A Systems Approach to Understanding Challenges in Preserving User Privacy

by

Mervine Anand Govada

B. Tech Electronics and Communications Engineering, Jawaharlal Nehru Technological University, 2008
MS Data Analytics Engineering, George Mason University, 2020

SUBMITTED TO THE SYSTEM DESIGN AND MANAGEMENT PROGRAM
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN ENGINEERING AND MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Mervine Anand Govada
System Design and Management
May 12, 2023

Certified by: Dr. Juanjuan Zhang
John D. C. Little Professor of Marketing at the MIT Sloan School of Management
Thesis Supervisor

Accepted by: Joan S. Rubin
Executive Director
System Design and Management Program

This page intentionally left blank

A Systems Approach to Understanding Challenges in Preserving User Privacy
By

Mervine Anand Govada

Submitted to the System Design and Management Program
On May 12, 2023, in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Engineering and Management

Abstract

In recent years, enterprises' collection and processing of personal data has raised significant concerns about customer privacy. Ensuring customer privacy is vital for ethical data use and building trust. However, enterprises may need to enhance their efforts to safeguard customer privacy effectively.

Customers have become increasingly aware of how businesses handle their personal information and the potential risks that come with it; they proactively seek businesses that prioritize privacy protection. However, customer trust in how enterprises protect customer data varies, emphasizing the need for businesses to be transparent and communicate clearly with customers about their data protection practices. Clear and concise communication can include privacy policies and obtaining informed consent from customers.

Enterprises typically use anonymization, encryption, data masking, pseudonymization, and access control to protect customer privacy. The thesis explores two key technologies to enhance customer privacy and increase customer trust in enterprises: Federated Learning and Differential Privacy.

Preserving customer privacy is essential for building trust with customers, ensuring ethical use of personal data, and compliance with regulations. Improving privacy from a technology standpoint might not necessarily result in the customers' desired outcome. Therefore, it is essential to take the entire system into account. A systems approach can aid in analyzing and understanding the challenges of holistically preserving customer privacy from the perspectives of the customer, enterprise, and other stakeholders. By adopting a systems approach, enterprises can identify potential risks and challenges within the system, gain a better understanding of interconnections and interdependencies, and develop more effective solutions.

The systems approach involves identifying and analyzing subsystems, goals, and interactions, allowing enterprises to view their data practices holistically and identify potential privacy risks. By using a systems approach and leveraging technologies such as Federated Learning and Differential Privacy, enterprises can take a customer-centric approach to reduce privacy concerns.

Thesis Supervisor: Dr. Juanjuan Zhang
John D. C. Little Professor of Marketing at the MIT Sloan School of Management

Acknowledgments

Completing this thesis would not have been possible without the support of many people who have contributed to my personal and academic growth. Looking back on my academic journey, I am deeply grateful for my wife, children, parents, friends, colleagues, and extended family's role.

First and foremost, I would like to express my deepest gratitude to my wife and children. Their unwavering love, support, and encouragement have been the cornerstone of my academic journey. Their patience, understanding, and belief in me have been a constant source of inspiration, and I could not have completed this journey without their support. My wife, Lenova, has always been my biggest cheerleader, and I am so grateful to her for her unwavering support, especially during challenging times. She has supported our family tremendously, sacrificing her career to take on more responsibilities and supporting me in my academic pursuits. My daughter Ariel, and son Asher, have been my constant source of joy and motivation. Their unconditional love and smiles have kept me going through long hours of research and writing. I am incredibly blessed to have them in my life.

I want to express my gratitude to my parents. Their support and belief in my dreams have been a driving force behind my professional and academic success. They supported my decision to pursue my education in the US after completing my undergraduate degree and have been a constant source of support and encouragement. I will forever be thankful to them. They have always been there for me, encouraging and supporting me during the most challenging times. I hope to make them proud of my achievements.

I want to express my appreciation to my academic advisor Joan Rubin. Her guidance, support, and expertise have been instrumental to my time at MIT. She has provided me with critical feedback and has encouraged and inspired me throughout my academic journey. I am honored to have had the opportunity to work with her and have learned so much from her expertise and insights.

I want to express my gratitude to my thesis supervisor, Dr. Juanjuan Zhang. Her guidance, support, and expertise have been invaluable to me throughout my research. She provided me with critical feedback, challenging questions, and creative insights that helped me improve my research and writing. She also supported me during the most challenging times, giving me the confidence and encouragement to keep going. I am honored to have worked with her, and I have learned so much from her expertise and insights.

Finally, I thank everyone who supported me on this academic journey. Their unwavering support, encouragement, and belief in me have helped me to achieve my goals and make my dreams a reality. I am grateful for their presence in my life, and I hope to pay forward their kindness by supporting others in their own academic or professional journeys.

Table of Contents

List of Figures

List of Tables

Chapter 1 - Introduction

1.1 Overview of The Thesis

User privacy has become more crucial due to the rapid development of technology for businesses that gather and analyze massive amounts of personal data. Businesses need customers' trust to succeed, but data and privacy breaches can quickly erode that trust. Enterprises must understand their challenges and how technology can help solve them and protect user privacy.

The thesis examines the challenges of protecting user privacy and how Federated Learning and Differential Privacy can be used to solve them. The thesis also discusses regulations, policies, legal and ethical considerations related to user privacy, and the importance of transparency and accountability. The thesis explores two technologies, Federated Learning and Differential Privacy, and how they can help reduce privacy risks and increase customer confidence. Federated Learning ensures that data stays on the user's device instead of being sent to a central server while allowing insights to be gained from the data. Differential privacy adds noise to the data, making it difficult to identify individual users while still providing valuable insights.

The thesis explores what customers want and how enterprises can enhance their privacy mechanisms while increasing user trust. Enterprises must demonstrate their commitment to user privacy by reviewing data collection practices and giving users control over their data. Businesses can build solid, long-lasting relationships by adopting a more holistic view and listening to customers. Understanding the challenges of safeguarding user information and implementing privacy-preserving technologies to leverage user data while protecting their privacy is vital.

The thesis takes a systems approach to holistically study the technologies, stakeholder needs, policies and regulations, and enterprises. The objective of putting user privacy first is to build a more secure and reliable digital environment. Businesses that take this approach can win customers' trust while displaying a dedication to moral behavior. This will benefit all parties as customers are likely to engage in online activities such as shopping, banking, and social media, leading to increased activity, higher revenues for businesses, and increased investment in technology and innovation.

1.2 Importance of User Privacy

User privacy has emerged as a fundamental right, demanding greater care and protection from businesses and policymakers. People increasingly rely on digital platforms for communication, transacting, and social interaction and leave their data open to abuse by businesses or misuse by malicious users. Businesses and policymakers must recognize its significance to safeguard it effectively; businesses and policymakers must work proactively towards safeguarding it while giving users greater control of their data; therefore, enterprises must prioritize this fundamental right to build a safer digital future for all.

As enterprises increasingly collect, store, and analyze personal data, user privacy has become a significant concern. In 2022 alone, the number of data compromises in the United States was 1802. Moreover, over 422 million individuals were victimized by cyber data violations [1] (see *Figure 1.*), such as breaches, data exposures, and data leaks in the United States.

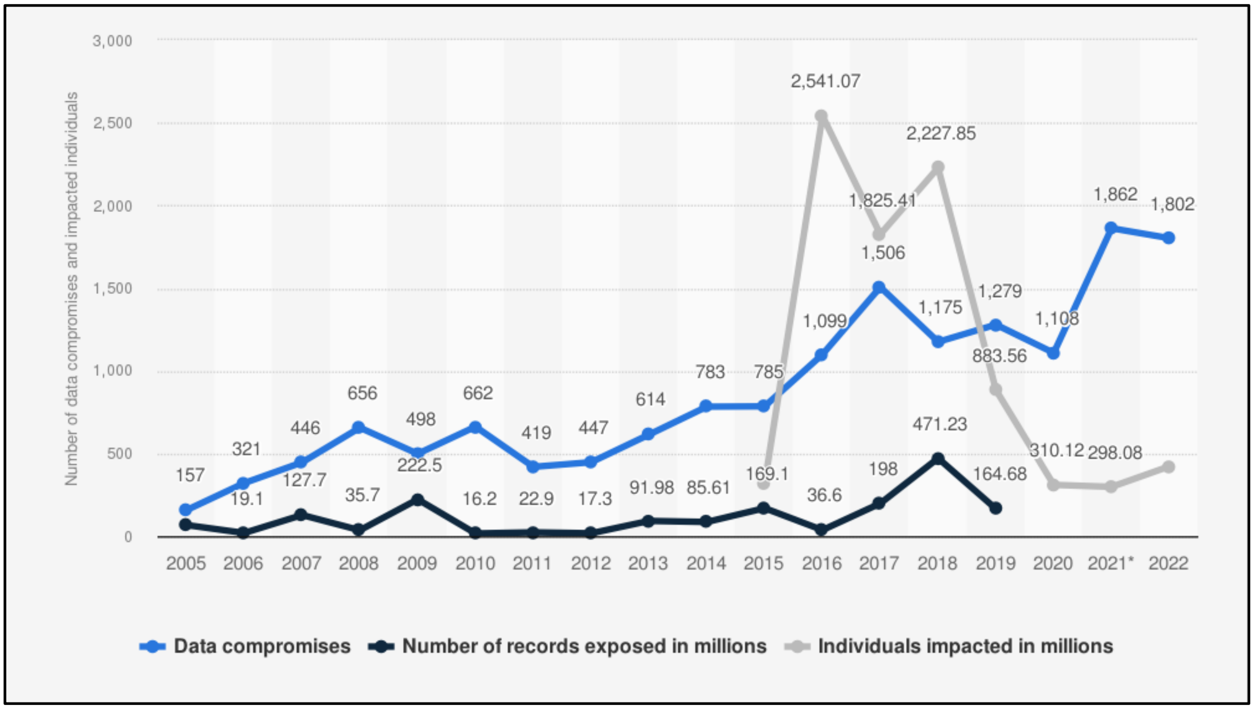


Figure 1. Annual number of data compromises and individuals impacted in the United States from 2005 to 2022

Table 1. shows the private data violation incidents in the US 2020-2022, by industry [2]. Health care, financial services, and manufacturing and utilities industries are prone to higher data violations.

Number of cases of data violation due to cyber-attacks in the United States from 2020 to 2022, by industry.

Industry	2020	2021	2022
Healthcare	306	330	344
Financial Services	138	279	268
Manufacturing and utilities	70	222	249
Professional Services	144	184	224
Education	42	125	100
Technology	67	79	86
Government	47	66	74
Non-profit/NGO	31	86	71
Retail	53	102	65
Transportation	21	44	36
Hospitality	17	33	34
Unknown		4	
Other	172	308	251

Table 1. Number of cases of data violation due to cyber-attacks in the US from 2020 to 2022

User protection refers to individuals' ability to manage how their data is collected, used, and shared. Personal data encompasses any information identifying an individual, such as name, address, telephone number, email address, or date of birth; more sensitive personal data, such as medical records or financial details, could also fall within this definition.

Dr. Sweeney's research [3] has revealed that a startling 87 percent of the United States population can be uniquely identified based on a combination of their date of birth, gender, and zip code. This illustrates the risk to individual privacy in today's digital world as seemingly irrelevant information can be combined to identify individuals with great accuracy uniquely. Thus, it underscores protecting personal information and more stringent data privacy laws and regulations to safeguard it better.

The importance of user privacy can be viewed from several perspectives, including legal, ethical, and economic. From a legal perspective, user privacy is protected by various laws and regulations, including the General Data Protection Regulation [49] (GDPR), the California Consumer Privacy Act [50] (CCPA), the Australian Privacy Act, Canadian Personal Information Protection and Electronic Documents Act [51] (PIPEDA), and Brazilian General Data Protection Law [52] (LGPD). These laws describe the rights of individuals regarding their personal data and impose obligations on enterprises to protect users' privacy. From an ethical perspective, user privacy is important because it reflects an individual's right to autonomy. Individuals should have the right to control their data, and businesses have an obligation to respect that right. The moral impact of a privacy breach can be severe, including reputational damage, loss of trust, and even financial loss in some cases. From an economic perspective, user privacy is critical to the functioning of a data-driven economy. Collecting and analyzing personal data allows enterprises to gain insights into customer behavior and preferences to improve products and services. However, the misuse of personal data can have negative consequences for individuals and businesses, including lost revenue, regulatory fines, and damaged reputations.

Technology's increasingly prevalent role in our daily lives underscores its significance regarding user privacy. Smartphones, social media platforms, laptops, smart TVs, and other digital devices have made personal data more readily available than ever. A 2019 study [53] reported that 81 percent of the 1,161 respondents (see *Figure 2.*) felt their data and personal information online was somewhat or very vulnerable to hackers, and only 2 percent indicated the opposite sentiment, they felt secure that their data was not vulnerable at all. Enterprises must increase their use of privacy-preserving technologies that protect user data while still permitting enterprises to leverage insights gained from user data for insights purposes. While protecting user privacy presents tremendous challenges, its protection should never be underestimated. These concerns include acquiring and retaining personal data, risks associated with data storage systems, and ethical considerations related to its utilization. Untangling personal data can be complex. Information may be distributed among various systems, databases, and applications making its management complex and time-consuming.

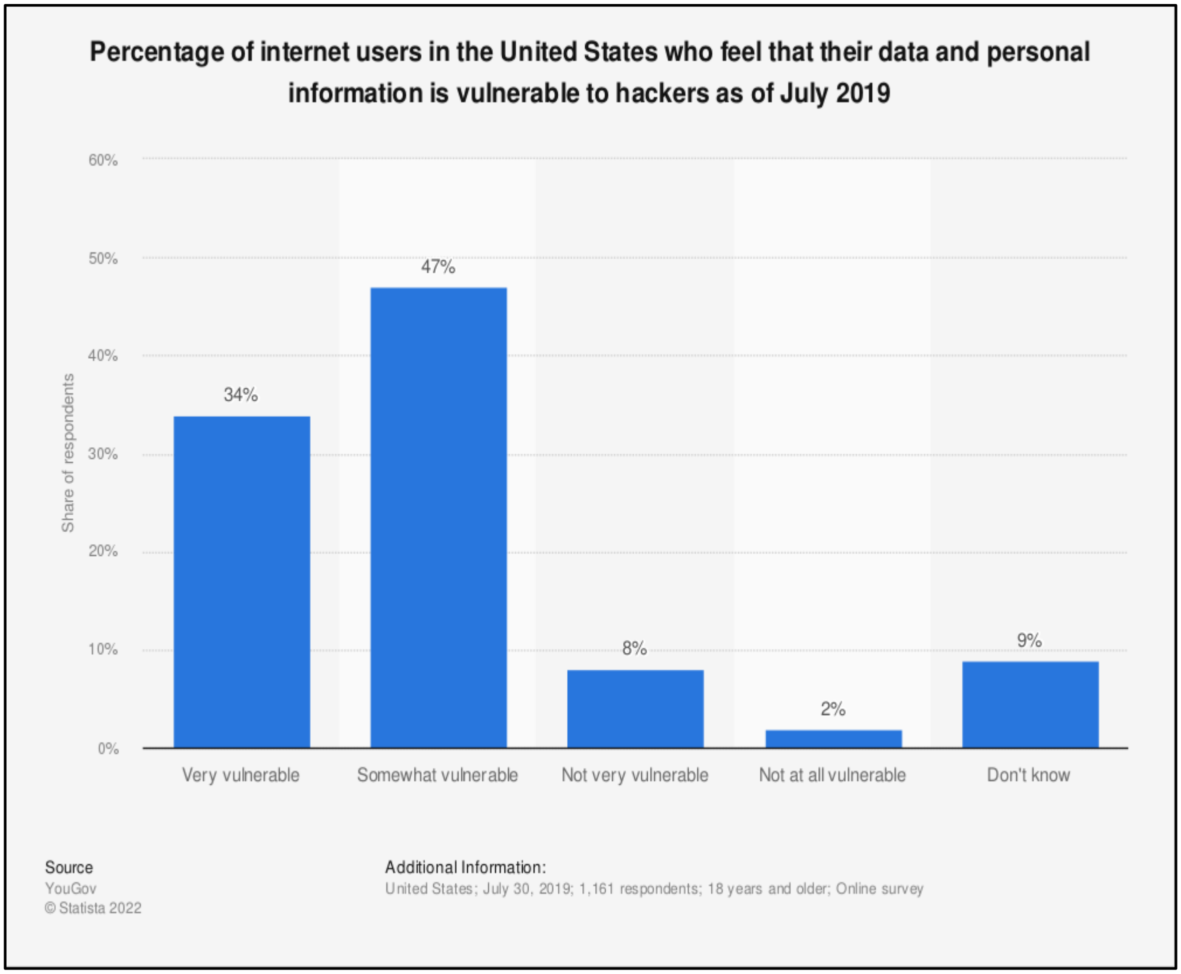


Figure 2. Percentage of internet users in the US who feel that their data and personal information is vulnerable to hackers.

At the same time, there has been an increase in data breaches, as even one breach can allow access to vast quantities of personal information without authorization from its rightful owners. Ethical considerations must also be factored into how data usage occurs. Enterprises must use personal data that meets users' preferences transparently, with no discriminatory treatment due to sensitive personal data being withheld from service. Enterprises cannot take measures that violate individuals' rights against discrimination in terms of service availability. And enterprises cannot make assumptions based on these variables that exclude people based solely on how that individual may have provided their data.

1.3 Challenges in Preserving User Privacy

Technology has transformed how people communicate, work, and live, yet due to widespread data collection and analysis practices that result in large volumes of personal information being gathered about individuals, enterprises, and governments. User privacy remains an increasingly pressing concern for all three entities involved.

Data collection poses one of the greatest hurdles to user privacy protection, with companies often collecting large volumes without explicit consent through cookies, tracking pixels, and other digital markers - such as user browsing history, search queries, location data, or sensitive personal information collected without explicit user knowledge and understanding. Regulating data collection requires significant user awareness as users need to recognize how much is collected or how their personal information may be utilized.

Data storage presents another daunting obstacle to protecting user privacy. Companies must ensure user data is stored safely to prevent unauthorized access or theft; this requires robust measures like encryption, access controls, and regular audits, but even with such safeguards in place, data breaches may still occur, resulting in significant harm to users.

Data sharing presents an obstacle to protecting user privacy. Companies often share user data with third-party entities like advertisers or analytics firms, increasing risk to individual data security. Users need to be made aware of who or why their data is shared; this may cause distrust among people sharing it and create user confidentiality and protection issues.

Protecting user privacy presents social and legal hurdles besides technical barriers. Users often desire more direct control of their data when providing online services that require user submission of personal information for usage; companies must also be more transparent regarding data collection practices leading to user mistrust.

Legal and regulatory frameworks relating to user privacy can be complex and vary widely across nations, making compliance challenging due to data that easily crosses borders. Companies must navigate these laws and regulations for the best compliance results, yet enforcement can sometimes prove challenging as companies must remain compliant.

Ensuring user control over their data can be an immense challenge, with users having a right to know exactly which data is being collected about them and shared. A survey [5] (see *Figure 3.*) conducted in December 2021 in the US discovered that 43 percent of consumers thought they had no say over companies sharing personal information with third parties, while 42 percent did not think they could prevent companies from collecting any of this personal information about themselves in the future.

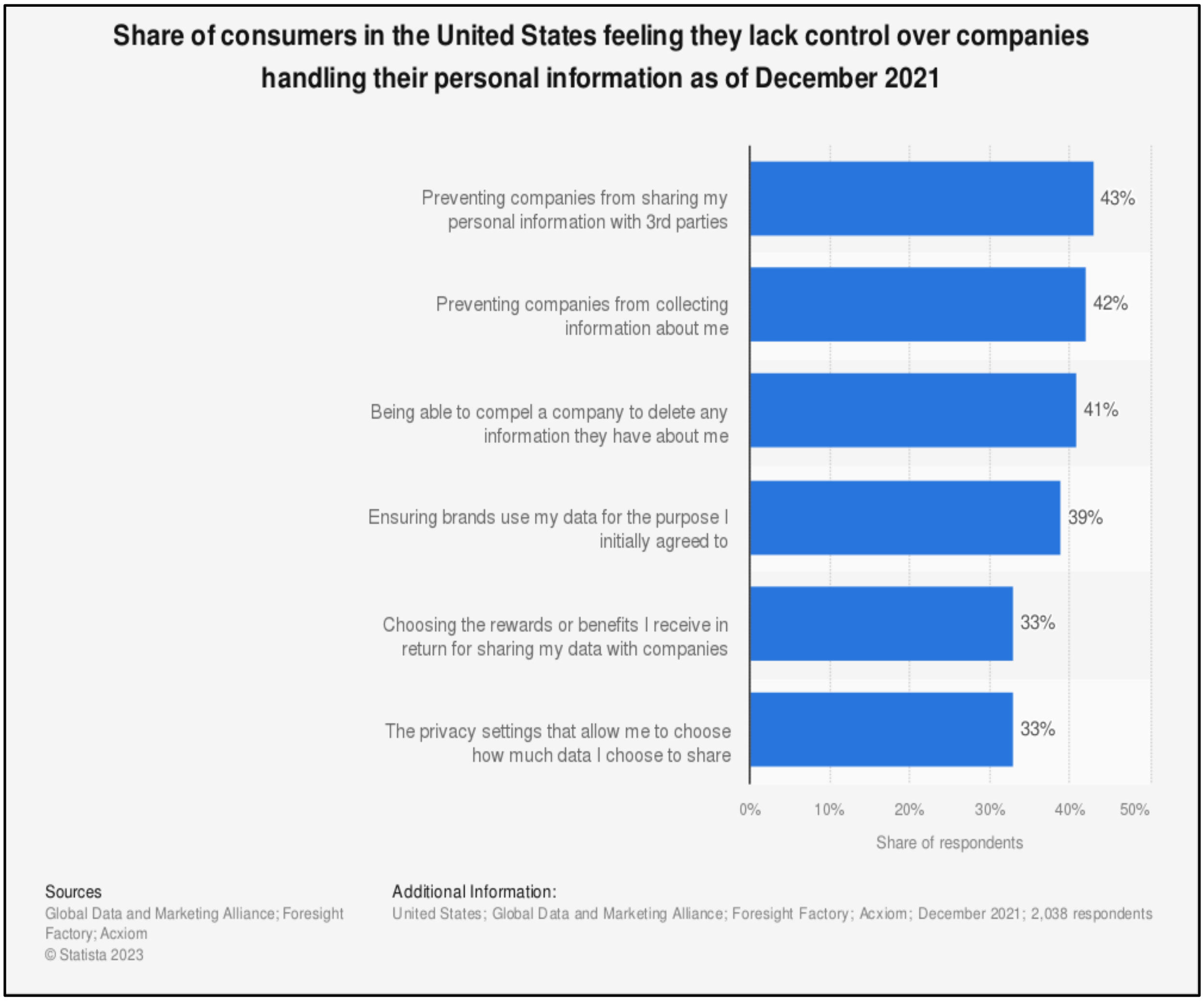
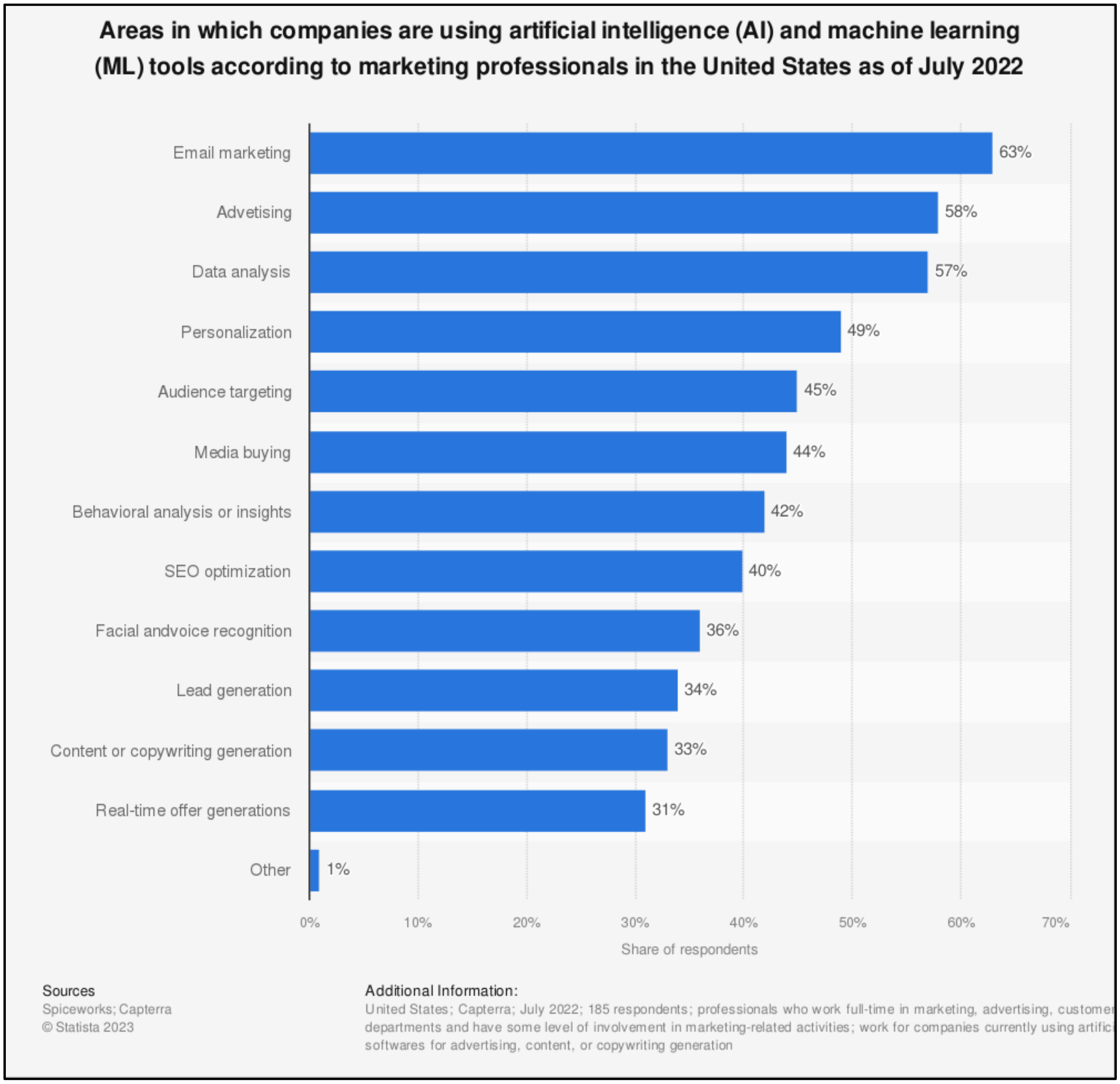


Figure 3. Share of consumers in the US feeling they lack control over companies handling their personal information (2021)

With the adoption of Artificial Intelligence (AI) and Machine Learning (ML), there are challenges when it comes to protecting user privacy. As these technologies use large volumes of data for training their algorithms, privacy concerns may arise if that data contains sensitive personal information that must be anonymized or processed appropriately. Companies have begun adopting AI/ML technologies across industries for various use cases - according to an October 2022 US survey [6] (*See Figure 4.*), 63 percent of marketers used these technologies as part of email marketing programs, while 58 percent employed them for advertising; 33 percent even said AI helped with copywriting tasks.



Sources
Spiceworks; Capterra
© Statista 2023

Additional Information:
United States; Capterra; July 2022; 185 respondents; professionals who work full-time in marketing, advertising, custome
departments and have some level of involvement in marketing-related activities; work for companies currently using artifici
softwares for advertising, content, or copywriting generation

Figure 4. Areas where US marketing professionals report the use of AI and ML tools.

User control over personal data, including its deletion or limitation, is vitally important. However, this can often prove challenging in practice when data is shared between multiple enterprises or sold to third parties. User privacy requires enterprises to balance user interests against those of themselves as enterprises as well as adhere to legal and ethical standards while utilizing new technologies that safeguard user rights. Enterprises should recognize these challenges and attempt to overcome them for sustainable businesses with data-driven businesses that work well for users and can build trust between themselves and users by working to address these problems head-on.

1.4 Statement of The Problem and Research Questions

The thesis analyzes the systemic issues facing enterprises and customers regarding user privacy. Employing a systems approach, this thesis attempts to identify how such issues influence customer perceptions and behaviors and explore technologies to help enterprises adopt customer-centric practices while mitigating privacy issues. Finally, this thesis offers information and recommendations for enterprises looking to increase user privacy.

Two technologies were examined, Federated Learning and Differential Privacy, to assist enterprises with taking a customer-centric approach to protect user privacy. Businesses can increase customers' trust, engagement, and loyalty by prioritizing transparency, user control, and informed consent, ultimately keeping customers at the center.

Research questions that the thesis explores further:

1. What are the major obstacles enterprises face when protecting user privacy, and how are these difficulties impacting customer perceptions and behaviors?
2. How can enterprises prioritize transparency, user control, and informed consent in their approach to user privacy, and what factors impact customer perception of these practices?
3. What strategies can enterprises employ to increase user privacy while balancing business needs?
4. How is Federated Learning used to maintain user privacy within enterprises, as well as what are its potential advantages and drawbacks?
5. What are the potential advantages and drawbacks of adopting a Differential Privacy approach in enterprises, and how might this strategy increase customer trust, engagement, and loyalty?
6. What factors drive or impede change within an enterprise environment?
7. What options exist to address customer issues?

Enterprises must adopt a systems approach to address user privacy challenges efficiently. There needs to be more than just the implementation of privacy-enhancing technologies at various points throughout the system. Although data protection technologies provide essential safeguards to user data, they only offer partial solutions by protecting individual infrastructure or information in isolation. A systems approach entails holistically studying all elements of a more extensive system and understanding that changes made to one aspect could cause unintended emergence elsewhere in its design. This approach offers a complete picture of the challenges of protecting user privacy, including stakeholders' roles, regulations and restrictions, and enterprise concerns. Adopting a systems approach to user privacy allows enterprises to better understand the interconnections and interdependencies within a system, identify risks or challenges more quickly, and develop more efficient solutions. Adopting such an approach helps enterprises build greater trust with users while strengthening reputations while guaranteeing a more secure digital future for all.

1.5 Thesis Structure

Chapter 1 introduces the topic of preserving user privacy. Chapter 2 discusses the challenges associated with collecting user data, risks related to identity theft, data breaches, online tracking, and legal and regulatory frameworks that govern user data privacy. Chapter 3 provides a literature review of privacy challenges and their impact on customer behavior. Chapter 4 focuses on Federated Learning as a solution to privacy concerns, analyzing its advantages and challenges. Chapter 5 explains Differential privacy as a solution to privacy concerns and its benefits and challenges. Chapter 6 discusses a customer-centric approach and provides a methodology, results, and strategies for enterprises to adopt this approach. Chapter 7 presents the system analysis. Chapter 8 summarizes the research, draws conclusions, and discusses study limitations and future research directions.

Chapter 2 - Understanding the Challenges in Preserving User Privacy

2.1 Types of User Data Collected by Enterprises

Enterprises collect user information from various sources to enhance their business, tailor marketing initiatives and make data-driven decisions. Collected data generally falls into three broad categories - personal data, behavioral data, and user-generated content (UGC).

Personal data encompasses information identifying an individual, such as their name, address, email address, telephone number, or social security number. Enterprises gather such data through various methods, including online forms, surveys, customer databases, or social media profiles, and use it for authentication, account opening, or communication purposes.

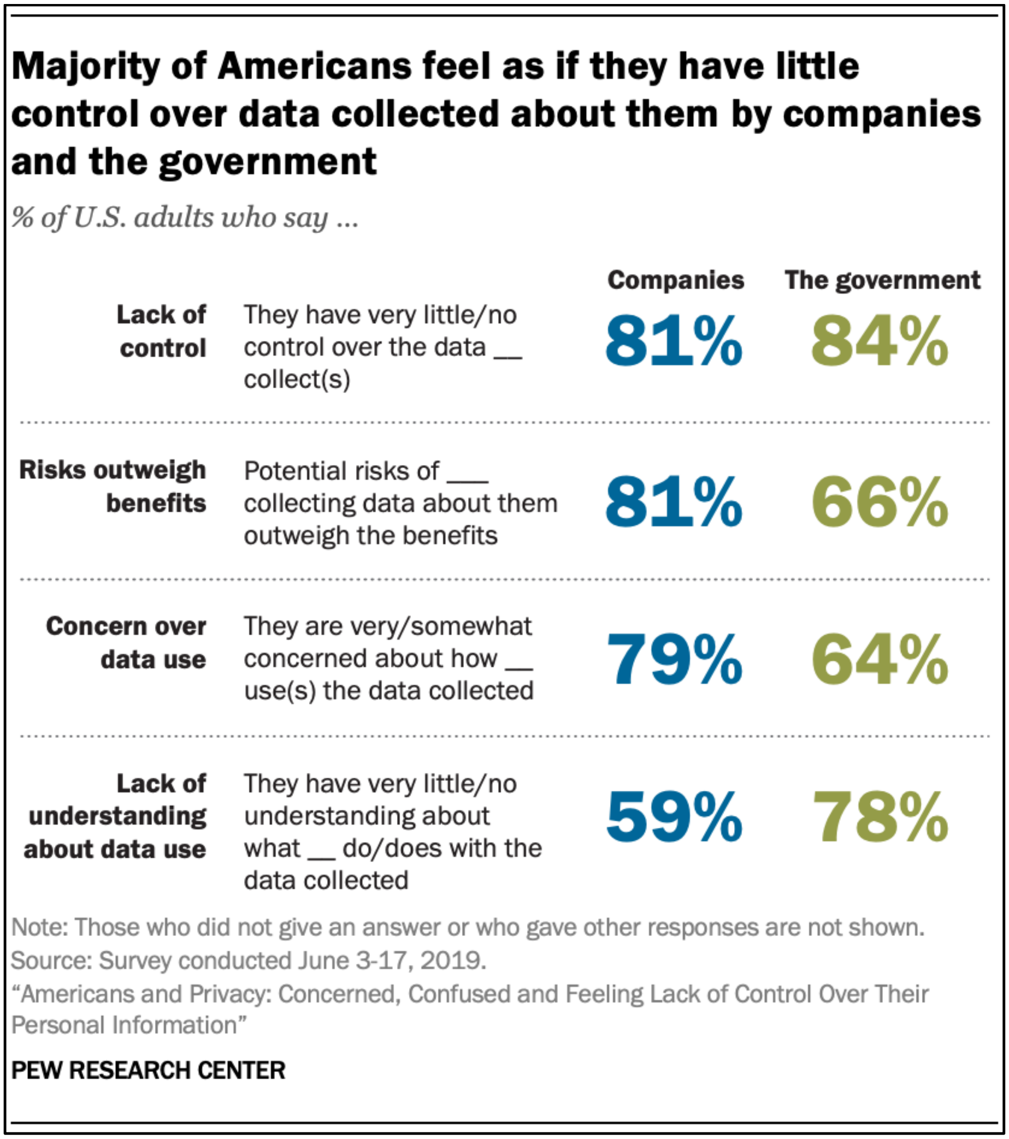
Behavioral data refers to how users engage with digital platforms like websites, mobile applications, and social media platforms such as websites. This could include apps used, content viewed, pages visited, time spent viewing pages or clicking links, search queries performed, and device identification details collected using tracking technologies like cookies, pixels, and beacons. Businesses use behavioral data analytics tools like these better to understand user behaviors, preferences, and interests, ultimately improving digital experiences and increasing engagement rates.

With the rise of social media platforms such as Facebook, Twitter, YouTube, and TikTok, user-generated content has become increasingly common; this includes all user-created materials, including comments, reviews, photos, and videos posted directly by its creators. Companies collect user-generated content from various sources, such as social media platforms, customer forums, and feedback forms. Although gathering this user data provides companies with invaluable insight, collecting it poses some unique challenges when protecting user privacy. One of the primary issues in business today is data breaches, any unauthorized access or theft of sensitive personal data by untrustworthy sources, whether that means accessing it directly or via third parties, such as hackers or employee negligence. While such breaches occur for various reasons (weak security measures and third-party breaches), employee negligence remains one of their leading causes. Data breaches can seriously affect businesses and users, including identity theft, financial fraud, reputational harm, and legal liability. One challenge of online tracking consists of monitoring user activity without their knowledge or consent; such monitoring could lead to privacy violations such as targeted advertising campaigns targeting specific areas, data profiling for marketing purposes, and identity theft.

2.2 How Customers Feel About the User Data That Enterprises Collect

According to a 2019 Pew Research Center survey [7] (*See Figure 5.*), approximately 81 percent of adult Americans indicated limited control over how companies gather their data, and 79 percent are worried about how those companies intend to utilize that data. Survey results also found that 72 percent of American adults considered control over who could access their personal information essential, with 69 percent concerned about companies using such data for advertising. Additionally, 64 percent of adults in the US reported needing to be made aware that personal information had been collected about them by companies and enterprises without their knowledge. At the same time, 61 percent felt the risks outweigh the benefits associated with data collection by these businesses. Overall, these results point toward significant privacy issues among most American adults.

A survey [8] (*See Figure 6.*) conducted in the United States in January 2023 about consumer attitudes toward brands' usage of AI and ML found that 45 percent of the participants did not understand how AI and ML technologies worked. Despite this, 73 percent of the respondents recognized the potential of AI and ML to impact customer experience (CX).



Majority of Americans feel as if they have little control over data collected about them by companies and the government

% of U.S. adults who say …

		Companies	The government
Lack of control	They have very little/no control over the data __ collect(s)	81%	84%
Risks outweigh benefits	Potential risks of ___ collecting data about them outweigh the benefits	81%	66%
Concern over data use	They are very/somewhat concerned about how __ use(s) the data collected	79%	64%
Lack of understanding about data use	They have very little/no understanding about what __ do/does with the data collected	59%	78%

Note: Those who did not give an answer or who gave other responses are not shown.
Source: Survey conducted June 3-17, 2019.
"Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information"

PEW RESEARCH CENTER

Figure 5. Pew Research - Control over data collected

Furthermore, 48 percent stated that they would interact with AI more frequently if it improved their CX with a brand by making it more seamless, consistent, and convenient. Customers' perceptions of AI and ML vary based on their experience, but they do not necessarily understand how they work.

Companies must find ways to educate customers on AI and ML technologies. Almost half of the respondents still need to understand them by teaching customers about using these technologies properly and improving customer experiences. Businesses must be open about data collection and usage practices to foster customer trust. This should involve clear communications to inform customers what data is collected, why, and how it is being utilized. Companies should establish strong privacy policies that adhere to existing regulations; they can employ privacy-enhancing technologies (PETs) to secure customer data while using it to enhance CX. Educating customers and using PETs allows companies to leverage data effectively to enhance customer experiences while protecting privacy.

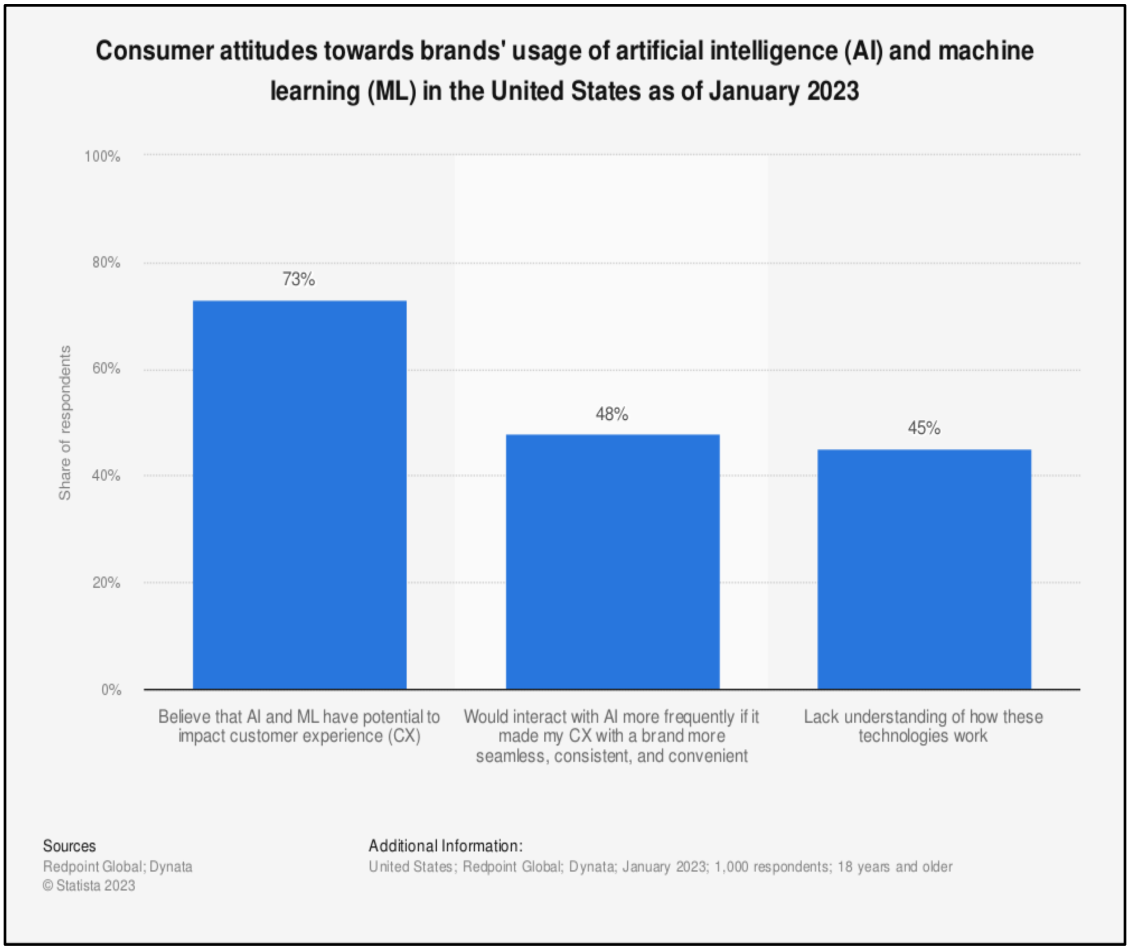


Figure 6. Consumer attitudes towards brands' usage of artificial intelligence (AI) and machine learning (ML) in the US (2023)

Data misuse arises when user data is collected or utilized in ways not agreed upon or disclosed. Misuse could take many forms, such as unclear policies, lack of user consent, and unethical practices that compromise individual privacy. Misuse of data by enterprises can result in loss of trust, damaged reputation, and potential legal ramifications. Aside from that, lack of transparency poses additional issues when enterprises must provide easy to comprehend details regarding their data collection and processing practices. Lack of transparency may cause user disorientation, distrust, and disagreement; without being fully informed about the usage and protection of personal data they share online, many may hesitate. An American study [9] (*See Figure 7.*) on online personal data use and privacy found that over half of the respondents in 2021 felt more concerned about protecting their online privacy, with 9 percent less concerned than before. This shows the urgency on how enterprises need to act.

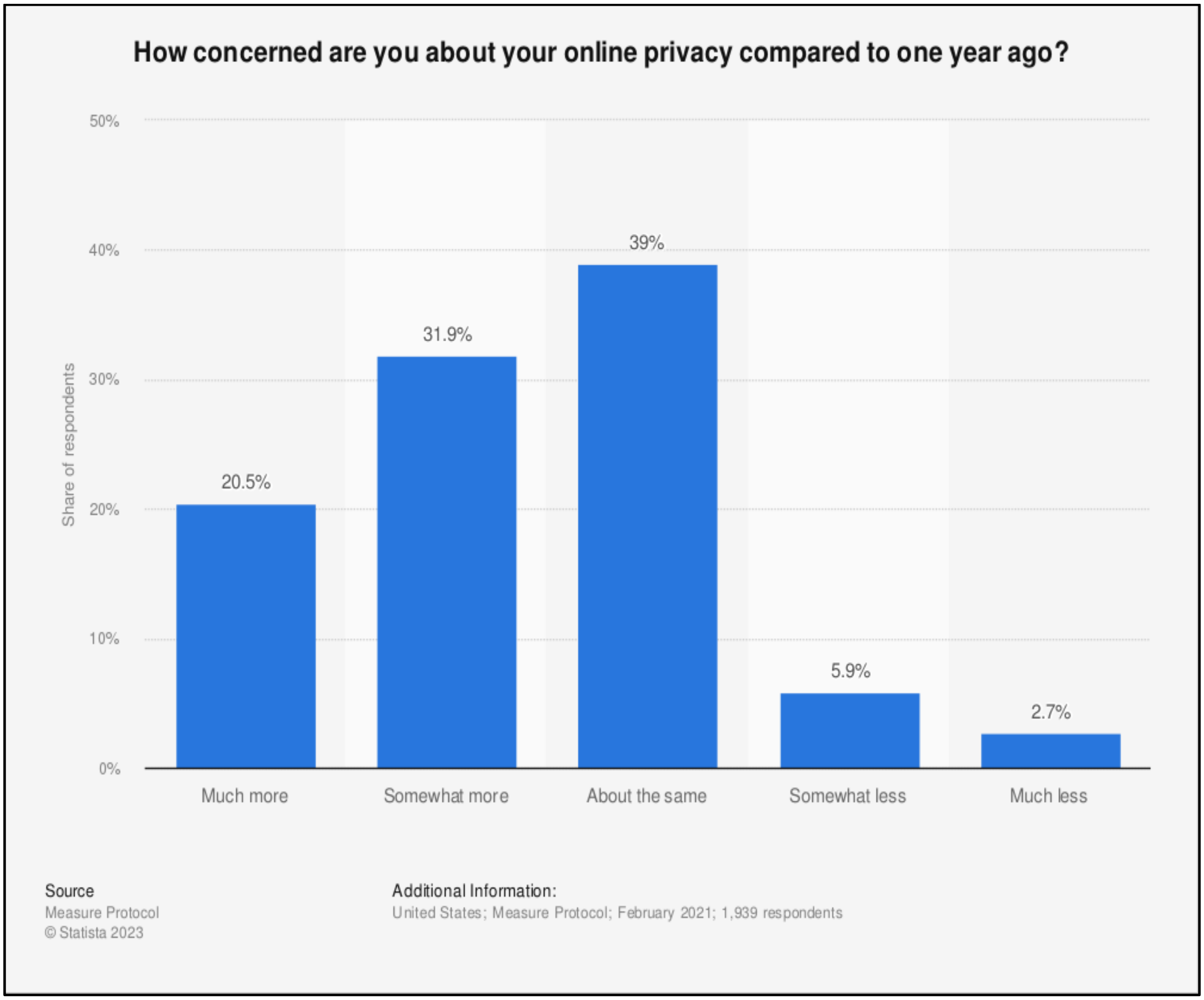


Figure 7. US users online privacy concerns compared to 2020 (2021)

Data retention presents challenges when companies store user data longer or insecurely than necessary or legally mandated. Improper storage increases privacy risks the longer the information remains stored, increasing its vulnerability to access or theft. Additionally, it could create compliance issues as many data protection laws require businesses to keep user information only as long as necessary and legally necessary. Regulatory compliance remains vital in protecting user privacy as companies must abide by various data protection laws, such as GDPR in Europe or CCPA in the United States, requiring enterprises to implement various protection measures, including anonymization, user consent, and data subject rights.

Enterprises gather user data from different sources, such as personal and behavioral records and user-generated content. While gathering this information can provide businesses with invaluable insights, collecting user data also poses many privacy threats that must be considered, including data breaches, online tracking tools that track individuals online, data misuse issues that hinder transparency as well a lack of control of collected user information.

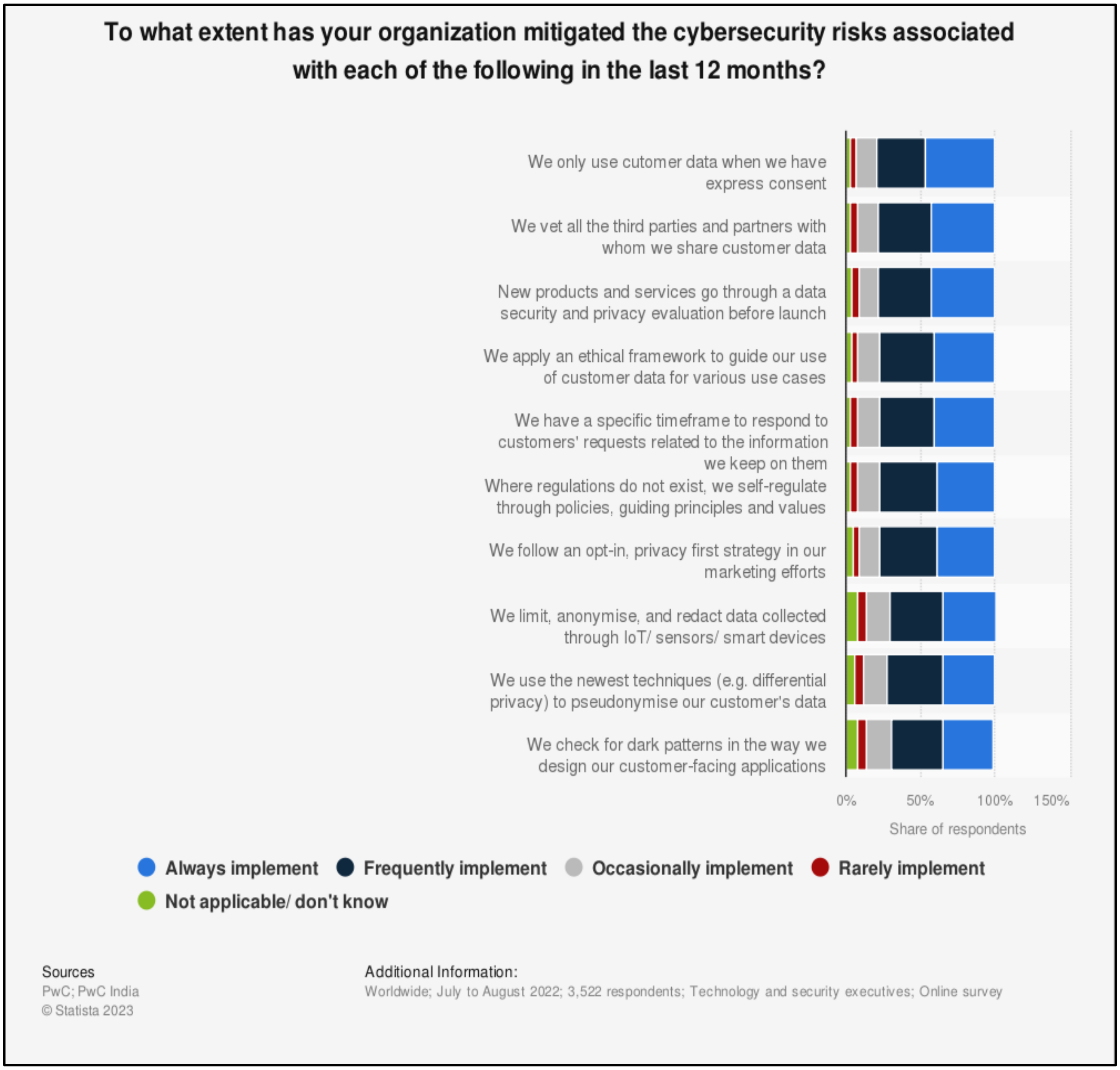
2.3 Risks Associated with Collecting User Data, Including Identity Theft, Data Breaches, And Online Tracking

Accumulating user data has become standard practice across various industries, from e-commerce and healthcare to financial services and social media. However, user data's widespread collection and usage have caused serious privacy and security risks and concerns, such as identity theft, data breaches, and online surveillance. Identity theft poses one of the greatest dangers when collecting user information. Cybercriminals may use user information like name, address, phone number, and social security number to assume another's identity and commit identity theft attacks that cause significant financial losses and damage their credit histories. Criminals could then use stolen identities for illegal activities like fraud and money laundering, leading companies to implement strong encryption and access control mechanisms when securely collecting and storing user data.

Data leakage poses another significant threat when collecting user data. Data breaches occur when unauthorized individuals access sensitive user information like credit card data, social security numbers, or email addresses collected for collection. Data breaches may occur for various reasons, including hacking, phishing, and malware attacks. Their consequences can be far-reaching: identity theft, financial loss, and reputational harm to users as well as legal ramifications that lead to lost customers and business reputational loss for businesses. As previously discussed, companies must implement stringent security measures like multi-factor authentication, intrusion detection and prevention systems, and employee awareness training to safeguard sensitive user data from breaches. As noted, online tracking presents another serious security threat regarding user data collection. Online tracking involves collecting users' online activities, such as browsing history, search query data, and location details to display targeted ads that meet a person's interests and preferences. Ad networks commonly employ this strategy to show targeted advertisements based on those interests; however, malicious actors could misuse this surveillance technology for criminal acts like identity theft and cyberstalking if left unmonitored by companies and allow their customers to opt out of further tracking practices. Companies must therefore disclose all practices associated with tracking users online while giving individuals an opt-out mechanism in case malicious actors misuse it for unlawful criminal purposes or criminal acts like identity theft and cyberstalking activities conducted online surveillance can use. Therefore, companies must disclose policies concerning tracking practices while giving customers options that allow opting-out options when possible. Hence, users have greater freedom to control what companies' monitoring efforts they engage into the opt-out option of tracking practices which could involve malicious actors engaging in criminal acts related to identity theft and cyberstalking.

User data collection has become widespread across industries but has also created significant privacy and security risks and concerns. Identity theft, data loss, or tracking are the primary threats to user data collection practices. To minimize such threats and protect user privacy and security effectively, companies should prioritize implementing strong security measures alongside transparency surrounding user data collection practices to mitigate these risks.

Respondents of an organization cybersecurity risk mitigation survey [10] (*See Figure 8.*) from 2022 reported that in the previous 12 months, their company used customer data only when given explicit permission to limit cybersecurity risks; 6 percent stated their organization limited, anonymized, or redacted data collected via IoT devices like sensors or smart devices frequently or rarely.


To what extent has your organization mitigated the cybersecurity risks associated with each of the following in the last 12 months?

Sources
PwC; PwC India
© Statista 2023

Additional Information:
Worldwide; July to August 2022; 3,522 respondents; Technology and security executives; Online survey

Figure 8. Global companies actions to minimize cyber risks (2022)

Privacy regulations could be violated, and individuals could become vulnerable to identity theft, stalking, and other unlawful actions. Furthermore, improper anonymization and redaction practices could expose confidential or proprietary data that would compromise the company's reputation and competitive edge; as a result, companies should implement sound data collection and management practices incorporating anonymization and redaction procedures to safeguard personal information as well as company interests.

Companies seeking to maximize data privacy and cybersecurity must regularly train employees and conduct audits and reviews of their practices, covering managing customer data and responding to security incidents during these sessions. Audits/reviews should cover reviewing privacy policies, data access controls, and security protocols for risks or areas for improvement - third-party audits can provide further reassurance if needed.

2.4 Current Legal and Regulatory Frameworks

Concerns over user data privacy have increased dramatically worldwide over recent years. Countries and international enterprises have implemented legal and regulatory frameworks designed to safeguard user data privacy - this section details them all.

One of the most influential data privacy and protection laws is the General Data Protection Regulation (GDPR), introduced by the European Union in 2018. Under this comprehensive regulation, companies that collect personal information on EU citizens regardless of location must seek explicit user permission before collecting their data, be transparent about its usage, and request its deletion - something companies were previously not required to do under previous data regulations.

The United States does not possess comprehensive federal privacy legislation, but individual states enact localized legislation regarding data collection. California introduced the California Consumer Privacy Act (CCPA) in 2018. Under CCPA regulations, companies that meet specific criteria. For instance, having annual gross revenues exceeding $25 million and collecting personal data of at least 50,000 customers annually fall under its scope, giving users rights such as knowing which data has been collected on them as well as request deletion and opt-out options of any further sale of that data to third parties.

Other countries have also implemented privacy laws. Canada introduced the Personal Information Protection and Electronic Documents Act (PIPEDA) in 2000; this statute covers private sector enterprises which collect, use, or disclose personal data as part of business activity. Under PIPEDA, companies must secure prior consent before collecting personal information and be transparent in their data practices. Users also have access to their own data and the option of rectifying or erasing it at their leisure. Asia-Pacific Economic Cooperation (APEC) also developed the Privacy Shield Framework in 2005. APEC's Privacy Framework offers member economies a set of principles they can use when crafting privacy legislation in their economies; emphasis is also placed on transparency, accountability, and user participation for protecting privacy.

Additionally, numerous industry self-regulatory frameworks exist. For example, Interactive Advertising Bureau (IAB) has introduced its online advertising framework, which features data collection and usage guidelines. This framework highlights the significance of transparency and user control. However, even with these systems in place, there remain difficulties when protecting user data - one major challenge is enforcing laws across borders. Imagine, for instance, that a company resides in one country but processes data from users in another; in such an instance, it could prove challenging to enforce both countries' privacy regulations simultaneously. Furthermore, technological change makes compliance with regulations complicated due to novel risks or challenges presented.

Legal and regulatory frameworks for user data protection can be complex and varied across countries and enterprises, each adopting its laws and guidelines to safeguard the personal information of its citizens. Although such systems provide essential protections for user data privacy, challenges remain in ensuring these laws and regulations work across borders as technology rapidly changes.

Chapter 3 - Literature Review

3.1 Overview of Privacy Challenges and Their Impact on Customer Behavior

Technology companies like Amazon, Google, Meta, Netflix, and Apple have revolutionized how customers engage with technology by increasing connectivity and access to data. However, while convenience may present advantages, this also creates issues surrounding privacy. Privacy issues arise when data collected or used without consent or knowledge by an individual may significantly influence customers' attitudes, preferences, and purchasing decisions - leading to changes in attitudes, preferences, or purchasing decisions altogether. The research investigates challenges such as 1) data breaches, 2) cyber security threats, 3) surveillance, and 4) online tracking impacting customer behavior.

Data breaches occur when an unauthorized person accesses sensitive data without authorization and misuses it, including identity theft, financial losses, and damaged reputations. As data breaches increase in frequency and severity. For instance, Equifax suffered one in 2017 that exposed [11] over 140 million customers' Social Security numbers, birth dates, and addresses.

Cyber security threats are any methods employed to attack computer systems, networks, and devices to inflict damage, steal information, or disrupt operations. Cybersecurity attacks come in various forms, such as malware, phishing scams, ransomware attacks, and denial-of-service attacks, which pose substantial financial and legal harm. They pose considerable financial loss, reputational risk, or legal liabilities for individuals and enterprises.

Surveillance refers to any act or behavior monitored without their knowledge and consent, often without them knowing or agreeing. Monitoring may take various forms, such as video surveillance, tracking cookies, or social media monitoring; legitimate uses include crime prevention or public safety improvements, while malicious purposes, such as identity theft, may also use surveillance technology.

Online tracking involves collecting data about an individual's activities for advertising or other commercial uses, typically cookies and web beacons. However, other tracking technologies could also be employed to collect this data. While tracking can provide invaluable data that improves user experiences by personalization of services tailored specifically for each user, its primary use, though, remains personalization: tailoring content and services specifically to individuals using user data collected during tracking sessions - but doing so often raises privacy issues due to collecting and using sensitive personal data in this process.

Impact on Customer Behavior

Privacy issues can erode customer trust in technology, brands, and institutions, when individuals perceive that their personal information must be adequately secured before using certain online services or technology solutions. If this becomes an issue for individuals, they may hesitantly opt-out.

Privacy concerns may result in reduced usage of online services, with individuals becoming reluctant to provide personal details or participate in activities online. This could seriously hamper businesses that rely on this form of engagement like social media platforms, e-commerce websites, and gaming companies that rely heavily on internet engagement for success.

Privacy concerns can also influence customers' purchasing patterns as customers become more selective about which brands and products they engage with and purchase. A study by Cisco [12] found that 84 percent of consumers said they would not do business with a company if they had concerns about the security of their data. These privacy worries have generated an increased need for protection with customers looking for ways to safeguard their data such as installing ads blockers, VPNs, or encrypted messaging apps to safeguard themselves online tracking or surveillance leading them down new paths of protecting themselves against being monitored online tracking or surveillance and tracking methods akin to be avoided altogether.

Privacy concerns have increased governmental and regulatory scrutiny, prompting governments and regulators to implement new standards designed to protect customer privacy. As previously discussed, the European Union introduced GDPR in 2018, giving individuals greater control of their personal data while setting strict requirements on companies that collect, use, or share it. Other states have similarly established similar regulations, including California Consumer Privacy Act or "CCPA." Privacy issues have long been a great concern among customers in today's technological environment, where data breaches, cyber security attacks, surveillance programs, online tracking technologies, and personalization pose substantial threats to privacy and security. These issues can have an enormous effect on customer behavior, leading to decreased trust and engagement, altered shopping patterns, increased privacy protection needs and regulatory ramifications, and customer education about these laws and regulations.

3.2 A Review of Privacy-related Concerns

Customers commonly view privacy as one of their top priorities regarding their personal data, with more customers becoming increasingly conscious about the risks involved when sharing such data and becoming less open to sharing such details with third parties. Studies show that customers appear more reluctant to provide this kind of personal information over time and share it freely as more potential risks become clear to them and, thus, more hesitation about sharing personal information publicly with companies and governments.

A study [13] conducted by the Journal of Consumer Research demonstrated that customers share personal information more freely with companies they trust and those transparent concerning data practices. They may even share more if they believe something valuable, like customized shopping experiences and exclusive discounts, is received in return. Another study [14] published in the Journal of Marketing has demonstrated that customers are more inclined to share personal data with companies who use it for customized customer experiences rather than simply marketing purposes. Customers were more willing to give out information if they could easily control its use and opt-out of sharing such details with businesses.

Culnan and Bies [15] discuss the significance of striking an equilibrium between economic and justice considerations regarding consumer privacy. While consumers may share personal information in exchange for discounts or personalized services, such as discounts or personalized recommendations from sellers, their right to control how this information is collected, used, shared, or destroyed should also be respected. Furthermore, identity theft risks must also be highlighted along with fraud. Lastly, they emphasize the need for companies to prioritize privacy when collecting, using, and sharing personal data, with clear and transparent communications from suppliers regarding data practices to their consumers.

Milne and Culnan [16] suggest using the Privacy Management Framework to evaluate electronic commerce privacy policies. According to Milne and Culnan, many privacy policies may need to be simplified or insufficiently protect personal data for consumers to navigate them effectively. The privacy management framework offers guidelines to evaluate privacy policies on their content, accessibility, and effectiveness. They apply a framework of privacy policies to an example sample of e-commerce privacy policies and find that many need revision to provide clear and understandable data collection and use information. Furthermore, they urge companies taking part in online commerce transactions to put greater attention and efforts toward strengthening and clarifying their own data use and privacy protection policies.

Another interesting paper is the Acquisti and Grossklags [17], which addresses privacy's role in individual decision-making and rational consideration of any associated choices. They assert that individuals often make privacy-related decisions based on insufficient or false assumptions regarding risks and benefits associated with sharing personal data. They present a framework for analyzing privacy decisions that consider rational and non-rational considerations, including social influence and cognitive biases. They outline the implications for policy making, suggesting that policymakers consider rational and non-rational influences that impact privacy-related choices.

Wang and Emurian [18] provide an in-depth examination of online trust, covering its concepts, elements, and implications. According to them, trust is an indispensable asset when conducting transactions online as it helps alleviate perceived risks and uncertainty associated with online interactions. An inclusive framework for understanding online trust, consisting of four essential components - credibility, reliability, intimacy, and self-orientation is proposed. They explore trust's implications for e-commerce, specifically how it influences consumer behaviors, vendor reputations, and overall online business success. They suggest future research areas, including developing new measures of online trust and exploring cultural or contextual influences that might influence trust online in different settings.

Li, Liang, and Wang [19] conducted a field experiment to assess the impacts of personalization, privacy concerns, and social influences on mobile advertising's efficacy. Study participants were recruited from a shopping mall and instructed to use a mobile app for clothing shopping using two types of ads: personalized and non-personalized. Study results demonstrated that personalized mobile ads were more successful in drawing in users and inciting purchase intentions than non-personalized ones; however, privacy concerns negatively impacted attitudes toward personalized ads as well as the purchase intentions of respondents. Finally, the study revealed that social influence positively moderated the relationship between personalization and purchase intentions suggesting it might mitigate some of the adverse reactions related to privacy concerns on personalized ads. They propose that mobile advertisers be mindful of users' privacy concerns when designing personalized ads; those designed should be nonintrusive, transparent, and provide clear value propositions for potential consumers. They suggest mobile advertisers utilize social influencers effectively when marketing personalized advertisements to their audience.

According to the papers, customers want more control over their data. As a result, companies should try to balance their demands with building client trust. This can be done through protecting privacy, using e-commerce to influence decision-making, and lowering the perceived danger of data breaches. Businesses can also employ personalized advertising while respecting the privacy requirements of their clients.

3.3 Existing Solutions and their limitations

Overall, customer awareness about the risks involved with sharing personal data has increased, prompting increased prudence from them when sharing it. Businesses prioritizing transparency and personalization while giving customers control over their information may find success in winning back their trust and earning their business.

Personal data collection, use, and sharing have become ever more frequent, making privacy laws and regulations even more vital than before. They serve to protect individual rights by outlining how companies and enterprises must collect, process, and utilize personal information. One of the critical pieces of privacy legislation, GDPR, stands out among others as one of its major provisions. The GDPR imposes numerous obligations upon companies who collect and process personal data, such as gathering consent from individuals for processing activities and providing transparent information on data processing practices; it also offers individuals rights to access and control their data, such as deletion.

US citizens also benefit from several federal and state privacy regulations in addition to GDPR, the CCPA. Under CCPA, Californians have the right to know what personal data is being collected about them, have any data deleted if requested, and opt out of having their information sold by companies. In addition, companies must provide clear privacy notices with reasonable security measures implemented as required under CCPA.

Privacy laws and regulations provide legal obligations on companies to safeguard personal data; however, their implementation can present several obstacles. Countries with weak legal frameworks or limited resources may need help enforcing them properly, while small businesses with limited resources may find it challenging to comply with them; jurisdictional issues become apparent for enterprises operating across multiple locations where privacy regulations differ significantly from country to country.

Education and awareness campaigns can provide nontechnical solutions that empower individuals to make educated choices regarding their privacy. Such campaigns aim to inform individuals about the risks associated with sharing personal information online, provide guidance for protecting personal data and introduce laws and regulations related to privacy-enhancing technologies - but these campaigns also come with their share of limitations. Education and awareness campaigns can only reach some people, those without active online communities or access to digital resources may need to be made aware of potential privacy risks; even once aware, individuals may only change their behaviors if it benefits them immediately to protect their data.

There are several protections that the laws can provide, but the customers are not aware of them. According to an April 2019 survey [20] (*See Table 2.)* of adults in the US, 58 percent had never heard of Payment Card Industry Data Security Standard (PCI-DSS), and only 16 percent knew its basics compared with 46 percent who knew about Health Insurance Portability and Accountability Act (HIPAA). Therefore, individuals, businesses, and governments must work collaboratively to address this challenge while upholding customer privacy in an ever-evolving digital sphere.

	I've never heard of it	**I've heard of it, but that's about it**	**I know the basics**	**I know a great deal and know what steps to take to ensure compliance**
General Data Protection Regulation (GDPR)	42%	26%	27%	5%
California Consumer Privacy Act (CCPA)	46%	31%	16%	6%
Health Insurance Portability and Accountability Act (HIPAA)	11%	19%	46%	25%
Payment Card Industry Data Security Standard (PCI-DSS)	58%	21%	16%	5%

Table 2. US familiarity with digital privacy regulations and guidelines 2019

Limited resources can also constrain educational and outreach activities. Designing and implementing effective education and awareness campaigns can be challenging, especially when resources are limited. An integrated approach that combines technical, legal, and educational solutions can effectively address privacy concerns. Such an approach may include technological solutions like encryption and privacy-enhancing technologies, legal solutions like privacy laws and regulations, and educational solutions like privacy awareness campaigns.

Current solutions to address privacy concerns have limitations that prevent them from eliminating privacy risks. Therefore, it is necessary to develop comprehensive and multidimensional approaches that take a systems approach that integrates technical, legal, and educational solutions to address privacy issues effectively. These solutions should also acknowledge the limitations of current approaches and strive to overcome them to protect personal privacy in the information era.

3.4 The Role of a Systems Approach in Preserving User Privacy

3.4.1 Systems Approach and How It Can Be Used to Analyze Privacy Issues

A systems approach [45] is a way of thinking that considers the relationships between the various components of a complex system. Regarding privacy concerns, a systems approach can help identify the stakeholders involved, and the data flows between them, and the potential privacy impacts. The systems approach is based on recognizing that privacy is not just an individual problem but a societal problem involving multiple stakeholders, including individuals, enterprises, governments, and regulators. By providing a comprehensive view of the privacy ecosystem, a systems approach can help identify gaps in current privacy practices and improve opportunities. Applying a systems approach to privacy issues requires identifying the various components of the privacy ecosystem and their interrelationships. See *Table 3.* for more information on stakeholder goals, responsibilities, and accountability.

By analyzing the relationships between these components, a systems approach can help identify areas of most significant privacy risk and areas where action is needed. For example, a system approach can reveal the dominance of a particular technology provider in the market and thus significantly impact the privacy practices of data controllers and data processors. A systems approach may also reveal that the legal framework in each jurisdiction needs to be stronger and therefore does not provide sufficient protection for everyone. Overall, a systems approach can help identify and address privacy issues in a more holistic and integrated manner, recognizing the complexity of the privacy ecosystem and the interdependence of its various components. By applying this approach, more effective and sustainable solutions to privacy problems can be developed that benefit all stakeholders.

3.4.2 The Key Components of a Systems Approach, Including Stakeholders, Goals, And Interactions

Stakeholder Salience Analysis

Stakeholder salience analysis allows companies and enterprises to quickly identify and prioritize stakeholders based on the level of influence and impact on data privacy policies and procedures. By understanding each stakeholder's needs and expectations, enterprises can design data privacy strategies that meet everyone's interests while remaining compliant with regulations.

Customers are the primary stakeholders since they use the products or services, and their privacy is directly affected by data protection practices and procedures. Employees are key stakeholders, as they are charged with processing customer data and protecting it from unintended access, use, or disclosure. Management is also an integral stakeholder as they must establish and uphold an information privacy program that complies with applicable laws and regulations while still reflecting their company values and reputation. IT, Legal, and Data Privacy Office (DPO) departments are primary stakeholders as they are accountable for implementing, monitoring, and enforcing privacy policies and procedures within their companies. Regulators also play an essential role in enforcing privacy laws and regulations and investigating violations against them by penalizing companies accordingly.

Privacy advocates, industry associations, partners/vendors, shareholders, and media/reporters are secondary stakeholders because their actions directly influence company data privacy policies and procedures. Privacy advocates can guide businesses on complying with privacy requirements and best practices while partners/vendors process customer data on behalf of companies; shareholders can hold management accountable for privacy or errors while media/reporters investigate breaches in privacy protection at companies as well as hold them responsible.

Stakeholders	Goals	Responsibilities	Accountable
Customers	Protect privacy, gain transparency, and control	Contact companies to provide personal information and exercise privacy rights	Receive timely notifications about data breaches
Employees	Access and use data for legitimate business purposes, maintain confidentiality and integrity of data	Interact with IT, report privacy breaches or incidents	Apply technical and organizational measures to protect data

Management	Comply with privacy laws and regulations, and allocate adequate resources and budget.	Establish and maintain privacy program, assess, and manage privacy risks	Develop a culture of privacy and data protection
IT Department	Protect data from unauthorized access, use, or disclosure, and ensure compliance with policies and procedures	Implement technical security measures, monitor data processing activities	Collaborate with other stakeholders
Legal Department	Ensure compliance with privacy laws and regulations, and manage legal disputes or investigations	Work with regulatory agencies, industry associations, or civil society enterprises.	Join an industry association or standards body
Data Protection Officer	Monitor privacy policies and procedures, ensure compliance with them, and act as a liaison between the company and other stakeholders	Monitor and report data processing activities, educate and train staff	Foster a culture of privacy and trust within the organization
Privacy Advocates	Improve consumer privacy, raise awareness of privacy risks and best practices, and provide feedback, criticism, or support to the company's privacy initiatives or policies	Communicate with the company and other stakeholders, and lobby for changes to laws or regulations	Advocate for privacy protections and data protection regulations
Regulators	Enforce privacy laws and regulations, and provide guidance and support to businesses and other stakeholders	Investigate and sanction companies that violate privacy laws, and collaborate with other regulatory agencies or international enterprises	Align privacy standards and practices
Industry Associations	Establish voluntary privacy standards or best practices, provide privacy training and support, and represent industry interests in discussions with	Develop privacy standards or best practices, provide privacy training and support, and represent industry interests in discussions with regulators or privacy advocates	Develop policies and practices that reflect community values and preferences, and advocate for privacy protections and data protection regulations

	regulators or privacy advocates		
Partners/Vendors	Comply with the company's privacy policies and procedures, and protect customer data against unauthorized access, use, or disclosure	Implement appropriate technical and organizational measures, and notify the company of privacy breaches or incidents	Protect customer data against unauthorized access, use, or disclosure
Shareholders	Obtain information about the company's privacy practices and performance, and hold management accountable for privacy breaches or errors	Strengthen privacy protections, and support responsible use of data within the company	Make informed investment decisions
Media/Reporters	Report on privacy news, events, and controversies. Investigate and expose company privacy violations or deficiencies, and educate the public about privacy risks and best practices	Report on privacy news, events, and controversies. Investigate and expose company privacy violations or deficiencies, and educate the public about privacy risks and best practices	Provide guidance and support to customers affected by privacy incidents
Civil Society Enterprises	Advocate for stronger privacy protections and encourage companies to adopt privacy-respectful practices that protect consumers' rights to privacy	Raise public awareness of privacy risks and issues and educate consumers on protecting their privacy	Advocate for stronger privacy protections and promote the adoption of privacy-respectful practices by companies. Lobby for the enactment of privacy legislation that protects consumers' privacy rights

Community	Address community concerns about privacy and demonstrate the company's commitment to responsible data stewardship and privacy protection	Engage with the community to understand their privacy concerns and communicate the company's data privacy policies and procedures	Host community events to discuss privacy issues and share information about the company's privacy practices. Use social media and other communication channels to engage with the community and address their privacy concerns. Participate in community initiatives that promote responsible data stewardship and privacy protection
Competitors	Monitor the privacy practices of competitors and use them to gain a competitive advantage by demonstrating a higher commitment to privacy protection	Conduct market research to understand competitors' privacy practices, identify areas of differentiation, and use them to inform their privacy strategy	Benchmark the privacy practices of the competitors and use the findings to improve their data privacy policies and procedures. Highlight the company's superior commitment to data privacy in the marketing materials and customer communication

Table 3. Stakeholders' analysis

Interactions Between the Stakeholders

Customers can contact the companies to provide personal information and exercise their privacy rights, such as accessing, rectifying, or deleting data.

Employees may interact with IT to access and use customer data for legitimate business purposes and receive training on privacy policies and procedures. Management may contact the DPO to establish and maintain a privacy program that complies with applicable laws and regulations and to assess and manage privacy risks.

IT departments may contact vendors or partners to implement technical security measures and ensure that data is protected in transit and at rest. The legal department may work with regulatory agencies, industry associations, or civil society enterprises to resolve legal or ethical issues related to data privacy or to manage legal disputes or investigations.

Privacy advocates can contact companies to provide feedback, criticize a company's privacy practices, or lobby for greater privacy protections. Regulators may contact companies to review or audit privacy practices, issue guidelines or sanctions, or enter settlement agreements.

Industry associations may contact companies to establish voluntary privacy standards or best practices or to share information and experiences on privacy issues. Shareholders may contact the company to express concerns about potential financial or reputational risks from a breach of confidentiality or to encourage the company to prioritize privacy and security.

Media may contact the company to report privacy violations or disputes or to indicate the company's privacy policies and practices. Civil society enterprises may contact us to improve privacy protections, engage in public campaigning or legal action, or raise awareness of privacy risks.

Competitors may contact the company to inquire about its privacy practices or to criticize or imitate such practices. Communities can contact the company to raise data privacy concerns or expectations or to start a conversation about the company's impact on the environment or local economy.

The intricate relationships between many stakeholders involved in safeguarding client privacy are depicted in Figure 9. Customers, corporations, government agencies, privacy specialists, and technology providers are just a few stakeholders. The figure emphasizes how various stakeholders are interdependent, with one stakeholder's actions affecting the others and creating a feedback loop.

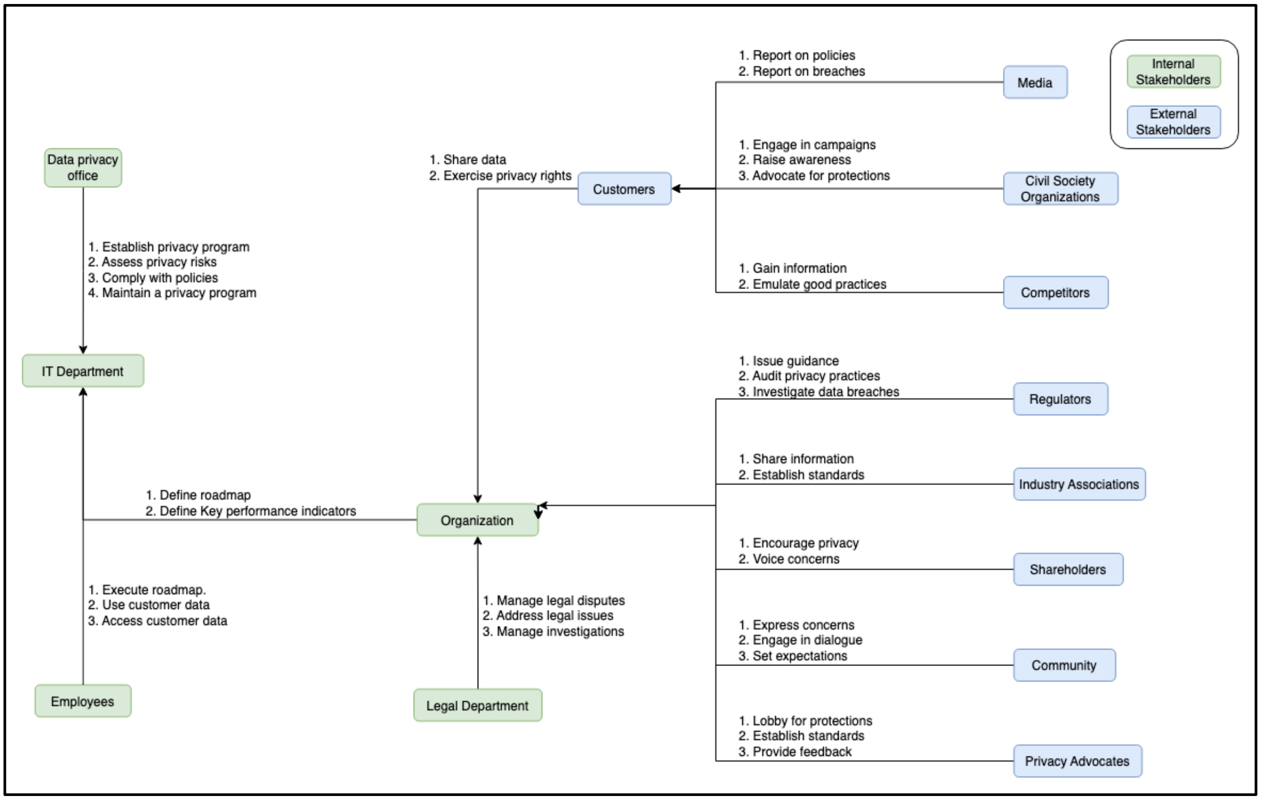


Figure 9. Stakeholders and their interactions

Goal Importance to the Business Vs. Stakeholders

Table 4. presents various goals related to data privacy that are important for businesses and stakeholders alike, suggesting that protecting customer privacy and complying with privacy regulations are both top priorities. To meet those requirements effectively, companies should collaborate with stakeholders to provide transparency and control to customers, report breaches promptly, and develop a culture of privacy within enterprises. Furthermore, an inclusive approach towards data protection that considers enterprises should also prioritize all stakeholder needs and goals. The table highlights this essential concept - taking an all-inclusive approach toward data privacy is imperative.

Goal	**Importance to Business**	**Importance to Stakeholders**
Protecting customer privacy	High	High
Complying with privacy regulations	High	High

Managing privacy risks	High	High
Maintaining data integrity and confidentiality	High	High
Educating and training employees	High	Medium
Transparency and control for customers	Medium	High
Reporting and responding to data breaches	Medium	High
Developing a culture of privacy within the organization	Medium	High
Collaborating with stakeholders	Medium	Medium
Legal disputes and investigations	Medium	Medium
Privacy standards and best practices	Medium	Medium
Accountability and transparency to shareholders	Low	High
Media scrutiny and accountability	Low	High

Table 4. Goal importance to the business

Note: The matrix assumes that all stakeholders are equally important to the business. However, some stakeholders may be more critical of a particular business than others, and the importance of the issues may vary accordingly.

3.4.3 How a Systems Approach Can Help Enterprises Better Understand the Privacy Implications

A systems approach to privacy program creation and implementation provides businesses with an effective method for developing comprehensive yet adaptable programs for changing situations. A systems thinking approach considers a business an interdependent system of components whose interactions interact and respond to external influences, thus shaping and shaping them internally. This approach can assist enterprises in identifying the source of privacy problems rather than only considering symptoms, as well as understanding any tradeoffs, feedback loops, and unintended outcomes of decisions or actions taken by their company. An adaptive systems approach utilizes principles and tools from systems thinking to analyze, design, implement, and assess privacy programs as wholes and parts. This approach helps enterprises create shared privacy visions among stakeholder goals and incentives and optimize data protection programs over time.

Companies using a systems approach can protect customer privacy in several ways. It helps companies identify critical components of an effective privacy program, such as data governance, risk management, incident response training, and awareness and accountability programs that help build coherent programs across enterprises that address privacy.

Mapping Data Flows: Adopting a systems approach can assist companies with mapping data flow from collection through processing across their enterprise. Companies can identify potential privacy risks like data breaches, unauthorized access, or accidental disclosure by understanding how customer data is collected, processed, stored, and shared. They can identify opportunities to enhance privacy through encryption, anonymization, and Federated Learning techniques.

Analyzing Stakeholder Goals and Motivations: Employing a framework approach, companies can identify the goals and motivations of all relevant parties involved with privacy initiatives, including customers, employees, regulators, partners, and shareholders. By understanding the different perspectives, expectations, and tradeoffs various stakeholders hold, companies can design privacy programs that accommodate all parties' needs while meeting legal and ethical guidelines. Measure Privacy Program Effectiveness: Companies using a systems approach can easily measure the success of privacy programs through metrics and feedback loops. Companies can gain insight into areas for improvement by evaluating how well their privacy programs perform in terms of outcomes like data breaches, complaints, or trust relationships. Feedback loops allow businesses to learn from past events or successes while applying them in future scenarios.

An approach utilizing systems theory can assist companies in creating and implementing an efficient privacy program that supports company values, goals, and legal requirements, as well as those of all its stakeholders. Businesses can gain the confidence necessary to easily protect customer privacy by understanding all the components comprising a comprehensive privacy program and its interdependencies. By giving companies a structured framework for evaluating, developing, and implementing privacy-enhancing technologies and strategies, companies can ensure their programs are efficient, sustainable, and meet the goals and values of stakeholders, including customers, employees, partners, and regulators. Through an interdisciplinary systems approach to safeguard customer privacy, they can prioritize protecting it at every step in data lifecycle management, mitigating potential risks while building customer trust and loyalty.

Adopting a systems approach to privacy can assist companies in better comprehending its effects by examining all stages of data lifecycle analysis, identifying risks to privacy preservation technologies, and engaging stakeholders to ensure transparency and accountability by taking an encompassing view of privacy protection while harnessing data as an engine of growth for innovation.

Chapter 4 - Federated Learning as a Solution to Privacy Concerns

4.1 An Overview of Federated Learning and Its Applications in Preserving User Privacy

The fundamental objective of Federated Learning is to build machine learning models based on data sets spread across many devices while preventing data loss and maintaining user privacy. In other words, Federated Learning is a machine learning technique that enables several devices to train a single model independently. This machine learning methodology is gaining traction because it can address the privacy concerns associated with traditional centralized machine learning methods. Data is acquired from numerous sources for processing and transmitted to a central server in the typical centralized machine-learning strategy. Because this strategy involves sharing sensitive data with other enterprises that could be vulnerable to data breaches or malicious intent, privacy concerns are raised.

Recent improvements have been overcoming statistical obstacles [21], [22] and enhancing security in Federated Learning [23], [24]. Additionally, studies are being done to improve the customization [25] of Federated Learning. This area of research is closely related to privacy-preserving machine learning because it considers data privacy in a decentralized collaborative learning environment. It would help expand the Federated Learning concept to a general concept for all privacy-preserving decentralized machine learning techniques to cover collaborative learning scenarios among enterprises. The

Federated Learning and federated transfer learning techniques were given a preliminary overview in Q, Yang, et al. [25].

Definition of Federated Learning

Define N data owners $\{F_1, ...F_N\}$, all of whom wish to train a machine learning model by consolidating their respective data $\{D_1, ...D_N\}$. A conventional method is to put all data together and use $D = D_1 \cup ... \cup D_N$ to train a model M_{SUM} . A Federated Learning[26] system is a learning process in which the data owners collaboratively train a model M_{FED}, in which process any data owner F_i does not expose its data D_i to others. In addition, the accuracy of M_{FED} , denoted as V_{FED} should be very close to the performance of M_{SUM}, V_{SUM} . Formally, let δ be a non-negative real number, if

$$| V_{FED} - V_{SUM} |< \delta$$

Say the Federated Learning algorithm has δ-accuracy loss.

To work correctly, Federated Learning must be jointly trained across various devices, including laptops, smartphones, and edge devices, without data being sent to a centralized server. Below is a brief overview of the Federated Learning process.

Device Selection: In Federated Learning, the initial stage is choosing the training process's participating devices. The devices could be computers, smartphones, or the Internet of Things (IoT). The devices must meet specific requirements, such as having minimum storage and processing power.

Data Distribution: The data distribution process begins after the devices have been decided. A portion of the data, which may be randomly picked from the entire dataset, is stored on each device. Typically, the data is divided into non-overlapping subsets, resulting in a distinct set of samples on each device.

Local Model Updates: Using their data, the devices update the model locally once the data has been distributed. The model parameters are updated based on the local data using an optimization procedure, such as stochastic gradient descent (SGD). The local updates are carried out decentralized without interacting with other devices or the central server.

For example, a local model update using SGD is given by:

$$w\_i(t+1) = w\_i(t) - lr * \nabla Q(w\_i(t), D\_i)$$

where w_i(t) is the model parameters at time t on device i, lr is the learning rate, ∇Q(w_i(t), D_i) is the gradient of the loss function Q with respect to the parameters w_i(t) computed on the data D_i held on device i, and w_i(t+1) is the updated model parameters.

Secure Aggregation: Once the local updates have been computed, the devices send them to a central server for aggregation. The server aggregates the updates using a secure aggregation [24] technique, such as secure multi-party computation (MPC) or homomorphic encryption, which ensures that the individual updates are not leaked to other parties. The aggregated update is then sent back to the devices. The aggregated update is given by:

$$\Delta(t) = \Sigma_i \delta\_i(t)$$

where Δ(t) is the aggregated update at time t, δ_i(t) is the local update computed by device i at time t, and the sum is taken over all devices.

Global Model Update: The devices use the aggregated update to update the global model [27], which is the model that is shared among all devices. The update is applied to the global model using an aggregation rule, such as averaging or weighted averaging, determining how the updates are combined. The global model is then sent back to the devices for further updates. The update is given by:

$$w(t+1) = w(t) - lr * \Delta(t)$$

where w(t) is the global model parameters at time t, lr is the learning rate, Δ(t) is the aggregated update computed by the central server at time t, and w(t+1) is the updated global model parameters.

Convergence: The process of local updates, aggregation, and global updates is repeated for a fixed number of iterations or until the model converges. The convergence criterion can be based on a validation set held out from the training data or a consensus among the devices. Once the model has converged, it can be used to predict new data. It's important to note that the specific equations used may vary depending on the optimization algorithm, model architecture, and communication framework used.

Federated Learning has many applications to protect user privacy. Some of the most common applications include 1) healthcare, 2) financial services, 3) urban development, 4) personalization, and 5) the Internet of Things (IoT).

Healthcare: Federated Learning can be used to create machine learning models that can assist in diagnosing a range of health issues. With Federated Learning, medical information can stay on the patient's device, lowering the possibility of data breaches and preserving patient privacy.

Financial Services: Federated Learning can be used in the financial sector to create machine learning models to detect fraud. The risk of data breaches can be decreased, and customer privacy can be preserved by allowing financial data to remain on customer devices with combined training.

Urban development: Federated Learning is a technique that can be used in smart cities to create machine learning models that can be used to predict traffic, track pollution, and respond to emergencies.

Personalization: Federated Learning can be used to create machine learning models that give users personalized recommendations. The risk of data leakage can be decreased, and user privacy can be preserved by storing user data on the user's device through shared training.

IoT devices: Federated Learning can be used to create machine learning models that aid in real-time data processing on edge devices like IoT devices.

Multiple use cases for Federated Learning can be applied, as it can use multiple devices to train a common model with Federated Learning without sharing their data with a central server. The privacy issues raised by centralized, conventional machine-learning approaches can be addressed by this method.

4.2 Privacy Improvement Through Federated Learning: Techniques and Advantages

One of the most critical advantages of Federated Learning is its ability to improve privacy. As previously mentioned, traditional centralized machine learning methods must collect data and send it to a central server for processing, potentially raising privacy concerns.

This section briefly reviews different Federated Learning techniques and identifies approaches and potential challenges for preventing indirect leakage. There are several kinds of Federated Learning techniques, including 1) Horizontal Federated Learning, 2) Vertical Federated Learning, and 3) Federated transfer learning.

Horizontal Federated Learning: In this method, the data is partitioned horizontally (*See Figure 10.*) so that each device holds a subset of the data points, and all nodes share all features. Each device or entity holds a portion of the data. Each device trains its local model on its data, transmitting only the updates to a central server, aggregating the updates to produce a new iteration of the global model. This allows the devices to work together to train a global model while maintaining the privacy of their own data. Horizontal Federated Learning is helpful when data is sensitive and cannot be shared with other entities due to privacy concerns. The design of a horizontal Federated Learning system is shown in Figure 12.

For example, imagine a hospital group interested in creating a machine-learning model to forecast patient outcomes for a particular illness. They are unable to exchange patient data among themselves due to privacy laws. In this situation, they might train a common model using horizontal Federated Learning without sharing the patient data.

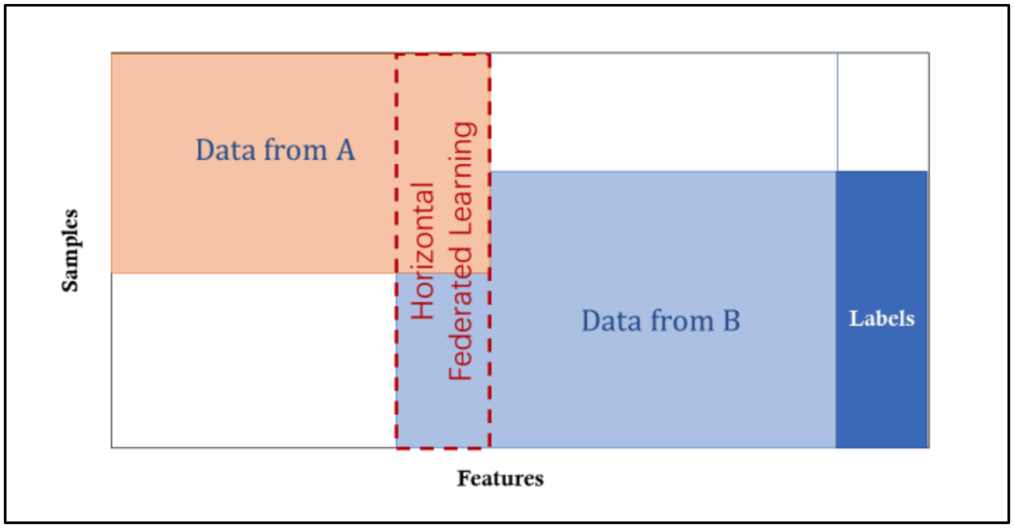


Figure 10. Horizontal Federated Learning [54]

Each hospital would save a portion of patient data and use a local optimization method to train a local model on that data. The updates from their local model would then be sent to a central server, combining them to produce a fresh iteration of the global model (*See Figure 11*). The process is repeated until the global model converges to an acceptable level of accuracy.

By working together in this way, the hospitals may create a model without compromising the confidentiality of the patient data. The generated model can forecast patient outcomes, enhancing clinical judgment, and eventually preventing unnecessary deaths.

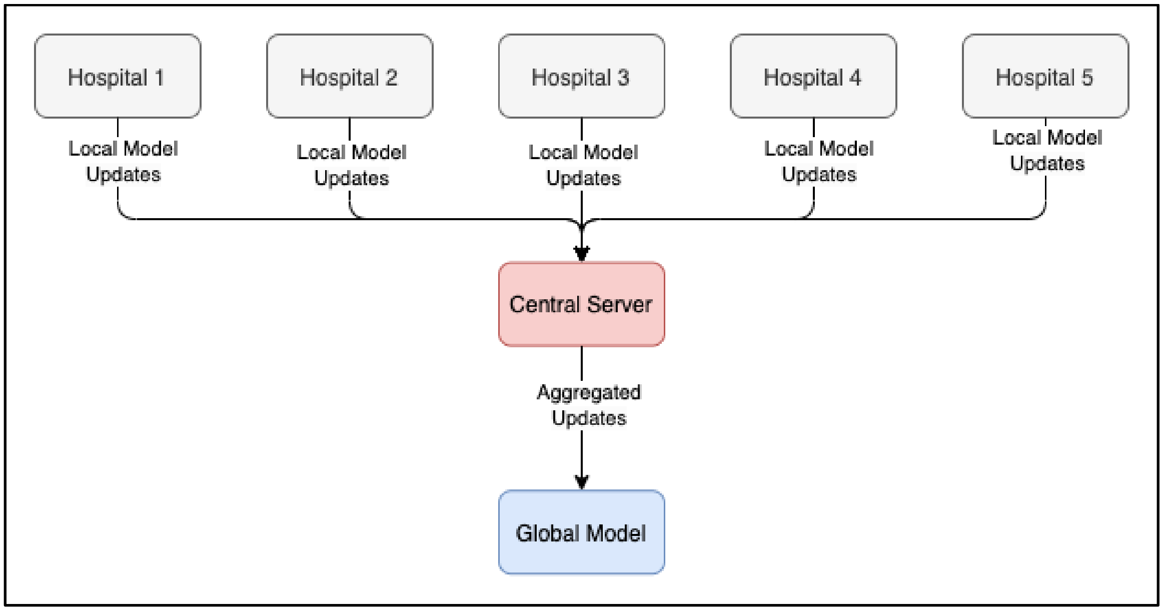


Figure 11. Example: Horizontal Federated Learning [by: Mervin Govada]

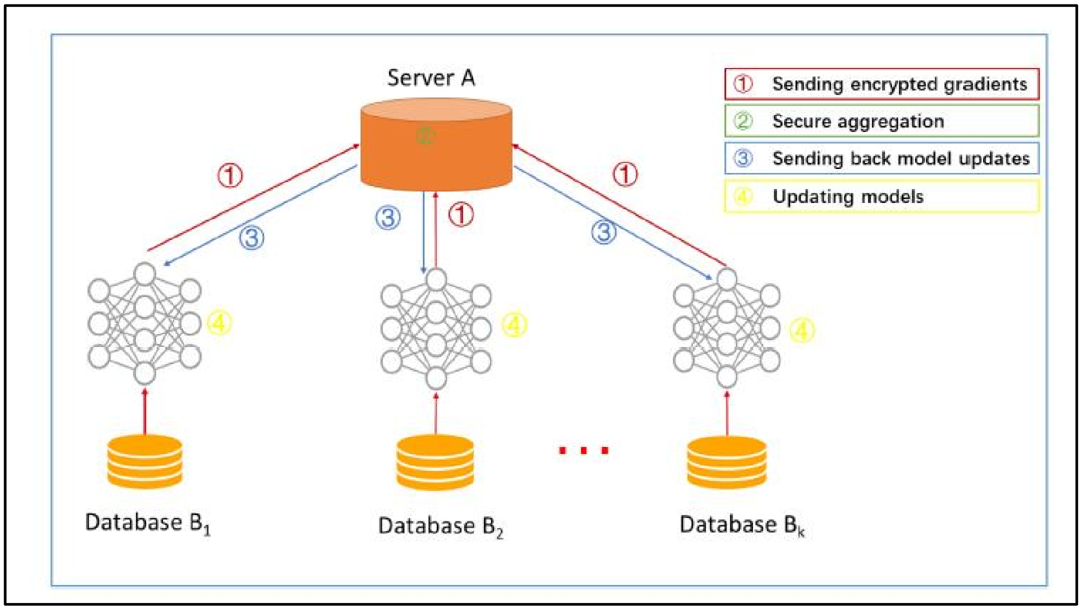
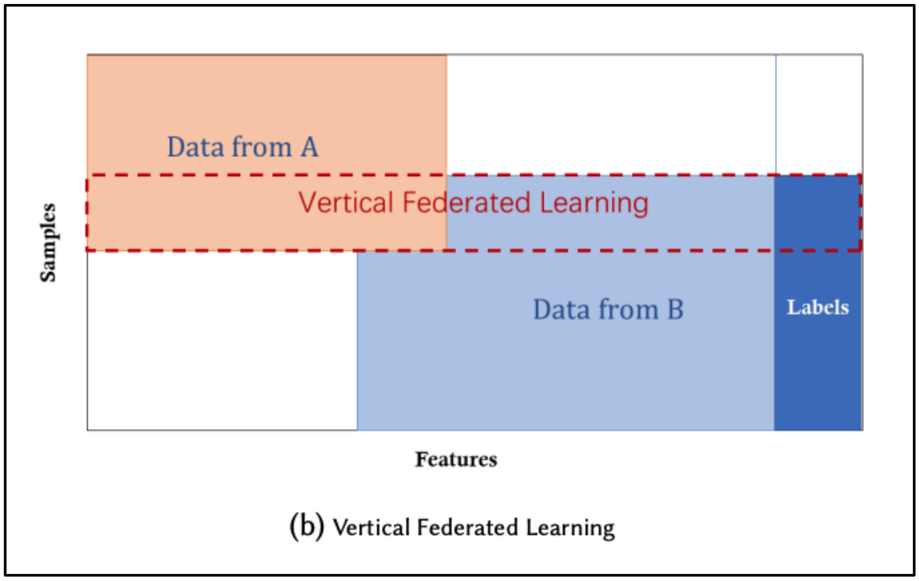


Figure 12. Architecture for a horizontal Federated Learning system [54]

Vertical Federated Learning: In this method, it enables many parties to jointly train a model using their own confidential data while maintaining the privacy and security of each party's data. Each participant in vertical Federated Learning has a subset of features (input variables) for a given set of data points, which are often kept on various servers or databases (*See Figure 13.)*. The parties can then work together to train a model that uses the data from each other's datasets without disclosing any sensitive information or sharing the raw data. When training a model that necessitates combining medical records from various hospitals or financial data from various institutions, this technique is beneficial when different parties may have access to different data types. Because the parties must coordinate and align their data to successfully train the model, vertical Federated Learning is typically more complicated than horizontal Federated Learning. The design of the vertical Federated Learning System is shown in Figure 15.

For example, in a situation where two banks, Bank A and Bank B, seek to work together and develop a machine learning model to detect fraudulent transactions could illustrate vertical Federated Learning in finance. Credit card transaction information that Bank A has access to includes transaction amounts, merchant categories, and client data like age and income. On the other side, Bank B gets access to details on bank transfers, such as account numbers, transaction times, and geographical details.



(b) Vertical Federated Learning

Figure 13. Vertical Federated Learning [54]

Both banks can use vertical Federated Learning to construct an extensive fraud detection model. Bank B can store local bank transfer data. At the same time, Bank A can keep local copies of its credit card transaction data. Without sharing their private information, they can then collaborate on the training process and share the model. The two banks can integrate their data sources in this fashion without disclosing them to one another, enabling the development of a more thorough and accurate fraud detection model (*See Figure 14).*

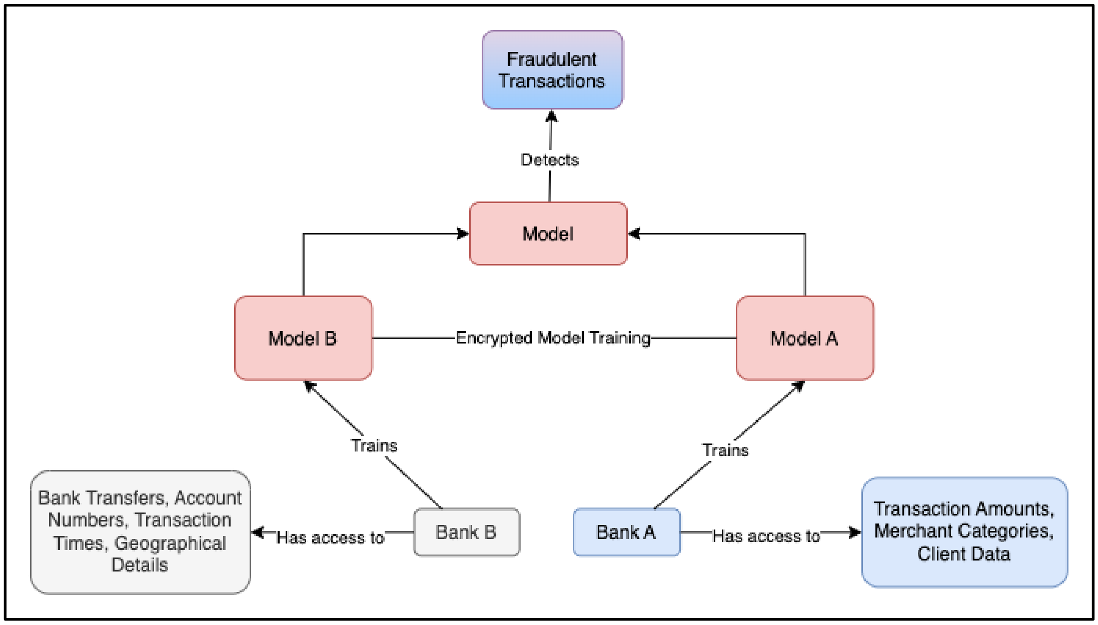

Fraudulent Transactions
Detects
Model
Model B
Encrypted Model Training
Model A
Trains
Trains
Bank Transfers, Account Numbers, Transaction Times, Geographical Details
Has access to
Bank B
Bank A
Has access to
Transaction Amounts, Merchant Categories, Client Data

Figure 14. Example: Vertical Federated Learning [by: Mervin Govada]

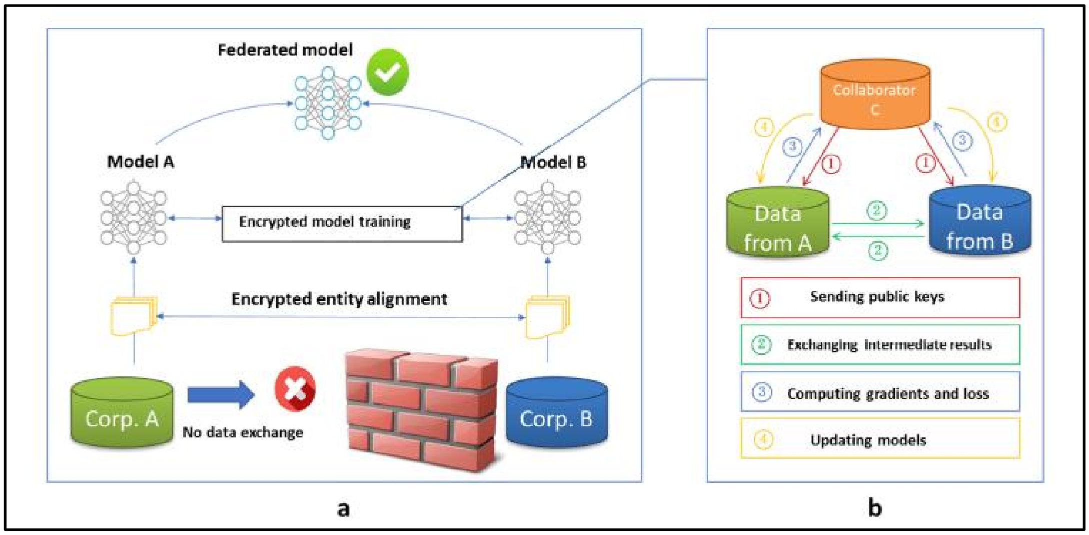


Figure 15. Architecture for a vertical Federated Learning system [54]

Federated transfer learning: This method combines Federated Learning and transfer learning. A pre-trained model is distributed to each participating device or node in this method, and each node then refines the model using its local data (*See Figure 16.)*. Federated transfer learning boosts model performance while preserving data security and privacy. Because it reduces the amount of data that must be transmitted between the nodes, this technique is beneficial when the participating devices have low computing power or bandwidth.

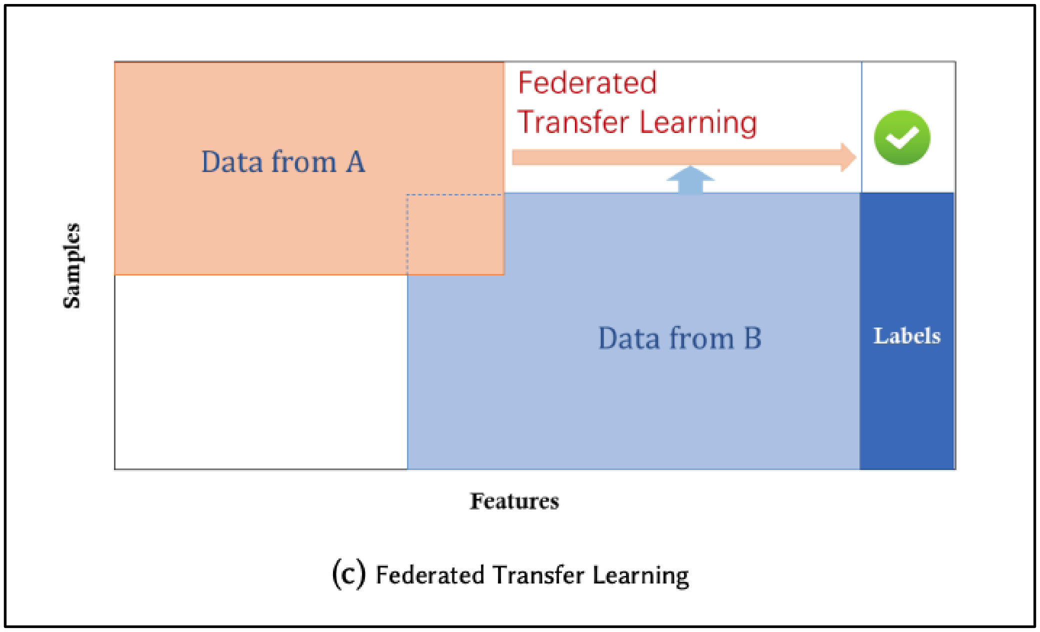


Figure 16. Federated Transfer Learning [54]

For example, consider a scenario where a hospital network wants to train a machine learning model to estimate a patient's risk of contracting a specific disease based on their medical history. The hospital, however, cannot provide the patient data to a central server for training because of privacy issues.
The hospital network can use federated transfer learning to solve this problem. The first step is to create a pre-trained model using a sizable dataset of medical records from different hospitals. The pre-trained model is then downloaded and further trained using the locally stored patient data from each hospital in the network. To produce a new, enhanced model, the updated models are then transmitted back to a central server for aggregation. This procedure is repeated regularly without disclosing private patient information to increase the maintainability of the model's accuracy. The example of a Federated Transfer Learning system is shown in Figure 17.

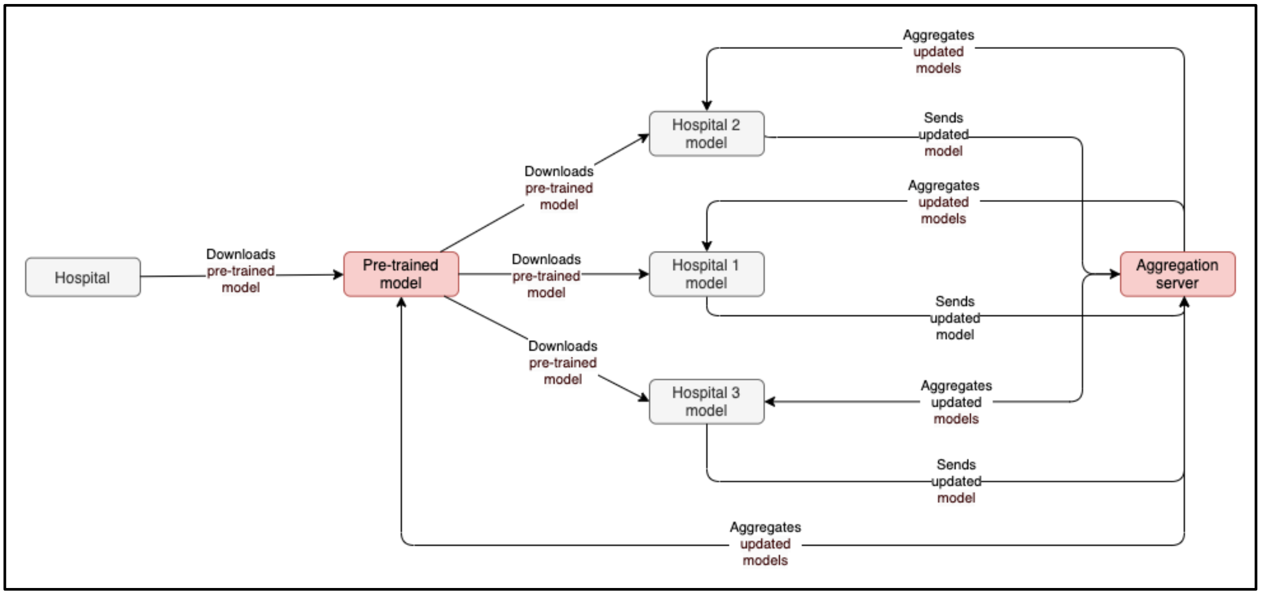

Aggregates updated models
Hospital 2 model
Sends updated model
Downloads pre-trained model
Aggregates updated models
Hospital
Downloads pre-trained model
Pre-trained model
Downloads pre-trained model
Hospital 1 model
Aggregation server
Sends updated model
Downloads pre-trained model
Hospital 3 model
Aggregates updated models
Sends updated model
Aggregates updated models

Figure 17. Example: Federated Transfer Learning [by: Mervin Govada]

In summary, horizontal Federated Learning is suitable when several parties have similar data and want to train a model while preserving privacy. Vertical Federated Learning is appropriate when parties have complementary data types, such as medical records and genetic information. Federated transfer learning is used when there is a lot of labeled data in one place that cannot be shared due to privacy issues, and it is used to improve a pre-trained model before distributing the knowledge to local models.

The decision of whatever strategy to employ will ultimately depend on the system's particular needs, including the type of data involved, privacy issues, and the available computing power of the resources.

Advantages Of Using Federated Learning

Federated Learning allows the creation of machine learning models that can be trained on private information without compromising security. For instance, Federated Learning can be applied to the healthcare sector to create machine learning models that can assist in diagnosing various health issues while maintaining patient privacy.

Traditional centralized machine learning techniques call for data to be gathered and stored on central servers, leaving it open to data breaches or malicious attacks. Data breaches are less likely to occur, and user privacy is preserved with Federated Learning because the device that created the data keeps it. Federated Learning can also create machine learning models that can withstand assaults. The fact that the data is still on the device makes it difficult for an attacker to successfully attack because they must compromise multiple devices to access it.

The benefit of scalability is another benefit of Federated Learning. Traditional centralized machine-learning techniques can restrict the data collected and processed by a single server. However, Federated Learning enables the processing of larger datasets by utilizing multiple units to train a single shared model. Furthermore, it helps the creation of scalable machine-learning models. Deploying models to numerous devices is simpler because no central server is required to store or process the data. After all, it remains on the device.

Cost efficiency is another benefit of Federated Learning. Significant resources are needed for centralized machine learning methods to collect and store large amounts of data. However, Federated Learning eliminates the need for pricey data storage solutions because data remains on the device. Additionally, it can lower the price of training machine learning models. Traditional centralized machine-learning techniques need much computing power to process large data. On the other hand, Federated Learning distributes the computational resources required to train the model across several units, lowering the price of model training.

Federated Learning can also increase accuracy. Traditional centralized machine learning techniques call for the collection of data and its transmission to a central server for processing, which can cause errors and delays. However, Federated Learning keeps the data on the device, enabling real-time processing and lowering the possibility of error. Additionally, Federated Learning can result in machine learning models with higher accuracy. Data can be trained on larger, more varied data sets thanks to the fact that it stays on the device, leading to models with higher accuracy.

Federated Learning offers a variety of benefits, such as improved data security, scalability, affordability, and accuracy. These benefits make Federated Learning a promising machine learning strategy, particularly in sectors where privacy and security are crucial, like healthcare, finance, and marketing.

4.3 Discussion of The Challenges Associated with Federated Learning

Federated Learning is an effective machine learning technique, but several issues must be resolved before it can be extensively used. The possibility of indirect leaking, when an attacker might infer personal information from a model's output even when the data is protected, is one of the critical problems. Homomorphic encryption, Differential Privacy, secure multi-party computation, and secure aggregation are just a few of the cutting-edge security and privacy techniques employed to address this. Even these approaches have drawbacks and trade-offs.

The variety of data sources used in Federated Learning presents another difficulty. These data sources' formats, distributions, and quality levels could differ, posing compatibility issues and reducing the accuracy and generalizability of the models developed using Federated Learning. Additionally, updating the model with new data might be challenging because the data may vary between devices. Instead, the model might need to be retrained using data from all devices.

Another difficulty for Federated Learning is selecting appropriate models and optimization strategies. Employing inappropriate models or techniques can result in poor model accuracy, slow convergence, excessive communication overhead, and inefficient computing. Edge computing and energy-efficient communication protocols have been proposed to address this issue.

Federated Learning requires confidence between participating companies in order for them to share data and collaborate in a way that safeguards data privacy and security. Contracts, prizes, and reputation-based systems can all be used to build and maintain trust between parties. The difficulties of Federated Learning have been addressed by adaptive Federated Learning and continuous learning, which allow updating models on a per-device basis using just local input. To ease communication problems, techniques like compression and data quantization can be used.

Differential privacy is a promising approach to control the level of privacy provided in this setting. Utilizing privacy parameters that can be adjusted to balance the privacy-utility trade-off allows for control. While homomorphic encryption is an additional method for securing privacy, it may be computationally expensive and impact scalability because calculations on encrypted data may take longer than those on unencrypted data. Differential privacy allows for more control over the level of privacy and is simpler to implement.

Federated Learning has the potential to make machine learning more secure, but several issues need to be resolved before it can be put into practice. These difficulties can be overcome by effective implementation using cutting-edge security and privacy techniques, suitable models and algorithms, trust between parties, and the application of adaptive and continuous learning.

Chapter 5 - Differential Privacy as A Solution to Privacy Concerns

5.1 Differential Privacy and Its Applications in Preserving User Privacy

Differential privacy is a framework for protecting the privacy of individuals in datasets containing sensitive information. It was first introduced in 2006 as a mathematical concept [28] for measuring privacy in statistical data analysis. Differential privacy protection has become a vital data protection tool in applications as diverse as health, finance, and social sciences.

The basic idea behind Differential Privacy is to limit how much information can be learned about a person from data. This is achieved by adding random noise to the data before analyzing it. The amount of added noise depends on a parameter called the privacy budget, which determines the level of privacy protection. The definition of Differential Privacy involves two functions: a data function f that maps data to results and a neighbor function n that measures the distance between two data points. Proximity functions define a distance metric that measures two data points' differences.

Differential privacy

A stochastic algorithm A is ε (epsilon) if the probability that the algorithm outputs a result S in a data set D is at most exp(ε) for two adjacent data sets D and D' and for any subset S of the output range. Privacy is multiplied by the probability that the data set D' will produce S results, where $\varepsilon > 0$ is a parameter controlling the privacy protection level. In other words, consider two data sets, D and D', that differ by only one data point, the probability that the algorithm will produce a particular result on D' should be very similar to the probability that it will produce the same result on D'. The parameter ε determines how much the probability can vary, and larger values of ε provide more variance and less privacy protection.

To achieve Differential Privacy, the random algorithm A usually adds random noise to the data before processing it. The amount of added noise depends on the privacy budget ε and the sensitivity of the data function f. The sensitivity of f is a measure of the change in the output of f when individual data points are removed or added to the data set. The amount of noise added to the data is proportional to the sensitivity f and inversely proportional to ε. This ensures that as ε decreases, the amount of noise added to the data increases, providing more robust privacy protection.

The basic equation for Differential Privacy is:

$$P(Q(D1) \in S) \leq \exp(\varepsilon) * P(Q(D2) \in S)$$

where,

$P(Q(D1) \in S)$ is the probability that the output of the query Q on dataset D1 belongs to a set S.

$P(Q(D2) \in S)$ is the probability that the output of the query Q on a neighboring dataset D2 (that differs from D1 in only one record) belongs to the same set S.

ε is the privacy parameter (Epsilon) that determines the level of privacy protection.

The concept of Differential Privacy can be defined as the degree to which the results of queries on a dataset change when elements of the dataset are removed. Data sets are different if the probability distribution of query results does not change significantly when a single record is deleted. In other words, changing or removing personal information will not significantly impact the analysis's overall findings. To protect the identities of the individuals in the data collection, Differential Privacy is done by adding random noise to the data. This noise is added to the data before analysis to protect all sensitive information. Input noise levels are carefully adjusted to ensure that analysis results remain accurate and valuable while protecting individual privacy.

Applications For Differential Privacy

Multiple industries, including healthcare, finance, personalization, and social sciences, have numerous uses for Differential Privacy. The applications that are discussed below are some examples.

Healthcare: Distinct privacy features are used to protect patient data while enabling researchers to analyze the data to gain insights. For instance, Differential Privacy can be used to analyze electronic health records to find patterns in the incidence of diseases, the efficacy of treatments, and patient outcomes without compromising patient privacy.

Differential privacy can be used in the financial sector to analyze financial data while maintaining confidentiality. Without disclosing the identities of the people involved in the transaction, it can be used, for instance, to analyze financial transactions and find fraudulent activity.

Social Sciences: The social sciences also employ Differential Privacy to safeguard subjects' privacy in extensive research or population studies. To analyze census data and find demographic trends and patterns without identifying specific people, for example, Differential Privacy can be used.

Personalization:
Companies can gather and analyze user data while minimizing the risk of disclosing sensitive information by employing Differential Privacy techniques. For example, Apple [29] protects user privacy while retaining the ability to offer customized services using Differential Privacy. This protects the privacy of Apple's users while enabling it to enhance its products and services, like Siri and Maps. Differential privacy is utilized, for example, when Siri suggests a word or phrase as a user type, to ensure that the suggestion is based on the user's input and not on sensitive information that might be extrapolated from the input.

5.2 Analysis of The Advantages of Differential Privacy

Differential privacy is a potent tool that safeguards sensitive data and offers helpful analytics. Improved privacy, security, and data accuracy are just a few of its many advantages over conventional data anonymization techniques.

Improves Privacy
By combining or removing personal identifiers, for example, layered privacy offers a higher level of privacy than conventional data anonymization techniques [30]. Even so, when combined with other publicly accessible data sources, these techniques may still expose individuals to re-identification. On the other hand, Differential Privacy [29] modifies the data by adding carefully calibrated noise that prevents the data from being used to identify specific people. Deep learning [31] with Differential Privacy can be used to achieve this. Because data about each individual may be generated from various combinations of inputs, it is impossible to identify the original input in datasets with Differential Privacy. Because of this, even if an attacker has access to data collection, they cannot tell which data belong to one person and which belongs to another.

Enhances Data Security

By providing data with an additional layer of security [32], Differential Privacy enhances data security. When noise is added, it becomes more difficult for attackers to extract useful information from data. This is particularly crucial in healthcare and finance, where data breaches can have severe repercussions.

Data breaches, for instance, can result in the disclosure of private medical information, which could negatively affect the patients. Data breaches in the financial industry can result in the disclosure of financial information, which can cause fraud and identity theft. Differential privacy adds an extra layer of security, making it more difficult for attackers to extract sensitive information from the data.

Impact on Accuracy

Differential privacy can still produce reliable analysis results despite the noise that has been added to the data [33]. The amount of additional noise is carefully adjusted to match the accuracy needed for privacy analysis. More noise can be added to the data and vice versa if exact results are required.

Enterprises can thus use critical data for insightful analysis. For instance, Differential Privacy can be used in the healthcare industry to analyze electronic health records to find patterns in disease, the effects of treatments, and patient outcomes without compromising patient privacy. Differential privacy is a technique [34] that can be used in finance to examine financial transactions, personalize ads, and spot fraudulent activity without disclosing the parties' identities.

Enables Flexibility

Differential privacy is a flexible framework that can be used for various data and analysis types. Due to its adaptability, it is perfect for various applications, including those in medicine, finance, and social science. For instance, Differential Privacy can be applied to census data analysis to find population trends and patterns without identifying specific people. To train a model on sensitive data without making the model reliant on the data, it can also be used in machine learning applications like image recognition or natural language processing.

Ensures Transparency

Operational transparency is also guaranteed by Differential Privacy. The parameter can be used to calculate the privacy level as a function of Differential Privacy. The maximum amount of privacy that can be lost during analysis is represented by this setting. Larger values increase confidentiality while adding more noise to the data. Thanks to this transparency, enterprises can choose the right level of privacy based on their data analytics requirements. Enterprises can also compare various privacy protection strategies to choose the best method for a specific application.

Differential privacy is an effective tool for safeguarding private information while delivering insightful analytics. Compared to conventional data anonymization techniques, it offers several significant benefits, including increased privacy, data security, accuracy, flexibility, and transparency. Differential privacy is the best choice due to these benefits.

5.3 Challenges Associated with Differential Privacy

Although Differential Privacy has many benefits for privacy, data security, accuracy, and flexibility, it also has some issues that must be resolved. The complexity of the model and the presence of noise are the two key issues.

Added Noise Reduces Accuracy

Differential privacy is a method for preserving privacy when processing data [35]. It must compromise between accuracy and privacy protection. Differential privacy adds noise to the data to safeguard privacy, but this noise also lowers the accuracy of the analysis's findings [36]. The parameter's value determines how much noise is introduced to the data. Less noise in the data is produced by larger values, which increases analysis accuracy but decreases privacy. In contrast, more minor levels result in more data noise and higher privacy but less precise analysis.

Therefore, finding the ideal value that balances accuracy and privacy is a challenge. This might be intimidating, especially when analytical precision is crucial, as in financial analysis or medical research. In some situations, customers have to settle for less privacy to get the necessary level of precision. Adding noise can make the analysis results unstable is another challenge. The study's findings can vary greatly depending on how much noise is added to the data. This makes comparing analysis results over time or between several datasets difficult.

These difficulties can be reduced using a variety of tactics. One strategy to improve analysis while protecting privacy is to use cutting-edge methods like machine learning. Even for vast levels, models that can handle noisy data better and produce more accurate analysis results can be created using machine learning. Utilizing ensemble methods is another strategy for boosting the accuracy of analysis results. Multiple analysis results with various noise levels are produced by ensemble methods, which then combine these results to produce a result. This strategy improves stability and dependability while reducing the impact of individual noise values on the analysis results. In addition to these technical solutions, organizational and legal measures can be taken to address various privacy issues. Data governance policies, for instance, can be implemented by businesses to control data collection, storage, and analysis. This policy ensures that private information is considered while handling sensitive data responsibly.

Increased Complexity of The Model

Another challenge with Differential Privacy is the complexity of the models used to generate the data. To ensure differential data privacy, models must be carefully designed with privacy protection requirements [37] in mind. This can be daunting, especially when the data is large or contains complex relationships between variables. Model complexity can also affect analysis performance. Complex models may require more computing resources and therefore take more time to generate analysis results.

Along with the model's performance and design issues, another concern is ensuring the model does not violate anyone's privacy. The model can occasionally be broken, allowing an attacker to change the raw data. When numerous enterprises or researchers share samples, this can be very challenging.

These challenges can be overcome using advanced techniques such as machine learning, ensemble methods, and the implementation of data governance policies. Research advances in Differential Privacy and future developments are expected to mitigate these challenges further.

5.4 Implement Differential Privacy with Federated Learning

In the context of Federated Learning, Differential Privacy can be implemented using either centralized or local methods.

Centralized Differential Privacy in Federated Learning involves adding noise to the gradients calculated by the client devices before sending them to the central server. This is done by aggregating the gradients from multiple client devices, adding noise to the aggregated gradients, and then updating the central model based on the noisy gradients. This method offers a stronger guarantee of privacy than local Differential Privacy, as it protects against attacks on the central server.

Local Differential Privacy in Federated Learning involves adding noise to the data or model parameters on each client device before sharing them with the central server. This method offers greater privacy protection for individual client devices, as it does not rely on trusting the central server with sensitive information. However, the noise added to each client's data or model parameters can make it more difficult to train an accurate model.

Both centralized and local Differential Privacy have their advantages and limitations, and the choice between the two depends on the specific use case and the privacy requirements. There are several mechanisms to implement centralized Differential Privacy in Federated Learning, including:

Adding noise to the gradients: This mechanism involves adding noise to the gradients of the models before they are sent to the central server. The noise is added to mask the contribution of each client's data to the overall model.

Adding noise to the data: In this approach, data from each client is processed with noise before being delivered to the central server. While maintaining the data's overall statistics, noise is added to make it more difficult to distinguish between individual data points.

Secure aggregation: A central server can aggregate model updates from all clients without having access to the raw data. This method ensures that the central server is protected from potential data breaches by ensuring that it never sees the raw data.

These mechanisms are designed to keep client data private while enabling the central server to build a solid model that adapts well to fresh input. There are several mechanisms to implement local

Differential Privacy in Federated Learning, including, Randomized response, Laplace mechanism, Exponential mechanism, Gaussian mechanism, and Objective perturbation.

Local Differential Privacy may be more appropriate for larger datasets or settings without a trusted central server. In contrast, centralized Differential Privacy may be the best option for small datasets with a trusted central server. A careful trade-off between privacy and utility is necessary to implement Differential Privacy in Federated Learning.

Differential privacy can be incorporated into Federated Learning through various mathematical constructs. The remainder of this section will provide a brief overview of the mechanisms that can be used.

Gaussian Mechanism:

The Gaussian Mechanism adds random Gaussian noise to the local updates sent by each device to the central server. The amount of noise added is determined by the desired privacy level and the sensitivity of the data being transmitted. Mathematically, the Gaussian Mechanism can be represented as follows:

Let f be the function to be privatized, and let x be the input data. To compute the output of f in a differentially private way. Let ε be the privacy parameter and let Δf be the sensitivity of f.

The Gaussian mechanism adds noise sampled from a Gaussian distribution to the output of f. Specifically, it outputs:

$$f(x) + N(0, \sigma^2)$$

where $N(0, \sigma^2)$ denotes a random variable drawn from a Gaussian distribution with mean 0 and variance σ^2.

The amount of noise added depends on the desired privacy level ε and the sensitivity Δf. The standard deviation of the Gaussian distribution is given by:

$$\sigma = \Delta f * \mathrm{sqrt}(2 * \ln(1.25/\delta)) / \varepsilon$$

where δ is the probability of the privacy guarantee failing (usually set to a small value, such as 10^{-6}).

Therefore, the differentially private output of f is:

$$f(x) + N(0, \sigma^2)$$

where σ is computed as above.

Laplace Mechanism:

The Laplace Mechanism [38] adds random Laplace noise to the local updates. The amount of noise added is determined by the desired privacy level and the sensitivity of the data being transmitted. Mathematically, the Laplace Mechanism can be represented as follows:

$$f(x) + Lap(0, b/\varepsilon)$$

where, $f(x)$ is the result of the data analysis algorithm on input data x

$Lap(0, b/\varepsilon)$ is a random variable drawn from the Laplace distribution with mean 0 and scale parameter b/ε

b is the sensitivity of the data analysis algorithm, which measures the maximum amount that the output of the algorithm can change if a single individual's data is added or removed

ε is the privacy parameter, which determines the level of privacy protection desired, with larger values of ε providing less privacy protection.

Exponential Mechanism:

The Exponential Mechanism [39] is a technique in Differential Privacy used to choose an output from a set of possible outputs in a way that preserves the privacy of individuals in the input dataset. The Exponential Mechanism works by assigning a score to each output in the set and then choosing an output with a probability proportional to the score.

More formally, let S be the set of possible outputs, and let f be a scoring function that maps each output in S to a real number. The Exponential Mechanism selects an output from S by choosing an output x with probability proportional to:

$$\exp(\varepsilon f(x) / (2\Delta f))$$

where, ε is the privacy parameter, Δf is the sensitivity of the scoring function f, and the denominator serves as a normalization factor to ensure that the probabilities sum to 1.

The sensitivity of a scoring function f is defined as the maximum amount that f can change when a single individual's data is added or removed from the input dataset. Formally, if d and d' are two input datasets that differ by at most one individual, then the sensitivity of f is given by:

$$\Delta f = \max \{ ||f(d) - f(d')|| \}$$

The Exponential Mechanism can be used to select outputs in a wide range of settings, including data analysis, machine learning, and social choice theory. By choosing outputs in a way that is sensitive to individual privacy, the Exponential Mechanism helps to ensure that data analysis and decision-making processes are robust and trustworthy while also protecting the privacy of individuals.

Randomized Response:

Sensitive data can be safeguarded in Federated Learning using a randomized response technique. It involves adding random noise to the responses to hide the real answer to a binary question. This method protects sensitive data, like medical records, while enabling helpful statistical analysis.

If a question is binary, like "Did you take part in the study?" a participant might not want to reveal their real response if it is delicate. By introducing random noise to the response, a randomized response can be used to provide privacy. The participant provides a probability of p for an honest response and a probability of 1-p for a random response. The Randomized Response mechanism can be modeled mathematically using the following notation:

$$f(x) = (x \text{ with probability } p) \text{ or } (\text{random response with probability } 1-p)$$

Objective perturbation:

It is a technique used in local Differential Privacy for adding noise to the model's gradient updates during training. The objective perturbation approach involves adding noise to the objective function before computing the gradient. The resulting noisy objective function is optimized using standard gradient-based optimization techniques.

Mathematically, objective perturbation can be represented as:

$$L'(\theta; D) = L(\theta; D) + \eta(\varepsilon)$$

where, $L(\theta; D)$ is the loss function of the model with parameters θ on the dataset D, and where $\eta(\varepsilon)$ is the noise function, which depends on the privacy parameter ε, and ε is the maximum amount of information leakage allowed. Typically, the noise function is modeled as Laplace noise, i.e.,

$$\eta(\varepsilon) = \text{Laplace}(0, \Delta L/\varepsilon)$$

where ΔL is the global sensitivity of the loss function, defined as the maximum amount the loss function can change when a single data point is added or removed from the dataset.

During training, the noisy objective function is optimized using stochastic gradient descent (SGD) or any other gradient-based optimization algorithm to update the model parameters. The gradient of the noisy objective function L'(θ; D) can be computed as:

$$\nabla L'(\theta; D) = \nabla L(\theta; D) + \nabla \eta(\varepsilon)$$

where, ∇L(θ; D) is the gradient of the original loss function L(θ; D) and ∇η(ε) is the gradient of the noise function η(ε). The gradient of the noise function is typically computed using the chain rule of differentiation and is proportional to the sensitivity of the loss function.

The objective perturbation approach involves adding noise to the objective function before computing the gradient during training. This approach is used in local Differential Privacy to ensure that the model updates are differentially private.

The best Differential Privacy mechanism for Federated Learning depends on factors such as data type, aggregation, and privacy level. That being said, some mechanisms may better suit certain types of data or use cases. For example, the Laplace mechanism is commonly used in local Differential Privacy because it provides a good balance between privacy and utility, and it is particularly well-suited to discrete data. The Gaussian mechanism may be better suited to continuous data. In contrast, the exponential mechanism may be better suited to scenarios where the output space is continuous, and privacy loss is a concern.

Because it introduces noise to the gradients of the objective function rather than directly to the data, the objective perturbation can be a useful approach for Federated Learning's local Differential Privacy implementation. This can decrease the amount of noise required to attain the desired level of privacy and increase the learning algorithm's accuracy. However, objective perturbation might require more computation than other mechanisms, and it might be more challenging to use in some circumstances.
In Federated Learning, Differential Privacy can be implemented by adding noise to the updates sent from the client to the server during training. The client trains the model locally [27] on the device and then sends updates to a central server that aggregates the model to improve it. However, the specific mechanism of the implementation depends on the use case.

Chapter 6 - The Customer-centric Approach

6.1 The Importance of Taking a Customer-centric Approach

Since the advent of AI technologies and Internet of Things (IoT) devices, user privacy has become the top of customers' concerns. Due to the expanding amount of data that businesses gather and the potential exploitation of that data, customers are becoming more conscious of the need to protect their personal information. To safeguard user privacy, a user-centric approach is required to prioritize users' wants and preferences during data collection and analysis. This part discusses the importance of maintaining user privacy from a customer-focused perspective.

Bélanger and Crossler [40] assert that privacy policies must be customer-focused in the digital age. According to Culnan and Bies [15], businesses must balance economic and just considerations when creating data collection and analysis processes. According to Xu et al. [41], who study the causes of privacy concerns, a customer-centric approach can help calm them.

Understand User Preferences and Expectations

One of the key benefits of a customer-centric approach to privacy is that companies can better comprehend user preferences and expectations about data collection and analysis. By interacting with customers and asking for their opinion, businesses can better understand the types of data people want to contribute and under what conditions. This knowledge can then be used to impact the design of data gathering and analysis procedures to ensure they meet user preferences and expectations.

Businesses could allow users only to accept location tracking if they want their location data to be shared. According to Li et al. [42], a customer-centric approach can help control privacy in IoT devices. Businesses can create better user interactions and boost user trust by concentrating on their privacy requirements.

Improve Transparency and Control

A focus on the needs of the client improves privacy transparency and user control. Companies can enable users to take control of their data by giving them clear information about what data is gathered, how it will be used, and with whom it will be shared. Because users will feel more confident that their data is being handled responsibly due to this transparency, it also helps to increase user trust.

A customer-centric strategy gives users more control over their data and precise information. Companies may, for instance, provide customers the choice to delete their data, set data usage restrictions, or refuse to collect any data. By offering customers control over their data, companies can show their dedication to privacy and increase user confidence. According to Milne and Culnan [43] (2004) analysis of methods for lowering online privacy hazards, a customer-centric strategy may help allay privacy worries.

Create A User-friendly Experience

Creating a user-friendly experience that makes it simple for customers to manage their privacy preferences is another aspect of a customer-centric approach to privacy. This can include easy-to-use privacy controls, a concise privacy statement, and a clear explanation of how information is used. Companies can ensure that users are more likely to use privacy features and feel more confident about giving data by integrating privacy into every aspect of the user experience.

A customer-centric strategy also involves building data collection and analysis methods to have the most negligible adverse effects on users. This can entail limiting the quantity of data acquired, protecting user identities, and avoiding intrusive tracking tactics. By performing these actions, businesses can show their dedication to user privacy and increase user confidence.

Respond To User Concerns

Responding to user complaints and feedback is another aspect of a customer-centric approach to privacy. Companies should take customers' complaints about specific data collecting or analytic procedures seriously and act to resolve them. This can entail altering the method used to gather the data, giving the individuals more control over it, or providing a more straightforward explanation of how the data will be used. Businesses may win customers' trust and prove their dedication to privacy by gently and promptly addressing user complaints.

6.2 Analysis of The Benefits of a Customer-centric Approach

Beyond preserving user privacy, a customer-centric approach to privacy has many advantages. Companies may boost trust, engagement, and customer loyalty by prioritizing user wants and preferences in the design and implementation of data gathering and analysis processes. To develop trust, engagement, and client loyalty, this section examines the advantages of a customer-centric approach.

Create Trust

A customer-centric approach to privacy has many advantages, one of which is that it can foster trust between businesses and customers. Companies may show their commitment to privacy and foster better relationships by prioritizing users' demands and preferences in data collection and processing. Users may be more inclined to disclose their data and use the business's goods and services. Users are more likely to share personal information to assist a business in improving its goods and services if they trust it. Users are more likely to criticize a product's features, usage habits, or prospective issues. By using this feedback, businesses can better understand their customers' needs and preferences to design more tailored experiences.

Increase Engagement

A customer-centric approach to privacy can also raise user engagement with a company's goods and services. Companies can improve the user experience by establishing data collecting and analysis procedures that reflect customers' needs and preferences. Raising users' propensity to use the business's goods and services can raise utilization and income. In addition to creating user-friendly data collecting and analysis procedures, businesses can set themselves apart from rivals by implementing privacy features. For instance, a business that values privacy will probably draw customers concerned about misusing their personal information. Companies may attract a more devoted user base and increase customer retention by differentiating in this way.

Build Customer Loyalty

Finally, a customer-focused attitude to privacy can encourage repeat business. Companies may offer a more personalized experience to fulfill users' demands by prioritizing user preferences and needs. As a result, there is a greater chance that users will stick around and use the business's goods and services in the long run, boosting sales and patronage. Companies can also foster a sense of ownership and investment in the data that customers supply by granting users control over their data. Users are more likely to feel invested in the business and increase their reliance on its goods and services when they perceive that they have control over their data.

6.3 Methodology

6.3.1 Research Design and Methodology

The objective is to comprehend how customers view data privacy and what factors they consider important when providing businesses with their personal information. Customers weigh various variables when determining whether to give their data, including the data being gathered, how it will be used, who will have access to it, and the degree of security measures in place. The goals are to understand client expectations for data privacy and assist businesses in meeting them.

The study design combines qualitative and quantitative techniques, such as data analysis and interviews. Interviews are used to acquire in-depth information on users' experiences and the difficulties of protecting their privacy. In-depth information regarding users' experiences with privacy in digital systems, worries, and attitudes toward current privacy solutions are gathered through interviews. Surveys are employed to collect and analyze information about privacy concerns and solutions. The purpose of the survey on customer privacy preferences was to gather information on the frequency of privacy infractions and customer perceptions of current privacy problems and potential solutions.

Combining these techniques made it possible to gain a deeper understanding of customers' difficulties. The techniques support one another and offer a thorough understanding to validate and bolster findings. Enterprises could better grasp the nature and scope of privacy concerns and look at more practical solutions by collecting data from diverse sources.

6.3.2 Data Collection and Analysis Methods

Two methods were employed to collect data, 1) interviews - which were conducted with a diverse group of individuals, and 2) a survey - which was administered to the general population that included a range of questions about data privacy and customer perceptions.

The interview participants were selected through an initial survey (see Appendix C). The survey is designed to gain a comprehensive understanding of the intricate factors that guide customer preferences regarding personalized ads and content. To ensure a diverse range of backgrounds among potential interviewees, the interview was created to incorporate different perspectives without being prescriptive. The final list included 14 individuals, 6 women and 8 men, with varied backgrounds; this included 2 technical professionals, 2 non-technical professionals, two parents, 2 Gen Xs, 3 millennials, and 2 Gen Zs. The interviews were conducted to collect more in-depth qualitative data from individuals, conducted remotely in a semi-structured format.

As a second data collection method, a survey was conducted to gather user preferences for online advertising and content delivery. Standardized questions were used, and the survey was available online to the general public. The survey aimed to gather anonymous data on various topics, including personalized vs. non-targeted advertisements, privacy concerns, payment methods, data storage, training procedures, and company reliability (see Appendix D).

6.3.3 Interviews

Interview Guide

The interview is designed to provide comprehensive steps to explore the customers' attitudes and behaviors regarding data privacy. The guide includes three main sections to explore the customer needs, 1) the objective, 2) the challenges and concerns, and 3) open-ended questions.

The interviews aim to learn more about people's attitudes toward and actions related to data privacy, as well as their beliefs regarding how enterprises and governments ought to handle personal data. The questions are designed to understand better how people safeguard their personal data online and what privacy-enhancing measures they would like to implement.

The interview focuses on gathering information on customer interactions with personalized ads and how businesses gather data. It contains specific open-ended questions that interviewers can use to discover more about the interviewee's background. These questions inquire about many things, such as the users' surfing patterns, interactions with adverts, data security, identity theft, and user experience. Interviewees get the opportunity to share their memories and experiences. Interviewees were given the opportunity to express themselves through three open-ended questions by sharing their experiences with data breaches, theft, and other personal experiences.

Interview Questions

1. What image or metaphor comes to mind when you see personalized content or ads?
2. What importance do you place on data privacy?

3. How comfortable are you letting companies have your personal information?
4. What details about yourself are you comfortable disclosing?
5. How do you usually safeguard your personal data online?
6. Have you ever had your data compromised or your identity stolen?
7. What notification method would you prefer in the event of a data breach?
8. Are you aware of the privacy practices of the businesses you do business with?
9. Have you ever declined to allow a business to utilize your data? Why, if so?
10. How much control would you like over the data companies collect about you?
11. Do you believe that companies are transparent with information about how they collect data?
12. Do you trust companies with your personal data?
13. Would you be prepared to pay more for additional privacy protections?
14. Have you ever been taken aback by how a business used your data?
15. What measures should businesses take to safeguard your privacy?

Conducting the Interviews

It is important to note that the interview guide is not prescriptive but gives an overall outline and helps probe into additional relevant topics. This helps gain a more comprehensive understanding of customer requirements, which can inform the customer needs from a diverse perspective and provide valuable insights.

The interview guide provides a strong starting point. During the interviews, the interviewee is asked essential questions about the topic. Permission to take notes and publish anonymized information was obtained from each interviewee. Post-interviews, the raw notes were transcribed to ensure accurate and detailed notes of the interviews. This was a critical step in the research process, allowing careful data analysis and drawing meaningful conclusions from it.

Scrubbing the Transcripts

The key pain points from the interviews are summarized and compiled into clusters in similar categories. Then similar needs are assorted and merged into L2 clusters and kept the most unique 2-3 needs within L1 needs. Additionally, a tally of each concern's frequency was conducted, which was used as a gauge when determining which concerns held more significance. See Appendix for a word cloud which represents the keywords that were used frequently such as privacy, personal data, breach, theft and consent.

6.3.4 Survey

Description Of the Sample Population

145 people participated in the poll and were asked to supply personal information. 47.22 percent of all respondents were between the ages of 45 and 60, making up most respondents. The next group was people over the age of 60, who made up 22.22 percent of all respondents. The survey successfully drew a wide range of ages, with some representation from younger age groups, as evidenced by the participants' age distribution.

Most respondents were female, at 58.33 percent of the total, while the proportion of male respondents was 41.67 percent. The survey garnered more female than male participants, which may indicate that the survey was more attractive to women.

Regarding household income, the survey's findings indicate that 25.69 percent of respondents, or most respondents, reported having an annual income of between $25,000 and $49,999. Those making between $50,000 and $74,999 annually came in second, making up 25.00 percent of all respondents. 9.72 percent of respondents, who were a relatively tiny fraction, said they made $150,000 annually.

Participants in the study came from a wide range of ages, genders, and household income levels overall. The poll might benefit from initiatives to draw more male participation and more participants from a wider variety of income levels, given the more significant number of female participants and the concentration of respondents in the middle-income categories.

6.4 Results

6.4.1 Analysis of Data Collected

Affinity Diagram from the Interviews

To gain a better understanding of the users' needs, concerns, and challenges from the interviews that were conducted, the transcribed notes were grouped into their needs and concerns and looked to identify the essential requirements. This information was used to create an affinity diagram (*See Table*

5.), which allowed us to group the statements into level 1 categories based on their relationships. This helped us find the main topics needed to dive deeper into. Further grouping of similar needs in level 2 categories based on their relationships was done. Doing this helped us learn more about our customers' needs.

By organizing statements into level 1 and level 2 categories, the most critical needs of our interview cohort were found. The focus was on those needs. The affinity diagram exercise was an excellent way to determine our customers' needs and how enterprises need to respond to feedback by delivering a service that meets their needs.

The insights reveal that people from different generations, including Gen Z to Boomers, value data privacy and take measures to protect their personal information. They also express concern about companies' data collection practices and the need for transparency and control over their data. Moreover, respondents exhibit a shared distrust of personalized ads and a desire for extra privacy protections. Views on government regulation of data privacy are mixed, with some believing in a balance between privacy protection and business operations.

Level 2	Level 1	Scrubbed Statements
Customers want control on personal information shared with businesses	● Customers want to have control over their data and who has access to it ● Customers want more strict privacy policies ● Customers want a more consent-based data sharing	● I feel like I do not have any control over the data that companies collect about me since my information is put in a lot of places ● Data privacy is a fundamental human right that should be protected
Customers want transparency in data collection practices	● Customers want their privacy to be respected and not violated by businesses or the government ● Customers want businesses to obtain their consent before collecting and using their data	● I don't think companies are transparent enough about their data collection practices ● I do not feel that companies are transparent about their data collection practices
Customers have a need for education and knowledge	● Customers want resources for educating on data privacy ● Customers want knowledge on increased awareness on the	● I think it's important for companies to be clear and upfront about how they're using customer data ● I feel the need for guidance and

	risks of data breaches	education to keep their personal information safe online ● I think the privacy policy should be condensed and put into words that people can actually understand
Customers want to limit the amount of personal information shared online	● Customers want opt-out options for data sharing ● Customers want to only provide personal information to trusted and reputable websites and businesses	● I am only comfortable sharing basic information like my name, profession, and email address, but would never share my phone number ● I prioritize the protection of my personal information and take active steps to ensure that my data is handled responsibly and ethically
Customers want standardized regulation of data privacy	● Customers want stricter regulations on businesses ● Customer want the government to imposing harsher penalties for data breaches	● I think government regulation can create clear rules for collecting, using, and safeguarding personal data, stopping privacy violations and data breaches that might harm personal information
Customers want accountability between government and businesses	● Customers want the government and businesses to work together to protect customer data. ● Customers want businesses and governments to be held accountable for any data breaches or misuse of data	● Companies should take proactive steps to protect customer privacy. ● Government oversight is necessary to regulate data privacy since companies have failed to handle personal data responsibly

Table 5. Affinity diagram

These many interviews may teach us a lot about people's attitudes and worries surrounding data privacy. The interviewees first voiced worries about their ability to regulate their personal data and how firms would use it. Second, they express general suspicion against businesses' data-gathering methods and call for transparency regarding the types of data being gathered and their intended uses. Finally, there are a variety of perspectives on how the government should regulate data privacy, with some favoring government intervention and others placing more responsibility on individuals to safeguard their data.

These insights underscore the complexity of data privacy and the need for a balanced approach that considers the perspectives and concerns of different generations and demographics. These findings highlight the importance of data privacy and suggest that individuals' perspectives on data privacy differ based on their demographic characteristics.

Discussion On the Survey Results

The survey looked at customer preferences and attitudes about privacy and tailored content. 145 persons in total took part in the survey. They were questioned regarding their preferences for various forms of personalized content, how frequently they engage with it, what motivates them, how important privacy is, and how comfortable they are with businesses utilizing their data to create personalized content. After the poll, 62.76 percent of respondents preferred receiving product recommendations, followed by personalized recommendations (55.86 percent) and personalized advertisements (29.66 percent). Only 12.41 percent of respondents did not select an answer. This shows that customers interact more with content customized to meet their needs and interests. In addition, 4.14 percent of respondents always participated with tailored material, 26.90 percent rarely engaged, 19.31 percent frequently engaged, and 49.66 percent occasionally engaged in it.

According to the survey, affordability, convenience, and recency were the most important factors to consider when deciding whether or not to engage with personalized information. This demonstrates that users value helpful content and are eager to interact with it if it satisfies their needs.

61.38 percent of respondents regarded privacy as very important, showing that people are becoming increasingly aware of their privacy rights and concerned about how their personal information is collected and used. When asked how comfortable people felt about businesses or brands utilizing their personal information to create personalized content, 30.34 percent of respondents said they were very slightly at ease, 27.59 percent said they were not at all at ease, and 14.48 percent said they were very at ease. Comparatively, 24.14 percent of respondents were impartial, while 3.45 percent selected "don't know". This demonstrates that many customers still need to be persuaded to provide businesses with their personal information and that they want their data to be treated securely and ethically.

When asked what motivates them to receive ads and personalized content, 36.55 percent of respondents said that ensuring their data is completely secure is the most important factor, followed by

31.03 percent who said that discounts or special offers on goods and services they care about, 20 percent who said that being able to customize ad types and content, and 10.34 percent who said that they enjoy viewing enhanced ads and content that is tailored to their interests. This implies that customers are more inclined to use tailored content if they desire greater control over their personal data and the possibility of receiving a benefit or reward.

The company's reputation was the most important decision factor when selecting whether to accept tailored adverts and content. This was followed by the advertising and content relevant to interests and the privacy concerns and perspectives noted above. This suggests that customers are more inclined to interact with tailored content if they believe in the brand and are concerned about how their personal information will be used or shared. In addition, when asked how much they trust businesses or brands that offer customized content, 41.38 percent of respondents said they do, followed by 13.79 percent who said they do so very much, 25.52 percent who said they were neutral, and 2.76 percent who said they don't know. Companies must try to earn customers' trust and guarantee that data collection and use are transparent. 53.10 percent of interviewees stated they wanted full control over personal information, while 28.97 percent said they wanted some control. Additionally, participants were concerned about data breaches (20 percent), misuse of personal data (40.00 percent), and unauthorized access to personal data (26.90 percent). This demonstrates that customers are concerned about the security of their personal information and want more control over it.

33.79 percent of participants said they have some faith in a company or brand to safeguard their personal information from abuse or unauthorized access when it comes to trust. 15.86 percent of respondents said they trusted them a lot, 20.69 percent said they were neutral, and 26.21 percent said they did not trust them at all. Regarding data use openness, 32.41 percent of respondents think businesses or brands are at least somewhat upfront about utilizing customer information to create personalized content. In contrast, just 29.66 percent are completely transparent. This again underlines how crucial honesty and trust are to developing strong client relationships.

46.21 percent of respondents who were asked about data collection expressed great worry about the gathering and use of their personal information for tailored content. 18.62 percent are neutral, and 32.41 percent are only slightly concerned. This demonstrates that customers are more conscious of the possible risks associated with sharing their personal data and want it to be handled ethically and responsibly. Participants were asked if they would pay more for privacy protection. Of those, 30.34 percent replied yes, while 37.93 percent said no. 46.21 percent of those who would pay stated they would pay no more than $5 per month. This demonstrates that while some customers are prepared to pay for more privacy, others still need to. Businesses must therefore find a balance between offering privacy services and keeping them inexpensive for their clients.

33.79 percent of respondents indicated they would pay for a VPN service as the specific privacy protection they would be willing to pay, while 24.14 percent said they would pay for an ad-free browsing experience. According to 33.10 percent of respondents, the assurance that a corporation will not share or sell personal data to third parties was the main motivator for paying for privacy. This demonstrates that customers want more control over how their personal information is used and want to ensure it is done ethically and responsibly.

The poll covers a range of topics related to data privacy, including privacy management, privacy concerns, trust in business practices, transparency, problems with data gathering, desire to pay for additional privacy measures, and the amount of money respondents are ready to spend. According to survey findings, respondents care about their privacy and are prepared to pay for privacy-enhancing services like VPNs, ad-free browsing, and social network privacy. These discoveries help businesses offer privacy services that address client preferences and worries. The poll also emphasizes how crucial it is to convey to customers that their personal information is secure and will not be disclosed to outside parties.

6.4.2 Summary of The Findings

Customer preferences and privacy concerns are expressed in the customer statements comprising the qualitative interview data. According to the statements, customers desire greater control over their personal information, greater transparency in data collection methods, education and awareness of data privacy, uniform data privacy regulations, and accountability between governments and corporations. Customers also want the ability to opt out of data sharing and only want to give their personal information to reputable and trusted websites and companies. According to the qualitative data, customers cherish their privacy and want businesses and governments to respect and defend it.

Customer opinions and preferences toward tailored content and privacy make up the survey's quantitative data. According to the survey, customers are more likely to interact with material that is specifically catered to their interests and requirements, with product recommendations emerging as the most popular type of personalized content. When interacting with personalized content, relevance dominated, followed by cost, ease of use, and freshness. The poll also revealed that customer responses about their preferences and data privacy worries comprise the qualitative data from the interviews. According to the declarations, customers desire greater control over their personal information, greater transparency in data collection methods, education and awareness of data privacy, uniform data privacy regulations, and accountability between governments and corporations. Customers also want the ability to opt out of data sharing and only want to give their personal information to reputable and trusted websites and companies. According to the qualitative data, customers cherish their privacy and want businesses and governments to respect and defend it.

The survey's quantitative data comprises customer opinions and preferences toward tailored content and privacy. According to the report, customers are more likely to interact with material that is specifically catered to their interests and requirements, with product recommendations being the most popular type of personalized content. When interacting with personalized content, relevance dominated, followed by cost, ease of use, and freshness. The survey also revealed that people value privacy and are growing more cognizant of their privacy rights. Customers still need persuading to give businesses access to their personal information, and they want that access to be protected and appropriately utilized. If users want more control over their personal data and expect to gain some advantage or reward, they are more inclined to consume tailored content. According to the survey, customers who trust a brand and care that their personal information won't be misused or shared with outside parties are more inclined to interact with personalized material.

Customers prefer personalized material that is pertinent to them and beneficial. However, they also value privacy and desire control over their personal information by combining qualitative and quantitative data in a triangulated manner. They seek resources to raise enterprises ' and the government's knowledge and awareness of the hazards to data privacy and transparency and accountability from both. Their level of trust heavily influences their decision to interact with tailored material, and they are concerned about the security and proper use of their personal data.

According to qualitative and quantitative evidence, customers prioritize privacy and desire greater control over their personal information. They favor responsibility between enterprises and the government, consistent legislation, data privacy education, and openness. Customers are more likely to interact with personalized material pertinent to their needs and interests if they have control over their personal data and if the company collecting it has their trust.

6.5 Strategies That Enterprises Can Use to Adopt a Customer-centric Approach

Companies that want to take a customer-centric approach to privacy must prioritize user demands and preferences when collecting and analyzing data. This makes it necessary to switch from traditional data collection methods that prioritize company interests to those that prioritize user interests. This research looks at mechanisms companies may employ, like user control, transparency, and informed consent, to implement a customer-centric strategy.

To adopt a customer-centric approach and provide goods and services that meet those needs, businesses must recognize and prioritize their customers' wants. However, it is crucial to remember that taking those actions alone might not produce the results the company desires. As a result, they must consider the system as a whole and each subsystem's interactions. Businesses can use system decomposition and the Design Structure Matrix [44] (DSM) to understand the system more holistically.

A complex system is decomposed into smaller, easier-to-handle components. By decomposing a system's numerous components, businesses can better understand how they interact. To better serve customers, this can be used to pinpoint areas where their demands are not being satisfied.

A DSM helps businesses illustrate the connections and interactions among a system's many parts. Businesses might identify possible bottlenecks or areas for development to better serve customers by visualizing these relationships.

Enterprises can use system decomposition and DSM to implement a customer-centric strategy by doing the following:

- Determine the system's elements that are pertinent to customer requirements. Products, services, procedures, and client interactions may all fall under this category.

- Separate each component into its smaller components. This can assist in locating specific areas for improvement.
- DSM can be used to visualize the relationships and interactions between the various components. This can assist in finding possible bottlenecks or places where processes should be improved to serve customers better.
- Based on client needs and the effect on the whole system, prioritize the areas that need improvement.
- Create and put into action plans to enhance the customer experience in the areas that are given priority. This could involve streamlining procedures, boosting customer relations, or revamping goods or services.
- Keep track of and assess how the modifications are affecting customer satisfaction and system performance. By doing so, the customer-centric strategy can be improved, and further areas for development can be found.

By using system decomposition and DSM, businesses can better understand their systems and how they engage with customers. By using this information, businesses can develop and put into practice plans that prioritize their customers' requirements.

System decomposition and the Design Structure Matrix (DSM), which will be covered in Chapter 7, are tools businesses can use. Thanks to the systems approach, they can improve user control, transparency, and informed consent. Transparency is essential because it enables businesses to acquire, use, and share correct data with users, building their confidence and engagement. Businesses can provide various data management solutions to enable users to regulate their data usage, which is also crucial. Informed consent is equally crucial, which allows users to opt-out and be informed about data collection and usage. Businesses may boost engagement, encourage client loyalty, and foster trust by prioritizing user demands and preferences.

Building great customer relationships can be facilitated by adopting a customer-centric approach to privacy. It can enhance brand reputation and customer loyalty by fostering a favorable perception, fostering trust, and encouraging repeat business and positive recommendations. Additionally, adhering to privacy laws and regulations can assist companies in putting a priority on user privacy, lowering the risk of data breaches, avoiding fines, and maintaining a positive reputation in the marketplace.

Businesses may lower the risk of data breaches and get more precise and pertinent data by employing transparent data procedures, user controls, and informed consent. This might lead to higher-quality data analysis and decision-making, resulting in better commercial results. Additionally, putting user privacy first and taking a customer-centric stance can set businesses apart from their rivals, giving them an edge in the marketplace and drawing in privacy-conscious clients.

Chapter 7 - System Analysis

7.1 System Decomposition

To manage a system's complexity, system decomposition [45] is crucial. Understanding how complex systems function and their various subsystems interact requires breaking the system into smaller, easier-to-handle parts. This makes the engineering process more effective and efficient by enabling the engineers to concentrate on system components.

System decomposition is how a complex system is broken down into more manageable, smaller components. When working with intricate systems like data privacy, it is vital to understand how the system works and how its many components interact. By breaking the larger, more complex system into smaller subsystems, engineers may focus on specific system components, streamlining and enhancing the design process.

System decomposition in the context of user privacy refers to segmenting a business's data privacy system into several subsystems, each with its own set of policies, processes, and controls. These subsystems consist of data collection, storage, processing, analysis, access, authorization, privacy controls, privacy compliance auditing, notification of privacy breaches, employee and customer training, and awareness-raising initiatives.

The data collecting and storage subsystem defines standards for data collection methods, accuracy, retention, and encryption. To ensure that data is secure both in transit and at rest, encryption and protection of that data are also crucial. This planning also includes the infrastructure required for data storage, including the necessary hardware and software, the data center's location, security measures, and backup and recovery protocols.

The data processing and analysis subsystem creates rules for data access limitations, classification, and analysis algorithms. Utilizing data anonymization methods allows for data privacy protection during analysis. Using Federated Learning systems, machine learning models can be trained while protecting data privacy. Data privacy can be protected through Differential Privacy systems, which restrict data usage and interruption.

The data access and authorization subsystem is responsible for establishing standards for authentication and access control methods. To assign access privileges based on job duties, role-based access control systems can be utilized. Multi-factor authentication and access request methods can be used to ensure data privacy.

Data collection, use, and deletion privacy policies are developed for data privacy controls in the subsystem. The rights of clients and users to privacy must also be considered. An auditing system for privacy compliance can be used to ensure that the business complies with data privacy rules. A privacy breach reporting system must be in place to find, prioritize, react to, and resolve data breaches.

To ensure that everyone knows the need for data privacy, programs for staff and customer education and awareness should be created. For privacy training programs to be effective, they must be developed, delivered, and evaluated. A privacy culture can be promoted within the organization using awareness campaigns, privacy culture, and feedback mechanisms.

Businesses' ability to manage the complexity of data privacy efficiently depends critically on system breakdown. It requires breaking the system into smaller subsystems, each with its own rules, procedures, and control mechanisms. By breaking the more complex system down into smaller subsystems, businesses can more effectively handle the complexity of data privacy requirements and guarantee that they collect, store, and exchange personal data safely and ethically.

This breakdown describes the different parts of a customer data privacy system, which is intended to guarantee that customer data is gathered, processed, and used in a way that respects each individual's right to privacy. The system is divided into several crucial sections, each of which has its own set of guidelines, procedures, and controls:

Data Collection and Storage: The focus of this component is on gathering and storing client data, including guidelines for data collection techniques, accuracy, storage needs, and infrastructure specifications. Guidelines for data encryption and protection are also included.

Data Processing and Analysis: This section covers the methods used to process and analyze customer data, including rules for data classification, access control, and algorithm selection and validation. It also includes data anonymization methods, a Federated Learning system for distributed data analysis, and a Differential Privacy system for safeguarding private data.

Data Access and Authorization: This section addresses how customer data is accessed and used, including role-based access control, access request and approval, and policies for data access and authorization. To protect privacy, it also consists of data perturbation techniques.

Data privacy controls: This section consists of rules and methods for guaranteeing that customer data is handled by privacy laws, such as privacy policies, data retention and deletion policies, and compliance audit procedures.

Training and awareness programs for customers and employees are also included in this component's description of the steps to be taken in the event of a privacy breach. These steps include incident identification and triage, incident response, and issue resolution.

This breakdown offers a thorough overview of all the elements needed to create a strong customer data privacy system, from data collection to storage, processing, analysis, and utilization. By following these tips, enterprises can ensure that customer data is managed better while also being compliant with regulatory standards.

0. Customer Data Privacy System
    1. Data Collection and Storage
        a. Data Collection Policy

i. Data Collection Methods
ii. Data Collection Accuracy
iii. Data Collection Retention

b. Data Storage Infrastructure
i. Hardware and Software Requirements
ii. Data Center Location and Security
iii. Backup and Recovery Procedures

c. Data Encryption and Protection
i. Encryption Methods
ii. Encryption Implementation

2. Data Processing and Analysis

a. Data Processing Policy
i. Data Access Controls
ii. Data Classification

b. Data Analysis Algorithms
i. Algorithm Selection and Validation
ii. Data Anonymization Techniques

c. Federated Learning System
i. Federated Model Training
ii. Federated Learning Optimization

d. Differential Privacy System
i. Data Usage Control
ii. Data Perturbation

3. Data Access and Authorization

a. Access Control Policy
i. Control Framework
ii. Policy Review & Implementation

b. Authentication and Authorization Mechanisms
i. Multi-Factor Authentication
ii. Access Request Methods

c. Role-Based Access Control System
i. Role Assignment

ii. Access Request & Approval

4. Data Privacy Controls
   a. Privacy Policy
      i. Data Collection Use
      ii. Data Retention and Deletion Policies
      iii. Privacy Rights of Customers and Users
   b. Privacy Compliance Auditing System
      i. Audit Planning and Execution
      ii. Audit Reporting and Follow-Up
   c. Privacy Breach Notification System
      i. Incident Identification and Triage
      ii. Incident Response and Resolution
5. Employee & Customer Training and Awareness
   a. Privacy Training Program
      i. Training Delivery
      ii. Training Curriculum
      iii. Training Evaluation
   b. Privacy Awareness Program
      i. Awareness Campaigns
      ii. Privacy Culture
      iii. Feedback Mechanisms

In Figure 18. the level 1 system decomposition of the Customer Data Privacy System is shown. This is the high-level view of the system, which shows the main components that make up the system. These components include the Data Collection and Storage, Data Privacy Controls, Data Processing and Analysis, Data Access and Authorization, and Employee & Customer training and awareness.

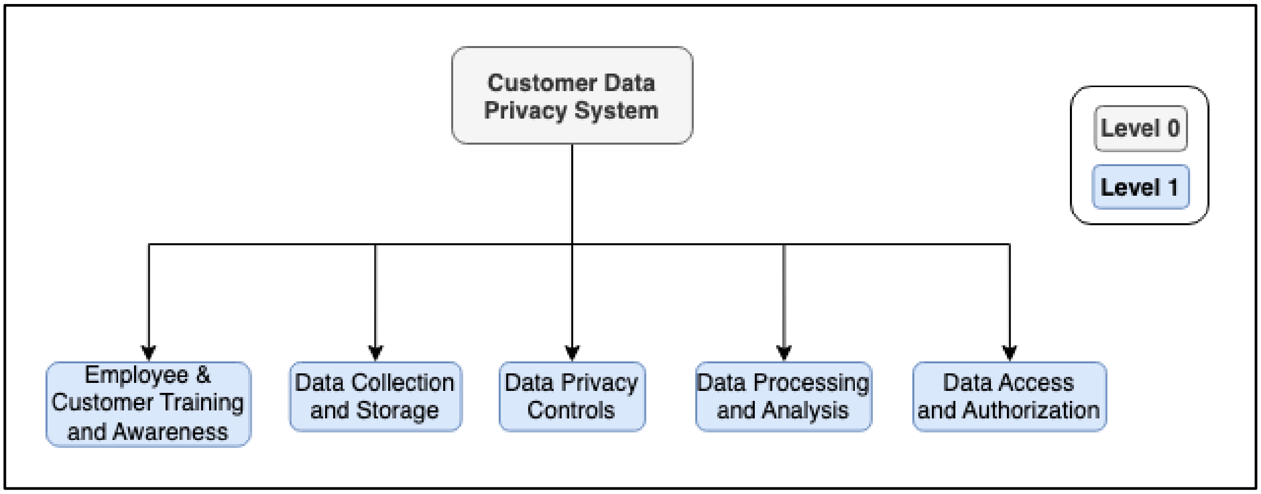

Customer Data Privacy System
Level 0
Level 1
Employee & Customer Training and Awareness
Data Collection and Storage
Data Privacy Controls
Data Processing and Analysis
Data Access and Authorization

Figure 18. Level 1 System decomposition [by: Mervin Govada]

Figure 19. shows the Customer Data Privacy System's level 2 system decomposition (see appendix A for a zoomed system decomposition). This more in-depth view of the system demonstrates how the major parts are further divided into more compact sub-parts. The Data Processing and Analysis, for instance, is divided into several smaller parts, including the Data Processing Policy, Data Analysis Algorithms, Federated Learning System, and Differential Privacy System. Other sub-components are similarly divided into several sub-components.

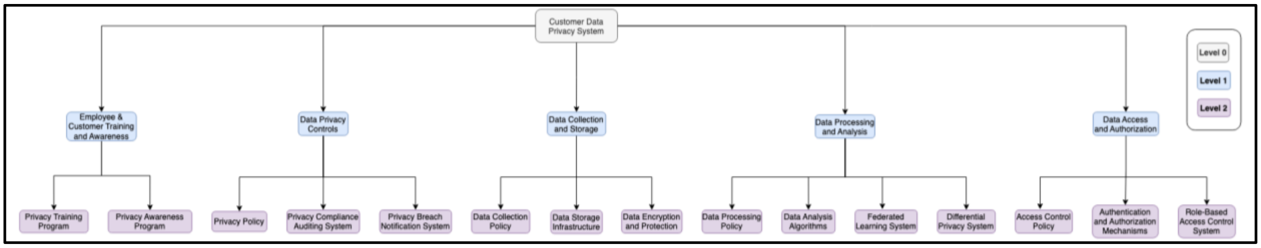

Customer Data Privacy System
Level 0
Level 1
Level 2
Employee & Customer Training and Awareness
Data Privacy Controls
Data Collection and Storage
Data Processing and Analysis
Data Access and Authorization
Privacy Training Program
Privacy Awareness Program
Privacy Policy
Privacy Compliance Auditing System
Privacy Breach Notification System
Data Collection Policy
Data Storage Infrastructure
Data Encryption and Protection
Data Processing Policy
Data Analysis Algorithms
Federated Learning System
Differential Privacy System
Access Control Policy
Authentication and Authorization Mechanisms
Role-Based Access Control System

Figure 19. Level 2 System decomposition [by: Mervin Govada]

A system boundary is a line that divides a system from its surroundings in the context of the systems approach, specifying what is contained inside the system and what is excluded. The system border for a customer data privacy system would consist of all the parts and procedures used in gathering, storing, analyzing, and exchanging customer data and the rules and regulations that control these operations. Any interactions with other systems or stakeholders, such as regulatory bodies or third-party service providers, that affect the confidentiality of customer data would also be considered part of the system border. The organization may ensure that all pertinent variables are considered when creating and executing the customer data privacy system and that any potential risks and obstacles are recognized and addressed by clearly defining the system boundary.

7.2 Design Structure Matrix (DSM)

A visual representation of the relationships between parts of a system or a product is called a Design Structure Matrix (DSM). It is a matrix where each row and column represent a component, and the cells stand in for the relationships or interactions between those components. DSM is useful for managing challenging product development projects because it enhances team communication, identifies potential issues early, and streamlines the design process.

The "Design Structure Matrix Method" (DSM method), which Eppinger [46] created, is a systematic method for using DSMs to manage the processes involved in product development. The DSM method entails creating a DSM of the system being designed, identifying the crucial components, and then using the DSM to streamline the development procedure by rearranging the dependencies between the components. The DSM method can shorten project lead times, lower project costs, and boost the final product's performance and quality.

DSM models and manages the intricate relationships between a system's various parts and subsystems. The DSM is a matrix-based method for representing the connections between complex subsystems. DSM is a square matrix, with the cells denoting the strength of the relationships between the subsystems and the rows and columns denoting the subsystems. An empty cell denotes an independent subsystem, while a value of "X" in a cell denotes a highly interdependent subsystem. The DSM matrix can visualize the system's structure, identify the highly interdependent subsystems, and examine the effects of changes to one subsystem on the other.

Browning, Eppinger, and Rowles [47] describe a DSM-based methodology for managing the design of complex systems. Managing the interdependencies among subsystems is critical for designing complex systems. The DSM-based methodology includes several steps, such as defining the system architecture, creating a DSM model of the system, identifying the subsystems with high interdependencies, and managing the dependencies between subsystems.

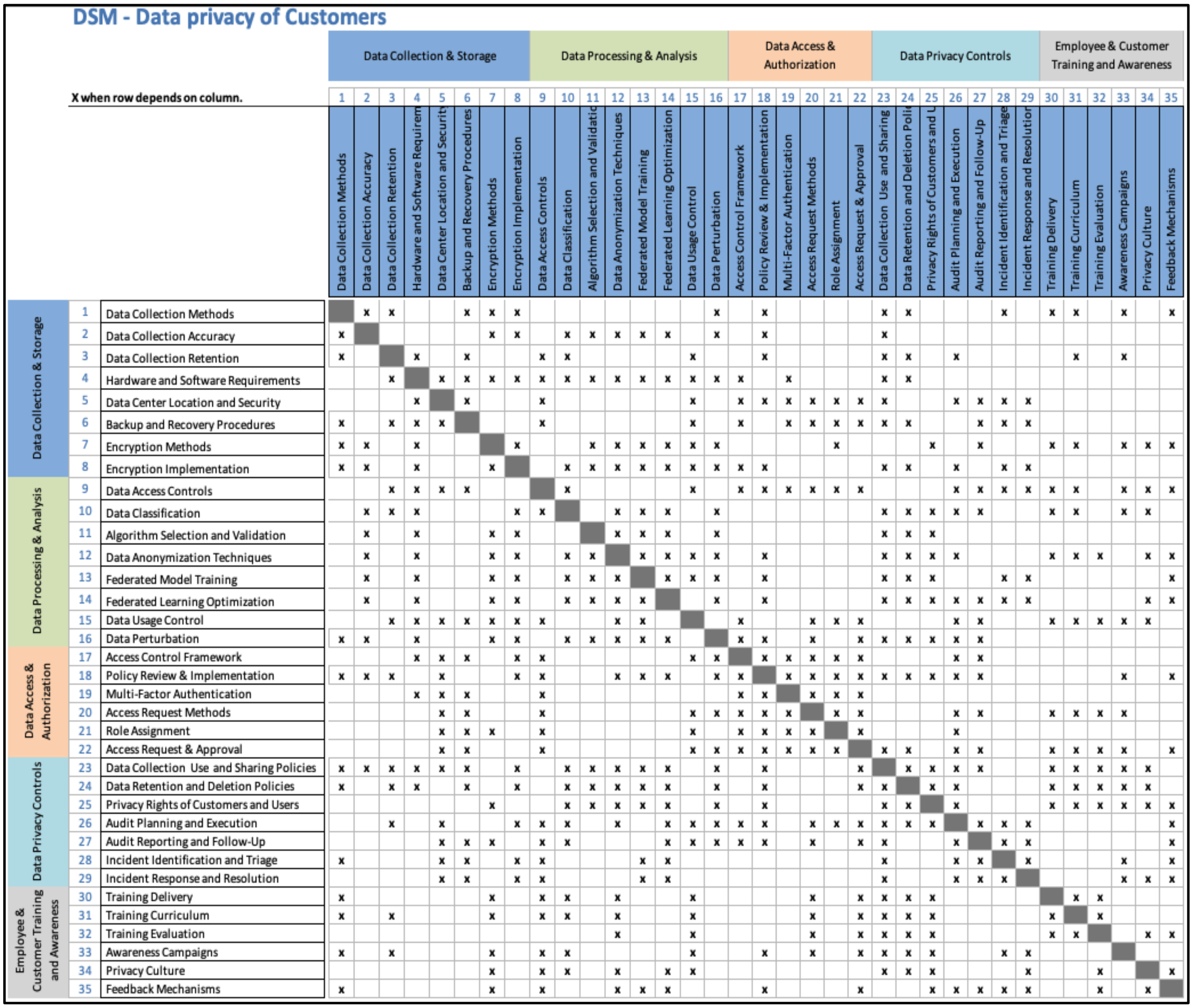
Eppinger and Browning [48] provide a comprehensive overview of DSM methods and their applications. It highlights the use of DSM in various stages of the design process, such as concept generation, product architecture design, and system integration.

DSM provides a structured approach for decomposing a system into subsystems, identifying its dependencies, and managing its complexity. The DSM in Figure 20. illustrates the dependencies between each Customer Data Privacy System component and how they impact the overall system. By understanding these dependencies, potential bottlenecks or areas of weakness can be identified that may need to be addressed to improve the system's performance (see appendix B for a zoomed DSM). Additionally, the DSM can help us identify opportunities for optimization and refinement and potential areas for expansion and growth. By using this tool to its fullest potential, enterprises can better understand the system and make informed decisions about managing and enhancing it over time

The Customer Data Privacy Matrix illustrates the different dependencies that affect customer data privacy. These dependencies include data gathering and storing, processing and analyzing, accessing and authorizing, enforcing data privacy regulations, and educating employees and clients. The matrix's rows correspond to the 35 dependencies in total. The columns depict many facets of data privacy, including access control frameworks, data gathering strategies, data collection retention, and encryption techniques.

The DSM offers insightful information regarding the complexity of data privacy and its different dependencies. The interconnectedness between many dimensions of data privacy is an important realization. For instance, the way data is collected (dependency 1) affects the correctness of the data (dependency 2), which affects data analysis (dependency 9). Similar to how restrictions on data usage (dependency 15) affect restrictions on data access (dependency 9), restrictions on data privacy (dependency 23) are then impacted. It is crucial to comprehend these relationships to create effective data privacy policies and practices.

.



DSM - Data privacy of Customers

X when row depends on column.

			Data Collection & Storage								Data Processing & Analysis								Data Access & Authorization						Data Privacy Controls							Employee & Customer Training and Awareness					
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
			Data Collection Methods	Data Collection Accuracy	Data Collection Retention	Hardware and Software Requirem	Data Center Location and Securit	Backup and Recovery Procedures	Encryption Methods	Encryption Implementation	Data Access Controls	Data Classification	Algorithm Selection and Validatio	Data Anonymization Techniques	Federated Model Training	Federated Learning Optimization	Data Usage Control	Data Perturbation	Access Control Framework	Policy Review & Implementation	Multi-Factor Authentication	Access Request Methods	Role Assignment	Access Request & Approval	Data Collection Use and Sharing	Data Retention and Deletion Polic	Privacy Rights of Customers and U	Audit Planning and Execution	Audit Reporting and Follow-Up	Incident Identification and Triage	Incident Response and Resolutior	Training Delivery	Training Curriculum	Training Evaluation	Awareness Campaigns	Privacy Culture	Feedback Mechanisms
Data Collection & Storage	1	Data Collection Methods	■	x	x			x	x	x								x		x					x	x				x		x	x		x		x
	2	Data Collection Accuracy	x	■					x	x		x	x	x	x	x		x		x					x												
	3	Data Collection Retention	x		■	x		x			x	x					x			x					x	x		x					x		x		
	4	Hardware and Software Requirements			x	■	x	x	x	x	x	x	x	x	x	x	x	x	x		x				x	x											
	5	Data Center Location and Security				x	■	x			x						x		x	x	x	x	x	x	x			x	x	x	x						
	6	Backup and Recovery Procedures	x		x	x	x	■			x						x		x		x	x	x	x	x	x			x	x	x						
	7	Encryption Methods	x	x		x			■	x			x	x	x	x	x	x					x				x		x			x	x		x	x	x
	8	Encryption Implementation	x	x		x			x	■		x	x	x	x	x	x	x	x	x					x	x		x		x	x						
Data Processing & Analysis	9	Data Access Controls			x	x	x	x			■	x					x		x	x	x	x	x	x				x	x	x	x	x	x		x	x	x
	10	Data Classification		x	x	x				x	x	■		x	x	x		x							x	x	x	x	x			x	x		x	x	
	11	Algorithm Selection and Validation		x		x			x	x			■	x	x	x		x							x	x	x										
	12	Data Anonymization Techniques		x		x			x	x		x	x	■	x	x	x	x		x					x	x	x	x				x	x	x		x	x
	13	Federated Model Training		x		x			x	x		x	x	x	■	x	x	x		x					x	x	x			x	x						x
	14	Federated Learning Optimization		x		x			x	x		x	x	x	x	■		x		x					x	x	x	x	x	x	x					x	x
	15	Data Usage Control			x	x	x	x	x	x	x			x	x		■		x			x	x	x				x	x			x	x	x	x	x	
	16	Data Perturbation	x	x		x			x	x		x	x	x	x	x		■	x	x		x		x	x	x	x	x	x								
Data Access & Authorization	17	Access Control Framework				x	x	x		x	x						x	x	■	x	x	x	x	x				x	x								
	18	Policy Review & Implementation	x	x	x		x			x	x			x	x	x		x	x	■	x	x	x	x	x	x	x	x	x						x		x
	19	Multi-Factor Authentication				x	x	x			x								x	x	■	x	x	x													
	20	Access Request Methods					x	x			x						x	x	x	x	x	■	x	x				x	x			x	x	x	x		
	21	Role Assignment					x	x	x		x						x		x	x	x	x	■	x				x									
	22	Access Request & Approval					x	x			x						x	x	x	x	x	x	x	■	x	x		x	x			x	x	x	x		x
Data Privacy Controls	23	Data Collection Use and Sharing Policies	x	x	x	x	x	x		x		x	x	x	x	x		x		x				x	■	x	x	x	x			x	x	x	x	x	
	24	Data Retention and Deletion Policies	x		x	x		x		x		x	x	x	x	x		x		x				x	x	■	x	x				x	x	x	x	x	
	25	Privacy Rights of Customers and Users							x			x	x	x	x	x		x		x					x	x	■	x				x	x	x	x	x	x
	26	Audit Planning and Execution			x		x			x	x	x		x		x	x	x	x	x		x	x	x	x	x	x	■	x	x	x						x
	27	Audit Reporting and Follow-Up					x	x	x		x	x				x	x	x	x	x		x		x	x			x	■	x	x						x
	28	Incident Identification and Triage	x				x	x		x	x				x	x									x			x	x	■	x				x		x
	29	Incident Response and Resolution					x	x		x	x				x	x									x			x	x	x	■				x	x	x
Employee & Customer Training and Awareness	30	Training Delivery	x						x		x	x		x			x					x		x	x	x	x					■	x	x			
	31	Training Curriculum	x		x				x		x	x		x			x					x		x	x	x	x					x	■	x			
	32	Training Evaluation												x			x					x		x	x	x	x					x	x	■		x	x
	33	Awareness Campaigns	x		x				x		x	x					x			x		x		x	x	x	x			x	x				■		
	34	Privacy Culture							x		x	x		x		x	x								x	x	x				x			x		■	x
	35	Feedback Mechanisms	x						x		x			x	x	x				x				x			x	x	x	x	x			x		x	■

Figure 20. DSM Customer Data privacy system [by: Mervin Govada]

The necessity of a comprehensive strategy for data privacy is another revelation from the DSM. The DSM emphasizes the significance of handling numerous data privacy issues, including data collection, storage, processing, access, and usage control. Additionally, it highlights the significance of policies and procedures for data retention and deletion, user and customer privacy rights, and the planning and execution of audits. A comprehensive approach to data privacy is required to ensure that all aspects of data privacy are adequately addressed and protected.

The DSM also stresses educating and raising customer and employee awareness. All stakeholders, including employees and customers, must take part in and cooperate with effective data privacy rules and practices. Training and awareness initiatives are required to ensure that workers and customers understand the value of data privacy and how to preserve it. These initiatives can strengthen the organization's privacy culture by preventing unintentional or deliberate data privacy breaches.

The DSM emphasizes reviewing and enhancing data privacy policies and procedures. Policies and procedures regarding data privacy must change to reflect the ongoing evolution of risks and difficulties. Through routine audits and reviews, it is possible to find improvement opportunities and ensure that data privacy policies and practices are still applicable and current.

In conclusion, the DSM offers a helpful framework for comprehending the intricacies of data privacy and the different dependencies that affect it. It emphasizes the significance of a thorough data privacy strategy considering all facets of data collecting, storage, processing, access, and usage management. It also emphasizes the need for customer and employee education, the necessity of continual evaluation, and the necessity of enhancing data privacy rules and procedures.

Chapter 8 - Discussion & Conclusion

8.1 Summary of Research, Key Points, And Impact on User Privacy and Enterprise Landscape.

The thesis analysis reveals the intricate and interrelated systems that support various aspects of customer privacy. We looked at each component and how they interacted with other components and examined the stakeholder demands and goals to understand the system thoroughly. The analysis explored potential new technologies, increased customer control, awareness programs, and standardized regulations and policies as potential effective measures. Stakeholders should improve customer privacy by taking focused steps and considering the potential unintended consequences and trade-offs brought on by an awareness of the system's interconnectedness and feedback loops.

The research indicates that to secure customers' personal information, privacy-enhancing technologies (PETs) must be used throughout the data lifecycle. PETs like Federated Learning and Differential Privacy help increase client data control while lowering the likelihood of data breaches and unauthorized access. Furthermore, implementing privacy legislation will strengthen the legal framework for data protection and improve decision-making processes for government enterprises. By regulating the practices that allow for data sharing between firms, the implementation of consistent regulations and standards will enable better protection of client privacy.

Finally, by being open and honest about data collection methods and notifying customers of how their data is utilized, businesses may increase customer trust. Promoting knowledge of privacy laws, data-gathering practices, and awareness can help protect customer data more effectively.

The research questions that were covered in this thesis are listed below.

1. What are the major obstacles enterprises face when protecting user privacy, and how are these difficulties impacting customer perceptions and behaviors?

Enterprises must balance the requirement to acquire and use user data for commercial purposes while reducing the risk of data breaches and unauthorized access to user data to safeguard user privacy. As discussed in Chapter 3, these issues may negatively impact customers' attitudes and behavior, resulting in a decline in interest in goods and services, reluctance to share personal information and mistrust of businesses. Some of the major difficulties that firms confront include the complexity of data privacy laws and the problem of balancing user privacy and corporate necessity. By weakening trust in companies, lowering the willingness to reveal personal information, and motivating customers to discontinue doing business with companies they feel are not respecting their privacy, these problems may affect customer behavior and attitudes.

2. How can enterprises prioritize transparency, user control, and informed consent in their approach to user privacy, and what factors impact customer perception of these practices?

The research indicates to prioritize user privacy, and enterprises should concentrate on openness, user control, and informed consent. Gaining customers' consent to data collection and use, giving them access to their data, and clearly outlining data collection and use policies are all ways to do this. The readability and clarity of privacy rules are just two crucial characteristics that can significantly impact how customers see these practices. The interviews and surveys clearly demonstrated what the customers expect; to gain the trust of their customers, businesses must provide users control over how their data is used and offer clear, straightforward information regarding data collection, usage, and sharing. The perceived value of the products or services and the perceived risk connected with disclosing personal information can affect how customers perceive these practices.

3. What strategies can enterprises employ to increase user privacy while balancing business needs?

To improve it, businesses must emphasize customer privacy and incorporate it into every aspect of their operations. This requires providing specific privacy policies and controls, implementing PETs discussed in Chapters 4 and 5, and conducting routine privacy assessments and audits to ensure compliance with laws and standards. Additionally, businesses should look to hiring privacy officers or teams, delivering privacy education and awareness campaigns for personnel, and using technology and third-party services to support privacy initiatives are all essential components of a comprehensive privacy program.

4. How is Federated Learning used to maintain user privacy within enterprises, as well as what are its potential advantages and drawbacks?

Businesses can train models on user data using Federated Learning without transferring the data to a central server, safeguarding user privacy. The benefits of Federated Learning include better user privacy and security, reduced risk of data breaches, and higher machine learning model accuracy due to the usage of varied data sets. However, Federated Learning has several disadvantages, such as the need for specialized infrastructure and knowledge, the potential for bias in the training data, and limitations on the types of machine learning models that may be trained.

5. What are the potential advantages and drawbacks of adopting a Differential Privacy approach in enterprises, and how might this strategy increase customer trust, engagement, and loyalty?

Differential privacy enables enterprises to collect and use data for operational purposes while giving users control over how their data is used and shared. This tactic may boost customer trust, engagement, and loyalty since people are more willing to interact with and support companies that respect their privacy. Differential privacy can help companies comply with privacy rules and reduce the risk of data breaches.

6. What factors drive or impede change within an enterprise environment?

As the organizational landscape develops, various forces are pushing for and against change in data privacy. As customers become more aware of the risks to their privacy, they are requesting greater openness and privacy protection. Due to new rules and the need to differentiate themselves from competitors, businesses are placing a higher priority on customer privacy. However, putting in place privacy measures is expensive and complex, and stakeholders who prioritize business goals over privacy frequently oppose change. Additional challenges include,

- Adapting to shifting privacy legislation
- Adding privacy controls in place across diverse business processes and systems
- Sharing customer data with third-party vendors while ensuring their commitment to data privacy

Businesses must develop a culture of trust and privacy protection by engaging with stakeholders, increasing management and employee knowledge of the threats to data privacy, and implementing adequate privacy controls.

7. What options exist to address customer issues?

Companies should address customer concerns about data privacy in various ways, such as increasing transparency, empowering users, providing tools and education, and implementing privacy-enhancing technologies. Customer data can be protected from unlawful access, use, or disclosure through organizational and technical safeguards. To enhance data privacy protections, cutting-edge technologies like Federated Learning and Differential Privacy can be deployed. Clear and thorough privacy rules, processes, and an incident management and response plan can be developed and implemented to address privacy breaches or occurrences promptly.

Establishing partnerships with trustworthy third-party vendors and partners who share a commitment to data privacy is possible. Regular privacy audits and monitoring can also be done to ensure that rules and regulations are followed. To promote a culture of privacy, programs for privacy education and awareness can be regularly delivered to customers and workers. Encouragement of participation with customers and the greater community demonstrates a commitment to appropriate data handling. Working with regulators and other stakeholders makes designing and executing privacy laws and standards that balance user privacy and business needs easier.

8.2 Limitations of The Study and Directions for Future Research

The study includes some limitations that should be considered, despite the results being enlightening. This study's sample size was restricted to respondents from the United States older than 18, which might have impacted the results. Considering this, future studies should employ quantitative methods to survey a substantial population from various backgrounds, including different ages, nations, genders, and online behaviors. This enables researchers to identify personality factors influencing people's opinions and privacy worries.

Investigating the relationship between a person's age and privacy concerns is important because it can show how their attitude toward privacy may evolve. Longitudinal studies that track people over time are required to understand better how privacy concerns vary over time and what elements may affect these changes.

The survey was also conducted online, which may have affected the results. Future research could gather richer and more complicated data on people's attitudes about privacy using alternative methodologies, such as focus groups in addition to the interviews.

It is significant to highlight that the study only looked at the data privacy technologies known as Differential Privacy and Federated Learning. The privacy of individuals may also be affected by other technologies like secure multi-party computation and homomorphic encryption. To better understand how different privacy-enhancing technologies affect users' privacy concerns, future studies might compare and contrast the effectiveness of these technologies.

Even though the survey provided valuable insights into Americans' privacy concerns, a similar study in other nations is necessary to determine whether privacy concerns vary depending on demographic factors. Researchers can discover links between demographic traits, online behaviors, and privacy concerns by examining extensive datasets from multiple populations.

The impact of public awareness- and education-raising initiatives on people's privacy concerns must also be considered. By examining how information campaigns can influence users' attitudes and behaviors toward privacy, researchers can learn more about the effectiveness of these campaigns and identify areas that need to be improved. To guarantee that their staff members understand the significance of data privacy and take the required precautions to secure their customers' data, businesses can also benefit from educating and training them on privacy issues.

8.3 Recommendations for Future Research and Practice

- Future research should adopt a more thorough and sophisticated methodology. Researchers can provide a more thorough understanding of the factors influencing privacy concerns by using quantitative methods, analyzing large datasets from different populations, investigating various privacy-enhancing technologies, and considering education and awareness-raising campaigns. This will help direct the creation of efficient privacy policies and practices that safeguard people's privacy and foster trust in online interactions.
- Developing new, more effective Federated Learning algorithms that work with other technologies, such as Differential Privacy, is being done. This can make it easier to scale this to various use cases while preserving the confidentiality and security of customer information and still allowing businesses to benefit from the pooled knowledge of many users.
- Research the privacy-accuracy trade-offs in Federated Learning using Differential Privacy vs. homomorphic encryption. The accuracy of the model could decline as noise and encryption levels rise. Future studies can examine the ideal equilibrium between accuracy and privacy in Federated Learning.
- Homomorphic encryption has promise as a method for protecting user data in Federated Learning. More research is necessary to develop valuable solutions for applications in the real world. Future research should aim to develop homomorphic encryption algorithms that are effective and scalable enough to be used in Federated Learning.
- Customers should be aware of the privacy risks of sharing their data during Federated Learning. Additionally, they should be informed about how their data is being used. Research should focus on developing efficient educational programs that users can understand. These initiatives should raise customer awareness and help them make wise data-sharing choices.
- Further research is needed on how the government can help improve customer data security and privacy by creating rules and regulations.
- Federated Learning can support sustainability efforts by increasing the effectiveness of machine learning research, lowering energy use for data transmission, and safeguarding individual privacy. Federated Learning may only address some sustainability issues. The effect of Federated Learning on sustainability requires further study.

References

[1] I. T. R. Center, “Annual number of data compromises and individuals impacted in the United States from 2005 to 2022 [Graph,” in Statista, 2023. [Online]. Available: https://www-statista-com.libproxy.mit.edu/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/

[2] I. T. R. Center, “Number of cases of data violation due to cyber attacks in the United States from 2020 to 2022, by industry [Graph,” in Statista, 2023. [Online]. Available: https://www-statista-com.libproxy.mit.edu/statistics/1318379/us-number-of-private-data-compromises-by-industry/

[3] L. Sweeney, “Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3.” Pittsburgh, 2000.

[4] YouGov, “Percentage of internet users in the United States who feel that their data and personal information is vulnerable to hackers as of.” Jul. 30, 2019. [Online]. Available: https://www-statista-com.libproxy.mit.edu/statistics/972911/adults-feel-data-personal-information-vulnerable-hackers-usa/

[5] G. Data, M. Alliance, Acxiom, and F. Factory, “Share of consumers in the United States feeling they lack control over companies handling their personal information as of December 2021 [Graph,” in Statista, 2022. [Online]. Available: https://www-statista-com.libproxy.mit.edu/statistics/1368913/us-consumers-not-having-control-companies-handling-private-data/

[6] Spiceworks, “Areas in which companies are using artificial intelligence (AI) and machine learning (ML) tools according to marketing professionals in the United States as of.” Oct. 01, 2022. [Online]. Available: https://www-statista-com.libproxy.mit.edu/statistics/1364707/ai-ml-usage-areas-us/

[7] P. R. A. P. C. R. Brooke Auxier, L. Rainie, M. Anderson, A. Perrin, M. Kumar, and E. Turner, 2019.

[8] R. Global, “Consumer attitudes towards brands’ usage of artificial intelligence (AI) and machine learning (ML) in the United States as of January 2023 [Graph,” in Statista, 2023. [Online]. Available: https://www-statista-com.libproxy.mit.edu/statistics/1364963/consumer-attitudes-ai-ml-brand-usage-us/

[9] M. Protocol, “How concerned are you about your online privacy compared to one year ago? [Graph,” in Statista, 2021. [Online]. Available: https://www-statista-com.libproxy.mit.edu/statistics/1228234/online-privacy-concerns-us/

[10] P. India, “To what extent has your organization mitigated the cybersecurity risks associated with each of the following in the last 12 months? [Graph,” in Statista, 2022. [Online]. Available: https://www-statista-com.libproxy.mit.edu/statistics/1318465/global-companies-actions-to-minimize-cyber-risks/

[11] “Equifax Says Cyberattack May Have Affected 143 Million in the US,” - N. Y. Times Sept., vol. 7, p. 2017.

[12] “Cisco 2019 Consumer Privacy Survey (Report.” [Online]. Available: https://www.cisco.com/c/dam/global/en_uk/products/collateral/security/cybersecurity-series-2019-cps.pdf.

[13] A. Acquisti and J. Grossklags, “What can behavioral economics teach us about privacy?,” J. Consum. Res., vol. 39, no. 3, pp. 513–518, 2013, doi: 10.1086/666547.

[14] U. Khan and Z. Rahman, “The impact of personalization on online customer behavior,” J. Mark., vol. 83, no. 1, pp. 98–116, 2019, doi: 10.1177/0022242918809183.

[15] M. J. Culnan and R. J. Bies, “Consumer privacy: Balancing economic and justice considerations,” J. Soc. Issues, vol. 59, no. 2, pp. 323–342, 2003, doi: 10.1111/1540-4560.00068.

[16] G. R. Milne and M. J. Culnan, “Using the privacy management framework to evaluate electronic commerce privacy policies,” J. Comput.-Mediat. Commun., vol. 8, no. 3, pp. 1–24, 2002, doi: 10.1111/j.1083-6101.2002.tb00122.x.

[17] A. Acquisti and J. Grossklags, “Privacy and rationality in individual decision making,” IEEE Secur. Priv., vol. 3, no. 1, pp. 26–33, 2005, doi: 10.1109/MSP.2005.8.

[18] Y. Wang and H. H. Emurian, “An overview of online trust: Concepts, elements, and implications,” Comput. Hum. Behav., vol. 21, no. 1, pp. 105–125, 2005, doi: 10.1016/j.chb.2004.10.012.

[19] X. Li, J. Liang, and Y. Wang, “Effects of personalization, privacy concern and social influence on the effectiveness of mobile advertising: A field experiment,” Internet Res., vol. 23, no. 5, pp. 562–583, 2013, doi: 10.1108/IntR-02-2012-0033.

[20] MediaPro, “How familiar are you with the following privacy regulations and guidelines?” Apr. 30, 2019. [Online]. Available: https://www-statista-com.libproxy.mit.edu/statistics/1050291/us-adults-knowledge-on-privacy-regulations-and-guidelines/

[21] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, “Federated Multi-Task Learning,” in Advances in Neural Information Processing Systems 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., 2017, pp. 4424–4434. [Online]. Available: http://papers.nips.cc/paper/7029-federated-multi-task-learning.pdf

[22] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, 2018.

[23] R. C. Geyer, T. Klein, and M. Nabi, “Differentially Private Federated Learning: A Client Level Perspective,” CoRR Abs171207557, 2017, [Online]. Available: http://arxiv.org/abs/1712.07557

[24] K. Bonawitz et al., “Practical Secure Aggregation for Privacy-Preserving Machine Learning,” in Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS ’17, New York, NY, USA: ACM, 2017. doi: 10.1145/3133956.3133982.

[25] F. Chen, Z. Dong, Z. Li, and X. He, “Federated Meta-Learning for Recommendation,” CoRR Abs180207876, 2018, [Online]. Available: http://arxiv.org/abs/1802.07876

[26] “Federated Machine Learning: Concept and Applications - arXiv Vanity.” [Online]. Available: https://www.arxiv-vanity.com/papers/1902.04885/

[27] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. Aguera, and Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Artificial Intelligence and Statistics, 2017, pp. 1273–1282.

[28] D. Cynthia, “Differential privacy,” in Proceedings of the International Colloquium on Automata, Languages, and Programming, Springer, 2006, pp. 1–12.

[29] Apple, “Differential privacy overview.” 2021. [Online]. Available: https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

[30] C. Dwork, “Differential privacy,” in Automata, languages and programming, Berlin, Heidelberg: Springer, 2011, pp. 1–12. doi: 10.1007/978-3-642-22006-7_1.

[31] M. Abadi et al., “Deep learning with Differential Privacy,” in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, ACM, 2016, pp. 308–318. doi: 10.1145/2976749.2978318.

[32] N. I. Standards and Technology, “Privacy risk management for federal information systems.” 2018. doi: 10.6028/NIST.SP.800-37r2.

[33] I. Goodfellow and O. Vinyals, “Qualitatively characterizing neural network optimization problems.” 2016. [Online]. Available: https://arxiv.org/abs/1412.6544

[34] “What Are Privacy-Enhancing Technologies (PETs) and How Will.” Aug. 11, 2021. [Online]. Available: https://about.fb.com/news/2021/08/privacy-enhancing-technologies-and-ads

[35] Y. Wang and X. Wu, “A survey on Differential Privacy,” J. Big Data, vol. 4, no. 1, pp. 1–30, 2017.

[36] C. Dwork, “Differential privacy: A survey of results,” in International Conference on Theory and Applications of Models of Computation, 2008, pp. 1–19.

[37] Ú. Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, K. Talwar, and A. Thakurta, "Amplification by shuffling: From local to central Differential Privacy via anonymity," J. ACM, vol. 66, no. 1, pp. 1–38, 2019.

[38] C. Dwork and A. Roth, "The algorithmic foundations of Differential Privacy," Found. Trends Theor. Comput. Sci., vol. 9, no. 3–4, pp. 211–407, 2014.

[39] K. Talwar, "The exponential mechanism: A tutorial." 2014.

[40] F. Bélanger and R. E. Crossler, "Privacy in the digital age: A review of information privacy research in information systems," MIS Q., vol. 35, no. 4, pp. 1017–1041, 2011.

[41] H. Xu, T. Dinev, H. J. Smith, and P. Hart, "Examining the formation of individual's privacy concerns: Toward an integrative view," MIS Q., vol. 35, no. 1, pp. 1–19, 2011.

[42] Y. Li, X. Li, X. Liang, and J. Huang, "A customer-centric approach to privacy management in IoT systems," IEEE Internet Things J., vol. 7, no. 2, pp. 1138–1148, 2020.

[43] G. R. Milne and M. J. Culnan, "Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices," J. Interact. Mark., vol. 18, no. 3, pp. 15–29, 2004.

[44] C. Huang and K. L. Mak, "Using design structure matrix to adopt customer-centric approach in enterprise operations," J. Ind. Prod. Eng., vol. 33, no. 5, pp. 282–291, 2016.

[45] E. F. Crawley, Design and Analysis of Large-Scale Engineering Systems. MIT OpenCourseWare, 2016. [Online]. Available: https://ocw.mit.edu/courses/aeronautics-and-astronautics/16-842-fundamentals-of-systems-engineering-fall-2015/lecture-notes/MIT16_842F15_lec06.pdf

[46] S. D. Eppinger, "Assessing design structurability: methods and implications," Manag. Sci., vol. 37, no. 6, pp. 726–743, 1991.

[47] T. R. Browning, S. D. Eppinger, and C. M. Rowles, "A design structure matrix-based methodology for managing the design of complex systems," Decis. Sci., vol. 36, no. 3, pp. 337–368, 2005.

[48] S. D. Eppinger and T. R. Browning, Design structure matrix methods and applications. MIT Press, 2012.

[49] "General Data Protection Regulation (GDPR) – Official Legal Text." [Online]. Available: https://gdpr-info.eu/.

[50] "California Consumer Privacy Act (CCPA." Dec. 08, 2021. [Online]. Available: https://oag.ca.gov/privacy/ccpa.

[51] "PIPEDA." [Online]. Available: https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/.

[52] -, “Planalto.” Aug. 14, 2018. [Online]. Available: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

[53] “Understanding Enterprises as Learning Systems.” Jan. 15, 1995. [Online]. Available: https://sloanreview.mit.edu/article/understanding-enterprises-as-learning-systems/.

[54] “ACM Trans. Intell. Syst. Technol., Vol. 10, No. 2, Article 12. Publication date: February 2019.”

Appendices

Appendix A: Zoomed System decomposition

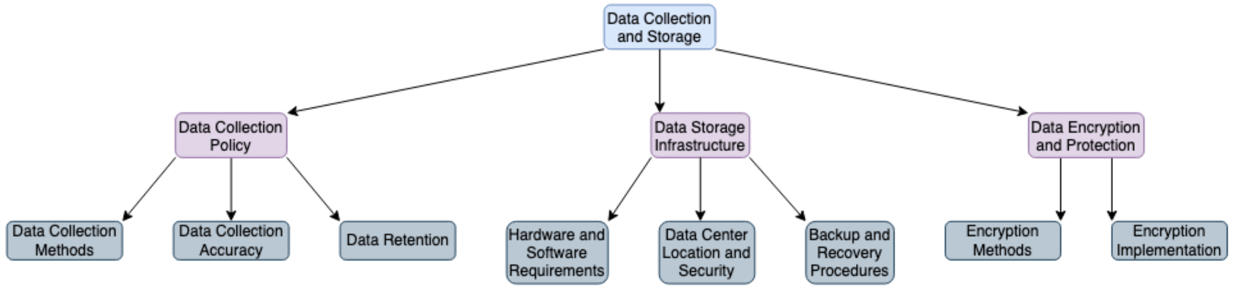

Data Collection and Storage
Data Collection Policy
Data Storage Infrastructure
Data Encryption and Protection
Data Collection Methods
Data Collection Accuracy
Data Retention
Hardware and Software Requirements
Data Center Location and Security
Backup and Recovery Procedures
Encryption Methods
Encryption Implementation

Figure 21. System decomposition (a)

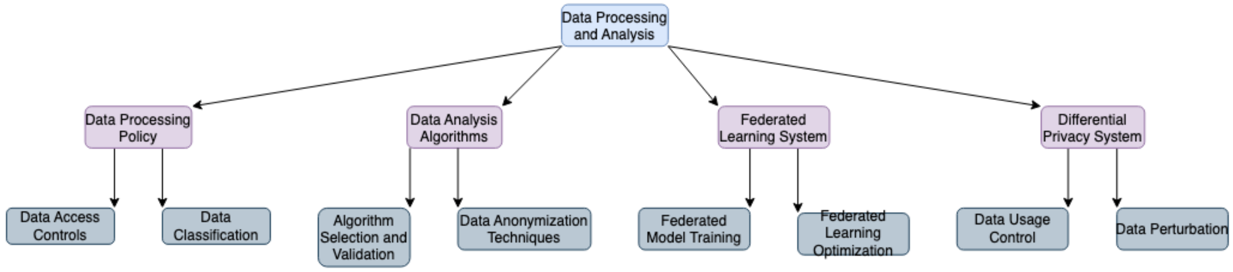

Data Processing and Analysis
Data Processing Policy
Data Analysis Algorithms
Federated Learning System
Differential Privacy System
Data Access Controls
Data Classification
Algorithm Selection and Validation
Data Anonymization Techniques
Federated Model Training
Federated Learning Optimization
Data Usage Control
Data Perturbation

Figure 22. System decomposition (b)

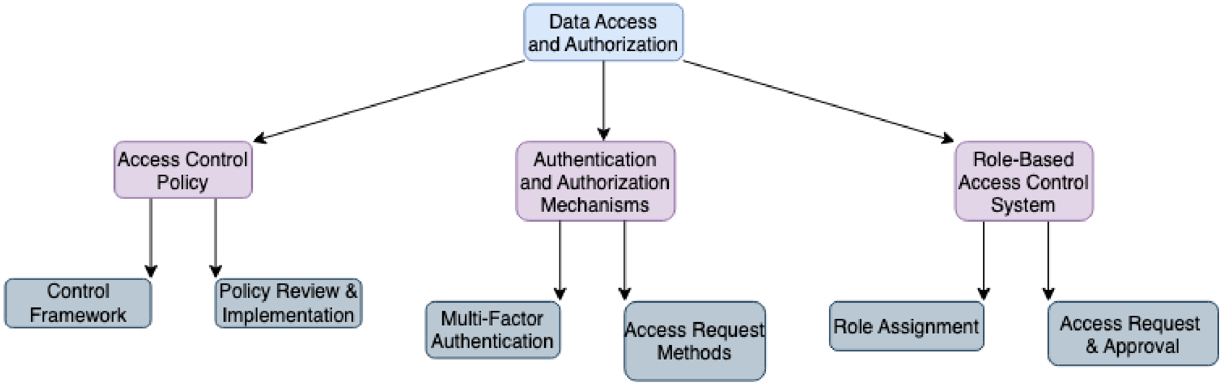

Data Access and Authorization
Access Control Policy
Authentication and Authorization Mechanisms
Role-Based Access Control System
Control Framework
Policy Review & Implementation
Multi-Factor Authentication
Access Request Methods
Role Assignment
Access Request & Approval

Figure 23. System decomposition (c)

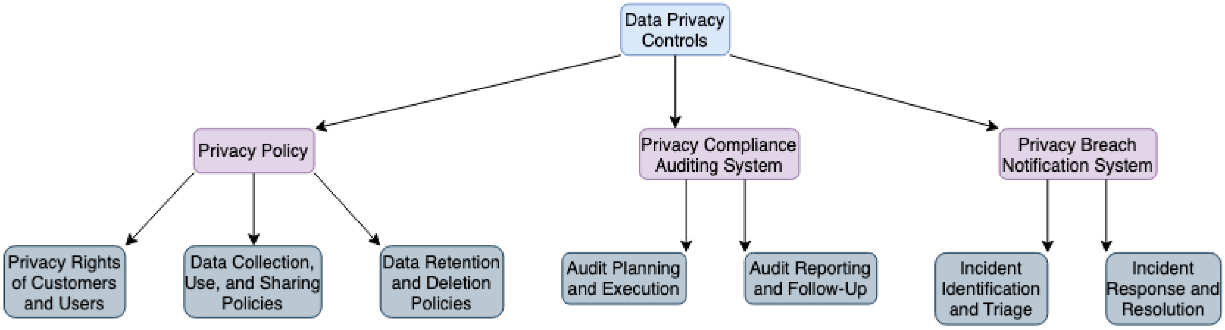

Data Privacy Controls
Privacy Policy
Privacy Compliance Auditing System
Privacy Breach Notification System
Privacy Rights of Customers and Users
Data Collection, Use, and Sharing Policies
Data Retention and Deletion Policies
Audit Planning and Execution
Audit Reporting and Follow-Up
Incident Identification and Triage
Incident Response and Resolution

Figure 24. System decomposition (d)

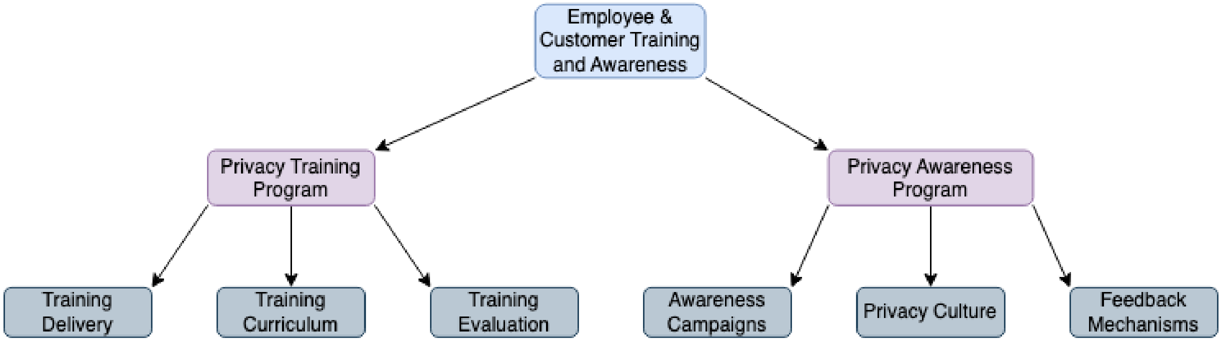
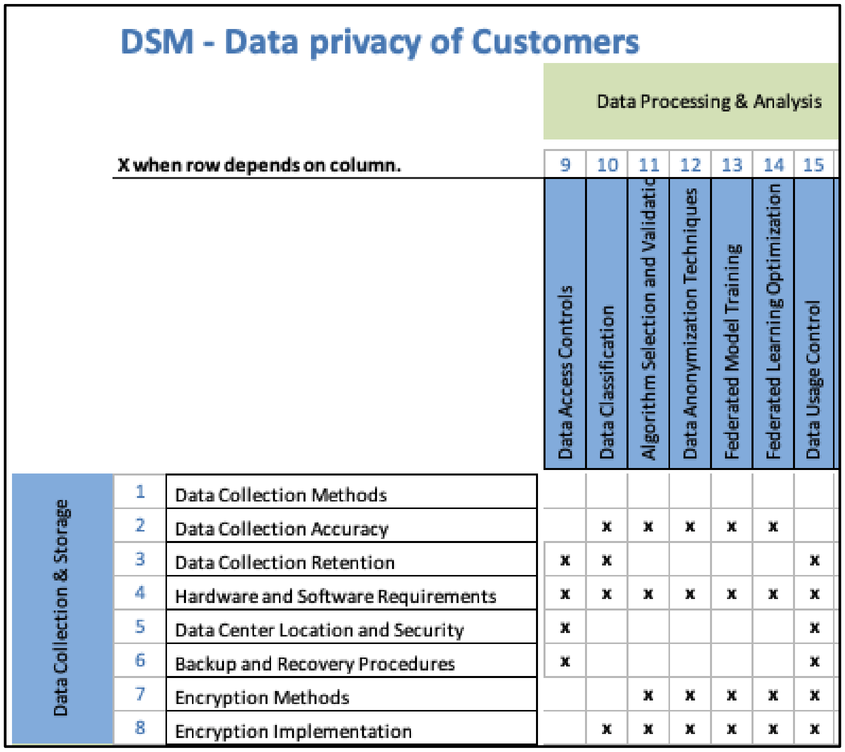


Figure 25. System decomposition (e)

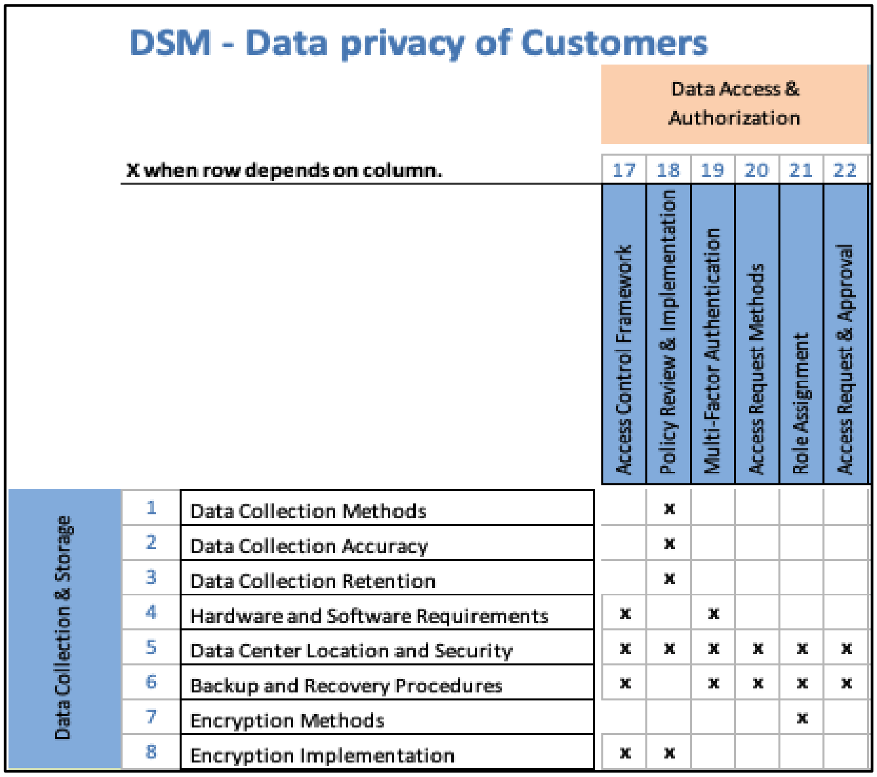
Appendix B: Zoomed DSM



DSM - Data privacy of Customers

			Data Processing & Analysis						
		X when row depends on column.	9	10	11	12	13	14	15
			Data Access Controls	Data Classification	Algorithm Selection and Validatic	Data Anonymization Techniques	Federated Model Training	Federated Learning Optimization	Data Usage Control
Data Collection & Storage	1	Data Collection Methods							
	2	Data Collection Accuracy		x	x	x	x	x	
	3	Data Collection Retention	x	x					x
	4	Hardware and Software Requirements	x	x	x	x	x	x	x
	5	Data Center Location and Security	x						x
	6	Backup and Recovery Procedures	x						x
	7	Encryption Methods			x	x	x	x	x
	8	Encryption Implementation		x	x	x	x	x	x

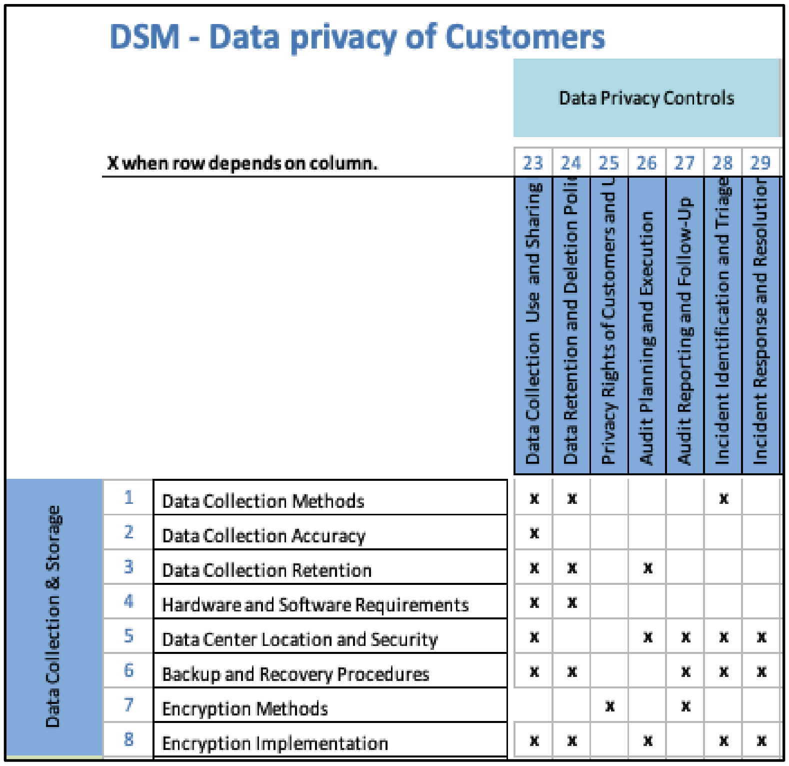
Figure 26. Dependencies of Data Collection & Storage on Data Processing & Analysis



DSM - Data privacy of Customers

			Data Access & Authorization					
		X when row depends on column.	17	18	19	20	21	22
			Access Control Framework	Policy Review & Implementation	Multi-Factor Authentication	Access Request Methods	Role Assignment	Access Request & Approval
Data Collection & Storage	1	Data Collection Methods		x				
	2	Data Collection Accuracy		x				
	3	Data Collection Retention		x				
	4	Hardware and Software Requirements	x		x			
	5	Data Center Location and Security	x	x	x	x	x	x
	6	Backup and Recovery Procedures	x		x	x	x	x
	7	Encryption Methods					x	
	8	Encryption Implementation	x	x				

Figure 27. Dependencies of Data Collection & Storage on Data Access & Authorization

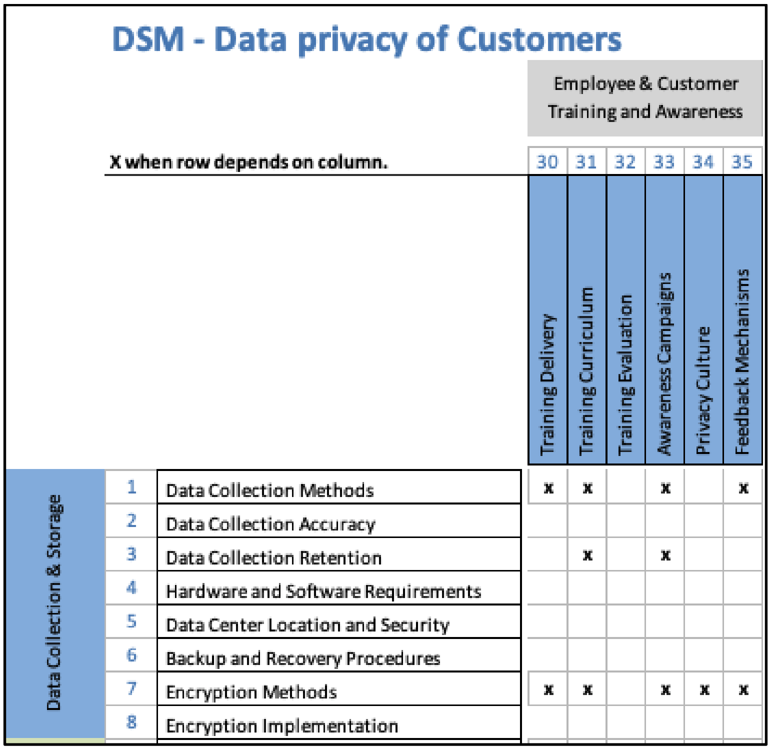


DSM - Data privacy of Customers

X when row depends on column.

			Data Privacy Controls						
			23	24	25	26	27	28	29
			Data Collection Use and Sharing	Data Retention and Deletion Poli	Privacy Rights of Customers and U	Audit Planning and Execution	Audit Reporting and Follow-Up	Incident Identification and Triage	Incident Response and Resolution
Data Collection & Storage	1	Data Collection Methods	x	x				x	
	2	Data Collection Accuracy	x						
	3	Data Collection Retention	x	x		x			
	4	Hardware and Software Requirements	x	x					
	5	Data Center Location and Security	x			x	x	x	x
	6	Backup and Recovery Procedures	x	x			x	x	x
	7	Encryption Methods			x		x		
	8	Encryption Implementation	x	x		x		x	x

Figure 28. Dependencies of Data Collection & Storage on Data Privacy Controls

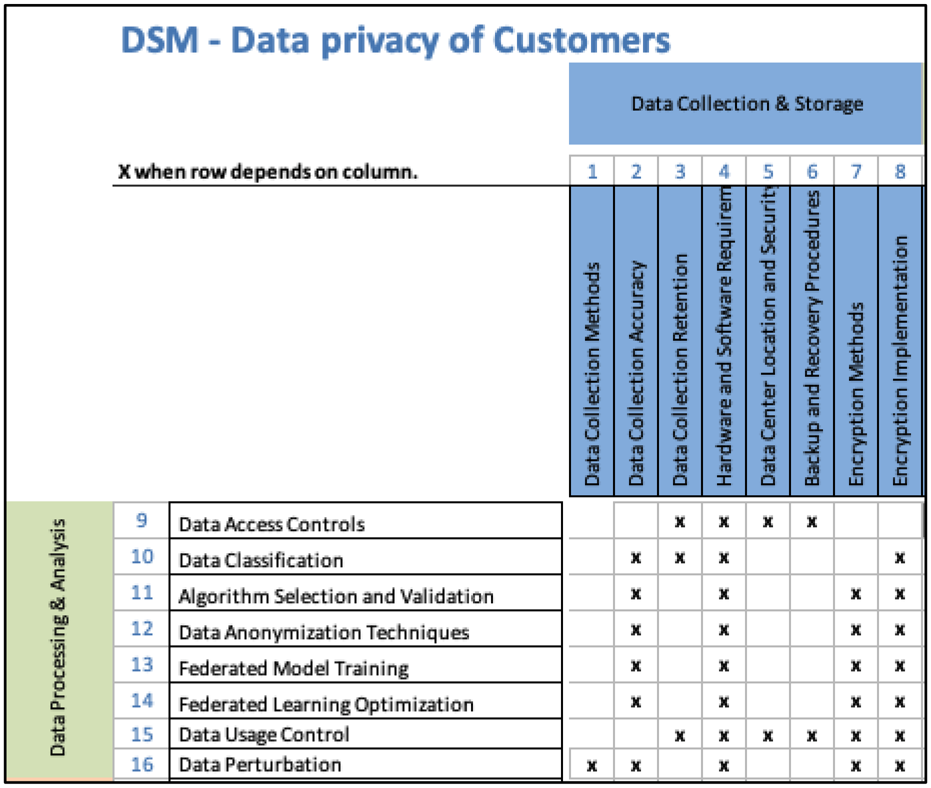


DSM - Data privacy of Customers

X when row depends on column.

			Employee & Customer Training and Awareness					
			30	31	32	33	34	35
			Training Delivery	Training Curriculum	Training Evaluation	Awareness Campaigns	Privacy Culture	Feedback Mechanisms
Data Collection & Storage	1	Data Collection Methods	x	x		x		x
	2	Data Collection Accuracy						
	3	Data Collection Retention		x		x		
	4	Hardware and Software Requirements						
	5	Data Center Location and Security						
	6	Backup and Recovery Procedures						
	7	Encryption Methods	x	x		x	x	x
	8	Encryption Implementation						

Figure 29. Dependencies of Data Collection & Storage on Employee & Customer Training and Awareness

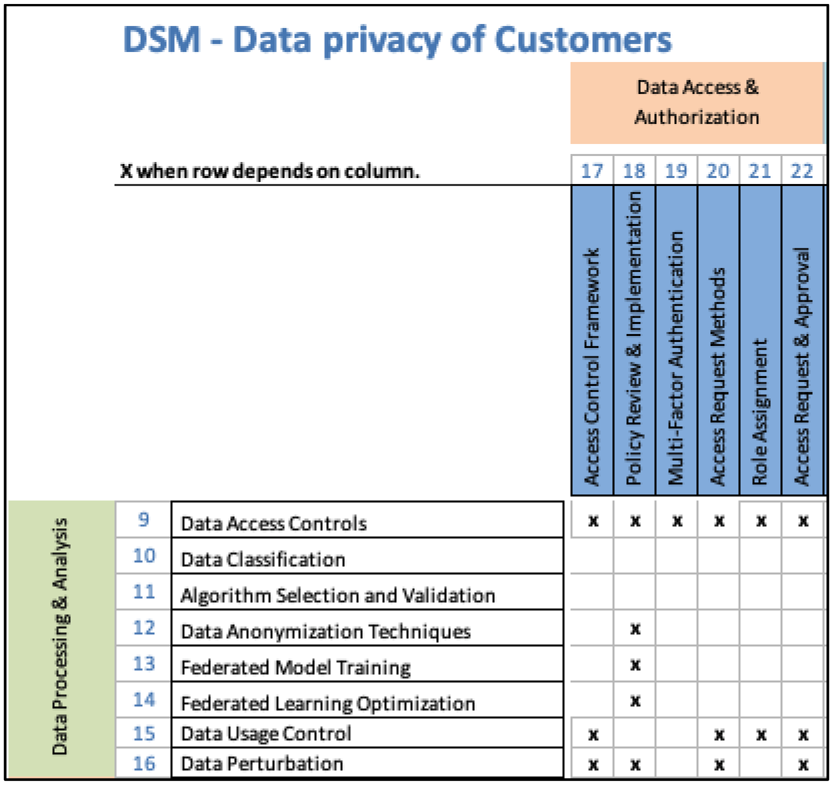


DSM - Data privacy of Customers

X when row depends on column.

			Data Collection & Storage							
			1	2	3	4	5	6	7	8
			Data Collection Methods	Data Collection Accuracy	Data Collection Retention	Hardware and Software Requirem	Data Center Location and Securit	Backup and Recovery Procedures	Encryption Methods	Encryption Implementation
Data Processing & Analysis	9	Data Access Controls			x	x	x	x		
	10	Data Classification		x	x	x				x
	11	Algorithm Selection and Validation		x		x			x	x
	12	Data Anonymization Techniques		x		x			x	x
	13	Federated Model Training		x		x			x	x
	14	Federated Learning Optimization		x		x			x	x
	15	Data Usage Control			x	x	x	x	x	x
	16	Data Perturbation	x	x		x			x	x

Figure 30. Dependencies of Data Processing & Analysis on Data Collection & Storage

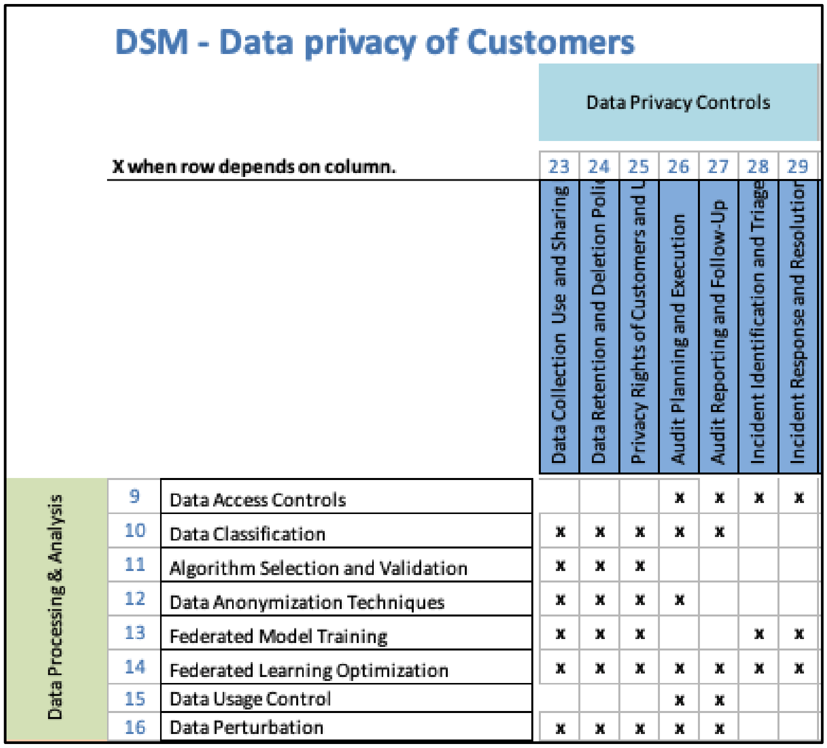


DSM - Data privacy of Customers

X when row depends on column.

			Data Access & Authorization					
			17	18	19	20	21	22
			Access Control Framework	Policy Review & Implementation	Multi-Factor Authentication	Access Request Methods	Role Assignment	Access Request & Approval
Data Processing & Analysis	9	Data Access Controls	x	x	x	x	x	x
	10	Data Classification						
	11	Algorithm Selection and Validation						
	12	Data Anonymization Techniques		x				
	13	Federated Model Training		x				
	14	Federated Learning Optimization		x				
	15	Data Usage Control	x			x	x	x
	16	Data Perturbation	x	x		x		x

Figure 31. Dependencies of Data Processing & Analysis on Data Access & Authorization

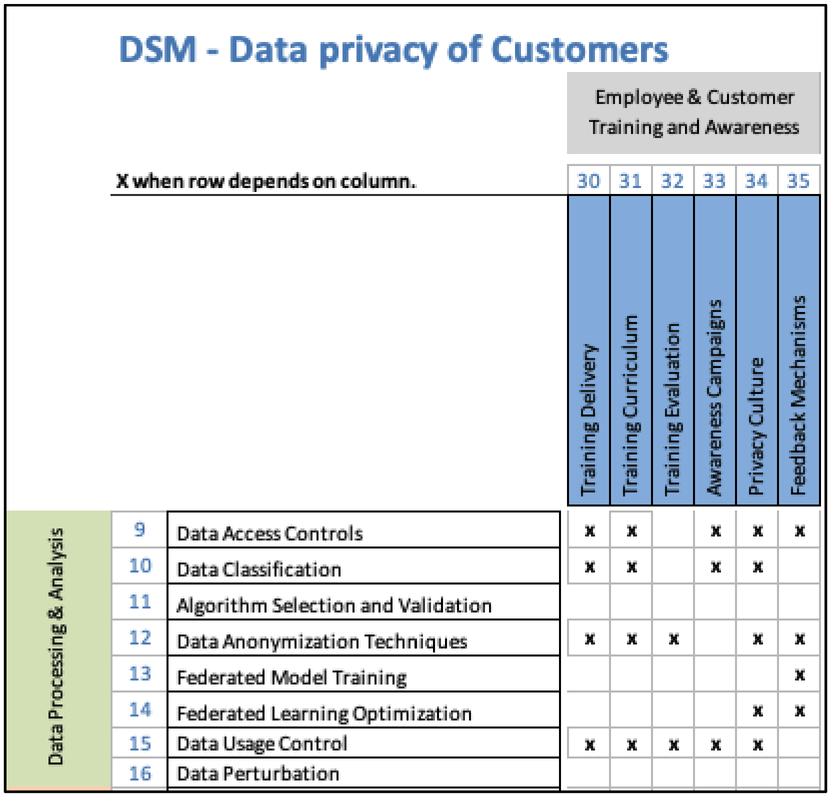


DSM - Data privacy of Customers

X when row depends on column.

			Data Privacy Controls						
			23	24	25	26	27	28	29
			Data Collection Use and Sharing	Data Retention and Deletion Poli	Privacy Rights of Customers and U	Audit Planning and Execution	Audit Reporting and Follow-Up	Incident Identification and Triage	Incident Response and Resolution
Data Processing & Analysis	9	Data Access Controls				x	x	x	x
	10	Data Classification	x	x	x	x	x		
	11	Algorithm Selection and Validation	x	x	x				
	12	Data Anonymization Techniques	x	x	x	x			
	13	Federated Model Training	x	x	x			x	x
	14	Federated Learning Optimization	x	x	x	x	x	x	x
	15	Data Usage Control				x	x		
	16	Data Perturbation	x	x	x	x	x		

Figure 32. Dependencies of Data Processing & Analysis on Data Privacy Controls

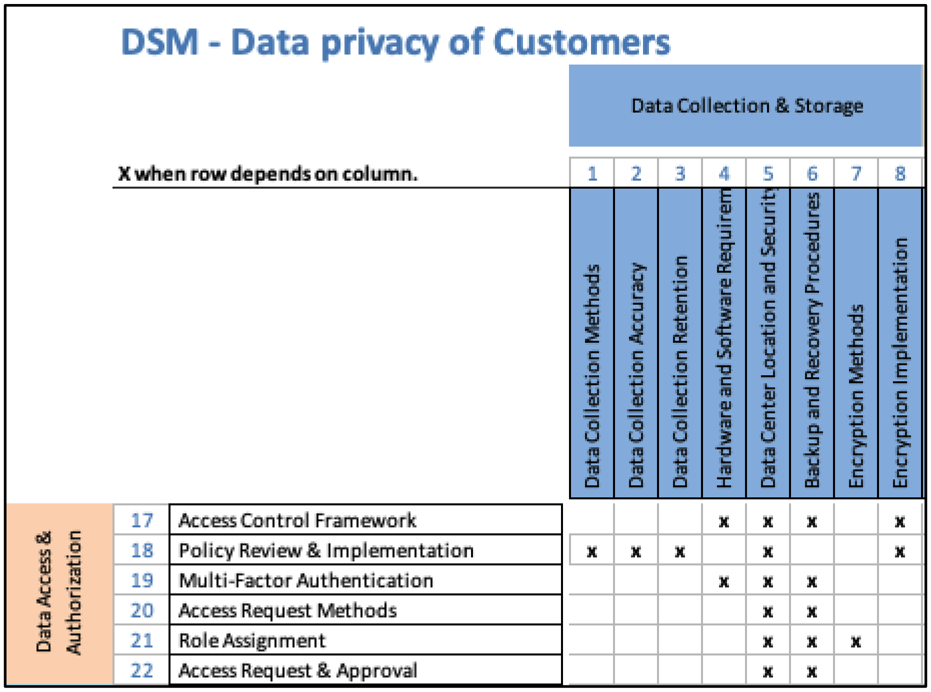


DSM - Data privacy of Customers

X when row depends on column.

			Employee & Customer Training and Awareness					
			30	31	32	33	34	35
			Training Delivery	Training Curriculum	Training Evaluation	Awareness Campaigns	Privacy Culture	Feedback Mechanisms
Data Processing & Analysis	9	Data Access Controls	x	x		x	x	x
	10	Data Classification	x	x		x	x	
	11	Algorithm Selection and Validation						
	12	Data Anonymization Techniques	x	x	x		x	x
	13	Federated Model Training						x
	14	Federated Learning Optimization					x	x
	15	Data Usage Control	x	x	x	x	x	
	16	Data Perturbation						

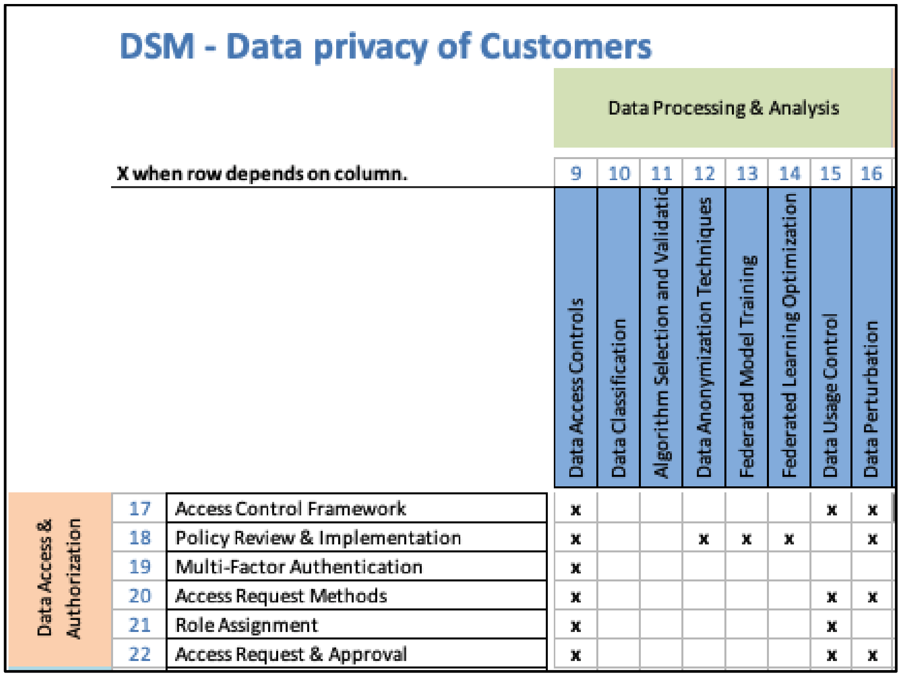
Figure 33. Dependencies of Data Processing & Analysis on Employee & Customer Training and Awareness



DSM - Data privacy of Customers

			Data Collection & Storage							
X when row depends on column.			1	2	3	4	5	6	7	8
			Data Collection Methods	Data Collection Accuracy	Data Collection Retention	Hardware and Software Requirem	Data Center Location and Securit	Backup and Recovery Procedures	Encryption Methods	Encryption Implementation
Data Access & Authorization	17	Access Control Framework				x	x	x		x
	18	Policy Review & Implementation	x	x	x		x			x
	19	Multi-Factor Authentication				x	x	x		
	20	Access Request Methods					x	x		
	21	Role Assignment					x	x	x	
	22	Access Request & Approval					x	x		

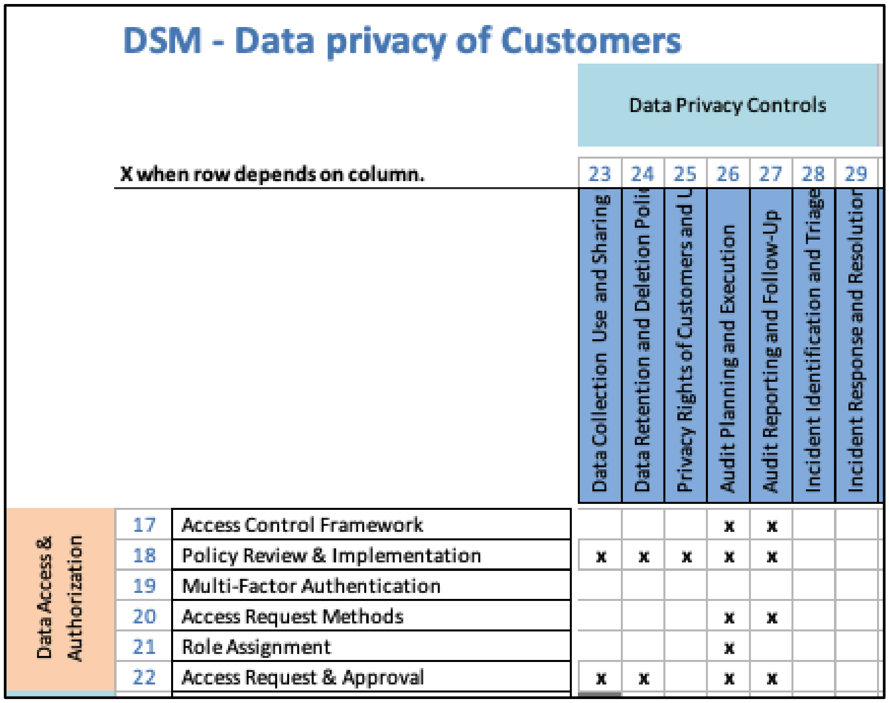
Figure 34. Dependencies of Data Access & Authorization on Data Collection & Storage



DSM - Data privacy of Customers

			Data Processing & Analysis							
X when row depends on column.			9	10	11	12	13	14	15	16
			Data Access Controls	Data Classification	Algorithm Selection and Validatic	Data Anonymization Techniques	Federated Model Training	Federated Learning Optimization	Data Usage Control	Data Perturbation
Data Access & Authorization	17	Access Control Framework	x						x	x
	18	Policy Review & Implementation	x			x	x	x		x
	19	Multi-Factor Authentication	x							
	20	Access Request Methods	x						x	x
	21	Role Assignment	x						x	
	22	Access Request & Approval	x						x	x

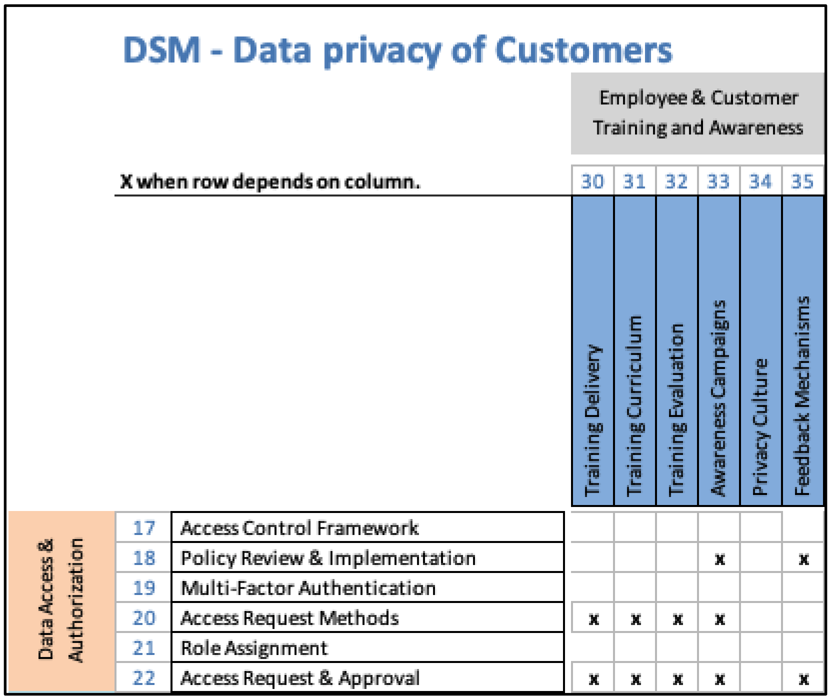
Figure 35. Dependencies of Data Access & Authorization on Data Processing & Analysis



DSM - Data privacy of Customers

			Data Privacy Controls						
		X when row depends on column.	23	24	25	26	27	28	29
			Data Collection Use and Sharing	Data Retention and Deletion Poli	Privacy Rights of Customers and U	Audit Planning and Execution	Audit Reporting and Follow-Up	Incident Identification and Triage	Incident Response and Resolution
Data Access & Authorization	17	Access Control Framework				x	x		
	18	Policy Review & Implementation	x	x	x	x	x		
	19	Multi-Factor Authentication							
	20	Access Request Methods				x	x		
	21	Role Assignment				x			
	22	Access Request & Approval	x	x		x	x		

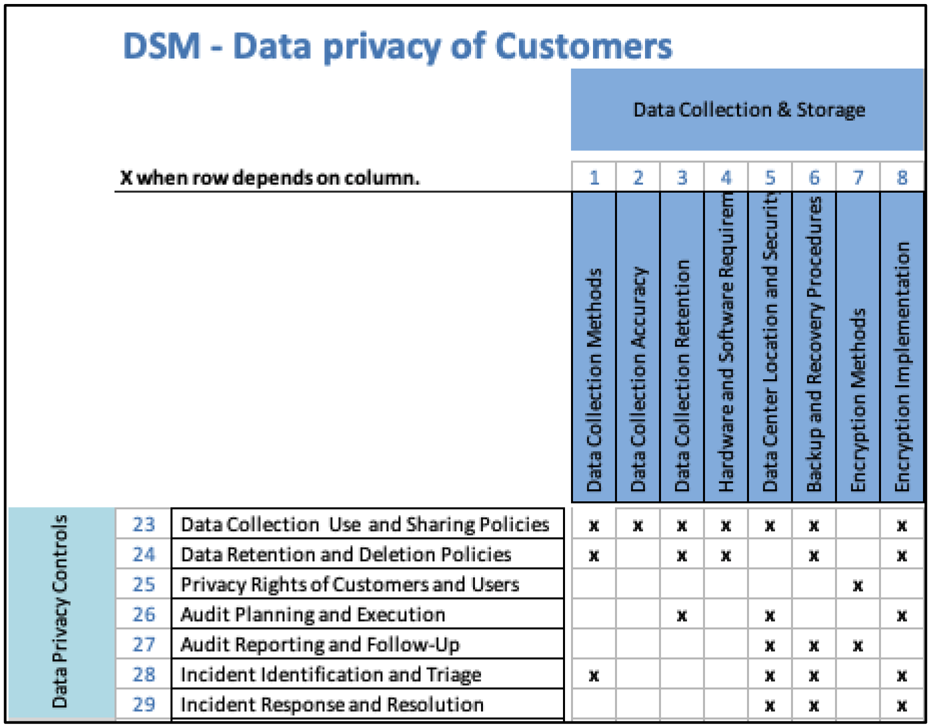
Figure 36. Dependencies of Data Access & Authorization on Data Privacy Controls



DSM - Data privacy of Customers

			Employee & Customer Training and Awareness					
		X when row depends on column.	30	31	32	33	34	35
			Training Delivery	Training Curriculum	Training Evaluation	Awareness Campaigns	Privacy Culture	Feedback Mechanisms
Data Access & Authorization	17	Access Control Framework						
	18	Policy Review & Implementation				x		x
	19	Multi-Factor Authentication						
	20	Access Request Methods	x	x	x	x		
	21	Role Assignment						
	22	Access Request & Approval	x	x	x	x		x

Figure 37. Dependencies of Data Access & Authorization on Employee & Customer Training and Awareness

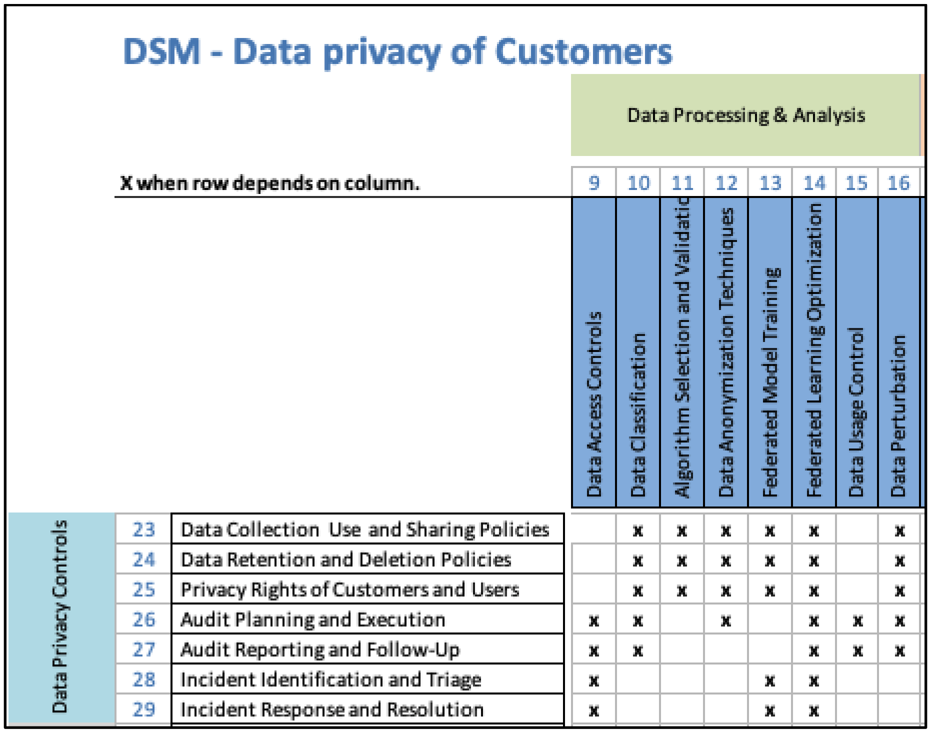


DSM - Data privacy of Customers

X when row depends on column.

			Data Collection & Storage							
			1	2	3	4	5	6	7	8
			Data Collection Methods	Data Collection Accuracy	Data Collection Retention	Hardware and Software Requirem	Data Center Location and Securit	Backup and Recovery Procedures	Encryption Methods	Encryption Implementation
Data Privacy Controls	23	Data Collection Use and Sharing Policies	x	x	x	x	x	x		x
	24	Data Retention and Deletion Policies	x		x	x		x		x
	25	Privacy Rights of Customers and Users							x	
	26	Audit Planning and Execution			x		x			x
	27	Audit Reporting and Follow-Up					x	x	x	
	28	Incident Identification and Triage	x				x	x		x
	29	Incident Response and Resolution					x	x		x

Figure 38. Dependencies of Data Privacy Controls on Data Collection & Storage

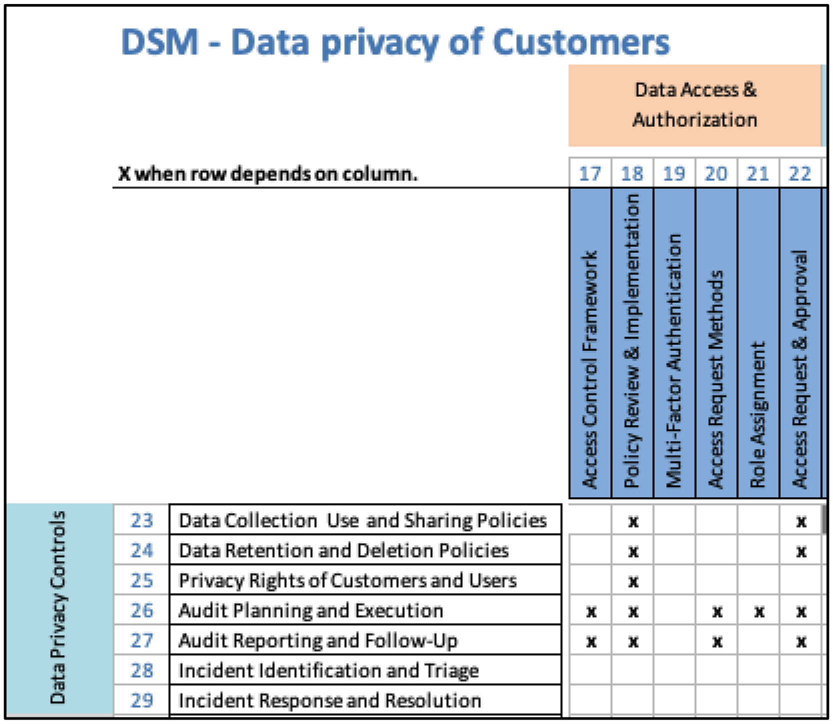


DSM - Data privacy of Customers

X when row depends on column.

			Data Processing & Analysis							
			9	10	11	12	13	14	15	16
			Data Access Controls	Data Classification	Algorithm Selection and Validatic	Data Anonymization Techniques	Federated Model Training	Federated Learning Optimization	Data Usage Control	Data Perturbation
Data Privacy Controls	23	Data Collection Use and Sharing Policies		x	x	x	x	x		x
	24	Data Retention and Deletion Policies		x	x	x	x	x		x
	25	Privacy Rights of Customers and Users		x	x	x	x	x		x
	26	Audit Planning and Execution	x	x		x		x	x	x
	27	Audit Reporting and Follow-Up	x	x				x	x	x
	28	Incident Identification and Triage	x				x	x		
	29	Incident Response and Resolution	x				x	x		

Figure 39. Dependencies of Data Privacy Controls on Data Processing & Analysis

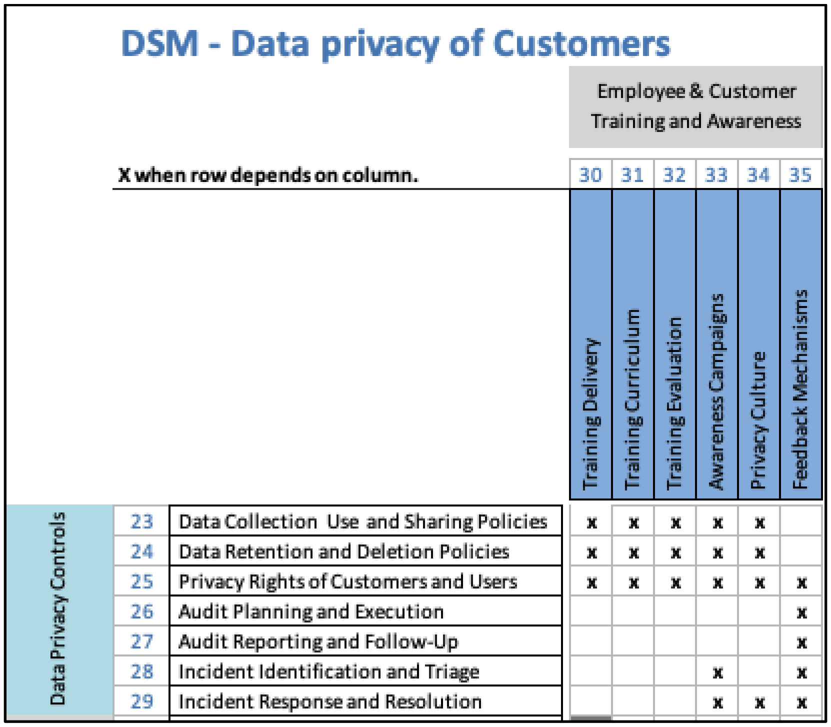


DSM - Data privacy of Customers

X when row depends on column.

			Data Access & Authorization					
			17	18	19	20	21	22
			Access Control Framework	Policy Review & Implementation	Multi-Factor Authentication	Access Request Methods	Role Assignment	Access Request & Approval
Data Privacy Controls	23	Data Collection Use and Sharing Policies		x				x
	24	Data Retention and Deletion Policies		x				x
	25	Privacy Rights of Customers and Users		x				
	26	Audit Planning and Execution	x	x		x	x	x
	27	Audit Reporting and Follow-Up	x	x		x		x
	28	Incident Identification and Triage						
	29	Incident Response and Resolution						

Figure 40. Dependencies of Data Privacy Controls on Data Access & Authorization

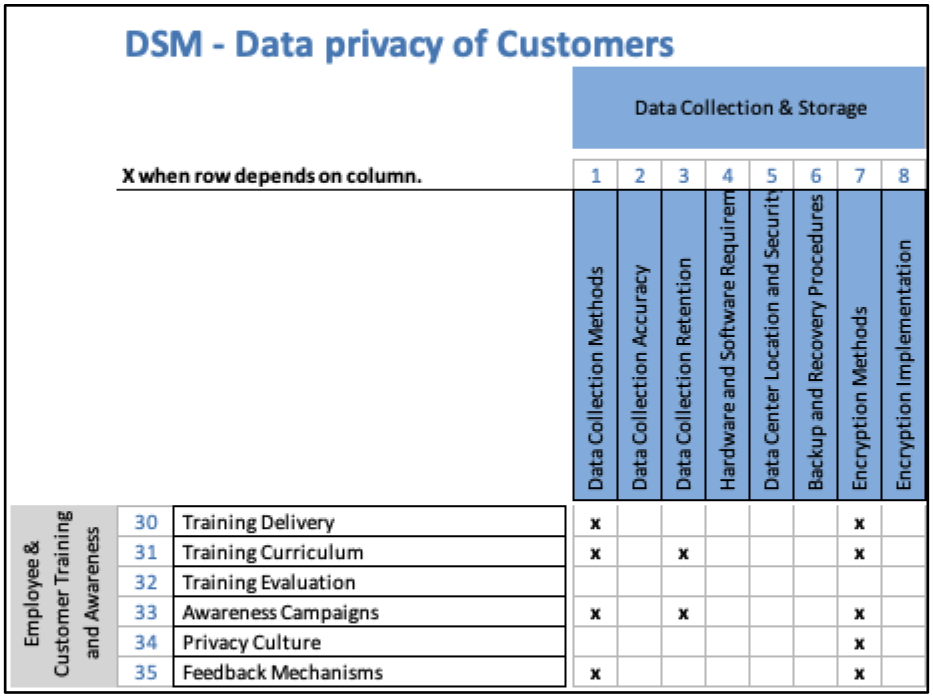


DSM - Data privacy of Customers

X when row depends on column.

			Employee & Customer Training and Awareness					
			30	31	32	33	34	35
			Training Delivery	Training Curriculum	Training Evaluation	Awareness Campaigns	Privacy Culture	Feedback Mechanisms
Data Privacy Controls	23	Data Collection Use and Sharing Policies	x	x	x	x	x	
	24	Data Retention and Deletion Policies	x	x	x	x	x	
	25	Privacy Rights of Customers and Users	x	x	x	x	x	x
	26	Audit Planning and Execution						x
	27	Audit Reporting and Follow-Up						x
	28	Incident Identification and Triage				x		x
	29	Incident Response and Resolution				x	x	x

Figure 41. Dependencies of Data Privacy Controls on Employee & Customer Training and Awareness

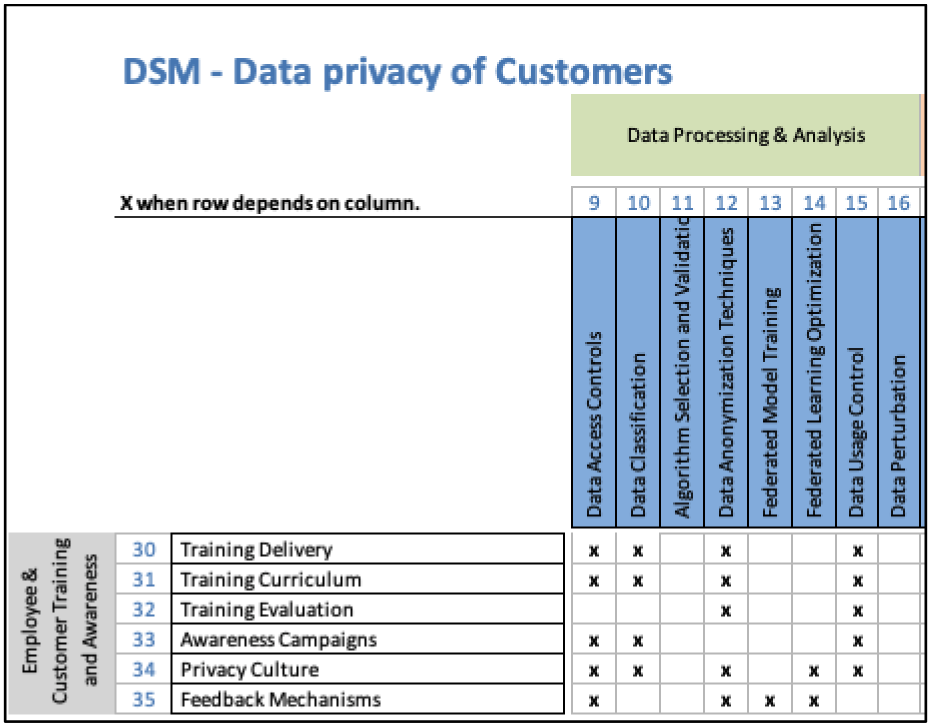


DSM - Data privacy of Customers

X when row depends on column.

			Data Collection & Storage							
			1	2	3	4	5	6	7	8
			Data Collection Methods	Data Collection Accuracy	Data Collection Retention	Hardware and Software Requirem	Data Center Location and Security	Backup and Recovery Procedures	Encryption Methods	Encryption Implementation
Employee & Customer Training and Awareness	30	Training Delivery	x						x	
	31	Training Curriculum	x		x				x	
	32	Training Evaluation								
	33	Awareness Campaigns	x		x				x	
	34	Privacy Culture							x	
	35	Feedback Mechanisms	x						x	

Figure 42. Dependencies of Employee & Customer Training and Awareness on Data Collection & Storage

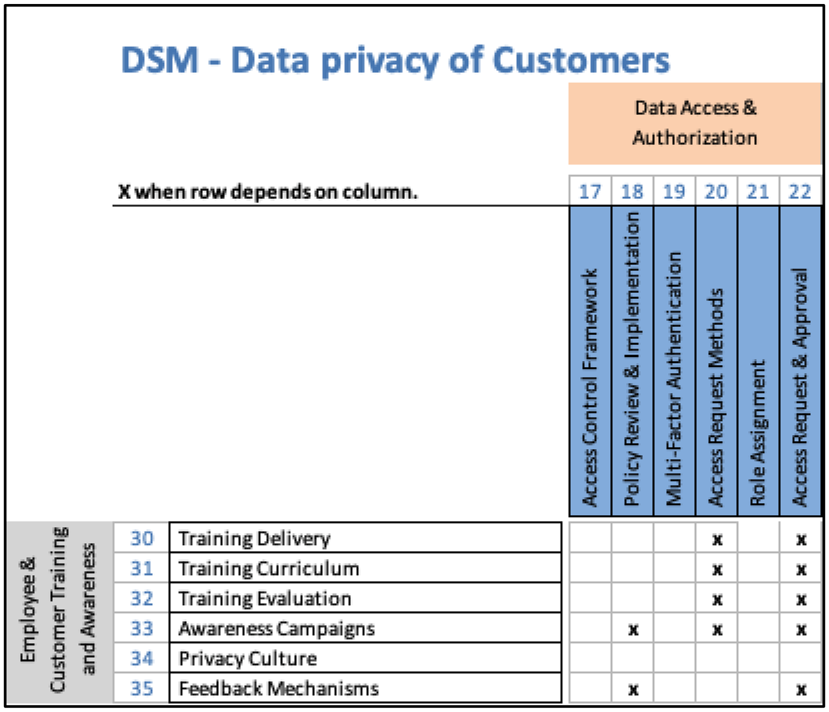


DSM - Data privacy of Customers

X when row depends on column.

			Data Processing & Analysis							
			9	10	11	12	13	14	15	16
			Data Access Controls	Data Classification	Algorithm Selection and Validatio	Data Anonymization Techniques	Federated Model Training	Federated Learning Optimization	Data Usage Control	Data Perturbation
Employee & Customer Training and Awareness	30	Training Delivery	x	x		x			x	
	31	Training Curriculum	x	x		x			x	
	32	Training Evaluation				x			x	
	33	Awareness Campaigns	x	x					x	
	34	Privacy Culture	x	x		x		x	x	
	35	Feedback Mechanisms	x			x	x	x		

Figure 43. Dependencies of Employee & Customer Training and Awareness on Data Processing & Analysis

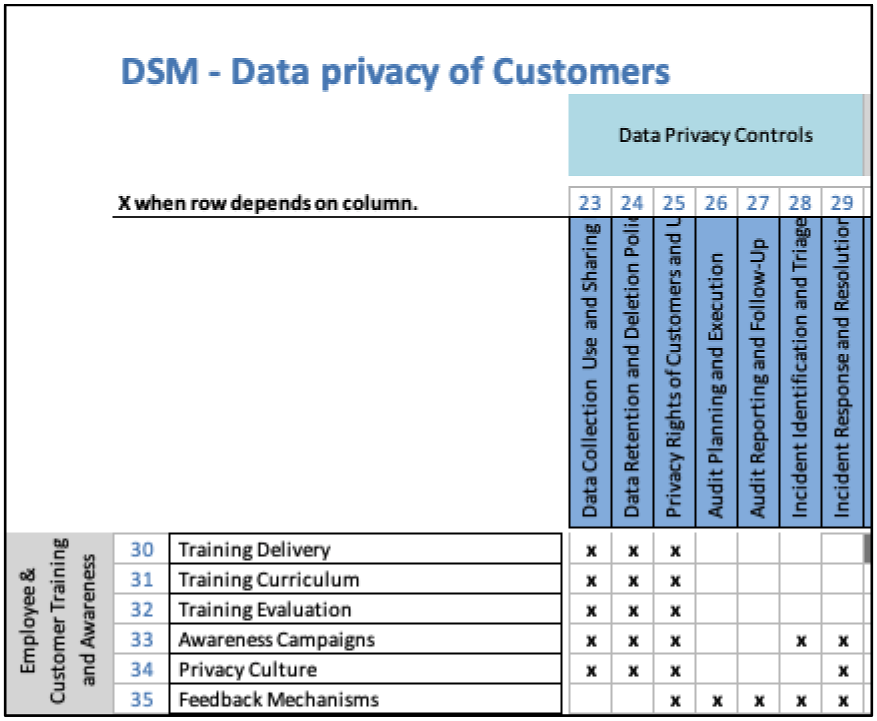


DSM - Data privacy of Customers

X when row depends on column.

			Data Access & Authorization					
			17	18	19	20	21	22
			Access Control Framework	Policy Review & Implementation	Multi-Factor Authentication	Access Request Methods	Role Assignment	Access Request & Approval
Employee & Customer Training and Awareness	30	Training Delivery				x		x
	31	Training Curriculum				x		x
	32	Training Evaluation				x		x
	33	Awareness Campaigns		x		x		x
	34	Privacy Culture						
	35	Feedback Mechanisms		x				x

Figure 44. Dependencies of Employee & Customer Training and Awareness on Data Access & Authorization



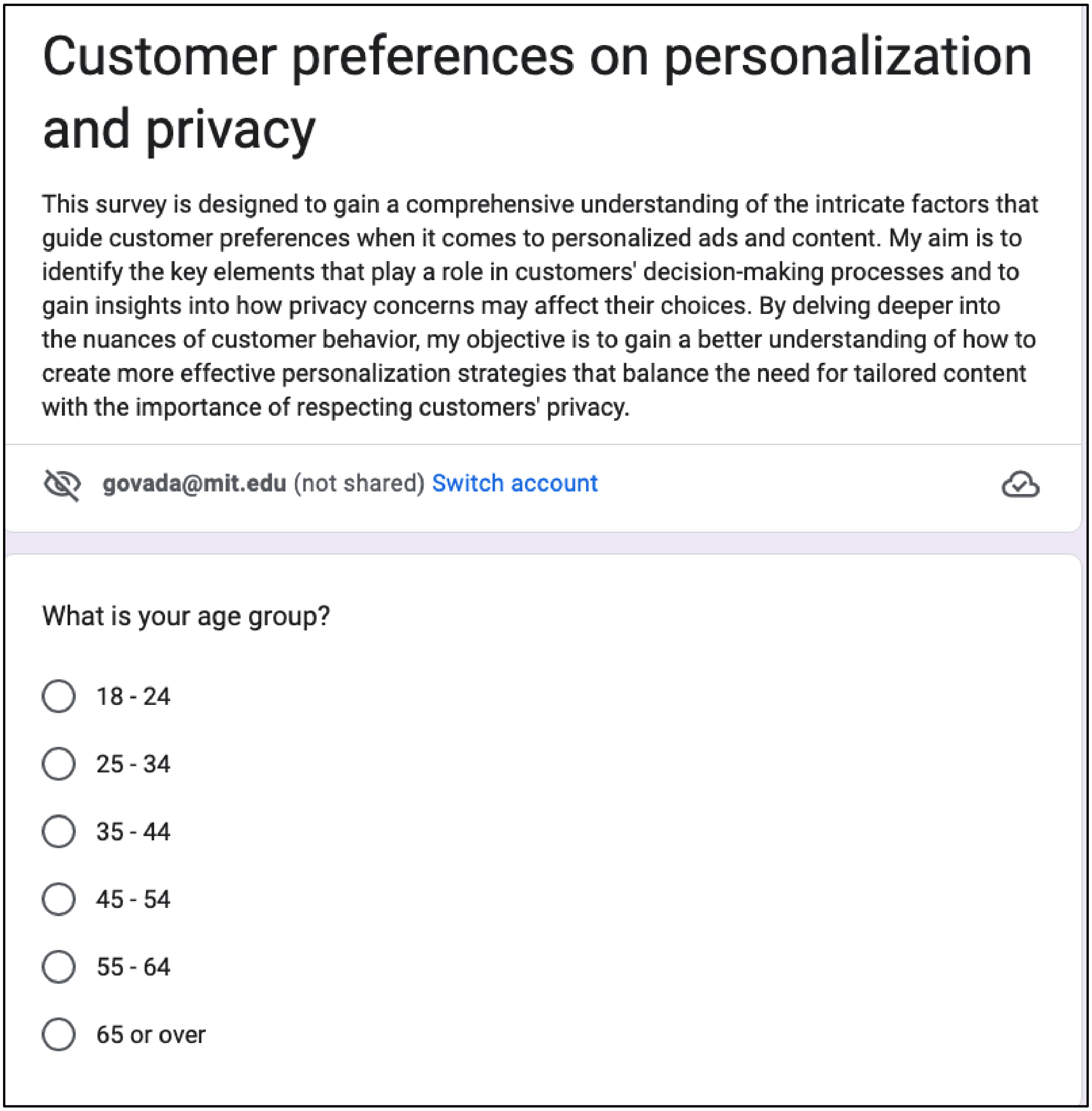
DSM - Data privacy of Customers

X when row depends on column.

			Data Privacy Controls						
			23	24	25	26	27	28	29
			Data Collection Use and Sharing	Data Retention and Deletion Poli	Privacy Rights of Customers and U	Audit Planning and Execution	Audit Reporting and Follow-Up	Incident Identification and Triage	Incident Response and Resolution
Employee & Customer Training and Awareness	30	Training Delivery	x	x	x				
	31	Training Curriculum	x	x	x				
	32	Training Evaluation	x	x	x				
	33	Awareness Campaigns	x	x	x			x	x
	34	Privacy Culture	x	x	x				x
	35	Feedback Mechanisms			x	x	x	x	x

Figure 45. Dependencies of Employee & Customer Training and Awareness on Data Privacy Controls

Appendix C: Customer preferences on personalization and privacy - Survey for interviews



Customer preferences on personalization and privacy

This survey is designed to gain a comprehensive understanding of the intricate factors that guide customer preferences when it comes to personalized ads and content. My aim is to identify the key elements that play a role in customers' decision-making processes and to gain insights into how privacy concerns may affect their choices. By delving deeper into the nuances of customer behavior, my objective is to gain a better understanding of how to create more effective personalization strategies that balance the need for tailored content with the importance of respecting customers' privacy.

govada@mit.edu (not shared) Switch account

What is your age group?

○ 18 - 24

○ 25 - 34

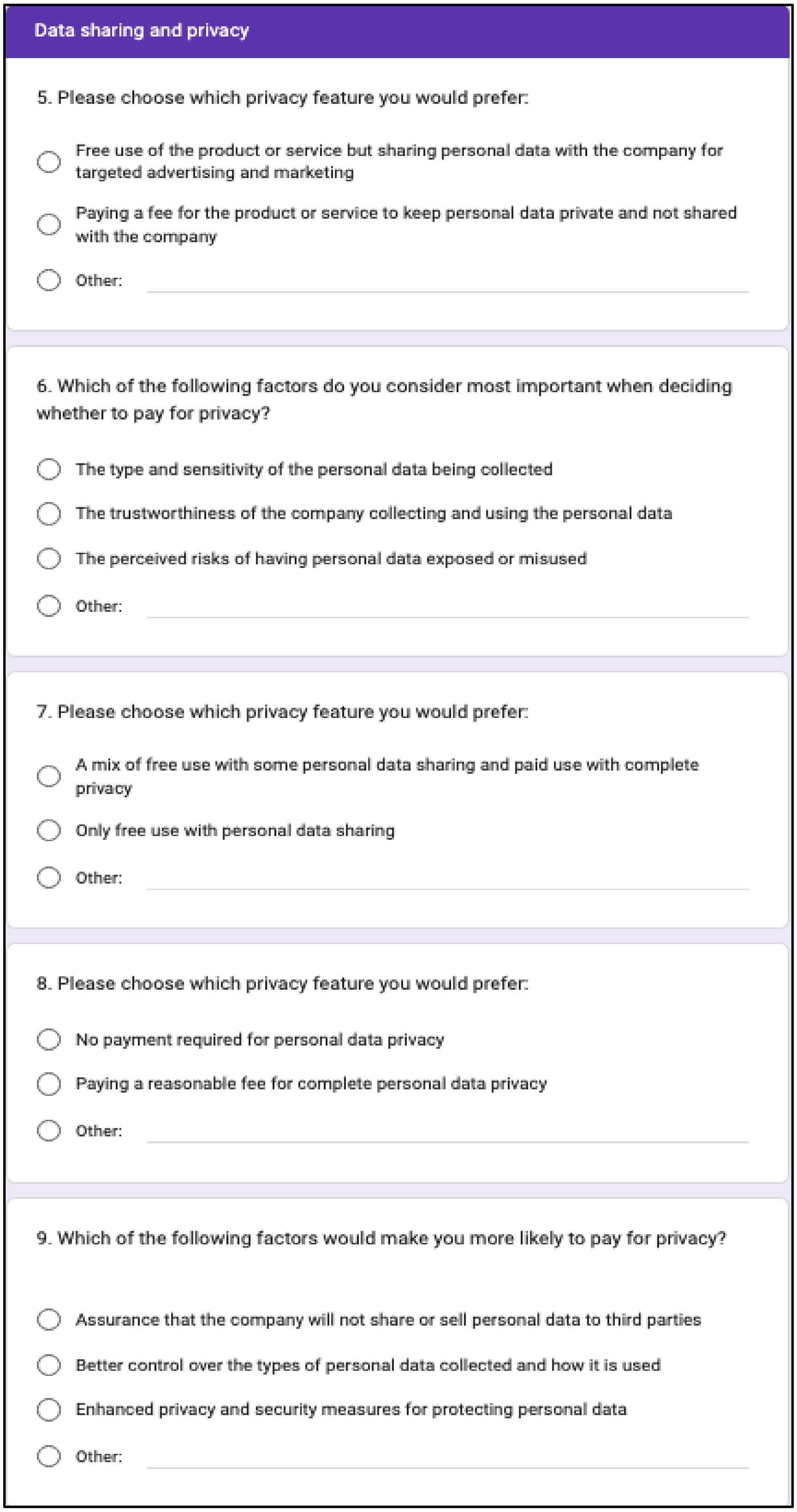
○ 35 - 44

○ 45 - 54

○ 55 - 64

○ 65 or over

Figure 46. Survey for interviews



Data sharing and privacy

5. Please choose which privacy feature you would prefer:

○ Free use of the product or service but sharing personal data with the company for targeted advertising and marketing

○ Paying a fee for the product or service to keep personal data private and not shared with the company

○ Other: ____________________

6. Which of the following factors do you consider most important when deciding whether to pay for privacy?

○ The type and sensitivity of the personal data being collected

○ The trustworthiness of the company collecting and using the personal data

○ The perceived risks of having personal data exposed or misused

○ Other: ____________________

7. Please choose which privacy feature you would prefer:

○ A mix of free use with some personal data sharing and paid use with complete privacy

○ Only free use with personal data sharing

○ Other: ____________________

8. Please choose which privacy feature you would prefer:

○ No payment required for personal data privacy

○ Paying a reasonable fee for complete personal data privacy

○ Other: ____________________

9. Which of the following factors would make you more likely to pay for privacy?

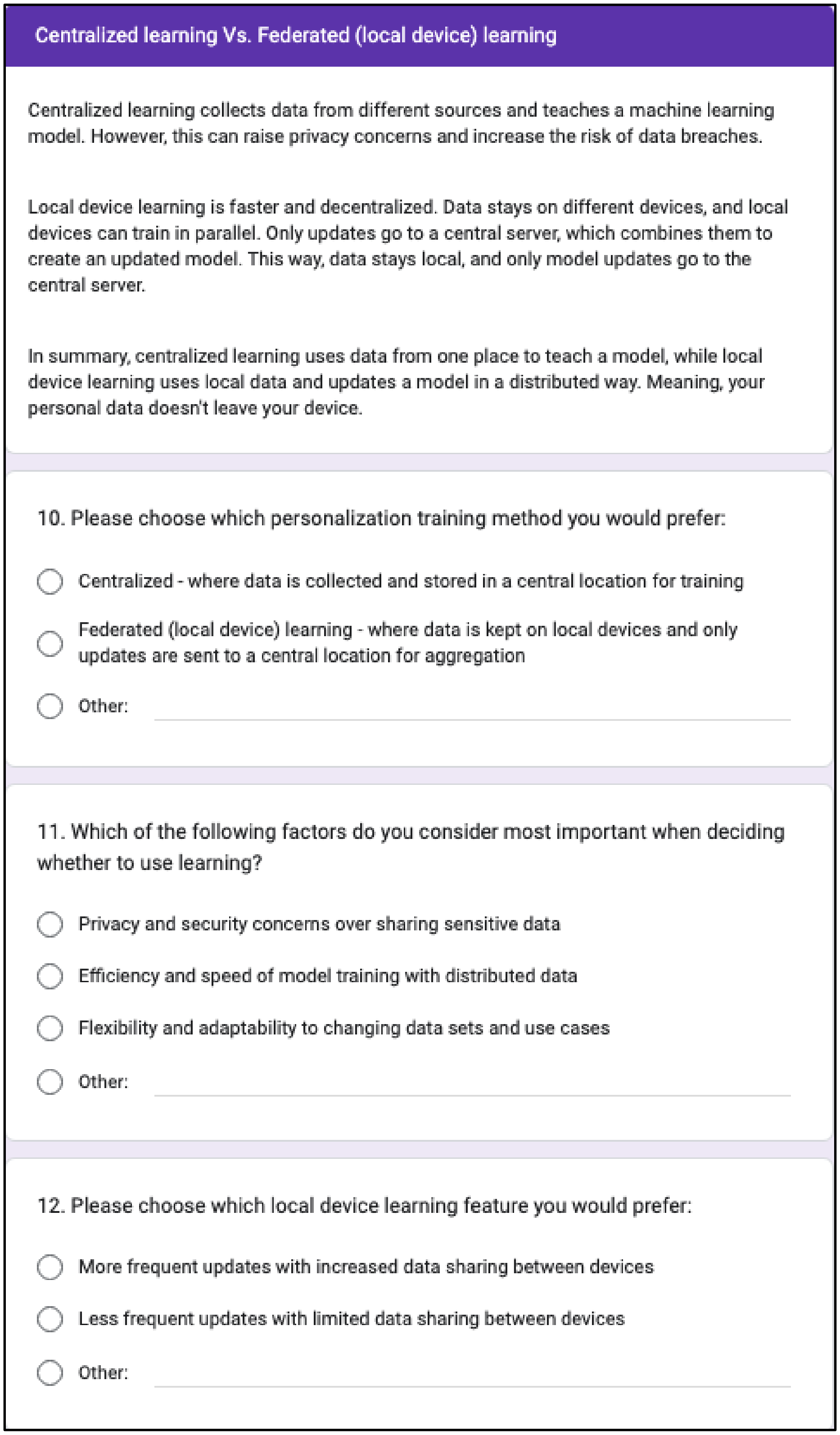
○ Assurance that the company will not share or sell personal data to third parties

○ Better control over the types of personal data collected and how it is used

○ Enhanced privacy and security measures for protecting personal data

○ Other: ____________________

Figure 47. Survey for interviews Personalized vs Randomized content



Centralized learning Vs. Federated (local device) learning

Centralized learning collects data from different sources and teaches a machine learning model. However, this can raise privacy concerns and increase the risk of data breaches.

Local device learning is faster and decentralized. Data stays on different devices, and local devices can train in parallel. Only updates go to a central server, which combines them to create an updated model. This way, data stays local, and only model updates go to the central server.

In summary, centralized learning uses data from one place to teach a model, while local device learning uses local data and updates a model in a distributed way. Meaning, your personal data doesn't leave your device.

10. Please choose which personalization training method you would prefer:

○ Centralized - where data is collected and stored in a central location for training

○ Federated (local device) learning - where data is kept on local devices and only updates are sent to a central location for aggregation

○ Other: __________

11. Which of the following factors do you consider most important when deciding whether to use learning?

○ Privacy and security concerns over sharing sensitive data

○ Efficiency and speed of model training with distributed data

○ Flexibility and adaptability to changing data sets and use cases

○ Other: __________

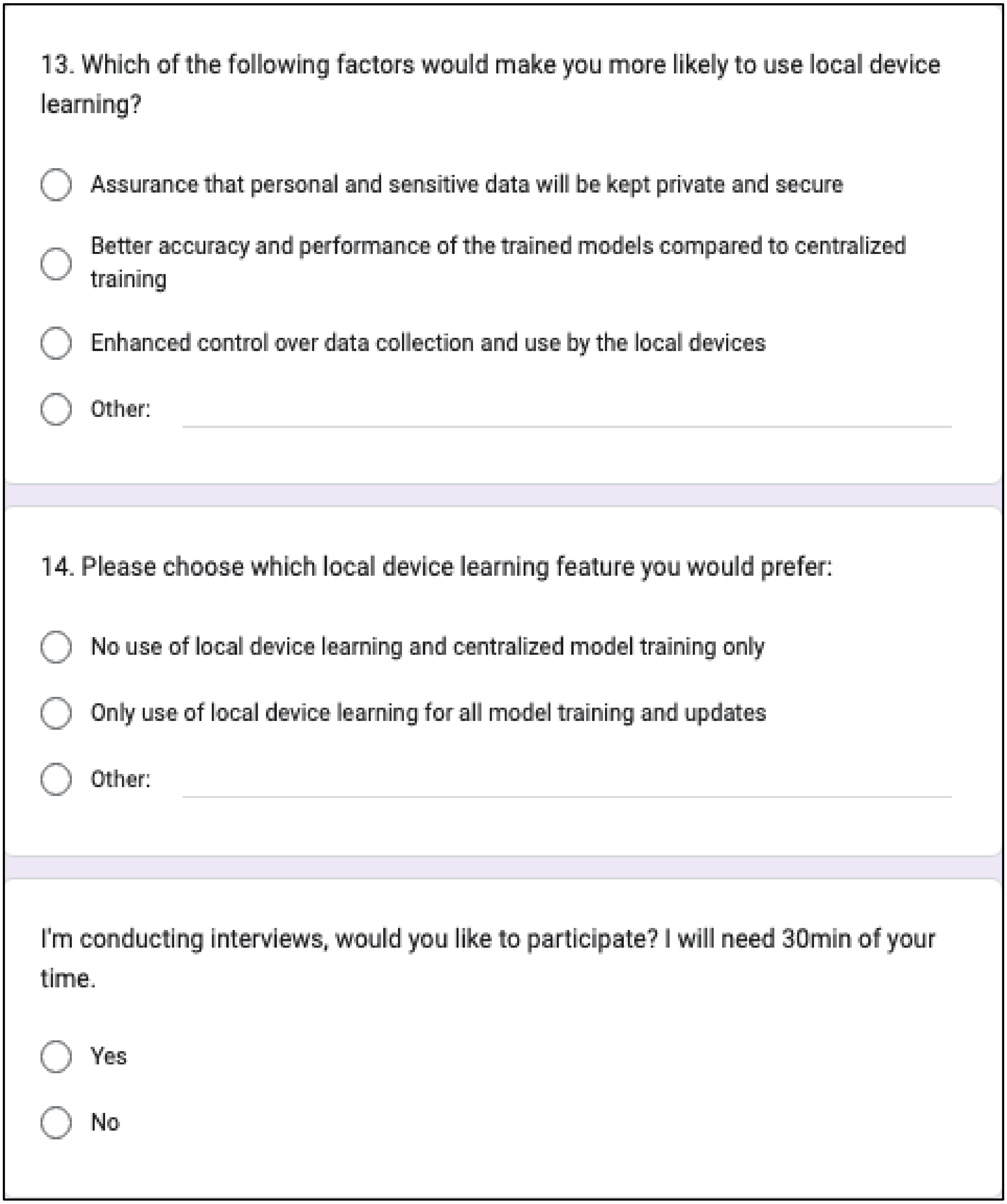
12. Please choose which local device learning feature you would prefer:

○ More frequent updates with increased data sharing between devices

○ Less frequent updates with limited data sharing between devices

○ Other: __________

Figure 48. Survey for interviews - Data sharing and privacy



13. Which of the following factors would make you more likely to use local device learning?

○ Assurance that personal and sensitive data will be kept private and secure

○ Better accuracy and performance of the trained models compared to centralized training

○ Enhanced control over data collection and use by the local devices

○ Other: ______________________

14. Please choose which local device learning feature you would prefer:

○ No use of local device learning and centralized model training only

○ Only use of local device learning for all model training and updates

○ Other: ______________________

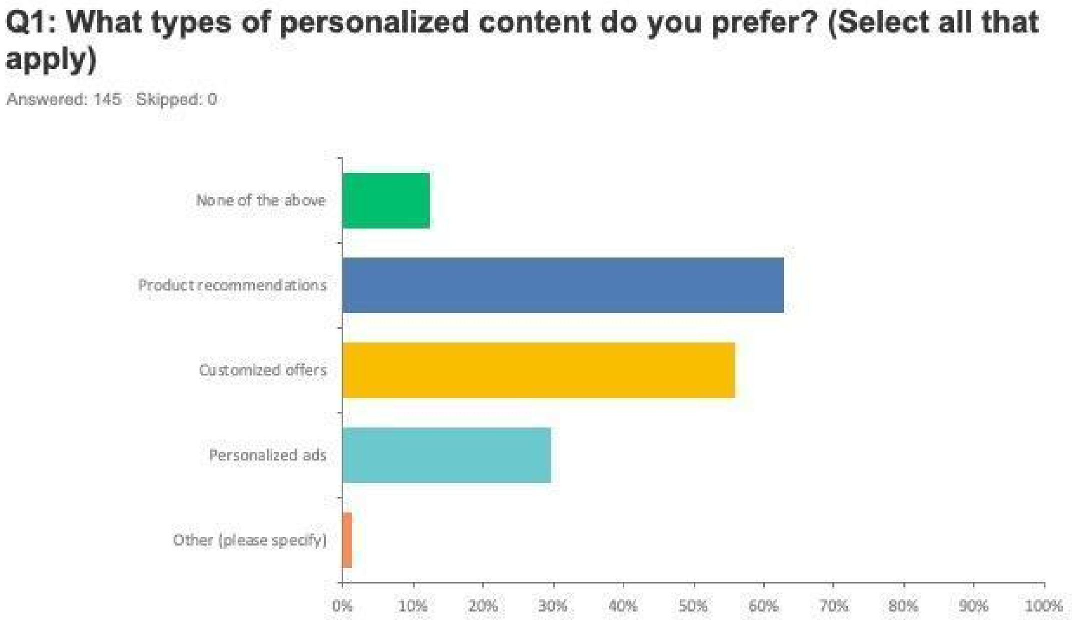
I'm conducting interviews, would you like to participate? I will need 30min of your time.

○ Yes

○ No

Figure 49. Survey for interviews - Centralized vs Federated Learning

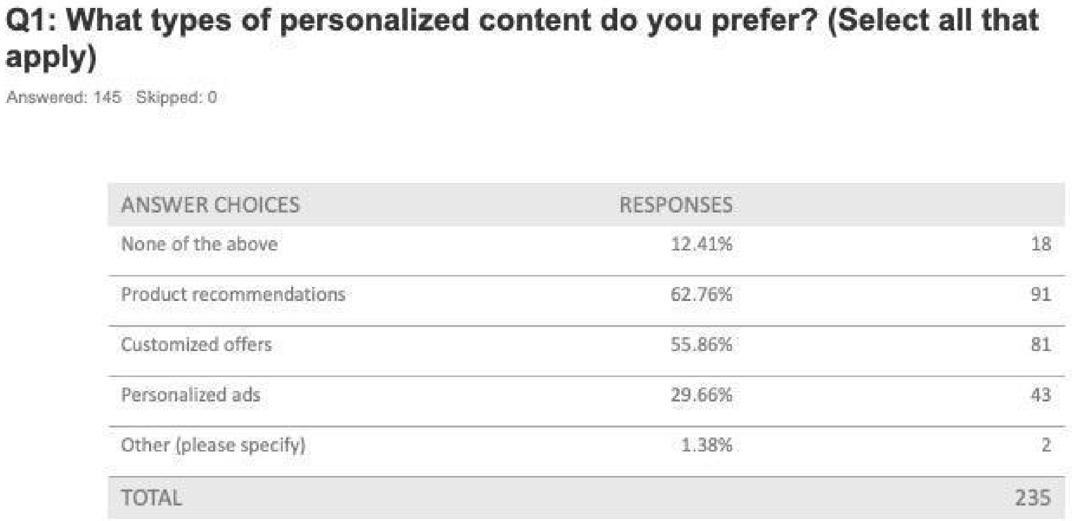
Appendix D: Customer preferences on personalization and privacy - Survey (general population)



Q1: What types of personalized content do you prefer? (Select all that apply)

Answered: 145 Skipped: 0

Figure 50. Content preferences (bar graph)

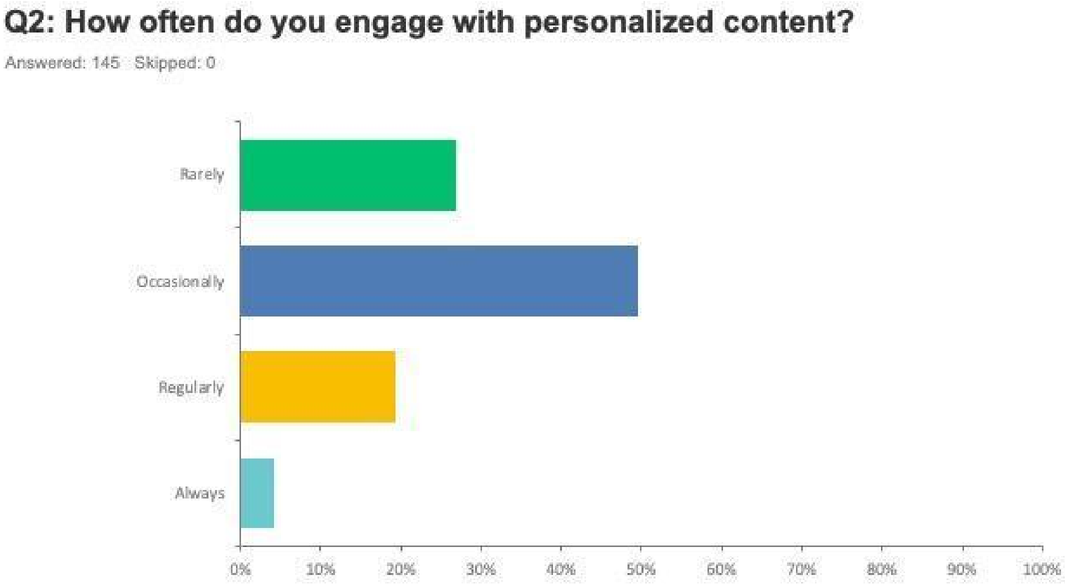


Q1: What types of personalized content do you prefer? (Select all that apply)

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
None of the above	12.41%	18
Product recommendations	62.76%	91
Customized offers	55.86%	81
Personalized ads	29.66%	43
Other (please specify)	1.38%	2
TOTAL		235

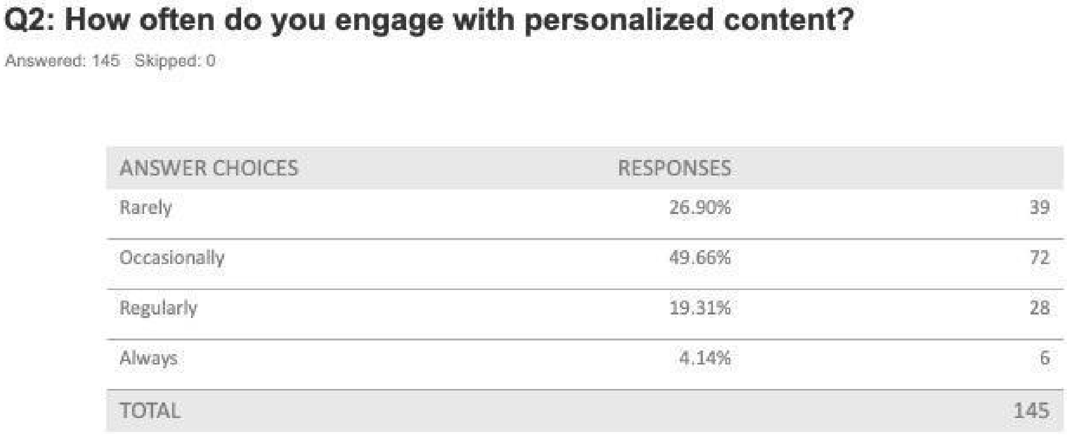
Figure 51. Content preferences



Q2: How often do you engage with personalized content?

Answered: 145 Skipped: 0

Figure 52. Personalized content engagement (bar graph)

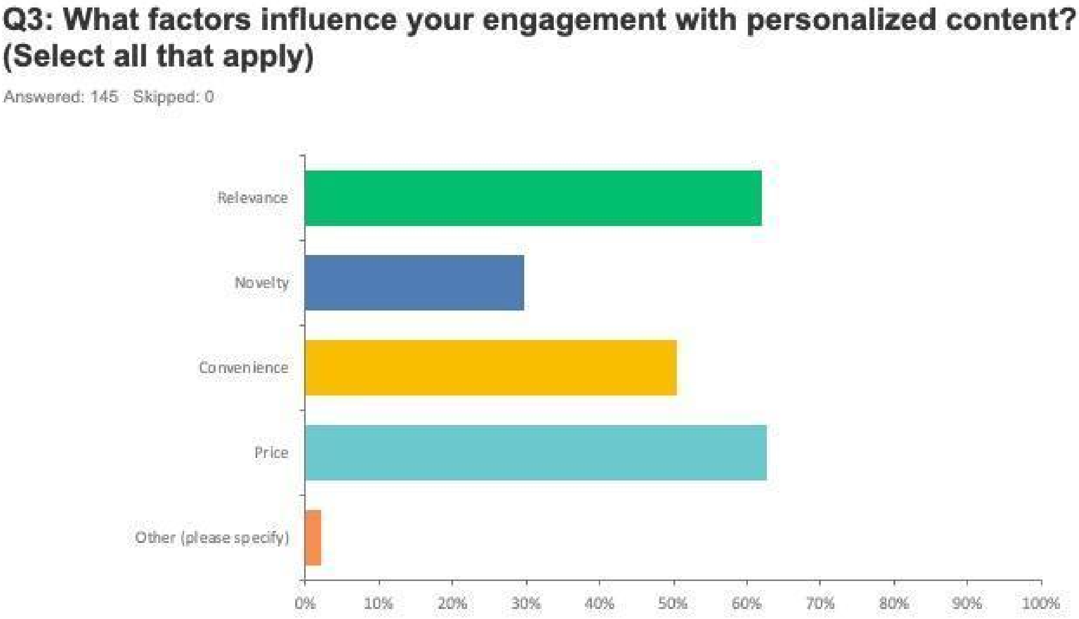


Q2: How often do you engage with personalized content?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Rarely	26.90%	39
Occasionally	49.66%	72
Regularly	19.31%	28
Always	4.14%	6
TOTAL		145

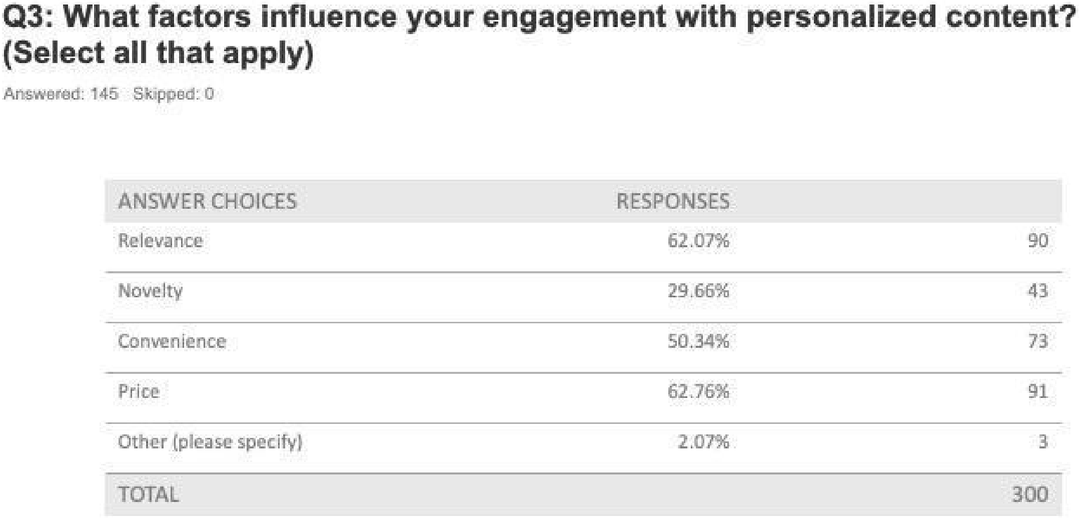
Figure 53. Personalized content engagement



Q3: What factors influence your engagement with personalized content? (Select all that apply)

Answered: 145 Skipped: 0

Figure 54. Factors influencing Engagement (bar graph)

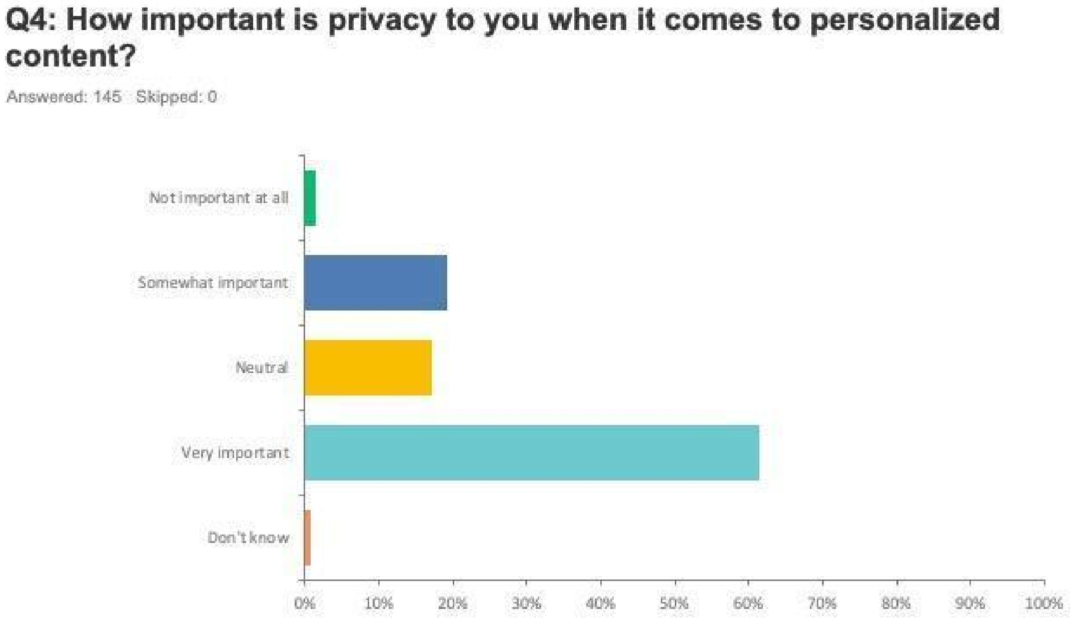


Q3: What factors influence your engagement with personalized content? (Select all that apply)

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Relevance	62.07%	90
Novelty	29.66%	43
Convenience	50.34%	73
Price	62.76%	91
Other (please specify)	2.07%	3
TOTAL		300

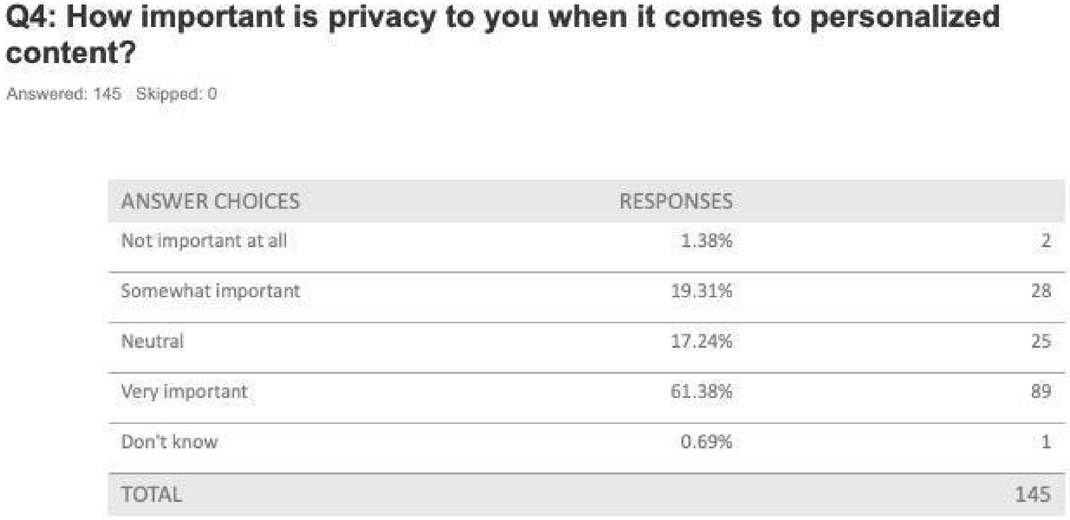
Figure 55. Factors influencing Engagement



Q4: How important is privacy to you when it comes to personalized content?

Answered: 145 Skipped: 0

Figure 56. Importance of Privacy regarding personalized content (bar graph)

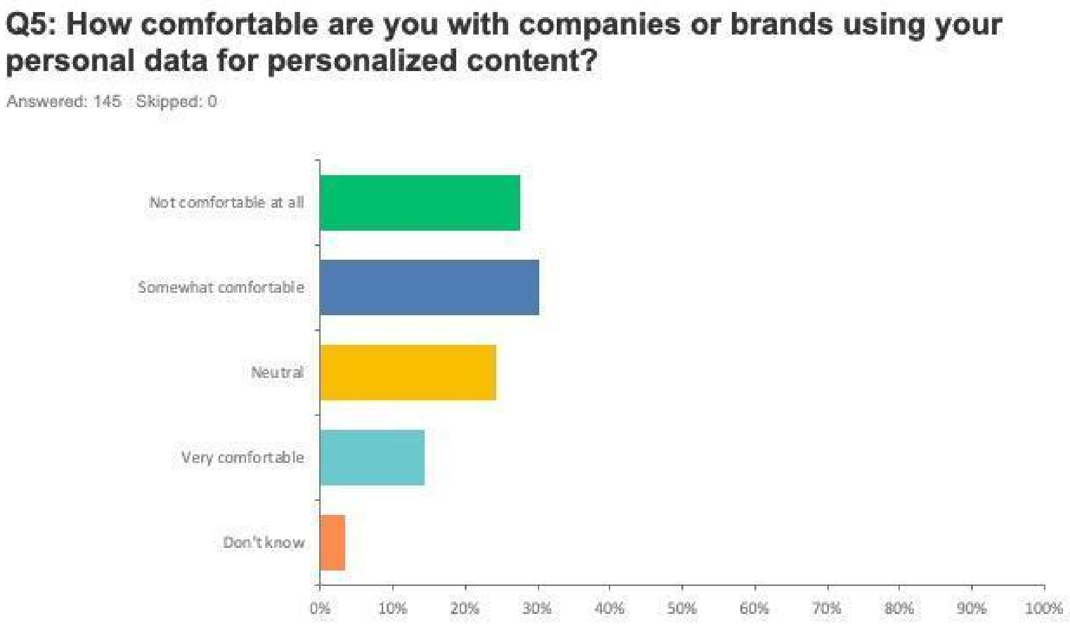


Q4: How important is privacy to you when it comes to personalized content?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Not important at all	1.38%	2
Somewhat important	19.31%	28
Neutral	17.24%	25
Very important	61.38%	89
Don't know	0.69%	1
TOTAL		145

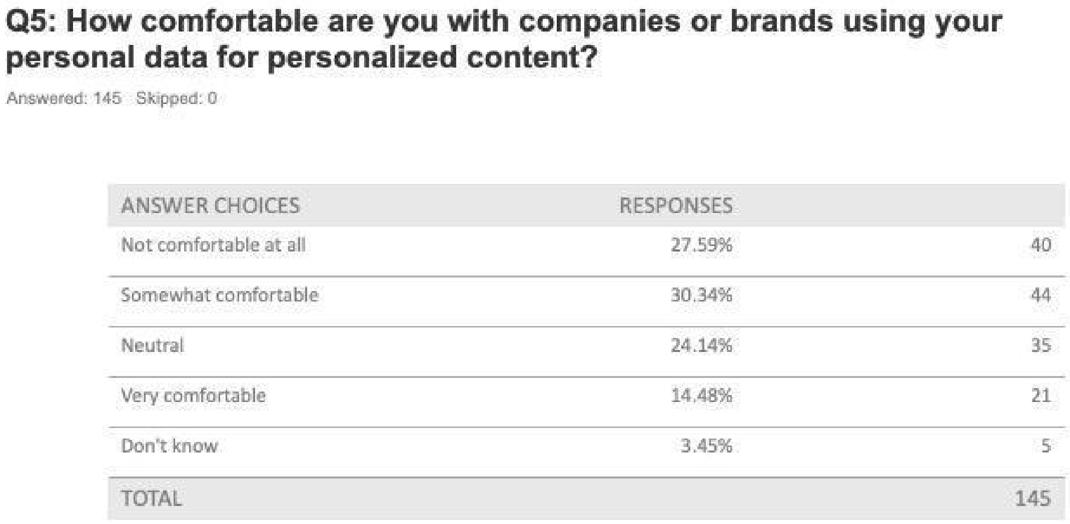
Figure 57. Importance of Privacy regarding personalized content



Q5: How comfortable are you with companies or brands using your personal data for personalized content?

Answered: 145 Skipped: 0

Figure 58. Comfort level with companies using personal data for personalized content (bar graph)

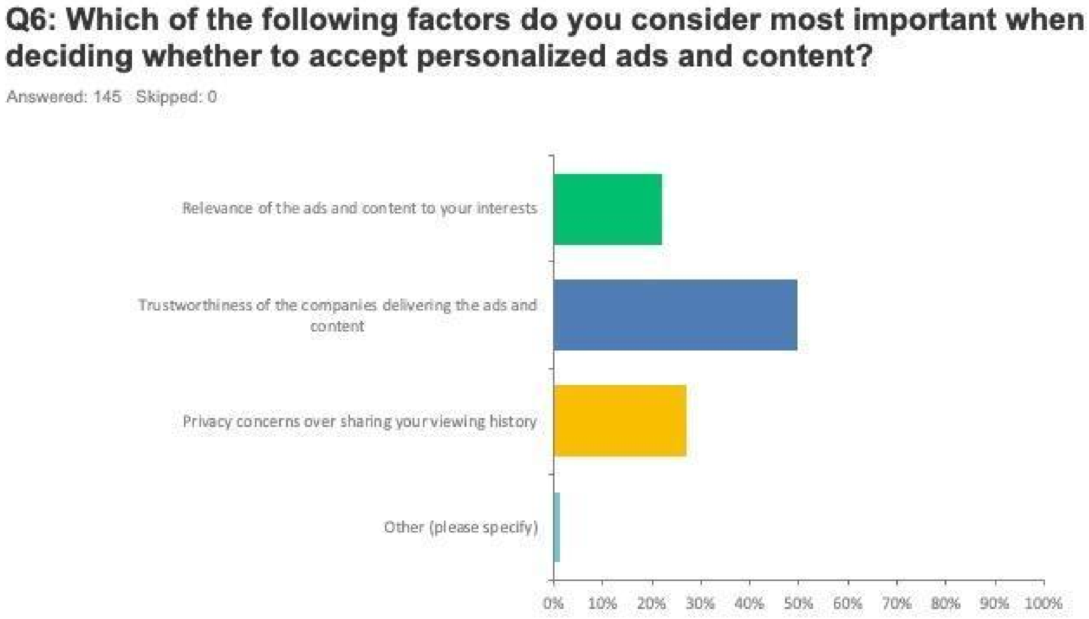


Q5: How comfortable are you with companies or brands using your personal data for personalized content?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Not comfortable at all	27.59%	40
Somewhat comfortable	30.34%	44
Neutral	24.14%	35
Very comfortable	14.48%	21
Don't know	3.45%	5
TOTAL		145

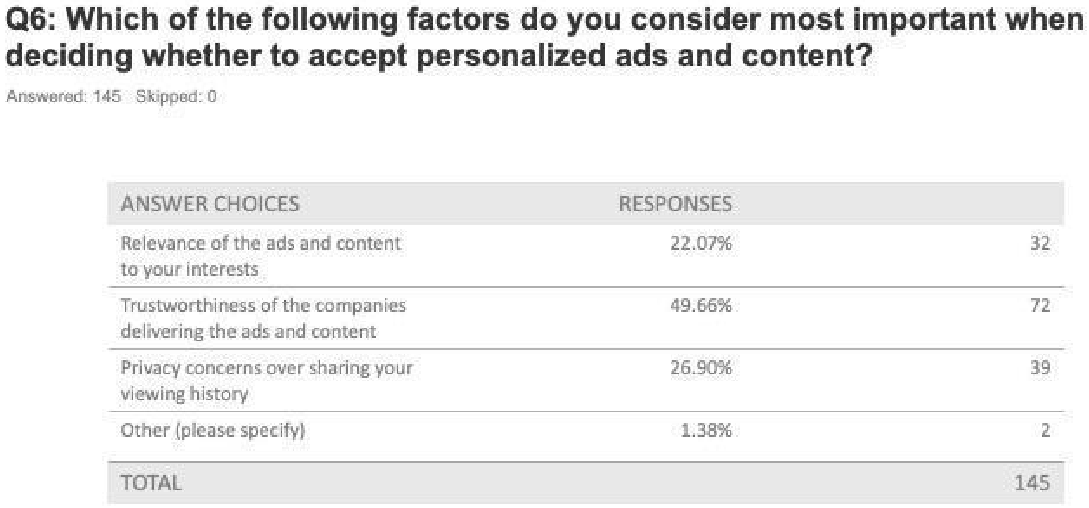
Figure 59. Comfort level with companies using personal data for personalized content



Q6: Which of the following factors do you consider most important when deciding whether to accept personalized ads and content?

Answered: 145 Skipped: 0

Figure 60. Factors considered whether to accept personalized content (bar graph)

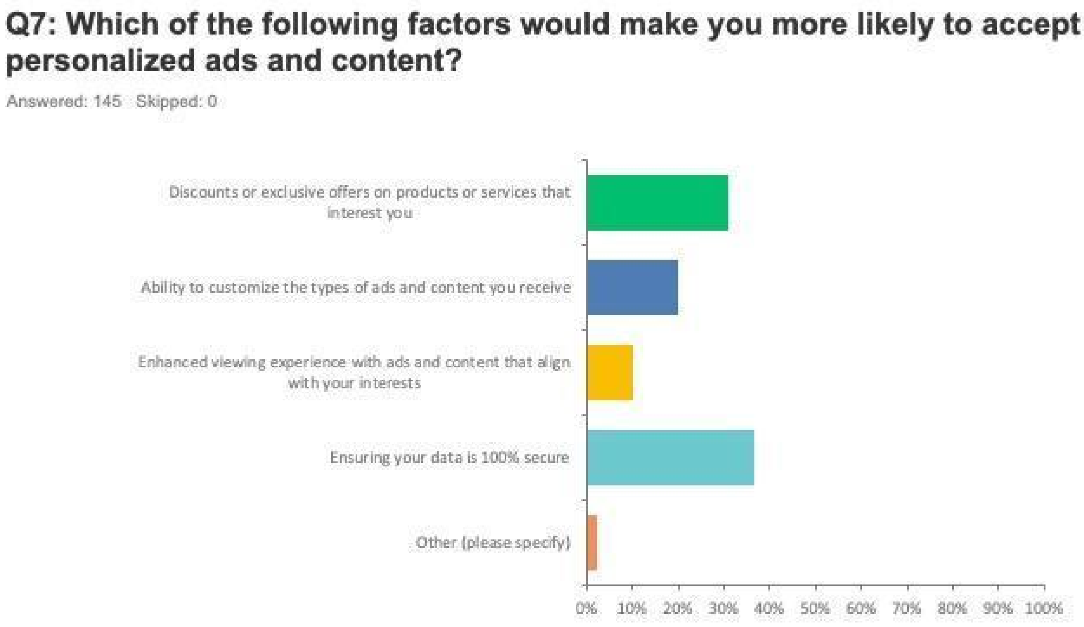


Q6: Which of the following factors do you consider most important when deciding whether to accept personalized ads and content?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Relevance of the ads and content to your interests	22.07%	32
Trustworthiness of the companies delivering the ads and content	49.66%	72
Privacy concerns over sharing your viewing history	26.90%	39
Other (please specify)	1.38%	2
TOTAL		145

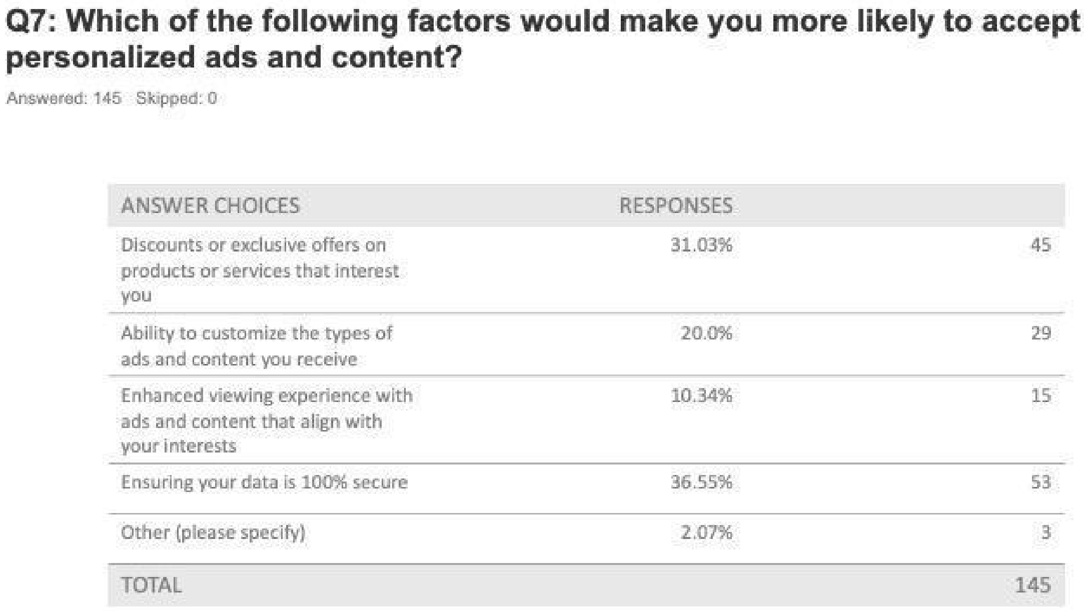
Figure 61. Factors considered whether to accept personalized content



Q7: Which of the following factors would make you more likely to accept personalized ads and content?

Answered: 145 Skipped: 0

Figure 62. Factors that would increase the acceptance personalized content (bar graph)

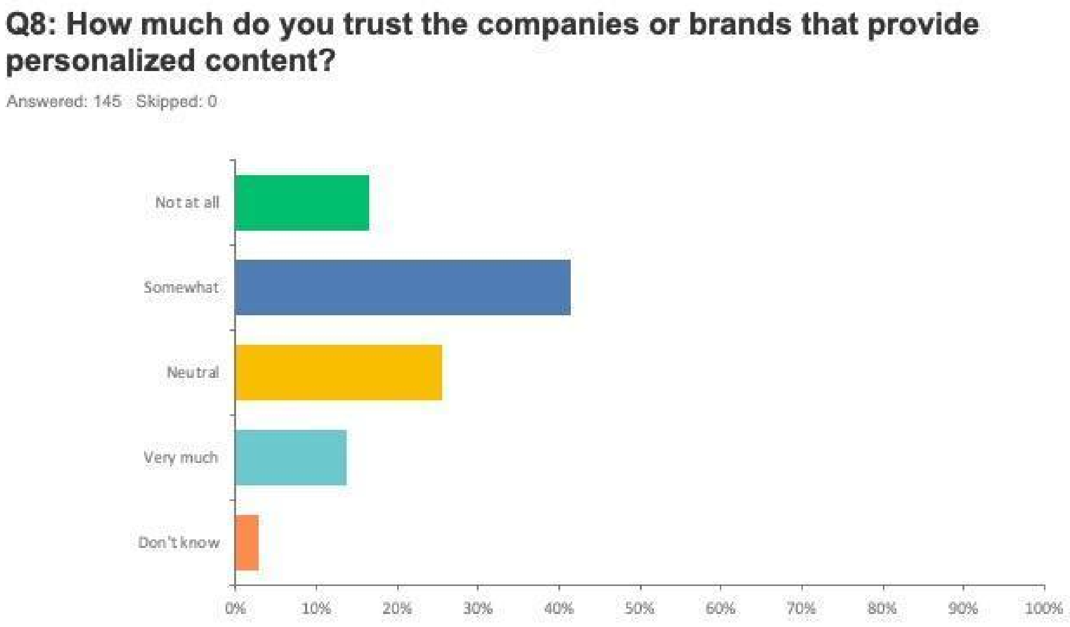


Q7: Which of the following factors would make you more likely to accept personalized ads and content?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Discounts or exclusive offers on products or services that interest you	31.03%	45
Ability to customize the types of ads and content you receive	20.0%	29
Enhanced viewing experience with ads and content that align with your interests	10.34%	15
Ensuring your data is 100% secure	36.55%	53
Other (please specify)	2.07%	3
TOTAL		145

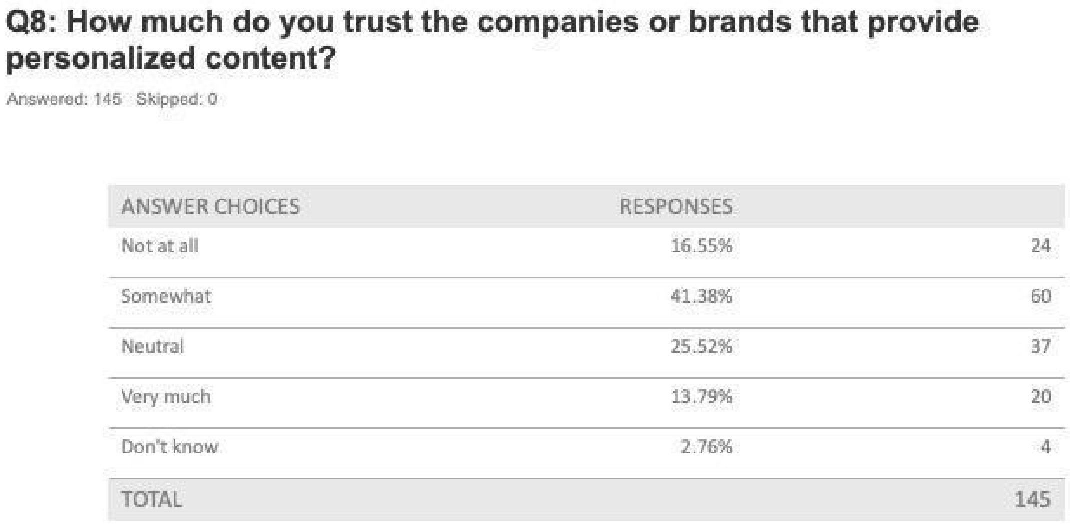
Figure 63. Factors that would increase the acceptance personalized content



Q8: How much do you trust the companies or brands that provide personalized content?

Answered: 145 Skipped: 0

Figure 64. Trust on companies or brands that provide personalized content (bar graph)

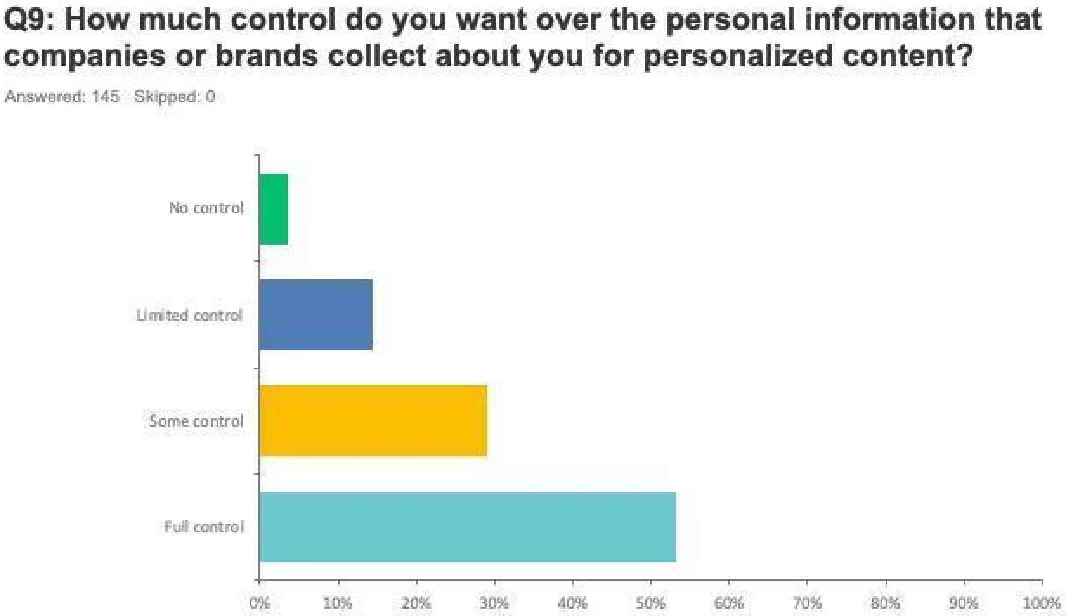


Q8: How much do you trust the companies or brands that provide personalized content?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Not at all	16.55%	24
Somewhat	41.38%	60
Neutral	25.52%	37
Very much	13.79%	20
Don't know	2.76%	4
TOTAL		145

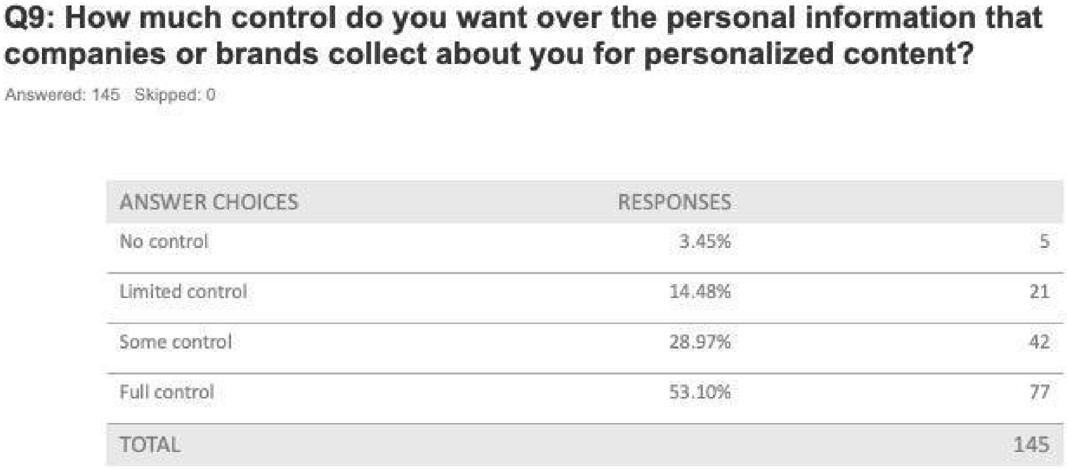
Figure 65. Trust on companies or brands that provide personalized content



Q9: How much control do you want over the personal information that companies or brands collect about you for personalized content?

Answered: 145 Skipped: 0

Figure 66. Control on data collected by companies or brands (bar graph)

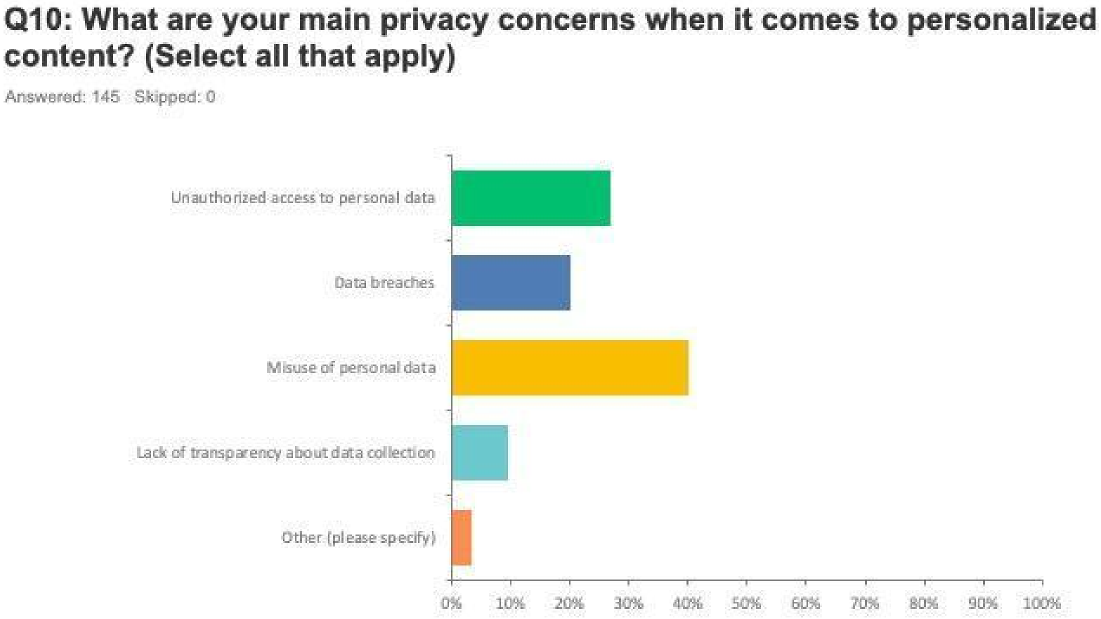


Q9: How much control do you want over the personal information that companies or brands collect about you for personalized content?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
No control	3.45%	5
Limited control	14.48%	21
Some control	28.97%	42
Full control	53.10%	77
TOTAL		145

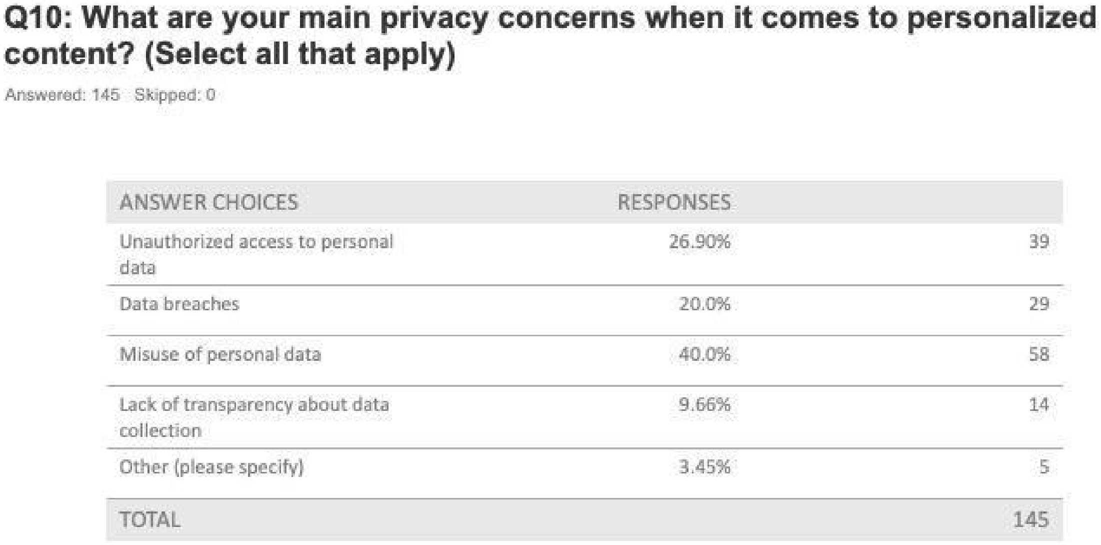
Figure 67. Control on data collected by companies or brands



Q10: What are your main privacy concerns when it comes to personalized content? (Select all that apply)

Answered: 145 Skipped: 0

Figure 68. Privacy concerns on personalized content (bar graph)

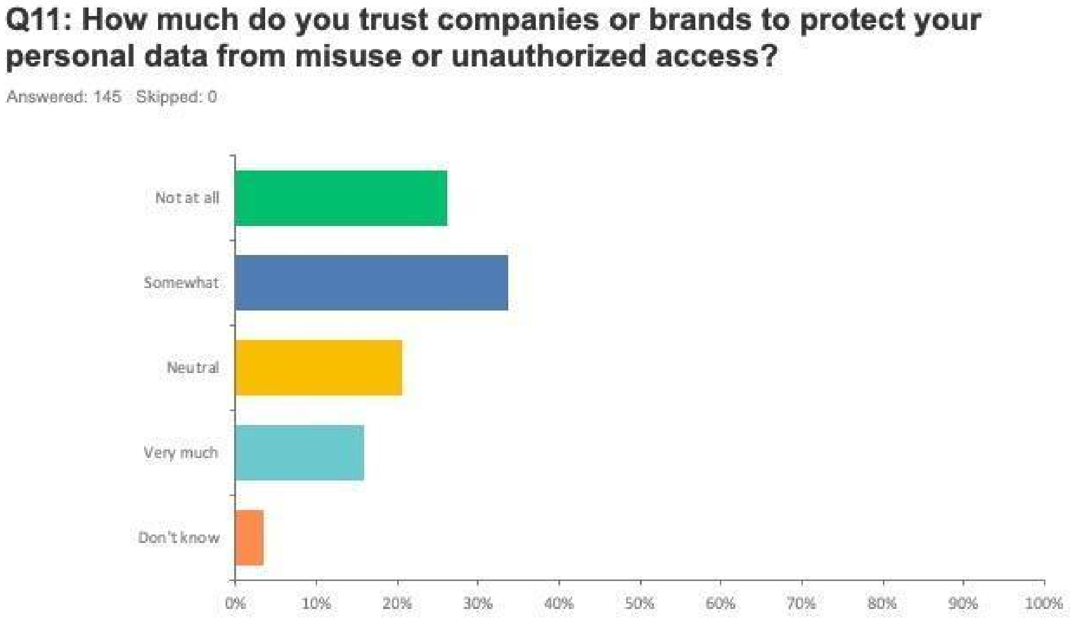


Q10: What are your main privacy concerns when it comes to personalized content? (Select all that apply)

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Unauthorized access to personal data	26.90%	39
Data breaches	20.0%	29
Misuse of personal data	40.0%	58
Lack of transparency about data collection	9.66%	14
Other (please specify)	3.45%	5
TOTAL		145

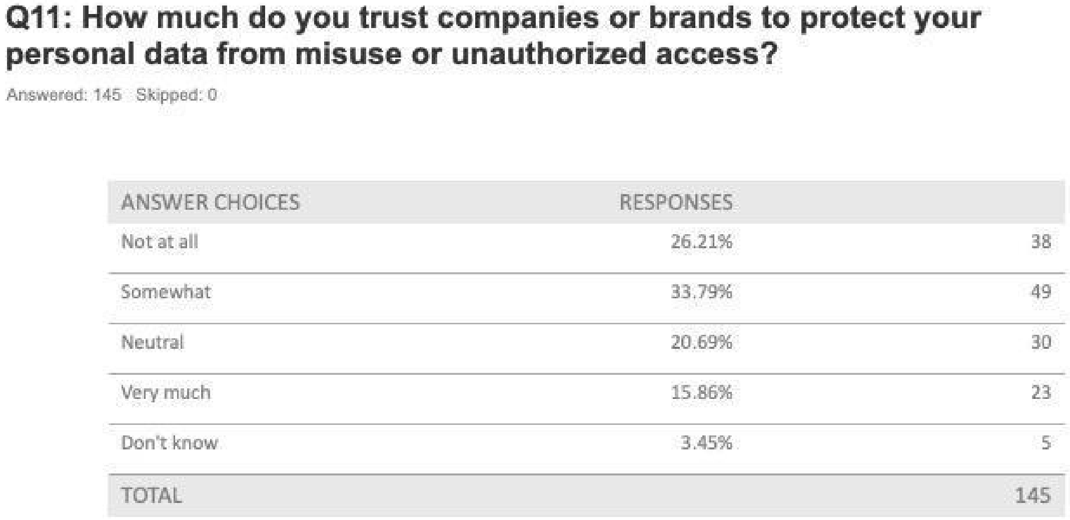
Figure 69. Privacy concerns on personalized content



Q11: How much do you trust companies or brands to protect your personal data from misuse or unauthorized access?

Answered: 145 Skipped: 0

Figure 70. Trust on companies or brands to protect personal data (bar graph)

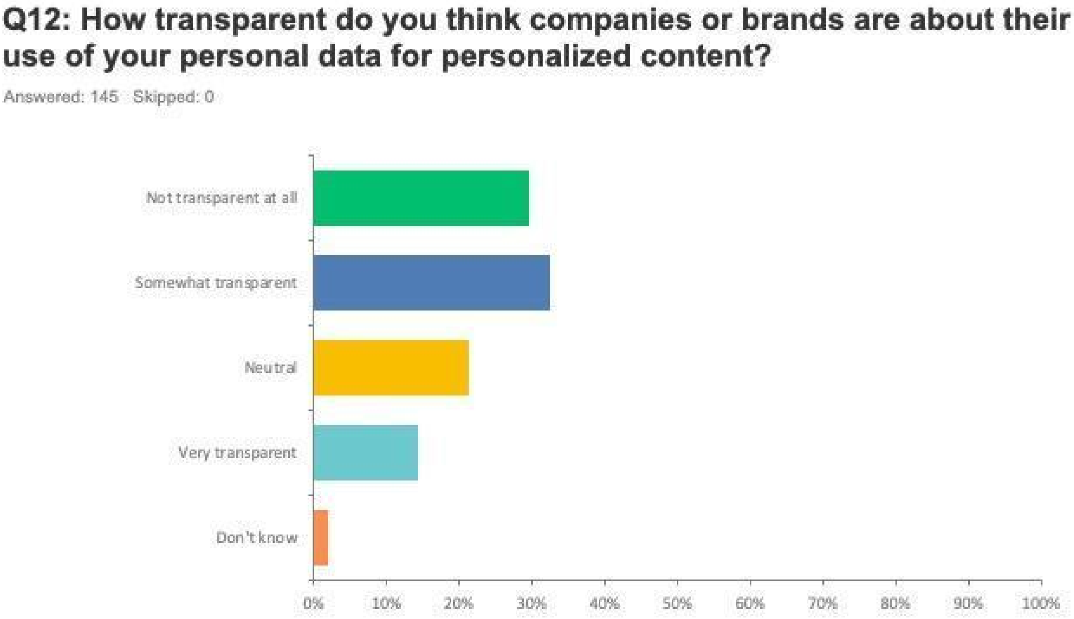


Q11: How much do you trust companies or brands to protect your personal data from misuse or unauthorized access?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Not at all	26.21%	38
Somewhat	33.79%	49
Neutral	20.69%	30
Very much	15.86%	23
Don't know	3.45%	5
TOTAL		145

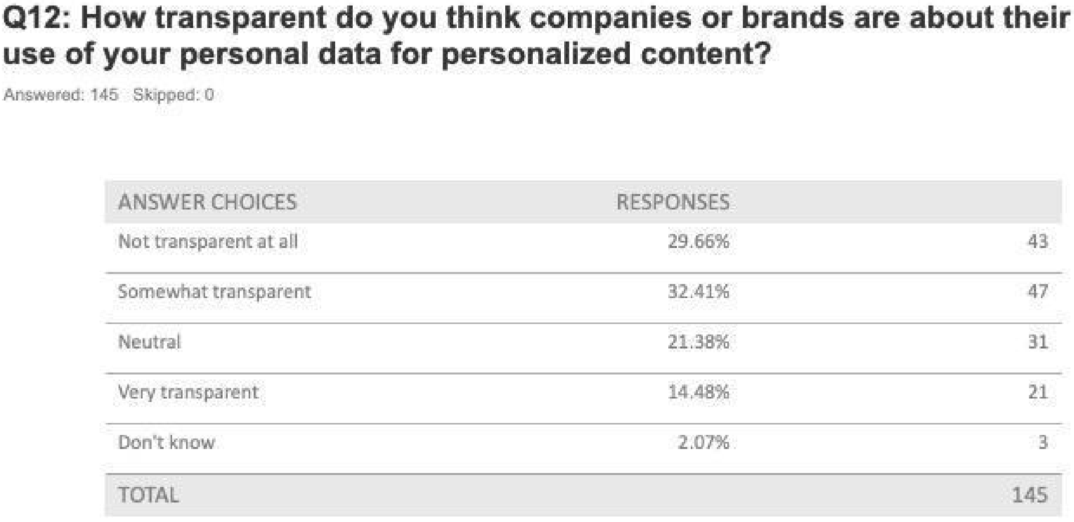
Figure 71. Trust on companies or brands to protect personal data



Q12: How transparent do you think companies or brands are about their use of your personal data for personalized content?

Answered: 145 Skipped: 0

Figure 72. Transparency about the use of personal data (bar graph)

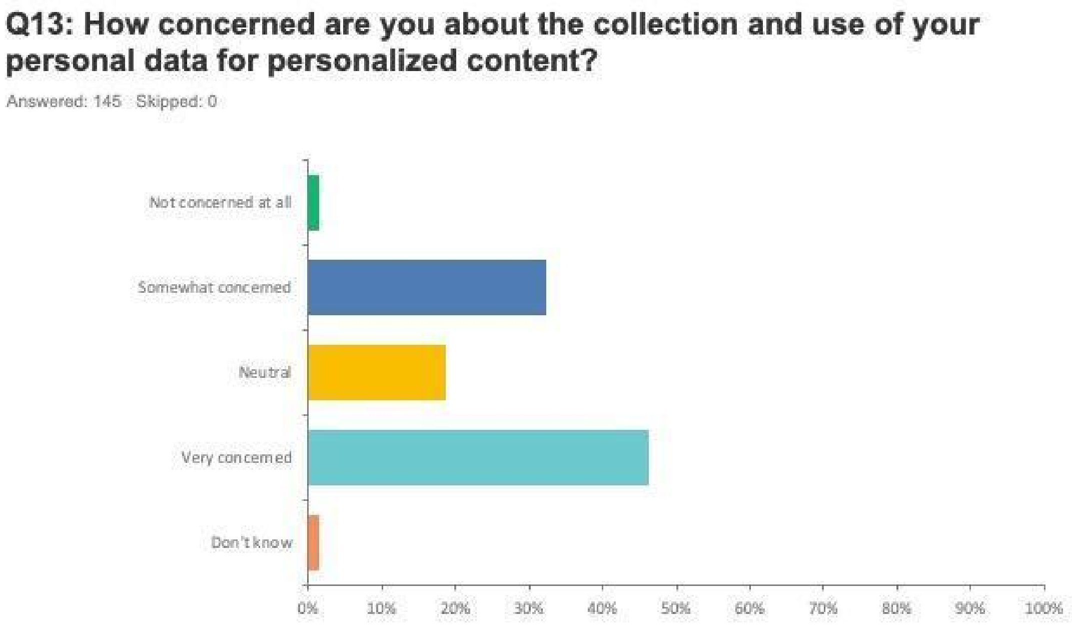


Q12: How transparent do you think companies or brands are about their use of your personal data for personalized content?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Not transparent at all	29.66%	43
Somewhat transparent	32.41%	47
Neutral	21.38%	31
Very transparent	14.48%	21
Don't know	2.07%	3
TOTAL		145

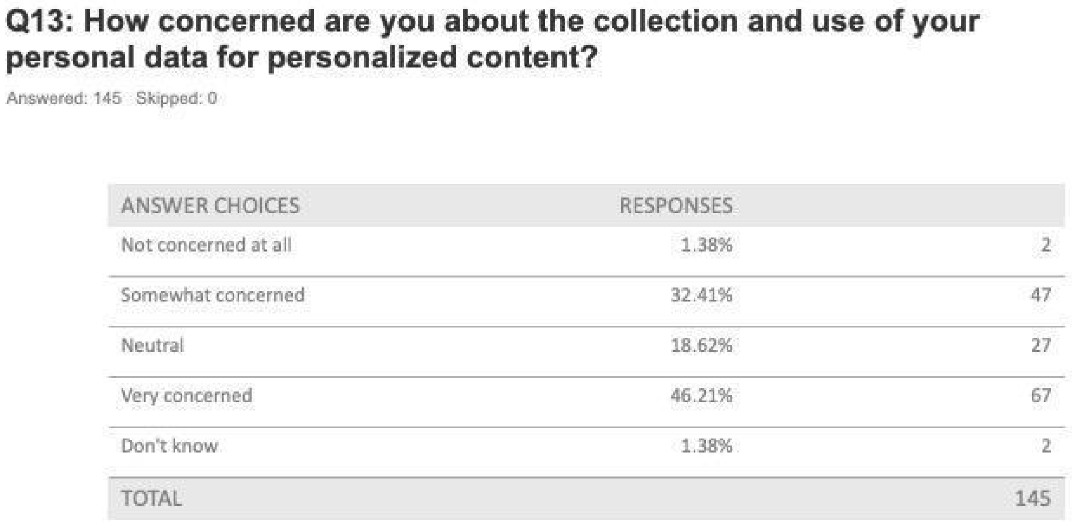
Figure 73. Transparency about the use of personal data



Q13: How concerned are you about the collection and use of your personal data for personalized content?

Answered: 145 Skipped: 0

Figure 74. Concern about the use of personal data (bar graph)

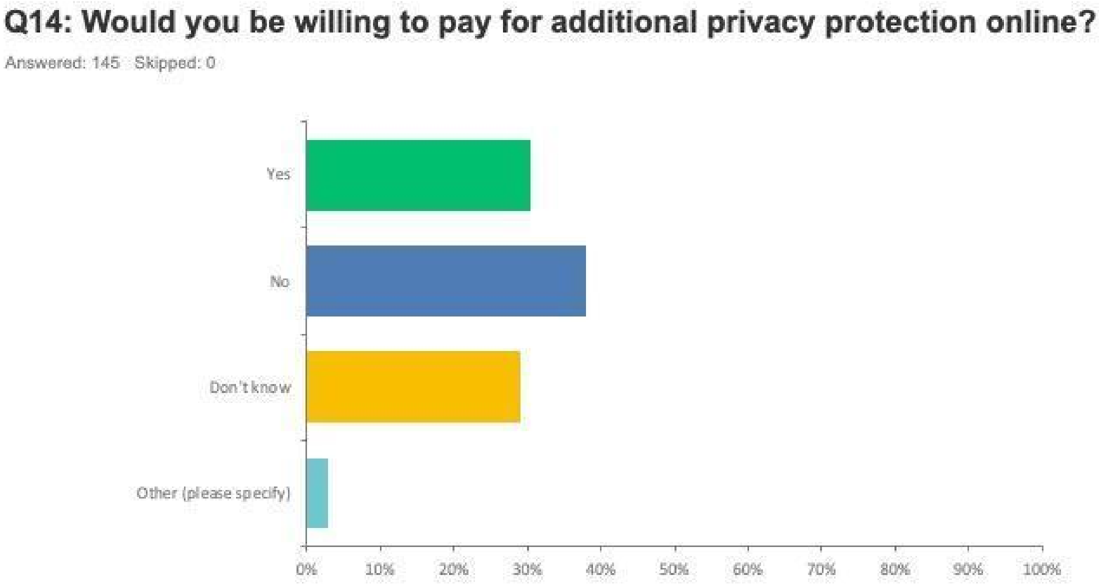


Q13: How concerned are you about the collection and use of your personal data for personalized content?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Not concerned at all	1.38%	2
Somewhat concerned	32.41%	47
Neutral	18.62%	27
Very concerned	46.21%	67
Don't know	1.38%	2
TOTAL		145

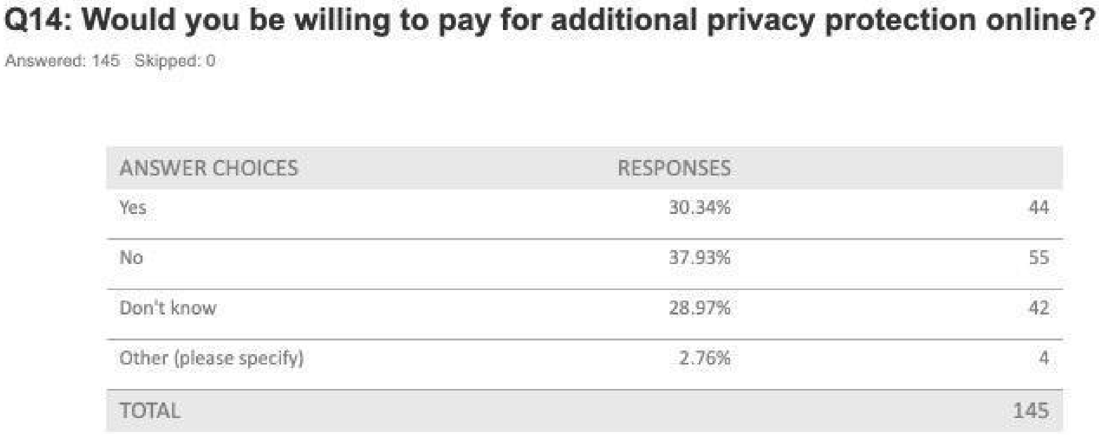
Figure 75. Concern about the use of personal data



Q14: Would you be willing to pay for additional privacy protection online?

Answered: 145 Skipped: 0

Figure 76. Pay for additional privacy protection online (bar graph)

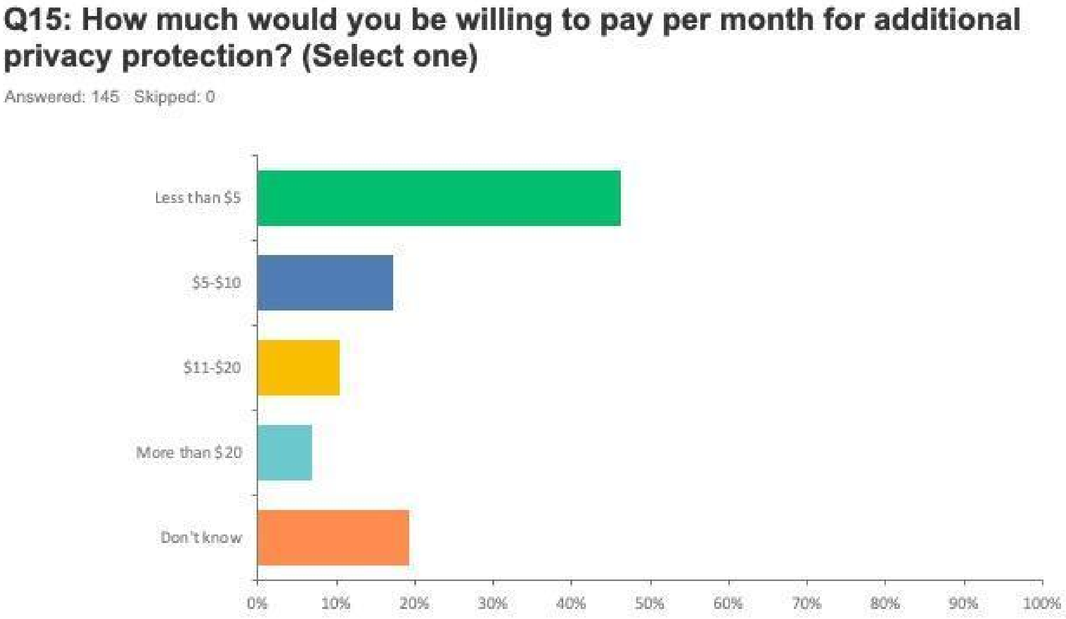


Q14: Would you be willing to pay for additional privacy protection online?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Yes	30.34%	44
No	37.93%	55
Don't know	28.97%	42
Other (please specify)	2.76%	4
TOTAL		145

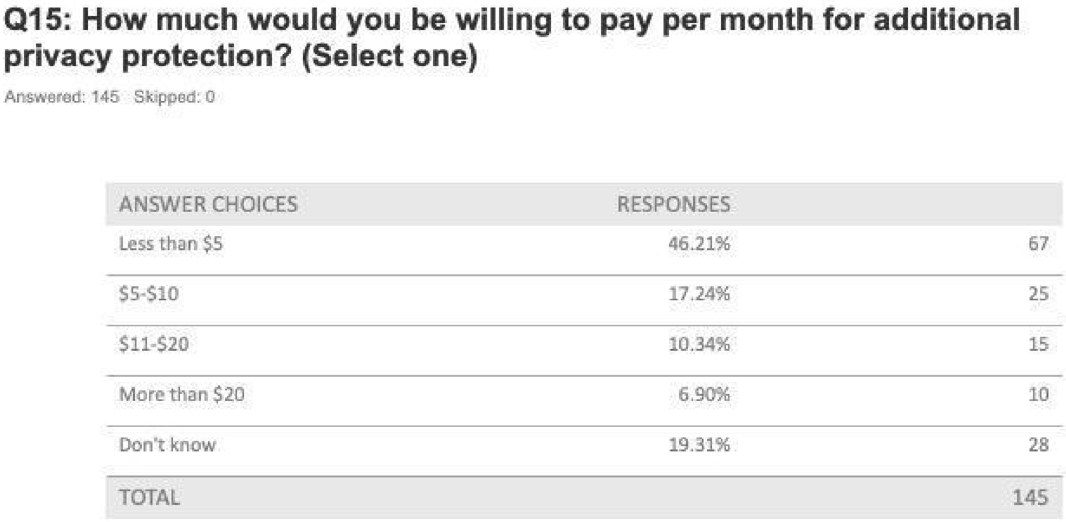
Figure 77. Pay for additional privacy protection online



Q15: How much would you be willing to pay per month for additional privacy protection? (Select one)

Answered: 145 Skipped: 0

Figure 78. How much to pay for additional privacy protection online (bar graph)



Q15: How much would you be willing to pay per month for additional privacy protection? (Select one)

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Less than $5	46.21%	67
$5-$10	17.24%	25
$11-$20	10.34%	15
More than $20	6.90%	10
Don't know	19.31%	28
TOTAL		145

Figure 79. How much to pay for additional privacy protection online

Q16: What specific privacy protection measures would you be willing to pay for? (Select all that apply)

Answered: 145 Skipped: 0

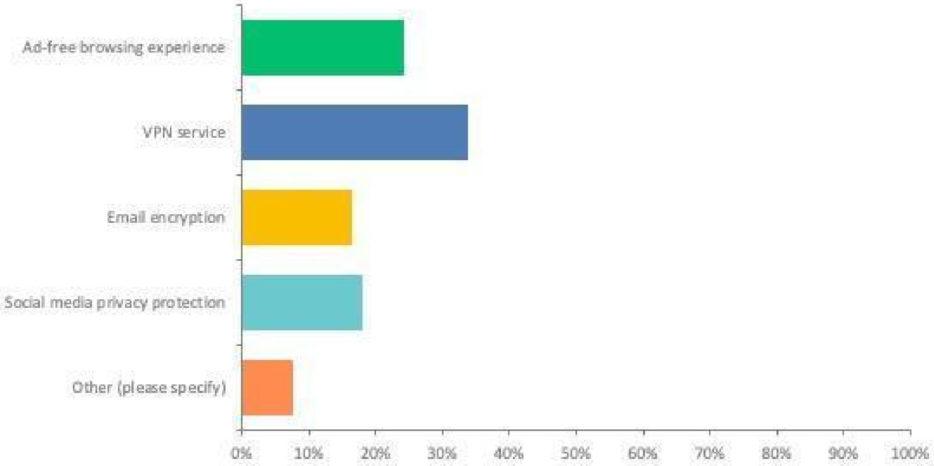
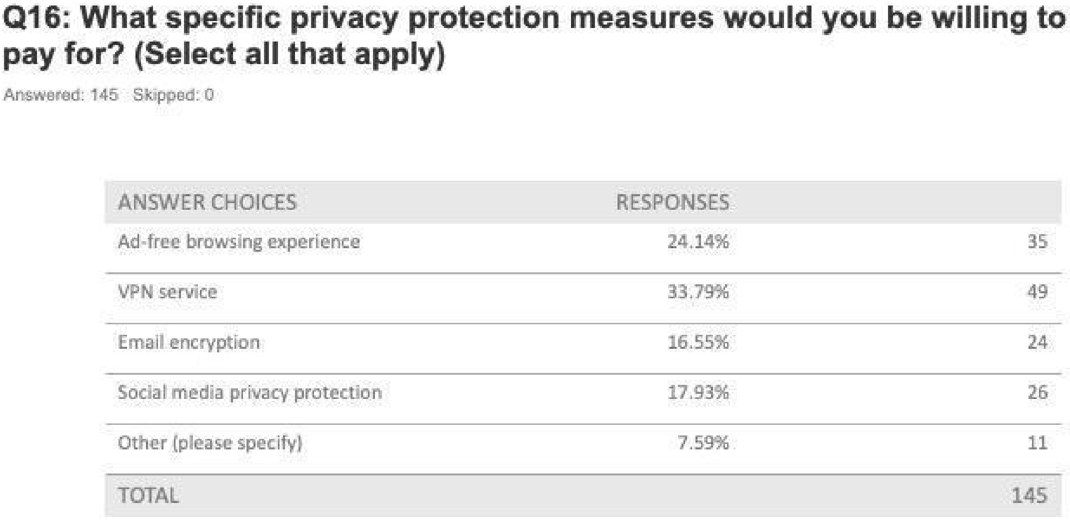


Figure 80. Features to pay for additional privacy protection online (bar graph)

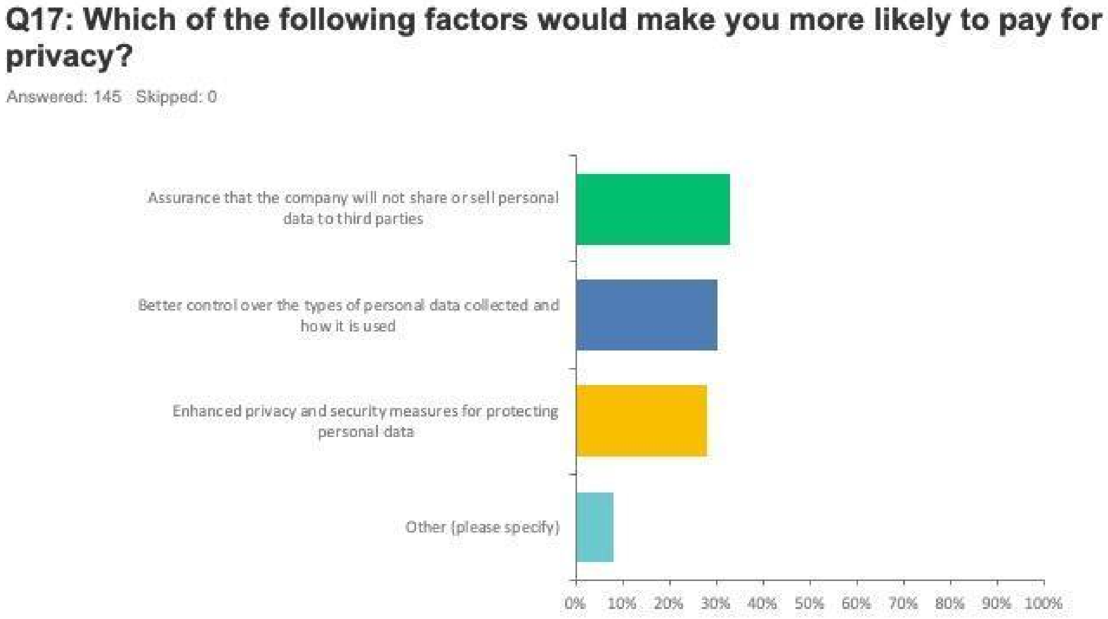


Q16: What specific privacy protection measures would you be willing to pay for? (Select all that apply)

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Ad-free browsing experience	24.14%	35
VPN service	33.79%	49
Email encryption	16.55%	24
Social media privacy protection	17.93%	26
Other (please specify)	7.59%	11
TOTAL		145

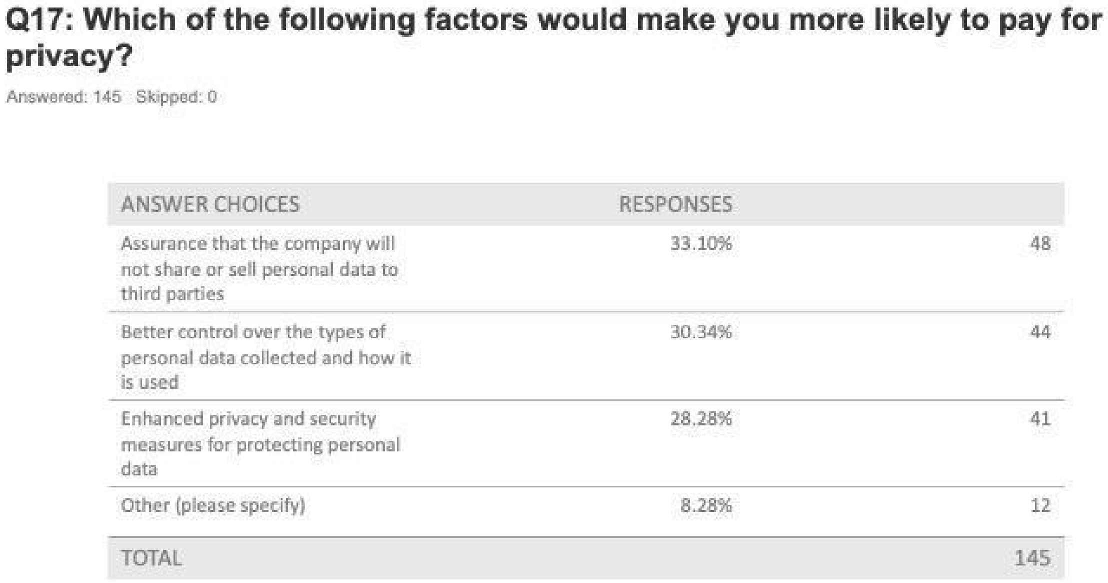
Figure 81. Features to pay for additional privacy protection online



Q17: Which of the following factors would make you more likely to pay for privacy?

Answered: 145 Skipped: 0

Figure 82. Factors that influence to pay for additional privacy protection online (bar graph)

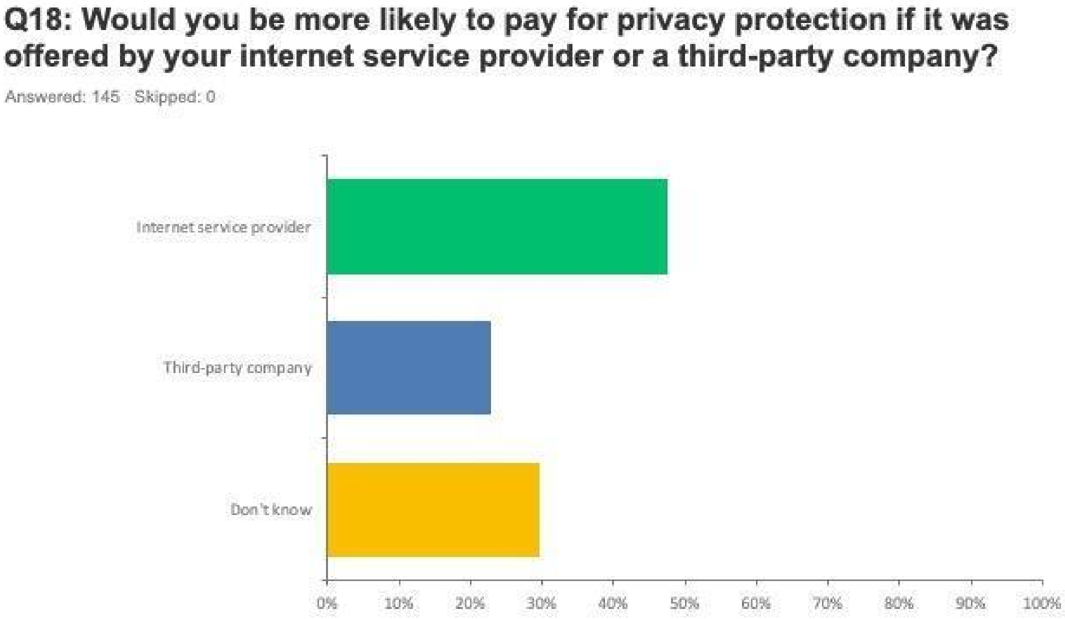


Q17: Which of the following factors would make you more likely to pay for privacy?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Assurance that the company will not share or sell personal data to third parties	33.10%	48
Better control over the types of personal data collected and how it is used	30.34%	44
Enhanced privacy and security measures for protecting personal data	28.28%	41
Other (please specify)	8.28%	12
TOTAL		145

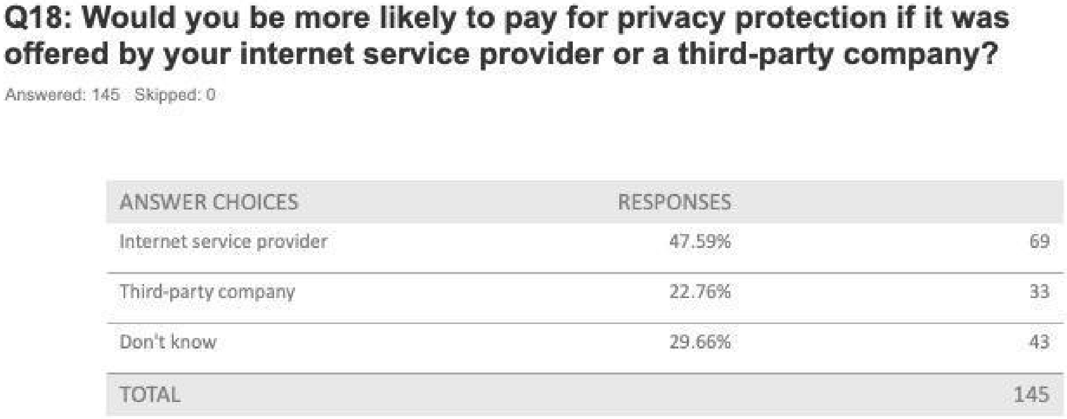
Figure 83. Factors that influence to pay for additional privacy protection online



Q18: Would you be more likely to pay for privacy protection if it was offered by your internet service provider or a third-party company?

Answered: 145 Skipped: 0

Figure 84. Pay for additional privacy protection online based on provider (bar graph)

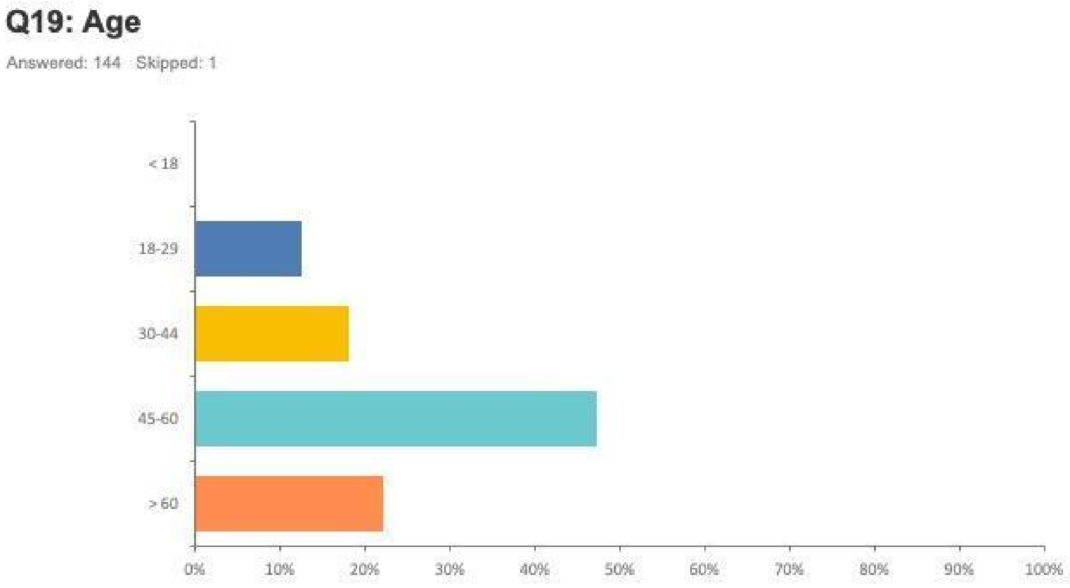


Q18: Would you be more likely to pay for privacy protection if it was offered by your internet service provider or a third-party company?

Answered: 145 Skipped: 0

ANSWER CHOICES	RESPONSES	
Internet service provider	47.59%	69
Third-party company	22.76%	33
Don't know	29.66%	43
TOTAL		145

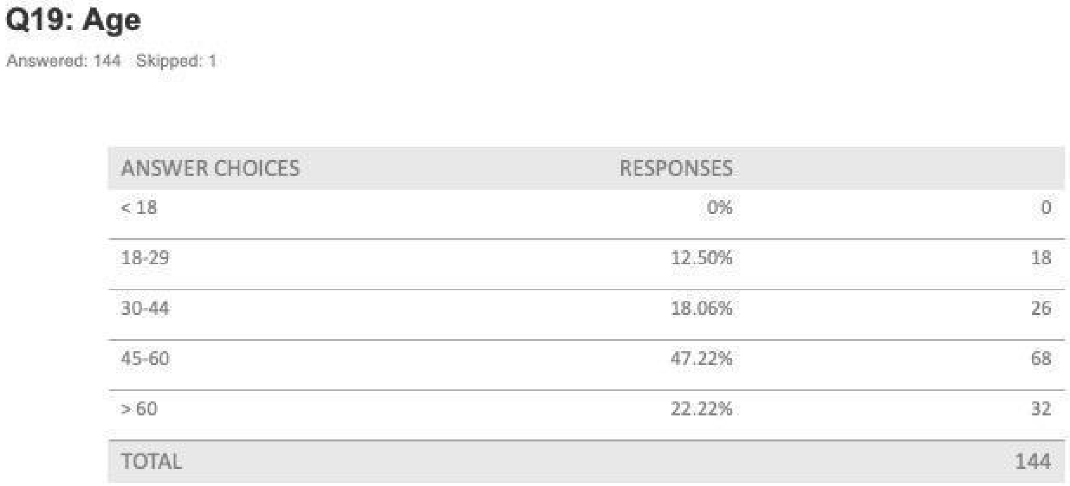
Figure 85. Pay for additional privacy protection online based on provider



Q19: Age

Answered: 144 Skipped: 1

Figure 86. Demographics: Age (bar graph)

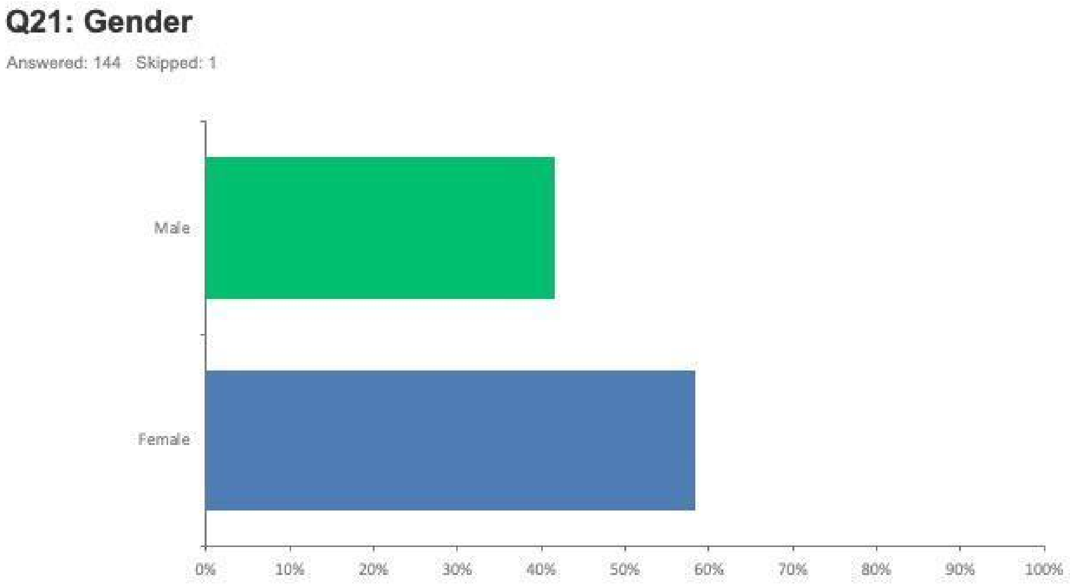


Q19: Age

Answered: 144 Skipped: 1

ANSWER CHOICES	RESPONSES	
< 18	0%	0
18-29	12.50%	18
30-44	18.06%	26
45-60	47.22%	68
> 60	22.22%	32
TOTAL		144

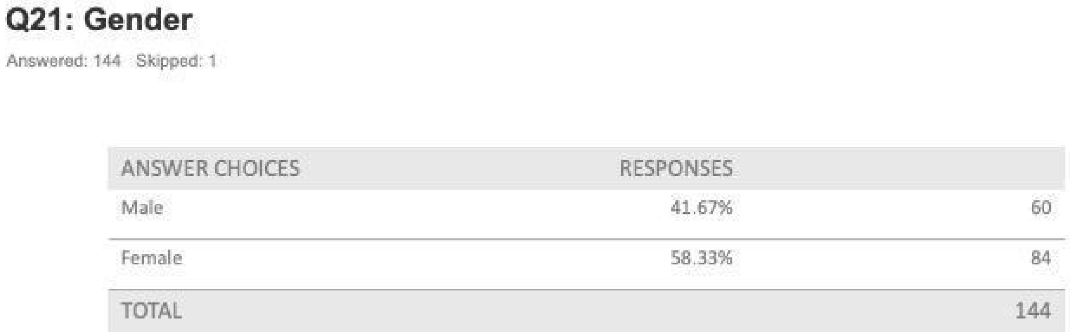
Figure 87. Demographics: Age



Q21: Gender

Answered: 144 Skipped: 1

Figure 88. Demographics: Gender (bar graph)



Q21: Gender

Answered: 144 Skipped: 1

ANSWER CHOICES	RESPONSES	
Male	41.67%	60
Female	58.33%	84
TOTAL		144

Figure 89. Demographics: Age

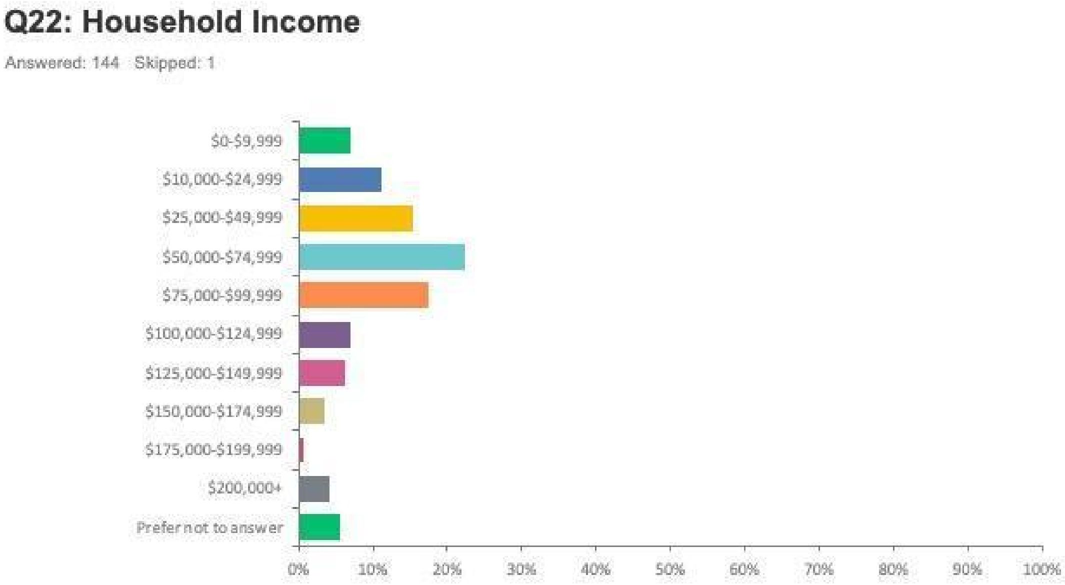
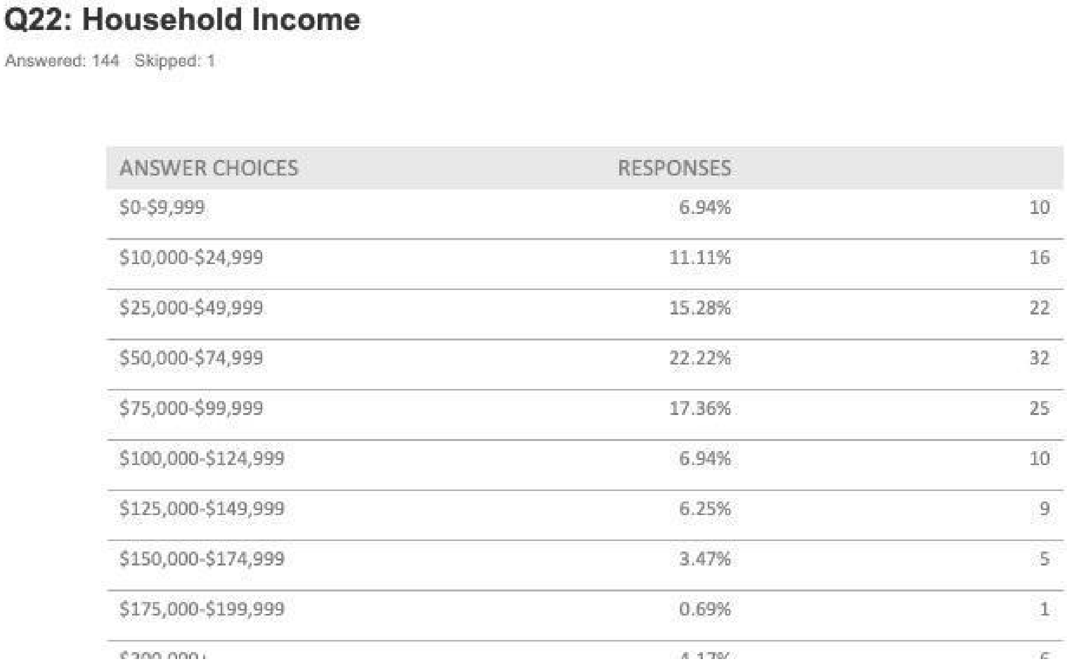


Figure 90. Demographics: Household income (bar graph)



Q22: Household Income

Answered: 144 Skipped: 1

ANSWER CHOICES	RESPONSES	
$0-$9,999	6.94%	10
$10,000-$24,999	11.11%	16
$25,000-$49,999	15.28%	22
$50,000-$74,999	22.22%	32
$75,000-$99,999	17.36%	25
$100,000-$124,999	6.94%	10
$125,000-$149,999	6.25%	9
$150,000-$174,999	3.47%	5
$175,000-$199,999	0.69%	1
$200,000+	4.17%	6

Figure 91. Demographics: Household income

Appendix E: Word cloud from the interviews

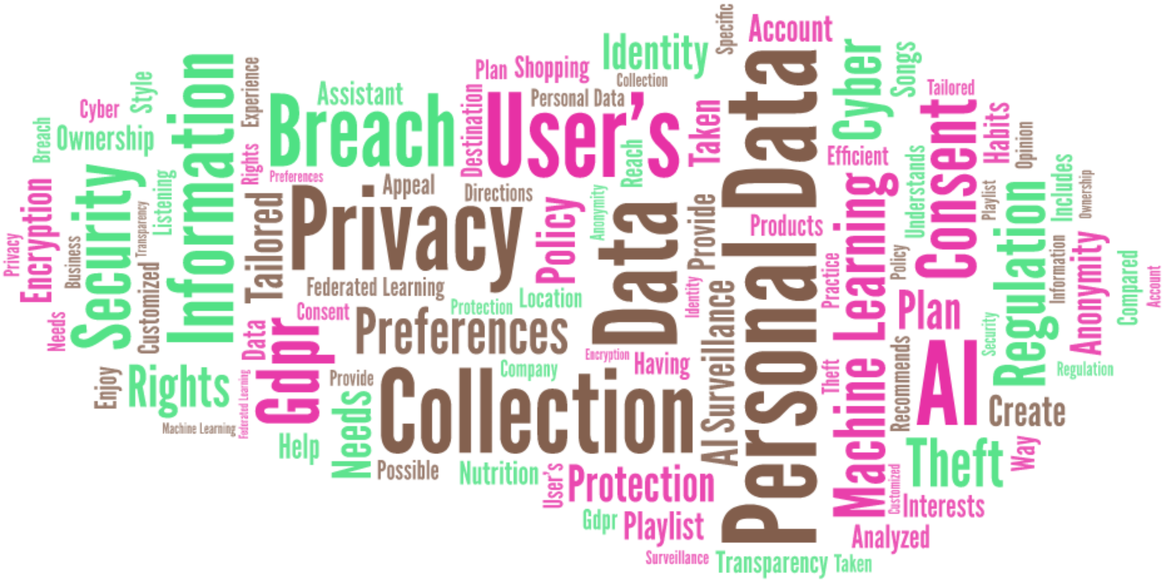


Figure 92. Word cloud from interviews