

Analyzing Data Privacy Concerns in Young Adults – Apprehensions of Engineering Students

Sakshi Kulkarni¹, Krutika Konde² and Dr. Mangesh Bedekar³

¹⁻³School of Computer Engineering & Technology, Dr. Viswanath Karad MIT World Peace University, Pune, India
Email: kulkarnisakshi26@gmail.com, krutika.konde@gmail.com, mangesh.bedekar@mitwpu.edu.in

Abstract— Personalization technologies can be a useful tool for improving user experience across a wide range of applications, but they also raise new privacy concerns. While the volume of personal data acquired by applications has continuously increased, user awareness of the ramifications has not followed the same trend. A Survey was conducted to better understand users' perceptions of the matter and their opinions in order to make better privacy-related decisions. The study focuses on topics such as user awareness of personalization and privacy, end user preferences for personalization and privacy, data misuse awareness, privacy awareness about continued usage of applications, and willingness to make better choices for enhanced privacy. The target demographic is between the ages of 19 to 22, sometimes known as "young adults."

Index Terms— Personalization, Privacy, Privacy Paradox, Privacy Policies, End User's License Agreements.

I. INTRODUCTION

As Internet usage has increased over the years, so has the importance of data privacy. The ability of a user to determine when, how, and to what extent personal data about them is shared with or expressed to others is referred to as data privacy. Data privacy relates to how data should be gathered, stored, managed, and shared with third parties as well as how existing privacy legislation should be implemented such as the California Consumer Privacy Act (CCPA) or General Data Protection Regulation (GDPR). Technology has the ability to know us better than we know ourselves, allowing it to give the best possible experience and elevate user's lives to new heights. Though this is an advantage of the technology, it is proving to be damaging the data privacy of end users [4, 5]. Personalization refers to a system that explicitly uses user information without their knowledge [7]. Often the end user may get confused between the idea of personalization and customization, although they may seem similar at user experience level. Personalization is the process of producing or customizing an application based on user's data to satisfy the demands of a specific user. When a user needs manual modifications to an application to match their needs or expectations, this is known as customization. Simply being online or using a mobile application generates a massive quantity of data, which is stored as digital footprints. Companies that employ technology to track end users exploit digital footprints as open resources. Most e-commerce sites, utilize cookies to retain users' browsing histories, which are subsequently used to track users and provide personalized offers depending on the information gathered. Although personalization provides significant benefits to end users, it does demand the disclosure of sensitive personal information. When users

discover that shopping lists are personalized based on their browsing and purchasing history, they may feel their privacy has been invaded. Data gathering, sharing, and usage are key points to the success and expansion of these billion-dollar firms, which rely primarily on advertising for revenue. Behavioral data, preference data, usage data, and other forms of activities are all examined, analyzed, and used to not only recruit and retain users, but also to keep them engaged for extended periods of time in order to monetize their activities through hyper-targeted advertisements.

Generation Z, those born between 1997 and now, are today's young adults. They have never known a world without the Internet, and they have always had access to all social media sites since they have always had internet and Wi-Fi at home. The Internet serves as the primary means for all of their daily activities, such as communication, connection, entertainment, online shopping, and online classes and homework, caused by the COVID-19 Pandemic situation. [3, 4] Privacy-related issues are surfacing to the point that we users at risk of losing control their data. The way a recommender handles privacy determines whether its users see it as a benefit or as a disadvantage (invasion of privacy) [10]. If the recommender solely utilizes this information to help users identify products to buy on an e-commerce portal, user will probably appreciate the feature and return to shop again. On the other hand, if the e-commerce portal sells user information to other companies, they land up bothering users [8].

II. LITERATURE SURVEY

The battle between digitization and privacy has recently become more crucial. The definition of privacy has fundamentally changed as a result of digitization. Previously, personal information clearly belonged to the individual. Anyone else who desired it had to go out of their way to get it, whether lawfully or unlawfully, by asking or looking. Users may benefit from a range of tailored recommendations, including areas of interest but this comes at the cost of user's privacy [11]. Personalization systems, on the other hand, can coexist peacefully alongside privacy preserving systems. Personalization can help privacy preserving systems safeguard a user's privacy in the way that user wants it to be protected. Users' viewpoints on how their privacy should be protected, as well as what data they choose to share with other users and what data they choose to keep private, should be addressed. Personalization has the potential to play a significant role in enforcing the user's privacy demands [1]. Personalization of a service requires the collection of a massive data. It must record online behavior of a user as frequently, as quickly, and as efficiently as possible. Services that are tailored to a user must also be learned [11]. According to a recent study on user data privacy, 97 percent of users [12] surveyed are concerned about securing their personal information. The existing systems do not guarantee that data is protected and that privacy is preserved. To secure user data, many countries have enacted data protection frameworks and or other regulations. The European Union (EU) was the first to propose the General Data Protection Regulation (GDPR), which serves as a legal foundation for safeguarding users' data within EU territory. In a similar vein, the Indian government recently is toying with the Personal Data Protection Bill-2019 (PDPB-2019) [9] to provide a legal framework for the protection of personal data held by companies doing business in India.

III. METHODOLOGY

The Data Protection Bill is a desperately needed legislation that would overhaul India's outdated and ineffective data protection policy. In comparison to current regulations, it would aid in the preservation of individual privacy rights and promote fair and transparent data use for innovation and growth, unlocking the digital economy. It has the potential to enforce accountability with data fiduciaries and processors and increase user awareness about their privacy [2]. A course called Cyber Laws was taught at a technical University in Pune and one of the topics in this course was the Data Protection Bill, which drew the student's attention the most. Following a thorough discussion of the bill and ample thinking why it is more important today than ever, as everything has shifted online as a result of the pandemic, it was noted that users have been unwittingly giving a lot more data to various applications than ever before. The major concern was how an individual's data privacy is handled in a world where practically everything has gone online. This very thought makes the end user more conscious of questions such as what does privacy imply to each end user? How much data does an application capture? How is end-user data collected, processed and protected? It was felt by the authors that it would be interesting to know what computer engineering students think about and how well they understand data privacy and associated terms.

IV. SURVEY

A survey on the topic of "Personalization and Privacy" was conducted. This survey was distributed to undergraduate Computer Engineering students at a technology university in Pune via Google Forms.

The survey was split up into five sections namely:

- Awareness about the topic personalization and privacy
- Preferences regarding personalization and privacy
- Awareness about data getting misused
- Awareness about privacy getting hampered but still using the applications
- Willing to make better choices for better privacy

The survey had a total of 17 questions. The survey had approximately 400 responses (close to 300 males, around 100 females) from department of computer engineering students aged 19 to 22 years old over the course of a month.

V. HYPOTHESIS

The survey intended to capture user (computer engineering students) perspectives on the topic "Privacy and Personalization".

The survey aimed to address the following hypothesis,

H1: Privacy is more important to users than personalization

H1.1: Perception of privacy differs according to gender

H2: Applications capture more data than they actually need

H2.1: Are users willing to make better choices for increased privacy?

H3: Users just skim read Privacy Policies or End User's License Agreements (EULA)

H1: Privacy is more important to users than personalization.

Privacy is of paramount importance for any individual. According to the Internet Society and Customers International [12], 69 % of users are concerned about how personal data is collected via mobile applications. Users are becoming more aware of how companies' use their data is one of the main reasons for the rise in privacy concerns.

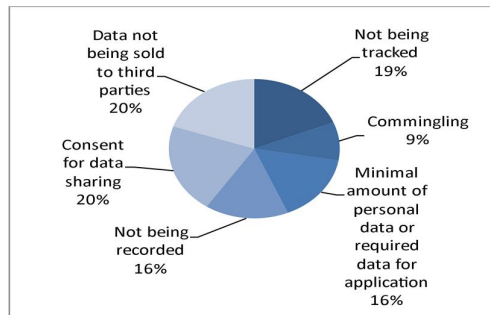


Fig. 1A. Depicts user's (males) perception of privacy

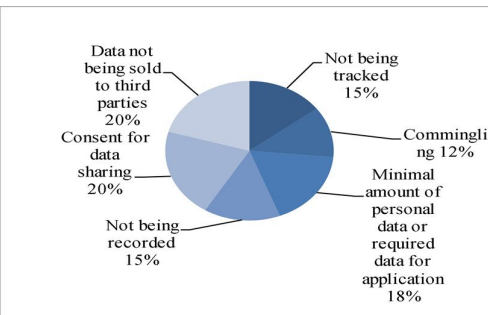


Fig. 1B. Depicts users (females) perception of privacy

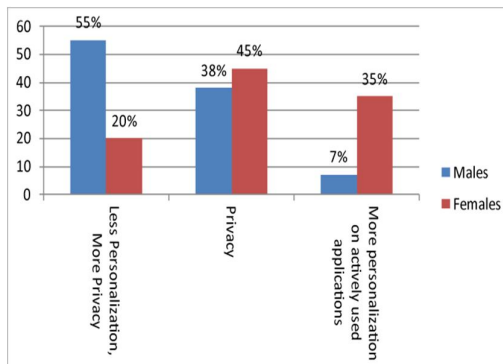


Fig. 1. Depicts users' privacy and personalization preferences

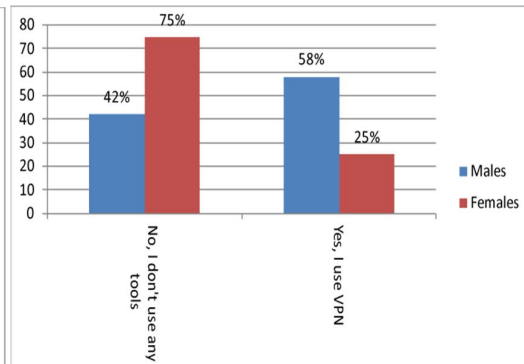


Fig. 3. Use of privacy-preserving tools

Observation: The primary context was to know more about what users prefer. It was observed that 55% of males choose more privacy than personalization, whereas 45% of females favor privacy. Based on the data in Fig. 2 it can be concluded that users value privacy more than personalization hence validating Hypothesis H1. When asked about their views on privacy, it's evident that both genders want the same things, such as asking for consent before sharing data, not being tracked, and so on. Given the data in Fig.1A and Fig.1B, it is evident that both genders have the same perception of privacy, disproving Hypothesis H1.1. According to the above Fig. 3, 58% of males say they use some form of privacy-preserving tool. In this survey results, it was a Virtual Private Network (VPN). While 75% of females indicated they do not use any privacy-preserving software.

Inference: Users regardless of their gender are increasingly concerned about their privacy but are not making use of free and online tools that safeguard their privacy. It can be concluded that both genders have the same perspective on privacy. It was apparent that males make efforts to protect their privacy, whilst females lack awareness of such tools (that are advantageous to all end users).

H2: Applications request more data than they actually need.

Users expect any free application to respect their privacy and not collect personal information or follow users' activities when they install it on their phone, especially if it's from a well-known company. Popular social media applications are the ones that collect the most data, food delivery applications do collect a lot of data as well. These apps capture more than just name, email address, and other necessary information. It can include a wide range of topics, including finances, health and fitness, browsing and search history, purchases, location, contact list, and other personal information.

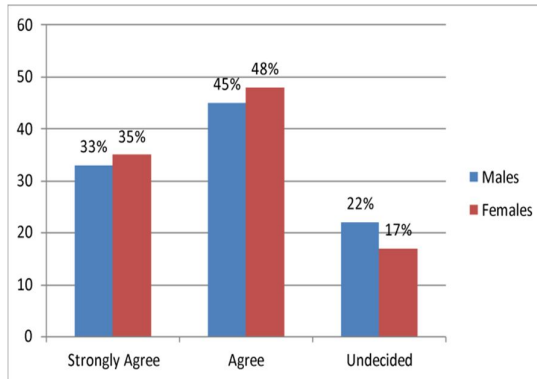


Fig.4. Depicts the user's perception to amount of data collected by various applications

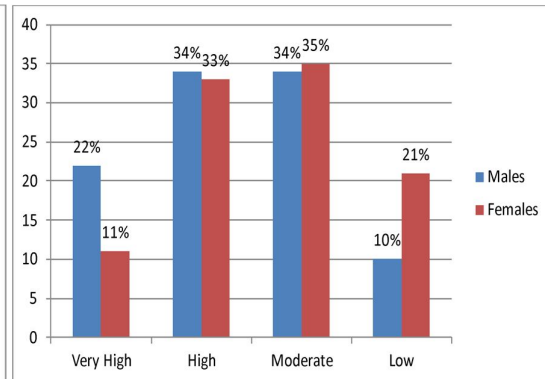


Fig. 5. Depicts the user's perception to amount of data that is misused

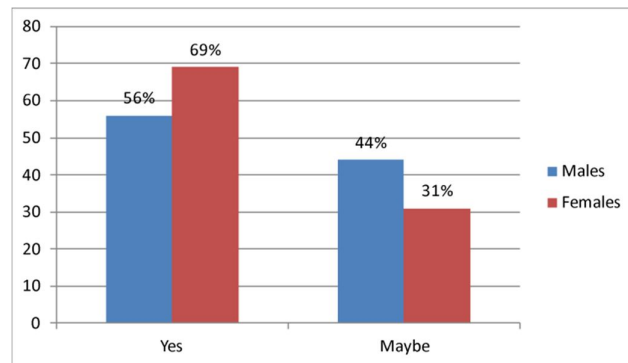


Fig.6. Depicts users' readiness to switch apps that provide more privacy

Observation: When asked if applications are requesting capturing more data than they require both genders, males and females, agreed with 48% and 45% (Fig. 4), respectively. Looking at the numbers in Fig. 4, it's evident that applications capture more data than they need, thus **validating hypothesis H2**. Using the primary findings from Fig. 5 68% of males and 67% of females believe their data is being misused. This implies that users must take necessary precautions to protect their privacy, such as utilizing a privacy-preserving tool,

switching to more privacy-friendly applications, etc. The survey also had questions concerning these applications. The survey results (69% males and 44% females) suggest that users are willing to switch to other applications if given better privacy among other applications. Here's a rundown of applications as suggested by the respondents and their alternatives: WhatsApp - Signal and Telegram, Chrome - Brave and Tor, Gmail - Yahoo and Outlook. By looking at the data in Fig. 6, it is evident that users are willing to switch to applications that provide greater privacy, **proving hypothesis H2.1**.

Inference: It can be concluded that both genders believe applications are requesting more data than they require. Users are aware that their data is being misused, yet they are unwilling to switch to applications that are more privacy conscious. The following are some of the reasons found for not migrating to other applications:

1. Users are very accustomed to use software.
2. The current application is more user-friendly.
3. Other programs aren't as user-friendly as this.
4. Other applications are unknown to many users.

H3: Users don't read privacy policies or end user's license agreements

Users are often confronted with service agreements and privacy policies. Users should be able to make educated decisions about regulations requiring these warnings, but current privacy policies fall short than expected. They're shockingly ineffective when it comes to enlightening users. According to research conducted in 2008, reading the privacy rules of all websites visited by the average American internet user would take 244 hours per year, which is why most users ignore them [9]. The survey had questions to understand if Computer Engineering Students read Privacy Policies and what the users reasoning are for not doing so.

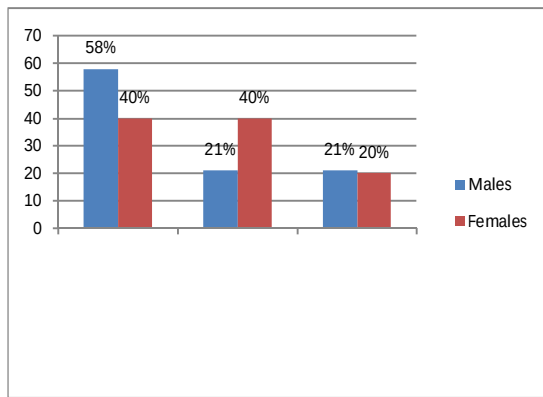


Fig. 7. Depicts whether or not users are to read privacy policies

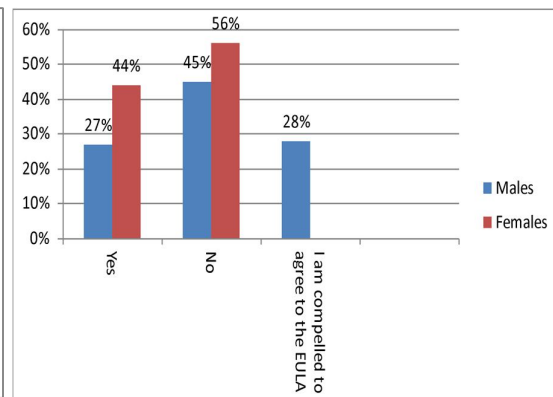


Fig. 8. Depicts whether or not users are to read EULA

End User License Agreements (EULA), are attempts to legally bind users to a set of harsh rules where users never sign their name, not physically endorse them. Usually, you won't see a EULA until after you've purchased the application [6]. EULAs are common in software and user devices these days, with millions of users clicking buttons that ostensibly commit them to agreements they never read and that frequently violate federal and state laws. In theory, these "contracts" are one-on-one agreements between companies and each of their users. These are more like legal requirements than user choices because practically every computer user in the world has been subjected to the take-it-or-leave-it conditions at some point or the other.

Some Common EULA terms that may harm users [6].

1. We are not responsible if this product messes up your computer
2. Do not criticize the product publicly
3. By signing this contract, you also agree to every change in future versions of it
4. EULAs are subject to change without notice
5. Using this product means you will be monitored

Observation: According to the Fig. 7, only 21% of males read privacy policies, whereas 40% of females responded that they read privacy policies. From Fig. 8, it is distinctively clear that 45% males and 56% females said they do not read EULAs. Fig. 7 and Fig. 8 demonstrate that users are reluctant in reading Privacy Policies or EULAs, **proving hypothesis H3**.

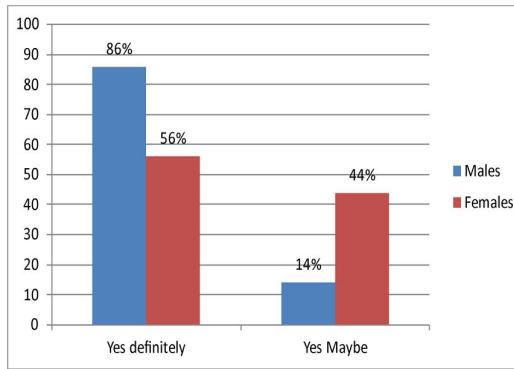


Fig. 9. Depicts users' views on how privacy policies should be framed

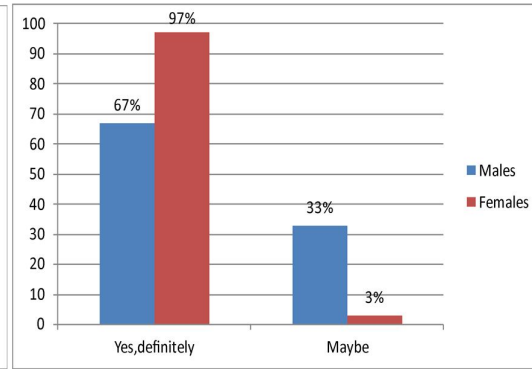


Fig. 10. Depicts users' views on whether or not a summary should be included on the first page of a privacy policy

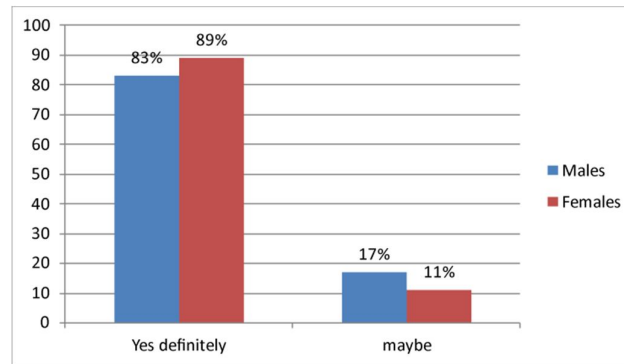


Fig. 11. Depicts user's views on whether or not a summary should be included on the first page of a EULA

Inference: It can be concluded that females are more concerned about reading privacy policies and consider reading those policies as a step to preserve their privacy. In addition, over 28% of males claimed they are compelled to click the consent button in order to use or purchase the application. The user must agree to the EULA in order to fully utilize the application's functions. The respondents indicated that the privacy policy or EULA should have a third option that says "I disagree, proceed with limited functionality" (akin to a trial version) in order to access required or needed services. It is also critical to comprehend why users don't read their privacy policies or EULAs. To understand this point there were a couple of questions to the survey to learn more about the reasons for this behavior and what can be done to encourage individuals to read and understand them. Even those users who go through privacy policies find it difficult to comprehend them because they frequently require higher-level reading and comprehending skills. Privacy Policies usually include a company's services, resulting in unclear phrases that make it difficult to understand what personal information is gathered, how it is used, and with whom it is shared. Most users are aware that their data is being captured, so they don't really require a privacy policy to inform them. Users are concerned about how this data is aggregated, processed and shared with third parties. Based on the respondent's inputs, as in Fig. 9, 10, 11, the authors, in this paper, propose a better way to establish simpler and easier-to-follow privacy regulations as indicated below,

1. Privacy Policies and EULA should be framed differently according to the application in order to make them more easily understandable to the user.
2. On the Initial page of the Privacy Policies and or the EULA, there should be a Summary of the key points that end-users are most concerned about.
3. Companies should conduct user acceptance testing of new privacy notifications to ensure that the final versions are clear and non-misleading, and that the options provided to the users are useful.
4. Before using the services, users should be asked if they wish to allow, restrict, or completely opt out of data sharing.
5. Traditional privacy policies and shorter user-friendly privacy notices can co-exist.

When a mobile application requests access to the phone's location or contacts, for example, the user is given the opportunity to choose "No." Users can use this type of system to get useful information and make informed decisions. They also encourage app developers to communicate with users about privacy more effectively.

VI. DISCUSSION

As most of the day-to-day activities have shifted to online mode, users are becoming more aware of the worth of their personal information than they previously were. According to the results of the survey, Males prefer Personalization with more Privacy, while females prefer Privacy only. Males who are concerned about their Privacy are putting an extra effort to protect their Privacy while using services online. However, there is a need to increase awareness in females about Privacy Preserving methods, as it has been observed in the survey that female users hardly take any steps to protect their privacy. Privacy is better to be covered in the early years of the engineering curriculum so that the users are more aware of privacy-preserving tools by the time they complete the course. Users are first required to be made aware of privacy issues, and then exposed to technologies that can help them protect their privacy. Switching to privacy-preserving browsers, like Tor and Brave, are required towards safeguarding user privacy. If switching to Tor or Brave is not convenient for the user, other techniques such as use of VPN, tunneling and use of proxy servers can be used instead.

According to the survey, users indicated that applications are capturing more data than they require under the name of personalization. Personalization is desirable, but users must retain control over their personal information at all times. Companies providing services should be open and clear about the data they capture, process and utilize. Companies should inform users about how much data has been collected and how it will be used. Every user should have the right to choose what personal data they want to share, as well as have the right to be forgotten. It is also required that companies should make Privacy Policies and EULAs more readable. According to the survey, users tend to skim read privacy policies or EULAs because they are too long to read and contain a lot of legal terminology that users cannot comprehend. To make Privacy Policies and EULAs easier to comprehend, the issues that concern users' privacy should be summarized on the first page and linked to their relevant pages. The user can read and agree to the points about which they are concerned. Privacy Policies and EULA should be tailored to the specific use of the application, so that users are aware of the rules and laws that apply to them when using the specific application. Whenever a user accesses a feature that collects data and uses it for another purpose, user consent should be taken. Privacy policies and EULAs written in this manner may be advantageous to both users and companies in terms of maintaining transparency.

Since infringement of user privacy is increasing, no amount of compensation can pay for the damages caused by it. The Government is required to take proactive action and enact data protection and privacy regulations at the earliest. These guidelines must be to safeguard the best interests of users. The Indian government's data protection bill is a long-awaited and much-needed bill for protecting user privacy. The bill mandates that all organizations undertake organizational changes in order to improve data security. These include privacy-by-design principles (an approach in which privacy is a key consideration in how the business is organized), security safeguards, and so on.

VII. FUTURE SCOPE

As VPNs typically do not provide thorough audit logs, users are not able to control and record the actions of every third-party vendor using the VPN. All of this is normally logged in the user's connection timings, and even then, the data is and monitored tracked in a separate log. If a breach or error occurs as a result of a third-party provider, proving who or what caused the problem is extremely challenging (applications accessed, user, purpose for access, etc.). The phrase "Privacy by Design" implies "data protection through technological design." Nevertheless, there is still some misunderstanding about what "Privacy by Design" involves and how to implement it.

Data protection by design: Entails the use of Pseudonymization (replacing personally identifiable material with artificial identifiers), encryption (encoding messages so only those authorized can read them) etc.

Data protection by default: Social media platforms should be encouraged to set users' profile settings in the most privacy-friendly mode by default, for example, limiting accessibility from the beginning of the account creation.

REFERENCES

- [1] Aïmeur, E., & Tremblay, A. (2018, July) Me, Myself and I: Are Looking for a Balance between Personalization and Privacy. *acm.org*, 115–119. doi:10.1145/3213586.3226221
- [2] Dhavate, N., & Mohapatra, R. (2020, January) A look at proposed changes to India's (Personal) Data Protection Bill. Retrieved from *iapp.org*: <https://iapp.org/news/a/a-look-at-proposed-changes-to-indias-personal-data-protection-bill/>
- [3] Kaaniche, N., Laurent, M., & Belguith, S. (2020, December) Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and survey. *sciencedirect.com*, 171. doi: 10.1016/j.jnca.2020.102807
- [4] Lightman, A. (2021, June) Retrieved from *nbcnews.com*: <https://www.nbcnews.com/think/opinion/teens-have-never-known-world-without-data-sharing-it-s-ncna1254332>
- [5] Liu, Y., Lin, F. Y., Post, Z. A., Ebrahimi, M., Zhang, N., Hu, J. L., Chen, H. (2020, November) Identifying, Collecting, and Monitoring Personally Identifiable Information: From the Dark Web to the Surface Web, IEEE International Conference on Intelligence and Security Informatics. doi:10.1109/ISI49825.2020.9280540.
- [6] Newitz, A. (2005, February) Dangerous Terms: A User's Guide to EULAs. Retrieved from <http://eff.org>.
- [7] Riedl, J. (2001, Nov-Dec) Personalization and privacy. IEEE. doi:10.1109/4236.968828
- [8] Schaub, F. (2017, October) Nobody reads privacy policies – here's how to fix that. Retrieved from *The Conversation*: <https://theconversation.com/nobody-reads-privacy-policies-heres-how-to-fix-that-81932>
- [9] Singh, R. G., & Ruj, S. (2020, May) A Technical Look at The Indian Personal Data Protection Bill. *arxiv.org*. doi:10.48550/arXiv.2005.13812
- [10] Spencer, J. (2019, November) Addressing the Privacy Paradox: Balancing Personalization with Data Privacy. Retrieved from *tealium.com*: <https://tealium.com/blog/privacy/addressing-the-privacy-paradox-balancing-personalization-with-data-privacy/>
- [11] Taylor, N., Papadopoulou, E., Gallacher, S., & Williams, H. (2013, January) Is There Really a Conflict between Privacy and Personalization? Proceedings of the 20th International Conference on Information Systems Development, ISD 2011, Edinburgh UK. doi:10.1007/978-1-4614-4951-5_1
- [12] Toch, E., Wang, Y., & Cramer, L. F. (2012, April) Personalization and Privacy: A Survey of privacy risks and remedies in personalization-based systems, *User Modeling and User-Adapted Interaction*, Volume 22, Issue 1-2, Pp. 203–220 doi:10.1007/s11257-011-9110-z