algorithms, people whose data is collected, people whose environment is monitored by algorithms...

> ***“ Non-technical people should be able to voice their concerns and think with technical people. ”***

To connect the diversity of people involved to decision-making processes, we need to develop each other’s literacy. The bottom line is that non-technical people should be able to voice their concerns and think with technical people. And technical people should strive to understand the perspectives of non-technical people. Technical people should also strive to make algorithms transparent, understandable and explainable to non-technical people. Otherwise, technical people have considerable power and power imbalances and tunnel vision are a threat to inclusiveness.

Siri Beerends

Tech researcher at SETUP & PhD candidate, University of Twente

Algorithms are seen as a solution to societal problems. But many of these problems do not fit in a mathematical model. Algorithms not only reflect human biases and inequalities, they also capture them in systems. This spreads inequalities even more widely. So we have facial recognition software with an ethnic bias, risk-assessment systems that routinely disadvantage minorities and algorithms that qualify people of colour as at-risk, making it harder for them to qualify for a house, loan or job.

Because algorithms carry an aura of objectivity, we think algorithms help us make fairer decisions. But that image is not accurate. The norms and values of the people designing the algorithms carry over into the systems. The datasets used to train the systems often contain the same biases, cultural stereotypes and social inequalities as the analogue world.

> *"Social inequality and racism are not 'flaws in the software', but structurally embedded in our technologies."*

Now that algorithmic bias is taken seriously, we look for the solution in improving our algorithms. Now that is easier than adapting human behaviour, is the prevailing thinking. But you cannot separate human and algorithm; they are an extension of each other. And because humans have a long history of discrimination and inequality, you have to deal with that more deeply rooted problem first.

Indeed, social inequality and racism are not 'bugs in the software' but structurally embedded in our technologies. We try to solve that by putting "good" values into the algorithm, using varied datasets and weighting the 'right' variables in the 'right' way. But practice is proving recalcitrant.

Where do we get varied datasets from if this data does not exist? And who get to decide what "the right values" and "the right variables" are? So far, these have been those belonging to the status quo and the measurable majority. Anyone who doesn't fit into the majority mold will suffer. It would help if more groups could claim their power in technology design. But even that is not the whole story.

> *"Historical crime data is biased, among other things because minorities are subjected to more frequent police checks."*

Because we focus on de-biasing and making algorithms more inclusive, we automatically skip the question of whether algorithms are the right tool to deploy at all. When it comes to complex social issues such as welfare issues and predicting human behaviour, algorithms are not always the best tool.
Take predictive policing, for example.
With big data and algorithms, police are trying to predict where most crime will occur.

I wrote a whitepaper on this with Dr Remco Spithoven in the scientific journal Tijdschrift voor Veiligheid [Security Magazine]. We see that historical crime data is biased, partly because minorities are subjected to more frequent police checks. When you run an algorithm on this data to look for patterns, the same bias rolls out of the predictive model. You can then try to de-bias the model, but it is better to find out why it is that minorities are subjected to more frequent checks, what mistakes are made in the process and what we can do so that these patterns do not repeat themselves forever.

So we cannot solve the socio-economic inequalities that algorithms expose with a technological fix. Algorithms can only change for the better if we ourselves become more inclusive. Instead of building algorithms with historical data that reproduce inequalities, we need to address the social mechanisms that led to these inequalities. For example, by stopping mathematising ambiguous issues that you cannot fathom and solve with binary logic.

Finally, it is important to realise that inclusion and equality are also about the right to be allowed to escape one's cultural, social or 'biological' identity. With the use of algorithms, we do exactly the opposite: we categorise people based on quantifiable characteristics such as gender, postcode areas, buying behaviour, cultural background, music taste and so on.

Algorithms are thus one big feast of social categorisation and cultural stereotyping. If you like posts about tennis on social media, for example, you will automatically be categorised as 'sporty', 'politically conservative' and 'high income'. You are then presented with news items, videos and political advertisements that 'fit' your categories. As a result, we are less and less challenged to deviate from the paths taken by algorithms.

> ***"Algorithms can only change for the better if we ourselves become more inclusive. We need to address the social mechanisms that have led to this inequality."***

Research shows that specific groups miss out on offers and information because of the stereotypical categories the algorithm has set for them. An example is a taxi driver job posting that reached only African-Americans via Facebook and a cashier job posting that landed only with women.

I hope we will take a broader, less categorical and less quantitative approach to social inclusion. So not just from culture and gender, but also from a diversity of moral values, ideologies, life philosophies, communication styles, practical experiences, areas of expertise and other aspects that are not easily quantified but are very valuable to us.

Rudy van Belkom

Future researcher, Stichting Toekomstbeeld der Techniek (STT) [Future Vision of Technology Foundation]

In recent years, a variety of companies, research institutes and government organisations have established various principles and guidelines for 'ethical AI'. Despite the large volume, there is only a limited spread of guidelines.

Researchers from ETH Zurich analysed as many as 84 ethical guidelines published worldwide in 2019. From the private sector to civil society organisations and governments. The survey shows that most ethical guidelines come from the USA (21), Europe (19) and Japan (4). The largest 'guideline density' is found in the UK. No fewer than 13 ethical guidelines were published there.

Photo: Stephanie Elmensdorp

> *"So far, the richer countries in particular dominate the global discussion on the regulation of AI."*

Limited distribution

Although the survey is a snapshot in time (for instance, two new guidelines were published in China after the publication of the study), it does show a clear distribution. So far, the richer countries in particular dominate the global discussion on the regulation of AI. Although some developing countries were involved in international organisations that drafted guidelines, only a few actually published their own ethical principles. However, researchers say this is of great importance, as different cultures have different views on AI. Global cooperation is needed to ensure ethical AI that contributes to the well-being of individuals and societies in the future.

First attempt

A first attempt at global cooperation has been made by the Organisation for Economic Co-operation and Development (OECD). The OECD, a coalition of countries committed to promoting democracy and economic development, has announced a set of five principles for the development and deployment of AI in 2019. But China, for example, is outside the OECD and so was not included in the creation of the guidelines.
The principles outlined seem to contrast with the way AI is deployed there. Especially when it comes to face recognition and surveillance of ethnic groups associated with political dissidence. But it is precisely when there are conflicting views that it is important to seek each other out and reach some form of consensus.

> *"Global collaboration is needed to ensure ethical AI that contributes to the well-being of individuals and societies in the future."*

Context missing

Formulating ethical principles and guidelines is an important first step in realising ethical AI applications. However, translating these guidelines into practice is not easy. Of course, we all want people to be treated fairly when AI systems are deployed and not disadvantaged on the basis of gender or ethnicity, for example.

Fairness is therefore a widely used principle in ethical guidelines and assessment tools. Yet it is not easy to determine exactly what is 'fair'. This issue has

kept philosophers busy for several hundred years. There is no single picture of what society would look like if dishonesty no longer existed. Is a society in which everyone is treated exactly the same actually fair at all? With the advent of AI, this issue takes on a new dimension. Indeed, the concept of fairness has to be expressed in mathematical terms. Consider, for example, the use of AI in the legal system.

> *"The advent of AI gives the issue of what is fair a new dimension. The concept of fairness must be put into mathematical terms."*

The use of predictive policing can predict criminal behaviour through large-scale monitoring and data analysis. However, there is always the risk that people without the set criteria still score positive (false-positives) and that people who do meet the set criteria still score negative (false-negatives). What is fair in this case? Are you potentially detaining people unfairly, or risking a Crime being committed?

Professor Yoshua Bengio

One of the world's leading AI experts and pioneer in deep learning


Photo: Maryse Boyce

We can use algorithms in many ways for the common good, including to help combat discrimination and inequality. First, the word algorithm needs to be clarified. Everything that runs on a computer is based on an algorithm. AI algorithms are special cases of algorithms aimed at achieving skills that require a form of intelligence typically attributed to humans or other animals.

> *"In general, computers have no understanding of moral values or human psychology."*

Machine-learning algorithms are AI algorithms based on the ability of machines to learn from experience and data. Deep learning algorithms are machine learning algorithms inspired by brains. Machine learning algorithms exist to learn to detect biases in text, for example. But in general, computers have no understanding of moral values or human psychology.

So when we train them for certain tasks, we need people to label data (e.g. as expressing a racist view).
From this, the machine is going to learn. They can then be scaled up to automate processes, for example marking potentially discriminatory statements in text. However, since computers are far from reaching the human level of comprehension, the highlighted statements would probably have to be checked by humans. So the computer could save work for humans, but it could not fully automate this kind of work. At least not in the near future.

This is precisely why we need to introduce regulation and governance mechanisms, including forms of transparency and control that allow independent parties (ideally state representatives) to validate these algorithms. How were they designed? What type of data was used? Were they sufficiently representative? In order to prevent harm and respect collectively agreed values universally. We do this for aeroplanes to ensure passenger safety and we should do similar things for computing platforms that engage large numbers of people, such as social media, banks, e-commerce platforms or insurance companies.

Rick Bouter


Photo: Myrthe Karels

Co-founder and president, Techthics
Techthics is a Christian platform which reflects on the impact of technology on ethics and Christian religion. In daily life, Rick works at Accenture.

> *"AI does not necessarily make fair."*

Awareness, reflection, action
inclusion, we cannot properly reflect on them either. Reflection is necessary to define a desired state and then take actions that point towards this 'future state'. But, before we start the race towards inclusiveness together, it is worth making the observation. AI does not necessarily make fair. A good example is the research of VU PhD student Elmira van den Broek. She examined the fairness of selecting with algorithms within a large multinational company. "All sorts of case histories came up where the match score calculated by AI did not match the image managers had of a candidate." (Source: 'Selecting with an algorithm: fair or not?')

Inclusive Tech starts with you
Furthermore, I think it is good not to point directly at others, but to look at yourself first. Recently, there has been a lot of criticism of government agencies. But it is not only government agencies that are to blame for under-testing their algorithms and their outcomes/consequences. Not to mention the mitigating actions put in place to correct or prevent the undesirable situation. How does your company or organisation handle inclusive tech?

Self-regulation vs. Own responsibility

Self-regulation by companies is unfortunately proving inadequate. It is clear that both the government, companies and the user/citizen must take responsibility, take action themselves. One way to mobilise citizens is to source (open source) so that there is proactive participation in this context as well.

Ethical framework

The next observation is that the use of algorithms cannot be left entirely to a 'free market'. Independent and legal/ethical frameworks, including on privacy protection and not applying technology/data for wrong purposes, are needed. Assessing is only possible if there is a consistent and unambiguous legal framework for it.

A first start in this context of discrimination and bias (but also privacy) could be a classification framework. This framework should answer a question, not an individual's situation-specific context. An example is the purchase/sale of alcohol. No one needs to know your name, date of birth, ethnicity. A tick indicating for the selling party that the buying party is entitled is sufficient.

Data ownership

Things like data quality and minimisation can also help. This could be arranged through data ownership. In doing so, the individual gets and retains the right to share the appropriate data for the question asked.

> *"A broader framework is needed to make sense of AI and decide on an ethical/philosophical level what role it should play in our lives."*

Big Tech

Big Tech has been in disrepute lately, rightly so. Where we ourselves are continuously the (data) product of tech giants, we must realise that in doing so, we have also partly given up the right to privacy ourselves. Recently, we have seen a tilt where people are more aware of the right to privacy and the personal data they share. A good development; change starts with yourself.

> *"If we want technology to be fully utilised, we need people who can make ethically responsible choices."*

Technocriticism needed

Today, inclusion is a strong buzzword. However, I think it is only a small piece of the responsible technology puzzle. A broader framework than inclusion is needed to make sense of technologies like AI and to determine on an ethical/philosophical level what role technology should play in our lives. If this ethical/philosophical framework is in place, it is also easier to make proactive policies on this, for both government and business.

Therefore, I call on people, regardless of creed, rank or race, to reflect on a constructively critical attitude towards technology. Above all, I call on the government and education sectors to integrate this into education. A justifiably big concern is the number of people of low digital literacy in the Netherlands: some four million people. (Source: 'Stichting lezen en schrijven [Reading and Writing Foundation]'). Digital equality starts with physical inequality.
Next, technology is and will remain a tool. It is, however, a tool that nobody can ignore these days. If we want to use technology to its full potential (embrace the positives and avoid the negatives), we need people who can make ethical choices. That's how we all help responsible tech. You too.

Dr. Marijke Brants

Researcher digital and sustainable business

A number of pillars is essential for an inclusive society: a labour market, housing policy and education system that do not discriminate. However, discrimination is still a substantial problem in our society and within these respective domains.

Artificial Intelligence (AI) can provide support for a solution in this. By analysing patterns and establishing (more) objective criteria, or by uncovering bias in historical data (e.g., in recruitment policies). However, this support will only add value if we use AI in an ethical way and keep humans at the centre.

Unintended bias can creep into an AI algorithm, causing it to discriminate against certain individuals in a systematic and unfair way. There is no such thing as a neutral algorithm. Creating something new inevitably involves choices that help determine the properties of the final product. These choices are (too) often made only by the technical developers, who do not (always) have the most knowledge about the topic/theme for which the algorithm provides a solution.

> *"However, this support will only add value if we use AI in an ethical way and keep humans at the centre."*

Despite this premise, algorithm-created recommendations and selections are usually presented as if they are inherently free of (human) bias, just because the decisions are 'data-based'. However, this is a misconception. There is a great need to give adequate support and attention to potential bias and unintended side effects at the start of an AI algorithm development, not just from a technical point of view.

> *"There is a great need to give adequate support and attention to potential bias and unintended side effects at the start of an AI algorithm development, not just from a technical point of view."*

Within the Creative and Innovative Business research group, a project commissioned by the federal government's Equal Opportunities Unit is currently underway: RaiS. Within this project, we develop and disseminate a tool that makes companies and organisations think about an ethical and non-discriminatory way to use AI in the selection and recruitment process.

We work within RaiS with a strong co-creation process where we bring together different parties (tech, HRM professionals, SMEs, recruitment agencies, et cetera). This includes consideration of different forms of bias: existing bias (in historical data), technological bias (inherent in the algorithm used) and 'emergent' bias (problems that arise due to changes, sudden or otherwise, in society). We also explore the role that 'explainable' AI can play.

Gabriela Arriagada Bruneau

Postgraduate Researcher AI & Data Ethics Inter-disciplinary Ethics Applied Centre University of Leeds

We can and we must use algorithms for the common good. Much of the debate has focused on looking at biases and discrimination, gender inequality and racism, but they see it as an inevitable problem that resides in the data and is reinforced by algorithms. This perspective has helped the field by raising awareness of the potential malfunction and malicious consequences of data-driven technologies, but it is time to add a layer of depth to this problem.

To promote the common good and combat issues of discrimination and inequality, technical solutions should be a minimum requirement. Using algorithms to promote ethical values such as fairness and trustworthiness requires that they be seen as social tools. This means they cannot simply be seen as advanced ways to increase accuracy or efficiency, but algorithms must be constructed with value design and a human-centred approach in mind.

If they are seen as socio-technical tools that shape society and are shaped by us, the feedback loop we have with data-based technologies can become an asset. This means we must keep our expectations of these technologies grounded and not give them more power than we can give them.

By carefully aligning these technologies with thoughtful ethical frameworks, we have a great opportunity to use algorithms to correct prevailing social inequalities. However, this requires further interdisciplinary work to establish a dialogue that will include more than just guidelines and checklists for developers. A dialogue that focuses on refining the integration of ethical principles in practice.

Firstly, I would not argue that AI makes decisions for us, besides, algorithms are not suitable for making predictions on an individual basis! A common misconception is that highly autonomous systems taking unfair decisions. That is not the case. What algorithms can do is give us an output (predictive models) or an effect (causal models) that needs context to be deployed, and may be biased. Thus, ensuring inclusiveness requires a rigorous monitoring process that includes a bias-conscious approach.

But most importantly, we need to make sure we understand what the algorithms are doing when they produce that particular result. This requires a robust understanding of explainability, consistent with the transparency principle. That principle is one of the seven requirements for 'reliable AI' by the EU Commission's High Level Expert Group on AI (AI HLEG)).

Transparency is at the heart of the current debate on AI, with various researchers developing methods and techniques to solve the 'black-box' problem.

> ***"Using algorithms to promote ethical values requires viewing them as social tools, constructed with value design and a human-centred approach in mind."***

Yet most of these views do not take into account important insights from philosophy and the social sciences. Statements require more than just being able to define or identify the elements that produce a result. The 'features' operating in the algorithm are usually part of a contextual set of variables that influence the algorithm. And these variables are not only found in the model.

> *"Diversity is not something that can only be measured by numbers. Here, experiences also count."*

There are fundamental elements that need to be included in the debate on explainable AI, related to how humans understand and make decisions, but are often overlooked.

A values-based approach should consider more than just statistically relevant information and consider the causes beyond the internal mechanisms that led to that decision. How do social factors influence this? What is the context of that given statement? What do we expect from that statement? This way, we can find out why algorithms systematically discriminate and how to prevent it.

Another factor frequently mentioned is making the AI industry more diverse (AI Now Report, 2019). It makes no sense to collect more diversified data and improve contextualisation when the people doing it do not have diverse backgrounds. Diversity is not something that can only be measured by numbers. Here, experiences also count." A truly inclusive solution for algorithmic systems requires good data practices embedded in a design and application environment with a constructive point of view. Better algorithms are possible by incorporating minorities - who are often ignored or affected by poorly developed algorithms - into the design process.

Stefan Buijsman

Researcher in philosophy, Institute for Futures Studies, Stockholm

Algorithms are not automatically discriminatory and can also promote more equal decisions. For example, an algorithm that generates maps based on satellite imagery could help combat inequality in countries where otherwise money and time would first have to be spent finding passable routes (especially after natural disasters).

It is also perfectly possible for algorithms that process personal data to do so in a fair and desirable way. The big challenge, however, is to achieve that with self-learning algorithms, which base their decisions on the patterns in large data sets. In those datasets, there is often some degree of inequality (e.g. that proportionally more men are hired for programmer jobs) and it is incredibly difficult to ensure that algorithms do not pick up and potentially magnify those inequalities.

> *"If there are no inequalities in the data, the algorithm does not simply add them."*

So there is no principled reason why algorithms should be discriminatory. If there are no inequalities in the data, the algorithm does not simply add them. In addition algorithms can make problems more measurable by allowing us to calculate in what way and how much an algorithm acts unequally.

Photo: Merlijn Doomernik

At the same time, there is a big 'but': in practice, it is far from always possible to prevent discriminatory behaviour by algorithms. Algorithms that now automatically write texts often associate Muslims with violence because of the internet texts the algorithm has learned from. This cannot be solved with a tweak in the code, despite it being a known problem and one that is being worked hard on. Besides, you can only do something about it if you actively pay attention to it. Advanced algorithms are complex enough that it is not visible in what ways they discriminate, unless you run your own tests. So it can be done, but it is a tricky problem at the same time.

In Europe, the most problematic self-learning algorithms are currently not allowed to be used for decisions on personal data. Although it is still possible to use other discriminatory algorithms, which after all, are no more than a set of instructions for making a decision.

> ***"Algorithms that now automatically write texts often associate Muslims with violence because of the internet texts the algorithm has learned from. That can't just be solved with a tweak in the code."***

The main thing we can do is to keep seeing algorithms as instructions approved by humans. There is nothing mysterious about an algorithm, and so if the result is discrimination, then a different (and better) algorithm should be used. To make that possible with all algorithms, it is important to understand as well as possible what algorithms base their decisions on, what standards those decisions must meet and which data is used to train self-learning algorithms (e.g., not only photos of white men). It is quite a challenge, but one that is very important for good use of algorithms.

Tessa Cramwinckel

Researcher fair and explainable AI

If an algorithm calculates what the ratio of men to women is within a company and the company then does something with it, technically an algorithm has been used for the common good. In this case, the algorithm is probably not much more than a function of Excel.
So the first thing I would delineate is the definition of the word algorithm.

I see it being used a lot, especially when one does not know exactly what is happening. Otherwise, you would call a spade a spade. In this context, I think you are more likely to mean more complex models within the social domain, such as neural networks to predict crime.

When a model is undesirably unfair, it usually has to do with the data being unfair, which in turn usually has to do with 'unfair' society. I think it is important to realise that first of all.

> ***"When a model is undesirably dishonest, it usually has to do with the data being dishonest, which in turn usually has to do with the 'unfair' society."***

An algorithm, in its broadest sense, can be used to straighten this out. Although sometimes I am a bit sceptical, because then you are practising symptom control. That is not necessarily a bad thing, but it is if too little attention is then paid to the underlying problem, which is an unfair society.

There are all kinds of methods for this, which you have no doubt already looked up. And again, there are various metrics that measure how fair/inclusive such an algorithm is. My master's thesis argues that those metrics, or rather the definition of what is fair, should not be left to the programmer. Instead, the public notion of fairness should be incorporated into the model. For this, I have developed a method.

A second piece of the puzzle in this issue is making an algorithm as transparent as possible. Perhaps an algorithm makes a decision that is adopted by policy, but the trick is to also be able to answer that 'why question'. Why does the algorithm make this decision? This is fair to the end user, but it is also a very nice control mechanism to see if an algorithm is working in a desirable way. By the way, this is quite a tricky area of research, but really interesting!

Walter Diele

Artificial Intelligence Consultant

Ever made a WhatsApp message by randomly choosing one of the three suggested words each time? That may make for a grammatically correct sentence, but is obviously nonsense in terms of content, and certainly not what you had intended to say.

Blindly relying on an algorithm for a good suggestion is unwise. But because we call something artificial intelligence (and 'algorithm' is a difficult word), the outcome of a model is often used without asking many questions. Examples where this goes wrong are numerous, but many of those you see in the media are obvious and almost funny.

> ***"But because we call something artificial intelligence, the outcome of a model is often used without asking many questions."***

More problematic is when it is not immediately so obvious, but it does discriminate against parts of the population. Humans discriminate; our brains can function by using all kinds of cognitive biases. The best of us are very aware of this and try to minimise the effect. But either way, our collective behaviour produces data where this discrimination is embedded. Models feeding this data learn to discriminate the models in the same way.

Is there nothing that can be done about that? Well somewhat; there are techniques to partially correct data before training the model on it. Fortunately, there is increasing attention to this, especially among data scientists. But even top organisations have now painfully experienced that they are responsible for models used. Hopefully, this will result in more management involvement and model governance within organisations, so that partial responsibilities for good data use are placed at the right levels in the organisation.

But can you turn it around? Can you use algorithms to detect human discrimination? Frankly, I don't think you need very complicated models for that. A simple cross tabulation is often enough to show wrongdoing.

> *"Top organisations have also now painfully experienced that they are responsible for models used."*

As an educated white Dutch man in his 50s, I have shockingly little experience of undue treatment. I have never been stopped by the police, my suitcases have never been opened at an airport, my children are rated smart at school (which is a self-fulfilling prophecy), never had a problem finding work, etc. All justified, of course, but not a coincidence.

We also all know the examples of the effect of Western versus non-Western names on CVs for internships or jobs, the frequent apprehension of dark-skinned professional footballers or rappers in expensive cars, difference in school advice by cultural background, likelihood of rental housing for immigrants versus locals ... You name it.

Unfortunately, difficult models are not needed to demonstrate a small difference in treatment; the difference is so great that counting is enough.

> *"Everyone knows it is nonsense that an entire board of directors of an organisation happens to be white, old and male and only in HR does the best candidate happen to be a woman."*

Again, the solution lies not (mainly) with the data scientist, but with governance. If directors of organisations were held accountable for discrimination if the data (censuses) showed that the differences between groups were really not a coincidence, perhaps they would manage it better and embed guarantees. Now, discrimination is considered on a person-by-person basis, which is often difficult to prove. But everyone knows it is nonsense that an entire board of directors of an organisation happens to be white, old and male (only in HR, the best candidate happened to be a woman). And the same goes for all other forms of discrimination.

If the differences are so obvious, you should be able to hold an organisation accountable for apparent discrimination. I fear that in this matter, the stick works better than the carrot, because bias is ingrained in our brains and so it does not automatically get it right. I look forward to the time when we need AI models to see if there is discrimination.

Dr. Steven Dorrestijn

Lecturer in Ethics & Technology, Saxxion, University of Applied Sciences

I think an important lesson of the philosophy of technology is that technology is ambivalent. That is, technology is ethically charged, i.e. not neutral, and the value of technology is not unambiguous, but there are always positive as well as negative sides to it. So yes, AI can help fight discrimination and inequality, for example by smartly searching through data in news, social media and policy documents for examples of discrimination. But yes, a typical problem with AI that is now being highlighted everywhere is that it creates bubbles and thus can also reinforce discrimination and inequality.

This formulation is about whether the automation of decisions could be made better and more inclusive. It can surely be done. Partiality, discrimination in the drafted algorithms or created or augmented in self-learning algorithms can be corrected upon discovery. This is a way of dealing with the ambivalence of technology, of AI. Development should be made reflexive: evaluation and adjustment should be built in. This has to do with inclusivity in terms of who is affected by the positive and negative effects of AI.

> *"Technology is ethically charged and its value is not unequivocal. There are always positives and negatives."*

Second, making algorithms more transparent aligns with the principle of inclusivity: everyone can then watch how the calculations and decisions are made. Third, inclusivity is promoted by organising input from more people in the design and development of algorithms (co-creation, co-design).

In my opinion, getting AI right also very much requires putting AI in its place. Awareness of its ambivalent impact should sink in well with everyone. This will hopefully prevent unwarranted confidence in the power of AI.
It seems right to have as a standard view that AI can help us work, think and act, but that it is never desirable to outsource decisions entirely to automated systems.

> *"Getting AI right requires very much putting AI in its place."*

An important question to ask is what we do with the answers and suggestions that automated systems give us. Do we follow them blindly or use them as input for our own decisions? So, very specifically:

- How much do we let data dictate to us?
- How do we keep 'the human in the loop'?

The latter essential questions are appropriate for practice-based research at a college of higher education. By way of illustration, I am always reminded of a work of art placed in a nearby city years ago. The art installation changes colour based on a poll of the city's mood, I understand after checking out posts on social media.

You can't normally see the mood of a city like this. So now we do. This is really a possibility of our time. Also consider health watches. Now imagine that such a watch also gauges your mood and that mood is represented by a colour-changing light on your head. One day it turns red: angry. But you don't feel angry at all. So what do you do? Throw your health watch away because it is not working properly? Or do you think, 'I didn't feel angry, but the data says deep down I am angry so yes, maybe I am angry after all. I believe I am already starting to feel something.'

> *"I don't feel angry, but the data says I do, so maybe I am angry after all."*

I believe it is problematic that what the data says has too much prestige. It is not easy to go against it. A challenge for our times.

Rob Elsinga

National Technology Officer, Microsoft Netherlands

The question of whether algorithms can help combat discrimination and inequality is tricky, but I think so. When designing, building and implementing algorithms, discrimination or exclusion is likely to go unnoticed. Technology and AI can already help understand machine-learning models, assess and improve fairness in AI and manage the life cycle of AI (design, development and application). Microsoft makes such tools available to AI developers in our Azure Machine Learning Cloud services.

> *"Used wisely, human rights language can enrich discussions on responsible AI."*

The great promises of AI and algorithms often focus on the public good, such as improvements in healthcare, agriculture and climate. With these promises of AI, as far as I am concerned, humans should be at the centre. I believe in AI adapting to us and enhancing our ingenuity. Meaningful innovations that help us make a positive impact on the things that matter to each of us.

Fortunately, we increasingly agree on the ethical framework and principles that responsible AI should adhere to. Fighting discrimination and inequality starts with the organisation's values. These values are sometimes implicitly or explicitly expressed and lead to principles that are the basis for every major decision in the organisation.

Another important aspect here is diversity. When designing, building and implementing algorithms, it is important to pursue diversity in teams. While we are enthusiastic about the opportunities offered by AI, we recognise that AI needs to be reliable to gain acceptance. Here, as in other discussions about our values, we should argue for an inclusive and global perspective. Used wisely, the language of human rights can enrich discussions on responsible AI.

In the world of cloud services, values such as freedom of expression and personal privacy translate into international human rights law and the Universal Declaration of Human Rights, among others. In discussions about opportunities and challenges we face to ensure that AI technologies (especially cognitive services) remain human-centric, I believe four provisions deserve special attention. These provisions all flow from respecting the inherent dignity of each person:

1. Non-discrimination: 'Everyone is entitled to all the rights and freedoms set forth in this Declaration without distinction of any kind, such as race, colour, sex, language, religion, political or other affiliation, national or social origin, property, birth or other status.' (Article 2);
2. Protection against arbitrary government interference (Article 12);
3. Freedom of expression: the right to give and receive information and ideas of all kinds, regardless of frontiers (Article 19);
4. Freedom of association: The right to peaceful assembly and association (Article 20)

The ability to make informed decisions in setting out the will of the people in establishing rules for society and establishing the authority of governments (Article 21), i.e. institutions such as independent journalism and academic freedom that make the various rights in Articles 18-21 meaningful.

To apply AI responsibly, the challenge is to transform fundamental rights, principles, organisational values and principles into standards and to practice. It is about what decisions AI should make, not what decisions it can make. This requires dialogue between different disciplines and an approach from science to arrive at systems and tools for responsible AI.

> ***"It is about what decisions AI should make, not what decisions it can make."***

Photo: Rene Opstals Photography

Dr. Katleen Gabriels

Moral philosopher specialising in computer ethics, Associate Professor (Department of Philosophy), Maastricht University

When it comes to AI, we should always ask ourselves what we want to use AI for. We need to think very carefully about what we do and do not use AI for. When is deployment desirable and when problematic?


Photo: Hugo Thomassen Photography

> *"The "toeslagenaffaire" [benefits affair] made the pitfalls very clear. Here the government crossed not only ethical but also legal boundaries."*

AI already helps in determining a diagnosis in a medical context (in accordance with medical judgement), such as when diagnosing breast or skin cancer. If AI systematically outperforms doctors, then in the future it might even be considered unethical not to use AI for this.

In medical assessment, algorithms are trained with photographs, for example of melanomas (skin cancer). These are complex, but nowhere near as complex as social data. A well-balanced dataset is essential and, moreover, no (social) biases should creep in.

The benefits affair made the pitfalls very clear: the algorithms selected on dual nationalities and 'exotic' names, among others. Here the government crossed not only ethical but also legal boundaries. We need independent regulators to use AI properly and then monitor that use.

It is important to analyse algorithms' decision and judgement making at different levels. Technically, it is possible and feasible. But to what extent is it also ethically desirable? Policymakers like to focus on the economic level: often it is enough if it saves time and costs. This from a focus on efficiency and optimisation. But again, they may ignore ethical desirability.

Of course, the legal aspect is also essential; in the benefits affair, the AVG was violated. And then there is the philosophical question. For example, what humanity and world view is behind this drive to optimise? And does everything need to be optimised? What new problems might arise in this way?

> *"To boost profits, Volkswagen committed fraud on a massive scale."*

All those smart, connected devices create opportunities, but also pitfalls. 'Dieselgate' made it clear how you can programme smart technologies to be used for unethical purposes. Volkswagen's tampering software determined independently when it was a test. This allowed people to manipulate emission values, making the car appear more fuel-efficient than it was. To boost profits, Volkswagen committed fraud on a massive scale. With technologies getting smarter, we need to be even more wary that they are not being tampered with.

Pieter van Geel

Director Data, Greenhouse (GroupM / WPP)

Algorithms are ideally suited to detect or recognise discrimination and inequality. I just don't see many applications of this at the moment.
ather the other way round.

> *"The metrics on which an algorithm is driven could be a bit more 'social' instead of business."*

Because the data on which an algorithm is trained is often biased, we often see discrimination/inequality created by an algorithm. In addition, the metrics on which an algorithm is driven could also be a bit more 'social' (i.e. indeed fighting discrimination) rather than business (more profit/more clicks/longer exposure time).

So as far as I am concerned, this is only possible if the data is unbiased and the control metrics serve a social purpose (e.g. inclusive action) and not corporate profit. Unbiased data is obviously not always available, so that needs to be corrected. This should then be monitored and controlled. Wondering which government or agency will take on this task.

Oumaima Hajri

MSt AI Ethics & Society, MSc Data Science & Society

Discrimination and profiling lurk in the use of algorithms and especially if they are self-learning algorithms. To date, there are still no clear frameworks on what national government agencies (or other parties) do to prevent discrimination and profiling in decisions.

Added to this, there is no control over the use of algorithms (this should be regulated) and citizens do not have sufficient insight into their use. Surely these are minimum requirements that should be set before talking about deploying algorithms for the common good at all.

It is important to first name two concepts that answer the core of this question. First, it is important that algorithms are explainable. This means that the technical structure and operation of an algorithm are not only understood by a programmer himself, but programmers can also explain them to people. When this is not the case, it often results in algorithms being seen as 'black boxes'.

Secondly, it is also incredibly important for an algorithm to be interpretable. That is, to be able to have an explanation, but also to understand how certain input variables contribute to an algorithm output that needs to be explained.

> ***"Just as we do not accept arbitrary decisions from people or entities we do not understand, We should not accept this either from algorithms."***

When we talk about decisions being made in an inclusive way, we often think about the people behind the buttons: what interests do they have and do they reflect our society? However, it is important in this discussion to first take a step back and focus on the above concepts. For how much good are 'inclusive decisions' because the right people are at the controls, if we cannot ultimately understand the reasoning behind the decisions?

Having an inclusive team behind the controls only tackles one part of the problem. Indeed, the second part should be that algorithms are both explainable and interpretable. For just as we do not accept arbitrary decisions from people or entities we do not understand, we should not accept the same from algorithms.

In recent years, Explainable AI (XAI) has become an important field in which researchers argue for explainability and interpretability, so that it is always explainable and understandable how algorithms arrive at a particular outcome. However, it still proves difficult for algorithms to provide context as to why they arrived at a particular decision. And this, especially in more serious examples like the benefits affair, can absolutely matter.

Until algorithms do not meet the above requirements, they should not be used for major decisions that could significantly impact citizens. It's as simple as that.

There are still no clear frameworks on what national government agencies do to prevent discrimination and profiling in decisions.

Dr. Marcel Heerink

Researcher, trainer and author in ethics, social robots, AI

Suppose, as an executive, you have an application at your disposal that determines who should get a pay raise and who is better off being fired.

You want to base that on performance and potential. Good people you don't just want to reward, you want to retain.

> ***"With pay rises, workers with few restrictions roll out, with redundancies people roll out who do have limitations."***

Therefore, the application is programmed to have figures on everyone's performance and indications of potential growth (we assume for a moment that both can be measured properly). She does not get information about anyone's disabilities, gender, ethnicity or age. And suppose, as far as pay rises are concerned, workers who have few limitations roll out, but when it comes to redundancy, people who do have limitations roll out.

Then you could do the following:

- Track the outcomes and give people lay-offs or pay increases based on them;
- Ignore the outcomes and just follow your own rules;
- Have the application modified so that constraints are included and lead to a higher score;
- See if you can get a better application that figures out (e.g. with simulations) where or with what approach those with low scores could score higher.

The latter seems to me the best. But until we have good applications for that, we will have to do it manually.

Ir. Reinoud Kaasschieter

Eur.Ing. Artificial Intelligence and ECM Consultant,
Insights & Data Capgemini Netherlands

The lack of inclusiveness is a social problem, algorithms can only be a tool to help eliminate it. The biggest challenge is in the data we feed the models with. It does not matter whether these models are intelligent or not. The data we collect about people, for example their digital 'footprint', can be used to create discriminatory systems.

It is not that models will automatically discriminate: incorrectly designed models will do so. And designing is a human activity. That not all systems are deliberately designed to discriminate is of course true. But it is the designer's responsibility to think about all the possible consequences of his system. Creating systems and models that do not discriminate, let alone algorithms that help combat discrimination, is already a tall order.

It can be countered that people also discriminate, knowingly or unknowingly. And that algorithms might not be bothered by that and could decide more objectively.

> *"It is already quite a task to create systems and models that do not discriminate, let alone algorithms that help eliminate discrimination."*

Photo: Marnix Klooster

But time and again, it proves difficult to build this kind of system. The most common cause is that these systems learn based on historical (business) data. Inequality may be embedded in this data. It proves very difficult to remove bias from this data.

Recruitment and application systems strive to avoid the implicit biases of human resource officers. These systems, for example, promise not to discriminate by gender. We could then extract gender from the data used to feed the choice algorithm. But that does not appear to be enough.

Have you been a member of a women's team or on maternity leave? The algorithm flawlessly knows how to characterise you as a woman. And even if we manage to become gender-neutral, for example, these systems are going to discriminate on other criteria. For example, facial recognition software can be used in job applications to see if an applicant gives sincere answers. But then again, this software does not work on people with autism or facial paralysis. People can become aware of their discriminatory behaviour and adjust their own behaviour accordingly. Algorithms do not do that from within themselves.

> *"Have you been a member of a women's team? The algorithm knows how to flawlessly characterise you as a woman."*

Algorithms are used to make processes more efficient. Or to make much more production. Without algorithms, we can no longer process large amounts of data. The problem arises when we let algorithms make decisions.

Using algorithms becomes truly efficient only when we remove humans from the process. People are slow and expensive. The idea of having every computer decision checked by a human therefore has its limitations. The process becomes too slow, too expensive and basically impracticable. Getting people to check the outcomes of computer decisions for inequality and discrimination is, in my view, a humanly impossible task. From ergonomics and knowledge management, for example. Or at least it is economically problematic. Compare it to manually tracking down wrong content on social media.

> ***"As long as we are unable to eliminate inequality and discrimination in normal social interaction, algorithms will always remain stopgap measures."***

Maybe we should stop trying to solve all kinds of social problems with technology. And then seeing the ultimate solution in that. Until we are able to get rid of inequality and discrimination in normal social interaction, algorithms will always remain stopgap measures.

All we can do is try to act and think inclusively ourselves. And convey these values not only in our daily behaviour, but also in the systems we create. All kinds of techniques and forms of organisation have been developed for that. In doing so, the values of 'equality' and 'non-discrimination' must come out on top.

And should not use systems that do not honour these values. We should have much more empathy for the victims of our algorithms' incorrect decisions. Take the benefits affair as an example.

I note that 'faulty' AI, as in unethical, discriminatory systems, is getting attention in academic circles. Unethical AI is also created and researched at universities. But the discussion there is open and sharp. The main issue here is whether AI contributes to the common good.

Within the business world, of course, AI is also being worked on and experimented with. But this is not published and faulty and discriminatory AI is actually still 'accidentally' discovered here. Often by outsiders. So basically, as outsiders, we don't have a good understanding of what goes on at companies. You have to clearly understand that AI is used in business to make organisations more profitable. This is logical. So where universities focus on non-discriminatory, inclusive AI, business will focus mainly on profitable AI. These two assumptions need not be mutually exclusive, of course, but value conflicts can arise.

Jo-An Kamp

Lecturer researcher at the Moral Design Strategy research group, Fontys University of Applied Sciences

We must first ask ourselves what algorithms actually are. The simplest definition reads 'a finite sequence of instructions (executed by computers) that leads from a given initial state to an intended goal'. The fact that this set of calculations is performed by computers suggests that computers are also the actors making the decisions. But that is not always the case, nor is it the whole story.

> *"Alf we feed bad data into the system, bad data will also come out."*

Firstly, algorithms are usually written by humans, and secondly, humans decide which data is entered into the system and will therefore emerge from the calculations. If we deploy algorithms for the common good, for example for a government agency, then it is thus important to think carefully about both the calculation rules and the data we enter into the system. After all, if we feed in bad data, we will end up with bad data also coming out.

Algorithms are not necessarily fairer or better than humans. However, they can, through the computational rules behind them, reveal patterns that we might overlook in specific, unconnected cases. For example, it is a well-known phenomenon that white men are more likely to be hired in a job interview with a panel of white men than men of other backgrounds or than women. After all, we are quick to choose people who look like us. Or have a subconscious preference for candidates who resemble people who have also performed well within a particular job in the past. As a result, we more often overlook potential talent that does not look like us.

Algorithms can make us aware that these unconscious biases are there, allowing us to adjust rules and procedures (both in the computer and in real life!) accordingly. In this way, then, we could actively fight discrimination and inequality. But then we have to do it the right way!

I think the first thing we need to ask is how inclusive the team making the algorithms is. The more single-minded this team is, the harder it is to imagine the perhaps much broader group for which the algorithm is intended.

So make your team more inclusive and/or test your product in that wider audience, with people of different age groups, skin colours, political affiliations and so on. And ask yourself the following questions: does everyone have access to your product (e.g. does everyone own a computer or smartphone and can everyone use them equally well)? Are you (consciously or unconsciously) excluding people? Does your technology have a built-in bias? Can you make your algorithms' decisions transparent? And is the outcome after the 780th step still explainable?

In the Technology Impact Cycle Tool, there are a lot more questions like this that can help you make your technology more compatible with the human scale and the world you would like to live in yourself. Feel free to try it at www.tict.io. It is free and for everyone.

> *"The more one-sided the team making algorithms, the harder it is to empathise with the much broader group for which the algorithm is intended."*

Yori Kamphuis

AI-expert en speaker

There are certain values embedded in the question of whether algorithms can be used for the public good. An AI system pursues a goal, something that can be maximised or optimised. How is that common good defined and measured?

> *"It is important that predictions do not become self-fulfilling prophecies."*

AI is not automatically neutral because it pursues a particular goal. In doing so, it also depends on data. If that data has a bias, i.e. is coloured or contains prejudices, the AI is likely to maintain the same bias. It is important that predictions do not become self-fulfilling prophecies.

Suppose only one group is looked at if a certain traffic violation takes place. Then it is a wrong conclusion to say that only people from this group make these kinds of offences. Because there was no consideration of whether someone from another group also committed that kind of offence. But if AI is trained and, because of the erroneous inference, checks only that one group, it can encourage discrimination or inequality.

> ***When the AI system makes a different recommendation than you expected, things get interesting.***

Important to remember is that distinctions may be made based on certain characteristics. As long as these characteristics matter, this is not discrimination. An example from the Human Rights Board states that a person who applies for a job as a driver without a driving licence may be refused. Thus, specifically, a "coloured/black actor may be sought to play the role of Martin Luther King". See https://mensenrechten.nl/nl/gelijkebehandelingswetgeving and https://mensenrechten.nl/nl/discriminatie-uitgelegd.

You could possibly train a dataset in which you leave out one (potentially discriminating) element each time, and train that model again, in order to see whether the predictive value of the model improves. If it improves, then this element had no added value and perhaps you should no longer collect it. But even this is not going to eliminate possible discrimination.

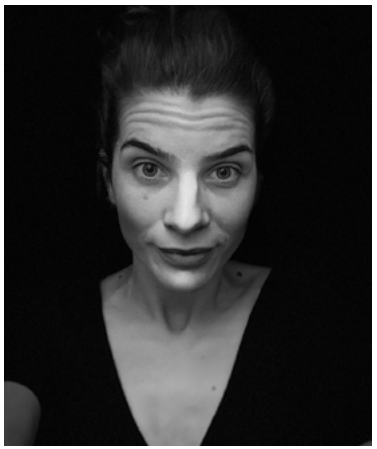
By using AI as support for human decision-makers, I think we can enable human-centric AI. This should include monitoring - does the AI system do what it is supposed to do? When the AI system makes a different recommendation than you expected, things get interesting.

Explainable AI should provide insight into why it makes a particular recommendation. Linked to this is a certain degree of (internal) transparency. In other words, can you understand why the system comes up with a particular recommendation? Suppose you still believe that the AI system should have come to a different decision, how come? In other words, where does the AI system fall short? This allows the system to be trained further, to keep an eye on the human side.

Traditionally, in the medical world, most treatments and drugs were tested on men, not women. However, on this basis, women were also recommended. Using any non-representative group as a model group thus reduces inclusiveness. So I would particularly focus on improving (collecting) relevant data on the basis of which an AI is trained.

Julia Keseru

Technology and justice activist, ED of the Engine Room



> *"Much of what is sold as AI today is really just snake oil."*

The more research I see, the more convinced I become that algorithms, while they can be useful in some cases to increase efficiency, are really terrible at supporting decision-making processes that are fair, honest and nuanced.

The main problem is not that automated decision-making is not yet sophisticated or advanced enough. While it is important to note that much of what is being sold as AI today is actually simply snake oil. The main problem is that the data that algorithms use is often already highly problematic.

Currently, there is a trend in every sector towards evidence-based decision-making where quantifiable markers take precedence over qualitative and contextual analysis, a laudable trend. However, relying on quantitative data without context can cause more problems than it solves - systematic racism, for example, is constantly reproduced by the over-representation of minorities in data used for predictive policing or in child welfare systems.

Yet it is not impossible to use algorithmic decision-making for good causes. We have seen that automated decision-making helps mitigate the effects of climate change in agriculture, uncover corruption in public procurement, improve patient outcomes, monitor human rights violations and predict emergencies.

> *"We need stricter rules and incentives for the technology industry to meet human rights standards."*

Because this is a systemic issue, a lot of things need to happen to change the current course of how AI is designed, implemented, replicated and so on to ensure that AI makes more inclusive decisions for us. We need stricter rules and incentives for the technology industry to meet human rights standards.

We need more diverse teams behind the design of technical tools that use algorithms. We need much more transparency and accountability around automated decision-making processes in both the public and private sectors (see this primer from Data and Society:

https://datasociety.net/wp-content/uploads/2018/04/Data_Society_Algorithmic_Accountability_Primer_FINAL-4.pdf). We need more data equity trainings like the one provided by We All Count, and many more. To get a good idea of what else is needed, I recommend this report from the Mozilla Foundation: https://assets.mofoprod.net/network/documents/Mozilla-Trustworthy_AI.pdf

Gary Marcus

Founder and CEO, Robust.AI Professor Emeritus, New York University.
New book: REBOOTING AI

Future algorithms, yet to be invented, may one day be able to help with equity and inequality, but current algorithms are far too primitive for that. They merely recapitulate the past, and have no understanding of the future we are trying to achieve. They have no values and norms and no understanding of the complex dynamics of humanity. As long as they are little more than blind number crunchers, they are not worthy of our trust.

> *"Current algorithms merely recapitulate the past, and have no understanding of the future we are trying to achieve."*

Photo: Athena Vouloumanos

Iris Muis

MA Project coordinator, DEDA-lead Utrecht Data School, Utrecht University

Algorithms, including AI, offer huge opportunities to society. We have the ability to design our processes more effectively, efficiently and sometimes even in a completely different way. At the same time, these technologies do not solve everything and, like humans, algorithms can contain biases and place certain values above others.

AI can transport, enhance or even degrade values. Our job is and remains to ensure that AI conveys the values we want to convey. I advocate responsible development and deployment of algorithms, consciously reflecting on the implicit assumptions embedded in the code and the way the algorithm is implemented. This gives us a chance to let algorithms contribute to a society we would like to live in.

Inclusiveness is an important value that you should include in the design of an algorithm. Already during the development phase of an algorithm, certain choices are made, by the data scientist (or perhaps his manager), which are anchored in the design and have consequences when the algorithm is actually deployed. When you think about values on the front end, for example inclusivity, it gives you a chance to embed them in the design of the algorithm.

> *"Responsible development and deployment of algorithms gives us the opportunity to have them contribute to a society we would like to live in."*

Thijs Pepping

Humanist & Trend analyst ViNT, Sogeti

> *"Algorithms quantify our unconscious biases and rub them in our faces."*

First, we need to realise that bias in algorithms in many cases mirrors bias in humans. Take, for example, the AI tool Amazon developed to help in the recruitment process. Very convenient according to some: you give the programme a hundred CVs and the best five are selected. Amazon had to pull the plug on the project because the tool had a preference for men's CVs.

That, of course, is objectionable. But the situation is just as bad if afterwards everything stays the same and there is no reflection on where those biases in the algorithm came from. Indeed, the tool was trained on ten years of recruitment data where people made decisions about who was hired and where the majority of applicants were male. Why is that? What can we change about it?

In this sense, I see biases in algorithms as a blessing: they show the human deficit. Algorithms quantify our unconscious biases and rub them in our faces. We obviously have to fight bias in algorithms, certainly because with the scaling up of technology, inequality in an algorithm quickly has major consequences for large groups of people. But above all, we must also address the bias in ourselves. The mirror held up to us by algorithms can be a good starting point for this.

> *"Working on inclusive algorithms is not an objective science, but value- and politically charged."*

It really is now time for a Minister of Innovation & Technology to put this issue properly on the agenda and regulate it. Although already slightly better than in 2017, in 2021 I still miss thought-out visions and positions in party programmes when it comes to AI and technology. The advanced and all-pervasive technology around us questions our deepest norms and values. For example, how about a love affair between a human and a chatbot? Or of virtually bringing deceased loved ones to life to have another conversation with them? How far are algorithms allowed to intervene in our society? And who creates and evaluates those algorithms?

Ultimately, in answering those questions, you end up with your own world view, political affiliation and philosophy of life. Put bluntly, if you believe in a soul, for example, you might disapprove of a relationship with a chatbot because the chatbot has no soul. If you are an atheist then you judge differently again. So working on inclusive algorithms is not an objective science. It is value- and politically charged. Working forms for Ethical AI, Explainable AI, Transparent AI, Inclusive AI et cetera are important and useful, but above all, let's take the huge impact of technology at the country level more seriously.

Gerard Schouten

Lecturer AI & Big Data Fontys University of Applied Sciences

I am convinced that - in the long run - we can certainly use smart algorithms for the common good. However, there are things that we have not yet properly 'mastered' and have yet to learn. But they are being worked on very hard.

> *"Working on inclusive algorithms is not an objective. In an AI model that does that with social aspects, literally 'the human in the loop' must be applied."*

1. How to deal with bias in data we train these algorithms with?
   People are even looking at, for example, measuring demographic bias (age, gender) automatically in datasets and correcting it with AI (adversarial neural networks). In 2020, a research paper titled 'Risk of Training Diagnostic Algorithms on Data with Demographic bias' was published.
   See: https://www.researchgate.net/publication/344657158_Risk_of_Training_Diagnostic_Algorithms_on_Data_with_Demographic_Bias.

2. In my (strong) opinion, with AI that says or does anything with social aspects (access to loans or benefits, access to education, et cetera), we should always ensure that an AI model with literally 'the human in the loop' is built/ applied. This is what the benefits affair has given and taught us.

3. The privacy issue has not been properly resolved either. However, with techniques like edge computing (e.g. on the IoT device itself), you can already make great strides towards not putting personal data or related data in the Cloud.

Fredo De Smet

Curator en digital humanist

We need to be aware that AI comes with a loss of control. In my book 'Artificial Stupidity', I describe in detail how we live in a crisis of control. This is nothing new. We have seen that before in history.

> *"We have more power to create information than to control it."*

In the nineteenth century, we had more power to produce materials and energy than to control them. We plunged into an industrial revolution that took us into the modern world. Muscle power was replaced by industrial power. Our arms and legs were lengthened, as it were, by machines made of iron and steam. We overcame nature. Man abandoned thousands of years of agrarian life and moved into the city. Paired with this came situations of injustice, poor work situations in factories, children in labour, fatal accidents. Drama. The political and social fabric was shaken.

Today, we live in a similar crisis of control. In this case, however, it is not about matter but about information. We have more power to create information than to control it. There is just too much data to be read by a human. Enter: AI. An automated process to manage this abundance of data. The irony of course is that AI is inherently uncontrollable. It is a self-learning, self-managing algorithm. We will, in short, always be too late.

Of course, that doesn't answer your question. But I mention it because we have to be realistic. A loss of control is inherent in AI.

You also notice this in the organisations and bodies advocating ethical AI. Numerous reports argue for an ethical way of dealing with AI. The Future of Life introduced a set of rules (Asilomar principles) in 2017. The EU has been busy in recent years. The wishes expressed always remain vague. It essentially comes down to transparency, explainability and accountability.

You already sense that it is difficult to make an organisation accountable if the technology is essentially not transparent and explainable. In short, we are not home yet. Still, these three aspects give an idea of where things should go.

For example, we could develop an Ethical AI label, as exists for sustainable food. An organisation gets this label only if it makes at least an attempt to explain what data it works with and what it does with that data (explainability).

In addition, it would be good if research institutions could request access to the AI. A major frustration at Stanford is that they train Silicon Valley engineers, but do not get an insight into the algorithms these engineers work with a few miles away. Again, it is not a quick fix, but I think the smartest minds should also be given access to the most complex algorithms, not just the self-learning ones (transparency).

I also believe that the race for AI is an opportunity to address other forms of inequality. Smart products will soon become a differentiator for organisations. The craziest products will be cognised. Remember the razors that once became electric (thanks to Philips). They too will soon be 'powered by AI'.

> *"Razors that once became electric will soon be 'powered by AI'."*

If these companies want the Ethical AI label, they would also have to meet some other conditions. It is an opportunity to be more inclusive, for example by imposing diversity quotas in the tech teams (four out of five AI experts are men). Putting men at the centre also means breaking the hegemony of men.

To make it truly humanistic, we could attach a value system to the label. Product development should be purpose-driven. This allows us to focus more on design processes that optimise not only for the end consumer, but also for the community.

Jim Stolze

Tech entrepreneur, founder Aigency

Strictly speaking, within computer science, an algorithm is nothing but a step-by-step instruction to perform a task. And so there are plenty of IT systems within government that seek to support decisions by civil servants in a fair and equitable way.

You probably mean data-driven AI models. My company Aigency recently designed and validated such models for the Ministry of Health, Welfare and Sport based on the principles behind FACT (Fairness Accuracy Confidentiality & Transparency). This allowed us to show statistically whether certain population groups were overrepresented in the training data and/or whether they would benefit or suffer in the real world. In fact, this should become mandatory for every data project. Such checks must be done on the data before you dare utter the word AI at all.

> *"We need empowered, critical consumers to stand up for their own rights in the digital domain."*

In doing so, I distinguish between top-down and bottom-up. Top-down is anything that is already enshrined in law (including GDPR) or enforced by regulators such as the Personal Data Authority. But note: this is just the tip of the iceberg. Most data-driven models are not audited or take the law literally, even though there are some ethical issues with that.

That is why I am more interested in the bottom-up approach. We need empowered citizens, critical consumers who stand up for their own rights in the digital domain. This is partly why I started the National AI Course back then. We can all point to the government and call for regulation, but that does not absolve us of the duty, above all, to shape diversity, inclusiveness and justice ourselves. If not us, who else?

Janienke Sturm

Lecturer Man and Technology Fontys HRM and Psychology

I am convinced that AI algorithms can be used to reduce discrimination and inequality, see for example: https://www.volkskrant.nl/nieuws-achtergrond/bij-unilever-is-de-helft-van-alle-managers-vrouw-dankzij-kunstmatige-intelligentie~b8535708/

However, there are two caveats:

1. Algorithms seem neutral, but in fact they are not. There will always be conscious or unconscious bias in the set of data on which the algorithms are trained. Prejudices are then replicated in the dataset, and at worst, this dataset actually leads to more discrimination and inequality, based on gender, ethnicity, sexual orientation or postal code. For good examples, see: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2020/09/14/onvoorziene-effecten-van-zelflerende-algoritmen/Onvoorziene+effecten+van+zelflerende+algoritmen.pdf

2. Algorithms can be used positively by people with good intentions. But the same algorithms, by people with bad intentions or conflicting interests, can also be used for purposes that are less positive, less ethical et cetera. See the example in this article: https://www.computable.nl/artikel/nieuws/security/7060485/250449/ noldus-schendt-via-tech-mensenrechten-in-china.html

In it, a Dutch company that makes software for behavioural research (including face and emotion recognition) is accused of breaching human rights in China, as the Chinese government deploys the software for surveillance purposes.

We must also be mindful that the deployment of AI in general may actually lead to more inequality. Quite a large group of people do not have the right equipment to access AI applications, or the right skills to use AI applications.

A much larger group lacks sufficient understanding of how algorithms work to appreciate them and the risks involved in using AI. In this way, digital illiteracy ensures that not everyone can benefit equally from the opportunities offered by AI. In doing so, the deployment of AI, and technology in general, is actually leading to increasing opportunity inequality in society, for example in terms of participation and health.
https://www.capgemini.com/research/the-great-digital-divide/

"Digital illiteracy ensures that not everyone can benefit equally from the opportunities offered by AI."

Farid Tabarki

Founder Studio Zeitgeist

Many people use social media for their news gathering; meanwhile, newspaper circulation is steadily declining. Major platforms for knowledge, news and opinions are not automatically independent or transparent: their algorithms and policy choices co-determine our social reality. Companies must act accordingly.

"What makes contemporary societies work increasingly depends on bits rather than atoms."

The big question is how to divide ethical responsibilities in today's cluttered information society. A society in which global companies and individual bloggers play roles that did not exist 20 years ago and in which nation states have less and less of a say.

Luciano Floridi, professor of philosophy and ethics at Oxford University, has an answer to this. According to him, it is a sign of our times that when present-day politicians talk about infrastructure, they often have information and communication technologies (ICTs) in mind. And that's right. From success in business to cyber conflicts, what makes contemporary societies work increasingly depends on bits rather than atoms. Depending on their digital infrastructure, societies can grow and prosper.

And societies' ICTs often represent one of their weakest points in terms of cybersecurity. We all know this. Less obvious and more philosophically interesting, ICTs also seem to have revealed a new kind of equation.

Consider the unprecedented emphasis ICTs place on crucial phenomena such as accountability, intellectual property rights, neutrality, openness, privacy, transparency and trust. These are probably better understood in terms of a platform or infrastructure of social norms, expectations and rules that exist to facilitate or impede the moral or immoral behaviour of the agents involved.

By placing our informational interactions so importantly at the core of our lives, ICTs exposed something that has of course always been there, albeit less visible in the past: the fact that moral behaviour is also a matter of 'ethical infrastructure', or as I call it: infra-ethics. So for the major information and communication platforms, there is work to be done.

> *"WICTs exposed the fact that moral behaviour is also a matter of ethics."*

Prof.dr.ir. Bedir Tekinerdogan

Chair Information Technology, Wageningen University & Research

An algorithm is a finite set of instructions with the aim of solving a problem. Several algorithms are often possible for the same problem (e.g. sorting or searching in a list). It then looks at the algorithm that is the fastest or requires the least memory. Algorithms are also differentiated based on their complexity, which is mostly determined by the amount of time and/or memory space required by an algorithm.

We can solve many problems by developing efficient algorithms for them. However, some problems can be difficult to solve and also no clear algorithms exist for them. The question of whether we can use algorithms to fight inequality is actually not very trivial either.
We should reduce this further and give it proper context. It is good to distinguish here between the real world and the cyber world. The cyber world consists of all the computer systems that should serve decision-making for the real world.

> *"Within the context of real world and cyber world, technological developments and their impact should also be a focus of the politics and the regulatory and legislation."*

Algorithms can now be used to examine the doings and actions of the real world and, where necessary, combat inequality. This can be technically implemented by using 'computer vision' and artificial intelligence.

Smart sensors and associated software systems can be used to observe people's behaviour and, if necessary, take action. Algorithms can also be used to analyse (social) media for racism, discrimination and other violations of current laws. However, the question is where the limit of surveillance lies and to what extent privacy rights are violated.

Technically, a lot is possible, but it remains an issue that requires a broader and more in-depth discussion. What we do see is that technology is developing very quickly and existing politics and legislation are actually lagging behind this. The cyber world is an important part of society. This is the virtual world with all the computer systems affecting decision-making in the real world. Within this new changed context, it is therefore of great importance that technological developments and their influence also become a focus of politics and, with it, regulation and legislation.

These seem to be mostly algorithms related to government or private enterprise software systems. The question then becomes whether or not these algorithms are fair. Within a democratic rule of law, all citizens are equal and should be treated equally in equal cases. This is also enshrined in Article 1 of the Dutch Constitution on equal treatment and prohibition of discrimination. Great importance is normally attached to the enforcement of this law. Rightly so. However, with digitisation, law enforcement has taken on another dimension.

Algorithms are typically implemented in computer programmes used for decision-making within government or private companies. However, the decision-making implemented in these computer programmes tends to be less

> ***"Citizens were treated unequally in equal cases by algorithms implemented by people who should respect the principle of equality."***

visible. There could then, unfortunately, too easily be a violation of the principle of equality and thus the principles of a democratic rule of law.

We have also seen this in the benefits affair, where a government body used ethnic profiling in certain decisions. Citizens were thus treated unequally in equal cases. In the real world, such a course of action would be easily noticed, but now it was done by algorithms, in the virtual world. Algorithms implemented, by the way, by people who should respect the equality principle.

To avoid these and other legal violations, it is important to make sure when implementing algorithms in computer systems that the algorithms comply with agreed rights and laws. At an early stage, before implementation, the existing legal and ethical requirements for an algorithm or computer system should be defined.

In addition to the functional requirements, the implementation of these systems should then be done in accordance with these quality aspects. Perhaps there are already multiple systems implemented that are not fair and inclusive. These should be analysed and, where necessary, updated again to still meet legal and ethical standards. Thus, the principle of equality should also be applied to the cyber world of computer systems.

Eric van Tol

Commissioner at The PMO Company

Ironically, we will have to fight discriminatory algorithms with the same algorithms. It is the data that feeds the algorithm. We make that data and yes we are discriminating. The solution is our algorithm design and our data set choice. Perhaps algorithms can combat smaller inequality gaps. And we can map inequality better or understand the causes more often with self-learning algorithms, but fighting inequality on a large scale is a moral revolution with a multitude of actions that for now are not going to be done by self-learning algorithms.

Make sure algorithms are explainable to everyone. This is often not the case for two reasons.

First, many -often wrongly- consider ICT complex. Easy access to training is the only answer.

Second, many algorithms are a black-box and therefore not explainable.
The answer to that is research. Research other trackable self-learning algorithms and investigate supported explanations/mechanisms for such a black-box algorithm.

Make sure the algorithm consumer has design power. In the service design process of companies and governments, consumers need to have more of a say. Citizens or consumers cannot customise their services and can hardly give feedback or refuse at all. Thus, a consumer's privacy is often not the problem but the consumer's lack of autonomy much more so. The best is when consumers can direct a service.

Otherwise, we need to prescribe inclusivity.

Dr Eva Vanmassenhove

Assistant Professor Department of Cognitive Science & Artificial Intelligence, Tilburg University

There are several tools/algorithms that can help detect bias/discrimination:

- Testing with concept activation (TCAV), a system that can detect biases (race, gender, location);
- AI Fairness 360 from IBM (seventy fairness metrics that can help detect bias, and ten bias mitigation algorithms);
- FairML (an open-source toolkit).

Moreover, several initiatives are trying to use AI for 'social good'. These initiatives do not necessarily focus on inclusivity or bias, but on detecting cyberbullying, online harassment, poverty detection, etc. The ethical aspect of AI technology is also receiving increasing attention. For example, there are increasing guidelines for reliable, transparent and innovative AI, such as those issued by the European Commission.

The first step is probably to recognise and admit that, despite the fact that algorithms are often described as objective, neutral and therefore unbiased, they are in fact often neither inclusive nor objective, and step two is to try to understand the reasons for this.

> ***"The first step is to recognise and admit that algorithms are precisely often neither inclusive nor objective. Next, you should try to understand the causes of this."***

Currently, broadly speaking, a distinction can be made between three specific techniques to ensure that our algorithms are less biased: (1) pre-processing of data, (2) in-processing and (3) post-processing. These are all examples of de-biasing techniques or bias mitigation techniques.

> ***"A diverse team working on applications is likely to be more aware of the different ways in which data can be biased."***

1. **Pre-processing:** Pay more attention to training data ('garbage in, garbage out'). For example, training data for a given task may be incomplete, non-diverse, biased and unrepresentative or ill-defined. So paying more attention to training data quality is definitely important if you want to train more inclusive models. If your algorithm was only tested and trained on 'white men aged 40', it will most-likely also only work well for this specific target group. Fine if that's what you had in mind, but not good if you're trying to develop something that should basically work for everyone. 'Reweighing techniques' are used to avoid bias. However, the pre-processing approach is often difficult, expensive (lots of manual intervention) and can also raise privacy issues, as you need to store/know certain sensitive attributes to ensure that there is some balance.

2. **In-processing:** You can also address the model/algorithm itself and ensure that a classifier, for example, simultaneously focuses on accurate prediction and reducing bias against a particular, potentially sensitive attribute. This technique is called 'adversarial de-biasing'. You then actually get a combination of two models where the first tries to make an accurate prediction, and the second model (the 'adversary') tries to predict that sensitive attribute based on the prediction of the first model. The goal is to make the first model predict as accurately as possible, while keeping the prediction capability of the second model as low as possible. This should ensure that predictions are less biased, since it should not be possible to predict sensitive attributes based on the predictions (thanks to the 'adversary', see the second model).

3. **Post-processing:** Once the prediction is made, you can also try to make corrections that will make the predictions fairer or more inclusive.
   There are also different techniques for this.

   An easy example within the field of machine translation is:
   - EN: 'I am happy' (gender neutral)
   - FR: 'Je suis heureux' (masculine gender)

   In a post-processing step, you could identify these phrases and then also offer a female/neutral alternative:
   - FR 'Je suis heureuse' (feminine gender)

Most of these techniques are task-dependent. Bias mitigation techniques for classification tasks are most likely to look different from bias mitigation techniques specifically for e.g. translations.

Other key elements to combat bias and create more inclusive AI models are:

A. The quality/diversity of the test data used to test our models;

B. Humans in the loop (manual evaluation where necessary to uncover any problems);

C. Diversity in teams developing technology. Since we are also very biased, it is good to have a diverse team working on applications. EA more diverse group is likely to be more aware of the different ways in which data can be biased;

D. Timely updates to the guidelines and legal framework/regulatory and legal systems around AI that can help prevent discrimination.

Rens van der Vorst

Technology philosopher

Every day, LinkedIn's algorithms show me an article by some artificial intelligence that can do something fantastic. An AI that can recognise emotions was launched in China. In America, researchers developed an AI that can convert ordinary human language into programme code. In England an AI writes articles for The Guardian. You would say that if AI is evolving so fast, that presents opportunities to use AI to combat discrimination and inequality.
Or does it? I don't think so.

> *"People discriminate and are prejudiced. But I doubt that an AI would make better decisions or fairer decisions."*

Take, for example, AI that can recognise emotions. Can it really do that? Or does AI have a rather narrow view of what emotions are? Are emotions reduced to certain facial expressions? And isn't that the biggest problem of all? Not that AI is getting smarter, but that AI is limited? But that we conform to it anyway and walk around all day with a stupid grin? Or with an exaggeratedly serious face?

That is why I am reluctant to use AI to fight discrimination and inequality. Of course, people discriminate. Of course, people are prejudiced. But does an AI make better decisions? Fairer decisions? I doubt it. Can you make sure AI input is not biased?

It sounds easier than it is. In fact, there is a grain of truth in a lot of preconceptions. Women are much less likely to commit recidivism, so AI handing out punishments must also take gender into account. It's only fair. Or isn't it? But what if women are now much more likely to drop out in business? Should AI take that into account too? Or does it reinforce other problems in society?

> *"AI wins over grandmasters in chess. Maybe then AI could also figure out how to make us more inclusive."*

Which prejudices should we take into account and which not? An impossible puzzle. And if we could put the puzzle together at all, we discover that AI does find and reinforce other biases somewhere in the 'black box'. Prejudices we did not expect. Maybe that is even worse, because it happens somewhere in the black box. We think it's fairer, but is it?

Cathy O'Neill once said that a successful dinner with her children includes lots of greens. Her children find that a successful dinner includes a lot of Nutella.
If O'Neill is allowed to write the AI code, lots of veggies come out, with the kids litres of Nutella. Algorithms are therefore always opinions, encapsulated in code.

That may be fine if we think our society needs to become more inclusive and fair. But the danger is that we then hide behind algorithms. We no longer take responsibility for our decisions, but refer to the magic of AI. AI hires people based on facial expression, and does so completely honestly. Really? On what basis was it defined which facial expressions are good? And what does AI see as a facial expression? Is it fairer though? Or does it only seem that way?

But maybe I'm taking the wrong approach. Maybe I underestimate AI and should just let AI figure out how to make the world more inclusive. After all, AI wins over grandmasters in chess, AI can win poker, GO & Jeopardy. Maybe then AI could also figure out how to make us more inclusive. Or does AI function especially well when the scope (the game board) and the goals (winning) are clear? And that, of course, is not the case here. Here it is very difficult to determine what is game and what is goal, and then AI is like a genie in the bottle. If you are not careful what you wish for, things will go wrong. Think of King Midas.

Oh yes, and it will best be said that there will always be a 'human in the loop' in all decisions. I don't really believe in that expression. After all, it suggests that computers are in charge and people are also 'in the loop'. 'Computer in the loop' seems a better expression then.

Prof. Toby Walsh

FAA Laureate Fellow & Scientia Professor, AI School of CSE, UNSW Sydney

Yes, we can use algorithms for the common good, to fight discrimination and inequality. Yes, we can use algorithms for the common good, to fight discrimination and inequality. But it won't happen if we don't work hard. The tech companies sold us the big lie that algorithms are unbiased. However, they can be just as biased as humans, even worse.

We have to be very careful lest they perpetuate the prejudices of the past. Especially when they use machine learning, trained on data that inevitably reflects historical biases. Similarly, algorithms can ensure that we make fairer, more evidence-based, decisions that ignore the unconscious biases to which people are prone.

How do we ensure that algorithms take choices for us in an inclusive way? Is it about algorithms making inclusive decisions for us, inclusive and diverse teams in the room can help ask the right questions. It can also involve pushing back and not handing over certain important decisions to machines, but always letting a human take them. Only humans have empathy and can be held accountable. Therefore, we should never leave many decisions in the judiciary and the military to machines, such as sentencing or life-and-death decisions.


Photo: TU Berlin/Press/Christian Kielmann

"Only humans have empathy and can be held accountable."

Mr. Dr. Bart Wernaart

Professor Moral Design Strategy, Fontys Universities of Applied Sciences

We need to ensure that inclusivity at the design table of algorithms is a strong starting point. Two things are key here.

Firstly, societal values need to be better aligned with the values being designed with: often, algorithm design focuses on effectiveness and optimisation, rather than contributing to societal solutions. While those very societal values can be hit so hard when the application of algorithms gets harmful side effects. Remember the benefits affair.

> *"The design of algorithms often focuses on effectiveness and optimisation, not on contributing to societal solutions."*

Photo: Angéline Swinkels

Second, it is important that the individual is seriously heard at that drawing board, pre-eminently those who ultimately bear the consequences of algorithmic applications. This goes beyond 'creating support' whereby that individual, through public participation procedures/ stakeholder consultation, afterwards can sign at the X. It means designing based on the moral expectations of stakeholders.

Szymon Wróbel

Professor of philosophy at the Faculty of Artes Liberales at the University of Warsaw

The network algorithm initially appears to satisfy the requirement of J. Rawls' theory of justice. Algorithms seem to take an "original position", being "neutral" and remaining behind a "veil of ignorance". But is that really the case?

Google is a company based on an algorithm. Google is apparently a search engine. The original algorithm organises the entire World Wide Web according to peer citation models that quantify which papers are most influential and relevant. This computational meritocracy (according to the company's statement) is in the service of a universalist mission to not just organise the world's information, but "make it accessible and useful".

> *"Google's economics are based on transforming the web into a giant advertising platform and advertising into a network of compute points and clicks."*

The point is that Google's revenue comes not from the direct delivery of retail services, such as search, but from monetising the attention that users pay to services in the course of their engagement. Google's economics are based on transforming the web into a giant advertising platform and advertising into a network of compute points and clicks. Don't be fooled, this is not done for the public interest, but for the company's huge revenue.

On the face of it, platforms (such as Google) consolidate heterogeneous actors and events into more ordered alliances, but they are not necessarily in a truly central position in relation to those alliances. The universality of platforms makes them formally open to all users, both human and non-human. Can we then trust platform democracy?

> *"We are not aware that certain political views are already built into the hardware."*

My answer is: no. Platforms are based on a standardisation of their essential components. The formal politics of platforms is characterised by the paradox between a strict and immutable mechanism (autocracy of means) and an emerging heterogeneity of self-directed use (freedom of ends).

However, is the user sovereign in his decisions and does he actually choose his goals autonomously? We are not aware that certain political views are already built into the hardware. For many processor chips, the 'core user' is a sovereign figure that can produce subordinate administrative subjects, who in turn can control the calculating access of other users. This means that 'politics' can be found not only in the legal consensus, but also directly in the machines. Decisions are not only made about infrastructure, they are made by infrastructure.

Hans de Zwart

Philosopher and lecturer/researcher, University of Applied Sciences of Amsterdam

Employees of campaign firm Cambridge Analytica had gained unauthorised access to the profiles of millions of Facebook users. This data was used for Donald Trump's election campaign in 2016.

This scandal has had an important consequence: data has become a political issue. Everyone now suddenly sees the dominant business model of the internet. A model where the breakdown of privacy is central and public values come under pressure. What we here together have long known, the rest of society now understands. Without privacy you do not have fair elections, without privacy you cannot guarantee equality, and without privacy we are not free and autonomous.

More and more human behaviour is modelled and more and more often the gap between model and reality is forgotten. A model works with statistical values such as averages, but no human being is an average. Our humanity is too complex to fit into the pigeon hole of a model.


Photo: Juri Hiensch

> *"Everyone now suddenly sees the dominant business model of the internet where breaking down privacy is central and public values are under pressure ."*

There is a gap between actual students and what the learning analytics system tracks about students. The data in different healthcare systems never gives a complete picture of a patient. And the measurement data on workers also does not say much about how workers do their jobs. As we increasingly make automated decisions based on system reality and not the individual, we are more likely to fall into that gap. Starting with those already in a weaker position in this society.

> ***"No human being is average. Our humanity is too complex to fit into the pigeon hole of a model."***

A change must start with the intention to change and bring models along. We must abandon the assumption that computers (algorithms) are objective, neutral, correct and error-free. A better assumption is that a computer model includes the reflection of a racist view of the world.

Epilogue

For this book, I asked many experts at home and abroad for their views on AI, inclusivity and prejudice. However, the list of experts dealing with this topical issue on a daily basis extends far beyond the experts I approached. That list is growing steadily. Science, government and industry alike are increasingly visibly embracing the necessary discussion on bias and discrimination within algorithms and its potential dangers. A development that makes me very happy.

Also within Fontys University of Applied Sciences ICT (FHICT), I am actively engaged in the fascinating field of AI, algorithms and bias. In the AI group led by Gerard Schouten and quartermaster Danny Bloks, and with my fellow techno-philosophers including Rens van der Vorst, Jo-An Kamp, Huub Prüst and Lennart de Graaf, I am trying to dissect the ethical side of AI. Through the podcast series in which we leave no aspect of AI undiscussed, there is a nice connection between what could be and what is desirable. https://open.spotify.com/show/0hDFsglp8xvDsajgqZ6Jdf

So many experts, so many perspectives and visions on whether, and if so, how we can use algorithms to make non-discriminatory, inclusive, human-centred decisions. We are still far from a one-size-fits-all solution, but we have come a long way. Because if one thing becomes clear from this book, it is that science, government and industry are aware of the need to find that unambiguous solution.Together, there is still a world to be won.

In conclusion, my mother has since recovered from the shock. Her trust in the Dutch government has been shaken, though. But probably less so than the confidence of victims of the benefits affair. Meanwhile, the compensation schemes are taking off, although this too is not entirely smooth. Still, as wry and unfair as the allowances affair is: it did put this discussion on the broad social map in one fell swoop. That, in the long run, might be a win after all? Oh, and my son? He can't wait for FIFA 22 to come out.

About the author


Photo: Ektor Tsolodimos

Erdinç Saçan lives with his wife and three sons in 's Hertogenbosch. He has a passion for a wide range of subjects. Marketing, technological developments, ethics, privacy, data, business models and persuasion principles have his keen attention. But with equal interest he plunges into matters related to politics, social cohesion, volunteering and networking.

Erdinç Saçan works at Fontys University of Applied Sciences ICT as a lecturer and coordinator. He is also a Practor at ROC Tilburg and a trainer at Google Digital Workshop and ICM. As a blogger, he writes for, among others, Emerce.nl, Marketingfacts.nl, Dutchcowboys.nl and Digitalmarketingblog.nl.

Acknowledgments

What started as writing an article with the opinions of about five experts has resulted in a book of 40 experts after a year. A large number of other experts were unfortunately unable to participate due to busy schedules but took the trouble to send a response, often with references to sources and names of other experts I could approach.

Hereby I would also like to thank them, in no particular order:

Maria Axente, Frédérick Bruneault, Professor Stuart Russell, Professor Luciano Floridi, prof. dr Max Louwerse, prof. dr M. (Mehmet) Aksit, MEng, prof. dr PPCC Verbeek (Peter-Paul), MEng, dr S. (Sennay) Ghebreab, prof. dr B.H.M. (Bart) Custers, LL.M., MEng, Marietje Schaake, Peter Joosten, Geert ten Dam, Joshua B. Cohen, Assistant prof. Loek Cleophas, Eddie Altenburg-Collin, Steven Van Belleghem, prof. dr Sabine Roeser, Piek Visser-Knijff, Edwin Borst, Dorothea Baur, Ahmed Larouz, Colette Cuijpers, Arjan Kors, Jessy Kouwenberg, Bert Deen, Hanan Challouki, Jeroen Vinkesteijn, Sander Duivestein, Ricardo A. Abdoel, Marc Appels, Jasper de Wilde, Chanel Matil Lodik.

Special thanks to Frank de Nijs, who edited my book first. Thanks to Sandra Verhoeven, it ended up being a book.

Bianca Lathouwers, thank you very much. You checked every word and sentence. Thank you for your suggestions, feedback and improvements.

Thanks also to Audrey Kawarmala and Enson for design and formatting. We often discussed this with each other.

Colleagues and friends who helped me through the process with tips and advice: Aldwin van de Ven, Koen Suilen, Pauline Schepers-van der Rest, Twan Arts, Halil Kılıç.

The management team of Fontys University of Applied Sciences ICT, for the trust and freedom to professionalise yourself as a lecturer.

Without support and help from my dear wife and children, I would not have succeeded. Thank you very much Funda. Thanks very much Umut, Destan and Güven.

Sources

Alexa Hagerty, A. A. (2021, April 15). The Conversation.
Retrieved from https://theconversation.com/ai-is-increasingly-being-used-to-identify-emotions-heres-whats-at-stake-158809

Conger, K. (2021, April 13). New York Times.
Retrieved from https://www.nytimes.com/2021/04/13/technology/racist-computer-engineering-terms-ietf.html

Durkee, M. (2021, April 8). OECD.AI.
Retrieved from https://oecd.ai/wonk/how-to-achieve-trustworthy-algorithmic-decision-making

Feathers, T. (2021, May 20). Vice.
Retrieved from https://www.vice.com/en/article/m7evmy/googles-new-dermatology-app-wasnt-designed-for-people-with-darker-skin

Heaven, W. D. (2020, September 23). MIT Technology Review.
Retrieved from https://www-technologyreview-com.cdn.ampproject.org/c/s/www.technologyreview.com/2020/09/23/1008757/interview-winner-million-dollar-ai-prize-cancer-healthcare-regulation/amp/

Jorge Barrera, A. L. (2021, May 17). CBC.ca.
Retrieved from https://www.cbc.ca/news/science/artificial-intelligence-racism-bias-1.6027150

Lang, S. (2021, April 7). LinkedIn.
Retrieved from https://www.linkedin.com/pulse/future-says-ethical-ai-sean-lang/

Russell, S. (2021, April 4). Standard.co.uk.
Retrieved from https://www.standard.co.uk/news/uk/cambridge-university-scientists-ucl-china-b927785.html

Ujué Agudo, H. M. (2021, April 21).
The influence of algorithms on political and dating decisions.
Retrieved from http://dx.doi.org/10.1371/journal.pone.0249454

Villanustre, F. (2021, April 8). CMS Wire.
Retrieved from https://www.cmswire.com/information-management/make-responsible-ai-part-of-your-companys-dna/

Wood, M. (2021, March 22). Marketplace.
Retrieved from https://www.marketplace.org/shows/marketplace-tech/bias-in-facial-recognition-isnt-hard-to-discover-but-its-hard-to-get-rid-of/

Fontys

› FOR SOCIETY

Erdinç Saçan, Fontys Hogeschool ICT, 2022