Accountability and Transparency in AI Systems: A Public Policy Perspective

R. Bahrevar and Khashayar Khorasani
Department of Electrical and Computer Engineering
Concordia University, Montreal, Quebec CANADA

- Artificial intelligence (AI) allows one to understand the ever-growing data that are being produced worldwide. Since AI systems can produce powerful models that can make sense of large amounts of data, this will stimulate the increase of investments in AI-based smart technologies in various areas. Scientific studies, industry, and job market analysts are designating significant interest in development of AI-based technologies.

- To ensure regulation of these systems has not failed to keep up with and sight of these advancements, here we address our perspective and recommendations on the barriers to ethics, security, fairness, and transparency of AI systems.

- After presenting several viewpoints regarding AI-of-Things (AIoT), we define our perspectives for a reliable and responsible AI system. Our goal is to discuss why developing a responsible AI should not be viewed as the sole responsibility of AI developers.

- Based on our detailed analysis of the state of the art, we present an accountability model for ethical AI, in which the technical aspects are as important as the ethical aspects. Our goal is to pinpoint and identify elements that are the source of deficiency and loopholes in the way of achieving a responsible AI.

- Two aspects concerning the tools required for regulation and design of an ethical AI are introduced. These two aspects can be viewed in terms of technical AI design regulation (transparency of the methodology) and ethical standards.

- Furthermore, a third aspect is introduced as the control and overviewing of AI systems. In terms of ethical aspect, it is argued that a framework for developers in terms of reminding them of their ethical responsibilities must be provided.

- People also need to have arrangements and tools for raising their voices against malicious or harmful AI systems.

- In terms of technical aspects, developers have to choose a methodology that is transparent with regards to data and must ensure that its model is not vulnerable to adversarial examples, so that it does not put citizens security and privacy at risk.

- Besides well-educated AI developers, to implement and regulate these aspects, we have recognized the need for specialized AI inspectors and legal experts. Specialized inspectors are not only mathematically well-educated, but also, they are specialized to recognize biases that are application-specific and may go unnoticed by an AI-forensic specialist that only is educated in the data mining aspects of the problem.

- The inspectors will be tasked to evaluate AI systems and grants standards in terms of maintaining safety, security, transparency, and privacy. The standardization is a path towards ethical by design AI system and can facilitate and enable the government to regulate these systems through legal experts or relevant authorities.

Abstract

Artificial intelligence allows us to understand the ever-growing data that is produced worldwide. Since AI systems can produce powerful models that can make sense of large amounts of data, this will stimulate the increase of investments in AI-based smart technologies in various areas. Scientific studies, industry, and job market analysts are designating significant interest in the development of AI-based technologies. To make sure the regulation of these systems has not failed to keep sight of these advancements, here we address our perspective and recommendations on barriers in ethics, security, fairness, and transparency for these systems.

After reviewing several papers and viewpoints regarding AIoT, we try to define our perspective of a reliable and responsible AI system. Here, we discuss why developing a responsible AI should not be viewed as the sole responsibility of AI developers. Based on the review of the literature, we present an accountability model for ethical AI, in which the technical aspect is as important as the ethical aspects. We try to pinpoint elements that are the source of deficiency and loopholes in the way of achieving a responsible AI.

1 Introduction

According to the IDC (International Data Corporation), the Global DataSphere[1] will increase from an estimate of 50 zettabytes in 2020 to 175 zettabytes in 2025 [1]. This will stimulate the increase of investments in AI-based smart technologies in various areas such as the autonomous system, recommender systems, biology, business, and politics.

The main problem with the AI systems is their learning structure, which is referred to as the black-box model [2]. These models that are commonly pre-trained will analyze the data through a complex mathematical structure and provide a concluding response as an output.

The lack of explainability of these systems is problematic in light of their accountability with regards to their resulting decision for different inputs. As a consequence, questions about ethics, transparency, privacy, security, and fairness of the decision-making process will surface, which are some of the characteristics of responsible or ethical AI [3].

For example, recommender systems may receive and process information such as a person's past website visitation behavior, search history, movies that have been watched or clicked by this person in the different platforms to advise the person of a certain product. Accessing a variety of information about a specific person is possible since there are numerous third-party companies that distribute the information between different platforms such as Amazon, Netflix, Google, and Facebook. Recommender systems may also use geolocation in smart cars' infotainment applications, which will transform them into a universal tracking device [4]. Data transparency, privacy, and security can be the main concern of infotainment technology. Here, the questions are: who are the third-party companies? Whom are they sharing this information with? Does this information reveal the user's identity or location?

Furthermore, there are other types of a recommender system, where the AI system will recommend hiring or firing of an individual [5, 6, 7], investing in a business [8], or granting parole [9, 10]. How can we make sure that the AI developer has trained a model that is not biased toward a company, ethnicity, or nationality? Can the AI system respect the privacy of individuals? Can an AI system learn to be considerate?

An AI developer can be one of the major causes of these issues by intentionally or unintentionally neglecting factors such as training, choosing proper design models, standard data selection and preparation, ethical constraints, or selection of representative inputs for their system.

Numerous studies, major tech developers, the global research and advisory firms have consensus that the next step for the advancement of AI-based technologies, is to achieve a responsible, or in another term, ethical AI [6, 11, 12]. But can an AI system learn to be fair or have consciousness?

The ideal situation is to achieve an "Ethical by Design AI" [6]. Based on our survey of the literature, we identified that there are two key viewpoints on the design of ethical AI. The first issue is regarded as the ethical constraint, which encompasses all the ethical, policy aspects, and concerns that must be constantly addressed with regards to currents standards on the ethical grounds [6, 12]. The second issue is referred to as the transparency of the methodology [2, 13, 14]. In this paper, we try to address achieving ethical AI through these viewpoints.

In terms of ethical or moral grounds, the issues can be viewed in terms of what the general public, and we as human beings believe is morally right. What steps should we take to be proactive with regards to undesirable events such as discrimination, loss of private information, defamation, and harm to the general public or the clients that are dealing with such systems? To deal with this, it is advised to use tools or

[1] The Global DataSphere: It is the summation of all the data, no matter created, captured, or replicated in any given year.

practices such as risk assessment [6], mandating offline development [15], and data regulation [16] to come up with the policies which protect the privacy and serves the interest of the people. In terms of technical aspects, it is essential to readjust our policies based on studies that are concerned with the transparency of the AI systems [17, 18, 19].

1.1 Ongoing Policy Concerns and Our Challenge

Major companies such as IBM, Google, Microsoft, NVIDIA, Hitachi, and ETRI are already investing in responsible AI. Governmental associations are also promoting ethical AI by increasing investment in AI-based technology as well as giving special attention to AI-centric education. In March 2017, the Quebec government, through the University of Montreal, set up a steering committee for the development of Quebec's artificial intelligence industrial and scientific cluster. The committee had the task to focus on the mathematical literacy of in training of AI students as well as promote responsible AI [20].

Institutes and agencies such as the European commission, European Union Agency for Cyber Security (ENISA), and Defense Advanced Research Projects Agency (DARPA) are also trying to pinpoint the possible elements that are necessary for the development of standards for ethical AI. In the following, we pinpoint some of the criteria introduced by [18]:

- "Cyber-hygienes" must be promoted. It means that cyber threats will not solely affect the technology but also the citizens.
- Raising awareness of citizens and business owners against cyber-attacks.
- Decreasing third-party dependency of the AI systems.
- Performing regular analysis and maintaining a "market observatory."
- Promoting cyber-security certification.
- Promoting international collaboration.
- Promoting regular consulting with standardization organizations.
- Thinking of flexible solutions, so the industry remains ahead of the threats.
- Promoting a high level of safety assurance for Information and Communication Technology (ICT) products.

Several researchers across the world are also focusing on developing an ethical framework for AI systems. In [6], the authors provide their framework regarding ethical AI. This study includes promoting proactive measures such as risk assessment and impact assessment to evaluate the effect of AI systems on society. The AI-related specialized legal platforms, promotion of ethical guidelines, and ethical standards are the additional concerns of this study and the other AI ethics framework researchers [6, 12, 21].

We have to make sure AI systems are beneficial, will not jeopardize people's civil rights, and can not be easily manipulated for malicious purposes.

In this paper, our recommendation specifically directed at concerns that include cybersecurity as well as the legal issues related to the AI systems. Our challenge is to present a structure that complements the mentioned policies and help to integrate the characteristics that are the basis of the development of ethical

AI. Our presented policy aims to lead to the accountability of the AI systems. We want to discuss when one talks about the accountably in AI, should the target be only the developers?

2 Axis of a reliable AI

Based on the review of the literature on the topics that consist of technological deficiency and ethical issues for the AI systems, we present three important factors that will lead to accountability in AI. In this section, we introduce these factors, and in the next section, we utilize them to establish the accountability model of a reliable AI.

2.1 Ethical Principal Regarding AI

The majority of policies regarding the ethical principal of AI systems have consensus on the accountability, transparency, contestability, security, fairness, privacy, and the net benefit of these systems [6, 12, 22]. Achieving these qualities will form the structure of an ethical AI [12]. Specialized jurisdiction system [6], risk [23], and impact assessment [24], data privacy [16, 25], data transparency [26], raising awareness through industry and academia [12], and diverse workforce [27] are some of the measures that are foreseen. The measures will assist to improve AI systems in such a way that they encompass the characteristics of an ethical AI.

An extensive survey of moral values and principles for AI systems can be found in [12]. Moreover, the main body and new recommendations of this paper are directed in the technological aspects of these systems. However, the moral code of conduct must be a necessity for every AI developer.

The developers should be informed of the necessary consideration that they should make to avoid misrepresentations that would lead to a possible bias by their product. The beneficence of ethical guidelines is that it aims to reduce the intentional or unintentional insertion of bias into the AI systems.

2.2 Transparency of Methodology

The issues regarding the AI systems necessitate the need for constraints that should be considered in the data preparation, training, and decision-making process of an AI system. As mentioned, ethics, privacy, security, transparency, and fairness are characteristics of a responsible AI. We have to promote tools that enforce the transparency of the AI systems. The selected tools should be an answer to the following questions:

- Data selection and preparation: does the AI system utilizes data that are unbiased and demonstrative of the related application?

Aside from data being representative of the population related to the application, we should be able to evaluate the logic behind the selected inputs of an AI system and the label that they will be assigned [28], which may seem extreme for most AI applications. However, considering AI applications such as surveillance and autonomous vehicle, safety and privacy of public should be of higher priority.

Without transparency in data selection, can the customers and standard providers trust these systems to make the right decisions in critical situations? On the other hand, increased transparency will also jeopardize the security of these systems since the developers are sharing sensitive data that can be used for attacks against these systems. Furthermore, one should seek to find a trade-off in terms of security, robustness, transparency, and privacy.

In input selection, purpose, relevancy, and necessity must also be considered in most AI applications. It will help us to avoid collecting unnecessary inputs that may cause the invasion of privacy. Knowledge of data selection will also provide the extra layer of confidence for people that assess the reliability of AI systems, especially in a critical application.

- Explainability: what is the reasoning behind generating a certain answer? Possible Solution: The issues regarding the AI systems necessitate the need for constraints that should be considered in the decision-making process and data preparation of an AI system. Achieving transparency is one of the means for evaluating the resulting decision of AI systems. One of the available tools for designing such systems is explainable AI (also referred to as interpretable AI). The explainable AI methodologies are just the first of many elements for achieving a transparent AI. It is an improvement tool that will help to improve transparency, reduce biases, and increase trust toward the performance of the AI system [13, 14].

  The added transparency by implementing the explainable AI methodologies can even help to resolve the legal arguments against the AI systems. There have been many cases where an AI system breached the privacy of the people. These cases resulted in the argument to be brought upon the court of law, where the AI system affected a career [5], privacy [29], or even a person's freedom [10]. An application that can have a non-negligible effect on an individual, civilians, organization, or government should be able to provide a document that includes every controversial decision plus an added explanation for that decision.

- Passive security and robustness: has the method been evaluated with different robustness tools?

  The idea here is that: how can we make the AI systems robust to the adversarial attack [30]."

  The vulnerability of the AI system to adversarial examples opens up the discussion about the robustness of these systems to such attacks.

  We have to promote tools that can verify the robustness of the AI systems based on precision and scalability [31]. It means that this tool must be able to analyze the output of an AI system over large sets of data while providing an acceptable precision. [31, 32]

  Robustness tools will provide the developers with a way to certify the response of their output to adversarial examples, which can be considered as another aspect of the standardization of AI systems.

- Active security and monitoring: how strong are the monitoring tools used for the detection of abnormal decisions?

  There are numerous established monitoring methods for the AI system [33, 34]. They should be an essential part of every one of these systems for the detection of abnormal events and security breaches. In [15], they introduced an adversarial monitoring system that can reinforce the security of such systems through a clone AI that is effective against adversarial examples. However, like the precedented methodologies, this observer is not answering the transparency issues with the AI systems.

  An interesting perspective that can be consider is to promote investments for developing detectors or observers that operate based on the explainable AI. Explainable AI methodologies are suggested to be utilized along-side a human operator to check the validity of the outcome [35]. However, an advanced and cost-effective solution can be through measures such as designing observers that automatically assess the logical reasoning of such systems.

  When the resulting explanation for AI system decision output can not be assessed, and non-secure operation of the investigated AI system can result in great compromise in terms of financial, safety, and

security of individual, general public, or government, perhaps one should consider the presented AI system unsafe and unauthorized to operate.

2.3 **Control through Education: Specialized Inspectors and Developers.**

How can we propagate and solidify the transparency of the methodology? We believe that the main element is education. The AI applications are in nature multidisciplinary. Gartner's view of best practices in AI is an assertion of this statement: "AI experts, such as data scientists, should avoid working in isolation. In addition, other stakeholders need to develop a mindset to embrace multidisciplinary collaboration. Business domain experts need to work together with AI experts to develop the right data science models, chatbots, image recognition, or any other AI application [36]."

To create a clearer vision, in the following, we highlight some of the important trends in Artificial intelligence:

- Business intelligence (BI): where an intelligent business model guides its clients such as executives or CEOs so that they can optimize their decision making [8, 37].
- How media will change in response to a pandemic [38]?
- Cognitive studies for understanding human depression [39].
- Combination of unmanned aerial vehicles and edge computing for disaster control [4].
- A recent report by Gartner urges the CIO and CDOs in governmental healthcare or other organization to benefit from AI's ability in early detection, containment, diagnosis, healthcare operation, and vaccine research and development. Other reports also indicate the growing need for involvement and revolutionization of AI in Healthcare [40].
- AI security and privacy issues [35, 41].
- Achieving transparent and non-discriminative AI. How to achieve digital ethics [42, 43]?
- Climate change: using machine learning for better analysis of the impact of the climate change [44].

Artificial intelligence needs a well-established multidisciplinary perspective to mitigate its risk and increase its benefits [20]. This perspective can be implemented in the form of AI specialized institutes. The outputs of these institutes will be trainees that are well educated, such that they not only possess the technical skills to devise an AI model, but also are aware of technical, ethical, and legal biases in their respective fields.

However, we believe that trainees should not be only limited to developers, but specialized investigators, and legal experts as well.

2.3.1 **Developers**

Universities should invest in AI developers that are prepared for the new wave of upcoming AI-based technologies.

AI sensory systems such as mask monitoring technologies with regards to the ongoing pandemic are being suggested [45] and even practiced in some countries [46]. As of now, these monitoring systems are implemented on a small scale. However, some studies suggest we should take this surveillance solution to the next level as when there is a pandemic the main priority should be given to public health [47]. On the other hand, mass surveillance may endanger the liberty and privacy of citizens. If a situation occurs that we may desperately need to employ such a device, are we ready for it?

Here, we suggest that in terms of developers' educations, we need more and more aware and educated professionals on the transparency, bias, and ethical issues of the AI systems. We recommend a few considerations in terms of training AI experts that may help us face the new wave of AI technologies.

- Offering students with courses that provides them with multidisciplinary training while establishing strong grounds on mathematical literacy. The students must become more critical of their models' decisions.
- Familiarizing AI students with courses related to designing interpretable AI systems.
- Familiarizing AI students with their social, legal, and ethical responsibilities.
- Encouraging and collaborating with newborn companies by providing them with interns and facilities.
- The coordinated attraction of investment from major stakeholders, by focusing on a major problem such as responsible AI, IoT, and AI security, health, and climate change.

2.3.2 **Inspectors**

One of the major problems in AI is the lack of specialized inspectors that can certify the safety and transparency of an AI system. Currently, many AI institutes such as Mila[2] in Quebec are offering education that generally familiarize students with problems such as bias and the interpretability of AI systems. However, the main problem here is that in such institutes, there are very few courses that are solely dedicated to recognizing ethical or technical biases in AI, and even fewer that are application-centric.

Since AI applications are embedded in almost every domain, we need specialized investigators that are capable of certifying the safety of an AI system with a view that is a combination of data mining and expert knowledge about the potential ethical or technical biases.

In AI institutes such as Mila, a student will receive an education that helps him to learn about the potential biases of the AI systems from a data mining perspective or some general concepts on the idea of ethical problems in the AI systems. Basically, they learn why an AI system should be interpretable, but not specifically what are the biases in the different AI platforms. The training that these students are receiving is not directly an answer to the biases with consideration of a specific application. Therefore, the view is broad rather than focused, while the industry may need a moderate combination of both.

The output of these institutes will be students that can design an AI system for a certain AI platform but not necessarily how to inspect its biases because they are biased to design by considering the potential biases and are not trained for the sole purpose of criticism.

[2] Mila: https://mila.quebec/en/

Credible organizations such as Gartner also weight in importance of AI inspectors: “Promote people skills. Fill or hire people in key AI roles related to AI ethics, governance, and policy. Look for privacy/brand remediation and AI behavior forensic specialists who can explain models and perform investigations when AI fails to reduce risk [48].”

Gartner also predicts: "By 2023, over 75% of large organizations will hire AI behavior forensic, privacy and customer trust specialists to reduce brand and reputation risk [48]." But, are we doing enough to make these AI forensic experts ready for their task?

Training AI inspectors will make us ready for the upcoming changes in AI technology, prepare us for stable governance of these systems, and accelerates the process of integrating AI systems in more aspects of the technology.

These potential inspectors should possess specific characteristics that are mentioned in the following:

- A trained inspector must have strong mathematical literacy to be able to recognize the deficiencies and common biases specific to an AI algorithm.
- These trainees must be of a multi-disciplinary nature that is specialized in a few essential ground model that recognizes different types of biases of AI models. Should be able to perform adversarial test, robustness test, and security test by purely mathematical and data mining knowledge [32, 34].
- A trained person should be able to introduce adversarial input that can result in misclassification by the AI systems [34]. They should be able to introduce these inputs through the available algorithm by purely data-mining knowledge, as well as experience-based input generated by the pre-defined simulators, or based on their knowledge about the application.
- They should be able to recognize what are the elements that are missing from the input or output of an AI system that can cause a legal, discriminatory, or ethical gap.
- They should also be trained specifically with regards to a certain application to be able to challenge or question the types of defined input or output, and demand adding the neglected input or output by the developer.
- A trained inspector should be aware of the IoT connected privacy, transparency, and security problems concerning the specific application of the AI system. Since every AI system can include the different realm of science such as biology, engineering, and medicine, we suggest the inspectors can be more valuable if they are domain-specific.

Furthermore, inspectors can be categorized into two groups:

1. The first group can become part of the AI company, help with identifying data breaches, unintended use of the AI system, or uncover undesirable bias in the system [48].
2. The investigators in the second group are certifiers that are not affiliated with the investigated AI company. They can perform model behavior forensic [48] or identify ethical problems associated with the AI system.

The tools that investigators utilize can be categorized into pre-built machine learning investigators [48] and expert-defined analysis.

For example, an AI system that does not consider a certain input or output in a specific application can be vulnerable to potential lawsuits or malfunction. Here, an inspector with enough knowledge of the application identifies and expresses the problem to the developers or the relevant authorities.

2.3.3 **Legal Experts**

Contestability is introduced as one of the core principles of AI systems in numerous AI ethics studies [3, 6, 49]. It is recommended that we need a specialized legal platform dedicated to AI systems, such that people can legally challenge the abuse or harm caused by these technologies [6].

One of the main problems regarding AI technologies is the lack of preparation in the legal systems to deal with the AI-related legal arguments. People, organizations, or government harmed from a potential AI sensory device or an AI algorithm used by media platforms such as recommender systems utilized in Facebook or Amazon have to utilize the legal route that is not necessarily prepared to evaluate the AI-related cases.

We believe the third element in control is the training of legal experts in AI. These legal experts, besides the law, are also familiarized with the bias in AI systems to be able to analyze the evidence presented to them by AI developers and AI inspectors. The legal experts are prepared in such a way that they can handle complex cases that may arise as the AI system's application propagates through every aspect of our life.

3 **Accountability Model**

We are going to go back to our main question: how can one regulate and control AI systems?

It should be noted, excessive regulation may delay the development of technology, and lack of regulation can lead to financial, health, security, privacy, and safety risks. Here we suggest excessive regulation should be installed based on the potential risk that an AI system would pose to society. For example, the AI system utilized for surveillance should not be treated the same way as a system used for entertainment suggestions. To acknowledge the difference, we may want to regulate them based on how much sensory information they will require in their processing system or how critical is the task they uphold. Here we argue that accountability issues may be resolved through three-element introduced as the axis of a reliable AI.

First, we addressed ethical frameworks as the first steps toward a responsible AI. These constraints can be handed to the AI developers, which they can use as their moral principal and duties for the development of responsible AI systems.

Next, transparency of methodology is argued as the second important factor in achieving the ethical by design AI. These transparencies can be ensured to some degree by understanding the nature of attacks, employing an explainable technique or adversarial AI (as an observer), utilizing defensive methodologies such as adversarial training for improving the AI system, and carrying out security certification tests on the AI systems. However, one cannot expect that an AI developer possesses all of this knowledge. Here we argue that these constraints and requirements must be worked through with the willing developers as part of a certification process that allows them to understand the standards through the guidelines provided to them. The certifications will be given to them, after carefully testing their respective AI system with the best tools available, by the assigned inspectors. Inspectors' job is to determine what level of safety they bring or what level of threat they pose to the public.

Therefore, in terms of accountability we can determine the ethical and legal responsibilities according to the following:

First, it is the clients who decide whether they are willing to use a tool with a certain level of standards or not. In a situation where the AI system has created a threat or harm toward the citizens, and the AI developer refused to provide any level of standard for their product, they also have to face the negative consequences. And lastly, those who follow the standards will be protected by the standard providers, and here the policymakers will be accountable and must work through to provide modifications for the constraints that will lead to an improved, robust, and reliable AI.

4 Conclusion

In this paper, initially, two aspects concerning the tools required for the regulation and design of an ethical AI are introduced. These two aspects can be viewed in terms of technical AI design regulation (transparency of the methodology) and ethical standards. Later on, the third aspect introduced as the control and overviewing of the AI systems. In terms of ethical aspect, it is argued that a framework for the developers in terms of reminding them of their ethical responsibilities must be provided. People also need to have arrangements and tools for raising their voice against malicious or harmful AI systems. In terms of the technical aspect, the developer has to choose a methodology that is transparent with regards to data and must ensure that its model is not vulnerable to adversarial examples so that it does not put citizens' security and privacy at risk. Besides well-educated AI developers, to implement and regulate these aspects, our study recognized the need for specialized AI inspectors and legal experts. Specialized inspectors are not only mathematically well-educated, but they are also specialized to recognize biases that are application-specific and may go unnoticed by an AI-forensic specialist that only is educated in the data mining aspect of the problem. The inspectors will be tasked to evaluate the AI systems and grants standard in terms of maintaining safety, security, transparency, and privacy by the AI systems. The standardization is a path toward ethical by design AI and can help the government to regulate these systems through legal experts or the relevant authorities.

References

[1] D. Reinsel, J. Gantz, and J. Rydning, "The digitization of the world from edge to core," *Framingham: International Data Corporation*, 2018.

[2] D. Castelvecchi, "Can we open the black box of ai?," *Nature News*, vol. 538, no. 7623, p. 20, 2016.

[3] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of ai ethics guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019.

[4] F. Al-Turjman, *Unmanned Aerial Vehicles in Smart Cities*. Springer, 2020.

[5] "Design, scope, cost-benefit analysis, contracts awarded and implementation associated with the better management of the social welfare system initiative." Senate Community Affairs Committee, Parliament of Australia, 2017.

[6] M. Zalnieriute and O. Gould-Fensom, "Artificial intelligence: Australia's ethics framework submission to the department of industry, innovation and science," *UNSW Law Research Paper*, no. 19-40, 2019.

[7] P. van Esch, J. S. Black, and J. Ferolie, "Marketing ai recruitment: the next phase in job application and selection," *Computers in Human Behavior*, vol. 90, pp. 215–222, 2019.

[8] D. Edge, J. Larson, and C. White, "Bringing ai to bi: enabling visual analytics of unstructured data in a modern business intelligence platform," *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–9, 2018.
[9] J. Ulenaers, "The impact of artificial intelligence on the right to a fair trial: Towards a robot judge?," *Asian Journal of Law and Economics*, vol. 11, no. 2, 2020.
[10] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine bias risk assessments in criminal sentencing," *ProPublica, May*, vol. 23, 2016.
[11] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, *et al.*, "Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations," *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018.
[12] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of ai ethics guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019.
[13] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
[14] A. Rai, "Explainable ai: from black box to glass box," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 137–141, 2020.
[15] M. Taddeo, T. McCutcheon, and L. Floridi, "Trusting artificial intelligence in cybersecurity is a double-edged sword," *Nature Machine Intelligence*, pp. 1–4, 2019.
[16] "Take control of your virtual identity," *European Commission*, 2019.
[17] D. M. Turek, "Explainable artificial intelligence (xai)." URL: https://www.darpa.mil/program/explainable-artificial-intelligence.
[18] "Regulation (eu) 2019/881 of the european parliament and of the council." URL: https://eur-lex.europa.eu/eli/reg/2019/881/oj.
[19] B. Mittelstadt, "Principles alone cannot guarantee ethical ai," *Nature Machine Intelligence*, pp. 1–7, 2019.
[20] "Strategy for the development of quebec's artificial intelligence ecosystem," *A mandate from l'Économie, Science et Innovation Quebec*, 2018.
[21] A. F. Winfield, K. Michael, J. Pitt, and V. Evers, "Machine ethics: the design and governance of ethical ai and autonomous systems," *Proceedings of the IEEE*, vol. 107, no. 3, 2019.
[22] "Statement on algorithmic transparency and accountability," *Association for Computing Machinery US Public Policy Council (USACM)*, 2017.
[23] A. Winfield, "Ethical standards in robotics and ai," *Nature Electronics*, vol. 2, no. 2, pp. 46–48, 2019.
[24] R. Clarke, "Principles and business processes for responsible ai," *Computer Law & Security Review*, vol. 35, no. 4, pp. 410–422, 2019.
[25] S. Dilmaghani, M. R. Brust, G. Danoy, N. Cassagnes, J. Pecero, and P. Bouvry, "Privacy and security of big data in ai systems: a research and standards perspective," pp. 5737–5743, 2019.
[26] T. A. Singlehurst, M. Kelley, A. Shirvaikar, C. T. O'Neill, M. May, and W. H. Pritchard, "eprivacy data protection who watches the watchers? – how regulation could alter the path of innovation," *Citi Global Perspectives & Solutions*.
[27] D. Schiff, J. Biddle, J. Borenstein, and K. Laas, "What's next for ai ethics, policy, and governance? a global overview," *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 153–158, 2020.
[28] L. Findlater, S. Goodman, Y. Zhao, S. Azenkot, and M. Hanley, "Fairness issues in ai systems that augment sensory abilities," *ACM SIGACCESS Accessibility and Computing*, no. 125, pp. 1–1, 2020.
[29] D. Alba, "A.c.l.u. accuses clearview ai of privacy 'nightmare scenario'."

URL:https://www.nytimes.com/2020/05/28/technology/ clearview-ai-privacy-lawsuit.html, 2020.

[30] N. Dalvi, P. Domingos, S. Sanghai, D. Verma, *et al.*, "Adversarial classification," *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 99–108, 2004.

[31] T. Gehr, M. Mirman, D. Drachsler-Cohen, P. Tsankov, S. Chaudhuri, and M. Vechev, "Ai2: Safety and robustness certification of neural networks with abstract interpretation," *2018 IEEE Symposium on Security and Privacy (SP)*, pp. 3–18, 2018.

[32] G. Singh, T. Gehr, M. Mirman, M. Püschel, and M. Vechev, "Fast and effective robustness certification," *Advances in Neural Information Processing Systems*, pp. 10802–10813, 2018.

[33] S. Qiu, Q. Liu, S. Zhou, and C. Wu, "Review of artificial intelligence adversarial attack and defense technologies," *Applied Sciences*, vol. 9, no. 5, p. 909, 2019.

[34] D. L. Marino, C. S. Wickramasinghe, and M. Manic, "An adversarial approach for explainable ai in intrusion detection systems," in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pp. 3237–3243, IEEE, 2018.

[35] D. Gunning, "Explainable artificial intelligence (xai)," *Defense Advanced Research Projects Agency (DARPA), nd Web*, vol. 2, no. 2, 2017.

[36] F. Choudhary, A. Chandrasekaran, and P. den Hamerl, "Organizational best practices for successful ai and ml initiatives." URL: https://www.gartner.com/en/documents/3981025/ organizational-best-practices-for-successful-ai-and-ml-i, 2020.

[37] P. Bhattacharya, "Artificial intelligence in the boardroom: Enabling 'machines' to 'learn'to make strategic business decisions," in *2018 Fifth HCT Information Technology Trends (ITT)*, pp. 170–174, IEEE, 2018.

[38] J. Snowdon, "Covid-19 spurs media and entertainment companies to transform." URL: https://www.ibm.com/ thought-leadership/institute-business-value/blog/ covid-19-media-and-entertainment, 2020.

[39] A. Jan, H. Meng, Y. F. B. A. Gaus, and F. Zhang, "Artificial intelligent system for automatic depression level analysis through visual and vocal expressions," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 3, pp. 668– 680, 2017.

[40] E. B. Pieter den Hamer, "How to use ai to fight covid-19 and beyond.." URL: https://www.gartner.com/en/documents/3983523/ how-to-use-ai-to-fight-covid-19-and-beyond, 2020.

[41] S. Dilmaghani, M. R. Brust, G. Danoy, N. Cassagnes, J. Pecero, and P. Bouvry, "Privacy and security of big data in ai systems: a research and standards perspective," pp. 5737–5743, 2019.

[42] L. Floridi, "Translating principles into practices of digital ethics: Five risks of being unethical," *Philosophy & Technology*, vol. 32, no. 2, pp. 185–193, 2019.

[43] F. Buytendijk, J. Hare, and L. C. Jones, "Digital ethics by design: A framework for better digital business." URL: https://www.gartner.com/en/documents/3953794/ digital-ethics-by-design-a-framework-for-a-better-digita, 2019.

[44] C. Huntingford, E. S. Jeffers, M. B. Bonsall, H. M. Christensen, T. Lees, and H. Yang, "Machine learning and artificial intelligence to aid climate change research and preparedness," *Environmental Research Letters*, vol. 14, no. 12, p. 124007, 2019.

[45] M. S. Hossain, G. Muhammad, and N. Guizani, "Explainable ai and mass surveillance system-based healthcare framework to combat covid-i9 like pandemics," *IEEE Network*, vol. 34, no. 4, pp. 126–132, 2020.

[46] J. Vincent, “France is using ai to check whether people are wearing masks on public transport.” https://www.theverge.com/2020/5/7/21250357/ france-masks-public-transport-mandatory-ai-surveillance-camera-software, 2020.

[47] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, “A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic,” *Measurement*, vol. 167, p. 108288, 2020.

[48] J. Hare and et al, “Predicts 2019: Digital ethics, policy and governance are key to success with artificial intelligence.” URL: https://www.gartner.com/en/documents/3895092/ predicts-2019-digital-ethics-policy-and-governance-are-k, 2018.

[49] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, *et al.*, “Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations,” *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018.