International Journal of Sciences and Innovation Engineering
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: ***www.ijsci.com***
Volume 03, Issue 10, October 2026
E-ISSN : 3049-0251
DOI: https://doi.org/10.70849/ijsci03102619825

A Study on Youth's Awareness and Understanding of Data Privacy While Using AI Chatbots in Maharashtra

Bhakti Rindhe[1] , Samruddhi Takalkar[2] ,Shweta Mohandule[3] , Vaishnavi Kate[4] , Dr. Vijay Dhawale[5]

[1,2,3,4] Students, Department of MCA, D.Y. Patil Institute of Master of Computer Applications and Management, Akurdi, Pune.

[5] Associate Professor, Department of MCA, D.Y. Patil Institute of Master of Computer Applications and Management, Akurdi, Pune.

Article Info

Article History:

Published: 11 Oct 2026

Publication Issue:
Volume 03, Issue 10
October 2026

Page Number:
435-446

Corresponding Author:
Bhakti Rindhe

Abstract:

The extensive usage of AI chatbots, such as ChatGPT and Google Gemini, has led to a complete transformation of the ways young people acquire information, conduct their academic assignments, and perform technical tasks. However, with the rise in popularity of these platforms, worries emerge regarding the protection of users' information while exchanging it on these sites. The purpose of this study is to identify the degree of awareness regarding the problem of data privacy among young people who use AI chatbots in the state of Maharashtra. The descriptive approach was applied in the research process, and primary data were collected through the use of Google Forms surveys. The sample consisted of 151 respondents aged 15-29 who utilized AI chatbots and exchanged personal information, privacy measures, risks, knowledge about the storage and training of the model, and data protection laws. Some of the AI systems that were in common usage include ChatGPT and Google Gemini, while the type of information that was in widespread sharing included academic information. However, there is one other important observation on the lack of privacy settings, the lack of awareness of storing conversations, and also the limited self-reported awareness of the Indian Digital Personal Data Protection Act, 2023 or similar data laws. The findings of this research suggest the need for proper digital privacy education, clear privacy information in different languages, proper data sharing, and awareness programs.

Keywords: AI Chatbot, Data Privacy, Digital Literacy, DPDP Act 2023, Generative Artificial Intelligence, Young Generation Awareness

1. Introduction

The use of the consumer-oriented AI models that gained popularity among the masses is responsible for the implementation of generative AI models, like ChatGPT and Google Gemini, by the youth of Maharashtra in their educational and technical aspects [21]. The changeover from the traditional AI models to the contemporary conversational AI models has an impact on the learning process, communication, and technical work of the youth. Selection of Maharashtra as the region for carrying out research work is justified by the high level of literacy and favorable conditions in the sphere of education and technology. However, this does not imply that the high level of literacy means the awareness of data privacy and AI models [24].

The goal of this research study is to address the problem of discrepancy between the perceived privacy of the user and interaction with the artificial intelligence interface. The following are some of the issues experienced by the users because of the privacy paradox since the user is expected to give out all types of information regardless of the fact [8]. Convenience provided by the artificial intelligence system makes the user oblivious of privacy concerns.

Copyright © 2026 The Author(s). CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/)

This research aims at conducting an empirical research to enable the development of a custom intervention to enhance digital literacy. This research focuses on assessing self-reported awareness of the Digital Personal Data Protection (DPDP) Act, 2023 in India and similar data laws, alongside broader data-privacy awareness [23]. This research will entail analysis of the type of information shared by youths, AI technology used, data retention and training, privacy practices, risks and awareness of the DPDP Act, 2023.

1.1 Statement Of Problem

Even without having adequate knowledge of the privacy of data in AI technology, there is a positive attitude towards the use of AI technology in the state of Maharashtra. Even though the young people might be aware of the technology that can be used in the Internet, they do not know how to go about the prompts and uploads in the AI technology [6]. The problem related to data, source code, personal data, and contextual data forms an extremely important issue [15]. The shortage of time for doing the academic task can make the person adopt an easy method without caring for privacy [8]. The problem related to privacy policies being in English can also be an issue [16],[17].

1.2 Objectives of the Research

1.To assess the awareness, understanding, and misconceptions of data privacy and digital safety among youth using AI platforms in Maharashtra, including their self-reported awareness of India's DPDP Act, 2023 or similar data laws.

2.To examine the relationship between educational background, sources of digital-safety knowledge, and actual privacy-related behaviour, including the use of chat-history controls, data-sharing practices, and privacy settings while using AI tools.

3.To identify gaps in AI data-privacy practices and develop evidence-based recommendations for improving digital literacy, privacy awareness, and institutional safety guidance among youth.

1.3 Hypothesis of the Study

H0 (Null Hypothesis): No significant difference in the practices of data privacy hygiene and risk awareness behavior among youths with different educational disciplines in Maharashtra; the current knowledge of the consumer is adequate enough to safeguard themselves in the process of conversational AI.

H1 (Alternative Hypothesis): Differences exist in data privacy security practices among people with different educational disciplines in Maharashtra, which involve knowledge gaps and deadline privacy paradox.

2. Literature Review

Previous studies that focus on the relations between youth and conversational AI can be analyzed from the perspectives of governance/policy, education/awareness, socio-cultural behavior, and technological risks to privacy. According to the authors of the source paper, the problem lies in the difference between the simple interface for the consumer and complex data gathering and processing techniques.

Governance and Policy Perspective: In Ng et al. (2026), the authors examine the policies and guidelines in universities concerning the privacy issues that arise due to generative AI. In Wang et al. (2025), the authors examine privacy issues and their implications in universities [1],[2],[17],[22]. It should be noted that the scope of the research includes the DPDP Act, 2023 in India and its implementation [23],[24].

Awareness and Educational Campaigns: Shah and Mehta (2026) pay attention to the awareness of students concerning the ethical problems of AI usage and stress the significance of educational campaigns [5]. Sebastian (2023) highlights the problem of privacy and data protection of ChatGPT and calls for user awareness [3]. As noted by the author, the role of formal education provided by universities is crucial for translating awareness into action [14],[19],[20],[21].

Sociocultural Issues: The privacy paradox is one of the behavioral issues addressed in the source research. Although the concerns of people in terms of corporate surveillance of personal data are rather high, people manage to perform risky behavior because of its immediate benefits [10],[8]. The time limit of performing the task or difficult formulation of privacy agreements may be connected to the privacy paradox [16].

Globalization, Media and Technology: Marchal et al. (2024) classify the abuse of generative AI and privacy practices [4]. Other literature referenced in the source bibliography focuses on digital literacy, vulnerabilities of information, data storage, and platform exposure to privacy [7], [9], [13]. This provides context for the study of how routine AI uses may lead to privacy exposure despite lack of malevolent intention.

2.1 Identified Gaps and Relevance to the Study

The most recent literature review conducted for the purpose of data privacy mainly targets Western scholars and world wide databases [1], [2], [18]; however, they do not offer sufficient attention to the socio-economic and linguistic background of the younger generation of Maharashtra [14], [15]. Moreover, the recent research mainly targets the issue of adoption of AI technology and the issue of privacy in isolation; however, there is no research that considers both the issues together. Only very few studies have been conducted to understand the effects of academic pressure and stress in the face of deadlines on the younger generation facing a long and complicated privacy agreement in English language [16]. Consequently, they are left with no choice but to sign a contract without even understanding what it means.

3. Research Methodology

This study uses descriptive research methodology to analyze the knowledge and understanding of youth regarding data privacy while using AI chatbots in Maharashtra. The primary and secondary sources of data have been considered.

3.1 Research Design:

In this study, there was an application of the descriptive cross-sectional design which was aimed at observing the current behavior of the youth in terms of their privacy online without causing any disturbance to the user process experience.

3.2 Primary Data Collection:

An organized questionnaire was made using Google Forms. The survey was aimed at youth ranging from the ages of 15 to 29 years residing in Maharashtra and consisted of questions on demographic variables, use of AI tools, information disclosed to chatbots, views regarding privacy, and privacy settings.

3.3 Secondary Data Collection:

The following sources that are employed in the purpose of research have been selected from a number of sources including law and academic sources. These include the book on privacy of generative AI, digital literacy, and Digital Personal Data Protection Act, 2023 [23].

1. Data Analysis:

Google Forms has been used for obtaining frequency and percentage distribution. The findings of the survey have been represented in a pie chart and bar graph. The study has been conducted keeping in mind the trends of usage of AI tools, information sharing, privacy settings, risk perception, and knowledge about data protection laws. For questions that allowed multiple selections, the percentages are not mutually exclusive and may add up to more than 100%.

3.5 Sample Size: A total of 151 survey responses were included in the analysis. Unless otherwise stated, the reported percentages are based on 151 respondents. For questions that allowed multiple selections, percentages are not mutually exclusive and may add up to more than 100%.

List of responses are available on given link : List Of Responses

4. Results

The survey results show how participants use conversational AI and report privacy-related practices, awareness, and risk perceptions. The following subsections present the findings. For questions that allowed multiple selections, the percentages are not mutually exclusive and may add up to more than 100%.

4.1 AI Chatbot Usage

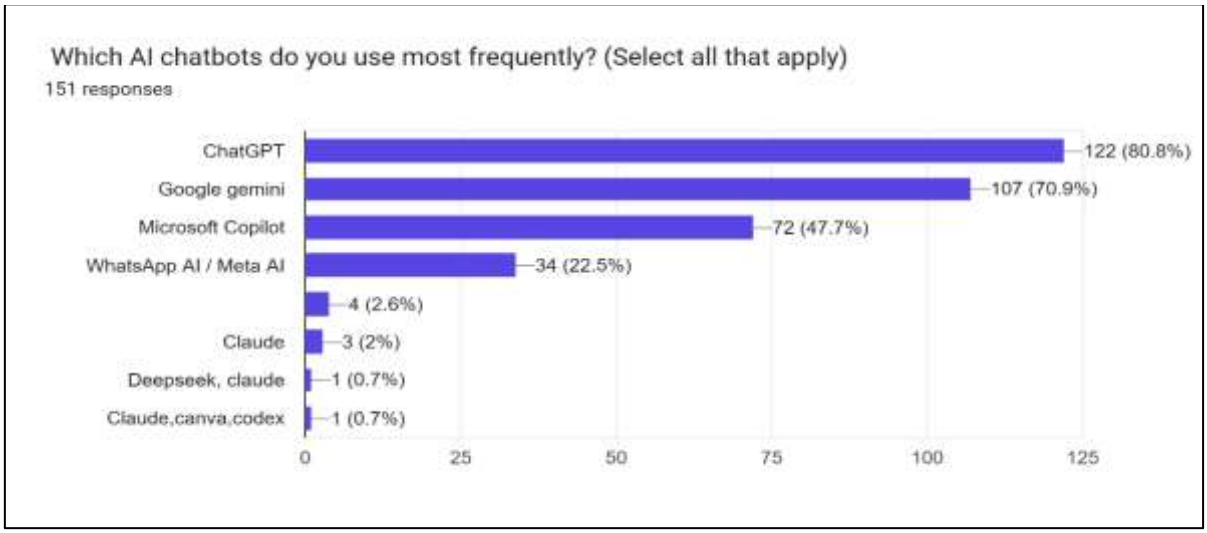


Fig. 1. AI chatbots selected by respondents.

Popular AI chatbots selected by respondents included ChatGPT, Google Gemini, and Microsoft Copilot. Because respondents could select all options that applied, the number of selections may exceed the total of 151 respondents.

4.2 Age Distribution

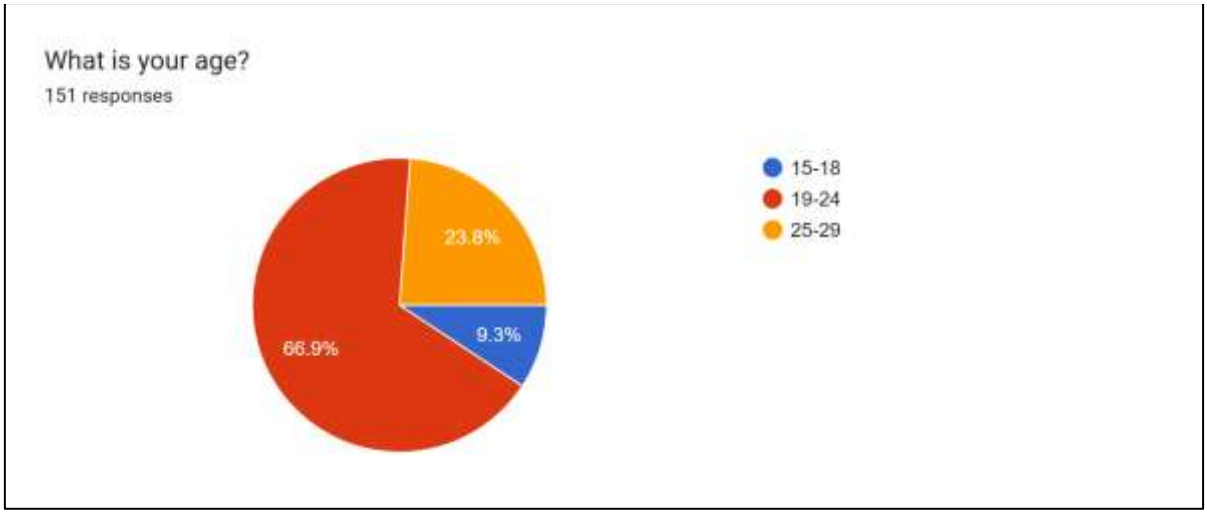


Fig. 2. Age distribution of survey respondents.

As per the outcomes of the analysis of the survey results, the most populous age group is that of 19-24 years (66.9%), followed by the age group of 25-29 years (23.8%) and 15-18 years (9.3%).

4.3 Types of Personal Information Shared

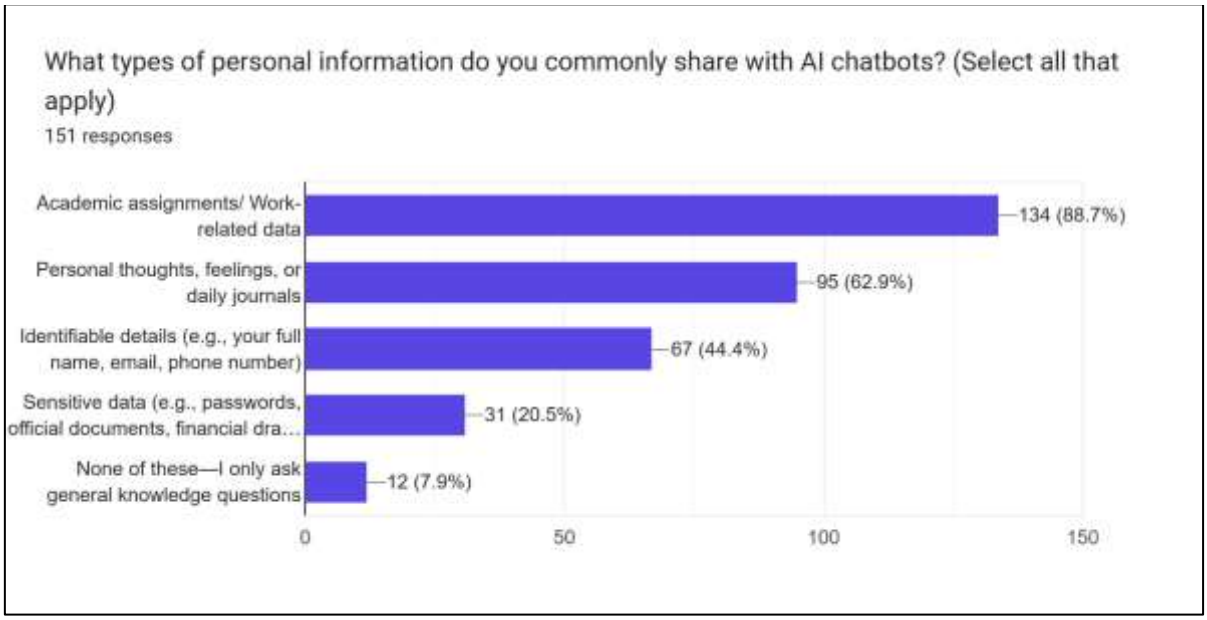


Fig. 3. Types of personal information commonly shared with AI chatbots.

Academic assignments/work-related data were reported by 88.7% of respondents. Personal thoughts, feelings, or daily journals were selected by 62.9%, identifiable details by 44.4%, and sensitive data by 20.5%. Only 7.9% selected "None of these—I only ask general knowledge questions." Since this was a multiple-selection question, responses are not mutually exclusive.

4.4 Privacy-Setting Behaviour

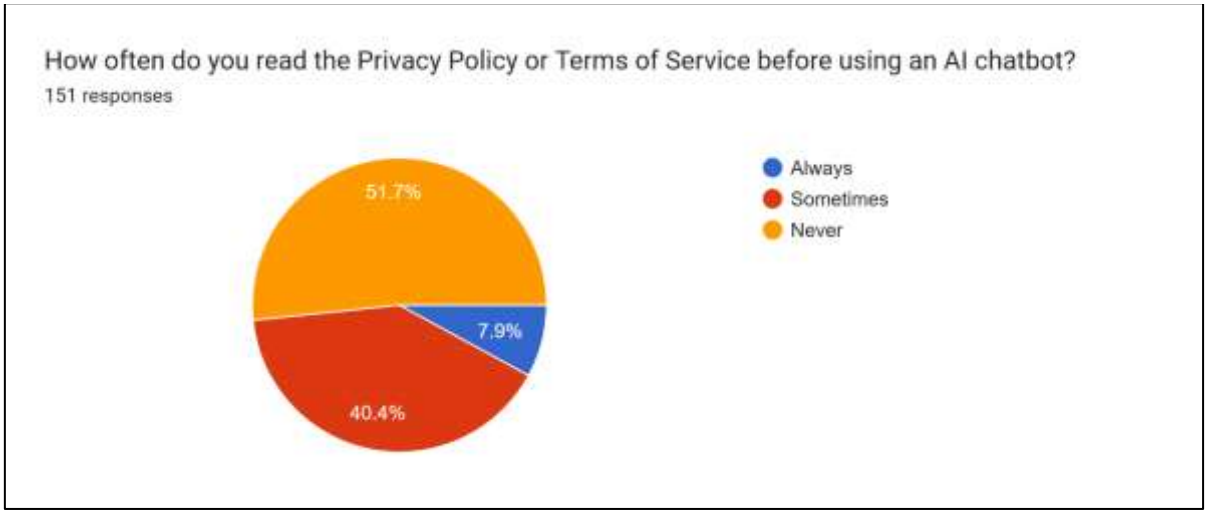


Fig. 4. Frequency of clearing chat history or turning off chat-history tracking.

The largest group of respondents (approximately 55.0%) selected "Never" in response to the combined question about clearing chat history or turning off chat-history tracking. A further 23.2% selected "Occasionally," 6.0% selected "Yes, regularly," and 15.9% reported that they did not know how to perform these actions. Because the survey combined two activities in one question, the results cannot distinguish between clearing chat history and turning off tracking.

4.5 Perception of Confidentiality

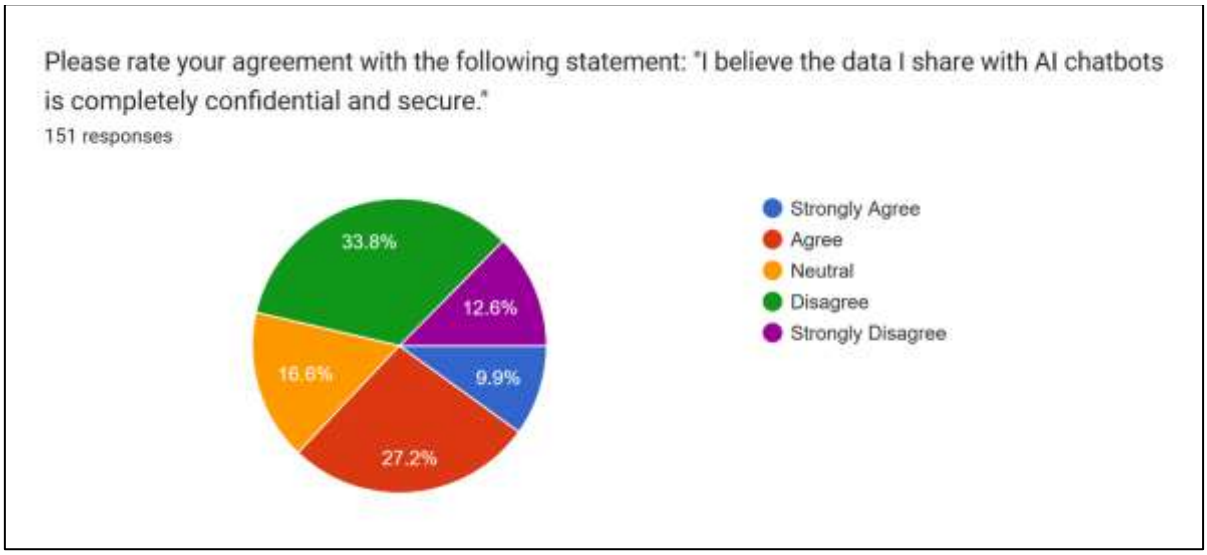


Fig. 5. Agreement with the statement that data shared with AI chatbots is completely confidential and secure.

The results are indicative of confusion and fears about privacy. The dominant response to the question is "Disagree," comprising 33.8%; while "Agree" is 27.2%, and "Neutral" is 16.6%. These responses indicate that there is no consistency among the users with regard to the degree of confidentiality of their AI interactions.

4.6 Awareness of Data Storage and Model Training

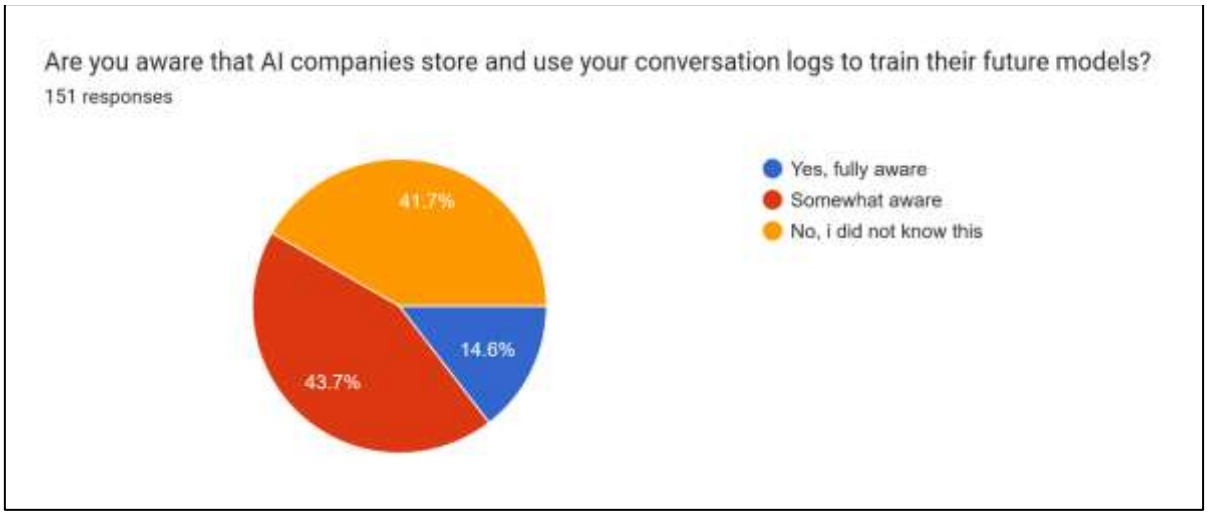


Fig. 6. Respondents' awareness of conversation storage and possible model training.

Just 14.6% said they were fully aware, whereas 43.7% were somewhat aware and 41.7% did not know about the issue. Therefore, over four-fifths of respondents were not fully aware. This finding corresponds with previous research on data collection and storage on platforms [12], [13]. The percentages reflect respondents' self-reported awareness in response to the survey wording; they do not establish that all AI companies store or use every conversation for model training.

4.7 Use of Fake Details

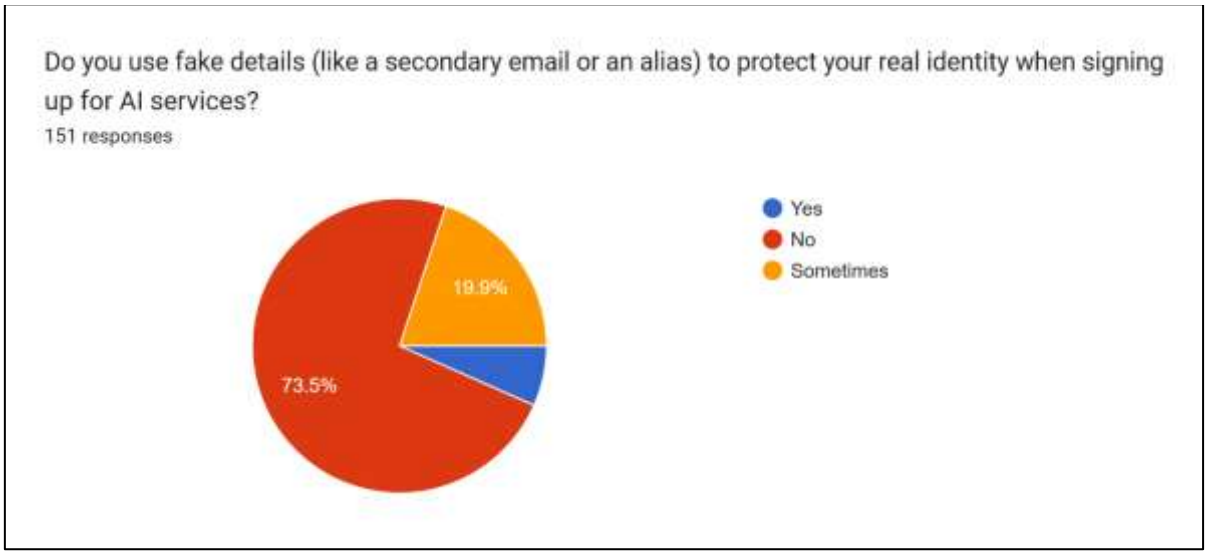


Fig. 7. Use of alternative details to protect identity when signing up for AI services.

Most respondents (73.5%) reported that they did not use alternative details, such as a secondary email address or an alias, to protect their identity when signing up for AI services. A further 19.9% reported doing so sometimes, while 6.6% reported doing so. These findings suggest that most respondents did not regularly use this particular method of identity protection.

4.8 Perceived Data Privacy Risks

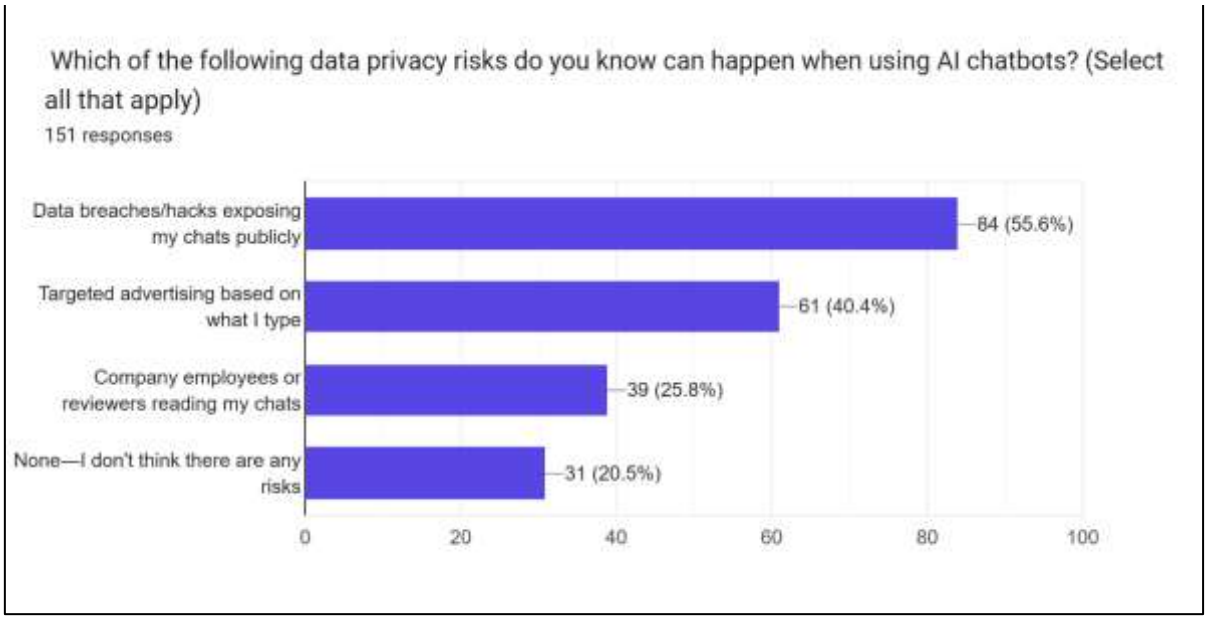


Fig. 8. Data privacy risks respondents believe may occur when using AI chatbots.

Data breaches or hackers exposing chat logs were selected by 55.6% of respondents, making this the most frequently selected risk. Targeted advertising based on user input was selected by 40.4%, while company employee or reviewer access to chats was selected by 25.8%. Approximately 20.5% selected the option indicating that they did not believe there were risks when using AI chatbots. The questionnaire did not provide a separate “unsure” option for this item. Because respondents could select multiple options, these percentages are not mutually exclusive.

4.9 Awareness of the DPDP Act

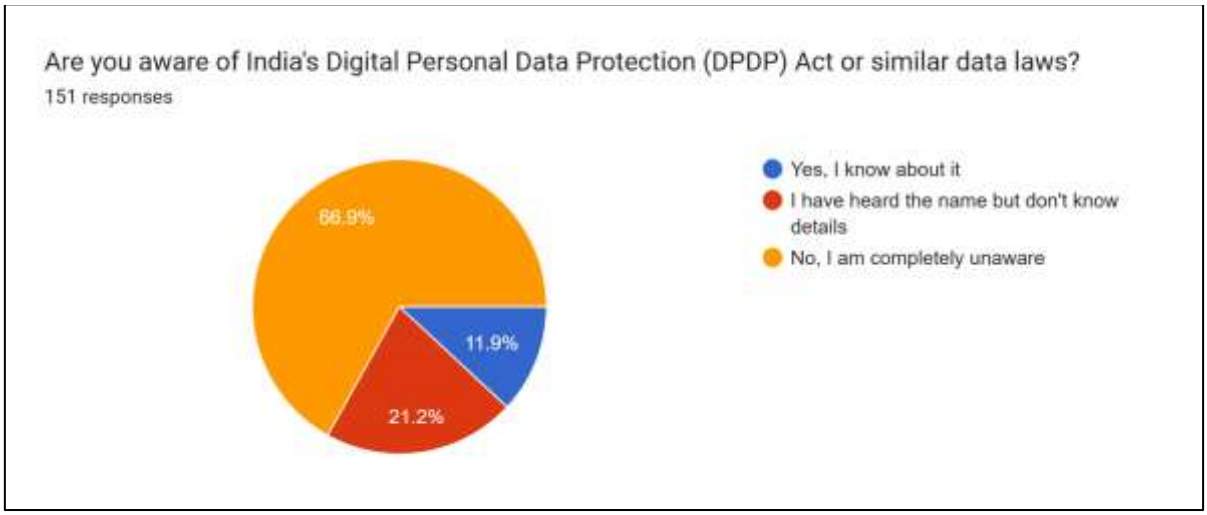


Fig. 9. Respondents' self-reported awareness of India's DPDP Act or similar data laws.

A total of 11.9% of respondents reported knowing about the DPDP Act or similar data laws, while 21.2% had heard the name but did not know the details, and 66.9% reported being completely unaware. Because the survey question referred to the DPDP Act or similar laws, these results cannot establish awareness of the DPDP Act alone. The findings nevertheless indicate a need for legal and digital privacy education among youth.

4.10 Preferred Ways to Improve Awareness

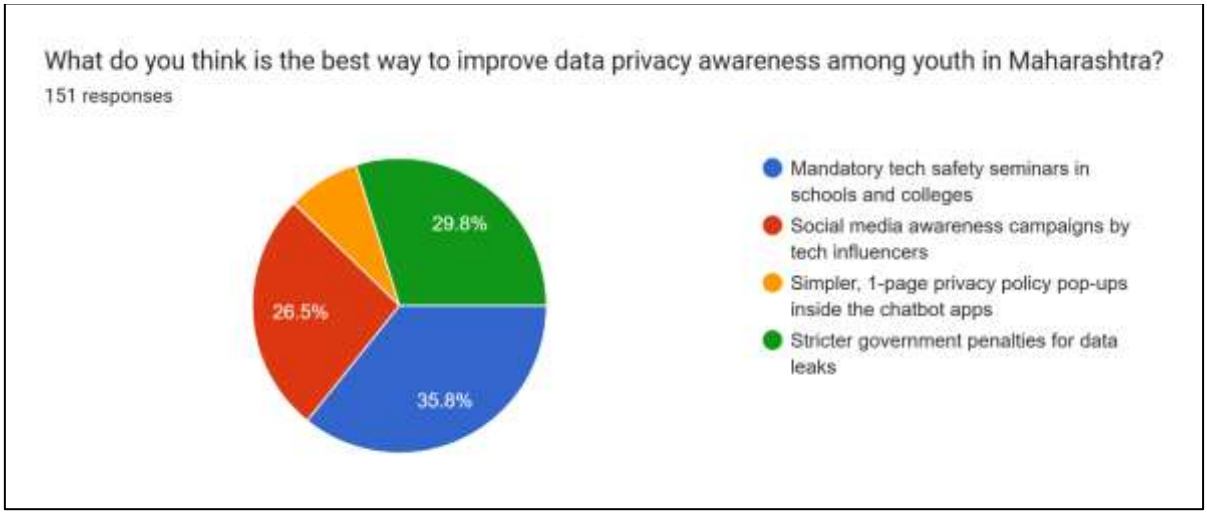


Fig. 10. Respondents' preferred approaches for improving data-privacy awareness among youth in Maharashtra.

The maximum amount (35.8%) was allocated for compulsory training for technology safety [19], [20], and penalties/supervision accounted for 29.8%, while the campus awareness program covered 26.5%, and one-page privacy policy was 7.9%.

4.11 Sources of Data Privacy Information



Fig. 11. Sources from which respondents learned about data-privacy rules and AI-related risks.

Social media was selected by 40.4% of participants, technology blogs/news articles/podcasts by 19.2%, school or college lectures/workshops by 21.9%, and friends or family conversations by 15.2%. In addition, 48.3% selected the option indicating that they had never actively received or found information about data privacy. Because the question allowed multiple selections, this option could be selected alongside a source of information; therefore, this result should be interpreted cautiously. Media coverage can also play a role in shaping privacy awareness [18].

5. Discussion

The conclusion from the survey data shows that the difference does exist in the degree of awareness and use of AI chatbots in the youth generation in the State of Maharashtra [1], [2]. When it comes to the usage level of the technology, it is rather high, although some discrepancy can be found with respect to the privacy issue [6].

The findings on the use of alternative details indicate that most respondents did not use secondary email addresses or aliases to protect their identities when signing up for AI services (Fig. 7). This result concerns self-reported privacy-protection behaviour and should not be interpreted as evidence that respondents used false identities to spread misinformation [14], [22]. The law-awareness results also require cautious interpretation because the survey asked about the DPDP Act or similar data laws (Fig. 9), rather than the DPDP Act alone.

It is necessary to bear in mind that the realization of the result that has been reached with the help of setting the privacy settings should be evaluated. Firstly, the majority of respondents do not want to delete logs and manage the process of their creation. Secondly, there are some respondents who simply do not know how to delete the logs [16]. It is necessary to emphasize the lack of information regarding the use of the logs by artificial intelligence organizations [12], [13].Another issue evident from the results is limited self-reported awareness of data protection laws. More than two-thirds of respondents (66.9%) reported being completely unaware of the DPDP Act or similar data laws, according to the wording of the survey question (Fig. 9).

However, while there is the fear of invasion of privacy among the current generation, it might be possible that the consequences of their actions will not be known to them [17]. It is absolutely essential for organizations to create a specific digital literacy program to educate individuals on data privacy [14], [19], [16].

The importance of the DPDP Act, 2023 underlines the need to address this awareness gap [23]. In the survey, 11.9% of respondents reported knowing about the DPDP Act or similar data laws, 21.2% had heard the name but did not know the details, and 66.9% reported being completely unaware. Because the question included similar laws, these results do not measure awareness of the DPDP Act alone.

From the perceived risk research results, it is clear that the respondents understand the danger of the data breach; nevertheless, awareness regarding the use of the data is inconsistent [11],[12]. Therefore, awareness programs have to

do much more than simply creating awareness [19],[20] but should rather educate about more practical issues such as data storage [12],[13], inspection, selective advertising, model improvement, and instant response [22].

5.1 Implications of the Study

For Software Developers: There must be privacy notices. Other languages and comprehensible information could help users make their decision [16], [22].

For Educational Institutions: The education institutions should offer information on privacy protection of AI in digital literacy or technology course curriculum [18], [19], [21]. They should also educate the students about the importance of deleting any unnecessary personal data from the prompt and privacy control before performing the task via AI.

For Policy makers/Government: The Digital Personal Data Protection Act, 2023 [23] should be simplified for students [17].

6. Conclusion

In the current research paper, the discussion will focus on the research carried out to examine the awareness and knowledge of the data privacy issues due to the use of the AI chatbot technology among the young people in Maharashtra. It can be seen that there would be a difference between the two variables since one variable would have an increase in the use of the technology while the other will be the knowledge of the data privacy issues.

The findings indicate uncertainty about conversation storage and model training, as well as limited self-reported awareness of the Digital Personal Data Protection Act, 2023 or similar data laws, as measured by the survey question. According to the findings, privacy practices may require more than reminders alone; young adults may also need assistance in minimizing shared data and understanding privacy threats.

Thus, the need arises for the educational sector, tech firms, and policymakers to work towards creating programs on digital literacy, privacy, and security which meet the requirements of the users. This would ensure that children become digital citizens instead of mere consumers of technology.

7. Limitations and Future Research

In this particular case, concerning the choice of the type of online surveying for the research at hand, the sample will be limited only to those people who can perform surveys through certain surveying devices. Taking into account the fact that, as it has already been stated above, in accordance with the sources presented, youth participation coming from rural areas, which have low levels of smartphone penetration rates, can be considered under-represented [15]. As far as future research in this area goes, there are several methods that can be used, namely, stratified sampling, technical and non-technical programs, longitudinal surveys both prior to and after raising awareness programs about digital privacy among young people [20]. Another direction in future research could be pop-ups when communicating sensitive information to AI Chatbot. Several questionnaire-related limitations should also be noted. The question about conversation storage and model training was worded as a general statement about AI companies; responses therefore reflect participants' self-reported awareness in response to that wording and do not establish that all providers store or use every conversation for model training. The question about sources of privacy information allowed multiple selections while also offering an option stating that respondents had never received such information, which could conflict with selecting a source. The questionnaire also asked how often respondents read a privacy policy or Terms of Service, but this item is not presented as a separate result in this manuscript. Future surveys should use neutral wording, ensure response options are consistent, and report findings for all analysed questionnaire items.

Acknowledgements

The authors would like to thank D.Y. Patil Institute of MCA & Management, Pune for providing us with all the facilities required to conduct this research.

Conflict of Interest

The authors would like to inform that there is no conflict of interest with regard to the research findings and their publication.

Authors' Biographies

The authors Bhakti Anil Rindhe, Samruddhi Santosh Takalkar, Shweta Lahu Mohandule, and Vaishnavi Namdev Kate are pursuing their MCA degree from D.Y. Patil Institute of MCA & Management, Pune. Dr. Vijay Dhawale is Associate Professor at D.Y. Patil Institute of MCA & Management, Pune. The interests of the authors include Data Science, Technological Advancements, and Social Sciences.

References

[1] B. Y. Ng et al., "Understanding how university guidelines address privacy and security issues of generative AI in academic settings," Proceedings on Privacy Enhancing Technologies, vol. 2026, no. 4, pp. 1055–1071, 2026.

[2] C. Wang et al., "University students' privacy concerns towards generative artificial intelligence: Causes and educational implications," Journal of Educational Computing Research, vol. 63, no. 2, pp. 210–235, 2025.

[3] G. Sebastian, "Privacy and data protection in ChatGPT," International Journal of Security and Privacy in Pervasive Computing, vol. 15, no. 1, pp. 1–14, 2023.

[4] N. Marchal et al., "Generative AI misuse: A taxonomy of tactics and insights from real-world data," arXiv:2406.13843, 2024

[5] P. Shah and R. Mehta, "Awareness of students regarding ethical use of AI," Asian Journal of Distance Education, vol. 21, no. 1, pp. 45–58, 2026.

[6] J. Das, C. DiRienzo, and D. Lanier Jr., "Corporate Compliance Data Networks and User Privacy Risks," Journal of Information Security, vol. 12, no. 2, pp. 88–96, 2024.

[7] J. Singh, A Study of Digital Literacy and Information Vulnerabilities in India, TechPress, 2025.

[8] A. Adhikary, User Interaction Hazards and Academic Deadline Pressures, Academic Press, 2024.

[9] N. Prakash and K. C. Vadlamati, "Conversational Interface Ingestion Profiles and Privacy Exposure," Security Review, 2025.

[10] R. Anupam, S. Kaur, and R. Singh, "An Empirical Analysis of Information Security in an Era of Globalization," International Journal of Technology, vol. 14, no. 3, pp. 202–211, 2025.

[11] V. Gyuracz, "Investigative Journalism's Role in Platform Privacy Awareness," Media and Technology Journal, vol. 22, no. 1, pp. 45–56, 2024.

[12] Two Six Technologies (TST) & University of California, Los Angeles (UCLA), "Estimating Data Asset Accumulation and Retention Cycles in Consumer Platforms," technical report, 2025.

[13] R. M. Kamble, "Data Retention Loops and User Privacy: Issues and Perspectives," Online International Interdisciplinary Research Journal, vol. 14, no. 5, pp. 112–120, 2024.

[14] S. R. G. Krishnan, "Reducing Digital Privacy Violations in India: The Role of Information Security Education," Journal of Technical Education and Social Practice, vol. 8, no. 2, pp. 134–142, 2025.

[15] N. Upadhyay, I. Paul, A. Vashistha, K. Bhatnagar, S. Kumar, and A. Gupta, Report on Digital Infrastructure Exposure in India, Symbiosis Law School Press, 2024.

[16] S. Lenka, Understanding the Relation between Platform Mining Patterns, Language Barriers, and Data Rights, Technical Law Publishers, 2024.

[17] A. Bhagat and S. Yea, "Towards Data Sovereignty Territories: Consent Frameworks and Platform Governance Constraints," Policy and Practice Review, vol. 11, no. 4, pp. 310–319, 2025.

[18] L. Hill and M. K. Hefner, "Successes and Challenges in Implementing a Data Privacy Screening Tool in Regional Higher Education Networks," Information Management Report, vol. 19, no. 2, pp. 77–85, 2025.

[19] D. M. Das et al., "Digital Safety Literacy Education — Pilot Study Integrated in Computing Curriculum," Educational Security Journal, vol. 16, no. 1, pp. 40–49, 2025.

[20] H. E. Fraley et al., "Systematic Review of Educational Interventions for Academic Networks on Algorithmic Privacy Risks," Journal of Information Security Literacy, vol. 10, no. 3, pp. 155–164, 2026.

[21] F. Miao and W. Holmes, Guidance for Generative AI in Education and Research. Paris, France: UNESCO, 2023.

[22] National Institute of Standards and Technology, "Cybersecurity, Privacy, and AI," NIST, 2026.

[23] Ministry of Law and Justice, Government of India, The Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023), The Gazette of India, 2023.

[24] Ministry of Electronics and Information Technology (MeitY), Information Technology Infrastructure Security Data – Maharashtra Region Profile, Government of India, 2024.