Navigating the Risks of Dark Data
An Investigation into Personal Safety

Author: Anshu Gautam
Supervisor: Despoina Fyntanoglou, Elissavet Kartaloglou
Date: 09/25/2023
Subject: Information System
Course Code: 4IK50E, 15credits
Level: Master
Department of Informatics
ABSTRACT

With the exponential proliferation of data, there has been a surge in data generation from diverse sources, including social media platforms, websites, mobile devices, and sensors. However, not all data is readily visible or accessible to the public, leading to the emergence of the concept known as "dark data." This type of data can exist in structured or unstructured formats and can be stored in various repositories, such as databases, log files, and backups. The reasons behind data being classified as "dark" can vary, encompassing factors such as limited awareness, insufficient resources or tools for data analysis, or a perception of irrelevance to current business operations.

This research employs a qualitative research methodology incorporating audio/video recordings and personal interviews to gather data, aiming to gain insights into individuals' understanding of the risks associated with dark data and their behaviors concerning the sharing of personal information online. Through the thematic analysis of the collected data, patterns and trends in individuals' risk perceptions regarding dark data become evident. The findings of this study illuminate the multiple dimensions of individuals' risk perceptions and their influence on attitudes towards sharing personal information in online contexts. These insights provide valuable understanding of the factors that shape individuals' decisions concerning data privacy and security in the digital era. By contributing to the existing body of knowledge, this research offers a deeper comprehension of the interplay between dark data risks, individuals' perceptions, and their behaviors pertaining to online information sharing. The implications of this study can inform the development of strategies and interventions aimed at fostering informed decision-making and ensuring personal safety in an increasingly data-centric world.

**Keywords:** Dark data, Hidden data, Big data, Unstructured data, Missing data, Privacy, Cybersecurity, Personal data, Data storage, Consumer protection
ACKNOWLEDGEMENTS

First and foremost, I extend my heartfelt gratitude to my esteemed thesis supervisors, Elissavet Kartaroglou and Despoina Fyntanoglou, whose unwavering support has been instrumental at every stage of this transformative journey. Their expertise, constructive feedback, and unwavering dedication have been a constant source of inspiration and motivation.

I would also like to extend a special acknowledgement to the esteemed Prof. Anita Mirijamdotter for her unwavering interest in this captivating topic and for wholeheartedly encouraging me to embark on this scholarly exploration.

Furthermore, my deepest appreciation goes out to all the remarkable participants who willingly became part of this study, generously sharing their perspectives and enriching this research with their valuable insights. Their contributions have been indispensable in shaping the outcomes of this investigation.

Lastly, I would like to express my profound gratitude to my cherished family and friends, whose unwavering support and encouragement have been the cornerstone of my academic pursuit. Their love, understanding, and presence have been an invaluable source of strength and inspiration, reminding me that a vibrant and supportive network of individuals is an invaluable asset for any researcher on their scholarly voyage.
## CONTENTS

1. **INTRODUCTION** .................................................................................................................. 7  
   1.1 Background ......................................................................................................................... 7  
   1.2 Related Studies .................................................................................................................... 9  
   1.3 Purpose Statement and Research Questions .......................................................................... 11  
   1.4 Topic Justification .............................................................................................................. 11  
   1.5 Scope & Limitations ........................................................................................................... 12  
   1.6 Thesis Organisation ........................................................................................................... 12

2. **LITERATURE REVIEW** ..................................................................................................... 14  
   2.1 Search Strategy .................................................................................................................. 14  
   2.2 Dark Data .......................................................................................................................... 14  
      2.2.1 Dark Data versus Other Data .......................................................................................... 15  
      2.2.2 Dark Data Sources ........................................................................................................ 16  
      2.2.3 Current state of Dark Data ............................................................................................ 17  
      2.2.4 Threat to Privacy and Safety ......................................................................................... 18  
      2.2.5 Mitigating the Risks ....................................................................................................... 19  
   2.3 Theoretical Framework ..................................................................................................... 19

3. **METHODOLOGY** ................................................................................................................. 21  
   3.1 Paradigm Approach ........................................................................................................... 21  
   3.2 Methodological Approach ................................................................................................. 22  
   3.3 Data Collection .................................................................................................................. 22  
   3.4 Interview Process ............................................................................................................... 23  
   3.5 Data Analysis ..................................................................................................................... 24  
   3.6 Reliability and Validity ...................................................................................................... 26  
   3.7 Ethical Considerations ........................................................................................................ 26

4. **EMPIRICAL FINDINGS** ..................................................................................................... 27  
   4.1 Themes ............................................................................................................................... 27  
      4.1.1 Dark data awareness ...................................................................................................... 27  
      4.1.2 Benefits of data sharing ............................................................................................... 28  
      4.1.3 Dark data risks .............................................................................................................. 29  
      4.1.4 Personal data protection .............................................................................................. 30  
      4.1.5 Accountability of Organizations .................................................................................... 32  
      4.1.6 Empowering Data Literacy ........................................................................................... 33  
   4.2 Summary of the Findings .................................................................................................. 34

5. **DISCUSSION** ...................................................................................................................... 35  
   5.1 Discussion on the Research questions ............................................................................... 35  
   5.2 Findings corresponding to the themes and related studies .............................................. 35  
      5.2.1 Dark data awareness ..................................................................................................... 35  
      5.2.2 Benefits of data sharing ............................................................................................... 36  
      5.2.3 Dark data risks .............................................................................................................. 36  
      5.2.4 Personal data protection .............................................................................................. 37  
      5.2.5 Accountability of Organizations .................................................................................... 38  
      5.2.6 Empowering Data Literacy ........................................................................................... 38  
   5.3 Insights from the Theoretical Framework Perspective .................................................... 38
6. CONCLUSION .............................................................................................................. 40
   6.1 Conclusion ........................................................................................................... 40
   6.2 Contribution ....................................................................................................... 41
   6.3 Future Research ................................................................................................. 42

References .................................................................................................................. 43

APPENDICES ............................................................................................................ 47
Appendix A ................................................................................................................. 47
Appendix B .................................................................................................................. 48
Appendix C .................................................................................................................. 49
List of Figures

Figure 1 – Thesis Structure ................................................................. 13

Figure 2 - Social Exchange Theory Framework ........................................ 20

List of Tables

Table 1 – Participant details ................................................................. 23

Table 2 - Summary of interview findings per theme .................................... 34

Table 3 - Research questions and correlating Themes ................................ 35

List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>BITS</td>
<td>Birla Institute of Technology and Sciences, Pilani</td>
</tr>
<tr>
<td>IDC</td>
<td>International Data Coorporation</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>IS</td>
<td>Information Systems</td>
</tr>
<tr>
<td>NGO</td>
<td>Non-Governmental Organization</td>
</tr>
<tr>
<td>OSN</td>
<td>Online Social Networks</td>
</tr>
<tr>
<td>PDA</td>
<td>Personal Data Attitude</td>
</tr>
<tr>
<td>SET</td>
<td>Social Exchange Theory</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

In this chapter, the groundwork for the master's thesis is laid. It starts by explaining the background and research setting and what has been stated in previous research. Continuing with the purpose statement and the research questions. Followed by the topic justification, scope and limitations as well as responsibility of the work. Lastly, a thesis organization is presented where the outline of the thesis is explained.

1.1 Background

The use of social media, internet of things (IoT) devices, and other digital technologies, has become pervasive in today's society which is generating large amounts of dark data (Umair et al., 2017). Today billions of users worldwide are actively participating in online social networks (OSN). These platforms offer opportunities for individuals to connect, communicate, and share personal information with others. Social media platforms collect vast amounts of data from users, including profile information, posts, comments, likes, shares, and browsing behavior (Boyd & Ellison, 2007). The amount of data generated and stored is growing at an exponential rate. According to a recent study by the research firm IDC (International Data Corporation) in collaboration with Seagate, the global data volume is predicted to reach 175 zettabytes by the year 2025 (Reinsel et al., 2018). Dark data, a significant portion of the total data generated, represents the information assets collected and stored by organizations in the course of routine business operations, undergoing collection, processing, and storage. However, these data sets are typically left untapped for alternative purposes such as analytics, fostering business relationships, or direct monetization (Moumeni et al., 2021).

Dark data also presents significant opportunities for organizations and individuals who are seeking to unlock the value of this data (Hand, 2022). While it can be employed for a range of lawful objectives, there also exist inherent risks to personal safety associated with dark data, such as social engineering attacks (phishing, spamming, online stalking etc.) (Adewole et al., 2017). The data might encompass delicate or personal information, including social security numbers, credit card details, or health records, which cybercriminals could exploit for financial motives or identity theft. Additionally, dark or unstructured data, like images, videos, or audio, can have a mix of information from different sources. For instance, a video may have subtitles or audio dubbing, and images may have tags and descriptions. This additional information is meant to provide more details about the object in the data, like the video (Wang et al., 2022). However, attackers can use this extra knowledge to figure out more information, such as a person's identity, location, social connections, and other sensitive details. For example, if an image has GPS information or text description, it can reveal personal information.

The data generated by these technologies often contains sensitive information about individuals, such as the data from deleted posts or accounts, hidden metadata, inferred or predicted data, and data collected from third-party sources. This kind of dark data can be used for various purposes, including user profiling, behavioral tracking, and influencing user behavior. This can result in individuals being subjected to intrusive and personalized ads or content that may compromise their privacy and personal safety. Moreover, data profiling can also lead to discrimination, exclusion, or stigmatization based on individuals' personal characteristics or behaviors, which may have negative consequences for their personal safety and well-being (Wu et al., 2011). Data breaches of social media platforms have resulted in the
exposure of users' personal data, such as names, email addresses, and even passwords, leading to identity theft, financial fraud, and other malicious activities (Rosen, 2018).

Additionally, cyberbullying and harassment can be facilitated by dark data when it is used to identify and target individuals for malicious activities. For instance, cyberbullies can leverage dark data to gather personal information about their targets, such as their location, interests, or contacts, and use it to harass, intimidate, or blackmail them (Patchin & Hinduja, 2006). Furthermore, unintended data sharing, such as sharing of sensitive or embarrassing information without individuals' knowledge or consent, can occur through dark data, leading to reputation damage or emotional distress (Baccarella et al., 2018).

Dark data, generated through online activity, networking websites or applications, can potentially lead to serious criminal activities as well that pose physical harm to individuals. For instance, it can be used for stalking and harassment, where geolocation data or personal information collected from social media can be used to track an individual's movements, leading to physical harassment or assault. Dark data can also reveal personal routines, habits, and possessions, making individuals vulnerable to physical assault or robbery. Moreover, dark data that includes personal information can be used for identity theft and fraud, where criminals can impersonate individuals and commit financial crimes. Dark data can also be used for extortion and blackmail, where compromising information can be used to extort money or favors from individuals. These criminal activities facilitated by dark data can result in physical harm, financial loss, reputational damage, and emotional distress (Wu et al., 2011; Baccarella et al., 2018). Therefore, it is crucial for individuals to be cautious about the information they share online and take appropriate measures to protect their personal safety in the digital age.

One area of particular interest is how individuals perceive the risks associated with dark data and how these perceptions influence their attitudes towards sharing personal information online. Understanding these perceptions and attitudes is crucial for informing policy, designing privacy-enhancing technologies, and promoting responsible online behaviors. For instance, research has shown that individuals' privacy concerns, trust in social media platforms, and perceptions of data collection and use influence their information disclosure behaviors and privacy protection strategies (Vitak, 2012). Moreover, perceived risk, including concerns about data security, privacy breaches, and identity theft, has been found to be a significant factor in shaping online purchase decisions (Lăzăroiu et al., 2020).

This research aims to contribute to the existing literature by exploring how individuals navigate the risks associated with dark data, with a focus on their perceptions, attitudes, and behaviors. By examining these factors, this research advances the understanding on the multifaceted nature of dark data, its implications for personal safety, and the ways in which individuals manage and respond to these risks in the context of online activity in general.

1.2 Related Studies

Previous research on the topic of how dark data can lead to criminal activities and physical harm to individuals has explored various aspects of this issue, utilizing different methodologies and sample populations.
The research conducted by Thomas et al. (2017) examined the risks associated with stolen credentials, specifically data breaches, phishing attacks, and malware, and how these risks can lead to criminal activities and harm individuals. The authors used a quantitative analysis approach, analyzing data from multiple sources, including public breach data, phishing kits, and malware samples. They employed various statistical techniques, such as clustering algorithms, machine learning, and network analysis, to analyze the data. The study did not involve human participants directly, as it relied on analysis of data from multiple sources related to data breaches, phishing attacks, and malware. The results of the study revealed that stolen credentials from data breaches were frequently used for criminal activities, such as account takeover, identity theft, and financial fraud. Phishing attacks were identified as a common method used to steal credentials, and the study identified different types of phishing attacks and their effectiveness. Malware, including keyloggers and remote access Trojans, were also identified as significant risks associated with stolen credentials. The findings highlighted that regrettably, a significant number of users remain unaware of the potential security risks associated with sharing their data while communicating on social media. These risks include privacy breaches (Mislove et al., 2010; Boshmaf et al., 2011), identity theft (Bilge et al., 2009), malware attacks (Baltazar et al., 2009), fake profiles (commonly known as sybils) (Cao et al., 2012; Stringhini et al., 2013) or socialbots (Boshmaf et al., 2011; Boshmaf et al., 2013; Elbashar et al., 2013), and instances of sexual harassment (Ybarra & Mitchell, 2008; Wolak et al., 2010) among others.

The research conducted by Fire et al. (2014) focused on the threats posed by OSN and proposed solutions to mitigate these risks. The authors used a review and analysis approach, where they examined existing literature, research studies, and industry reports related to online social networks and their associated risks. They critically analyzed the findings from these sources to identify common threats, vulnerabilities, and potential solutions. The research did not involve human participants directly, as it was a literature review and analysis study. The authors reviewed and analyzed a wide range of research studies, reports, and literature related to online social networks and their associated risks. The results of the study highlighted various threats posed by online social networks, including identity theft, privacy breaches, cyberbullying, social engineering attacks, and spreading of misinformation. The authors also identified potential solutions to mitigate these risks, such as improving privacy settings, user education and awareness, regulation and policy enforcement, and technological solutions. The study emphasized the need for a multi-pronged approach involving users, service providers, regulators, and policymakers to address the risks associated with online social networks effectively.

In another study conducted by Dwyer et al. (2007) shed light on the trusting nature of Facebook and MySpace users, who exhibit confidence not only in the social networking platforms themselves but also in other users within these networks. This trust fosters a culture of information sharing and relationship building. Furthermore, recent research (Boshmaf et al., 2011) has revealed that a significant number of users on OSN disclose personal and intimate details about themselves, their friends, and their relationships. Such information may be shared through photos, or directly provided in the form of home addresses and phone numbers, underscoring the openness and vulnerability of OSN users.

Moreover, as revealed by Elbashar et al. (2013), Facebook users have been found to readily accept friend requests from individuals whom they do not personally know, but who share mutual friends. Little do they know that by accepting these requests, they unwittingly expose their private information to complete strangers. This information can potentially be exploited
for malicious purposes, causing harm not only in the virtual realm but also in the physical world. The stakes are even higher when it comes to young children and teenagers, who are inherently more susceptible and exposed, heightening the risks associated with such online behavior.

The study done by Chugh & Guggisberg (2020), analyzed the popular psychological thriller series You on Netflix, which depicts stalking and dating violence through social media. The researchers used a qualitative content analysis approach to identify themes and messages throughout the 10 episodes of the first season. Three overarching themes emerged: erotic stalking, controlling behavior in dating relationships, and a lack of understanding about cyber safety. The series highlights concerning behaviors, such as stalking, coercive control, violence, and questionable masculine attitudes, and suggests that the dangers of oversharing of the personal data online and cyber safety are often overlooked. The study found that the series is an excellent educational tool for raising awareness about online safety and security in the context of the personal data shared online. The study recommends that educating community members and professionals to enhance knowledge and understanding of partner stalking is required to address stalking and other forms of dating violence. The study also highlights the importance of controlling privacy settings and the responsibility of social media platforms to simplify these settings to protect vulnerable young women in particular from being exploited and abused.

In the study conducted by (Addae et al., 2017), they aimed to create a tool, called Personal Data Attitude (PDA) scale, to understand how people feel about their personal information. Researchers looked at different aspects to make sure the tool is reliable and useful. They developed the scale by trying different things and checked if it could show how people think about their personal data. The study also looked at information about privacy and personal data rules. The results of the study showed that people feel differently about their personal data depending on what kind of data it is. For example, some people feel that health information is more private and important than information on social media. The study also found that when people are worried about their privacy, they also think more about the value and security of their personal data. The researchers think that the tool they made can help in understanding how people think about their personal data and how they try to protect it. This could help in making things like online security better in the future. Overall, this study helps to learn more about how people see their personal data and can be useful for making the internet safer.

In another study conducted by Raman et al., (2015) discusses the concerns about privacy during online shopping, many users do not take adequate measures to protect their personal information. It also explores their perceptions of online privacy during the shopping experience. The study found that IT graduates exhibited greater awareness of privacy issues compared to non-IT graduates, but no significant gender differences were observed. The study underscores the need to raise awareness among young online shoppers about privacy issues, given the increasing use of e-commerce in India. However, in their study, Jiang, Yang & Jun (2013) highlighted the significance of convenience as a primary motivator for the adoption of online shopping. They identified five key dimensions of online shopping convenience. These dimensions encompass: access, search, evaluation, transaction convenience and post-purchase experience. Search convenience specifically focuses on user-friendly websites, a variety of search options, and the ability to quickly locate desired products. Additionally, in the research by Aldas-Manzano et al. (2011) the findings underscore the pivotal roles of satisfaction with banking websites, trust, and perceived risk as vital factors in fostering loyalty to online
banking services. Importantly, the study reveals that an individual's loyalty to a banking website is closely tied to their levels of trust and perceived risk. However, as the perception of risk increases, the correlation between satisfaction and loyalty diminishes, with risk perception emerging as the most prominent inhibitor of loyalty in this analysis.

In their paper, the authors (Hande & Mane, 2015) studied how to keep cloud data secure and analyzed different aspects like making sure the data is accurate and private, and how the data is used. They talked about security issues, like when many people use the same cloud service, and explained models to keep data safe. The paper also talked about two important things: security and accountability. Security is about keeping data safe from threats. With cloud computing, the service provider is trusted to manage security. But that means control over the data is lost. Also, since data is spread across different locations, it needs methods to manage it correctly. Whereas, accountability means making sure that the data is used properly and transparently. If the data is altered in some way, there should be a record of it. This helps build trust between the user and the cloud service provider. Accountability also helps providers give control over how the data is utilized. However, in another study (Bertino et al., 2019) explained that comprehensive data transparency processes are crucial for investigating and preventing unfair data usage. Data transparency refers to the ability of individuals to access all information about data used in processes and decisions that impact them. It involves disclosing not only the data itself and its purposes but also metadata like collection methods. However, most organizations don't offer mechanisms for users to access detailed information about how specific data is actually used.

As a deeper exploration is conducted into the world of data and online privacy, it becomes evident from the sources above that while many people are familiar with the concept of big data, the term 'dark data' still remains unknown to a large number of individuals. Hence it is crucial to understand the risks associated with dark data and the steps that can be taken to protect privacy during any digital activity. Consequently, there exists a desire to gain insight into the general public's perception related to dark data and their awareness of the steps that can be taken to protect themselves.

1.3 Purpose Statement and Research Questions

This study aims to explore individuals' perceptions concerning the risks linked to dark data, encompassing privacy infringements, identity theft, cyberstalking, and other potential hazards to personal safety and security. This qualitative study targets university students, encompassing both males and females, with the objective of gaining insight into their broader online experiences. The research delves extensively into participants' attitudes regarding the online sharing of personal information. It seeks to elucidate the types of information they are willing to share, their apprehensions regarding privacy and security, and the strategies they employ to navigate their online presence. The questions that are addressed through this master thesis are:

**RQ1:** How do individuals perceive the risks of dark data?

**RQ2:** How do these perceptions influence their attitudes towards sharing personal information online?
1.4 Topic Justification

In today’s digital world, people are generating huge amounts of data by using the internet. This data is being generated through multiple channels, such as, IoT devices, banking/online shopping websites, or Online social networks (OSN). The people using the very popular social networking websites, like, Facebook and Twitter, commonly share free-text formats, such as comments, posts, blogs, and tweets, with their chosen social groups. However, this unstructured data that is ignorantly being shared with other users, may contain sensitive information, including breakups, hospitalization, sexual orientation, political discussions, religious beliefs, employment details, personal opinions and much more (Aghasian et al., 2020). This can lead to significant risks such as online victimization, unintended fame, employment problems, and surveillance. Therefore, understanding people's perception related to dark data and their steps towards online privacy cannot be overstated. With the exponential growth of digital data, organizations and businesses are increasingly relying on dark data to make critical decisions. However, this often results in a breach of personal privacy and security. Therefore, it is crucial to investigate the awareness level of people regarding dark data and online privacy, and identify the steps they take to protect their privacy online. By gaining insight into the public's awareness level and behavior towards dark data, improvement areas where education and awareness efforts are required can be identified (De Los Santos & Klug, 2021). This study seeks to fill the current knowledge gap pertaining to the presence and hazards of Dark Data. Its objective is to bridge this gap by investigating how individuals perceive the risks linked to dark data and how it influences their willingness to share personal information online. The study holds the potential to yield valuable insights that can lead to practical recommendations for strengthening data security. It also underscores the importance of education, accountability, and transparency in tackling the challenges posed by dark data in the digital age.

1.5 Scope & Limitations

The scope of this study is to delve into the level of public awareness concerning the risks associated with dark data and the measures they adopt to safeguard their online privacy. This study expands upon existing knowledge on how dark data generated from online activity can affect individual privacy and safety. Furthermore, it captures the viewpoints and suggestions of these individuals regarding initiatives aimed at raising awareness on this subject.

This study employs the convenience sampling approach. Convenience sampling was chosen due to its practicality and accessibility, particularly when targeting specific groups such as university students. This method allows to efficiently gather data from a diverse set of participants within the confines of this research constraints (Schneider et al., 2012). The sample in this study sample primarily consists of university students both males and females, as they were readily accessible for participation. While convenience sampling offers advantages in terms of ease of recruitment, it’s essential to acknowledge its limitations, including the potential for selection bias and limited generalizability.

The limitations of this study include the fact that it is based on a qualitative study, which tend to focus on small sample sizes, and while they provide rich, detailed information, they may not be representative of the larger population. The study is also limited by the honesty and accuracy of the responses provided by the participants. Finally, the study only explores the perception of a small set of students and does not delve into the perception of experts or
professionals in the field of data privacy. However, despite these limitations, qualitative research is valuable in providing in-depth understanding of individuals' perspectives and experiences, which can be used to inform policy and practice.

1.6 Thesis Organisation

This thesis is organized into six chapters, as illustrated in Figure 1. Each chapter is subdivided into sections to facilitate a coherent reading experience. The second chapter, titled "Literature Review," provides the theoretical underpinnings of this thesis, including an exposition of the theoretical framework employed. Chapter three, "Methodology," delineates the research's nature, elucidates the data collection process, and expounds upon the data analysis methodology. The fourth chapter, "Empirical Findings," presents the data acquired from the study participants. Chapter five, "Discussion," critically examines and interprets the empirical findings. Finally, chapter six, "Conclusion," serves as the culmination of the thesis, encapsulating the study's outcomes, implications and limitations, while also suggesting avenues for future research.

![Figure 1: Thesis Structure](image)
2. LITERATURE REVIEW

This chapter provides a comprehensive literature review on the research topic, focusing on Dark data. It presents an overview of Dark data, including its sources and how it differs from other types of data. The chapter also examines the current state of Dark data and explores its potential threats to personal safety. Furthermore, it discusses strategies to mitigate the risks associated with Dark data. Finally, the chapter discusses the theoretical framework applicable to this specific research scenario.

2.1 Search strategy

The literature search strategy used for constructing the literature review on this topic involves multiple steps. The first step involved identifying relevant databases and search engines such as Scopus, Google Scholar, JSTOR, and Science Direct. Keywords such as "dark data," "unstructured data," "personal safety," "social media data," and "big data" were used to search for articles, research papers, and reports relevant to the topic.

Next, very clear inclusion and exclusion criteria were created with the goal of limiting the review to only relevant and recent studies. The criteria included a publication date of 2007 or later, peer-reviewed articles in English, and studies focused on the impact of dark data on privacy and personal safety. Once the initial search was completed, a comprehensive review of the abstracts and full-text articles was conducted to ensure that they met the inclusion criteria and to identify any additional relevant studies that were not found in the initial search. The collected data was then assessed and sorted as per the research queries, topics and aims of the study.

Additionally, a hand search was conducted by reviewing the reference lists of the selected articles to identify any additional studies that may have been missed in the initial search. This was followed by a citation search to identify any studies that have cited the selected articles, thereby ensuring that the literature review was comprehensive and up-to-date. The literature search strategy used for the construction of the literature review on the chosen topic was systematic and thorough, ensuring that only relevant and up-to-date studies were included in the review.

2.2 Dark Data

Dark data, which refers to data that is not systematically organized and often goes unused and unrecorded, is a largely overlooked and underutilized resource that has the potential to offer many opportunities (Heidorn, 2008). Much of the existing but unused data generated by the IoT devices remains "dark," offering a wealth of data-driven opportunities for businesses to capitalize on. The data that is unknown can be just as significant as the data that is known if there is an aspiration to make accurate conclusions (Hand, 2022). Dark data also presents significant opportunities for organizations and individuals who are seeking to unlock the value of this data. For example, research has shown that organizations can gain valuable insights into their operations and customers by analyzing dark data (Gandomi & Haider, 2015). This can help organizations make informed decisions, improve their operations, and increase their competitiveness. In addition, individuals can also benefit from the value of dark data. By analyzing and utilizing their personal data, individuals can gain greater control over
their personal information and use this information to make informed decisions about their health, finances, and personal relationships.

To take advantage of dark data, businesses must first increase awareness of it within their organizations and allocate the necessary resources to exploit it. However, even when organizations are aware of the data being generated, they may struggle to understand its practical applications. Many companies understand that this information is important, but are uncertain about how to use it effectively. This can be due to factors such as a resistance to change, outdated beliefs about how to run the business, or a lack of resources for exploratory data analysis as a result of prioritizing current operations and meeting customer demands (Gimpel, 2020a).

Despite the numerous opportunities that remain unexplored the presence of dark data creates privacy and security risks, as sensitive information may be left unsecured.

2.2.1 Dark Data versus Other Data

Dark data differs from all other kinds of data in its unique characteristics and attributes. It is essentially a subset of big data that remains unutilized, unstructured, and unanalyzed due to its complexity, volume, or lack of a clear use case. This data is generated by various sources, including social media, web traffic, customer transactions, machine logs, and IoT devices, among others.

One of the key differences between dark data and other types of data is that it is typically unstructured, meaning that it lacks a predefined data model or schema. Unlike structured data, which is organized and stored in a specific format, dark data can exist in many forms, such as text, images, videos, or audio files (Wang et al., 2022). This makes it challenging to process and analyze this data using traditional database management tools or analytics techniques. Another significant difference between dark data and other data is its sheer volume. It is estimated that dark data accounts for up to 90% of all data generated by organizations. This represents a vast amount of untapped potential for insights, cost savings, and competitive advantages. However, the sheer volume of dark data can also be overwhelming, leading to difficulties in storage, processing, and analysis.

Dark data is also characterized by its ambiguous value proposition. Unlike structured data, which has a clear use case, dark data often lacks a defined purpose or immediate value. It is generated and collected in the background, without a clear intention of how it will be utilized in the future. However, this does not mean that it is useless data. On the contrary, dark data can provide valuable insights and inform business decisions if properly analyzed and processed (Hand, 2022). Lastly, dark data is often associated with risks and concerns related to privacy, security, and compliance. Since it is unutilized, it is also often left unsecured, creating vulnerabilities and potential opportunities for malicious actors to exploit. Additionally, as more data protection regulations are enacted worldwide, organizations must be careful when processing and analyzing dark data, as it may contain sensitive or personal information.

Overall, dark data differs significantly from other types of data due to its unstructured nature, vast volume, ambiguous value proposition, and associated risks. As organizations continue to generate and collect data, they must prioritize exploring and analyzing dark data to leverage its potential value while mitigating the risks.
2.2.2 Dark Data Sources

Often unstructured or semi-structured data, such as social media posts, emails, images, videos, and other digital content that remains unused falls under the category of Dark data. The sources of dark data are diverse (Moumeni et al., 2021) and can include both internal and external sources. According to Moumeni et al. (2021) these sources in detail are:

- **Internal Sources**

  Backups: Organizations often make regular backups of their data to prevent loss in case of system failures or cyber-attacks. These backups often contain outdated, redundant, or irrelevant data that is never used, but can still be considered dark data.

  Archives: Archived data refers to information that is stored for long-term preservation. This data is often not actively used or analyzed, but can still contain valuable insights. Examples of archives include email archives, log files, and server backups.

  Legacy Systems: Legacy systems refer to outdated software or hardware that is no longer in use, but still contains data that is not migrated to newer systems. This data may not be compatible with current systems or require significant effort to migrate, and is often left untouched.

  Application Logs: Application logs refer to the records of events that occur within software applications. These logs can contain information such as user actions, system errors, and performance metrics. Although they are often not used for analysis, they can provide valuable insights into application usage and performance.

- **External Sources**

  Social Media: Social media platforms generate vast amounts of data, including posts, comments, likes, and shares. This data is often unstructured and difficult to analyze, but can provide valuable insights into customer behavior, preferences, and sentiment.

  Public Records: Public records include government records, court records, and other publicly available information. This data can be useful for background checks, due diligence, and research, but is often not analyzed in depth.

  Sensor Data: Sensor data refers to data collected by sensors such as temperature sensors, motion sensors, and GPS sensors. This data is often collected in large volumes and requires advanced analytics to extract insights.

  Web Data: Web data refers to information collected from web pages, such as HTML pages, XML files, and other structured or unstructured content. This data can be used for web scraping, search engine optimization, and market research.

Organizations must be aware of the sources of dark data and take steps to manage and analyze this data effectively to derive valuable insights. By doing so, organizations can gain a competitive advantage, identify new opportunities, and improve their decision-making processes.
2.2.3 Current state of Dark Data

Dark data, which is frequently created but not examined or utilized, is a result of technological constraints or organizational shortcomings. With the surge in data production from digital devices and the IoT devices, the amount of dark data is rapidly growing. Questions regarding how to manage data stop many businesses from capitalizing on their dark data. Concerns about data security, cyber threats, customer approval of data use, legal uncertainty around privacy policies present a massive barrier to success (Gimpel & Alter, 2020b).

Currently, the amount of dark data is estimated to be several times larger than the amount of structured data, and its growth is projected to continue at a rapid pace. The sheer volume of data being generated by people and machines, combined with the fact that it is often stored in multiple, disconnected repositories, makes it increasingly difficult to manage, analyze, and make sense of. According to a 2019 survey conducted by Splunk, a data analytics company, nearly two-thirds of organizations believe that at least half of their data is "dark" or "unclassified" (Moumeni et al., 2021). One of the biggest challenges with dark data is that it is often difficult to identify and access. This is because it is usually stored in disparate locations and formats, including legacy systems, file shares, and cloud services. Furthermore, much of this data is unstructured, meaning that it lacks a consistent format or structure that makes it easy to analyze using traditional tools. Despite these challenges, there are a number of technologies and techniques that can be used to unlock the value of dark data. These include machine learning, natural language processing, and other forms of artificial intelligence (AI), which can help to identify patterns and insights in unstructured data. Other tools, such as data catalogs and data lakes, can help to organize and manage dark data in a way that makes it more accessible and easier to analyze.

The potential benefits of unlocking dark data are significant. By gaining insights into previously untapped data sources, organizations can gain a competitive advantage, improve operational efficiency, and make better decisions (Gimpel, 2020a). For example, analyzing social media data can provide insights into customer sentiment, while analyzing sensor data from industrial equipment can help to identify maintenance issues before they become critical. Despite the potential benefits of unlocking dark data, many organizations struggle to do so effectively. This is due in part to the challenges of identifying and accessing dark data, but also because of the skills and expertise required to analyze it. To address these challenges, organizations need to invest in the right tools and technologies, as well as in the skills and expertise of their employees. This may involve hiring data scientists and other data specialists, or partnering with outside firms that can help to unlock the value of dark data. In conclusion, the current state of dark data is one of rapid growth and increasing complexity. As the amount of unstructured data continues to expand, organizations are facing significant challenges in identifying, accessing, and analyzing it (Gimpel & Alter, 2021). However, with the right tools and technologies, organizations can unlock the value of dark data and gain insights that were previously unavailable.

2.2.4 Threat to Privacy and Safety

Dark data, being the information that is collected but not used, poses a significant threat to personal safety in today's world. Dark data generated from online activity can potentially be used to perform crimes against ordinary people. Dark data generated in the digital space from social media, online shopping, internet banking etc. can include personal information such as
names, addresses, phone numbers, dates of birth, relationship status, expected location and other sensitive information. The potential risk to personal safety comes from the fact that this data can be used by malicious actors for various purposes, ranging from identity theft, fraud, cyberbullying and other crimes (Umair et al., 2017). One of the most significant threats from dark data comes from the information collected by social media platforms. These platforms collect a large amount of data about their users, including their location, search history, and the content they interact with. This data is often stored indefinitely, and users have little control over how it is used. Malicious actors can exploit this data to target users with phishing scams or other forms of cyber attacks. As discussed, in addition to social media data, dark data can also come from various other sources such as online shopping websites, fitness trackers, and other IoT devices. These devices collect vast amounts of data about individuals, including their daily routines, medical history, and personal habits. With the help of personal information that has been obtained through dark data, criminals can create fake identities, open bank accounts, and take out loans in someone else's name, causing significant financial harm to the victim.

Another example is physical theft. Criminals can use dark data to determine when someone is likely to be away from their home, which can make them a target for burglaries or other forms of theft. In some cases, this information can even be used to gain access to a victim's home, potentially putting them in danger. There is also the risk of stalking and harassment. If a perpetrator gains access to someone's personal information through dark data, they can use it to commit serious crimes such as stalking or even murder (Baccarella et al., 2018). For instance, information gathered from social media, online shopping or browsing history, and other online activities can reveal a person's location, daily routines, and personal habits. This data can be used to stalk or harass someone, or even plan a physical attack. This can be extremely distressing for the victim and can have a significant impact on their mental health and physical safety.

Moreover, dark data can also be used to commit fraud against individuals. For example, criminals may use personal information obtained through dark data to apply for credit cards or loans in the victim's name, leaving the victim with the debt. Overall, the potential for offline crimes using dark data highlights the importance of taking steps to protect personal information and online privacy (Ghaiumy Anaraky et al., 2019). With the ever-growing prevalence of dark data generated from social media accounts, it is essential to remain vigilant and take necessary precautions to protect your personal information. Awareness of these risks is vital. This can include using strong passwords, regularly monitoring accounts for suspicious activity, and being cautious about the information shared on social media platforms.

### 2.2.5 Mitigating the Risks

Organizations can mitigate the risks posed by dark data in terms of personal safety by implementing various strategies and best practices. As per the research and consulting firm Gartner Inc., one of the key strategies is to have a proper data management plan in place. This includes regular monitoring and reviewing of the data being collected, storage and usage policies, and ensuring that the data is being processed and stored in a secure manner (Panetta, 2019). Protecting sensitive data is an important part of any business, which is why it's essential to use security measures such as encryption, secure data storage, access control and data masking. Implementing security protocols and measures can help protect against data breaches and unauthorized access to confidential information.
It is essential for individuals to be aware of the risks associated with dark data and to take steps to limit the amount of personal information that is available online. Additionally, governments and organizations must also take steps to ensure that personal information is protected and that appropriate laws and regulations are in place to prevent the misuse of dark data. To mitigate the threat posed by dark data, there is an urgent need for better data protection regulations. Such regulations should ensure that users have more control over their data and that companies collecting data are more transparent about their data collection practices (Tarkowski & Vogelezang, 2021). Additionally, there is a need for better awareness and education programs to help individuals understand the risks posed by dark data and how to protect themselves from it.

Additionally, dark data can be employed for multiple purposes like sentiment analysis, risk assessment, and market analysis. This can give organizations an edge over their rivals as well as help them make informed decisions. Utilizing dark data appropriately will undoubtedly benefit them. Organizations must proactively tackle the potential risks of dark data to be able to fully exploit its opportunities. Investing in this decision now could bring great rewards for them later on. Organizations can strategically manage dark data in order to maximize value and reduce risk (Moumeni et al., 2021). This requires proper data retention and disposal planning, as well as the implementation of adequate security measures. When done properly, the organization can benefit from this valuable but often overlooked resource.

2.3 Theoretical Framework

The proposed theoretical framework for this research is informed by the privacy and security literature. This literature provides a useful framework for understanding the risks associated with dark data, as well as the awareness level of individuals about its existence.

When the synthesized literature on this topic is observed from the IS theory point of view, the Social Exchange Theory (SET) seems to be a great fit. This theory proposes that individuals engage in social relationships based on the perceived benefits and costs associated with those relationships. In the context of dark data, individuals may engage in online behavior that generates dark data in exchange for perceived benefits such as access to online services, convenience, or social connection.

SET is a theoretical framework in social psychology that explains human behavior as a result of the exchange of resources or benefits between individuals or groups (Cook et al., 2013). According to this theory, individuals engage in social interactions because they believe that these interactions will result in a net positive gain of benefits or resources (Urbonavicius et al., 2021). These benefits or resources can be material, social, or psychological in nature. The theory suggests that individuals evaluate the outcomes of a social exchange based on three key factors: the perceived benefits of the exchange, the perceived costs of the exchange, and the comparison level, which is the standard against which individuals judge the fairness of the exchange. The comparison level is influenced by past experiences, social norms, and cultural values. Figure 2, depicts SET adjusted in context to this study.
Perceived Benefits correspond to the perceived advantages that individuals associate with sharing their personal information online. These benefits might include personalized services, convenience, and staying connected with others.

Perceived Costs relate to the perceived risks and costs associated with sharing personal information online, such as concerns about privacy breaches, identity theft, cyber threats etc.

The comparison level represents the standard against which individuals judge the fairness of the exchange. This could reflect individuals’ past experiences with online data sharing, adherence to social norms regarding privacy, and their cultural values regarding data protection. It's essential to consider that these factors influence how individuals assess the benefits and costs of online data sharing.

The outcome represents individuals' overall attitudes and behaviors regarding online data sharing. It's influenced by their perceptions of benefits, costs, and the comparison level. If individuals perceive the benefits as outweighing the costs and consider the exchange fair according to their comparison level, they may be more inclined to share personal information online.

By using the Social Exchange Theory framework, it can be better understood how individuals evaluate the risks and benefits of sharing personal information online, which is a fundamental aspect of this research. It provides a structured way to analyze and interpret participants' attitudes and behaviors in the context of data privacy and dark data awareness. SET is particularly relevant to this research study as it allows for an exploration of individuals' attitudes and behaviors towards dark data in the context of personal safety. The theory suggests that individuals engage in social interactions based on a cost-benefit analysis, which can be applied to the decision-making processes related to online behavior and personal data.
sharing (Ghafari et al., 2019). In the context of this study, the theory can help to explain how individuals evaluate the risks and benefits of using online platforms and sharing personal data. Overall, SET provides a useful lens through which to analyze the costs and benefits of online behavior and can inform users about safer online practices.
3. METHODOLOGY

This chapter introduces the philosophical tradition and methodological approach that underpin this study. It provides a detailed description of the procedure for data collection. The chapter proceeds with the analysis of the collected data, followed by a discussion on the reliability and validity of the study. Ethical considerations are also addressed in the concluding section of this chapter.

3.1 Paradigm Approach

In information systems (IS) research, a paradigm refers to a set of beliefs and assumptions that guide the understanding and study of a particular phenomenon. It influences the research design, methodology, data collection, and data analysis methods used in a study (Creswell, 2013). There are three main paradigm approaches in research, namely positivist and interpretivist and Critical realist. These approaches have different ontological, epistemological, and methodological assumptions about the nature of reality, knowledge, and the process of researching it.

The positivist paradigm, as described by Creswell and Creswell, refers to a philosophical framework that views knowledge as objective and based on empirical evidence. This approach seeks to identify cause-and-effect relationships and to develop generalizable laws and theories. In positivist research, the researcher often assumes a detached and neutral stance, seeking to explain the phenomenon under study by controlling for extraneous variables and testing hypotheses through systematic data collection and analysis. The positivist paradigm is often associated with quantitative research methods, such as surveys, experiments, and observational studies (Creswell and Creswell, 2018). In the positivist view, the reality being studied exists objectively, independently of the researcher's interpretation, and can be described and measured through quantitative methods. This paradigm emphasizes the importance of objective, reliable, and valid data, and aims to establish universal laws and generalizations that can be applied to a wider population. The positivist paradigm is also concerned with the reduction of bias and error in research, and seeks to achieve a high degree of accuracy and precision in data collection and analysis.

The positivist paradigm is contrasted with interpretivist paradigm is a philosophical approach that is rooted in the belief that reality is socially constructed. This paradigm asserts that knowledge cannot be gained through objective means, but rather through the interpretation of individual experiences and subjective understandings. In interpretive research, the researcher seeks to understand the meaning that participants attribute to their experiences, behaviors, and actions (Walsham, 2006). The focus is on the subjective experience and how it is shaped by the larger cultural, social, and political context. The interpretivist paradigm is commonly used in the social sciences, particularly in the study of human behavior, culture, and communication. In this type of research, the researcher may use qualitative methods such as in-depth interviews, observation, or focus groups to gather data (Klein & Myers, 1999). The goal of interpretive research is not to test hypotheses or make generalizations, but to achieve a deeper understanding of the lived experiences of individuals and the meaning they attach to those experiences. According to (Creswell and Creswell, 2018), interpretive research is often guided by the principles of constructivism, which holds that reality is constructed through the interaction between individuals and their environment. This interaction is not simply a passive reflection of reality, but an active process of interpretation, meaning-making, and
understanding. In interpretive research, the researcher aims to understand the world from the perspective of the participants and to provide a rich, nuanced description of their experiences.

The third most common paradigm in IS research is the critical realism paradigm, as described by Creswell and Creswell, is a philosophical approach to research that emphasizes the need for a critical examination of the underlying assumptions and biases in the research process. It recognizes that reality is complex and multi-layered, and the understanding of reality is shaped by perceptions, experiences, and beliefs. In this paradigm, the researcher takes a critical and reflexive approach to the research process, exploring not only the phenomenon being studied, but also their own biases, assumptions, and perspectives (Myers & Klein, 2011). Critical realism also recognizes the importance of multiple sources of evidence and the need for triangulation, or the use of multiple methods and perspectives, to provide a more complete and accurate understanding of reality (Creswell and Creswell, 2018). Overall, critical realism provides a nuanced and complex view of reality and encourages researchers to critically examine their own beliefs and assumptions, as well as those of the broader social and cultural context in which the research is being conducted. For the research proposal on Dark data and the risks it poses towards personal safety of the ordinary people, an interpretive paradigm is proposed. This is because the study aims to explore the subjective experiences and perspectives of individuals with regards to dark data. The interpretive paradigm allows a detailed and in-depth exploration of these experiences and perspectives, which is essential for understanding the impact of dark data on personal safety.

When deciding on the research approach, it is essential to consider the paradigm employed as it influences the research design, collection of data and analysis methods used in a study. In this research, the interpretive paradigm is applied because the study aims to explore the subjective experiences and perspectives of individuals with regards to dark data. The interpretive paradigm allows for a detailed and in-depth exploration of these experiences and perspectives, which is essential for understanding the impact of dark data on personal safety.

3.2 Methodological Approach

The two primary methodological approaches to conduct research are qualitative and quantitative research methods. Qualitative research methods are exploratory and aim to understand the underlying reasons, opinions, and motivations of individuals or groups. These methods typically involve collecting data through open-ended interviews, focus groups, observations, and other non-standardized data collection methods (Elo & Kyngäs, 2008).

Quantitative research methods, on the other hand, are structured and aim to measure numerical data and test hypotheses. These methods typically involve collecting data through surveys, experiments, and other standardized data collection methods. The data is then analyzed using statistical techniques to identify patterns and relationships between variables.

This research employs a qualitative research approach to investigate the perceptions of individuals, including both male and female students, regarding the concept of dark data as collected from their online activities. The selection of qualitative research methodology is deemed suitable for this study due to its capacity to conduct an in-depth and thorough exploration of the subjective experiences and perspectives of individuals concerning dark data.
3.3 Data Collection

The choice of participants in this study is essential for several reasons. First and foremost, the selected qualitative interpretivist approach aims to understand individuals' perceptions and interpret their responses effectively. To achieve this, it is crucial to engage participants who can provide rich and nuanced insights into the subject matter (Denscombe, 2017). Additionally, the selection of participants aligns well with the research objective, which is to comprehensively grasp students' perspectives regarding dark data. Therefore, students and library visitors from diverse backgrounds were chosen as suitable participants. Their varied experiences and viewpoints make them valuable contributors to the study, ensuring a comprehensive understanding of the subject.

To gather this diverse range of perspectives, the data collection involved personal and audio/video interviews with participants from different settings. These included students from Linnaeus University (Department of Informatics), students at BITS, Pilani (India) (Department of Computer Science), and visitors to Helsingborg City Library (Lund university students). The chosen participants included university students both males and females. This balanced gender distribution was aimed at ensuring a diverse and representative sample.

The recruitment process was conducted meticulously to maintain consistency and transparency. For students at Linnaeus University and BITS, Pilani, participants were approached via email. Similarly, for personal interviews with visitors to Helsingborg City Library, individuals were approached directly in the library setting. They were provided with a clear explanation of the study's aim and the interview procedures.

Approximately 15 people were requested to participate, resulting in interviews with 6 individuals. This approach ensured that the chosen participants adequately represented the diverse perspectives required for this study.

The interviews were conducted in the form of open-ended questions, which allowed the participants to freely express their thoughts and opinions. The interviews were conducted in English, and were audio recorded. To safeguard the privacy of the participants, any personal information disclosed during the interview has been maintained in strict confidence.

3.4 Interview Process

The interviews for this study were conducted between 25th April 2023 and 10th May 2023, using a semi-structured approach. The mode of the interviews varied, with some being conducted in person and others being conducted over audio/video conferencing. The participants were contacted through email and appointments were made based on their availability. The interviews were conducted in English and lasted between 30 minutes to an hour. The profiles of the interviewees are presented in Table 1. It is noteworthy to mention that their identities have been anonymized to preserve confidentiality.

<table>
<thead>
<tr>
<th>Name</th>
<th>University</th>
<th>Interview Date</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annika</td>
<td>Linnaeus University</td>
<td>28-Apr-23</td>
<td>35 Mins</td>
</tr>
<tr>
<td>Andreas</td>
<td>Linnaeus University</td>
<td>1-May-23</td>
<td>55 Mins</td>
</tr>
<tr>
<td>Chris</td>
<td>Linnaeus University</td>
<td>2-May-23</td>
<td>57 Mins</td>
</tr>
<tr>
<td>Felice</td>
<td>Lund University, Helsingborg Campus</td>
<td>4-May-23</td>
<td>43 Mins</td>
</tr>
</tbody>
</table>
Table 1: Participant details

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Date</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sourav</td>
<td>BITS Pilani</td>
<td>6-May-23</td>
<td>1 Hr 4 Mins</td>
</tr>
<tr>
<td>Swati</td>
<td>BITS Pilani</td>
<td>7-May-23</td>
<td>32 Mins</td>
</tr>
</tbody>
</table>

Prior to the interviews, the participants were sent a consent form (Appendix A) outlining the purpose of the study and their rights as participants. The consent form also included information about the recording of the interviews, and all participants provided their consent for the interviews to be recorded. Before commencing the interview process, participants were fully informed about the nature and purpose of the study. They were also made aware that the recordings of the interview would be kept confidential and would only be used for research purposes. Moreover, the participants were informed that their identities would be kept anonymous, and the recordings would be deleted once the study was completed. It was emphasized to the participants that they had the freedom to withdraw from answering any question that they felt uncomfortable with or to withdraw from the whole interview process altogether.

During the interviews, the participants were asked a series of open-ended questions related to their perceptions of dark data and its impact on their personal safety (Appendix B). The questions were designed to allow the participants to provide detailed and in-depth responses, and to share their personal experiences and perspectives on the topic.

The use of semi-structured interviews allowed for flexibility in the questioning process, while also ensuring that all participants were asked the same core questions. During the interviews, there was a positive and interactive atmosphere that enabled open and honest communication. The participants were engaged and willing to share their opinions, experiences, and concerns regarding the use of dark data. Overall, the interview process was conducted in a professional and respectful manner, and all participants were given the opportunity to share their thoughts and feelings on the topic. The experience was informative and insightful, and I am grateful for the opportunity to have conducted these interviews.

3.5 Data Analysis

Thematic analysis is utilized for the purpose of data analysis in this qualitative study. Thematic analysis is a widely used and flexible method for analyzing qualitative data that allows researchers to identify and analyze patterns, themes, and meanings within the data. Thematic analysis involves a six-step process, which includes familiarization with the data, generating initial codes, searching for themes, reviewing and refining themes, defining and naming themes, and producing the final report (Braun & Clarke, 2006).

1. Familiarization with the data: This step involves reading and re-reading the interview transcripts to gain an overall understanding of the data and to become familiar with the content, context, and meaning. This process helps to identify patterns, trends, and initial ideas for themes.

2. Generating initial codes: This step involves identifying initial codes, which are short descriptive labels that capture the essence of the data. These codes can be generated inductively, based on the data itself, or deductively, based on existing theoretical or conceptual frameworks.
3. Searching for themes: This step involves identifying overarching themes that emerge from the initial codes. Themes are patterns or meanings that emerge from the data and provide an interpretive account of the phenomenon being studied.

4. Reviewing themes: This step involves reviewing and refining the potential themes to ensure they accurately reflect the data. Themes should be internally coherent, distinct from each other, and relevant to the research question.

5. Defining and naming themes: This step involves defining and naming the final themes. Each theme should have a clear definition and a descriptive name that captures its essence.

6. Producing the report: This step involves writing up the analysis, including a description of the methods used, the themes identified, and the interpretation of the data. The report should be clear, concise, and transparent, providing a detailed account of how the themes were derived and how they relate to the research question.

Thematic analysis is a useful method for analyzing qualitative data because it is flexible and can be adapted to fit the research question and data. It allows for a detailed and systematic analysis of the data, while also allowing for the identification of new and unexpected themes.

Personal and audio/video interviews were conducted with the participants, and were recorded using the voice recorder app on the mobile phone. Each interview was conducted in a quiet room to ensure good sound quality. After the interviews, all of the recordings were transcribed verbatim. This involved listening to the recordings multiple times, and typing out the exact words that were spoken. This was a time-consuming process, but it allowed for a detailed analysis of the data.

To find and record the initial codes in this study, at first all the interviews were transcribed word for word. Then, the transcripts were imported into Microsoft Word and read through them multiple times to identify any recurring patterns or themes. Every line of the transcripts was carefully read and any phrases or sentences that related to the research question were highlighted. A list of codes was created in a separate document, with each code receiving a name that encapsulated the core idea or concept represented by it. This process, although time-consuming, involved meticulous reading of each transcript line multiple times and comparing them to the initial codes to ensure accurate coding of all pertinent text. This method allowed for the comprehensive coverage of various perspectives and experiences regarding the research question.

After the generation of the initial code list, the process of deriving themes commenced. Each code was reviewed multiple times to identify commonalities and disparities among them, leading to the grouping of potential themes. Patterns within the data that hinted at emerging themes were also explored. Once potential themes were identified, the data underwent further review to guarantee that each code was appropriately assigned to the corresponding theme. This entailed an iterative process of cross-referencing between the initial codes and emerging themes to ensure their consistency.

In naming the themes, descriptive labels were chosen to capture the core essence of each theme. This selection involved picking words or phrases that best conveyed the underlying meaning of the codes within each theme. It was also essential to maintain concise and clear differentiation between the theme names. Overall, the process of deriving themes from the
initial codes necessitated a methodical and iterative approach to reviewing and analyzing the data. This approach demanded meticulous attention to detail and a commitment to engage in reflective and critical data analysis.

3.6 Reliability and Validity

Reliability is a fundamental concept used to assess research quality, particularly in quantitative studies (Eisner, 1991). However, this research is qualitative in nature, where evaluation serves as a means to elicit information from collected data. According to Stenbacka (2001), reliability is associated with "explaining" in quantitative research, while in qualitative research, it focuses on "generating understanding." In the context of qualitative paradigms, terms such as Credibility, Confirmability, or Dependability are deemed more appropriate than reliability (Lincoln & Guba, 1985). In line with these definitions, this qualitative study on the risks of Dark data pertaining to personal safety employed a careful participant selection process. Both male and female participants were included, and individuals who were knowledgeable about data or actively used digital technologies were specifically targeted to facilitate better understanding.

Similar to reliability, validity encompasses different perspectives among researchers and scholars. Kvale (1989) presents three approaches to defining validity in qualitative research: as investigation, communication, and action. Creswell and Miller (2000) suggest that validity is influenced by the researcher's perception and paradigmatic assumptions. Lincoln and Guba (1985) propose alternative terms such as quality, rigor, and trustworthiness for qualitative research, instead of validity. In this study, semi-structured interviews were conducted for data collection. Various scientific sources cited in the references were utilized to ensure validity. Additionally, to enhance the validity of the interviews and the overall research, participants were provided with a brief overview of the research topic and insights into Dark data. This approach aimed to maintain the validity of the research findings.

3.7 Ethical Considerations

In this study, ethical considerations were given the utmost importance to ensure that the participants' rights and privacy were protected. Firstly, an informed consent (Appendix A) was obtained from all the participants before conducting the interviews (Mohd Arifin, 2018). The participants were provided with a detailed explanation of the study's purpose, the data collection process, and the measures taken to maintain their privacy and confidentiality. They were informed that their participation was voluntary, and they had the right to withdraw at any time during the interview.

To maintain anonymity and confidentiality, all participants' names and any identifiable information were replaced with pseudonyms in the final report. The recordings were deleted after transcription, and the transcripts were only accessible to the researcher to maintain confidentiality. Additionally, the participants were assured that the information provided by them would be used solely for the purpose of this research and would not be disclosed to any third party without their explicit consent (Orb et al., 2001). The data collected was analyzed objectively, and the results were reported in a manner that did not allow the identification of individual participants.

Furthermore, during the data analysis process, care was taken to ensure that the data was not interpreted in a manner that would harm or disrespect the participants. Any potential biases or
preconceived notions were also acknowledged and addressed during the analysis process (Creswell, 2009). Overall, this study followed ethical principles and guidelines to ensure that the participants’ rights and privacy were protected throughout the research process.
4. EMPIRICAL FINDINGS

This chapter presents the empirical findings derived from the interviews conducted in the study. The detailed description of the empirical findings concludes the chapter, providing an overview of the insights gained.

The empirical findings chapter presents the results and analysis derived from the interviews conducted for this study. It offers comprehensive insights into the themes that are derived. Utilizing a qualitative research approach, this study employed interviews and discussions as the primary methods of data collection. The gathered data underwent a rigorous examination to identify recurring patterns. Notably, the findings extracted from the student interviews yielded an initial set of 36 codes, which were subsequently categorized into 9 distinct categories (Appendix C, Table 2). These categories ultimately facilitated the identification of 6 overarching themes (Appendix C, Table 1), forming the basis for data analysis in this study.

- Dark data awareness
- Dark data risks
- Benefits of data sharing
- Personal data protection
- Accountability of Organizations
- Empowering Data Literacy

4.1 Themes

The below findings shed light on the current state of dark data and its implications for individuals. The study aimed to explore and understand the perceptions, challenges, and practices associated with dark data in order to highlight areas of concern and identify potential solutions. The following themes emerged from the analysis of the collected data:

4.1.1 Dark Data Awareness

The findings reveal varying levels of awareness among individuals regarding dark data. While some participants demonstrated a solid understanding of dark data and its implications, others had limited knowledge or misconceptions. The awareness of dark data varied across different sectors and industries.

When asked, what came to their minds when they hear the term Dark data, Chris said, "Once you have used the data that you collect, and looking at the specific data that you actually use, after you have used the data, and you still continue to store the data without the intention to reuse it will constitute to dark data."

Whereas, Felice explained that,

"There is a lot of data stored somewhere that is not directly connected to what the organizations need for analysis etc. but there is a lot of data that they don’t use or need but they are still storing it. Basically, I feel that it shouldn't be allowed to store so much data about me."


In Andreas’s words, ”It is something which is hidden, which is not easily visible, something is hidden and nobody's paying any attention on that particular data. Because a large numbers of various businesses and organizations are daily collecting large amounts of data that and are storing it. Out of that they select certain data which they think is useful for their businesses to run smoothly and profitably. But the data that they do not pay much attention on that, but that data on which they do not pay any attention that I call dark data.”

Sourav mentioned that,

"Dark data is something that is recorded and stored without real-time utilization or purposeful analysis. It may include call recordings, transaction records, unused data sets, etc. This data you're not using to enhance anything at your end, but still you have to keep this data safe. But, all the required encryption is an additional cost to these organizations."

Swati demonstrated lesser awareness of the exact term 'Dark data’ and she told that she thinks it is,

"Some information that is stored, somewhere in the dark or maybe, information which is stored and not used."

The conversations with the participants reflect varying levels of understanding and perspectives on the concept of dark data.

4.1.2 Benefits of data sharing

Through the interactions with study participants, a subtle perspective emerges, highlighting their awareness of the inherent risks associated with data sharing. However, it is evident that despite these concerns, individuals tend to perceive data sharing as advantageous in certain aspects. The participants acknowledge the potential for personalized services tailored to their preferences. Chris in his own words tells,

"I find it very convenient to have mega corporations such as Microsoft, Google and Amazon knowing stuff about me, as they are able to deal with my data in a better way. I truly desire the comfort of businesses out there being able to target my needs and send me targeted commercials of the things I’m planning to buy. It makes my life more comfortable. So, I think that is horribly hard to have an opinion on whether they should be able to collect my data, not only the data that I actively know that I consent about."

Whereas, some believe that it is the facilitation of staying connected with loved ones regardless of distance that makes them share some of their information on the various OSN avenues. Felice shared in her interview that,

"Yeah, we do share our pictures and information a bit as well because it's also a good way to connect with people at home. Yeah, but you need to be aware of what you post and how it looks."

The participants also expressed the comfort of remote access and the seamless convenience of performing tasks from the comfort of their homes. For example, Sourav mentioned,
"I have literally stopped visiting banks for any kind of service that I need. If I have to open a deposit or change the branch of my bank account or transfer money to someone etc., I can do all that online. Now I mostly shop for food, groceries and most of the other things online. It's really so convenient. But yeah I also am aware of so many online scams and try to keep a check on what and where I share my data.”

This collective sentiment emphasizes the positive impact that data sharing can have on various aspects of modern living, even as participants remain conscious of its associated risks.

4.1.3 Dark Data Risks

The study identified several risks associated with dark data, including data breaches, cyber attacks, and potential misuse of personal information. Participants expressed concerns about the security and privacy implications of storing unutilized data, highlighting the need for proper risk assessment and mitigation strategies.

For example, Felice doesn’t like the companies to collect her data. She says,

"I think it's a bit scary how much they collect through the IoT devices that we use such as webcams, Alexa, Siri or fitness trackers etc. So, if it was only for your own use or support, it will be okay. But it's not only that, it's also used to sell you something. That I don't like.”

She also added that, "If you look at your iPhone, that starts to tell you that you're probably on your way to Ica now because you always do that on Saturday at this time of the day. Yeah. I mean, that could also be used by other people to see when your house is empty normally on a Wednesday or normally on a Saturday or whatever. Everybody can use it. So no, I don't like that.”

Andreas states that,

"Sometimes there can be bad people who know information about you from the data that is available. They can call you and pretend that they know some of your loved one very well and then can say that your loved one has met with an accident somewhere or some kind of wrong information they may give to lure you to go to a place where they can take advantage of you or harm you.”

However, Sourav told that he believed that numerous organizations engage in the practice of collecting our personal data and subsequently selling it, including phone numbers and email addresses, to other entities. Unfortunately, there are instances where fraudulent companies misuse our data to contact us with the intention of perpetrating scams. In his own words,

"During a phone conversation, my friend's mother received a call from an individual claiming to be from her bank. The caller informed her that certain information was required to keep her account active, warning that it would become inactive otherwise. Filled with panic, she noticed SMS messages on her phone indicating transactions taking place in her account. When she questioned the caller about the transaction amounts, the person dismissed her concerns and insisted on obtaining the OTP (One Time Password) immediately to prevent account inactivity. Fortunately, my friend, who happened to be nearby, overheard the conversation and promptly disconnected the call. I’m sure it was some scammer, trying to fool her.”
Chris went on to share,

"The malicious intent comes from very different places, that is offering a product or service that doesn't exist, targeting my personal preferences, or collecting data to get to know my vulnerabilities, how to fool me in one way or the other that that is malicious use."

He also told something which is really concerning, in his words,

"A few years ago, there was a major scandal involving a house alarm company, possibly Verisure, where it was revealed that employees had created a subculture within the company, where they shared live feeds from cameras installed in people's houses. Unfortunately, some of these videos were leaked to the public, leading to a breach of privacy and trust. The scandal arose due to the misuse of these live feed videos. There are other companies as well, where you have active monitoring for your safety in your house or in your car or whatever. You find that the employees at the company that is supposed to keep you safe, they are actually stealing your data and the app really is stealing sensitive parts of you and your life and, and sharing it, unscrupulously. And that is something that I really; I think that that should be punished, too."

Swati mentioned a incident where the data that was shared innocently on some OSN, was used to harass people. She states,

"You know, there was some girl on Facebook, who ranted about a popular political leader in Mumbai and then people came to her house to beat her up."

Annika in her own words told,

"One of my colleagues, she was stalked and then there is a lot of data online. So it's quite easy to track what you're doing or where you are and If somebody wants to stalk me. I don't want them to know where I live right? That is why I don’t like my address or any personal information to appear on Google."

From the responses of almost all the participants it can be understood that the collection and storage of dark data, including conversation records on IoT devices, daily activity records from phones or fitness trackers, financial scams, and personal videos getting leaked, pose significant risks to personal safety. These risks encompass various forms of online and offline threats, such as stalking, online harassment, and physical harassment. It is crucial to recognize and address these risks to ensure the protection and well-being of individuals in an increasingly data-driven yet naive world.

4.1.4 Personal Data Protection

Participants emphasized the importance of robust data security measures and personal data protection practices. They expressed the need for individuals and organizations to prioritize data security, implement encryption techniques.

Chris goes on to inform that he takes several precaution to protect his data such as,

"I believe that I have a higher than average level of awareness regarding passwords, and security and in logging on. I do try to maintain two factor authentication. I do use a
complicated password, also I don't reuse passwords. And I also take care, whether to say the credit card information or not on the website, I'm actually thinking twice before I allow that to happen.” He also says,

“It is the responsibility of data storage providers in ensuring the security and protection of stored data and if a breach occurs, the storage provider should be held accountable. To prevent the misuse of dark data we need regulations and enforcement focused on the usage and consumption of data rather than just its storage.”

Annika expressed that she doesn’t share her data or pictures at a lot of OSN, she tells,

“I don't share my data on every site. Only limited what I really think I need and I know as much as possible. I protect my data as well to just share with friends. And I am very aware of what I share where I share it. I never post holiday pictures or pictures of the kids and locations especially when I'm still there. It's also about what kind of pictures of the kids you post, to protect them as well. These days, it is quite easy to use those pictures and other circumstances. There are a lot of pedophile networks that use harmless pictures and morph them.”

On the same lines, Felice says,

“I tried to share my email address as little as possible. And I also teach the kids that they shouldn't put their email address on every website. Only if you have a serious interest should you use your email address. I also don't accept all the cookies actually. Sometimes rarely, rarely I accept them or if given a choice, like functional analytical or necessary cookies then I suppose that I select only the necessary cookies. I try to protect my data.”

Sourav says that in this highly digital world it is most essential to be aware of the data risks and to proactively take steps to protect our data as best as possible. In his words,

“It is crucial to be cautious about the basic steps when dealing with personal information online. For instance, I’m mindful of emails that may be junk or spam and refraining from clicking on any suspicious links. Whenever I share any personal information, always ensure that the website begins with HTTPS, indicating a secure connection. Another important thing to protect is the credit card details, where after inputting the numbers, they should be masked except for the last four digits. Failure to observe these security measures on a website may indicate that it is fraudulent. So, I keep an eye on that as well.”

Andreas also says,

“It is very much necessary that you do not share your personal information about your family, your location, your address, and when you are working online it is essential to protect your bank information, your passwords to keep yourself financially safe. To protect the data on your hardware, you can install some good antivirus software. I scan my device periodically after every 7 to 10 days. I keep my apps updated and use strong passwords, the passwords which could be a collection of alphabets, numeric and special characters and also sufficiently long passwords. And I keep changing my password from time to time and also I do not use the same password on different platforms.”
Swati mentioned that she has set the privacy settings of her social networking apps to share her pictures only with the friends that she has accepted. She also states,

"People should avoid sharing their geolocation, otherwise anybody can trace your location and in today's world there are so many criminals just roaming around in our society. If they come to know your location, they can harass you or cause some physical harm as well."

The findings underscored the significance of protecting sensitive information from unauthorized access and potential threats. Overall, these conversations emphasize the awareness and proactive measures individuals take to protect their personal data and ensure their safety in an increasingly digital world.

4.1.5 Accountability of Organizations

Accountability emerged as critical factors in managing dark data. Participants highlighted the importance of clear communication and transparency from organizations regarding data collection, storage, and usage practices. They expressed the need for organizations to adhere to ethical standards, provide comprehensive user agreements, and be accountable for safeguarding personal data.

Sourav states that,

"They can summarize the terms of agreement in a paragraph, that this is what it is. Organizations can improve transparency by presenting key points in a concise manner, allowing users to understand how their data will be collected and used. They need to clearly communicate with the users about their data practices."

Chris also talks on the same lines, he says,

"So, the companies should share the kind of intended usage in a more straightforward manner. It's fairly general, what they say, may we use your data for generating some likes or get through specific ads or, but the usage is not very clearly stated. I think that that could be more properly presented."

Swati mentions,

"What companies share in terms of agreement, should be a more understandable language that a layman can understand. So, they can crisp it down to like 10 bullets that are the highlights of what you have agreed on so that we know what we are signing up for."

She also suggests,

"There should be some sort of data optimization I guess that they can definitely do. If they don’t need something, it should be discarded. So, they should have some sort of a routine that they can only keep the data for this much time, then any data after that should be discarded in a right way so that it cannot be reused or misused."

Andreas also goes on to discuss that, it’s necessary for the organizations to have some routine setup to destroy the data that has been stored for a long time. In his own words,
"It is very necessary for the organizations or businesses that even if they are not using that dark data, they should destroy that data or periodical basis because keeping that data is also cost, enormous cost for their storage. So they are paying the money for storing the data and at the same time data is posing a threat if cyber criminals penetrate through this data and misuse it in some way."

Felice also mentions the same thing, she mentions,

"In my opinion, it is important that only the data I explicitly consent to be saved should be stored, rather than everything I do. Additionally, there should be regular deletion of this data to ensure the privacy and security of my data."

Lastly, Annika says that,

"The organizations should invest in research and proactive measures to identify at a minimum cost, how much of dark data they need to store and how much of the dark data they can discard while ensuring compliance with regulations."

The study found that participants stressed accountability, clear communication, and transparency from organizations in managing dark data. The participants express concerns about the lack of clarity in terms of agreements, the need for organizations to clearly communicate data usage, and the importance of using language that is easily understandable to users.

4.1.6 Empowering Data Literacy

The participants expressed a shared belief in the necessity of spreading awareness about the existence of dark data and its potential risks. They emphasized the importance of educating individuals from childhood to adulthood about data protection and privacy. The participants suggested various approaches, such as integrating the topic into school curricula through fun activities or conducting workshops for adults in their workplaces. They emphasized the need for proactive efforts to instill a culture of data awareness and responsible data practices at an early age and throughout individuals' lives. In their own words,

Felice says, "Some online education is maybe an option. It shouldn't be too negative, though, because otherwise people tend to lose attention because everybody likes to be socially active these days."

Sourav tells, "Maybe the NGOs or government agencies should conduct workshops for their own cities or states where they conduct these kinds of education workshops to let others who are not that fluent in technology, to make them understand what are the impacts and how they can save themselves."

Annikas adds by saying, "Maybe some local authorities can conduct small workshops to make the general public aware of, what are the criticalities behind just blindly doing something on the internet because internet is everywhere. So maybe just make them understand how, how the data is processed, how the data is captured and how is it used."

But Andreas says that the awareness has to be inculcated from early childhood. In his words, "Kids should also be made aware of the risks of sharing data online, maybe by making this as
“a part of their curriculum where they understand in smaller chunks would be better. They are very active on the internet these days, without being taught about what’s out there. And I also think that kids listen to teachers more than their parents. So it’s better that it comes from that source.”

Overall, these conversations highlight the importance of transparency, user understanding, and responsible data practices in the collection and storage of personal data. They also emphasize the need for data optimization and regular data deletion to protect privacy and prevent misuse. The conversations collectively underscore the need for improved data governance, user consent, education and proactive measures to ensure the security and privacy of personal information.

4.2 Summary of the Findings

Based on the findings presented, the collected data revealed six distinct themes. These themes were discerned in alignment with the data collection process and in context of the thesis objectives. The ensuing table (Table 2) showcases these themes along with a concise overview of the corresponding findings for each theme.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dark data awareness</td>
<td>• Dark data is mostly unstructured data which is not being used and stays hidden</td>
</tr>
<tr>
<td></td>
<td>• Potential value of dark data</td>
</tr>
<tr>
<td></td>
<td>• Neglect of dark data</td>
</tr>
<tr>
<td>Benefits of Data Sharing</td>
<td>• Personalized services provided by the organizations</td>
</tr>
<tr>
<td></td>
<td>• Stay connected with family and friends</td>
</tr>
<tr>
<td></td>
<td>• Convenience of doing things from home</td>
</tr>
<tr>
<td>Dark data risks</td>
<td>• Personal safety and security risks</td>
</tr>
<tr>
<td></td>
<td>• Dark data misuse for fraud</td>
</tr>
<tr>
<td></td>
<td>• Risks to Children's Safety from pedophiles</td>
</tr>
<tr>
<td></td>
<td>• Cyberbullying, online harassment or even stalking</td>
</tr>
<tr>
<td></td>
<td>• Security risks of dark data as cyberattacks or data breaches</td>
</tr>
<tr>
<td>Personal data protection</td>
<td>• Best Practices and Personal Data Protection</td>
</tr>
<tr>
<td></td>
<td>• Balancing privacy, entertainment and convenience</td>
</tr>
<tr>
<td></td>
<td>• Data retention and intention</td>
</tr>
<tr>
<td></td>
<td>• Data ownership and consent</td>
</tr>
<tr>
<td></td>
<td>• Minimizing risks through proper data handling such as periodic data destruction and data governance</td>
</tr>
<tr>
<td>Accountability of</td>
<td>• Transparency and Terms of user agreements</td>
</tr>
<tr>
<td>Organizations</td>
<td>• Transparency of data collection and usage</td>
</tr>
<tr>
<td></td>
<td>• Data Storage responsibility and Use by Companies</td>
</tr>
<tr>
<td>Empowering Data Literacy</td>
<td>• Educating Users through workshops</td>
</tr>
<tr>
<td></td>
<td>• Spread awareness through education or campaigns</td>
</tr>
</tbody>
</table>

*Table 2: Summary of interview findings per theme*
5. DISCUSSION

This chapter explores the observational discoveries derived from the data collection process in this research. The findings extracted from the data are discussed in detail, shedding light on the key observations made.

5.1 Discussion on the Research questions

The empirical findings provide valuable insights into the current landscape of dark data awareness, risks, data security, and transparency practices within organizations. The identified themes contribute to the existing body of knowledge and can guide the organizations, individuals in developing ways to address the challenges associated with dark data. It is crucial to foster awareness, establish robust data protection measures, and promote transparency and accountability to mitigate the potential risks posed by dark data. The themes identified in the empirical findings offer a strong foundation for addressing the research questions posed in this study. The table 3 below illustrates the correlation between themes and the corresponding research questions they address.

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1: How do individuals perceive the risks of dark data?</td>
<td>1. Dark data awareness</td>
</tr>
<tr>
<td></td>
<td>2. Benefits of Data Sharing</td>
</tr>
<tr>
<td></td>
<td>3. Dark data risks</td>
</tr>
<tr>
<td>RQ2: How do these perceptions influence their attitudes towards sharing personal information online?</td>
<td>4. Personal data protection</td>
</tr>
<tr>
<td></td>
<td>5. Accountability of Organizations</td>
</tr>
<tr>
<td></td>
<td>6. Empowering Data Literacy</td>
</tr>
</tbody>
</table>

*Table 3: Research questions and correlating Themes*

5.2 Findings corresponding to the themes and related studies

5.2.1 Dark data awareness

The interactions with the participants of this study revealed that awareness regarding dark data varies among individuals and organizations. While some participants exhibit a strong grasp of the concept and its implications, others possess limited knowledge or misconceptions. Participant responses reflect on a range of perspectives over this concept. One participant's view aligns with the definition of dark data as unutilized and unrecorded information, suggesting that data left unused after its initial purpose constitutes dark data. Another participant questions the need for storing excessive data, particularly data not directly required by organizations, emphasizing the potential redundancy in accumulating such data. This sentiment resonates with the notion that dark data is often misused and there is an eminent need for better data management and handling techniques to prevent data breaches and misuse (Heidorn, 2008). Furthermore, a participant's characterization of dark data as concealed and overlooked data draws a parallel with the idea that dark data is often unseen.
amidst the extensive data collected by businesses. This aligns with the understanding that much of the data generated by the Internet of Things remains untapped (Hand, 2022). Another participant’s view underscores the significance of data security even for unutilized information, highlighting the need to safeguard dark data despite its lack of immediate use. Lastly, a participant’s perception of dark data as stored yet unused information reinforces the notion that this unanalyzed data remains a resource to be explored. These insights collectively exemplify a diverse range of viewpoints on the concept of dark data. The broader literature further supports the idea that dark data, often overlooked, has untapped potential for both businesses and individuals (Gandomi & Haider, 2015). The understanding that untapped data can be as meaningful as utilized data further emphasizes the value inherent in exploring dark data (Hand, 2022). This aligns with the understanding that, when harnessed, dark data can offer opportunities for valuable insights and informed decision-making across various domains, from organizational operations to personal well-being (Heidorn, 2008; Gandomi & Haider, 2015).

5.2.2 Benefits of Data Sharing

The study participants’ nuanced stance on data sharing, despite their awareness of associated risks, aligns with the understanding that data sharing can be advantageous in specific contexts. Their perspectives highlight the potential benefits of personalized services and enhanced connectivity, reflecting a positive outcome of data sharing in modern living. This sentiment resonates with the notion that dark data, often overlooked or underutilized, can offer valuable opportunities for organizations (Moumeni et al., 2021). This untapped resource can contribute to enhanced operational efficiency, improved decision-making, and competitive advantage. Analyzing hidden insights within dark data can provide organizations with a deeper understanding of customer behavior and market trends, guiding strategic decisions. Moreover, leveraging dark data can lead to more personalized customer experiences, driving engagement and satisfaction. The potential for innovation, cost savings, and operational efficiencies also arises from the analysis of dark data, allowing organizations to identify underutilized resources and optimize processes (Hobart, 2020). These opportunities, facilitated by advanced analytics and data management strategies, parallel the innovative application of dark data in the health sector discussed in Ahmed et al.’s (2020) book, where proactive health predictions are made using credit card data. This convergence of perspectives underscores the versatile potential of data utilization in both individual and organizational contexts.

5.2.3 Dark Data Risks

The participants’ responses provide a diverse spectrum of concerns regarding data privacy and security. One participant raises apprehensions about personal data collected via IoT devices being utilized for profit, underlining the unsettling reality that such data isn’t solely used for support but also sold to third parties, potentially leading to scams. Another participant recounts an incident where data misuse resulted in an attempted scam, showcasing the potential pitfalls of unrestricted data sharing. Instances of data misuse for targeted harassment, stalking, and manipulation underscore the tangible threats associated with the unchecked sharing of personal data. These narratives reflect the delicate balance between the convenience of data sharing and the ensuing vulnerability individuals face.

These participant responses align with the research conducted by Thomas et al. (2017), which examines risks tied to stolen credentials like data breaches, phishing attacks, and malware. This correspondence emphasizes participants’ concerns about data misuse leading to criminal
activities like identity theft and financial fraud. Similarly, the study by Fire et al. (2014) underscores the dangers posed by online networks, including identity theft and privacy breaches. Participants' accounts of data misuse for harassment and manipulation mirror these findings, illustrating the real-world perils of personal data being exploited maliciously.

The study also addresses the elevated vulnerability of young users, mirroring participants' apprehensions about the perils faced by adolescents in the online domain. This is inline with Dwyer et al.’s (2007) research which sheds light on users' trust in social networking platforms and their propensity to share personal data, correlating with participants' remarks on sharing information on platforms like Facebook, where users inadvertently expose private details to potential risks.

Several participants expressed apprehension about sharing data that might be exploited for malicious intent. This sentiment correlates with the findings of Elyashar et al. (2013), who discussed the potential risks associated with accepting friend requests from unfamiliar individuals on social media platforms, including incidents of cyberbullying, encouragement of self-harm, physical assault, and even fatal outcomes.

Collectively, these studies and participant insights underscore the urgent need to confront data privacy and security challenges. Both participants and research studies underline the potential dangers of data misuse and exploitation, underscoring the significance of informed decision-making, conscientious data practices, and robust security measures to shield individuals in the digital era. The findings of this study emphasize the importance of user education, policy enforcement, and technological solutions in safeguarding personal data and ensuring a secure digital environment.

5.2.4 Personal Data Protection

The insights gained from participant's perspectives in the previous section vividly underscore the imperative nature of robust data security and personal data protection practices. Their accounts collectively underscore the urgency for both individuals and organizations to prioritize data security, employing strategies like encryption methods and stringent authentication protocols. For example, one participant meticulously detailed their approach to fortifying password security and the cautious management of credit card information. Another participant stressed the pivotal role of data storage providers in ensuring security, suggesting an emphasis on regulations regarding data usage. This resonates with the suggestions by (Thomas et al., 2017), which offer a comprehensive array of measures that individuals can adopt to enhance their online security during various online activities. These include utilizing strong and unique passwords, enabling two-factor authentication, prioritizing secure websites, and exercising caution against phishing attempts. Moreover, the selective approach to sharing information, especially concerning children, was highlighted by another participant. Multiple participants shared active steps they take to safeguard their data, including prudent handling of emails, utilizing secure website connections, and shielding credit card details. Additionally, participants outlined strategies encompassing hardware protection, software updates, and the use of robust, frequently changed passwords. The discussions also touched on setting privacy controls on social media platforms, judicious geolocation sharing, and avoiding the use of public Wi-Fi. This is harmoniously aligned with the recommendations provided by (Fire et al., 2014) about regular software updates, careful management of privacy settings, and avoiding public Wi-Fi for sensitive tasks are also emphasized. Strategies such as securing home Wi-Fi networks, backing up data, scrutinizing email correspondence, and minimizing
the sharing of personal information are recommended. Furthermore, reviewing app permissions, avoiding suspicious downloads, employing reliable security software, monitoring accounts, and staying informed about online threats form a vital part of the overarching approach.

Collectively, adhering to these measures can significantly mitigate the risk of cyberattacks, identity theft, and unauthorized data breaches, aligning with the overarching concerns illuminated by the participant's perspectives and advocating for a safer digital landscape.

5.2.5 Accountability of Organisations

The participants' recurrent emphasis on accountability and transparency in managing dark data resonates strongly with the findings from Hande & Mane (2015) and Bertino et al. (2019). Both studies delve into the crucial factors of security and accountability in data management. Hande and Mane discuss security challenges and propose models for maintaining data accuracy and privacy within cloud computing. They underscore the trade-off between data security and control when using cloud services and stress the importance of accountability for data usage. Similarly, Bertino et al. (2019) highlight the significance of data transparency, enabling individuals to access information about data usage impacting them. This aligns with participants' calls for clear communication and comprehensible user agreements. These connections reinforce the collective notion that accountability, transparency, and security must be prioritized when managing dark data. Organizations can mitigate the risks posed by dark data in terms of personal safety by implementing various strategies and best practices. As per the research and consulting firm Gartner Inc., one of the key strategies is to have a proper data management plan in place (Panetta, 2019). This includes regular monitoring and reviewing of the data being collected, storage and usage policies, and ensuring that the data is being processed and stored in a secure manner. Protecting sensitive data is an important part of any business, which is why it's essential to use security measures such as encryption, secure data storage, access control, and data masking. Implementing security protocols and measures can help protect against data breaches and unauthorized access to confidential information. As elucidated by Tarkowski & Vogelezang (2021), better data protection regulations are imperative to ensure users' control over data and companies' transparency in its collection. Implementing these principles aligns with the participants' proactivity in safeguarding their personal data. Furthermore, recognizing the potential of dark data, organizations can harness it for insights, an approach supported by Moumeni et al. (2021). Properly managed, dark data can yield strategic advantages and informed decisions, ultimately highlighting the importance of prudent data management.

5.2.6 Empowering Data Literacy

The participants' unanimously emphasize on raising awareness about dark data's existence and risks. In this context, individuals evaluate their actions, such as sharing personal information online, based on the perceived benefits and costs of these actions.

When participants stress the importance of awareness initiatives and education, they are essentially highlighting the benefits of being informed and proactive in safeguarding personal data. They understand that by gaining knowledge about the risks associated with dark data and practicing responsible online behaviors, they can minimize potential costs, such as privacy breaches or cyber threats. This aligns with the Social Exchange Theory, as individuals...
are making informed decisions to maximize benefits (data security and privacy) while minimizing costs (risks and vulnerabilities). Furthermore, the reference to Chugh & Guggisberg's (2020) study reinforces this connection. The study's analysis of the consequences of oversharing personal data, particularly in the context of stalking and dating violence on social media, serves as an example of the potential costs associated with online actions. This underscores how individuals' awareness and education about online safety, as advocated by the participants, can help them avoid such negative outcomes.

5.3 Insights from the Theoretical Framework Perspective

The conversations and findings from the study also shed light on the relationship between personal data sharing and the social exchange theory, which hinges on three key factors: perceived benefits, perceived costs, and the comparison level, against which individuals assess the fairness of the exchange. Despite these concerns, individuals tend to perceive data sharing as advantageous in certain aspects of their lives, a perception that aligns with the principles of the Social Exchange Theory.

Participants in this study recognize the potential for personalized services and targeted advertisements, a notion that exemplifies the perceived benefits of data sharing. One participant articulates that mega corporations like Microsoft, Google, and Amazon, collecting and utilizing his data, enhance his daily life by tailoring their services to his preferences. For instance, in a study by Gimpel (2020a), it is analyzed that dark data can serve various purposes, such as sentiment analysis, risk assessment, and market analysis, providing organizations with a competitive advantage and the ability to make well-informed decisions. Properly leveraging dark data undoubtedly holds the potential to be highly beneficial for both individuals and organizations, which signifies a favorable comparison level. This illustrates a willingness to engage in data sharing as a form of exchange, where individuals offer their data in return for more personalized and convenient experiences, aligning with the core premise of the Social Exchange Theory.

Moreover, some participants emphasize the social aspect of data sharing, highlighting its role in maintaining connections and relationships, even across great distances. Their perspective echoes the idea of a reciprocal exchange within social networks, where the comparison level involves maintaining these social bonds. By sharing personal information, individuals create connections and strengthen their social bonds, aligning with the Social Exchange Theory's emphasis on relationships as a core element of exchange. Furthermore, Dwyer et al. (2007) and Boshmaf et al. (2011) reveal the trusting nature of social network users, demonstrating their willingness to share information within the online social sphere. This trust forms the foundation of social exchange within these platforms, where individuals offer personal data in exchange for the benefits of connection and interaction. Moreover, the study by Addae et al. (2017) introduces the concept of a PDA scale, emphasizing the nuanced nature of how people perceive and value their personal data, which factors into their comparison level. It acknowledges that the perceived value of personal data varies depending on the context and type of information, contributing to the cost-benefit calculation in data sharing. It also underscores the link between individuals' concerns about privacy and their actions, aligning with the theory's emphasis on rational decision-making in exchanges.

Participants in the study also expressed their appreciation for the convenience of remote access and online services, particularly in the context of online banking and shopping. They highlighted how data sharing contributes to the perceived benefits of making everyday tasks
more convenient, which represents a positive outcome of the exchange. This convenience, highly valued by participants, can be considered as part of an exchange equation. In return for their data, individuals gain access to the efficiency and convenience of online services, a favorable outcome. This collective sentiment underscores the multifaceted nature of data sharing, where individuals carefully weigh the potential benefits against perceived risks, factoring in the comparison level and expected outcome of the exchange.

Moreover, a study conducted by Raman et al. (2015) delved into privacy concerns during online shopping, revealing that many users fail to take adequate measures to protect their personal information, which highlights the perceived risks associated with the exchange. The research also explored participants’ perceptions of online privacy during the shopping experience. It found that IT graduates exhibited greater awareness of privacy issues compared to non-IT graduates, although no significant gender differences were observed. This highlights the need to raise awareness among young online shoppers about privacy issues, especially given the increasing use of e-commerce in India, as part of risk assessment and management within the exchange. In contrast, another study by Jiang, Yang & Jun (2013) emphasized the significance of convenience as a primary motivator for the adoption of online shopping. The study identified five key dimensions of online shopping convenience, encompassing access, search, evaluation, transaction convenience, and post-purchase experience, elements that influence the comparison level. Notably, the dimension of search convenience focused on user-friendly websites, various search options, and the ability to quickly locate desired products, contributing to the perceived benefits of the exchange. Furthermore, research by Aldas-Manzano et al. (2011) underscored the crucial roles of satisfaction with banking websites, trust, and perceived risk as pivotal factors in fostering loyalty to online banking services. It revealed that an individual's loyalty to a banking website closely correlated with their levels of trust and perceived risk. However, as the perception of risk increased, the relationship between satisfaction and loyalty diminished, with risk perception emerging as the most prominent inhibitor of loyalty in the analysis, reflecting the importance of perceived costs and risks within the exchange.

This aligns with the core premise of the Social Exchange Theory, which suggests that individuals engage in social interactions, including data sharing, based on a rational calculation of costs and benefits. Additionally, participants discussed the importance of balancing data usage and deletion, seeking a fair exchange that minimizes risks and costs while optimizing benefits. Overall, the findings demonstrate how individuals navigate data sharing decisions based on the perceived costs, benefits, trust, and reciprocity, aligning with the principles of the social exchange theory, where the comparison level and expected outcome play crucial roles in shaping their choices.
6. CONCLUSION

The final chapter concludes by providing a concise review of the paper, addressing the research questions, and offering reflections on the contribution of the research. Furthermore, potential avenues for further research are suggested.

6.1 Conclusion

The purpose of this study was to gain insights into the perceptions and practices of individuals regarding their online privacy and the risks associated with dark data. The research questions that were examined were,

RQ1: How do individuals perceive the risks of dark data?

RQ2: How do these perceptions influence their attitudes towards sharing personal information online?

To address these inquiries, semi-structured interviews with a range of university students were conducted. This group included participants from Linnaeus University and Lund University, as well as students from BITS, Pilani in India. Both male and female students were included to ensure a varied representation. Analyzing the insights gained from these interviews, initially 36 distinct codes were identified. These codes were later grouped into 9 specific categories, which played a pivotal role in uncovering 6 themes,

- Dark data awareness
- Dark data risks
- Benefits of data sharing
- Personal data protection
- Accountability of Organizations
- Empowering Data Literacy

The themes Dark Data Awareness, Dark Data Risks, and Benefits of Data Sharing, directly address the first research question (RQ1). Through engaging with participants, it became apparent that awareness regarding dark data varies significantly among individuals. Some participants exhibited a robust comprehension of the concept and its implications, while others displayed limited knowledge or misconceptions, showcasing a diverse spectrum of perspectives. Additionally, participants described dark data as hidden, overlooked, and concealed within the vast data repositories of businesses, emphasizing its elusive nature. They also stressed the importance of maintaining data security, even for data that remains unused, highlighting the necessity of safeguarding dark data despite its immediate lack of application. This aligns with the understanding that, when appropriately leveraged, dark data can yield valuable insights across various domains, encompassing organizational operations and personal well-being. Furthermore, participants' nuanced stance on data sharing, coupled with their awareness of associated risks, underscores the notion that data sharing can yield advantages within specific contexts. The participants' concerns regarding data privacy and security further underscore the delicate equilibrium between the convenience of data sharing and the vulnerability it can introduce. These concerns align with existing research findings, reinforcing the pressing need to address challenges related to data privacy and security.
Answer to the second research question (RQ2), is derived through the themes Personal data protection, Accountability of Organizations and Empowering Data Literacy. The participants’ emphasis on robust data security and personal data protection practices underscores their heightened awareness of the potential risks associated with sharing personal information online. Their accounts demonstrate a proactive approach towards safeguarding their data, with strategies like fortifying password security, cautious management of credit card information, selective sharing of information (especially concerning children), and utilizing various protection measures such as secure website connections and robust passwords. These actions reflect a conscious effort to counteract the potential threats of data breaches, cyberattacks, and identity theft.

The participants' recurring focus on accountability and transparency in managing dark data further illustrates how their perceptions shape their attitudes towards sharing personal information. Their calls for clear communication, comprehensible user agreements, and the implementation of security measures align with their concerns about data privacy and the need for responsible data management. This proactive stance towards data protection indicates that their perceptions of the potential risks directly influence their approach to sharing personal information online. Additionally, the participants' unanimous emphasis on raising awareness about dark data's existence and risks, underscores how their perceptions drive their commitment to educating themselves and others about online safety and security. This awareness-oriented approach reflects their acknowledgment of the potential dangers associated with oversharing personal data online. Overall, the participants' perceptions shape their attitudes towards sharing personal information online, guiding them towards more cautious and security-conscious behaviors in the digital landscape.

In conclusion, this study has illuminated a multifaceted understanding of individuals' perceptions and practices concerning online privacy and the challenges associated with dark data. Their viewpoints have consistently emphasized the importance of education, starting from childhood, to cultivate a culture of responsible data practices. The recommendations provided by the participants, coupled with insights from various research studies, collectively underline the urgent need for a proactive approach to data protection and security. Moreover, the participants’ keen focus on accountability and transparency in data management aligns with the broader research landscape, emphasizing the pivotal role of these factors in addressing data-related challenges.

6.2 Contribution

This study has made significant contributions to the understanding of how individuals perceive and respond to the risks associated with dark data and its impact on their attitudes towards sharing personal information online. Firstly, it has provided valuable insights into the diverse range of perspectives individuals hold regarding dark data, highlighting the variation in awareness levels and perceptions among different participants. This nuanced understanding is crucial for tailoring education and awareness initiatives to specific needs and knowledge gaps.

Secondly, this research has shed light on the proactive measures individuals take to protect their personal data online. Their strategies, such as fortifying password security, not oversharing their personal data, staying vigilant of sharing any digital information and advocating for clearer user agreements, offer practical recommendations for enhancing online security practices.
Lastly, the study underscores the importance of accountability and transparency in data management, aligning with broader research findings. This highlights the need for organizations to adopt responsible data management practices and clear communication with users to ensure data privacy and security.

In summary, this study contributes to the existing body of knowledge by providing a comprehensive exploration of individuals' perceptions and practices concerning dark data and online privacy. It offers practical recommendations for enhancing data security and underscores the significance of education, accountability, and transparency in addressing the challenges posed by dark data in the digital era.

6.3 Future Research

The sample used for this research primarily consisted of university students. While this demographic provided valuable insights into the perceptions and practices of this group, it may not fully represent the broader population. Future research could benefit from including a more diverse and inclusive range of participants to account for variations in perceptions across different demographics.

The study also touched on generational differences briefly, but future research could consider how different age groups perceive and manage online privacy. Given that younger generations often have distinct online behaviors and attitudes, understanding these generational differences can aid in tailoring educational efforts and privacy policies accordingly.

Secondly, while this study emphasized the importance of education in enhancing data protection practices, it did not delve deeply into the effectiveness of various educational interventions. Future research could evaluate the impact of diverse educational initiatives, considering factors such as learning preferences and cultural background.

Lastly, this study adopted a qualitative approach through in-depth interviews. While qualitative research offers a deep understanding of individuals' perceptions and practices, it may lack the generalizability that quantitative research can provide. Incorporating quantitative surveys into future studies can validate and quantify the findings, offering a more comprehensive perspective.

In conclusion, while this study provides valuable insights into individuals' perceptions and practices concerning online privacy and dark data, there is ample room for further research in this vital field. Addressing the study's limitations and pursuing these future research opportunities can lead to a more comprehensive understanding of how individuals navigate the intricate landscape of dark data and online privacy. This in-turn can lead to more informed policies, practices, and educational initiatives.
References


APPENDICES

Appendix A

Consent form for the interview

Dear Participant,

I would like to invite you to participate in my research study on the topic of "Navigating the risks of Dark Data – An investigation into personal safety". The aim of this study is to gain insights into the perceptions and practices of individuals regarding their online privacy and the risks associated with dark data.

If you agree to participate, a personal, audio, or video interview will be conducted with you, which will take approximately 30-45 minutes of your time. During the interview, a few questions related to your online activities, the steps you take to protect our privacy, and your perception of dark data will be asked.

Please note that your participation in this study is voluntary and you have the right to withdraw at any time without any consequences. Your responses will be kept confidential, and your anonymity will be maintained by assigning you a pseudonym in the data analysis. The data will be stored securely and will only use it for research purposes and this data will be destroyed after completion of the study. This study adheres to the ethical guidelines provided by the Swedish Research Council for conducting research on human subjects. Your rights and interests are protected throughout the study.

You are kindly requested to carefully read and understand the information provided in the consent form before giving your consent to participate in the study. Please sign at the bottom of this form to acknowledge that you have understood the information provided and voluntarily agree to participate in the study.

If you have any questions or concerns, please do not hesitate to contact me. Your participation in this study will be greatly appreciated, and I hope that this research will contribute to a better understanding of online privacy and the risks associated with dark data.

Thank you for your participation.

Sincerely,
Anshu Gautam

Participants name & signature

..................................................

Date:
Appendix B

Interview Questions

1. Have you ever heard of the term "dark data" (What comes to your mind when you hear the word 'Dark data')?

2. Do you think dark data is different from any other kind of data?

3. Do you believe that dark data can pose any kind of risk towards personal safety of people?

4. In your opinion, what can be the risks that might be associated with dark data?

5. Can you tell me your views in general, about sharing personal information online, I mean do you take any steps to protect your personal information?

6. If yes, what steps do you take to protect your data?

7. Have you or anyone you know has ever experienced any negative consequences related to dark data?

8. What information or education do you think would help people better understand the risks associated with dark data?

9. Do you think that organizations should be more transparent about their collection and use of dark data?

10. How do you think organizations should handle dark data to minimize risks?
## Appendix C

### Table 1: Themes and Categories

<table>
<thead>
<tr>
<th>Themes</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dark data awareness</td>
<td>Dark Data Awareness and Understanding (1, 2, 3, 4, 5, 6, 7)</td>
</tr>
<tr>
<td>Benefits of Data Sharing</td>
<td>Enhanced Convenience and Connectivity through Data Sharing (34, 35, 36)</td>
</tr>
<tr>
<td>Dark data risks</td>
<td>Personal Security and Safety Risks (15, 16, 17, 18, 19, 20)</td>
</tr>
<tr>
<td>Personal data protection</td>
<td>Data Management, Security, and Privacy (8, 9, 10, 11, 12)</td>
</tr>
<tr>
<td></td>
<td>Personal Data Protection (13, 14)</td>
</tr>
<tr>
<td></td>
<td>Proper Data Handling (24, 25, 26, 27, 28)</td>
</tr>
<tr>
<td>Accountability of Organizations</td>
<td>Transparency of Organizations (21, 22, 23)</td>
</tr>
<tr>
<td></td>
<td>Data Storage and Usage by Companies (32, 33)</td>
</tr>
<tr>
<td>Empowering Data Literacy</td>
<td>Education and Awareness Initiatives (29, 30, 31)</td>
</tr>
</tbody>
</table>

### Table 2: Codes and Categories

<table>
<thead>
<tr>
<th>Codes</th>
<th>Codes</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dark data awareness</td>
<td>Dark Data Awareness and Understanding</td>
</tr>
<tr>
<td>2</td>
<td>Unstructured nature of dark data</td>
<td>Dark Data Awareness and Understanding</td>
</tr>
<tr>
<td>3</td>
<td>Dark data understanding</td>
<td>Dark Data Awareness and Understanding</td>
</tr>
<tr>
<td>4</td>
<td>Dark Data Definition</td>
<td>Dark Data Awareness and Understanding</td>
</tr>
<tr>
<td>5</td>
<td>Culture and Dark Data</td>
<td>Dark Data Awareness and Understanding</td>
</tr>
<tr>
<td>6</td>
<td>Potential value of dark data</td>
<td>Dark Data Awareness and Understanding</td>
</tr>
<tr>
<td>7</td>
<td>Neglect of dark data</td>
<td>Dark Data Awareness and Understanding</td>
</tr>
<tr>
<td>8</td>
<td>Data retention and intention</td>
<td>Data Management, Security, and Privacy</td>
</tr>
<tr>
<td>9</td>
<td>Data ownership and consent</td>
<td>Data Management, Security, and Privacy</td>
</tr>
<tr>
<td>10</td>
<td>Proper management of dark data</td>
<td>Data Management, Security, and Privacy</td>
</tr>
<tr>
<td>11</td>
<td>Data protection measures</td>
<td>Data Management, Security, and Privacy</td>
</tr>
<tr>
<td>12</td>
<td>Research to Minimize Risks</td>
<td>Data Management, Security, and Privacy</td>
</tr>
<tr>
<td>13</td>
<td>Best Practices and Personal Data Protection</td>
<td>Personal Data Protection</td>
</tr>
<tr>
<td>14</td>
<td>Balancing privacy, entertainment and convenience</td>
<td>Personal Data Protection</td>
</tr>
<tr>
<td>15</td>
<td>Personal safety and security risks</td>
<td>Personal Security and Safety Risks</td>
</tr>
<tr>
<td>16</td>
<td>Dark data misuse for fraud</td>
<td>Personal Security and Safety Risks</td>
</tr>
<tr>
<td>17</td>
<td>Risks to Children's Safety from pedophiles</td>
<td>Personal Security and Safety Risks</td>
</tr>
<tr>
<td>18</td>
<td>Cyber bullying, online harassment or even stalking</td>
<td>Personal Security and Safety Risks</td>
</tr>
<tr>
<td>19</td>
<td>Security risks of dark data as cyber attacks or data breaches</td>
<td>Personal Security and Safety Risks</td>
</tr>
<tr>
<td>20</td>
<td>Data Risks and concerns</td>
<td>Personal Security and Safety Risks</td>
</tr>
<tr>
<td>21</td>
<td>Transparency and Data Sharing</td>
<td>Transparency of Organizations</td>
</tr>
<tr>
<td>22</td>
<td>Transparency and Terms of user agreements</td>
<td>Transparency of Organizations</td>
</tr>
<tr>
<td>23</td>
<td>Transparency of data collection and usage</td>
<td>Transparency of Organizations</td>
</tr>
<tr>
<td>24</td>
<td>Privacy Concerns and Government Surveillance</td>
<td>Proper Data Handling</td>
</tr>
<tr>
<td>25</td>
<td>Data handling and minimization</td>
<td>Proper Data Handling</td>
</tr>
<tr>
<td>26</td>
<td>Minimizing risks through proper data handling such as periodic data destruction and data governance</td>
<td>Proper Data Handling</td>
</tr>
<tr>
<td>27</td>
<td>Data protection measures</td>
<td>Proper Data Handling</td>
</tr>
<tr>
<td>28</td>
<td>Minimizing Risks and Data Handling</td>
<td>Proper Data Handling</td>
</tr>
<tr>
<td>29</td>
<td>Educating Users</td>
<td>Education and Awareness Initiatives</td>
</tr>
<tr>
<td>30</td>
<td>Awareness and education</td>
<td>Education and Awareness Initiatives</td>
</tr>
<tr>
<td>31</td>
<td>Spread awareness through education or campaigns</td>
<td>Education and Awareness Initiatives</td>
</tr>
<tr>
<td>32</td>
<td>Data Storage and Use by Companies</td>
<td>Data Storage and Usage by Companies</td>
</tr>
<tr>
<td>33</td>
<td>Data Collection through Online Platforms and IoT Devices</td>
<td>Data Storage and Usage by Companies</td>
</tr>
<tr>
<td>34</td>
<td>Personalized services provided by the organisations</td>
<td>Enhanced Convenience and Connectivity through Data Sharing</td>
</tr>
<tr>
<td>35</td>
<td>Stay connected with family and friends</td>
<td>Enhanced Convenience and Connectivity through Data Sharing</td>
</tr>
<tr>
<td>36</td>
<td>Convenience of doing things from home</td>
<td>Enhanced Convenience and Connectivity through Data Sharing</td>
</tr>
</tbody>
</table>