Expanding Horizons
Business, Management and
Technology for Better Society
20–22 May 2020
Online Conference

Management,
Knowledge and Learning
International Conference 2020
Technology, Innovation
and Industrial Management

Accountability in AI as Global Issue

Elisabetta Azzali
Pegaso International, Italy
amasingdrake@gmail.com

Abstract

Artificial Intelligence are a global phenomenon increasingly crucial in economic development and leadership in this field. At first to make possible effective communication on this argument definition are needed as a first step of knowledge in order of regulate this and safeguard citizens' rights. In 8 April 2019 High Level Expert Group on Artificial Intelligence set up by EC published this text. "Artificial Intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions with some degree of autonomy to achieve specific goals".AI base system can be purely software-based acting in Virtual world (e.g. voice assistants, image analysis, software, search engines, speech and face recognition systems) or I can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or internet of things applications)".

BACKGROUND

Researchers prefers use the notion of rationality, this refers to the ability to choose the best action to take in order to achieve a certain goal, given certain criteria to be optimized and the available resources of course, rationality is not the only ingredient in the concept of intelligence, but it is a significant part of it. The simplest human behaviour is ascribed to intelligence, while even the most complicated insect behaviour is never taken as an indication of intelligence. Psychologists generally do not characterize human intelligence by just one trait, but by the combination of many diverse abilities. Research in AI has focused chiefly on the following components of intelligence: learning, reasoning problem solving, perception and using language. There are a number of different forms of learning as applied to artificial intelligence. The elementary pattern is learning by trial and error. For example, a simple computer program for solving mate in one chess problems might try move at random until mate is found the program. The program might then store the solution with the position so that the next time the computer encountered the same position, it would recall the solution.

The simplest is learning by trial; this simple memorizing of individual items and procedures known as rote learning is relatively easy to implement on a computer more challenging is the problem of implementing what is the problem what is called generalization. Generalization involves applying pat experience to analogous new situations. For example, a program that learns the past tense of a word such as jump unless previously had been presented with jumped, whereas a program that is able to generalize can learn the "Add ed" rule and so form the past tense of jump based on experience with similar verbs. Reasoning consist in to draw inferences appropriate to the situation. Inferences are classified as either deductive or inductive. The most significant difference between these forms of reasoning is that in the deductive case the truth of the premises guarantees the truth of the conclusion, whereas in the inductive case the truth of the premise lens support to the conclusion without giving absolute assurance. Inductive reasoning is common in science where data are collected, and tentative models are developed to describe and predict future behaviour until the appearance of anomalous data forces the model to be revised. Deductive reasoning is common in mathematics and logic, where elaborate structures of irrefutable theorems are built up from a small set of basic axioms and rules. There has been considerable success in programming computers, to draw influences, especially deductive inferences[1]. However, true reasoning involves more than just drawing inferences, it involves drawing inferences relevant to the solution of the particular task or situation. This is one of the hardest problems confronting AI. Problem solving particularly in Artificial Intelligence may be characterized as a systematic search through a range of possible actions in order to reach some predefined goal or solution.

Problem solving methods divide in special purpose and General Purpose. A special purpose method is tailor made for a particular problem and often exploits specific features of the situation in which the problem is embedded. In contrast, a general-purpose method is applicable to a wide variety of problems. One General Purpose technique used in AI is means - end analysis. A step by step, or incremental reduction of the difference between the current state and the final goal the program selects actions from a list of meanings. All about the simplest human behaviour is ascribed to intelligence, while even the most complicated insect behaviour is never taken as an indication of intelligence. Psychologists generally do not characterize human intelligence by just one trait but by the combination of many diverse abilities. Research in AI has focused chiefly on the following components of intelligence: learning, reasoning problem solving, perception and using language. There are a number of different forms of learning as applied to Artificial Intelligence. The elementary pattern is learning by trial and error. For example, a simple computer program for solving mate in one chess problems might try moves a random until mate is found. The program might then store the solution with the position so that the next time the computer encountered the same position it would recall the solution. There has been considerable success in programming computers to draw inferences, especially deductive inferences. However, true reasoning involves more than just drawing inferences; it involves drawing inferences relevant to the solution of the particular task or situation. This is one of the hardest problems confronting A.I. Problem solving, particularly in Artificial Intelligence, may be characterized as a systematic search through a range of possible actions in order to reach some predefined goal or solution. Problem solving methods divide into special purpose and general-purpose methods is applicable to a wide variety of problems. One general purpose technique used in AI is means end-analysis consist in a step by step or incremental reduction of the difference between the current state and the final goal. The program selects actions from a list of means. In the case of a simple robot this might consist of pick up, put down, move forward, move back, move left, and move right until the goal is reached. Many diverse problems have been solved by artificial intelligence programs. Some examples are finding the winning move (or sequence of moves) in a board game devising mathematical proofs and manipulating "virtual objects" in a computer-generated world - perception. In perception the environment is scanned by means of various sensory organs, real or artificial, and the scene is decomposed into separate objects in various spatial relationships. Analysis is complicated by the fact that an object may appear different depending on the angle from which it is viewed, the direction and intensity of illumination in the scene, and how much the object contrasts with the surrounding field. At present time, artificial perception is sufficiently well

[1] History of Computing CSEP 590A University of Washington December 2006

advanced to enable optical sensors to identify individuals, autonomous vehicles to drive at moderate speeds or the open road, and robots to roam through buildings collecting empty soda cans. One of the earliest systems to integrate perception and action was Freddy a stationery robot constructing at the University of Edinburgh (during 1966 - 1973). Language is a system of signs having meaning by convention. In this sense language need to be confined to the spoken word - traffic signs for example, form a mini language it being a matter of convention that a particular sign on a signal means “hazard ahead”. In some countries it is distinctive of languages that linguistic units possess meaning by convention, and linguistic meaning is very different from what is called natural meaning, exemplified in statements such as “those clouds mean rain and “the fall in pressure means the valve is malfunctioning an important characteristic of full - fledged human languages in contrast to bird calls and traffic signs is their productivity. A productive language can formulate an unlimited variety of sentences. It is relatively easy to write computer programs that seem able, in severely restricted contexts, to respond fluently in a human language to questions and statements. Although none of this programs actually understands language they do not have genuine understanding in order whether or not one’s behaviour but also one’s specific culture in the fundamentals of learning. It is relatively easy to write computer programs that seem able, in severely restricted contexts to respond fluently in a human language. To question and statements when involves in genuine understanding, if even a computer that uses language like a native human speaker is not acknowledged to understand ‘. There is no universally agreed upon answer to this difficult question. According to one theory, whether or not one understands depends not only on one’s behaviour but also on one’s history, in order to be said to understand, one must have learned the language and have been trained to take one’s place in the linguistic community by means of interaction with other language users. In 1957 two researchers Allen Newell, researcher at the Rand Corporation, Santa Monica California, Herbert Simon a psychologist and computer scientist at Carnegie Mellon University Pittsburgh Pennsylvania. Summed up the top down approach in what they called the physical symbol is sufficient, in principle, to reduce artificial intelligence in a digital computer and that, moreover, human intelligence is the result of the same type of symbolic manipulations. During 1950's and 60s the top down and bottom up approaches were pursued simultaneously, and both achieved noteworthy, if limited results. Employing these methods AI research attempts to reach one of three goals: strong AI, Applied AI, or cognitive simulation. Strong AI aims to build machines that think (the term strong AI was introduced for this category of research in 1980 (by the philosopher John Searle[2] of the University of California at Berkeley). The ultimate ambition of strong AI is to produce a machine whose overall intellectual ability is indistinguishable from that of a human being. Applied AI, also known as advanced information processing, aims to produce commercially viable “smart” systems for example “expert“ medical diagnosis systems and stock - trading systems applied artificial intelligence has enjoyed considerable success as expert systems. In cognitive simulation computers are used to test theories about how the human mind works, for example, theories about how people recognize faces or recall memories. Cognitive simulation is already a powerful tool. In both neuroscience and cognitive psychology. Connectionism, or neuron like computing, developed out of attempts to understand how the human brain works at the neural level, in particular how people learn and remember. Table 1 describe most relevant biases in AI, particularly relevant in face [3]recognition technology.

[2] The Rediscovery of the Mind. Contributors: John R. Searle - Author. Publisher: MIT Press. Place of publication: Cambridge, MA. Publication year: 1994.

[3] University of Melbourne, 2018, Biometric Mirror highlights flaws in Artificial intelligence- http://www.eurekalert.org/pub

Table 2 Unconscious bias definitions

Short cut biases	**Impartiality biases**	**Self - interest biases**
Availability bias	Anchoring bias	Ingroup/outgroup bias
Base rate fallacy	Bandwagon bias	Sunk cost bias
Congruence bias	Bias blind spot	Status quo bias
Empathy gap bias	Confirmation bias	Not invented here bias
Stereotyping	Halo effect	Self - serving bias

Main factors in promoting AI development are availability of good quality data, software, hardware. Good quality data involve ethics and privacy concerns. E.C. on 8 April 2019[4], the LEG on AI presented ethical guidelines for trustworthy artificial intelligence. According to the guidelines, trustworthy AI should be: (1) Lawful - Respecting all applicable laws and regulations. (2) ethical -Respecting ethical principles and values. (3) Robust- A technical perspective while considering its social environment. The Guidelines put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy. A specific assessment list aims to help verify the application of each of the key requirements.

- Human agency and oversight - AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms needed to be ensured, which can be achieved through human in-the-loop, human on-the-loop, and human in command approach.

- Technical Robustness and safety - AI systems need to be resilient and secure. They need to be safe ensuring a fall-back plan in case something goes wrong as well as being accurate, reliable, and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.

[4] See: https://eu.europa.eu

- Privacy and data governance: besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, considering the quality and integrity of the data, and ensuring legitimized access to data.

- Transparency - the data system and AI business models should be transparent. Traceability mechanisms can help achieving this moreover, AI systems end their decisions should be explained in a manner accepted to the stakeholder concerned human need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.

- Diversity, non-discrimination, and fairness. Unfair bias must be avoided as it could have multiple negative implications, from the marginalization of vulnerable groups to the exacerbation of prejudice and discrimination, fostering diversity. AI systems should be accessible to all, regard less of any disability, and involve relevant stake holders throughout their entire life circle.

- Societal and environmental well-being: AI systems should be accessible to all, regardless of any disability and involve relevant stakeholders throughout their entire life circle. AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should consider the environment, including other living beings, and their social and societal impact. Should be carefully considered.

- Accountability- Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms. Data and design processes plays a key role therein, especially in critical applications. Moreover, adequate, and accessible redress should be ensured.

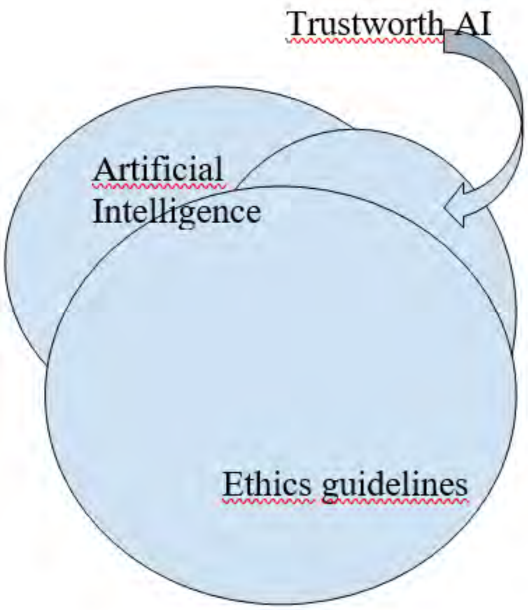
Europe became a leader in ethical AI. Guidelines are not compulsory, but moral suasion and

competitive advantages nudge other countries competitors in follow this example. Ethics lacks mechanisms to reinforce its own normative claims. Enforcement of ethical principles may involve reputational losses in the case of misconduct or restrictions on memberships in certain certification bodies. However, at global level is needed a certification program in order of a more transparent market.

Discussion

On November 13,2019 the European Data protection board published its draft Guidelines 4/2019 the "Guidelines " on the obligations of data protection by design and by default set out under Article 25 of the EU General Data Protection Regulation (GDPR) Text ART 25 GDPR requires all data Controllers irrespective of their size , to implement appropriate technical and organizational measures and necessary safeguards which are designed to implement the basic data protection in Article 5 of the GPR , as well as individuals data protection rights laid down 12.22 of the GDPR as well as well as individuals freedoms set out in recital 4 of the GDPR and the EU Charter of fundamental rights ("data protection by design") ; and appropriate technical and organizational measures for ensuring that, by default only personal data , which is necessary for each specific purpose of the data processing is processed(data protection by default"). Both this requirement serves the same objective, i.e. the effective implementation of the GDPR data protection principles on individual's data protection rights and freedoms regarding the processing of their personal data Guidelines are directly addressed to data controllers, they explicitly recognized data processors and technology providers as key enablers for PDbDD and further provide "Recommendations and how data controllers, data processors and technology providers can cooperate to achieve D.P.bDD as well as leverage it as a competitive advantage. We have summarized. The GDPR requires data controllers to implement data protection by design when they are in the process of determining which means or design elements (e.g. the architecture, procedures, protocols, layout and appearance) must be incorporated into the data processing the amount of personal data collection, extent of process, period of storage accessibility.

Table 3: Trustworthy AI embedded in Algorithms



Strong EU attention in ethics move also other AI strong competitors like US. On November 18, 2019 members from four senate committees realised a set of "core principles" for federal privacy legislation. The principles cover several issues across four categories to protect consumer privacy: establish data safeguards, (2) Invigorate competition, (3) strengthen consumer and civil rights, and (4) impose real accountability. These breaks down more specifically as follows: - Establishing data safeguards - minimization - collection of data must be minimized so it is narrowly tailored to its authorized use. - Abuse prevention - harmful deceptive and abusive collections and uses of data must be prohibited. - Sharing limits - Rules must be established to limit data sharing to that which is necessary to carry out purposes expected and authorized by consumers. -Security: organizations need higher standards for the way they retain and secure data., Invigorate competition. -Market power checks - Consumers must be able to prevent their data from being commingled across separate business within an enterprise and ensure those restrictions apply to data obtained and through mergers and acquisitions. - Data portability- Consumers should be provided with the rights to know, access, delete, correct, and restrict the transfer of their data. Consumers must also receive heightened protections such as a "Do-not-track" right. Civil rights protections- Consumers must have transparency into algorithmic decisions that result in bias or discrimination and have the ability to challenge such decisions. -Impose real accountability: Corporate accountability, the burden of protecting privacy must be shifted from consumers to companies. - Federal (US) enforcement and rulemaking. State and private remedies.

Facial recognition technology is upon us; phones and computers can now be unlocked with your specific facial structure using this gimmick, but it can also scarily be used to identify individuals from security cameras, which are increasingly becoming omnipresent

As a matter of fact, Australian governments have been actively pushing for the implementation of this technology with everyday utilization such as public transport payment and government services. But what about the implications of this technology in healthcare? From better diagnosis of rare conditions to ethnic discrimination, will it revolutionize the healthcare system or be the harbinger of a Black Mirror-worthy reality? We will analyse these questions, short for facial recognition technology, is a method used to identify a person based on their specific facial features like bone structure and skin texture. Its functional algorithm relies on existing databases which it dives into to compare those features in order to output a result. Such software has been in use to identify law offenders or to visualize how missing children might look like as adults but not really in the healthcare sector, that is until recently. This is because with time FRT became more sophisticated (thanks to a larger database of faces) and accessible (even your phone has it!), and is now becoming increasingly attractive in medicine thanks to the numerous ways it can be implemented in this sector from cutting down on paperwork to helping physicians in diagnosis. Facebook prompts you with suggestions about people to tag in a photo this is a prime example of FRT in action – the software can identify someone based off their unique facial features. Now imagine going to your local hospital in the near future for that sore throat that has been bothering you for over a week. Instead of going through the waiting lines for administrative purposes. A virtual assistant will scan your face in a matter of seconds and assign you to your doctor. In so doing, the algorithm can even detect other irregularities like signs of depression and will inform your doctor of such a possibility. Such applications are far from being mere products of the mind. An oft-cited example is Face2Gene, an app used by clinicians that can detect rare genetic conditions like Cornelia de Lange syndrome where patients have characteristic facial features but that can be missed by Physicians because they simply might not have come across it during their clinical practice. This was the case with Omar Abdul-Rahman[5], a medical geneticist, who, thanks to the app, recommended the family of a three year old boy to order a specific genetic test for Mowat-Wilson syndrome which returned positive. “If it weren’t for the app I’m not sure I would have had the confidence to say ‘yes you should spend $1000 on this test,” this app helped the family from making further expenses which might not correctly identify the condition and would further delay the appropriate care that the young boy required. There has even been a recent study based on the deep-learning algorithm Deep Gestalt, a facial image analysis framework, which powers Face2Gene.The algorithm was shown to outperform clinicians in diagnosing syndromes like Noonan syndrome. Deep Gestalt even correctly identified conditions in its top ten list 91% of the time. “It’s like a Google search,” the study’s co-author Karen Gripp tells Pediatrics. With such a high success rate and the ease of using the app, such a comparison is not far-fetched. It would not even be far-fetched to speculate what this technology could lead to or do in the future. Below are our top 3 potential and highly anticipated use of FRT in healthcare in the (near?) future: Smart mirrors “Mirror, Mirror on the Wall, Am I Healthy?” Asking this question to your mirror might soon be possible.

[5] Valentine, M., Bihm, D., Wolf, L., Hoyme, H. E., May, P. A., Buckley, D., … Abdul-Rahman, O. A. (2017). Computer-Aided Recognition of Facial Attributes for Fetal Alcohol Spectrum Disorders. *Pediatrics*, *140*(6), e20162028. doi:10.1542/peds.2016-2028

By combining FRT into a seemingly simple mirror might soon be possible. By combining FRT into a seemingly simple mirror with a built-in camera and existing technologies would not even be far-fetched to speculate what this technology could lead to or do in the future. Below are our top 3 potential and highly anticipated use of FRT in healthcare in the (near?) future: Smart mirrors "Mirror, Mirror on the Wall, Am I Healthy?" Asking this question to your mirror you like SkinVision skin analysis and Nuralogix transdermal optical imaging technique to measure blood pressure and stress level, a quick scan can reveal a lot by simply looking at your own reflection (or by asking your mirror!). Such a smart mirror could advise you to get that new mole on your cheek checked, recommend meditation in case your stress level is higher than usual and refer you to your doctor if there is any abnormal fluctuations in your blood pressure. More than just for patients' health, FRT can also be used for the wellbeing of healthcare practitioners themselves. Medscape's 2019 report found that 44% of physicians feel burned out, 11% were colloquially depressed and 4% suffered from clinical depression. These can be identified via facial analysis and subsequently addressed, like suggesting stress-relieving measures like yoga or even vacations before they further affect the health of healthcare providers. Not only about health the applications of such technology can further be adapted to other areas within the healthcare system. For example, unlawful people like insurance imposters, drug seekers and criminals can be easily identified from a given database and dealt with accordingly. FRT as a security measure does not lie just in one's health but also in that of staffers medical conditions and environment is laud. The answer to this question is: it depends on who you are. While using FRT for aiding in identifying able, the worrying implications of the same seemingly benevolent technology are manifold. As the technology picks up steam and gets used more widely by clinics and physicians, the increasing amount of biometric data collected will present as a real responsibility for those in its possession. Advertising companies would jump on this gold mine and make offers that many cannot refuse. Hackers will find new content to hold hostage and ask ransoms from. With greater amount of personal data comes greater responsibility. The healthcare system will have to double down on its efforts to securely store them and adhere to privacy protection rules like the Health Insurance Portability and Accountability Act (HIPAA) which protects "full-face photographs and any comparable images" and offers the standard for de-identification so that such "health information is not individually identifiable". Inherent bias in databases Even with tightly controlled security over those data, the database itself might be put into question. As we discussed extensively in a recent article, the A.I. behind such technology as facial recognition aren't immune to human-induced bias. We identified three main reasons behind biased algorithms namely judgmental data sets, deeply ingrained social injustices and unconscious or conscious individual choices. These can subsequently affect results of diagnostic methods relying on FRT by discriminating patients based on their sex, ethnicity and even their accent. Additionally, the same study that showed Face2Gene's prowess noted that "these technologies is providing a safer work identified only a few disease phenotypes, limiting their role in clinical settings, where hundreds of diagnoses must be considered". For instance, a 2017 study found that "Face2Gene showed a better recognition rate for DS in Caucasian (80%) compared to African (36.8%)" when assessing Down Syndrome in children. Such limitations are important considerations if Even if facial recognition technology has shown impressive usage in healthcare, it still needs to be worked on to be a must-have tool for every physician. It is just not there yet. It is one thing to have a wide database to be able to identify conditions from, but it is another to have one which is representative of every human being, irrespective of their background. While definitely a challenging goal and one that will involve complex processes and years of hard work, isn't this an ideal that we should aim to achieve? We aim to have fair and reliable algorithms to aid in healthcare provision.

CONCLUSION

European Union early activity in call for an expert panel composed by Academia, Companies, Developers, successfully involves public opinion at global level. This pattern so needs to be replicate at Global level where a unique certification Authority is needed. This should be better for Companies, Consumers, National Governments, and all stakeholders. Trustworthy AI can manage energy and reduce pollution, provide care in mountainous areas, while respect research and intellectual right patent. As Nobel prize Joseph Stieglitz with a group of economists summarizing the deliberation of a panel of experts on the measurement of economic performance and societal progress at the OCD proposed a new dash boards of metrics to access society's health including measures of inequality, environmental sustainability and how people feel about their own lives. Al least we need add at this list transparency in algorithms and informed consent as worldwide truly binding legal framework.

REFERENCES

Baldwin, J. F.; Martin, T. P. and Pilsworth, B. W. (1995). *Fril-Fuzzy and Evidential Reasoning in Artificial Intelligence.* New York, USA: John Wiley & Sons, Inc.

University of Washington. (2006). *History of Computing CSEP 590A*

Abrassart, C., Bengio, Y., Chicoisine, G., De Marcellis, N., Warin, M, Dilhac, Gambs, S., Gautrais, V. et al. (2018). *Montreal Declaration for responsible development of Artificial Intelligence*.

Amodei, D., Olah, C., Stainhardt, J., Christiano, P., Shulman, J. and Manè, D. (2017). Concrete problems in AI safety. *ArXiv*, 1-29.

Ananny, M., (2016). Toward one ethics of algorithms: convening, observation, probability, and timeliness. *Science, technology & human values, 41*(1), 93-117.

Bendel, O. (2017). The synthetization of human voices. *AI & Society Journal of knowledge culture and communication,* 82, 737.

Valentine, M., Bihm, D., Wolf, L., Hoyme, H. E., May, P. A., Buckley, D., … Abdul-Rahman, O. A. (2017). Computer-Aided Recognition of Facial Attributes for Fetal Alcohol Spectrum Disorders. *Pediatrics*, *140*(6), e20162028. doi:10.1542/peds. 2016-2028

Stiglitz, J. E., Fitoussi, J. P., Durand, M. (2019). *Measuring what counts: the global movement for wellbeing* (2019). Amazon.

Stieglitz, J. E. (2010). *Mis-measuring our lives ed*. The New Press

Stilgoe, J. (2017). Machine learning, social learning, and the governance of self-driving cars. *Social Studies of Science*, *48*(1), 25-56.

University of Melbourne. (2018). *Biometric Mirror highlights flaws in Artificial intelligence*. Retrieved from: http://www.eurekalert.org/pub.

The IEEE Global Initiative on ethics of autonomous and intelligent systems. (2019). *Etically alligned design: a vision for promoting human wellbeing with autonomous and intelligent systems*. 1-24

Van den Hoven, J (2017). *Introduction in Van den Hooven, J. Miller, S and Pogge, T. (ed). Designing in ethics*. Cambridge University Press

Floridi, L., Cowls, J., Beltrametti, M. *et al.*. (2018). AI4People - an ethical framework for a good AI Society: opportunities, risks, principles, and Recommendations. *Minds & Machines,* 28, 689. https://doi.org/10.1007/s11023-018-9482-5

Hagendorf, T. (2019) Forbidden knowledge in machine learning reflections on the limits of research and publication. Cornell University. Retreived from: https://arxiv.org/abs/1911.08603v1

Veale, M. and Binns, R. (2017). Fairer Machine Learning in the real world: mitigating discrimination without collecting sensitive data. *Big data & society*, *4*(2), 1-17.

Vakkuri, V, and Abrahamsson, P. (2018). Key concepts of ethics of Artificial Intelligence. *Proceedings of the 2018 IEEE International Conference on engineering, technology, and innovation.* 1-6.

Vsoughi, S., Roy, D. and Sinan, A. (2018). The spread of true and false news online. *Science*. *359* 6380, 1146-51.

Searle, J. R. (1994). *The Rediscovery of the Mind*. Cambridge, MA. MIT Press.