Decisions, Decisions

By Bram Vanderborght

A few weeks ago, I was asked by a journal to review a paper with a title that was very far from my research interests. Because this had happened several times previously with this particular journal, I decided to contact the editor, who replied that "an automated program selects reviewers." I have no problem that a machine does a preselection. Finding suitable reviewers is a tough problem, and any help is welcome. Furthermore, in our society, software is used, for example, to check for plagiarism, but it is a human that does the final assessment.

With the evolution of robotics and artificial intelligence (AI), machines make more and more decisions. This leads to relevant applications that improve our life, economy, and society in ways that, of course, extend beyond the publishing industry: self-driving vehicles, better diagnoses that help to detect cancer, data analyses that assist insurance companies, camera images processed in real time to estimate dangers posed to crowds at major events, and so on. With all of these applications, algorithms make decisions for and about us. But, as Ann Nowé (Vrije Universiteit Brussel), Katleen Gabriels (Maastricht University), and I argued in *Knack*, algorithms should not exercise their power over us: people decide whether to delegate decision making to algorithms and under what conditions. Those conditions must be more transparent.

Digital Object Identifier 10.1109/MRA.2019.2921468
Date of publication: 5 September 2019

Decisions are often so complex that they are difficult to grasp in instructions executed by a computer. Calculating the statistical probability that someone is working in a certain research field is still manageable. But what about a multitude of data? In such cases, self-learning algorithms—computer algorithms that can teach themselves a decision-making process—offer relief. For this, you have to feed the computer a lot of data, from which algorithms learn to recognize patterns. Because of the enormous amount of data and a lot of computing power, decisions are getting better and are often more accurate than those made by an expert having years of experience. After all, a computer is more adept and thorough than a human in searching huge amounts of information. Unfortunately, the computer often cannot explain why a certain decision was made. The decision-making process the machine mastered is not transparent for people. Unlike a human expert, such a system cannot reflect on the decision or its consequences.

If landlords refuse to rent out their home to an immigrant couple, then they can rightly be accused of discrimination. But when an algorithm comes to that decision in an unfathomable way, there seems to be an objectivity that often does not exist. There is a real risk of discrimination when the algorithm bases a decision on a trait such as gender or ethnic origin. Despite these risks, such information is offered today to algorithms. In the United States, AI is used to estimate the likelihood of criminals' reoffending. Research has shown that they give an unreasonably high score to blacks and too low a score to whites.

Often, indirect links are used by the algorithm, such as links between hobbies and gender. Even though these relationships are not particularly strong, they can be strengthened by another weak relationship, such as that between occupation and gender. Consequently, gender is indirectly included in the decision. Amazon ultimately abandoned its AI recruitment tool because the software showed a preference for men. After all, that software "learned" from the curricula vitae the company received, which underrepresented women. The algorithms even favored letters that used typically masculine words.

There is a growing need for AI systems that can provide an explanation to the user and make clear which criteria were taken into account to arrive at a decision. That is a first step in the direction of true, interpretable AI that allows people to be proactively informed about all the nuances and relevant patterns behind the decision. Which aspects of your profile led you to be chosen as a reviewer, awarded an insurance policy, provided a risk assessment, or given medical treatment? As consumers, patients, and citizens, we have the right to receive that information in clear language. Although the transparency

(continued on page 13)

shown that not all proposed fairness criteria can be optimized at the same time, therefore limiting guarantees that an algorithm will always be fair. An adequate enumeration of protected attributes and their proxies (e.g., ZIP codes are correlated with race in the United States, thus leading to redlining practices that exacerbate inequalities) might be difficult without, for example, working with sociologists who can help put real-world abstract data into categories and contextualize data in particular socioeconomic and cultural environments. Several companies are releasing fairness tool kits to be integrated into software-development pipelines early in the process. Industrial standards, such as the IEEE Ethics Certification Program for Autonomous and Intelligent Systems within the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, are also being proposed [12]. However, to date, the methodology for certification is not known, and it is not clear whether such certification is enough to truly avoid biased systems.

More holistic approaches that encompass a wider range of disciplines have recently been advocated as necessary during the design of complex AI or robotic systems [9] as a way not only to provide technical solutions but also to better understand the socioeconomic and cultural contexts of use of the technology. This includes taking special care in the construction, documentation, and validation of data sets and making the push for more transparent algorithms that can be monitored and audited during real-world operations, perhaps by independent institutions.

Wide societal acceptance and trust of robotic systems will require a concerted effort involving sociologists, ethicists, philosophers, and technologists to ensure fairness and build trust in deploying autonomous and interactive systems. Perhaps Mark Twain wasn't thinking about bias in robotics, but he said it best when he noted, "What gets us into trouble is not what we don't know. It's what we know for sure that just ain't so"!

References

[1] J. Vincent, "Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech," *The Verge*, Jan. 12, 2018. [Online]. Available: https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai

[2] I. D. Raji and J. Buolamwini, "Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products," in *Proc. AAAI/ACM Conf. AI Ethics and Society*, 2019. [Online]. Available: http://www.aies-conference.com/wp-content/uploads/2019/01/AIES-19_paper_223.pdf

[3] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks," *ProPublica*, May 23, 2016. [Online]. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[4] A. Datta, M. C. Tschantz, and A. Datta, "Automated experiments on ad privacy settings: A Tale of opacity, choice, and discrimination," in *Proc. Privacy Enhancing Technologies*, 2015, pp. 92–112. doi: 10.1515/popets-2015-0007.

[5] S. Myers West, M. Whittake, and K. Crawford, "Discriminating systems: Gender, race and power in AI," AI Now Institute, New York Univ., New York, 2019. [Online]. Available: https://ainowinstitute.org/discriminatingsystems.pdf

[6] A. Howard and J. Borenstein, "The ugly truth about ourselves and our robot creations: The problem of bias and social inequity," *Sci. Eng. Ethics*, vol. 24, no. 5, pp. 1521–1536, Oct. 2017. doi: 10.1007/s11948-017-9975-2.

[7] H. Han and A. K. Jain, "Age, gender and race estimation from unconstrained face images," Michigan State Univ., East Lansing, Tech. Rep. MSU-CSE-14-5, 2014. [Online]. Available: http://biometrics.cse.msu.edu/Publications/Face/HanJain_UnconstrainedAgeGenderRaceEstimation_MSUTechReport2014.pdf

[8] J. Kleinberg, S. Mullainathan, and M. Raghavan, Inherent trade-offs in the fair determination of risk scores. 2016. [Online]. Available: https://arxiv.org/abs/1609.05807

[9] M. Whittaker et al., "AI now report," AI Now Institute, New York Univ., New York, 2018. [Online]. Available: https://ainowinstitute.org/AI_Now_2018_Report.pdf

[10] J. Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," Reuters, Oct. 9, 2018. [Online]. Available: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

[11] B. Wilson, J. Hoffman, and J. Morgenstern, Predictive inequity in object detection. 2019. [Online]. Available: https://arxiv.org/abs/1902.11097

[12] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethics in action." [Online]. Available: https://ethicsinaction.ieee.org/

FROM THE EDITOR'S DESK *(continued from page 4)*

of the reasoning process was an essential part of AI expert systems in the 1980s, the research domain of explainable AI is still in its infancy for systems that learn from so-called big data. A thorough knowledge of AI algorithms is necessary to guarantee aspects of justice toward individuals and groups as well as to prevent bias.

As usual, this September issue of *IEEE Robotics and Automation Magazine* includes articles submitted on a variety of topics rather than focusing on a particular theme, as do special issues. Calls for upcoming special issue papers focus on deep learning and soft robotics. Please check the Society website for more information. To support the reproducibility of robotics and automation research, the IEEE Robotics and Automation Society (RAS) waives, for two years and a maximum of five articles per year, open access fees for reproducible articles (designated *R-Articles*), the first of which appears in this issue. Enjoy your reading!