Ana SOKOLOVSKA,* Ljupco KOCAREV*

FAIRNESS, PRIVACY AND ACCOUNTABILITY OF ALGORITHMS AND THEIR IMPACTS ON HUMAN RIGHTS

Abstract: This paper contributes to the general worldwide debate that recently emerged as response to the increased development of technologies that use artificial intelligence in people's daily lives. We tackle the issue of right to protection of individuals with regard to the processing of personal data and on the free movement of such data, the situations when that right is violated by the behavior of algorithms, and their accountability and fairness. Computer programs for autonomous decisions, in which machine learning methods are embedded can disrupt the fairness, accountability, and privacy. Several questions related to the behavior of algorithms will be addressed: Do computer programs that bear decisions are fair? Who is responsible for the decisions they are making on their own? Whether the information published on Internet is safe and being protected? To what extent, humans, that develop ICT technologies, are responsible for the unwanted consequences of their products? Drawing on different documents for privacy protection, we critically investigate the human rights to privacy at a time when growing demand for information, the rapid flow of information and the mass use of information and communication technologies in all areas of modern life require strong models for the protection of privacy and personal data of citizens, and taking responsibility for contempt, abuse and violation of human rights. The issues of human rights in the globalized world are addressed by considering a human-centered approach which could resolve some of the challenges our society faces today.

Key words: *technology, autonomous decisions, human rights*

INTRODUCTION

The Internet, the Web, the increased presence of Information and Communication Technology (ICT) in everyday life, the enlarged complexity of these technologies and the new opportunities they offer, have raised a number of questions about fairness, privacy and accountability of computer algorithms. Who is responsible for the decisions that algorithms are making on their own? Are personal data "attached" on internet safe? Who is responsible when the data is lost or when they contain errors? How can we protect ourselves from data abuse?

* Macedonian Academy of Sciences and Arts, Skopje, Macedonia

Computers have the potential to help address some of the biggest challenges that society faces. Advances in ICT technologies have opened up new opportunities for progress in many areas. Smart vehicles may save hundreds of thousands of lives every year worldwide, and increase mobility for the elderly and those with disabilities. Smart buildings may save energy and reduce carbon emissions. Precision medicine may extend life and increase quality of life. Smarter government may serve citizens more quickly and precisely. These are just a few of the potential benefits if the technologies are developed with an eye to their benefits and with careful consideration of their risks and challenges.

ICT technologies become ubiquitously embedded in our economies and societies, bringing both benefits and challenges. As ICT technologies move toward broader deployment, technical experts, policy analysts, and ethicists have raised concerns about unintended and undesired consequences of their widespread adoption, in particular regarding their decision-making abilities. Experts forecast that rapid progress in the field of specialized artificial intelligence will continue, and that machines will reach and exceed human performance on more and more tasks. The use of algorithms for consequential decisions about people, often replacing decisions made by human-driven bureaucratic processes, leads to concerns regarding how to ensure justice, fairness, and accountability of the algorithms.

However, the Internet, the Web and the ICT technologies have major shortcomings and we, as a civilization, are facing enormous challenges in addressing these shortcomings. Without diminishing all the benefits that the ICT technologies brings with us, this paper analyzes their drawbacks related to the increased information flow as a result of using the new technical and technological achievements in the globalized world, and their impact on human rights and personal data.

Every search we type into Google, every "like" on Facebook, everything we do, both on and offline is stored and analyzed. While each piece of such information is too weak to produce a reliable prediction, when tens, hundreds, or thousands of individual data are combined, the resulting predictions become more accurate. Recently, scholars have shown that easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a psychological profile of an individual (including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender [1]). These psychological profiles can be used in a political campaign for delivering targeted and highly persuasive messages to people on social media. Cambridge Analytica is a political tech company that creates psychological profiles by using publicly available data from a range of different sources, and exploiting them for political purposes. According to Alexander Nix (CEO of Cambridge Analytica), there are "somewhere close to four or five thousand data points on every individual" in US resulting in models of "personality of every adult across the United States, some 230 million people", New York Times, 20 Nov, 2016 [2]. Epstein and Robertson have presented evidence from five experiments in two countries (US and India) suggesting the power and robustness of the search engine manipulation effect (SEME) such that biased search

rankings, which can be masked so that people are not aware of the manipulation, can shift the voting preferences of undecided voters [3].

Data manipulation and unauthorized collection and use of personal data (for creating sophisticated models of user's personalities), raise ethical and privacy issues. By combining advances in Law and Computer Science, we show that recent development of ICT technologies and powerful artificial intelligence algorithms, go beyond the existing legal framework, so it is necessary to build new legal systems / models that will focus on the fairness and responsibility of computer programs that increasingly make own decisions, and on the privacy of citizens by regulating the unauthorized use of their personal data. Moreover, this paper shows that the contemporary concepts of fairness, accountability and privacy developed in computer science offer new and unexpected solutions that may need to be used by policy makers.

AUTOMATED DECISION MAKING ALGORITHMS: CHALLENGES

When Uber's self-driving car ran a red light in San Francisco, or when Google's photo app labeled images of black people as gorillas or when the Massachusetts Registry of Motor Vehicles' facial-recognition algorithm mistakenly tagged someone as a criminal and revoked their driver's license, are all examples of algorithms that made wrong decisions and did not fulfill their obligation well. The enormous growth of artificial intelligence (AI) has led to an explosion in the number of decision-making algorithms which are now more responsive to various requests, not only by answering simple questions, but through automated decision-making processes.

The dilemma we set is figuring out what to do about these problematic algorithmic outcomes. Who is responsible, or to what extent and for how long the people who have developed IT technologies are responsible for the unwanted consequences of their products? Many researchers and academics are actively exploring how to increase algorithmic accountability. However, accountability mechanisms and standards that govern and regulate decision-making processes are not keeping with new technologies and the development of computer science. The current framework of responsibility and fairness is not well adapted for situations in which a potentially wrong or unjustified decision stems from a computer.

Automated decision making algorithms raise significant challenges and numerous dilemmas, not only by policy makers, but also for society as a whole, on how to protect fundamental rights and human dignity, in terms of rapidly changing technology, especially the right to protection of personal data as a consequence of the emergence of a large number of social networks, online accounts etc., where users leave their data online that can easily be abused. HiQ Labs (https://www.hiqlabs.com/), the data gatherer company, has been processing publicly available data from LinkedIn (LI) and using it to train AI models. LI had taken technological steps to protect LI members' ability to control the information they make available on LinkedIn, and prevent HiQ from continued scraping. Further attempts to circumvent such protections would be a violation of the Computer Fraud and

Abuse Act [4], but HiQ' lawyer argues that HiQ is scraping only publicly available data: never sought privately stored information, which is confirmed by a court's ruling. Where does our data privacy end, and where does our data publicity start? Where is the line between private data and publicly available data? The unauthorized collection, processing, and usage of personal data is a constant generator of human rights violations and affects the whole world.

LAW FRAMEWORK FOR FAIRNESS, PRIVACY AND ACCOUNTABILITY

Several EU documents, including the Charter of Fundamental Rights of the European Union, the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), and the Treaty on the Functioning of the European Union, contain general provisions for protection of different human rights, especially governing accountability for the violation of the human rights, and protection of the privacy of the people.

Perhaps the longest debate on human rights of people relates to the right to data privacy. Many people "attach" some of their personal data online in order to enable online use of specific services. According to the legal regulations, personal data will not be processed for any other purpose, except for the purpose for which they were collected. But when all the information is found online, the question arises as to how to protect ourselves from misusing our data? People are often unaware that protection of their rights is envisaged. In continuation, we would rather retain to analyze the right to protect people's data, at a time when almost all of our life is available on the Internet. Due to rapid technological development and globalization, new challenges have emerged for the protection of personal data.

In January 2012, the European Commission proposed a comprehensive reform of the rules for the protection of personal data in the European Union (EU). On 4 May 2016, the official texts of the Regulation and the Directive are published in the Official Journal of the European Union in all official languages [5]. The EU Member States need to put them in place, i. e. to harmonize their national laws. The protection of personal data should be carried out in accordance with the principles for the protection of personal data contained in Directive 95/46 / EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and the free movement of such data. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data has to be applied from 25 May 2018. Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data will replace Directive 95/46 / EC of 1995, and General Data Protection Regulation (GDPR) is directly applicable. This means that they apply like national legislation.

The text of the GDPR makes it clear that the EU intends to have greater extraterritorial effect. The GDPR anticipates effective protection of data subjects' rights

in a digitalized and globalized world, while at the same time allowing the processing of personal data, including sensitive data, for scientific research. GDPR sets an important precedent: its success, or failure, will have repercussions that extend well beyond Europe. The GDPR's objective is to establish a uniform legal framework for the protection of personal data in the EU member states and equal competitive conditions for the processing of personal data. To help find a common solution to the danger of personal data abuse, not only from humans, but also from the machines which are aware of and able to act upon its surroundings and which can make decisions.

The regulation expands the definition of personal data such that data privacy will encompass other factors that could be used to identify an individual, such as their genetic, mental, economic, cultural or social identity. The Regulation will require that personal data held must be documented and include where it came from and with whom it is shared. The definition of personal data will become broader, bringing more data into the regulated perimeter, recalling that it is not easy at all to give a definition of personal data at a time of rapid development of technology and the need to protect more and more personal data of people.

Activity on social media may be instant, but the unintended consequences for children when they post something online can last beyond childhood. The Regulation will bring in special protection for children's personal data. If information is collected about children (anyone under 16 years old) then parental consent will be required in order to lawfully process their data. The consent document must be laid out in simple terms, and it is likely that the consent will be required to have an expiry date. Where the consent is for processing a child's data the privacy notice and the consent must be written in language a child can understand.

There are also potentially significant new rights for individuals, including the right to erasure ("right to be forgotten") and the right to data portability. Data portability or right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided. This is an enhanced form of subject access where organizations, businesses and institutions have to provide the requested data electronically and in a commonly used format. The Regulation also requires that data subjects should have the "right to be forgotten", or an obligation for the data controller to delete information as soon as it is no longer needed for processing. The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay.

What is different from the Directive is that under the Directive the vast majority of privacy regulations applied directly only to entities established in the EU or that used equipment in the EU. In contrary, GDPR would apply to any business, regardless of which region of the world is, that offer goods or services (even for free) to individuals in the EU or that monitors individuals located in the EU. For all the demands that it will make on resources through the preparation and implementation periods, GDPR should be still be seen as a positive step

for businesses. If a business is not in the EU, they will still have to comply with the Regulation. Non-EU controllers and processors who deal with EU subjects' personal data must comply with the new Regulation. Although enforcing regulation beyond EU borders will be a challenge, those providing products or services to EU customers, or processing their data, will face sanction under the Regulation if an incident is reported. In Germany, for example, the government has drafted a law that would fine social networks up to 50 million euro for failing to remove fake news or hate speech.

The GDPR's Article 22 establishes the right of individuals not to be subject to an automated decision-making process where those decisions have "a legal effect" or "a similar, significant effect" on the individual. Whether these changes are good or not, we will see over time. As an immediate next step, organizations must ascertain whether they have adequate resources and expertise in such key areas as finance, information technology, compliance, risk, legal and IT service management for preparation and implementation of the proposed changes.

Focusing on the European Union's perspective on the overall situation and in particular on the issue of fairness, privacy, and accountability of algorithms, it is good to mention that the legislation also paves the way for third party inspections of algorithms or 'algorithm audits'. If implemented properly, the algorithm audits supported by the GDPR could play a critical role in making algorithms less discriminatory and more accountable. Many researchers have proposed several potential methods to address algorithmic accountability: front-end and back-end process. The front-end method involves ensuring certain values are encoded and implemented in the algorithmic models that tech companies build. For example, tech companies could ensure that concerns of discrimination and fairness are part of the algorithmic process. On the backend, you could imagine that developers build the systems and deploy them without being totally sure how they will behave, and unable to anticipate the potential adverse outcomes they might generate. What you would do is to build the system, feed it a bunch of examples, and see how it behaves by analyzing how the system operates based on a variety of inputs/examples. The judicial decisions of the European Court of Human Rights have been predicted to 79% accuracy using an artificial intelligence by automatically analyzing case text using a machine learning algorithm; and we think we can all agree that the algorithms are only going to become more prevalent and powerful. It is time academics, technologists and other stakeholders to determine a concrete process to hold algorithms and the tech companies behind them accountable, and that must be done in the collaboration between computer scientists and lawmakers.

SOLUTIONS FROM COMPUTER SCIENCE

This section addresses procedural regularity and fairness of algorithms. The enormous growth of artificial intelligence has led to an explosion in the number of decision-making algorithms having the ability to automatically learn and improve from experience without being explicitly programmed. Although these

algorithms are machine-made, the result could still be biased or unfair. Moreover, the automated decision-making algorithms should fulfil the requirements for procedural regularity meaning that: (1) the same policy or rule was used to render each decision; (2) the decision policy was fully specified (and this choice of policy was recorded reliably) before the particulars of decision subjects were known, reducing the ability to design the process to disadvantage a particular individual, (3) each decision is reproducible from the specified decision policy and the inputs for that decision; and (4) if a decision requires any randomly chosen inputs, those inputs are beyond the control of any interested party.

One of the basic premises of an automated decision-making process is that the process should be known to individuals (decision subjects). Even when a part of the process is secret (for example, the process is not fully disclosed as it is protected intellectual property), the automated decisions should be reached following requirements of procedural regularity. We now describe three computer science methods that can provide accountability for procedural regularity: commitment schemes, zero-knowledge proofs, and fair random choices. These methods can guarantee that the decision-making process satisfies the requirements for procedural regularity even when the parts of the process and/or the data input to the process are secret.

A commitment scheme is a cryptographic algorithm that allows person to commit to a chosen value (or chosen statement) while keeping it hidden to others, with the ability to reveal the committed value later. The cryptographic commitment is a digital equivalent of a locked box held by a third party. This locked box does not reveal anything about the commitment contained in it until the key for the locked box is released so the third party can open it. It is possible to compute a commitment for any digital object (e. g., a file, a document, the contents of a search engine's index at a particular time, or any string of bytes). Two basic properties are essential to any commitment scheme: the binding property — having given away the box, the committer cannot anymore change what is inside, and the hiding property — the third party can tell what is inside only when the committer provides the key. Commitments can ensured that the computer systems employ the same decision policy for each of many decisions, and that implemented rules are fully determined at a specific moment in time.

A zero-knowledge proof is a method by which one party (the prover) can prove to another party (the verifier) that a given statement is true, without conveying any information apart from the fact that the statement is indeed true. It satisfies three properties: (1) Completeness: if the statement is true, the honest verifier (that is, one following the protocol properly) will be convinced of this fact by an honest prover; (2) Soundness: if the statement is false, no cheating prover can convince the honest verifier that it is true, except with some small probability; and (3) Zero-knowledge: if the statement is true, no cheating verifier learns anything other than the fact that the statement is true. If a computer system (an automated decision maker) commits to (1) the specific policy, (2) the inputs used for a particular decision based on the policy, and (3) the outcome which is the result of the application of the policy to the inputs, a non-interactive zero-knowledge proof can prove

that these values correspond to each other (without revealing the policy and the input used). For a challenged outcome, a court can force the legal entity, responsible for designing, developing and deploying the decision-making, to reveal the actual policy and input used and, thus, the court can verify the published commitments, providing digital evidence that the computer system was honest about its announced decision. Moreover, by employing a commitment to the same policy in decisions for multiple decision subjects, the computer system can demonstrate that it provides a consistent policy comprehensively to all subjects.

An automated decision process very often contains random choices; in this case, the fairness of the randomness used in computer systems should be verifiable. Random numbers can be computed in a deterministic, pseudorandom way, allowing the computer system that makes random choices to be made fully reproducible and reviewable. Pseudorandom sequences typically exhibit statistical randomness while being generated by an entirely deterministic algorithm, often a computer program or subroutine, which in most cases takes random bits as input (the seed of the generator). If pseudo-randomness is used, the automated decision maker has to be prevented from tampering with the seed value, as it fully determines all random data accessed by the program implementing the decision policy. The simplest prevention should involve a decision subject entering a short random number as part of the input for their decision. The automated decision-maker should generate a seed value using a combination of (1) a random value from a trusted third-party, (2) a random value chosen by the decision-maker, and (3) a participant or decision-specific identifier that cannot be changed or controlled by the decision-maker.

Methods described so far ensure that automated decisions are reached following requirements of procedural regularity. The algorithms generated through machine learning may turn out to be discriminatory. First, machine learning models can be discriminatory if the algorithms are trained on historical examples that reflect past prejudice or implicit bias, or on data that offer a statistically distorted picture of groups comprising the overall population. Second, machine learning models can be discriminatory through feature selection, that is, through the choice of inputs. Three types of choices about inputs could be of concern: (1) using membership in a protected class directly as an input; (2) considering an insufficiently rich set of factors to assess members of protected class with the same degree of accuracy as non-members; and (3) relying on factors that happen to serve as proxies for class membership. Third, machine learning models can be designed so that intentional discrimination is hidden as one of the above-mentioned forms of unintentional discrimination. Hardt [6] and Dwork et al. [7] propose a "catalog of discriminatory evils", where each entry is a form of intentional discrimination that is increasingly more difficult to detect.

We now describe how computer science techniques may be used to avoid outcomes of decision-making processes that could be considered discriminatory. Computer scientists have recently provided various definitions of fairness in machine learning. The concept of fairness is captured by the principle that similarly situated people are given similar treatment: any two individuals who are similar

with respect to a particular task should be classified similarly. Thus, a fair algorithm (classifier) will give similar participants a similar probability of receiving each possible outcome. This is an individual-based fairness introduced by Dwork et al. [7] which is centered around the notion of a task-specific similarity metric describing the extent to which pairs of individuals should be regarded as similar for the classification task at hand. The metric is assumed to be public, open to discussion and continual refinement, and can even be externally imposed, for example, by a regulatory body, or externally proposed, by a civil rights organization.

The concern that machine learning algorithms can be discriminatory through feature selection has been recently addressed with the concept of group fairness. The key question renders to how to use a sensitive attribute such as gender or race to maximize fairness and accuracy, assuming that it is legal and ethical. Group fairness has a variety of definitions, including conditions of statistical parity, class balance and calibration. In contrast to individual fairness, these conditions constrain, in various ways, the dependence of the model (classifier) on the sensitive attributes. Recently, Dwork et al. [8] provide a simple and efficient decoupling technique for addressing sensitive attributes, or more generally, the problem of having too little data on any one group. They proved that the technique can be added on top of any black-box machine learning algorithm. A different way to define fairness is to say that an algorithm's outcome does not allow predicting whether the subject was a member of a protected group or not. Thus, fairness can be seen as a form of information hiding requirement similar to privacy. Indeed, Dwork et al. [7] have observed that the definition of fairness in [9] is a generalization of the notion of differential privacy [10].

CONCLUSIONS

Recent advances in AI and machine learning have enabled computers interpret and analyze data automatically, making them active subjects in the knowledge discovery and decision making processes. The majority of the data now being generated by electronic devices and computers is for consumption by other computers. This, in turn, has scaled decision-making processes: it is becoming increasingly common for a computer to make decisions. The shift from humans towards automated decision-making processes has raised a multitude of issues ranging from the costs of incorrect decisions to ethical and privacy issues. Not surprisingly, these issues have recently been addresses by scholars in several disciplines, including social science, law, public policy, and computer science.

In April 2016 the European Parliament adopted a set of comprehensive regulations for the collection, storage and use of personal information, the GDPR. This regulation has been described as a "Copernican Revolution" in data protection law, "seeking to shift its focus away from paper-based, bureaucratic requirements and towards compliance in practice, harmonization of the law, and individual empowerment" [11]. As it stands, transparent and accountable automated decision-making is not yet guaranteed by the GDPR, nor is the right to explanation of algorithmic decisions. At best, data subjects will be granted a 'right to be

informed' about the existence of automated decision-making and system functionality. However, if we cannot trace the processes of an AI and ascertain how it reached a certain decision, we must at the very least make transparent all elements of the process that were controlled by humans. As we argued above, we must demand that AI is programmed according to the most rigorous transparency, fundamental values or basic ethical and societal principles.

As algorithms are increasingly used for decisions, the social, ethical, and legal values converted in these decision-making processes are the subject of increasing study, with fairness being the main concern. Different notions of fairness for both individual and group fairness have been suggested, showing how to design fair algorithms when it is ethical and legal to use a sensitive attribute (such as gender or race) in machine learning systems. In general, scientists approach trust and assurance of computer systems differently than policymakers, seeking strong formal guarantees or trustworthy digital evidence that a system works as it is intended to or complies with a rule or policy objective rather than simple assurances that a piece of software acts in a certain way [12].

We need a system that allows citizens (data subjects) to maintain their privacy and control over their own data even as more data are produced every day. Policy (law) makers should seek to learn more about the implementation of automated decision-making systems in order to ensure that existing laws and legal frameworks are effectively implemented in response to the challenges posed by automated decision making in the various spheres of their applications. Machines are not humans, and probably will never be. A solution in near future may be giving machines a degree of personhood (legal status to "electronic persons") — much in the same way that corporations are legally regarded as persons — so that companies can be held accountable for the actions they take on their own.

REFERENCES

[1] W. Youyou, M. Kosinski, and D. Stillwell: "Computer-based personality judgments are more accurate than those made by humans" Proceedings of the National Academy of Sciences (PNAS), vol. 112 no. 4, 2015, p. 1036–1040.
[2] McKenzie Funk: "The Secret Agenda of a Facebook Quiz" New York Times, 2016 https://www.nytimes.com/2016/11/20/opinion/the-secret-agenda-of-a-facebook-quiz.html?mcubz=0
[3] R. Epstein, and R. E. Robertson: "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections" Proceedings of the National Academy of Sciences vol. 112, 2015, p. 4512–4521.
[4] https://energy.gov/sites/prod/files/cioprod/documents/ComputerFraud-AbuseAct.pdf
[5] http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf
[6] M. Hardt: "A Study of Privacy and Fairness in Sensitive Data Analysis" Dissertation. Princeton University, 2011.
[7] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel: "Fairness through awareness" ITCS '12 Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 2012, p. 214–226.

[8] C. Dwork, N. Immorlica, A. T. Kalai, and M. Leiserson: "Decoupled classifiers for fair and efficient machine learning" 2017, arXiv: 1707.06613 [cs. LG]
[9] C. Dwork: "Differential Privacy" Proceedings of the 33rd International Colloquium on Automata, Languages, and Programming, 2006, p. 1–12.
[10] C. Dwork, F. McSherry, K. Nissim, and A. Smith: "Calibrating noise to sensitivity in private data analysis" In Proc. 3rd TCC, 2006, p. 265–284.
[11] C. Kuner: "The European Commission's proposed data protection regulation: A Copernican revolution in European data protection law" Bloomberg BNA Privacy and Security, vol. 6, 2012, p. 115.
[12] J. A. Kroll, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G Robinson, and H. Yu: "Accountable algorithms" U. Pa. L. Rev., vol. 165, 2016, p. 633.