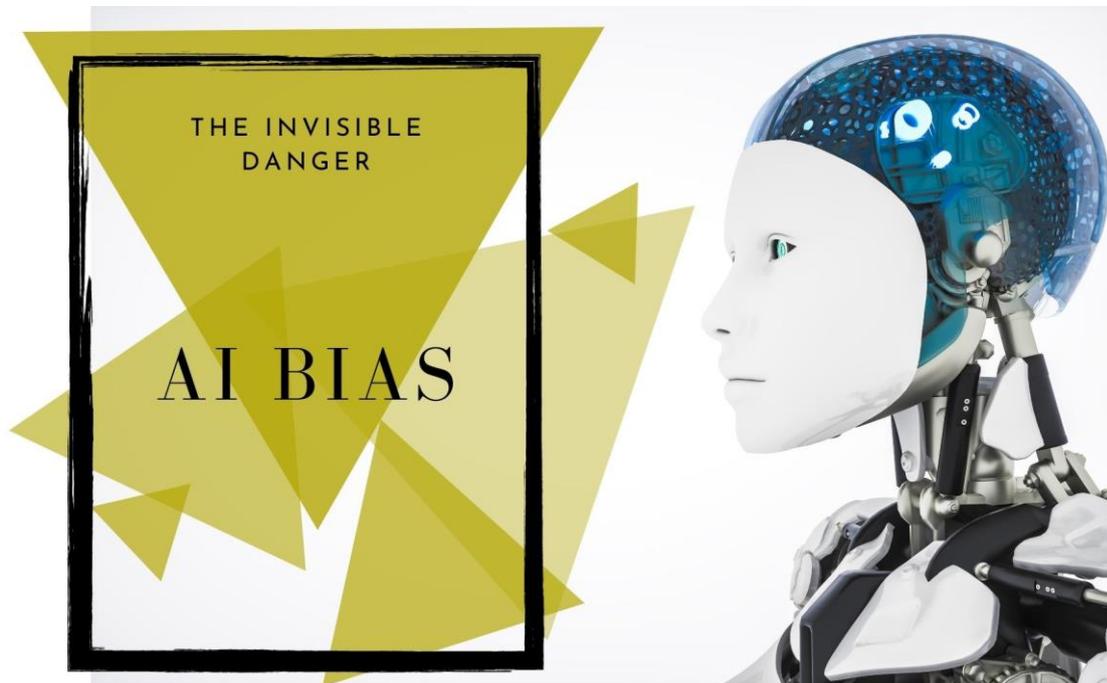




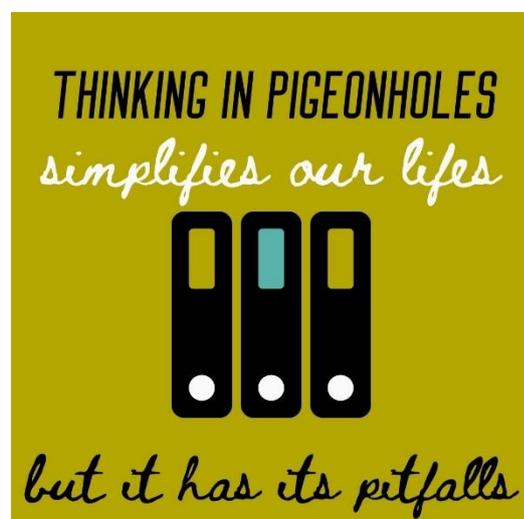
THE INVISIBLE DANGER: AI BIAS



We women are particularly familiar with prejudices; they accompany us from the cradle to the casket. Prejudices are the reason why women are still rare in leadership positions. Erasing prejudice is a slow process because pigeonholing, and shallow categorization is an important part of our evolutionary history. Even in our newest technologies, bias can be found. Algorithms, which increasingly influence and determine our lives, are a major barrier on the way to a prejudice-free(er) future.

WE ALL THINK IN CATEGORIZATIONS ...

... because the automatic classification of people and situations has a distinct evolutionary benefit: It simplifies decision-making processes and thus saves mental capacity and enables us to respond swiftly to recognizable scenarios. However, the biggest danger in categorizing people is it isn't always fair. After all, we only recently learned the difference between structural and unconscious biases. Both are attitudes and beliefs that lie outside our consciousness.



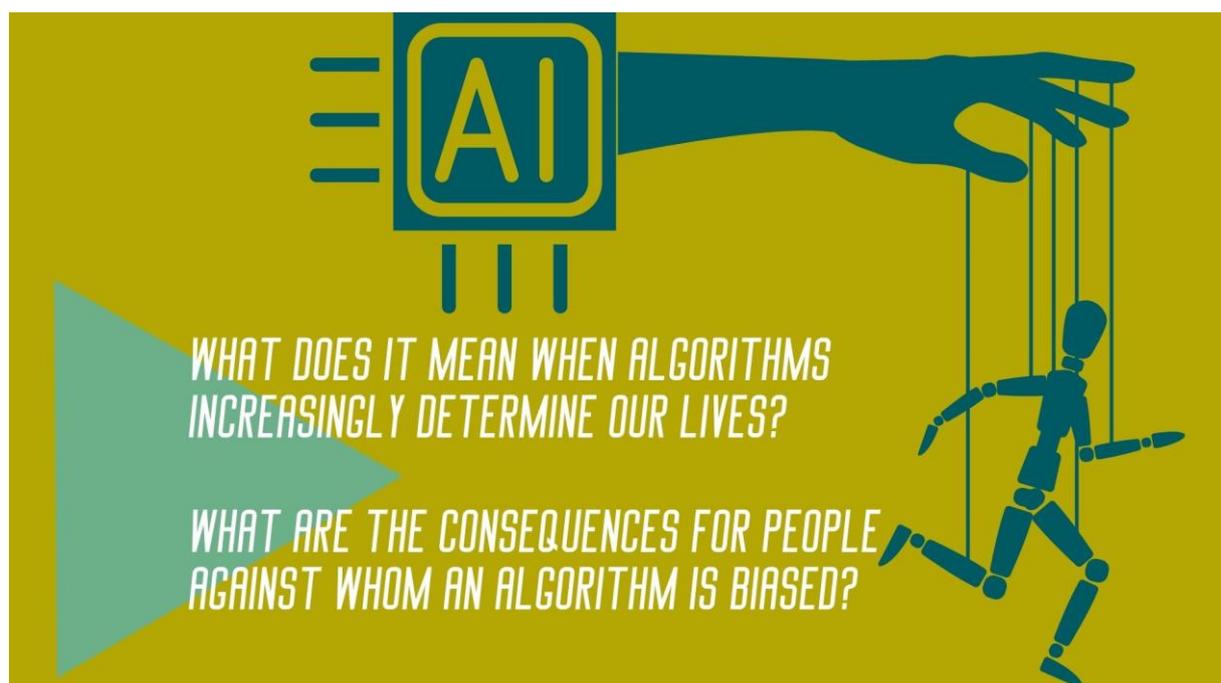
Structural biases are collective, system-defined prejudices that we usually don't question, but affect all our lives on a daily basis. Want an example? In the auto industry, safety testing is done primarily with male dummies. As a result, women are 47% more likely to be seriously injured in an accident and 17% more likely to die in an accident.

Unconscious biases are our individual and unique categorizations acquired through experience. They are effectively the footprint of structural bias on an individual. While we cannot resolve structural biases on our own, we can take concrete steps with respect to our individual Unconscious Bias. But there is a new and strikingly important form of human categorization that was only recently recognized and far more elusive: AI Bias.

THE NEW DIMENSION OF PIGEONHOLE THINKING: AI BIAS

We live in a new digital age, and era of computers and artificial intelligence (AI). All social media platforms, most search engines and news websites are built on algorithms. AI systems are used, among other things, in most image recognition systems used by large companies, such as the image captioning AI used by Facebook, which recognizes whether a person, animal or object is in an image. It not only identifies people in pictures, but also recognizes whether the person is smiling, wearing accessories, standing or sitting, and the total number of people in a group picture. There is little doubt about the power of AI systems. The data analysis and pattern recognition enabled by Deep Learning enables AI to diagnose early-stage cancer with greater accuracy than human doctors. AI can save lives, change lives, but also destroy lives. And therein lies the problem. Above all, AI has one effect: it allows confirmation bias to grow uncontrolled and unchecked and shows us the world as we like it. Every day it confirms our beliefs, opinions, and preferences. We live in a perpetual echo chamber that is designed, coordinated, and updated constantly by AI. Worst of all it's fun. That is terrifying.

Algorithmic biases are a completely new dimension, and we will not be able to avoid having to deal with them more closely. Artificial intelligence (AI) is shaping all our lives and we need to ask ourselves two questions in particular:



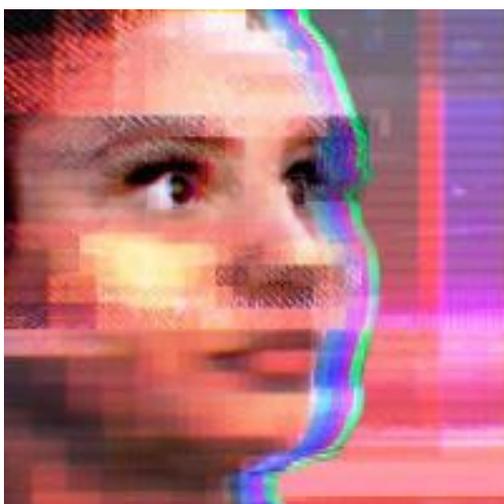
In many large companies, fundamentally important decisions, such as personnel decisions, are supported by artificial intelligence. For example, AI pre-selects who will be invited for interviews. The knowledge of the algorithms that carry such decisions can be fed by flawed, unrepresentative, and prejudiced past experiences. This has the potential to inhibit positive future development towards an unbiased future. Explained less abstractly, when an algorithm decides who should be shortlisted for an advertised executive position, it preemptively eliminates people with immigrant backgrounds, people with disabilities, and women because it relies on data-based empirical values. For an algorithm, middle-aged white men are a safe choice. AI systems cannot address ethical and social concerns unless they are programmed accordingly. And that's exactly what's lacking in the male-dominated IT industry.



Read also this <https://www.jdsupra.com/legalnews/watson-as-a-warning-current-limits-and-9792780/>

TECHNICALLY EXPLAINED

Machine learning, a subset of artificial intelligence, is dependent on the quality, objectivity, and size of the training data used to learn. Faulty, poor, incomplete, and especially outdated data lead to inaccurate predictions and result in "garbage in, garbage out," a joking phrase from computer science that states that if the input is invalid or not meaningful, a computer is very likely to produce an invalid or non-meaningful output. Biases in artificial intelligence machine learning are due in part to problems caused by biased individuals designing and/or training these machine learning systems. Second, incomplete, erroneous, biased, and outdated datasets are used to train and/or validate the machine learning systems. Types of cognitive biases that can unintentionally affect algorithms include stereotyping, bandwagon effect, priming, selective perception, and confirmation bias.



The failure of AI

Tay is a classic example of artificial intelligence failure and how learning mechanisms take on a life of their own, making them difficult to control. Tay was an artificial intelligence chatbot originally released by Microsoft Corporation via Twitter on March 23, 2016. It sparked controversy when the bot began posting inflammatory and offensive tweets via its Twitter account, prompting Microsoft to shut down the service just 16 hours after its launch.

How did this happen? Some Twitter users started tweeting politically incorrect expressions and taught Tay inflammatory messages revolving around common topics on the Internet, such as 'red-pilling' and 'Gamergate.' The robot began posting racist and sexually provocative messages in response to other Twitter users. Artificial intelligence expert

Roman Yampolskiy commented that Tay's misbehavior was understandable because it was imitating the intentionally offensive behavior of other Twitter users and Microsoft had not given the robot an understanding of inappropriate behavior.

IT'S NOT THE ALGORITHM'S FAULT

AI bias is generally unintentional. Artificial intelligence is not evil. But the consequences of its decisions can be significant: poor customer service, lower sales and revenue, unfair and even illegal actions. And: they can even lead to potentially dangerous conditions.

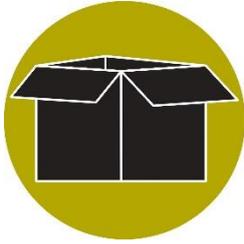


YouTube Tipp: Lo and Behold Clip - The Evolution of AI
https://www.youtube.com/watch?v=_ChGhnbCy6g

In a conversation with director Werner Herzog about the dangers of AI, Tesla founder and AI skeptic Elon Musk gave a compelling example: Suppose an artificial intelligence was tasked with maximizing the value of an investment portfolio. Suppose also that the system's creators did not clearly specify how to achieve that goal. Theoretically, in this case, the machine intelligence could invest more in defense stocks, triggering a war in the worst-case scenario.



This makes it all the more important to check the data used to train machine learning models for lack of completeness and cognitive bias. It is imperative that the data represent all of humanity regardless of its origin, background, gender, and culture. It is up to the data scientists who develop these algorithms to design the data samples in such a way that AI bias is at least minimized, and there needs to be oversight and decision makers who evaluate when it is appropriate and when it is inappropriate to apply machine learning technology.



WATCHOUT! THE BLACKBOX PROBLEM:

The greatest danger of AI-systems is the "Blackbox Problem" -- we are now unable to fully understand why the algorithms behind AI work the way they do.

Our brain is the most vital, but at the same time the most complex and enigmatic of our organs. Artificial intelligence, which represents the pinnacle of human technological development, is – despite us creating it – in many ways a mystery. We do know a great deal about our brains, and we can predict with some certainty how they will respond to various stimuli. Likewise, we can know with some certainty what results an AI algorithm will produce given certain inputs. The mystery facing scientists and researchers is not that of the output, but how it is generated. This lack of knowledge about the inner workings of the "black box" of artificial intelligence is the biggest hurdle in AI development and will be with us for a long time, if not forever.

Harvard Business Review says: *"One of the biggest sources of anxiety about A.I. is not that it will turn against us, but that we simply cannot understand how it works. The solution to rogue systems that discriminate against women in credit applications or that make racist recommendations in criminal sentencing, or that reduce the number of black patients identified as needing extra medical care, might seem to be "explainable A.I." But sometimes, what's just as important as knowing "why" an algorithm made a decision, is being able to ask "what" it was being optimized for in the first place?"*

At the moment, the IQ of artificial intelligence is still below the average value of human intelligence, but AI's intelligence is steadily increasing and will inevitably overtake ours, making it even more incomprehensible and significantly more uncontrollable.



Read more about on that topic :

https://en.wikipedia.org/wiki/Artificial_general_intelligence

The economic philosopher Anders Indset, in an interview about his book 'Quantopia' warned of the dangers of artificial intelligence, even goes so far as to say:

'The belief that we simple apes could build and dominate an external digital super god is the height of narcissism.'

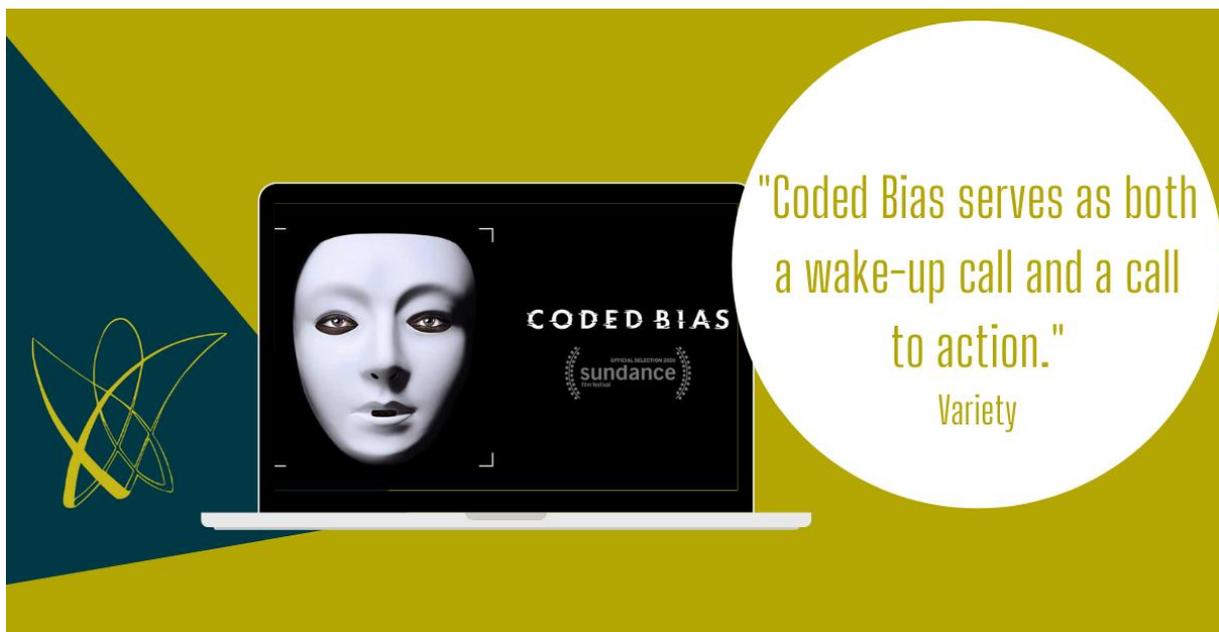
The widespread use of artificial intelligence will be a particular challenge for humanity. Organizations that use AI will need to disclose the human decisions behind the design of their AI systems, as well as what ethical and social concerns they have considered, and how well they have monitored the results of those systems for traces of bias or discrimination. We need models we can trust. Achieving transparency in AI systems is critical. And we need one thing above all else: more diversity in computer science.

But what does that mean in concrete terms? What can I do?

- 1 We must become digitally literate; intensively engage with the technologies we use and understand them as well as possible. We must not leave the field to the men.
- 2 It is our mission and task to question AI models, we must insist that job advertisements are written in a gender-neutral way and that colleagues review the selection of AI and that diverse data is used as a basis.
- 3 A conflicting issue for many but incredibly important: AI learns from our language. Therefore, it is even more important to gender. Changing the way we use language is an important milestone in the elimination of bias.

Film recommendation on the topic

CODED BIAS



In the documentary *Coded Bias*, filmmaker Shalini Kantayya follows computer scientist Joy Buolamwini of the M.I.T Media Lab and data scientists, mathematicians, and watchdog groups around the world to expose the discrimination caused by facial recognition algorithms that are now prevalent in all aspects of daily life. This documentary can be seen on Netflix.

"When you think of A.I., it's forward-looking, but A.I. is based on data, and data is a reflection of our history."

Joy Buolamwini