Artificial Intelligence as a Tool for Enhanced Data Integrity and Data Security

Lakshmi Priya Vinjamuri, Aghanaashaa Amal and Aditya Chahar

January 3, 2023

ARTIFICIAL INTELLIGENCE AS A TOOL FOR ENHANCED DATA INTEGRITY AND DATA SECURITY[1]

ABSTRACT

A strong and impregnable system is needed in the modern era of big data analytics and management to handle the problems of data integrity and security. Data is a company's most valuable asset, and if it is changed or removed, there is no way to properly know how, when, or by whom. This has a significant impact on the decisions that are made based on data and the analysis of accessible data. It is frequently claimed that data is essential to an organization's success and that "data is money." Because data has such a large impact, it is important to make sure it is preserved properly. This is true regardless of the industry a company operates in.

As a result, data management involves both data integrity and data security concerns, which call for an integrated tool that can function in tandem and concurrently guarantee both the integrity and security of the data.

The article makes an attempt to recommend artificial intelligence (AI) as a tool for improved and guaranteed data security and integrity while identifying the various areas that AI contributes to and addressing the gaps that need to be filled in order to overcome the difficulties of using AI tools with respect to data in general.

KEYWORDS: Data integrity, security, artificial intelligence, tools, company's survival

INTRODUCTION

It is ideal to comprehend the topic of artificial intelligence with the goal of ensuring and enhancing data integrity and security, thus it is important that we comprehend the phrases used to describe the objects.

Data correctness is determined by the data's integrity, which is defined as "the reliability and trustworthiness of data throughout its lifecycle ascertaining its validity or non-validity." For example, error checking and validation are typical techniques for guaranteeing data integrity as part of a process.

On the other hand, data security—often confused with integrity—primarily focuses on how to reduce the risk of disclosing information on intellectual property, especially under an NDA (Non-disclosure agreement) or confidentiality clause, the various documents used in the

[1] AUTHORS-
1. DR. LAKSHMI PRIYA VINJAMURI, ASSOCIATE ROFESSOR, LAW COLLEGE DEHRADUN, UTTARANCHAL UNIVERSITY, DEHRADUN 248007.
2. MISS. AGHANAASHAA. A., SOFTWARE ENGINEER, NINE LEAPS TECHNOLOGIES, BANGALORE, INDIA
3. MR. ADITYA CHAHAR, MBA, STUDENT, BITS, PILANI, RAJASTHAN, INDIA
4. MISS ALISHA, STUDENT, BJMC, UTTARANCHAL UNIVERSITY, DEHRADUN, INDIA

course of business or trade that are typically of a confidential nature, including information about people's health and medical records, email addresses, and trade secrets protected by intellectual property laws. Permissions management, data classification, identity and access management, threat detection, and security analytics are some of the emerging fields in data security.[2][1]

Although statistics from a report by KPMG show that most of the top senior executives are sceptical of the use of the technology of AI in protecting the integrity of the data and data per se when it comes to the employment of AI by their organisations for making crucial decisions that are binding on the company and dependent on data and its analytics, artificial intelligence has emerged as an effective tool to ensure both the integrity and security of data.

"Only 35% of people claim to have a high level of confidence in the use of data and analytics by their firm. 92% of people are worried about how data and analytics may affect an organization's reputation.

Just 35% of the 2,190 senior executives questioned by the company believe they have a high level of trust in how their company uses data and analytics. About two-thirds of them have some reluctance or outright mistrust in their data and analytics, which indicates a high level of anxiety about the hazards associated with data, analytics, and AI.[3]

Before dwelling further into divulging in the use and application of AI, it is important to understand the actual essence of artificial intelligence.

ARTIFICIAL INTELLIGENCE FOR DATA INTEGRITY (DI) & DATA SECURITY (DS)

Artificial Intelligence (AI), a subfield of computer science, can be defined as the area within computer science that focuses on simulating human intelligence. Its primary objective is to develop computer systems that can carry out tasks that are typically done by humans. These abilities include the ability to perceive visual and aural information, to learn and adapt, to reason, to identify patterns, and to make decisions. Machine learning, predictive analytics, natural language processing, robotics, and other related techniques and technologies are all referred to together as "AI" in this context..[4]

“As artificial intelligence develops, it increases the potential for using personal information in ways that violate privacy concerns by boosting the speed and capability of personal information analysis.”[5] Artificial intelligence applications have a profound and significant impact on many facets of society interaction, including human experience and national security. The potential of artificial intelligence extends to areas of knowledge expansion,

[2] https://www.varonis.com/blog/data-integrity

[3] https://www.zdnet.com/article/most-executives-dont-trust-their-organizations-data-analytics-and-ai/

[4] https://ovic.vic.gov.au/privacy/resources-for-organisations/artificial-intelligence-and-privacy-issues-and-challenges/

[5] https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/

prosperity, and solutions to global challenges while assisting the private and public sectors to address the issues of supercharged cyberattacks, defamation campaigns, weakened social cohesion, and subverted individual rights, leaving the legal frameworks in their infancy and naked state to address the issues of algorithmic bias and privacy.

The flip side of the coin that needs to be thoroughly examined is the potential for technologies powered by AI-enabled systems to increase surveillance and repress dissent, enable human rights violations, collect personal data, and further impose social control throughout the world. Legal and technological experts are to be held accountable for incorporating the tenets of international law, rules of technical standard bodies, and the tenets of open society.[6]

In addition to the technological advancements driving the IP (intellectual property domain), we are witnessing a paradigm shift in the application of artificial intelligence in the 21st century as a result of the Turing Test. This rapid development in the field of AI is due to improved algorithms, increased networked computing power, and improved ability to capture and store an unprecedented amount of data. We have subconsciously and unwittingly incorporated AI into our real-world experiences and apps, which has allowed AI to become a part of our everyday lives. The main characteristic of AI is that once it begins to function, it ceases to be referred to as AI and becomes a common form of computing. Examples include an automated voice on the other end of the phone or being recommended a restaurant or a movie based on your preferences and prior searches. These examples concentrated on the established aspects of our daily lives and frequently ignored the AI techniques of speech recognition, natural language processing, and natural language understanding.

AI is different in addressing data security and integrity

While privacy concerns are usually always a major concern when using new technology, the scope and applicability of AI present a particularly challenging scenario. Big data's effects can be considered as an extension of AI's in some ways, but AI technology has the capacity to handle enormous volumes of data while also using it to learn, build adaptive models, and make actionable predictions, often without the need of clear or comprehensible procedures.

The creation of AI technology entails a sizable risk that the preconceptions and prejudices of the people and businesses who produce it will have an impact on the AI's performance. For government organisations looking to deploy neural networks for decision-making, there are hurdles because to unintended consequences brought on by biases and opaque outcomes. This is necessary and, as a result, helps us comprehend the possibility of prejudice and how it interacts with privacy.

The ability to automate each of these processes is a crucial point of distinction between AI and current analytics tools. Furthermore, incorporating AI into existing technologies has the potential to fundamentally modify how they are currently used and how privacy is protected.

[6] https://www.lawandsecurity.org/security-privacy-and-innovation-reshaping-law-for-the-ai-era/

However, a network of cameras may become a considerably more intrusive tool when used in conjunction with facial recognition software. AI may alter how people communicate with robots in the future. Human-sounding voices found in digital assistants like Alexa and Siri are examples of anthropomorphic interfaces that may give rise to new privacy concerns. According to social science study, people frequently treat technology like a living being. [7]

Much of the information privacy discourse around AI has not accounted for the growing power asymmetries between institutions that accumulate data and the individuals who generate it.

Traditional views of information privacy are predicated on the idea that humans are the major data controllers and were not created to compete with AI's computational power, which defies conventional notions of data collection and processing. 15 AI has never before posed such a fundamental challenge to the way we now perceive ideas like informed consent, notice, and what it means to access or govern personal information. As these ideas develop, strategically including privacy concerns as a component of an ethical framework could help create AI that does not compromise information privacy.[8]

Personal information protected by the PDP Act and other information privacy legislation is one of the key factors to take into account with regard to data privacy. The definition and interpretation of personal information are subject to change along with society and legal standards. The identifiability principle underpins the integrity of data relating to personal information, particularly in the context of data, it is challenging to distinguish between real and fake information in terms of what is identifiable and what is not. It gets harder to determine whether a given piece of data is "identifiable" as the amount of data available grows and technologies for processing and combining it advance; looking at a piece of data in isolation is incompatible with AI technology and no longer a true reflection of whether it can be deemed "personal information."

Collection, purpose, and use of the information or data are the three factors that must be taken into account to ensure that privacy issues are addressed. Doing so is possible by implementing the OECD principles. Collection restriction, purpose articulation, and use restriction are the three enduring pillars of information privacy derived from the OECD.

[7] Stanford University, 'Artificial Intelligence and Life in 2030', *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, Section III: Prospects and Recommendations for Public Policy, September 2016, available at: http://ai100.stanford.edu/2016-report; Kate Darling, *Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects*, 2012.

[8] Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker & Kate Crawford, 'AI Now 2017 Report', *AI Now*, 2017, available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf, p 28

Limitations on collection: Personal information should only be gathered when it is absolutely necessary; only by fair and lawful procedures; and, when appropriate, with the individual's knowledge or consent. The data being fed into AI systems is frequently not gathered in a traditional transaction where consumers consciously give their personal information to someone who is asking for it due to technological advancements in IoT devices, smart phones, and online tracking.

In fact, a lot of people are frequently unaware of the extent to which information about them is being gathered from their gadgets and subsequently used as input for AI systems. Limiting the acquisition of personal information is inconsistent with how AI systems and the devices that gather data to enable them work, but gathering such enormous volumes of information entails inherent privacy hazards.

Purpose specification: At the time of collection, the individual should be informed of the reason for collecting their personal data. Most organisations follow the purpose specification concept by explaining the reason for the collection (often through a collection notice). This idea is really put to the test by AI's capacity to derive meaning from data that goes beyond the purposes for which it was originally gathered. In some circumstances, organisations might not always be aware of how the data will be used by AI in the future. By utilising too inclusive collection notices and privacy policies in an effort to "catch-all," there is a danger of collecting more data than is necessary "just in case." their privacy obligations, but it is deceptive and at odds with the collection limitation principle's fundamental purpose. Additionally, it erodes people's capacity to exert meaningful control over their personal data.

On the other hand, AI might be used to improve people's ability to declare their preferences for how their personal information is utilised. For instance, it is not out of the question to envision systems that can recognise the privacy preferences of their users and apply various restrictions to the data that is gathered about various people. In this approach, AI may play a key role in the development of personalised, preference-based models that may even more successfully than the existing model of notice and consent achieve the transparency, consent, and reasonable expectations goals of information privacy law.

Use limitation: Unless there is permission or legal justification to do otherwise, personal information should only be used or disclosed for the reason for which it was obtained. The overarching objective of these linked principles is to reduce the amount of information that any one organisation possesses about a person and to make sure that the information is handled in a manner that is compatible with that person's expectations. All three of these ideas are profoundly challenged by AI. The use limitation principle works to ensure that personal information is only used for the intended purpose after it has been gathered. In general, organisations are also allowed to use a person's personal information for a secondary purpose that they believe the person would "reasonably expect." In light of the fact that in many cases, the result of doing so would be unknown to the individual, it is unclear whether information used as input data for an AI system may be regarded a "reasonably expected secondary purpose." AI has the ability to draw attention to patterns and relationships in data that people would have missed, and it may also suggest new applications for that data. When organisations use AI technology, it may be challenging to ensure that personal information is being used for the reason for which it was obtained when these concerns of purpose definition are combined.

The notion that a reasonably expected secondary purpose for the use of information would be extremely broad may be prompted by the premise that people, especially young people or "digital natives," are becoming less worried about their information privacy. This isn't always the case. According to research by the Boston Consulting Group, the privacy of personal information continues to be a top concern for 75% of customers in most nations, and consumers between the ages of 18 and 24 are only somewhat less careful than older generations.[9] This suggests that, despite technology becoming more pervasive, people are not automatically becoming less concerned about how their personal information is being used. As a result, they may not always view the use of their personal information by AI as a secondary purpose that is to be reasonably expected. AI is anticipated to make it more difficult to distinguish between primary and secondary purposes, which may require reevaluating the applicability of the use limitation principle.

ARTIFICIAL INTELLIGENCE AND PRIVACY – ISSUES AND CHALLENGES

The article makes an effort to present a realistic picture of the application of artificial intelligence and the difficulties posed thereby for the legal framework addressing the issues of artificial intelligence in relation to information privacy, which are still under investigation in Toto on a global platform. The emphasis is solely on presenting a high-level understanding of AI, or rather, a critical and unbiased perspective of the technology's application, particularly in the public sector, and the difficulties this poses in relation to and in reference to information privacy, a key component of data security.

The challenges imposed in addressing the data integrity and security in combination or in isolation range from the issues surrounding narrow, general and super AI[10]deliberately programmed to be confined to and competent in one specific area referred commonly as augmented intelligence, 'artificial general intelligence' (**AGI**), super-intelligence, big data[11] inclusive of the types and scale of information included, searching online, sharing and transmitting of data and information on a regular basis, browsing a smart phone intentionally

[9] *Boston Consulting Group*, November 2013, available at: https://www.bcg.com/publications/2013/marketing-sales-trust-advantage-win-with-big-data.aspx

[10] Nick Bostrom, 'How long before superintelligence?', *Linguistic and Philosophical Investigations*, Vol. 5, No.1, 2006, pp 11-30

[11] Report of the Special Rapporteur on the right to privacy, prepared for the Human Rights Council, A/72/43103, October 2017

or not, machine learning [12] and the concept and technology driving the domain of deep learning[13]

The United Kingdom's Information Commissioner's Office has summarised the complex and fundamental relationship between the three technological facets of machine learning, artificial intelligence, and big data. *“Big data is a resource that can be challenging to use. Big data can be viewed as having value, and machine learning is one of the technical methods that supports and facilitates AI.”*[14]

Though it shouldn't be thought of as a solution to all the administration and enforceability issues facing different functionaries on different functionalities, AI is promising in a variety of fields, including the private, public, and governmental sectors. This includes a focus on information management, including privacy, data security, and ethics. [15].

PRIVACY CONSIDERATIONS

It is important to analyse the privacy and security issues raised by the use of artificial intelligence technology with an emphasis on the difficulties associated with information privacy and information security, which is a subset of data security. An analysis of the difficulties might provide the crucial springboard needed to address significant issues with information privacy.

The 1980 *OECD Guidelines on the Protection of Privacy and Trans-border Flows of Personal Data* serve as the foundation for most information privacy laws. The *Privacy and Data Protection Act of 2014*, for example, and other privacy laws around the world continue to uphold eight of these rules (PDP Act). *Even though the OECD Guidelines have been extraordinarily effective in advancing information privacy legislation globally, AI poses problems for the guiding concepts of the Guidelines*.

While the old ideas and ideologies of privacy are being questioned, AI should be used in a way that addresses some of the privacy concerns. For instance, the use of AI technology is likely to result in fewer people actually needing access to raw data in order to work with it. This could reduce the risk of privacy breaches caused by human error and enable consent, in which people receive personalised services based on privacy preferences that have been learned over time. The "status quo" of privacy protection may need to be reviewed in light of

[12] Will Knight, ‘The Dark Secret at the Heart of AI’, *MIT Technology Review,* 11 April 2017, available at https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/

[13] ICO, *Big Data, artificial intelligence, machine learning and data protection*, 2017, p 11

[14] The UK Information Commissioner’s Office (ICO) Big Data, artificial intelligence, machine learning and data protection, 2017, p 8.

[15] Hila Mehr, ‘Artificial Intelligence for Citizen Services and Government’, *Harvard Ash Center for Democratic Governance and Innovation*, August 2017, https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf, p 10

the rising usage of AI, but this does not imply that privacy would vanish or become unimportant.

The framework that information privacy gives for choosing how we should utilise new technology ethically is a key component of information privacy. The long-term viability of AI will depend on taking technological ethics into account and resolving privacy issues. The construction of socially responsible AI that can help in the long run in the creation of public value will be encouraged by a balance between technological progress and privacy concerns.

Together, AI severely challenges the purpose specification, collection limitation, and usage limitation rules. The current understanding of information privacy through these principles may no longer be applicable due to mass data collection, which frequently occurs through methods that are not obvious to individuals, vague or deceptive collection notices, and an assumption that people are more comfortable with the secondary use of their information than they actually are. AI offers the potential to fundamentally alter how conventional privacy rules are implemented, though.

A machine learning algorithm, for instance, might be trained on vast volumes of data in a secure environment before being released, which would boost data security. We will need to adjust how we apply established privacy principles as a result of the widespread usage of AI, but it is unclear whether this will result in higher or lower privacy protection requirements. There is potential for organisations to improve collection notice practises and give people the opportunity to have a more nuanced and informed interaction with organisations regarding the use - and secondary use - of their information by taking privacy into consideration as a foundational element within an ethical framework for developing AI.

CHALLENGES AND GREY AREAS OF AI FOR DATA PRIVACY AND SECURITY

Data gathered using AI also presents privacy concerns including freely given informed consent, the ability to opt out, data collection limitations, explanations of the nature of AI processing, and even the ability to remove data upon request. However, given a possible spillover effect, how would human subjects of the acquired data even be aware that data was being collected on them, allowing them to contact organisations about their own data or request that it be deleted?

This loss of trust was greatly exacerbated by the Cambridge Analytica affair, which involved the psychographic profiling of Facebook users, as well as the privacy implications of artificial intelligence. Threats to democracy are still being fueled by AI manipulating democratic levers. Another illustration is the US company Clearview AI's violation of Canadian privacy laws in collecting images of Canadian adults and even children for mass surveillance and facial recognition without their consent and for commercial sale. This only serves to erode public confidence in the ability of entire nations to responsibly handle privacy and AI-related issues. In Australia and the United Kingdom, investigations into Clear view AI are being conducted separately and concurrently.

Some of the most sensitive data is personally identifiable information (PII) and protected health information since "[data] is the lifeblood of AI" (PHI). Therefore, we must investigate how much AI uses PII, PHI, biometrics, and whether the appropriate caution has been taken to ensure, for example, that the levers of democracy are not twisted.

REFERENCES

1. https://www.varonis.com/blog/data-integrity
2. https://www.zdnet.com/article/most-executives-dont-trust-their-organizations-data-analytics-and-ai/
3. https://ovic.vic.gov.au/privacy/resources-for-organisations/artificial-intelligence-and-privacy-issues-and-challenges/
4. https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/
5. https://www.lawandsecurity.org/security-privacy-and-innovation-reshaping-law-for-the-ai-era/
6. Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker & Kate Crawford, ‘AI Now 2017 Report’, *AI Now*, 2017, available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf p3
7. Toby Walsh, *It's Alive! Artificial Intelligence from the logic piano to killer robots*, Latrobe University Press, 2017, p 60.
8. Stanford University, ‘Artificial Intelligence and Life in 2030’, *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, Section III: Prospects and Recommendations for Public Policy, September 2016, available at: http://ai100.stanford.edu/2016-report; Kate Darling, *Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects*, 2012.
9. Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker & Kate Crawford, ‘AI Now 2017 Report’, *AI Now*, 2017, available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf, p 28
10. *Boston Consulting Group*, November 2013, available at: https://www.bcg.com/publications/2013/marketing-sales-trust-advantage-win-with-big-data.aspx
11. Nick Bostrom, ‘How long before superintelligence?’, *Linguistic and Philosophical Investigations*, Vol. 5, No.1, 2006, pp 11-30
12. Report of the Special Rapporteur on the right to privacy, prepared for the Human Rights Council, A/72/43103, October 2017
13. Will Knight, ‘The Dark Secret at the Heart of AI’, *MIT Technology Review,* 11 April 2017, available at https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/
14. ICO, *Big Data, artificial intelligence, machine learning and data protection*, 2017, p 11
15. The UK Information Commissioner's Office (ICO) Big Data, artificial intelligence, machine learning and data protection, 2017, p 8.
16. Hila Mehr, ‘Artificial Intelligence for Citizen Services and Government’, *Harvard Ash Center for Democratic Governance and Innovation*, August 2017, https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf, p 10
17. John Rose, Christine Barton, & Rob Souza, ‘The Trust Advantage: How to Win with Big Data’