Journal of Media Horizons
ISSN: 2710-4060 2710-4052 Volume 6, Issue 2, 2025

THE IMPACT OF AI AND BIG DATA ON PRIVACY OF YOUTH: INVESTIGATING CHALLENGES, SOLUTIONS AND FUTURE DIRECTIONS

Laraib Noor[1], Aniba Ameen[2], Itba Tahreem[*3]

[1,*3]*Department of Media and Development Communication, The University of Punjab, Lahore, Pakistan*
[2]*Department of Journalism Studies, The University of Punjab, Lahore, Pakistan*

[*3]itbatehreem@gmail.com

DOI: https://doi.org/10.5281/zenodo.15743075

Keywords
Data Privacy, Big data, Artificial Intelligence (AI), Digital Literacy, Privacy Challenges, Youth.

Article History
Received on 18 May 2025
Acceptsed on 18 June 2025
Published on 26 June 2025

Corresponding Author: * Itba Tahreem

Abstract
Big data comes with big problems. The rapidly growing usage of big data and AI, revolutionize the digital media landscape especially for youth. This study focuses on how youth privacy is affected by AI and big data. The study addresses the key challenges faced by youth with these growing technologies. The research paper examines the complicated landscape of privacy and security concerns, highlighting the relevance of the topic as well as the problems. The paper discusses privacy concerns of youths in the contemporary society exhibiting serious problems with complicated privacy policies, inability to control privacy settings and undue insistence on sharing more information on social networking sites than is necessary with the emergent technologies in AI and big data practices. A semi-structured questionnaire designed for in-depth interviews from university students. Based on the results, educational programs are needed for developing digital media literacy of the population, and regarding the further research, new-age technology impact and different populations' perceptions of privacy should be investigated further.

INTRODUCTION

Artificial Intelligence and Big data are two of the most significant revolutionary innovations of this decade that transform business and people's life. AI, recognized for the capacity to perform tasks usually requiring human intelligence like learning, perceiving, understanding and decision making, has wide usage in different areas including health sector, finance and educational sector (Russell & Norvig, 2021). On the contrary, big data refers to the large amount of data produced through online activities, distinguished by its diversity, velocity and veracity (Mayer-Schönberger & Cukier, 2013).

Though these innovations come with many advantages, they also pose notable risks particularly regarding the data privacy of youth. Data Privacy is the protection of PII, Personally Identifiable belongings to individuals which is the major risk faced by youth digitally (Steve W.,2024). As the user's engagement on these platforms increases, the significant privacy concerns also increase.

Digital platforms such as social media, gaming and educational applications gather personal data. Data collection procedures frequently lack transparency leading to huge data trails that can be utilized for reasons beyond their original aim without user's consent and awareness (Zuboff, 2019). The rationale of the study is to provide comprehensive analysis of threats and challenges that AI and Big data pose to youth privacy.

The research is conducted to explore

RQ1-What main privacy concerns can AI and Big data generate for youth?

RQ2-How do Social Media Platforms impact user's privacy by collecting and analyzing data through AI and big data?

RQ3-What are the privacy challenges youth are facing in the era of AI and big data?

RQ4-What role can digital literacy play in minimizing the privacy risks?

Privacy Risks posed by AI and Big Data

The technological evolution in communication encourages collecting information and dissemination across multiple application domains, extensively increasing big data. Personal data about individuals, whether accessible or masked in data, is vulnerable to the invasion of multiple privacy attacks and the risk of disclosure (Tao W., Zhigao Z., Mubashir R., et al., 2019).

The surveillance potential associated with AI and Big data is the most crucial privacy risk to youth. As seen earlier, AI systems depend on a large amount of data resulting from users' activities, behaviors and online interactions. The constant monitoring leads to the buildup of highly sensitive information of individuals, most of which users might not have awareness of or have not consented to be shared (Zuboff, 2019). Research emphasizes the fact that big data can pose privacy risks to individuals, even if their personal information is only indirectly involved or gathered (Matzner 2014; Raschke et al. 2014; Sa´nchez and Viejo 2017).

Literature concludes that turning youth into products by commodification of personal information is a major privacy risk associated with AI and Big data. Data is frequently considered a valuable asset that can be purchased, sold and exchanged. This practice not only violates privacy but also exploits young users (Livingstone & Sefton-Green, 2016).

Data analytics has undergone a revolution by big data in the contemporary world. Information that was once considered useless and discarded a few years back is now valued highly. Due to the enormous volume, variety and amount of big data, the standard security approaches cannot be applied on it. Big data can hold a large amount of personal information about users, thus at a high risk of security (Anjana G., Nikita C., 2014).

Data Misuse and Exploitation

The term "data misuse" describes the exploitation of personal data. It involves unauthorized or unethical handling of other people's information, which is often gathered without users' consent or agreement. Solove (2008) in his studies highlighted various data misusage practices such as data disclosure to unauthorized people and selling personal data to third parties. It has been pointed out in the literature that data misuse frequently happens in circumstances where the use of data is known to the accountable and the users are unaware of how their data is handled (Nissenbaum, 2010).

Targeted advertising is the most popular type of data exploitation (Zarsky, 2016). Today, big companies gather customers' data to identify changes in trends and make suggestions according to their online behavior. Not all companies ensure data security in their databases. Mismanagement of data during conveyance and storage poses a hazard to social ethics (IbrahimA., Ismail M., 2021).

Isaak and Hanna (2018), in their research on user data privacy demonstrated the Cambridge Analytica Scandal in 2010s, in which the personal data of millions of Facebook users gathered without their consent and later repurposed for political advertising, showed the potential for data abuse on a global scale. The previous literature shows that algorithm bias results in prejudiced profiling, monitoring and targeted advertising which violates users' right to privacy. These biases are privacy invasive for marginalized groups of youths vulnerable to biased algorithms (Eubanks, 2018).

The "filter bubbles" are a direct result of data exploitation in which people are exposed only to the information that makes them more alike and confirms their pre-existing beliefs (Sunstein, 2018).

Youth Perception and Privacy Awareness

As the interaction with digital platforms that rely on Ai and Big data increase, the youth perception of privacy has molded. The survey findings revealed that youth today appreciate privacy; however, they are not clear of what privacy means in the context of these technologies. Many young people's privacy concerns

are rapid and centered on social interactions, such as managing their digital footprint or maintaining who gets access to their social media profiles (Livingstone, 2008; boyd & Marwick, 2011).

Research indicates that youth using social media, search engines, e-commerce sites and other similar applications and websites are ignorant of how much Ai and big data technologies are interwoven into these platforms (Madden et al., 2013). This lack of awareness can further lead to a gap between what they consider private, and their actual online activities, that may compromise privacy (Raynes-Goldie, 2010).

Impact of Digital Literacy on Privacy

Zuboff (2019) in his studies reveals that while the conventional media seek to control and manage us only by their language and visuals, this new form threatens and oversteps much further, by manipulating the fundamental structures of the people's mind. Under this category, it is seen the conspicuous relationship between general user skills and the way they manage their personal information in the online environment. AI and big data coupled with their attributes, which include data collection, analysis and processing as well as making of vital decisions, makes digital literacy even more important than before. Digital literacy involves recognizing of the collection and use of own data knowledge of privacy and the ability to control or adjust settings (Hargittai & Hsieh, 2010).

Digital literacy also influences how people approach and respond to potential threats to privacy from big data and artificial intelligence. Individuals with greater technological proficiency are better equipped to weigh the pros and cons of easy access versus privacy and make informed choices about sharing personal data (Pew Research Center, 2014). Whereas, digital illiteracy would equally lead to low privacy consciousness. People who are unfamiliar with the complexities of the digital environment are less likely to realize that their privacy is being violated. This less awareness can lead to increased exposure to privacy invasion and exploitation (Madden et al., 2013).

Role of social media in privacy

In this case, social media is central to governance of privacy with regards to youth in subject areas of AI and big data. Such platforms are built for the express purpose of sharing and iteration, at the cost of personal privacy. A study suggests that people that use social media; mainly the young ones, often fail to take advantage of the privacy settings, thus accidentally sharing data (Hargittai, 2010).

Furthermore, boyd & Marwick (2011) question the ways in which personal details of users can be collected as well as used by social media firms, can this information be misused and exploited? For example, targeted advertising and content creation as prompted by algorithms exert control over the behavior of users, issues to do with consent become problematic (Zuboff, 2019). Youth want to be accepted by peer groups hence force themselves to reveal information to the public causing them to be at a wrong footing on privacy threats (Livingstone, 2008).

Surveillance Capitalism Theory

Surveillance Capitalism Theory, proposed by Shoshana Zuboff (2019), examines how digital platforms commercialize and sell user data by collecting, analyzing and using it for profit, frequently without users' knowledge or permission. By using artificial intelligence (AI) and big data technologies, this data is processed in order to predict and shape user conduct while being turning it into a valuable asset. The exploitation of personal information in this way undermines serious privacy issues.

In the terms of Privacy, youth are the prime target of data exploitation. Youth create massive volumes of personal data as a frequent user of online platforms and social media, which are constantly gathered by AI-driven systems. With the use of this data, comprehensive profiles are created, allowing businesses to target consumers with ads and change their behavior. It not only exploits users' data without consent but also create a new form of social and economic exploitation (Couldry & Mejias,2019).

The theory offers a path through which different facets of structural power asymmetry in digital spaces may be studied. The issue is compounded by the low levels of education among many of the young people, who often cannot comprehend complex policies on privacy measures, avoid exposure of data or fully grasp how their information is being used. Cohen (2017) believes that the processing of identity data through data mining and data exploitation by AI systems violate privacy rights due to literacy.

This study aims to analyze the elements of the social relations of a new generation that would have to face problems connected with privacy and power relations while experiencing the digital environment through the use of Surveillance Capitalism Theory. The theory was useful in learning how self-organizing technologies that consume large amounts of data erode personal privacy. It can also look at how ordinary and often masked surveillance features incorporated in the AI applications infringe on the privacy of youth. It also draws the light on the predomination of big data and AI and how it has worsened this power differential and made it even more difficult for young people to reclaim agency concerning their information.

Research Methodology

For this study, the researcher used the qualitative research method. Qualitative methodology is the inquiry of analyzing data in the form of opinion, experiences and behaviors. Also, it provides a more comprehensive review of problems, for example data privacy, while involving individual perceptions and interpretations (Creswell & Poth, 2018). Phenomenology research strategy was used in the study to capture the participation and perception that youths have about privacy risks and how they interact with social platforms that integrate AI and big data techniques (Moustakas, 1994). This approach helped the researcher to realize the understanding of youths' privacy, consent, data exploitation, and their handling of privacy issues.

The target population for this study included youth aged between 18 and 25 who were social media and AI and big data technology users. Students from mass communication department of Punjab university was sample to include a gender and diverse range of the population in the research study.

The purposive sampling technique means that the subjects were chosen according to the following criteria: the participants should have some experience in using digital platforms; their awareness of data privacy issues varies. A number of studies have shown that with at least 20-25 participants it is possible to obtain a diverse range of responses that can be considered sufficient for sample (Guest et al., 2006).

Semi structured interviews were conducted in order to get more comprehensive and a qualitative data. The most suitable form of data collection is the semi-structured interviews, because they yielded detailed information which could be comprehensive and flexible allowing of new themes in the course of the interviews (Kvale & Brinkmann, 2009). Open-ended questions were used in the interview so the participant can express their opinions. The interview was conducted through zoom calls and each interview will take about 45 to 60 minutes and recorded by the consent of each participant. Thematic analysis is an approach which is versatile and structured in its approach when it comes to identifying and analyzing patterns or themes in data collected from interviews and/or focus-group discussions mainly qualitative (Braun & Clarke,2006).

Findings and Discussions

Theme: Privacy Risks in the Age of AI and Big Data

RQ1: What main privacy concerns can Ai and Big data generate for youth?

As reported earlier based on structured interviews, youths have raised a number of privacy issues around AI and big data. Some participants have concerns about collection, storage and use of their personal information without their consent. One common idea again and again addressed was privacy violation through artificial intelligence algorithms operationalized by tech companies, specifically social media. One participant put it this way: "It is as if every move one makes on this site is monitored." Such a statement is in line with the increasing concern of youths on the level of scrutiny by algorithms of their virtual identity.

An important issue which participants discussed was the behavioral tracking for advertising purposes. Several people complained about how some adverts that appear in their feed are usually in line with their most recent searches or conversations, it feels like someone is always watching. One participant summed this up by saying, 'I feel like they're always listening', because these adverts are usually tailored to the users' online presence. This goes hand in hand with the existing literature where it is pointed that through surfing the web users build detailed profiles favorable to the surveillance (Andrejevic 2014).

Participants also noted that theft of identity and data loss were equally another major concern they foresaw. Youth are very concerned on issues of identity theft

and loss of name, addresses, and financial details. Having more questions about it, one of the participants said, 'I think my information might be hacked.' They quickly realized, too, there is so much information being saved about me that I don't know how to guard. This fear is not unfounded, as the literature shows that major data breaches have grown common in the age of AI and big data, and now people's information is exposed to exploitation (Solove, 2020).

A majority complained of not having adequate knowledge in privacy settings on these sites and others admitted to leave default settings unaltered; rendering them more vulnerable to other privacy risks. Due to lack of easily understandable and easily visible privacy settings, youth feel that they do not control their personal data which is supported by the literature (Cohen, 2019).

Theme: Social Media Sites as Facilitators of the Erosion of Privacy

RQ2: How do Social Media Platforms impact user's privacy by collecting and analyzing data through Ai and big data?

Most of the participants in the in-depth interviews agreed with the argument that social media sites play a considerable role in compromising privacy. A common concern was the manner in which these platforms harvest large volumes of individual data, with little consent from the users. One participant keenly said, "It's like giving out part of myself without being fully aware of it. I also do not believe they know how much they actually know about me." Such a sentiment emerges with concerns most people have over data collection from their activities on social media platforms.

Targets of specific advertisement also emerged as a major concern of the participants in relation to privacy. One of the interviewed participants said, "I have this kind of feeling that somebody is always observing me. It got rather creepy to note that ad pop us for things I just discussed." Such perception resonates with what the existing literature deems prospective, whereby algorithms monitor individuals' usage habits so that marketing strategies can be relative to the user (Tufekci, 2015). Consequently, participants believed that they were under surveillance while they were online, and indeed experienced increased surveillance panic.

While conducting interviews, another theme that arose was the over sharing on social media. Some participants described how emphasizing or dramatic presentation of an experience was a common procedure when influenced by friends or other people. That is another reason why people share the personal things because everyone shares everything. But when the researcher asked the question what happened to that information. This is further in agreement with Livingstone (2008), who says that the rapid sharing of information through SNS has a tendency of violating privacy features of particularly young people who may lack adequate foresight on eventual repercussions of some of their online activities.

Respondents acknowledged the fact of sharing information that is almost impossible to manage afterward, or to delete. I know I have written things in the past that I don't appreciate being written now, but I know that they are out there. Another interviewee had this to say, 'It's scary to even imagine how that could come back to haunt me.' This fear is well explained by Gonzales (2016) under the conventional digital footprints, meaning that every action one takes online can affect his/her future chances.

Overall, the participants showed considerable concern over data harvesting and use, advertising, sharing too much information and an inability to erase an online footprint.

Theme: Challenges Youth Face in Navigating AI-Driven Privacy Risks

RQ3: What are the privacy challenges youth are facing in the era of Ai and big data?

Specifically, four main issues concerning privacy risks arising from the use of AI and big data among young people were identified through detailed interviews with the youth participants. One of the common trends that could be identified was the opaqueness of so-called privacy policies in social media networks. A number of participants stated that they would like the website to address the common issues that elicit users' frustration and confusion resulting from the lengthy and complex terms of service polices applied to social network sites as well as other online services. Another participant said, "I don't usually get to read such terms

and policies either. They're too long and boring. I simply agree with all the terms stated when I want to use the app." Such opinion reflects a general trend among the young population who often neglect important aspects of privacy and thus, unwittingly agree to the data collection practices (McDonald & Cranor, 2008).

Privacy that was mentioned by the participants as one of the challenges included the challenge of handling different account privacy levels. In our interviews, the participants reported frustration at the sheer number of choices and the absence of clear interactive tools. The youth said: "There are so many settings, I can't even remember what most of them are for". It is very uncomfortable to feel protected when at times I cannot even understand how to protect myself let alone my information. This difficulty in managing privacy settings reiterated findings from literature that suggest that most users are often struggling with adequate skills in the management of their privacy online (Tufekci, 2015).

Participants also pointed towards the speed of technological advancement and the speed at which it happens is always faster than how people adjust themselves. Young people stated feelings of concern about novel technologies and their dangers as a reason for their feelings. "Every time I think that I know everything that is going on, something new pops up and I have to start learning all over again", said one of the participants. Such perception aligns with previous findings that suggest that, due to the high rate of development of artificial intelligence and big data tools, a user put himself under pressure to learn about his/her rights and protections concerning privacy (Kirkpatrick, 2016).

An example of such questions was the one that aimed at determining whether participants recognized the role of peer pressure in their privacy practices, to which they gave their positive answer. Largely, many young people out of pressure went ahead to give out their personal details in order to be accepted in their relevant social groups. As one participant explained it, "Everyone is publishing everything so you feel compelled to do the same." I think about what that might do for my privacy." This behavior supports the Boody and Marwick's (2011) assertion that risky sharing of information among youth caused by social interaction on social media.

Theme: The Role of Digital Literacy in Enhancing Privacy Awareness

RQ4: What role can digital literacy play in minimizing the privacy risks?

A series of interviews with young participants of the project focused on the importance of digital competence in increasing people's privacy literacy and reducing privacy threats posed by AI and big data. Some of the key points were made stating that a better digital competence makes participants aware of necessary knowledge and skills to be informed about various digital fields. A participant said, "The more knowledge I gain of these platforms the more competent I feel in handling issues to do with privacy and even what gets to be informed to the public."

One of the respondents said: "If you had classes that would teach us about privacy and how we can shield ourselves when using the internet, it would be very helpful." This desire correlates with the literature because the integration of digital literacy instruction into content area curricula has been supported as a way to promote synthetic thinking and informed decision making related to privacy (Hollis, 2018).

Also, all participants pointed out that digital competencies do not only increase knowledge but also result in proper use of the Internet. Another participant said: "Understanding the consequence certainly rein in one from posting something on the social networking sites. I think what might be the consequences on me later. This concurs with similar research that suggests that enhanced Internet proficiency reduces vulnerability to privacy intrusions on the Net (Hargittai, 2010).

Limitations

From the numerical standpoint, 62 % of people use the internet across the globe, but this study confines to a region only. These findings may not be transferable to other nations. The data is obtained through purposive sampling and the age limit is restricted. Although this research yields important findings concerning privacy risks that youths encounter in the era of AI and big data, this study has a few weaknesses. First, the sample size was quite small and it may not contain good mix of the youths from different age groups. Furthermore, the sample is based on self-reports of participants from in-depth interviews therefore the participants may not provide

a true and accurate picture of actual behaviors and concerns regarding their online presence. Finally, the use of only qualitative data in this study rules out generalization of study results in a larger population, implying the need for future quantitative study.

Future Directions

Further research prospects of this study concern the enlarging of the study in order to define the correlation between digital literacy and privacy among different groups of young people and in different regions. Such research could give an understanding of how privacy perceptions change with time due to technological enhancements. It is proposed that introducing quantitative methods can help expand the range of analysis and provide a more comprehensive picture of the privacy difficulties faced by youth. In addition, exploring the contribution of other novel technologies in developing privacy threats like artificial intelligence and machine learning forms the core area of interest for subsequent research in the subsequent period.

Conclusion

Finally, it beacons the centrality of privacy concerns to youth in the era of AI and big data; this paper captures important aspects of their views and realities. They complained of the challenges in understanding privacy policies, the challenges of controlling privacy, high penetrative influence of social media and the pressure it puts on them to post private details which may be important to others. However, they do find that such technical competency plays a critical function in developing privacy consciousness among the youthful population to manage such difficulties efficiently. Engaging young people with privacy threats and safe use of the Internet can help them become digital literacy and protect their own data. Thus, it is high time educators and policy makers paid close attention to promoting conscious use of digital environment among student, offering them tools to protect their privacy in the digital environment, taking into account the further development of informatics technologies.

REFERENCES

Andrejevic, M. (2014). Infoglut: How too much information is changing the way we think and know. Routledge.

boyd, d., & Marwick, A. E. (2011). Social privacy in networked publics: Teens' attitudes, practices, and strategies. In Privacy, technology, and youth: A report on the issues and recommendations (pp. 1-23). University of California, Berkeley.

Cohen, J. E. (2017). Law for the platform economy. UC Davis Law Review, 51(1), 133-204.

Cohen, J. E. (2019). Between truth and power: The legal constructions of informational capitalism. Oxford University Press.

Couldry, N., & Mejias, U. A. (2019). The costs of connection: How data is colonizing human life and appropriating it for capitalism. Stanford University Press.

Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). Sage Publications.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Gonzales, A. N. (2016). Digital footprints: Understanding the impact of sharing personal information online. Journal of Computer-Mediated Communication, 21(4), 210-225.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. Field Methods, 18(1), 59–82.

Hargittai, E. (2010). Digital na(t)ives? Variation in Internet skills and uses among members of the 'Net Generation'. Sociological Inquiry, 80(1), 92-113.

Hargittai, E., & Hsieh, Y. P. (2010). Digital literacy and privacy management. New Media & Society, 12(4), 737-754.

Hollis, V. (2018). The role of digital literacy in protecting online privacy: Implications for educational policy. Journal of Educational Technology Systems, 47(2), 232-249.

Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer, 51(8), 56-59.

Kirkpatrick, D. (2016). The Facebook effect: The inside story of the company that is connecting the world. Simon & Schuster.

Kvale, S., & Brinkmann, S. (2009). Interviews: Learning the craft of qualitative research interviewing (2nd ed.). Sage Publications.

Livingstone, S. (2008). Taking risky opportunities in youthful content creation: Teenagers' use of social networking sites for intimacy, privacy, and self-expression. New Media & Society, 10(3), 393-411.

Livingstone, S., & Sefton-Green, J. (2016). The class: Living and learning in the digital age. New York University Press.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., & Beaton, M. (2013). Teens, social media, and privacy. Pew Research Center, 21(1055), 2-86.

McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. I/S: A Journal of Law and Policy for the Information Society, 4(3), 543-568.

Moustakas, C. (1994). Phenomenological research methods. Sage Publications.

Nissenbaum, H. (2011). Privacy in context: Technology, policy, and the integrity of social life. Journal of Information Policy, 1, 149-151.

Raschke, R., Krishen, A. S., & Kachroo, P. (2014). Understanding the components of information privacy threats for location-based services. Journal of Information Systems, 28, 227–242.

Raynes-Goldie, K. (2010). Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. First Monday.

Sanchez, D., & Viejo, A. (2017). Personalized privacy in open data sharing scenarios. Online Information Review, 41(3), 298–310.

Solove, D. J. (2008). The new vulnerability: Data security and personal information.

Solove, D. J. (2020). Understanding privacy. Harvard University Press.

Sunstein, C. (2018). #Republic: Divided democracy in the age of social media. Princeton University Press.

Tufekci, Z. (2015). Algorithmic sociality: How social media shapes our lives and relationships. Social Media + Society, 1(2), 1-2.

Wang, T., Zheng, Z., Rehmani, M. H., Yao, S., & Huo, Z. (2019). Privacy preservation in big data from the communication perspective—a survey. IEEE Communications Surveys & Tutorials, 21(1), 753-778.

Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. Science, Technology, & Human Values, 41(1), 118-132.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.