Data Handling Ethics: Literature Review

In an era marked by growing reliance on digital technologies and widespread use of personal information, anxieties about data privacy and security have intensified significantly. The surge in high-profile data breaches, coupled with ongoing debates over corporate handling of personal data, has sparked significant public discourse regarding the adequacy of existing protective measures for individual privacy rights. This literature review revolves around various themes related to data breaches, new technologies, and laws and regulations governing data handling practices, and aims to answer this big question: How do high-profile data breaches and controversies surrounding the use of personal data by companies impact public perceptions regarding data privacy and security, and what measures can be recommended to address and mitigate these concerns effectively? Data breaches are a growing concern due to the increased volume and sophistication of cyberattacks, highlighting the need for robust data protection measures. Laws and regulations like the California Consumer Privacy Act (CCPA) and General Data Protection Regulation (GDPR) aim to safeguard consumer privacy but can be challenging for companies to fully comply with, particularly smaller businesses with limited resources. Advancements in technology, such as AI, offer potential solutions for enhancing data privacy and enforcing compliance but also introduce new risks like data leaks and privacy violations. Companies must prioritize ethical data handling practices and transparency in their operations to build trust with consumers and mitigate potential legal and reputational risks associated with data breaches.

It is also important to direct this paper towards some target audience. In our case, it will be all users of technology and smaller tech companies. Everyone must be aware of ongoing current affairs regarding technology to keep themselves and their loved ones safe and secure from threats. Additionally, all tech companies must be fully aware of the laws regarding user data as it can put them into severe legal issues. When problems arise with companies messing up, and letting data go loose, it is not just bad for us, but also just as bad for the companies. Recovering from these issues causes companies to lose lots of money, lose their credible reputation, and they may even get involved in legal cases with the law. While they may be able to gain back their losses, it is difficult to gain back trust. With the negative toll on their reputation, people will be less likely to buy from the company, negatively impacting the economy.

**Themes and Discussions**

**Data Breaches**

Over the past few years, the number and level of data breaches have risen significantly. Such breaches can be direct, indirect, exposed, or inferred with private information, pointing out insistent challenges over the issue of how to safeguard personal data within big data systems. The ethical and security considerations prove that powerful measures need to be ensured to safeguard consumer privacy, along with a guarantee of responsible data handling practices.

The empirical article "Understanding Privacy Violations in Big Data Systems" by Shamsi and Khojaye, explains this very issue: the violation of privacy in big data systems. Shamsi and Khojaye identify four categories to classify the breaches: direct, indirect, exposure, and inference of private information. It is emphasized by the fact that the concern and violation concerning privacy are on the rise both in a business and government environment since huge volumes of data are gathered, fused, and analyzed. The article puts its emphasis not only on the benefits derived from the use of Big Data but also on the growing danger concerning the leakage of private and sensitive data. An example of this includes an automatic email scan and consequently retrieval of some personal information. They conclude with strategies for enhancing privacy in big data systems and stress the need to implement a robust security design. This is a seminal paper that works as an indispensable source in realizing how serious these data breaches are and hence, essential to find out and prevent the same.

As seen in an empirical article titled “Big Data Technology and Ethics Considerations in Customer Behavior and Customer Feedback Mining”, The writer, Deng, discussed the ethical importance of handling the data of customers. He emphasized how it is very important for businesses to be able to balance the use of big data technology with moral values. Deng stated, “With the improved capability and availability of big data collection, storage, access and sharing, and big data analytics and deep learning fever, in particular, NLP enabled unstructured text understanding, ethics consideration in the entire process of customer behavior and customer feedback mining become more urgent and prominent.” (Deng) He then continued talking about his concern about this due to customer feedback and behavior analysis, and in addition, he later highlighted the urgency of bearing in mind ethics in the total procedure of data mining, particularly in the context of client satisfaction as well as privacy.

Another study done by IT Professionals Jawwad Shamsi, and Muhammad Ali Khojaye titled “Understanding Privacy Violations in Big Data Systems” explored the four main categories of privacy breaches and violations in big data systems. The categories they identified were direct

violations, indirect violations, exposure, and inference of private information. Their research spoke about the liabilities and weaknesses of big data systems. They also stress the need for more deterring security measures that should be put in place to protect restricted information and data more strongly, stating “they should strive to make big data storage systems more secure. This should be done by strengthening cyber-defense systems and by proposing efficient methods for encryption and storage. Intrusion detection and prevention should be enhanced to limit data breach incidents.” (Shamsi and Khojaye)

The article “Big Data Security Problems Threaten Consumers’ Privacy” by Jungwoo Ryoo, an Associate Professor with expertise in information sciences and technology, provides an in-depth analysis of the security and privacy implications of big data (as well as meta-data), highlighting the risks and consequences of data breaches. He gives two instances where immense quantities of user data were breached, stating “Due to the sheer scale of people involved in big data security incidents, the stakes are higher than ever. When the professional development system at Arkansas University was breached in 2014, just 50,000 people were affected. That’s a large number, but compare it with 145 million people whose birth dates, home and email addresses, and other information were stolen in a data breach at eBay that same year” (Ryoo). Ryoo advocates for measures to enhance security and protect consumer privacy by enhancing company transparency: “Transparency is the key to letting us harness the power of big data while addressing its security and privacy challenges. Handlers of big data should disclose information on what they gather and for what purposes. In addition, consumers must know how the data is stored, who has access to it, and how that access is granted.” (Ryoo)

**New technologies**

Although the rise of technology has been great for the market, it has also been bringing in more issues. The invention of AI has drastically impacted the world of technology, but with every good thing also comes some bad. For example, although AI can help us in spotting who is stealing data from certain software, it can also be used to do more damage and at certain times, even cause data leaks. On the other hand, other regular household smart devices could also be doing us harm. Although we may not realize it, devices such as our smart home AC units could be collecting data on us without us being aware of it.

An article written by Chibba and Cavoukian in 2015, titled “Privacy, Consumer Trust and Big Data: Privacy by Design and the 3 C'S”, highlighted the risks of major businesses in managing big data. They introduced the idea of Privacy by Design (PbD), which was an approach

that stresses continual and frequent protection even when scaling. The article also promotes a possible approach to secure and moral business scaling that could guarantee the goal of data privacy in the oncoming years of big data. After their research, they concluded that they could accomplish these goals by “embedding or coding privacy preferences into the technology itself, to prevent the privacy harms from arising. This is eminently within our reach. No doubt, it will require innovation and ingenuity, through communication, consultation, and collaboration (3C's), but if we are to continue with existing technological progress in an increasingly connected world, it will be essential to maintain our future privacy and freedoms.” (Chibba)

Advancements in technology can also be used to our advantage when enforcing compliance with data privacy regulations in the era of big data and online surveillance: "The battle is ongoing...laws worldwide...regulate who gets that data and how they can use it" (Pande). The article "AI could constantly scan the internet for data privacy violations, a quicker, easier way to enforce compliance” written by Karuna Pande Joshi, an Assistant Professor of Information Systems with experience in data privacy, presents ideas for the efficient use of artificial intelligence to enforce compliance with data privacy regulations. She offers a detailed analysis of AI-based solutions for interpreting data privacy regulations in the digital age, discussing the challenges of understanding complex regulations, such as the lack of terminological knowledge or context behind written laws and policies. Advocating for the adoption of innovative technologies, she proposes the use of ontology and knowledge graphs as effective tools to simplify AI/tech-related current affairs for common users. Joshi emphasizes that these technologies are essential for addressing the challenges posed by evolving regulatory frameworks.

**Laws and Regulations**

There are many rules and laws surrounding how certain companies should handle their data. Notable examples include the California Consumer Privacy Act (CCPA) in the United States and the General Data Protection Regulation (GDPR) in Europe, both of which impose comprehensive procedures to enhance privacy rights and consumer protection. Despite these well-defined legal frameworks, many organizations struggle with trying to keep up with all these advancements in new technology.  De and Imine explored the troubles that were linked to gaining valid consent for directed advertising. This case study, done in 2020, focused on the application Facebook. Throughout the article, the importance of user consent under the EU General Data Protection Regulation (GDPR) is emphasized and leads to morality and ethics

being questioned. The study questions whether or not Facebook follows these laws and consent requirements and talks about the penalties or possible concerns of non-compliance which could include substantial fines. The article states, “The absence of consent may mean an unlawful data processing and a lack of control of the user (data subject) on personal data. However, consent mechanisms that do not fully satisfy GDPR requirements can give users a false sense of control, encouraging them to allow the processing of more personal data than they would have otherwise.” (De and Imine). They highlight the importance of these legal measures that are put in place and used to ensure that data privacy and security are not being compromised, in the perspective of directed advertising procedures and methods. Targeted advertising is a major source of income for large tech companies such as Google and Facebook. They question the ethics of targeted advertising and its compliance with the GDPR, also highlighting that the absence of proper consent is a potential violation of data processing laws. As big data is becoming more popular, businesses must be aware of the laws and regulations they must follow when collecting and distributing the personal data of users, such as meta-data, preferences, and connections.

Facebook's revenue model is heavily reliant on advertising. This is a rough overview of their design process:

1. “Powerful data controllers such as Facebook and Google collect a lot of personal data about their users from different sources including the users themselves, data brokers, other websites and applications.” (De and Imine)
2. This platform uses an inference algorithm to develop user profiles, categorizing the information into interests, demographics, and behaviors based on the collected activities and other personal data.
3. Advertisers select specific user profiles they wish to target and initiate an advertisement on the platform.

The platform then uses a selection algorithm to align the advertiser's desired user profiles with actual users, identifying a group that meets the specified criteria.

However, there are severe consequences set for not adhering to proper safety regulations regarding big data and meta-data. There are laws such as the GDPR which are a deterrent to irresponsible and misuse of data. It is safe to share information as long as laws are strong enough to deter companies from misuse.

Additionally, there is one over-arching concern of every technology-utilizing person: Personal information handling (meta-data). You may have probably experienced receiving suspiciously targeted advertisements while scrolling through social media and thought it was a coincidence. Unfortunately, it is not. The article “How Companies Learn What Children Secretly Want," written by Faith Boninger and Alex Molnar, are respected researchers at the University of Colorado Boulder, specializing in education policy. Boninger and Molnar discuss the collection of student data by educational technology companies which they say is used for marketing purposes. They also explain the methods used to collect data, highlighting the potential consequences of targeted advertising on children's well-being. The article raises awareness about children’s privacy concerns and emphasizes stringer monitoring of companies' data practices in educational settings, provides an in-depth analysis of how educational tech companies gather and utilize children's data for marketing purposes, and emphasizes the use of targeted advertising and its negative effects on children's health and well-being.

Another similar study was done regarding addressing the lack of awareness among Americans regarding the predictive capabilities of companies utilizing their data. "Sixty-seven percent of smartphone users rely on Google Maps to help them get to where they are going quickly and efficiently." Thus, it is important that information regarding user location is kept secure (Rader). Emily Rader emphasizes in her article “Most Americans don’t realize what companies can predict from their data” the lack of awareness among Americans regarding the predictive capabilities of companies utilizing their data. Rader advocates for stronger privacy-related regulation to create transparency and accountability in data usage.

**Conclusion**

In conclusion, personal data security incidents have increased within big data systems. Classified into four categories, Shamsi and Khojaye's article identifies the mounting concern of a privacy violation in business and the government sector. Furthermore, there must be designs around strong privacy and security, as well as prevention of transgression by regulatory bodies like CCPA and GDPR. New technologies like AI come with both opportunities and threats, thus making it necessary to practice data handling ethically through initiatives like Privacy by Design. In other words, even with such legislations and regulations as the General Data Protection Regulations (GDPR), companies find it difficult to show adherence, this means that there must be transparency and accountability in data use if consumer privacy is going to be safeguarded.

Coming back to the original question regarding how high-profile data breaches and controversies surrounding the use of personal data by companies impact public perceptions regarding data privacy and security and what measures can be recommended to address and mitigate these concerns effectively, addressing data breaches and upholding privacy standards necessitates a multifaceted approach, which would include enforcing strict legal frameworks like GDPR and CCPA, using AI to our advantage when securing user data and systems, and continuous education for businesses and consumers about data rights and protections. By integrating robust legal measures, innovative tech solutions, and ethical practices, trust in digital systems can be rebuilt, allowing new technologies to be harnessed effectively without compromising user privacy. With strong security measures and cyber/AI crimes deterrents set, and spreading awareness of technological current affairs, everyone will remain safe from harm.

**Citations**


- Ryoo Associate Professor of Information Sciences and Technology at Altoona campus, Jungwoo. “Big Data Security Problems Threaten Consumers’ Privacy.” *The Conversation*, 18 Jan. 2024, theconversation.com/big-data-security-problems-threaten-consumers-privacy-54798.

- Shamsi, Jawwad A., and Muhammad Ali Khojaye. “Understanding privacy violations in Big Data Systems.” *IT Professional*, vol. 20, no. 3, May 2018, pp. 73–81, https://doi.org/10.1109/mitp.2018.032501750.

- De, Sourya Joyee, and Abdessamad Imine. “Consent for Targeted Advertising: The Case of Facebook.” *AI & SOCIETY*, vol. 35, no. 4, 12 May 2020, pp. 1055–1064, https://doi.org/10.1007/s00146-020-00981-5.

- Deng, Xin. “Big Data Technology and Ethics Considerations in Customer Behavior and Customer Feedback Mining.” *2017 IEEE International Conference on Big Data (Big Data)*, Dec. 2017, https://doi.org/10.1109/bigdata.2017.8258399.

- M. Chibba and A. Cavoukian, "Privacy, consumer trust and big data: Privacy by design and the 3 C'S," 2015 ITU Kaleidoscope: Trust in the Information Society (K-2015), Barcelona, Spain, 2015, pp. 1-5, doi: 10.1109/Kaleidoscope.2015.7383624.

- Karuna Pande Joshi Assistant Professor of Information Systems. “Ai Could Constantly Scan the Internet for Data Privacy Violations, a Quicker, Easier Way to Enforce Compliance.” *The Conversation*, 24 Jan. 2024, theconversation.com/ai-could-constantly-scan-the-internet-for-data-privacy-violations-a-quicker-easier-way-to-enforce-compliance-128973.

- Boninger Research Associate in Education Policy, Faith, and Alex Molnar Research Professor. “How Companies Learn What Children Secretly Want.” *The Conversation*, 5 July 2023, theconversation.com/how-companies-learn-what-children-secretly-want-63178.

- Rader Associate Professor of Media and Information, Emilee. “Most Americans Don’t Realize What Companies Can Predict from Their Data.” *The Conversation*, 1 Mar. 2024, theconversation.com/most-americans-dont-realize-what-companies-can-predict-from-their-data-110760.