ISSN: 2754-6659

Journal of Artificial Intelligence & Cloud Computing

SCIENTIFIC Research and Community

Review Article

Open Access

Responsible Artificial Intelligence on Large Scale Data to Prevent Misuse, Unethical Challenges and Security Breaches

Chandra Sekhar Veluru

Tracy, United States

ABSTRACT

Artificial Intelligence, AI, has recently grown exponentially and has transformed tremendously and some of the sectors that primarily benefitted are education, public administration, environmental management, and workforce management. The element of AI enhances data analysis, decision-making, and automation, and it is, according to its promises, set to do this with more efficiency and accuracy. As AI technology rapidly proliferates, it raises ethical concerns about its potential misuse, bias, and security. However, stringent ethical practices have emerged, and the emphasis remains squarely on accountability, transparency, fairness, privacy, and security. And that is what this research paper delves deeper into, to assess how it remains effective in addressing the associated risks of AI technologies. It is therefore important to explore the key principles for responsible practices of AI, including how it is implemented and their effects across industry sectors. Along with these fundamental principles, the paper presents various case studies, which cover a series of fields to give an ample view of practical applications and success of these ethical standards. For instance, AI in education offers personalization in learning while keeping fairness and transparency, AI in public administration ensures accountability in making decisions, and the environmental management department gets to have AI application, offering sustainability to the environment. In the workforce management sector, AI enhances workforce management with ethical guidelines to ensure fairness and no bias during recruitment and evaluation processes. This establishes that some important themes entail continuous monitoring, data practices that are diverse and inclusive, and techniques for AI that can help in explaining, hence promoting transparency and trust.

***Corresponding author**

Chandra Sekhar Veluru , Tracy, United States.

Received: January 08, 2024; **Accepted:** January 15, 2024; **Published:** January 23, 2024

Keywords: Artificial Intelligence (AI), Ethical Frameworks, Accountability, Transparency, Fairness, Privacy, Security, Bias Mitigation, Explainable AI (XAI), Responsible AI Practices, Ethical AI Deployment, AI Governance

Introduction

Artificial Intelligence has recently experienced a decade-long exponential rise and, with its formerly unimaginable bounds of possibilities for data analysis, decision-making, and automation, has already been a game changer for a few industrial sectors. In sectors such as healthcare or finance, transportation, or education, AI has silently crept in and started paying dividends by quickly increasing efficiency and accuracy in running large volumes of data. Machine learning algorithms have also had a significant leap in advances, which enables AI systems, in interfacing with advancements in computing power and data availability, to adjust and learn in the execution of tasks of increasing complexity with less human control.

For instance, AI systems in health care can diagnose diseases with the same accuracy as human experts. For example, AI systems, such as those developed by IBM Watson Health, can scan billions of medical data to help diagnose and plan treatment [1]. The domain of finance operations that AI-based algorithms target includes performing high-frequency trading at the stock market, fraud detection, and automated customer service with the help of chatbots [2].

Transport has accordingly been drastically transformed with AI driverless cars and intelligent traffic management as well, just as much as the sphere of education, which is using AI algorithms to perform experiences of personalized learning that are subject to change following the needs of a student [3].

In other words, while AI has many advantages, the rapid dissemination of AI technologies also carries various possible risks and challenges, including the misuse and abuse of AI-based systems, ethical concerns, and security weaknesses. Such threats are magnified manifold as security risks increase with the integration of AI in these critical systems and their corresponding decision-making processes.

Uncontrolled malicious employment may be easily possible in cyber-attacks, unfair decision-making without careful consideration, and privacy vulnerability without proper treatment of sensitive data. The potential risks include misuse and abuse of AI-based systems, ethical concerns, and security weaknesses; hence, its wide adoption calls for responsible AI practice with proper ethical standards, transparency, and security measures to mitigate possible risks involved in the comprehensive adoption process. In other words, challenges such as misuse and abuse of AI-based systems, ethical concerns, and security weaknesses call for responsible practices of AI with adequate ethical standards, transparency, and security measures to mitigate possible risks in the process of its wide adoption [4].

One common misuse is that the AI in the deepfake technology is used to create very authentic but false content, be it audio or video, and it is used to spread misinformation, thus creating a massive threat to public security and trust. Such are some of the examples in which some sort of decisions give way to biased AI algorithms, which, in the case of recruitment, will mean discriminating against some demographic groups if the training data does not represent a large enough share of the different sections of the populous [5]. All such cases set the dire need for responsible AI frameworks. Responsible AI is the ethical, transparent, and safe development and use of AI technologies. It is a way to ensure that AI technologies are aligned with respect for the principles related to furthering benefits towards societies and reducing the creation of harm. Essential elements of responsible AI include accountability, fairness, transparency, and security. There is an accountability requirement for AI systems to be designed and operated so that ultimate responsibility lies with some human or organizational agent.

Fairness checks AI systems so that the system shows no bias against a group or individual. Therefore, transparency is a property that makes AI decision-making processes understandable to users and stakeholders. This, in return, enables the aspect of trust and openness. Lastly, securing the AI means protecting it from tampering and protecting its data from attacks and unauthorized access. In essence, security guards the systems and their data [6].

A surge in interest in accountability about AI systems used in organizations, especially with regulations such as the European Union's GDPR and the proposed Artificial Intelligence Act in which organizations are held liable for decisions and impacts made by their AI systems, has been created. Initiatives looking at fairness are being driven by organizations such as the AI Now Institute, under which recommendations for fairness in algorithms have been initiated, and now, Google and Microsoft are all working towards reducing biases in their models. This is about transparency, which is being promoted through open research and disseminated to the rest of society by organizations such as OpenAI and other players in that field. Security is a significant concern to research agencies and scholars, who provide guidelines for securing these AI systems against cyber threats. The main objective of this proposal is to show a review question to inquire into the current position concerning responsible AI, including its effectiveness in preventing misuse, unethical challenges, and security. It should be comprehensive in at least three aspects: the extent to which the principles of responsible AI are being implemented across various industry sectors, the challenge points of their adoption, and the results.

This will connect to a review of best practices,current gaps in approaches, and future areas of improvement through publication and case studies.

Objectives of the Review

The primary objective of this review is to evaluate the effectiveness of responsible AI in mitigating risks associated with AI technologies. Specifically, the review will focus on:

- To identify and evaluate the existing responsible AI practices and frameworks.
- To analyze the effectiveness of these practices in preventing misuse of AI.
- To explore how responsible AI addresses unethical challenges.
- To assess the impact of responsible AI on mitigating security breaches.

Artificial Intelligence Misuse

Although AI technologies developed rapidly and have been used across different industries to bring remarkable changes, cases of their misuse remain an important issue. Several ways in which AI can be misused are represented herein. First is deepfake technology that creates fake but hyper-realistic video and audio content. The term "deepfakes" has been visible during misinformation campaigns and has contributed significantly to public trust and security threats. At the 2020 U.S. presidential election, deepfake videos were used to manipulate attitudes on a grand scale, and they polluted the information space with deceit and disarray.

Moreover, AI-driven cyber-attacks are becoming sophisticated. They can be unethically applied to automating phishing attacks and adding to their efficiency by the personalization feature in messages through the data obtained from social media Oliveira, et al. This level of personalization masks a person from being able to detect any phishing, which therefore poses an increased potential for security breaches to the users, which internally has increased the need for solid measures in matters of security against the rise of cyber-threats, like coming up with an AI-detect system to sniff such acts and absorb further cyber-attacks in real-time.

Several frameworks and guidelines were developed to handle these concerns regarding misuse. For example, the Artificial Intelligence Act by the European Union tries to enforce strict compliance regulations on all high-risk AI applications. This includes issues regarding undertaking risk assessments, transparency, and accountability mechanisms [7]. Such rules are fundamental to prevent potential misuse by developing standards and holding developers fully accountable for the actions and impacts of their systems.

Dealing with Unethical Challenges is a significant thrust of AI ethics, which is generally troubled by concerns of bias and fairness. AI systems can perpetuate and even magnify existing biases if not adequately managed. This is most evidently seen with automated hiring systems, which may cause biased outcomes because of the biased training data they have been fed. As per Dastin, one of Amazon's AI recruiting tools is biased in selecting male candidates over female candidates due to historical data reflecting gender imbalances in large tech industries.

Organizations are resorting to machine learning techniques that have embedded fairness. This ranges from preprocessing the data by removing any biases to designing inherently fair algorithms and postprocessing the outputs to ensure equal outcomes. Companies like Google and Microsoft are increasing their investment in fairness research and implementing tools to audit and mitigate biases in AI models.

Another critical area of ethical AI is transparency. Clarity in making forecasts by AI systems creates trust with users and stakeholders and gives a channel for exercising oversight. Techniques for Explainable AI (XAI) work toward improving the interpretability of AI models by providing precise and understandable explanations for their decisions. For example, the LIME (Local Interpretable Model-agnostic Explanations) technique explains predictions of models to users by approximating them with simpler interpretable models of the same kind Ribeiro, et al. Another example from health care is applying XAI, in which AI decisions will be based on patient results or outcomes. An example is IBM Watson Health, an excellent representative of a large-scale application of XAI technology in explaining how AI treatments are derived, enabling

the physician to make informed decisions while enhancing the trust and use of AI systems in clinical practice [8].

Making Better Security Measures One concern around AI is that many questions about securing these technologies arise as they become highly integrated into critical infrastructure and services. The most serious vulnerabilities to AI systems include adversarial attacks. An attacker with a shallow knowledge of the model parameters can generate adversarial examples to induce incorrect decisions in the model. It is very important to identify such attacks to prevent the misuse of applications such as autonomous vehicle software and medical diagnostics [9].

Researchers have been developing brutal techniques for protecting AI systems, adding to the model's robustness against adversarial attacks. Some methods, such as adversarial training, involve training models on clean and adversarially perturbed examples to increase their robustness and thus attain safety properties [10].

The methods ensure that the data are obscured through the injection of noise in such a quantity that the aggregate-level analysis of data is permitted, ensuring further that sensitive data is obscured and precluding the extraction of information that may apply to any individual. Companies putting AI into the robustness and reliability of their self-driving systems, such as Tesla in its autonomous driving vehicles and Waymo, apply profound techniques in their products that usually encompass strict testing in simulation environments, accurate testing in mixed environments, and the use of redundant systems for detecting and responding to potential failures [11].

Case Studies

Healthcare

AI is rapidly attracting a lot of interest in health-related operations and studies. However, its application has to be managed effectively to avoid interfering with the many problems of ethics security. A good case study includes the implementation of AI for COVID-19 diagnostics. In this regard, AI systems have been developed and applied in the analysis of medical images to predict cases of COVID-19 with high accuracy, contributing to the prediction of early detection and treatment planning. However, rapid implementation can cause issues with data privacy and ethical use of patient information Wang, et al. On these grounds, care providers have instilled responsible AI practices through transparent decisions regarding AI and patient data security using advanced data encryption models. Additionally, these ethical guidelines for AI in health were developed to ensure the consideration of patient consent and responsible use of insights derived from the use of AI.

Finance

The other fundamental innovative area where high-frequency trading has evolved concerns AI technologies, which operate with low-latency transaction execution at very high speeds. However, in its view, these systems may also be a source of risk to market integrity and flash crashes. For example, the 2010 Flash Crash, where the Dow Jones Industrial Average plunged and recovered within minutes, shows this kind of light-speed trading is potentially dangerous Kirilenko, et al. As a result, financial regulators around the globe have implemented measures to make HFT systems more transparent and accountable. These include requirements that the firm keep comprehensive records of its algorithms used for trading purposes and be regularly audited to ensure compliance with market rules Brogaard, et al.

Education

AI in education can be used to drive home the point that personalized learning is accurate and tailored according to the capabilities of each student.

However, it is also essential that fairness is maintained at all costs to avoid bias and make these inequalities worse. For example, it should be designed to give all students equal opportunity, regardless of their background or ability [12]. Ed-tech companies address these through the adoption of fairness-aware algorithms and including diverse stakeholder groups in the design and testing processes of AI systems, such an approach might ensure inclusivity in the design of AI-driven educational tools for all students' equal benefit [13].

Self-Driving Cars

There are giant steps in the transport sector featuring autonomous vehicles, yet ensuring these AVs are safe and reliable is the most important thing. The fatality resulting from an accident involving a self-driving Uber vehicle reinforces the need for rigorous testing and validation of AV systems [14]. AV companies are developing advanced simulation techniques to run many tests regarding safety under different conditions. They also implement redundancy measures, e.g., including more sensors and fail-safe mechanisms in the design so that it will work safely even if a system crashes [15].

Deepfakes

Deep fake technology could be used in legitimate applications in the entertainment and educational sectors but poses numerous ethical and security risks. For instance, a deep fake was made on the video clip that ran political circles viral with fabricated information meant to turn the opinion of the voters. The content on the video by the former President of the United States, Barack Obama, was created as a deepfake in an attempt to showcase to the world the potential dangers that are associated with deepfakes and the power that the technology possesses in creating critically disruptive content [16].

To fight back against such misuse of technologies, researchers are developing detection models to localize manipulated media accurately. Besides, social platforms like Facebook and Twitter have moved even further in detecting and deleting deepfakes to avoid turning the situation into a big issue. Although the proliferation of AI technologies across sectors is promising, the potential to create value is equally associated with ethical and security risks. There are also important reasons related to accountability, fairness, transparency, and best security practices for having responsible AI in place to be prepared to address these challenges [17].

Public Administration

AI can augment public administration decision-making and service delivery. Yet, without a proper ethical guideline, very much misused, and it will not ensure the working of AI systems is transparent and fair. Wirtz, et al. emphasize this importance and document the role of transparency in the decision-making of AI systems and public administrators' accountability and interpretability expected from them in these systems [18]. They also ensure the interpretation of AI systems through citizen participation based on public values. For instance, in the area of public safety, better predictive analytics could lead to a better result, but in the absence of proper ethical frameworks, such systems could lead to biased policing. When representative data is used, these guidelines must ensure fairness, transparency, and

accountability in the resulting predictive models.

Environmental Management

AI has the potential to make tangible contributions to sustainability in the management of the environment. Ethical considerations are, in the process, vital so that the technologies are applied correctly, devoid of any environmental injuriousness and harm to already marginalized communities. Vinuesa, et al. explore the role of AI in reaching the UN Sustainable Development Goals (SDGs) while advocating for the formulation of ethical guidelines that can guide the development and application of the same in managing the environment [19]. These are liable to foster and secure transparency, fairness, and accountability for sustainability and social justice. A practical example would be climate modeling and prediction, which comes with ensuring the accuracy and representativeness of the training data, transparency, and accountability in the process of use of ethical AI .

Workforce Management

AI in workforce management can be used to optimize the recruitment process, performance evaluation, and employee training. Ethical frameworks are then required to avoid bias and ensure employees' fair treatment. According to Binns, the problem when using AI in recruitment is that a system needs to be fair and enhance accountability [5]. It is also important that no biases are implemented in the AI system that result in some form of discrimination based on certain demographic characteristics.

Candidates need to be informed of how AI systems have reached their decisions. Ethical guidelines in workforce management also need to refer to the data privacy problem. Employee data needs to be secured, and AI systems must disclose the use and processing of the provided data. According to Holstein, et al. in AI ethics in workforce management that is of proper form, the data about employees and the relevance of the notions of fairness and transparency are discussed [20].

Furthermore, the exploitation of AI to its maximal potential and risks is still controllable with robust frameworks, regulations, and technological solutions. The use cases and practical implementations described above are examples of how responsible holistic approaches will ensure AI technologies are used ethically and safely for the good of society.

Ethical Practices in Artificial Intelligence

Accountability and Transparency

Accountability and transparency are some of the mainstays of any ethical AI framework. Such principles are set to ensure AI systems are designed and run by accountable organizations. Transparency means the processes involved in selecting an appropriate model of the algorithm, or similar decision-support tools and what those models are, for any given AI system, ascertainable behavior, so that stakeholders can see and observe the decisions being made by AI systems Mesbahi, et al. On the contrary, recent research has found the need for accountability in AI. For instance, Raji, et al. noted that practices recommend the establishment of an end-to-end framework for internal algorithmic auditing to close the AI accountability gap [21]. This work highlights the need for ongoing measurement and examination auditing to provide assurances of compliance with responsible AI standards. Similarly, as can be observed by Chen, et al. there are discussions of Explainability and Transparency in AI, where they offer ways to interpret obscured reasoning of complex AI models to allow understanding of gaps between the machine decisions and humans [8].

Fairness and Non-Discrimination

Fairness has deep implications in AI, with a need to make sure that AI technologies do not spread or amplify biases, which is a fundamental principle, as biases in training data and/or algorithms may result in discriminatory outcomes. These can be alleviated in methods such as data preprocessing for debiasing, learning algorithms that are ensured to be fair, and postprocessing under constraints of fairness.

Consistent with the above, Mehrabi and Holstein et al. have presented their study up to now on bias within machine learning methods to address bias, specifically the importance of fairness-aware algorithms and datasets that are more diverse and representative. More recently, Holstein, et al. discussed the efforts of big techs toward reducing bias [22].

Privacy and Data Protection

Privacy and data protection are critical to preventing the misuse of AI. Adequate AI application frameworks underscore ethical considerations in applications dealing with sensitive information.

Differential Privacy and Secure Multi-party Computation are two ways data can be protected when training and analyzing AI. Abadi, et al. explain differential privacy, such that data perturbs itself by injecting noise to make it less capable of revealing any information about any individual beyond a certain point while being useful [23]. More recent work improved this, as discussed by Bittau, et al. so that the tools for differential privacy are scalable and practical within real AI systems [24]. These contributions balance the needs of data utility with privacy preservation.

Safety and Security

This safeguards the functioning of AI, it is secure and free from misuse, unauthorized access, and adversarial attacks through robust security measures for both AI model and their processed data. For instance, in one of the earliest works, Papernot, et al. presented easily and showed how AI systems can conveniently be fooled into any evaluation [25].

Therefore their work has resulted in a call for more sophisticated and robust AI techniques through adversarial training that brings more robustness to model against such attacks. In another trend, Madry et al. outline the approaches through which AI models have been trained to be stronger against adversarial examples, ensuring the security and safety of the AI systems underutilization.

Conclusion

The rapid adoption of AI across sectors has shown transformational potential and has heightened the urgency for the development of stringent ethical frameworks to avoid misuse and address associated pitfalls. The major ethical challenges are biased and discriminatory systems, non-transparent and unaccountable systems, privacy intrusion, and security risks. The response to these issues is a multidimensional approach, and it should amalgamate within its scope of work comprehensive ethical guidelines, continuous monitoring, and engagement of stakeholders. Efforts should be made to ensure accountability and transparency of AI systems which will promote trust and enable effective oversight by the public. Accountability of AI systems should be achieved by ensuring end-to-end frameworks for internal algorithmic auditing. The accountability of AI is increased by a system of internal auditing to cover all algorithmic activities.

Transparency is also brought by the use of explainable AI models designed and developed to make the process of AI decision-making understandable and accessible to users and stakeholders. AI will neither cause nor propagate any bias if the proposed fairness-aware algorithms and diverse, representative datasets are guaranteed. AI is to be conceived as designed and implemented to ensure equal outcomes and adequately treat biases in the training data, which are inherently present.

Privacy and data protection will also be guaranteed if sensitive information is protected fully using methods like differential privacy and secure multi-party computation. Balancing utilities with data privacy preserved is very important for guarding an individual's information.

Adversarial attacks and unauthorized access to AI systems are rendered ineffective because their robustness is achieved by means of adversarial training. Use cases and practical applications in education, public administration, environmental management, and workforce management delineate the successful integration of ethical AI frameworks. For example, AI-based personalized learning in education needs to be sensitive to issues of fairness and openness in order not to magnify education disparities. In public administration, the quality to better serve the decision-making process can be achieved through adversity if the AI is to be clear and responsible. AI-based applications in environmental management help expedite the goals of sustainability, given they are deployed in an ethical way among the environmental and social harm guidelines. In administrative manpower management, ethical practices in AI assure fair recruitment and evaluation processes that protect the freedom of the employees and avoid discrimination. In a nutshell, there lies the ethical use of AI to reap its benefits and minimize the risks of deploying AI techniques. Organizations can harness the true potential of AI in a responsible and transformative manner with strong and comprehensive ethical frameworks that will maintain accountability, transparency, fairness, privacy, and security principles. Inclusive data practices, the development of explorable AI techniques, and constant surveillance regarding the use are the at-hub elements of trust buildup and assurance that AI technologies are always used for the good of society. A positive collective impact on society will be a strong tool in the hands and a driving force for innovations while keeping ethical principles [26-34].

References

1. Davenport M, Kalakota K (2019) The potential for artificial intelligence in healthcare. Future Healthcare Journal 6: 94-98.
2. Russell SJ, Norvig P (2020) Artificial Intelligence: A Modern Approach. 4th ed. Pearson https://aima.cs.berkeley.edu/.
3. Goodman B, Flaxman S (2017) European Union regulations on algorithmic decision-making and a 'right to explanation'. AI Magazine 38: 50-57.
4. Luckin R, Wayne Holmes (2016) Intelligence unleashed: An argument for AI in education. International Journal of Artificial Intelligence in Education 26: 731-735.
5. Floridi L, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, et al. (2018) AI4People-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines 28: 689-707.
6. Chesney R, Citron D (2019) Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. Foreign Affairs 98: 147-155.
7. Binns R (2018) Fairness in machine learning: Lessons from political philosophy. Proceedings of the 2018 Conference on Fairness Accountability and Transparency 149-159.
8. Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nature Machine Intelligence 1: 389-399.
9. (2021) European Commission. Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Brussels https://resourcecenter.cis.ieee.org/government/eu/cisgovph0040.
10. Whittaker M, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, et al. (2018) AI Now Report 2018. AI Now Institute at New York University https://ainowinstitute.org/wp-content/uploads/2023/04/AI_Now_2018_Report.pdf.
11. Brockman G (2019)) Open AI: Ensuring AI benefits all of humanity. Journal of Artificial Intelligence Research 65: 355-399.
12. Subramanian S (2020) Securing AI: Addressing Security and Privacy Issues in Artificial Intelligence. IEEE Security & Privacy 18: 21-29.
13. Wang B, Li Z, Qian Y (2020) Detecting Deepfake Videos Using Temporal Features and Attention Mechanism. IEEE Access 8: 135494-135505.
14. Raji K, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, et al. (2020) Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. Proceedings of the 2020 Conference on Fairness Accountability and Transparency https://arxiv.org/abs/2001.00973.
15. Papernot G, Patrick McDaniel, Somesh Jha, Matt Fredrikson, Berkay Celik Z, et al. (2018) The Limitations of Deep Learning in Adversarial Settings. IEEE European Symposium on Security and Privacy 372-387.
16. Madry A, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, Adrian Vladu (2018) Towards Deep Learning Models Resistant to Adversarial Attacks. International Conference on Learning Representations (ICLR) https://arxiv.org/abs/1706.06083.
17. Dwork C, Roth A (2014) The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science 9: 211-407.
18. Bojarski M, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, et al. (2016) End to End Learning for Self-Driving Cars. arXiv preprint arXiv:1604.07316 https://arxiv.org/abs/1604.07316.
19. Holmes S, Bialik Maya, Fadel Charles (2019) Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Learning Technologies 15: 85-102.
20. Zawacki Richter O, Victoria I Marín, Melissa Bond, Franziska Gouverneur (2019) Systematic review of research on artificial intelligence applications in higher education – Where are the educators?. International Journal of Educational Technology in Higher Education 16: 1-27.
21. Goodall N (2020) Can you program ethics into a self-driving car?. IEEE Spectrum 53: 28-58.
22. Shladover S (2018) Connected and automated vehicle systems: Introduction and overview. Journal of Intelligent Transportation Systems 22: 190-200.
23. Cellan Jones B (2020) Deepfake Obama: You won't believe what he says in this video! BBC News.
24. Nguyen T (2020) Deep Learning Models for Detecting Fake News in the Context of COVID-19. IEEE Access 8: 191587-191599.
25. Chen X (2021) Explaining Neural Networks with Hierarchical Interpretability. IEEE Transactions on Pattern Analysis and

Machine Intelligence 43: 1987-2001.

26. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A (2021) A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys (CSUR) 54: 1-35.
27. Holstein K, Wortman Vaughan J, Daumé III H, Dudik M, Wallach H (2021) Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems https://arxiv.org/abs/1812.05239.
28. Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, et al. (2020) Deep Learning with Differential Privacy. ACM SIGSAC Conference on Computer and Communications Security https://arxiv.org/abs/1607.00133.
29. Bittau A, Erlingsson Ú, Maniatis P, Mironov I, Raghunathan A, et al. (2021) Differential Privacy at Scale: Practical Algorithmic Privacy Techniques. IEEE Symposium on Security and Privacy.
30. Papernot N, McDaniel P, Goodfellow I, Jha S, Celik ZB, et al. (2021) Practical Black-Box Attacks Against Machine Learning. Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security https://arxiv.org/abs/1602.02697.
31. Wirtz BW, Weyerer JC, Geyer C (2020) Artificial Intelligence and the Public Sector-Applications and Challenges. International Journal of Public Administration 43: 1200-1211.
32. Richardson R, Schultz JM, Crawford K (2021) Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. NYU Law Review 94: 192-233.
33. Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, et al. (2020) The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications 11: 1-10.
34. Rolnick D, Donti PL, Kaack LH, Kochanski K, Lacoste A, et al. (2020) Tackling Climate Change with Machine Learning. ACM Computing Surveys (CSUR) 53: 1-36.