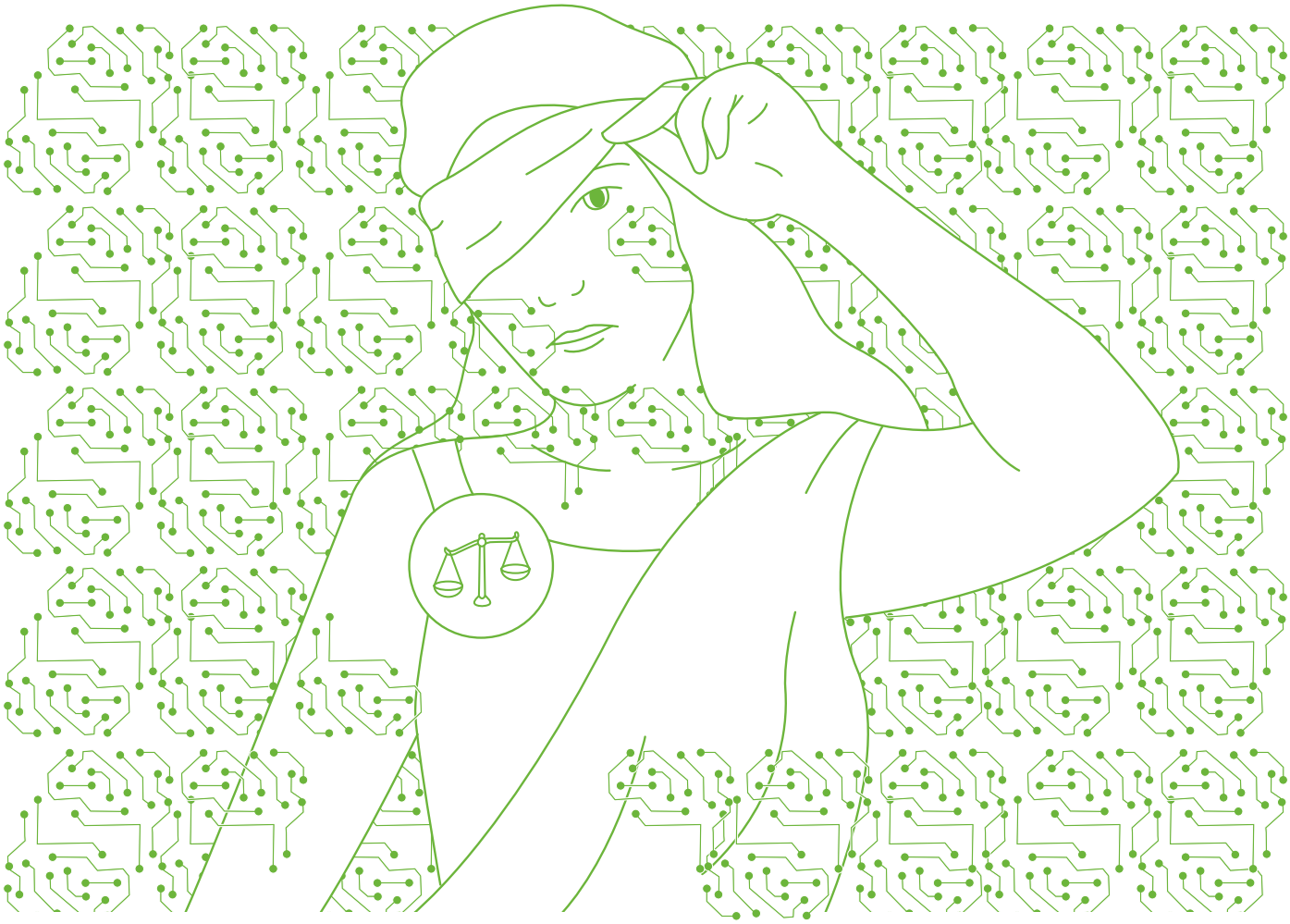
Artificial intelligence is no more discriminating than the data with which it is fed

Interview with Henrike Schlottmann
Jan Kuhlen



W.I.R.E.
THINK TANK FÜR WIRTSCHAFT, WISSENSCHAFT & GESELLSCHAFT

The use of artificial intelligence (AI) as the next step in algorithm-based decision-making is spreading across industries, and with it the risk of discrimination against individuals and groups. Henrike Schlottmann and Jan Kuhlen from the Zentrum für digitalen Fortschritt (centre for digital progress) explain in an interview what can be done to reduce unwanted unequal treatment. Their assessment focuses on the hypothesis that artificial intelligence simply reflects social bias and that discrimination can be easier to correct by systematising the errors – in both digital and analogue life.

Why is artificial intelligence discriminatory?
Jan Kuhlen: Discrimination means the equal treatment of unequal persons or the unequal treatment of equal persons – without justification. In principle, technology is not discriminatory, unless human beings draw discriminatory conclusions from it, or train it so that the results automatically lead to discrimination. The same applies to artificial intelligence. Unwanted discrimination can occur during its use, because it's unclear how an artificial neural network comes up with its results. There's currently a lack of awareness of the relevant criteria for assessing the challenges.

What forms of discrimination are known in the use of artificial intelligence?
Henrike Schlottmann: Discrimination can take place at four levels. It depends on the training data, the selected algorithms, the team and the objectives. It is a person who chooses discriminatory targets or deploys teams who have blind spots. When AI is used, discrimination is also usually based on protected attributes such as sex, skin colour or religion. In principle, this is the same problem as in analogue life, because the algorithm fed with data merely reflects the bias. And this already exists in society.

Can you give us an example?
Henrike Schlottmann: A classic example is the recruiting algorithm used at Amazon[1], which seemed discriminatory during the application process: It came to the conclusion that applications from women should be given a lower rating. This is because the algorithm was trained on an unbalanced data set – based on applications received by Amazon over the last ten years. Most of the applications came from men. The system therefore concluded that men were the preferred candidates. Another example is predictive policing, where automated techniques are used in law enforcement to identify potential criminal activity. In the USA, for example, predictions were made about relapse rates with an inherent bias in the training data – where non-whites were systematically discriminated against on the basis of their skin colour[2].

By "rational discrimination" do you mean the optimisation of the individual at the expense of the structure of solidarity, as sociologist Stefan Selke does?
Jan Kuhlen: Life logging, i.e. the collection of large amounts of data from the daily life of an individual, leads to more measurement and evaluation of human beings, which could also result in a greater measurement of society. This way, individuals can be compared with one another on a larger scale. Critics say this measurement robs us of some of our privacy.
Henrike Schlottmann: The term "rational" has a systematic element to it. With regard to discrimination, this means that the bias already exists in society and is reflected by the algorithm, but the application systematises the whole concept. In other words, it's no longer a human being who decides on a case-by-case basis, but a system that acts systematically. At the same time, this makes discrimination more visible and correctable.

1: Artificial intelligence is (still) discriminatory: https://www.zeit.de/arbeit/2018-10/bewerbungsroboter-kuenstliche-intelligenz-amazon-frauen-diskriminierung
2: Predictive policing: tracking future crimes: https://www.bpb.de/dialog/netzdebatte/238995/predictive-policing-dem-verbrechen-der-zukunft-auf-der-spur?type=galerie&show=image&i=238997

Are there applications in which rational discrimination is particularly rife?
Jan Kuhlen: With health data collected by apps, the risk of rational discrimination is particularly high. However, this is not necessarily the case with discrimination by artificial intelligence. In this case there is primarily a risk of discrimination with regard to the evaluation of persons, particularly with protected attributes.

What is the relationship between life logging and the use of artificial intelligence?
Henrike Schlottmann: On its own, life logging or big data is not artificial intelligence, but it creates the basis for intelligent analyses.
Jan Kuhlen: If applied in this way, artificial intelligence can use this data to make deductions and generalisations that are not based on causality. For example: anyone who gets up at 7 am every morning and jogs three kilometres is more successful.

Is there also rational discrimination in analogue life?
Jan Kuhlen: Yes. If an insurance company sends me a questionnaire, the risk evaluation is also based on this data. However, evaluations can be fine-tuned with big data and are easier to systematise using artificial intelligence, which also leads to better results. They make it possible to take the individual risk situation into account. The whole system is currently being laid bare through artificial intelligence. So far too little attention has also been paid to rational discrimination.

You've written a position paper[3] calling for binding rules for the use of AI in order to prevent misuse.
Henrike Schlottmann: We'd like to see a global ethics council put in place to ensure that certain rules and guidelines apply to everyone involved.

How could this prevent discrimination?
Jan Kuhlen: We must try to prevent discrimination on all four of the levels mentioned above. Teams must be selected in such a way that there are no blind spots, data must be divided into learning and review data and, in case of doubt, checked using black box analyses – an indirect approach that, for example, subsequently produces explanatory models through visualisation. Even if the mechanics of an algorithmic black box – i.e. a closed system – is unknown, the behaviour, i.e. inputs and outputs, can be checked.
Henrike Schlottmann: The systems must be as clear and comprehensible as possible. You should be able to understand how decisions are made. For example, the algorithms could be audited to check the data for bias. However, the tools required for this always depend on the individual definition of fairness and still require a great deal of optimisation.

Do we always have to be able to argue with causalities in order to give plausible reasons for decisions?
Jan Kuhlen: When it comes to unavoidable government decisions, we expect causal arguments that are transparent and comprehensible. Decisions like these should not be made on the basis of probability. But artificial intelligence is always based on probability values. So caution should be exercised when using AI. However, we cannot always expect a technology to produce a fully traceable output from an input. We don't expect that from living creatures either – not even human beings. Society will have to get used to the fact that, while technology in laboratory conditions doesn't deliver 100% causal results, it is much more precise overall than human actions.

3: How artificial intelligence affects freedom, justice and solidarity: https://d-64.org/wp-content/uploads/2018/11/D64-Grundwerte-KI.pdf

Where can AI be used today without hesitation?
Jan Kuhlen: There are already numerous technologies with automated learning components that work well. For example, translation, video analyses and voice recognition. Most people don't have to worry about how an algorithm works to achieve good translation results. What's important is that the sentence is well translated. With voice recognition – for example speech input that is converted into text – most users don't need to understand why a word is recognised.

But why can't we argue that technologies that are not comprehensible should be completely banned because they could – without justification – always appear discriminatory?
Jan Kuhlen: If this question were applied to people, we would also have to restrict people's decisions, because they could always appear discriminatory without justification. On the whole, artificial intelligence is an achievement with enormous potential. But as I said, in some areas, for example where government decisions are made, its use needs to be examined more closely.

The introduction of a global ethics council seems illusory, since normative principles don't apply worldwide, as clearly demonstrated by the Moral Machine online game deployed by the Massachusetts Institute of Technology (MIT)[4]: Test persons were asked to make a series of ethical decisions regarding fictional self-driving car crashes and to decide which consequences were justifiable in their eyes. The results show that ethical principles vary greatly geographically.
Jan Kuhlen: Certain rules or standards have become established worldwide. I don't consider this out of the question for artificial intelligence either. For example, the coalition agreement between CDU, CSU and SPD calls for the prohibition of autonomous weapon systems[5]. As with atomic weapons, this is only possible, of course, with worldwide agreement. In the same way, we could globally agree to identify bots or avatars as machines rather than people. We must also agree when decisions have to be comprehensible and transparent. Of course, the more countries and cultures that are involved, the harder it will be to create a common ethical code of values.

What happens if we don't succeed?
Henrike Schlottmann: We can imagine what the world would look like if the USA or China were to establish its own rules by looking at China's social credit system. China is currently trying to set up a form of surveillance technology that can immediately identify each individual, even in a vast mass of people. You're rewarded for good behaviour and punished for bad behaviour. We have to ask ourselves: Is this the kind of world we want to live in in the future? I believe we should strive to discuss certain rules on an international level because artificial intelligence is a global issue. AI applications are available around the globe, and the internet gives everyone access to this market.

You're also calling for data to be handled based on the concept of solidarity. What do you mean by this?
Jan Kuhlen: It's about using data to get meaningful added value. In other words, non-personal data should be made available to society – in a show of solidarity – with the aim of creating social innovations geared towards the common good. This way, for example, sim-

4: Moral machine: can an autonomous car act ethically?: https://www.dw.com/de/moral-machine-kann-ein-selbstfahrendes-auto-ethisch-handeln/a-46045294-0
5: Coalition agreement between CDU, CSU and SPD: https://www.bundesregierung.de/breg-de/themen/koalitionsvertrag-zwischen-cdu-csu-und-spd-195906

plifications brought about by artificial intelligence can also benefit public service. Today, many ideas still fail due to a lack of data material or a lack of preparation, which consumes a lot of time and resources. For instance, the German economy is currently having problems obtaining sufficient quantities of high-quality data material. At the same time, a great deal of knowledge about German consumer behaviour has been stored with American tech giants such as Facebook or Google. Although data material is available in the German economy, it is not shared and only covers a specific area.

What do you suggest?
Jan Kuhlen: One solution, for example, would be to create intelligent data platforms to train or exchange data. This calls for common standards so that all stakeholders can upload their data to a platform and make it available to others - either for a fee or free of charge.
Henrike Schlottmann: Of course, it must be ensured that the data is not personalised, otherwise certain data could be assigned to certain persons. And not everyone wants information about themselves to be freely available. Depersonalising data is difficult because it is not enough to just remove the name. Systems must be put in place to prevent other attributes, for example street or education, from being triangulated in order to identify individuals.

Do you also think that solidarity is necessary when it comes to health data?
Jan Kuhlen: Collecting and sharing health data is problematic, because it contains highly sensitive information. After all, it's about a person's state of health. There's a reason why we have doctor-patient confidentiality. It is socially accepted that this information does not have to be shared. In principle, however, it's not a question of sharing personal data, but of depersonalising it. At the same time it's important that the data systems are not attacked, otherwise we would have a problem with personal self-determination. Everyone needs the same health protection, no matter whether they go jogging at 7 in the morning or not. Statutory health insurers must not be allowed to make any algorithmically justified differences, be it in relation to the range of benefits on offer or the costs.

Solidarity is a human concept. What does this mean for the development of AI systems?
Jan Kuhlen: Technology is essentially neutral, but can of course also be used in solidarity. If, for example, AI were to improve the management of electricity grids and thereby reduce electricity consumption, everyone would benefit. All uses that are oriented towards the common good can, and should, be promoted.
Henrike Schlottmann: In the end, they are also political decisions. One thing is clear: If I have an objective that is not based on solidarity, the result will not be based on solidarity. If, for example, an algorithm is used to determine sportiness or body weight and a worse health insurance tariff is offered as a result, this is not a system based on solidarity. AI can therefore only demonstrate solidarity if the algorithm developed by humans specifies objectives based on solidarity.

Finally: What social benefits do you see from AI?
Jan Kuhlen: We have to take concerted action against the negative portrayal of artificial intelligence. One big success is that it helps automate processes and simplify tasks, because this creates a lot of space for new tasks and creativity. We need to be aware that the digital revolution is an industrial revolution.
Henrike Schlottmann: I'd like to come back to the issue of discrimination, because AI also has its positive sides here: The algorithms highlight certain forms of discrimination that are prevalent in society, and systematisation can be used to take action against it.

Henrike Schlottmann and **Jan Kuhlen** head up the Artificial Intelligence working group at D64 – Zentrum für digitalen Fortschritt (centre for digital progress), a non-profit digital-political association. Schlottmann has an MSci in mathematics from University College London. She worked for several years for an international management consultancy and oversaw projects in the field of digitalisation and innovation. As co-CEO of ProjectTogether she helps social entrepreneurs get their projects off the ground. Kuhlen is a lawyer and sociologist. At his law firm KUHLEN Partnerschaft von Rechtsanwälten he advises start-ups, agencies and SMEs on issues of corporate and commercial law. Jan Kuhlen is a member of the German Federal Government's "Artificial intelligence" study commission and the artificial intelligence council of experts.