

To cite this article: Sidharth Jain (2024). ARTIFICIAL INTELLIGENCE AND PRIVACY: AN INVESTIGATION ON HOW ARTIFICIAL INTELLIGENCE-BASED DATA MINING IS AFFECTING THE PRIVACY OF USERS, International Journal of Education and Social Science Research (IJESSR) 7 (3): 220-227 Article No. 945, Sub Id 1473

## ARTIFICIAL INTELLIGENCE AND PRIVACY: AN INVESTIGATION ON HOW ARTIFICIAL INTELLIGENCE-BASED DATA MINING IS AFFECTING THE PRIVACY OF USERS

Sidharth Jain

The Shriram Millennium School Noida, Sector 135  
Noida, 201304, Uttar Pradesh, India

DOI: <https://doi.org/10.37500/IJESSR.2024.7316>

### ABSTRACT

Artificial intelligence has become an integral part of everyone's lives in today's world. Due to the rapid advancements in the field of artificial intelligence, there hasn't been sufficient mainstream analysis about its cons: especially privacy invasion. This demands focused attention on artificial intelligence (AI) driven data mining, a process which calls ethics and user privacy into question. Ultimately, there must be important compromises made, and in this case it's between privacy and growth.

**Research Question:** *Artificial intelligence is being used by machines, individuals, corporations and governments all over the world. Its capabilities and potential are truly extraordinary and life-changing; however, this does not mean that privacy and individuality can be sacrificed just for the sake of growth. Artificial intelligent technologies continue to grow through data mining, which enables to train on new and unique data. But the question that arises is whether AI-driven data mining serves users better or is it overriding their privacy? Is it fair for the entirety of the human population to have their sensitive information used just for the sake of growth and innovation? These and more questions will be attempted to answer in the course of this paper.*

**KEYWORDS:** Artificial intelligence (AI), data, big tech, privacy, growth, AI based data mining, cyber-attacks, data leaks.

### 1. INTRODUCTION

Artificial intelligence is the current buzz all over the world, but have layman users really given a thought about the privacy issues which are brought about by this technology? The realization that AI based technologies can lead to sensitive and confidential data being used just to train and improve the software is extremely important. The concept of data mining isn't new, it dates back to 1950 where the foundational algorithms were hypothesized, sowing the roots of data mining. But it wasn't a concept until the 1980s, when the term data mining was coined. (Thomas) In the recent times, data mining has been revolutionized due to artificial intelligence, enabling the creation of such software that can collect, process and analyze massive datasets on its own. Due to the immense capabilities

data mining holds, it is extremely important to analyze AI based data mining, and truly understand its consequences.

## **2. MATERIALS AND METHODOLOGY**

This research is mainly qualitative research which aims to investigate the ethical issues of artificial intelligence and artificial intelligence-based data mining. Additionally, this research incorporates quantitative research through the means of survey, through which important statistics were unraveled. The survey subjects consisted of educated individuals from across the globe. The data obtained from the survey was analyzed through using the process of data encapsulation and visual data representation.

### **2.1 Introduction to artificial intelligence**

Artificial Intelligence (more commonly known as AI) has revolutionized every aspect of today's world, whether it be content generation, sciences, business, or any other field. The core of AI relies on simply data, the more data the AI model has access to, the better it can be trained to analyze trends and patterns, leading to it being accurate and reliable. But, the type of data being used is on the discretion of the companies who own the AI software, giving birth to multiple questions centered around the privacy of the user. In the present era, technology is being used in almost all aspect of one's life, ranging from social media to even bread toasters, and this allows our data to be recorded in 0's and 1's. At times users aren't even sure about the terms and conditions they are agreeing to, giving out permissions to record their data through the device's cameras and microphone. What do the companies get out of this? Companies used to get the ability to sell never-ending lists of data in return monetary benefit, but with the development of AI technologies these lists of data can be used to analyze future trends which leads to a greater monetary benefit and market power.

As the capability of artificial intelligence is continuing to grow exponentially day by day, an individual's private data has a greater risk of being use manipulatively without their awareness of the same. Is it fair for the entirety of the human population to have their sensitive information used just for the sake of growth and innovation, which is a question I will be discussing in the paper.

### **3. Background on AI based data mining**

AI based data mining is an automated process in which systems use data which they are allowed to access, analyze the data, extract facts and predictions which would be helpful for the company. The findings would allow business to predict the future, and make themselves ready for it in order to maximize profit. In addition, AI based data mining is not only used by businesses: various governments are also actively using it for surveillance. The aim of the surveillance of our data (conducted by either public or private bodies) is to automate us. Our individual traits such as our voices, physical features and even emotions are being rendered into data. We are being programmed to function according to the predictions through the analysis of our data, we are being automated

(Zuboff). AI is the fuel for big tech analytics and the Internet of Things: their only motive is to create consumer behavior profiles and sell them specific goods and services (Manheim and Kaplan).

On the contrary, AI based data mining is not only negative for consumers, in some aspects it is beneficial and contributes to technological growth. With the help of such software models, consumers are being exposed to information that they would find more appealing and relevant based on their interests and behavior, allowing uniquely curated content made for each individual. Data mining is not only limited for the advertisement of goods and services, various other fields can also embrace the advantages it poses: In healthcare predications through analysis of data sets can contribute to accurate diagnosis and treatment of diseases leading to the betterment of our community and increases in social welfare.

### **3.1 Threats to privacy imposed by AI based data mining**

Privacy, by definition, is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others (Westin). Analyzing data and curating information out that analysis is more harmful than most of us think it is, big tech companies know us inside-out. This makes us ever so vulnerable to data leaks and cyber-attacks, which would put our confidential data in the wrong hands. The quantitative data present is eye opening: every 1 in 2 American internet users had their account breached in 2021 alone, 14 accounts were breached every second from July- September 2022. The hourly cost rate of cybercrimes has increased from \$2054 in 2001 to \$787,671 in 2021, this indicated the direct proportionality between increase in AI data mining and cybercrimes (Griffiths). “Would cybercrime even exist if our data was not being stored by big companies?” is a question that requires interpretation of the fundamentals of data mining and how companies are being careless about how they are storing our individuality.

### **3.2 Results of survey**

As mentioned before, I have conducted a survey which was aimed to determine the awareness of people about privacy threats due to AI data mining, whether the respondents have had their confidential data leaked and if they feel that AI mining models pose any benefits for them. Over 500 individuals had responded to my survey, from which 20% were between the ages 13 and 19, 50 % were between the ages 20 – 35, 20% were between the ages 35 – 50 and 10% were ages above 50 (Figure 1). Only 33.3 % of the respondents were aware about the privacy threats posed by AI data mining, which depicts the lack of awareness about privacy threats due to the storage of our data in our community (Figure 2). This statistic further highlights the need of the hour: awareness about threats posed by AI data mining. 70% of the respondents had been victims of data leaks, which they found out with the help of the website: haveibeenpwned.com (Figure 3). This percentage backs the argument that organizations and governments are not maintaining the security of the data that is being collected, this is the root of various cyber-attacks ranging from vishing, spear phishing, smishing to identity theft, fraudulent loan scams and sim-jacking. This primary data allows us to gain an insight

into the understanding society has about AI data mining and how our society is being adversely affected by it.

Age

514 responses

 Copy

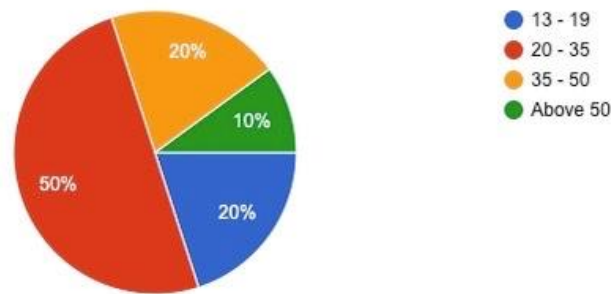


Figure 1: Age Divisions

Are you aware about privacy issues due to your data being analysed by companies ?

514 responses

 Copy

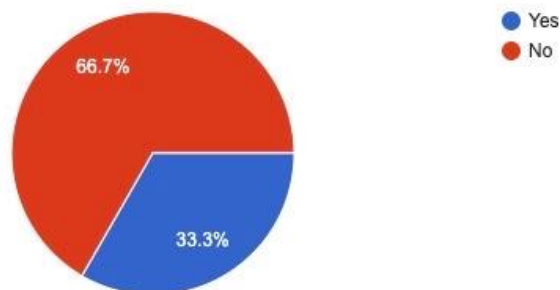


Figure 2: Awareness in the society

Has your data ever been leaked ? You can check this by going to [haveibeenpwned.com](https://haveibeenpwned.com).



514 responses

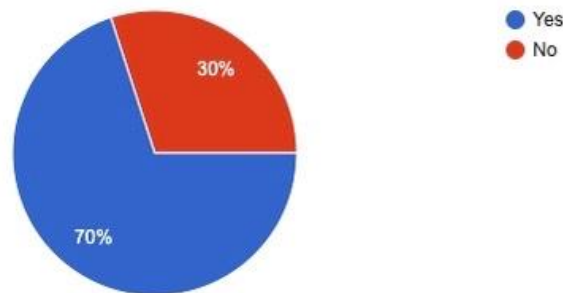


Figure 3: Data Leaks

### 3.2 Contribution to growth by AI data mining

AI data mining does have their benefits which lead to better efficiency, accuracy and automation of work. AI models have the ability to work 24/7, without any breaks and illogical errors, something which practically humans are not capable of. This profit maximization technique will help businesses to invest more in various other sectors, such as research and development leading to growth in other sectors as well. As stated, before various other fields will also benefit from AI data mining, some include: Healthcare, Education, Sports and Music. In healthcare the development of virtual private nurses can be established, allowing accurate diagnosis and care from your own house without the need of ever going to the hospital. The essential data can be recorded via smartwatches which are dynamic and multi-functional. Administrative workflow, such as documentation, can be made more efficient as without the need for a human to document the required information an AI model can do it through patterns it has recognized from the documentation it was trained with (IBM).

Education is an effective example of a field other than business where AI data mining can cause immense growth and innovation. Consider a classroom, where along with a teacher, there is a virtual class assistant. This assistant has access to all your mistakes, lesson times and conceptual errors, allowing for apt analysis of your holistic academic profile. This will enable that assistant to curate courses specifically for you, allowing you to maximize your understanding and thus grades. This would revolutionize the field of education, and multiple new functionalities can be added allowing students to reach their maximum potential, a level which not all teachers can guide all students to. On a broader picture, this would have a positive long run impact too, as the workforce would be highly productive and intellectual, leading to new ideas, more innovation and further growth of humanity. But this example has various limitations due to the assumptions made, as there is possibility of data leaks, biased AI results and various others.

Through the survey I had conducted, I was able to deduce the surprising fact that people find AI data mining helpful as it provides them access to personalized content on the internet, making the internet more appealing and useful for the users. (Figure 4). This backs the idea that AI based data mining has immense capabilities if used for social welfare and benefit. Maximum social welfare can be attained through using highly secure databases, encryption of the collected data, multi-factor authentication required to access the data and continuous supervision of the collected data.

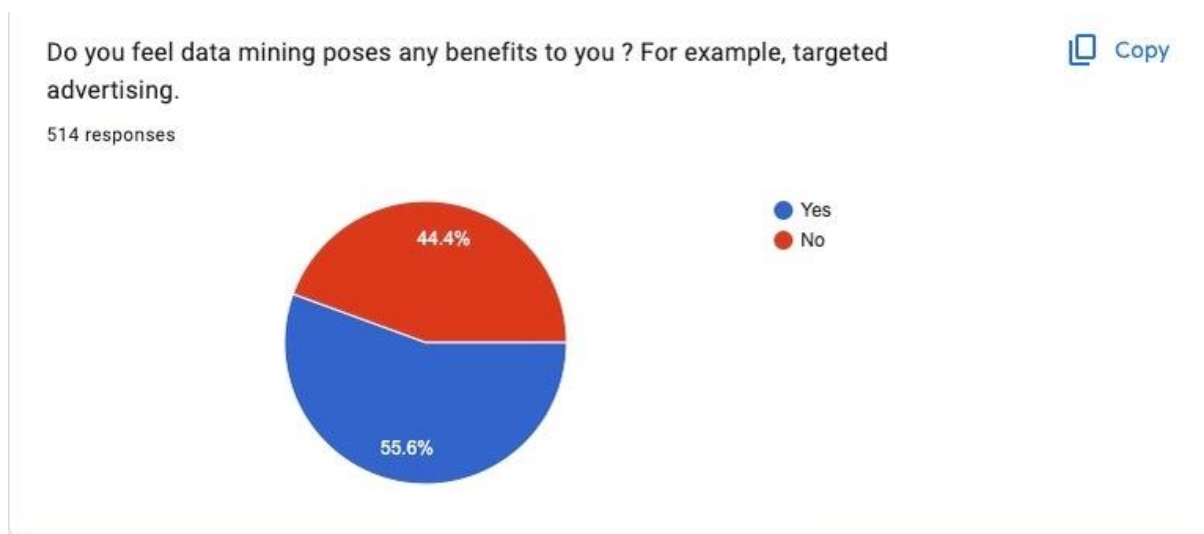


Figure 4: Whether AI data mining is helpful to users

#### 4. Consequences of AI based data mining on privacy and growth

The question which arises next is: “Is privacy more important than the growth of humanity?”. This is a controversial question, which leads to different perspectives. This question can be made more vivid with the help of a thought experiment : If privacy is not given important and only growth in terms of economic growth and technological growth is focused upon then after a certain point in time, wouldn't the AI models become so powerful the general public have to resort to either death or accepting the fact that AI driven models would control their lives and reach such a stage where data leaks wouldn't even matter as there is nothing left to hide. Would this be considered as growth? On the contrary, focusing solely on privacy and imposing various restrictions on AI data mining may lead to a decrease in incentives to produce such technology hindering growth of our civilization. Another point to be noted is that after a point no government or organization would be able to restrict AI, due to the simple fact that it would already have had access to a wide database and it would not be possible to determine where to restrict AI, in what way and in what field. The approach that needs to be taken should be as such that neither the growth of AI models is compromised, nor the privacy of sensitive information of users is compromised.

## 5. CONCLUSION

Coming back to the question I had stated at the beginning of this paper: “Is it fair for the entirety of the human population to have their sensitive information used just for the sake of growth and innovation?”. My answer to this question is no, through the research I have conducted and the knowledge I have gained through the process of writing this paper I strongly believe that privacy and growth should go hand in hand, because without privacy there is no growth and without growth there is not privacy. Some solutions I would like to propose for the development of such an AI data miner that can continue to provide useful analytics and quantitative data and still maintain the privacy of the general public are: data anonymity, encrypting data and proper data consents. Data anonymity would mean that the data of an individual is recorded and stored, but not their personal information like name, email address and phone number. This would allow the results to be unbiased, the data to be unlinked to an individual and ergo greater privacy. Secondly, encrypting data is a necessity, without properly encrypting data privacy cannot be attained, such strong encryption keys need to be innovated which are near to impossible to crack, the time and resources use to achieve this would be highly beneficial in the long run. Lastly, proper data consent should be provided, not overwhelming 100-page terms and conditions which beat around the bush. The data consents should clearly state the data that is being recorded, its uses and how it may affect the user, this practice will allow the user to make more informed decisions. To sum it up, AI data mining is currently breaching our privacy, adversely affecting us without our knowledge but with the help of the points stated in this paper and the solutions I have proposed I truly believe that such AI data mining can be done which doesn't breach one's privacy, and in doing so neither does it comprise on technological growth, allowing humanity to have the best of both worlds.

## 6. REFERENCES

- [1] Thomas, Sam. “The Future of Data Mining Services: How Ai Is Reshaping the Landscape?” Data Science Society, 29 Apr. 2024, [www.datasciencesociety.net/the-future-of-data-mining-services-how-ai-is-reshaping-the-landscape/](http://www.datasciencesociety.net/the-future-of-data-mining-services-how-ai-is-reshaping-the-landscape/).
- [2] Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Public Affairs, 2020.
- [3] Manheim, Karl, and Lyric Kaplan. “About.” Hein Online, 8 Mar. 2021, [heinonline.org/HOL/LandingPage?handle=hein.journals%2Fyjolt21&div=4&id=&page](http://heinonline.org/HOL/LandingPage?handle=hein.journals%2Fyjolt21&div=4&id=&page)
- [4] Westin, Alan F. *Privacy and Freedom* by Alan F. Westin. Atheneum, 1967.
- [5] Griffiths, Charles. “The Latest Cyber Crime Statistics (Updated May 2024): Aag It Support.” AAG IT Services, 3 May 2024, [aag-it.com/the-latest-cyber-crime-statistics/#:~:text=It%20is%20clear%20that%20the,hour%20worldwide%20has%20also%20increased](http://aag-it.com/the-latest-cyber-crime-statistics/#:~:text=It%20is%20clear%20that%20the,hour%20worldwide%20has%20also%20increased).
- [6] Education, IBM. “The Benefits of AI in Healthcare.” IBM Blog, 3 Oct. 2023,



[www.ibm.com/blog/the-benefits-of-ai](https://www.ibm.com/blog/the-benefits-of-ai)

inhealthcare/#:~:text=AI%20can%20help%20providers%20gather,to%20help%20improve%20drug%20safety.