Ethics in AI and Automated Decisions

Filippo Bianchini
Studio Legale Bianchini
Perugia, Italy
e-mail: info@bianchini.legal

***Abstract*—The use of Artificial Intelligence is increasingly pervasive in our lives and this poses technological, legal and ethical problems; the definition of the term itself is not unique and evokes different contexts. The protection of human dignity and the safeguarding of fundamental rights and freedoms pass through a correct information on the use both of personal data and of algorithms for processing them, which must be knowable and their results contestable. The ethical approach to these issues is particularly relevant where it does not exclude the law but manages to overcome it by ensuring that it can keep pace with a tumultuous technological development. A proposed solution is that of an assessment that considers the ethical and social impact as well as a complete possibility of access by the data subject.**

Keywords-artificial intelligence; ethics; personal data; GDPR; data protection; automated decision; profiling.

I. INTRODUCTION

John McCarthy first coined the term "Artificial Intelligence" (AI) in 1955 when he invited a group of researchers from a variety of disciplines including language simulation, neuron nets, complexity theory and more to a summer workshop called the "Dartmouth Summer Research Project on Artificial Intelligence" to discuss what would ultimately become the field of AI [1].

Artificial intelligence systems are becoming increasingly common in everyday life, strongly influencing the habits and behaviour of both individuals and communities. Nowadays, we increasingly speak of "datafication" (the trend to turn a phenomenon into a quantitative form, i.e., into data) [2], a phenomenon that enables us to analyse and store enormous amounts of data, thus setting the stage for the *Big Data* economy. This notion has been traditionally outlined by D. Laney using the so-called 3V-model, i.e., volume, velocity and variety [3]. A fourth "V" can be identified in veracity, or truthfulness. In turn, these combined features generate a fifth one: value, profit.

Datafication makes it possible to correlate the collected data for profiling purposes: on the one hand, this enables the profiling controller to tap their informative potential, with benefits in terms of streamlining and savings; on the other hand, it seriously threatens the rights and freedoms of the individual, with potential repercussions not only on their behaviour but also on their knowledge, choices and feelings. In addition, new "inferred" data are generated from the first batch, and they too require protection.

Major concerns arise involving complex algorithms, which can process considerable amounts of data and are therefore increasingly used to dig into the personality of the individual and lay bare its innermost recesses, thus enabling their users to make potentially impactful decisions on the data subject. Suffice it to think about the negative legal and personal implications that the processing of incorrect or outdated data can have for a given individual [4]: this may well trigger a *garbage in, garbage out* mechanism, whereby the processing of poor data inevitably leads to misleading results.

Today, modern dictionary definitions focus on AI being a sub-field of computer science and how machines can imitate human intelligence (being human-like rather than becoming human). The English Oxford Living Dictionary gives this definition: "The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages" [5]. Merriam-Webster defines artificial intelligence this way: "1. A branch of computer science dealing with the simulation of intelligent behaviour in computers. 2. The capability of a machine to imitate intelligent human behaviour" [6]. The Encyclopedia Britannica states, "artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings" [7] Intelligent beings are those that can adapt to changing circumstances.

The Council of Europe offers the following definition of AI: "A set of sciences, theories and techniques whose purpose is to reproduce by a machine the cognitive abilities of a human being. Current developments aim, for instance, to be able to entrust a machine with complex tasks previously delegated to a human" [8]; while the European Commission gives this one: "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-based systems can be purely software-based, acting in the virtual world (e.g., voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g., advanced robots, autonomous cars, drones or Internet of Things applications)" [9].

Arend Hintze, an assistant professor of integrative biology and computer science and engineering at Michigan State University, categorizes AI into four types, from the

 ISBN: 978-1-61208-721-4

kind of AI systems that exist today to sentient systems, which do not yet exist [10]. His categories are as follows:

Type 1: Reactive machines. An example is Deep Blue, the IBM chess program that beat Garry Kasparov in the 1990s. Deep Blue can identify pieces on the chess board and make predictions, but it has no memory and cannot use past experiences to inform future ones: it analyses possible moves - its own and its opponent - and chooses the most strategic move.

Type 2: Limited memory. These AI systems can use past experiences to inform future decisions. Some of the decision-making functions in self-driving cars are designed this way. Observations inform actions happening in the not-so-distant future, such as a car changing lanes. These observations are not stored permanently.

Type 3: Theory of mind. This psychology term refers to the understanding that others have their own beliefs, desires and intentions that impact the decisions they make. This kind of AI does not yet exist.

Type 4: Self-awareness. In this category, AI systems have a sense of self, have consciousness. Machines with self-awareness understand their current state and can use the information to infer what others are feeling. This type of AI does not yet exist.

The rest of this paper is organized as follows. Section II describes the contribution of the Convention 108+ and the Regulation (EU) 2016/679. Section III describes the contribution of the High-Level Expert Group on AI. Section IV describes the contribution of the Organization for Economic Co-operation and Development. Section V explores some practical uses of AI. The conclusion closes the article.

II. THE CONTRIBUTION OF THE CONVENTION 108 AND THE REGULATION (EU) 2016/679

Earlier this January, the Consultative Committee of the Convention for the Protection of Individuals with regard to the Processing of Personal Data (Convention 108 [11]) has published its "Guidelines on Artificial Intelligence and Data Protection" [12]. In this document the Committee confirms that "The protection of human dignity and safeguarding of human rights and fundamental freedoms, in particular the right to the protection of personal data, are essential when developing and adopting AI applications that may have consequences on individuals and society". The Committee also acknowledges that the development of AI should be based on the principles of the Convention 108, signed in Strasbourg on 28 January 1981 and recently modernised as Convention 108+

While the core principles contained in Convention 108 have stood the test of time and its technologically-neutral, principle-based approach constitutes an undeniable strength, the Council of Europe considered necessary to modernize its landmark instrument.

The modernization of Convention 108 pursued two main objectives: to deal with challenges resulting from the use of new information and communication technologies and to strengthen the Convention's effective implementation [13].

A. Information and Access to Personal Data

The principles enumerated in the Convention are the basis of the current legislation on the protection of personal data and, in particular, of the "Regulation (EU) 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, and repealing Directive 95/46/EC" (General Data Protection Regulation, GDPR [14]), as it can be seen by a quick comparison of Article 5 in both texts (see Table I below).

TABLE I. COMPARISION BETWEEN CONVENTION 108 AND GDPR

Convention 108+	GDPR
3. Personal data undergoing processing shall be processed lawfully.	1. Personal data shall be: (a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
4. Personal data undergoing processing shall be: a. processed fairly and in a transparent manner;	
b. collected for explicit, specified and legitimate purposes and not processed in a way incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes is, subject to appropriate safeguards, compatible with those purposes;	(b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');
c. adequate, relevant and not excessive in relation to the purposes for which they are processed;	(c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
d. accurate and, where necessary, kept up to date;	(d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');
e. preserved in a form which permits identification of data subjects for no longer than is necessary for the purposes for which those data are processed.	(e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
	(f) processed in a manner that ensures appropriate security of the

	personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

Articles 13(2) and 14(2) of the GDPR say that "the controller shall provide the data subject with the following information necessary to ensure fair and transparent processing in respect of the data subject: (g) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject". However, it is not entirely clear why the safeguards were limited to fully automated decision-making processes.

Further protection is granted by the right of access under Article 15(1) GDPR, which provides for the possibility of obtaining information about "(h) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject".

B. Automated Individual Decision-Making, including Profiling

Moreover, Article 9(1) of the Convention 108+ has that "Every individual shall have a right: a. not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration" while Article 22(1) of the GDPR says that "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".

In my opinion, the expression "right not to be subject" must be considered addressed to the data controller, being automatically applicable (by default) except in the case of exceptions foreseen in paragraph 2. This way, AI applications should always allow meaningful control by data subjects over the data processing and related effects on individuals and on society.

III. THE CONTRIBUTION OF THE HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG)

The High-Level Expert Group on Artificial Intelligence (AI HLEG) is an independent expert group that was set up by the European Commission in June 2018. The group has been set up in order to support the implementation of the European strategy on Artificial Intelligence, including the elaboration of recommendations on future-related policy development and on ethical, legal and societal issues related to AI [15]. Moreover, the AI HLEG will serve as the steering group for the European AI Alliance's work, interact with other initiatives, help stimulate a multi-stakeholder dialogue, gather participants' views and reflect them in its analysis and reports.

The group has given its definition of AI: "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.

As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)" [16].

In its "Ethics guidelines for trustworthy AI" [17], the AI HLEG has found that "trustworthy AI has three components, which should be met throughout the system's entire life cycle:

1. it should be lawful, complying with all applicable laws and regulations;
2. it should be ethical, ensuring adherence to ethical principles and values; and
3. it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm".

Speaking about ethics, it is necessary to observe that "laws are not always up to speed with technological developments, can at times be out of step with ethical norms or may simply not be well suited to addressing certain issues. For AI systems to be trustworthy, they should hence also be ethical, ensuring alignment with ethical norms".

And what these norms are? The AI HLEG has specified four principles, in form of ethical imperatives:

i. Respect for human autonomy
ii. Prevention of harm
iii. Fairness
iv. Explicability

In my opinion, all these explications should be considered *by design and by default* (that is "both at the time of the determination of the means for processing and at the time of the processing itself", see Article 25 GDPR), so as to establish a real protection for the individual.

IV. THE CONTRIBUTION OF THE ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD)

On 22 May 2019 the Organization for Economic Co-operation and Development (OECD) and partner countries formally adopted the first set of intergovernmental policy guidelines on Artificial Intelligence, agreeing to uphold international standards that aim to ensure AI systems are designed to be robust, safe, fair and trustworthy [18].

The OECD's 36 member countries, along with Argentina, Brazil, Colombia, Costa Rica, Peru and Romania, signed up to the OECD Principles on Artificial Intelligence at the Organization's annual Ministerial Council Meeting that took place in Paris and was focused on "Harnessing the Digital Transition for Sustainable Development". Elaborated with guidance from an expert group formed by more than 50 members from governments, academia, business, civil society, international bodies, the tech community and trade unions, the Principles comprise five values-based principles for the responsible deployment of trustworthy AI and five recommendations for public policy and international co-operation. They aim to guide governments, organizations and individuals in designing and running AI systems in a way that puts people's best interests first and ensuring that designers and operators are held accountable for their proper functioning.

In summary, the Principles state that:

1. AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being;
2. AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society;
3. there should be transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them and can challenge outcomes;
4. AI systems must function in a robust, secure and safe way throughout their lifetimes, and potential risks should be continually assessed and managed;
5. organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

The OECD recommends that governments:

a) facilitate public and private investment in research & development to spur innovation in trustworthy AI;
b) foster accessible AI ecosystems with digital infrastructure and technologies, and mechanisms to share data and knowledge;
c) create a policy environment that will open the way to deployment of trustworthy AI systems;
d) equip people with the skills for AI and support workers to ensure a fair transition,
e) co-operate across borders and sectors to share information, develop standards and work towards responsible stewardship of AI.

V. PRACTICAL USES

As any human instrument, AI can be used in good as well as in bad ways.

A. Good Applications of AI

Artificial intelligence is beginning to be applied in the medical setting and has potential to improve workflows and errors, impacting patients and clinicians alike.

The European Commission, for instance, has a strong history of project involving AI aimed at improving people's quality of life.

Among others:

- DE-ENIGMA: using play to help autistic children recognize and express emotions.
  A humanoid robot known as Zeno helps to teach school-aged autistic children, who have additional intellectual disabilities or limited spoken communication, to express emotions. It will be able to process children's movements, vocalizations and facial expressions in order to adaptively present activities linked to emotions, and engage in feedback, support and play.
- Alfred: a virtual assistant helping older people stay active.
  The project created a virtual "butler" to which people can talk, ask questions or give commands, and developed systems to encourage older people to socialize by suggesting and managing events, to monitor their state of health, and to help them stay physically and mentally active via personalized games. It produced 25 apps, both for immediate use and to inspire developers interested in designing new services that target the needs of senior citizens.
- Bots2Rec (Robots to Re-Construction): using robots to clear asbestos and keep workers safe.
  The project is developing robots that can clear asbestos – which, when inhaled by humans in the form of fibers or dust, can cause serious lung diseases – from contaminated buildings. The robots act autonomously in a building's rooms, but an operator can also control them to perform specific tasks, with the help of a virtual representation of the site.
- MURAB (MRI and Ultrasound Robotic Assisted Biopsy): using AI to detect cancer.
  The project is developing technology that will make it possible to take more precise and effective biopsies (tissue samples) and diagnose cancer and other illnesses faster. It is creating a robot that will scan a patient's body using a combination of Magnetic Resonance Imaging (MRI) and ultrasound technology and select the right location for a biopsy.
  This will be quicker and more comfortable for patients and will have the potential to identify early-stage signs of cancer that conventional ultrasounds may not pick up as well as reduce the likelihood of false negative results.

Recent studies show that facial analysis technologies measured up to the capabilities of expert clinicians in syndrome identification. However, these technologies identified only a few disease phenotypes, limiting their role in clinical settings, where hundreds of diagnoses must be

 ISBN: 978-1-61208-721-4

considered. A group of researchers presented a facial image analysis framework that quantifies similarities to hundreds of syndromes using computer vision and deep-learning algorithms [19]. On the final experiment reflecting a real clinical setting problem, this structure achieved 91% top-10 accuracy in identifying the correct syndrome on 502 different images. The model was trained on a dataset of over 17,000 images representing more than 200 syndromes, curated through a community-driven phenotyping platform.

B. (Possible) Bad Applications of AI

There is also a dark side in facial recognition. On May 2019, San Francisco has become the first city in the United States to ban the use of facial recognition technology by the police and local government agencies. The "Stop Secret Surveillance" ordinance [20], set to take effect one month later, also requires city agencies to gain the board's approval before buying new surveillance technology and an audit of any existing surveillance tech in use by the city. The ban does not cover use of the technology by individuals or businesses.

Critics of facial recognition say the technology is not reliable enough to be in the hands of law enforcement. The American Civil Liberties Union (ACLU) is one of many civil-rights groups supporting the ordinance. Matt Cagle, a technology and civil liberties attorney at the ACLU of Northern California, said that this technology "provides government with unprecedented power to track people going about their daily lives. That's incompatible with a healthy democracy" [21].

Concerns about the technology aren't unfounded. In a study published by the MIT Media Lab earlier this year [22], researchers found facial analysis software made mistakes when identifying the gender of female or dark-skinned individuals.

All this echoes the "Correctional Offender Management Profiling for Alternative Sanctions" (COMPAS) risk assessment [23], a presentencing investigation report (PSI) – the documents that typically provide background information on offenders to sentencing courts – mainly known for the "State v. Loomis" case [24] [25].

The COMPAS was analyzed by ProPublica, a Non-Governmental Organization, which found that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk [26].

VI. CONCLUSION

Following an exploration of the different meanings of AI, the present work has described the various contributions offered by the Convention 108+, the Regulation (EU) 2016/679, the High-Level Expert Group on AI and the Organization for Economic Co-operation and Development. Subsequently it has presented some practical applications of AI, both good and (possibly) bad.

Postulating the absolute value of the human beings and the protection of their personal data as a consequent fundamental right, it is necessary to observe that the Data Protection Impact Assessment (DPIA) introduced by Article 35 GDPR can evolve into a Privacy, Ethical and Social Impact Assessment (PESIA), which takes into account not only the aforementioned data protection but also its ethical and social impact, i.e., the collective nature of the risk [27].

Furthermore, it seems appropriate to establish a full 'right to explanation', whereby the data subject is not only made aware of the rationale behind the algorithm's automated decision-making but is also given a full explanation of the outcome – and thus the specific decision taken.

Thus, the compliance with ethical norms, as well with positive ones, will have the effect of expanding the protection of natural persons with regard to the processing of personal data, notably in relation to automated individual decision-making processes, with the advantage of not necessarily having to wait for a legislative provision which may arrive too late to regulate the tumultuous technological development.

ACKNOWLEDGMENT

The Author sincerely thanks Mr Nicola Fabiano for his constant inspiration in the field of Data Protection.

REFERENCES

[1] J. McCarthy, M. L. Minsky, N. Rochester and C.E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence". Available from: http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html [retrieved: 2019-05-31]

[2] V. Mayer-Schönberger and K. Cukier, "Big Data: a revolution that transforms how we work, live, and think", Houghton Mifflin Harcourt, 2012

[3] D. Laney, "3D Data Management: Controlling Data Volume, Velocity, and Variety". Available from: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [retrieved: 2019-05-31]

[4] C. O'Neil, "Weapons of Math Destruction", Crown 2016. She is also author of the blog mathbabe.org

[5] Online, available from: https://en.oxforddictionaries.com/definition/artificial_intelligence [retrieved: 2019-05-31]

[6] Online, available from: https://www.merriam-webster.com/dictionary/artificial%20intelligence [retrieved: 2019-05-31]

[7] B.J. Copeland , "Artificial intelligence", Available from: https://www.britannica.com/technology/artificial-intelligence [retrieved: 2019-05-31]

[8] Online, available from: https://www.coe.int/en/web/human-rights-rule-of-law/artificial-intelligence/glossary [retrieved: 2019-05-31]

[9] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels, 25.4.2018 COM(2018) 237 final, online, Available from: https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-237-F1-EN-MAIN-PART-1.PDF [retrieved: 2019-05-31]

[10] A. Hintze, "Understanding the four types of AI, from reactive robots to self-aware beings", online, available from: https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616 [retrieved: 2019-05-31]

 ISBN: 978-1-61208-721-4

[11] Details of Treaty No.108, Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, online, available from: https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/108 [retrieved: 2019-05-31]

[12] Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108), "Guidelines on Artificial Intelligence and Data Protection", online, available from: https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8 [retrieved: 2019-05-31]

[13] Council of Europe, "Explanatory Report to the Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of PersonalData", online, available from: https://rm.coe.int/cets-223-explanatory-report-to-the-protocol-amending-the-convention-fo/16808ac91a [retrieved: 2019-05-31]

[14] Regulation (EU) 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, and repealing Directive 95/46/EC (General Data Protection Regulation), online, available from: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=IT#d1e1797-1-1 [retrieved: 2019-05-31]

[15] High-Level Expert Group on Artificial Intelligence, online, available from: https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence [retrieved: 2019-05-31]

[16] AI HLEG, "A Definition of AI: Main Capabilities and Disciplines", online, available from: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341 [retrieved: 2019-05-31]

[17] AI HLEG, "Ethics Guidelines for Trustworthy AI", online, available from: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477 [retrieved: 2019-05-31]

[18] OECD, "Recommendation of the Council on Artificial Intelligence", OECD/LEGAL/0449, online, available from: https://legalinstruments.oecd.org/api/print?ids=648&lang=en [retrieved: 2019-05-31]

[19] Y. Gurovich, Y. Hanani, O. Bar, N. Fleischer, D. Gelbman, L. Basel-Salmon, P. Krawitz, S.B Kamphausen, M. Zenker, L.M. Bird and K.W. Gripp, "Identifying facial phenotypes of genetic disorders using deep learning", Nature Medicine 25, 60–64 (2019), available from: https://arxiv.org/abs/1801.07637 and also https://www.nature.com/articles/s41591-018-0279-0#article-info and also available at https://www.eff.org/files/2019/05/07/leg_ver3.pdf [both retrieved: 2019-05-31]

[20] Online, available from: https://sfgov.legistar.com/View.ashx?M=F&ID=7206781&GUID=38D37061-4D87-4A94-9AB3-CB113656159A and also available from: https://www.eff.org/files/2019/05/07/leg_ver3.pdf [both retrieved: 2019-05-31]

[21] K. Conger, R. Fausset and S.F. Kovaleski, "San Francisco Bans Facial Recognition Technology", online, available from: https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html [retrieved: 2019-05-31]

[22] I.D. Raji and J. Buolamwini, "Actionable Auditing:Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products", online, available from: http://www.aies-conference.com/wp-content/uploads/2019/01/AIES-19_paper_223.pdf [retrieved: 2019-05-31]

[23] T. Brennan, D. William and E. Beate, "Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System." Criminal Justice and Behavior 36, no. 1 (January 2009), available from: http://www.northpointeinc.com/files/publications/Criminal-Justice-Behavior-COMPAS.pdf [retrieved: 2019-05-31]

[24] Supreme Court of Wisconsin, State of Wisconsin, Plaintiff–Respondent, v. Eric L. Loomis, Defendant–Appellant, 881 N.W.2d 749 (Wis. 2016), online, available from: https://caselaw.findlaw.com/wi-supreme-court/1742124.html [retrieved: 2019-05-31]

[25] "Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing", Harvard Law Review, March 2017, Vol. 130, No. 5, available from: http://harvardlawreview.org/wp-content/uploads/2017/03/1530-1537_online.pdf [[retrieved: 2019-05-31]

[26] J. Larson, S. Mattu, L. Kirchner and J. Angwin, "How We Analyzed the COMPAS Recidivism Algorithm", online, available from: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [retrieved: 2019-05-31]. For the whole story, see: 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", online, available from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [retrieved: 2019-05-31]

[27] Virt-EU project, "What's the PESIA framework?", online, available from: https://medium.com/@VIRT_EU/whats-the-pesia-framework-912bf0a12a4e [retrieved: 2019-05-31]

 ISBN: 978-1-61208-721-4