UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

VISUAL PRIVACY MITIGATION STRATEGIES IN SOCIAL MEDIA NETWORKS AND SMART ENVIRONMENTS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

JASMINE DEHART
Norman, Oklahoma
2023

VISUAL PRIVACY MITIGATION STRATEGIES IN SOCIAL MEDIA NETWORKS AND SMART ENVIRONMENTS

A DISSERTATION APPROVED FOR THE
SCHOOL OF COMPUTER SCIENCE

BY THE COMMITTEE CONSISTING OF

Dr. Dean Hougen, Chair

Dr. Deborah Trytten

Dr. Song Fang

Dr. Angela Zhang

Acknowledgements

I wish to thank my committee members, who were generous with their expertise and time. First, I would like to thank my committee chairperson, Dr. Dean Hougen, for your willingness to support me and guidance in my dissertation journey. Thank you, Dr. Song Fang, Dr. Deborah Trytten, and Dr. Angela Zhang, for agreeing to serve on my committee and supporting me throughout this process. A special thanks to Dr. Christan Grant for his countless hours of reflecting, reading, encouraging, and, most of all, patience throughout this process.

I want to acknowledge and thank my collaborators, the School of Computer Science, the Gallogly College of Engineering, OU's Graduate College staff, the GEM Fellowship, and the SMART Scholarship for Service for supporting me in conducting my research and providing any assistance requested.

Lastly, I would like to thank my family and friends. It is with your help and support that I have come this far. Whether you have read my drafts or asked me to repeat my explanations twice, it is because of you that I can showcase my dissertation today. Thank you, thank you all!

Contents

List of Figures

List of Tables

Abstract

The contemporary use of technologies and environments has led to a vast collection and sharing of visual data, such as images and videos. However, the increasing popularity and advancements in social media platforms and smart environments have posed a significant challenge in protecting the privacy of individuals' visual data, necessitating a better understanding of the visual privacy implications in these environments. These concerns can arise intentionally or unintentionally from the individual, other entities in the environment, or a company.

To address these challenges, it is necessary to inform the design of the data collection process and deployment of the system by understanding the visual privacy implications of these environments. However, ensuring visual privacy in social media networks and smart environments presents significant research challenges. These challenges include accounting for an individual's subjectivity towards visual privacy, the influence of visual privacy leakage in the environment, and the environment's infrastructure design and ownership. This dissertation employs a range of methodologies, including user studies, machine learning, and statistics to explore social media networks and smart environments and their visual privacy risks. Qualitative and quantitative studies were conducted to understand privacy perspectives in social media networks and smart city environments. The findings reveal that individuals and stakeholders possess inherited bias and subjectivity when considering privacy in these environments, leading to a need for visual privacy mitigation and risk analysis.

Furthermore, a new visual privacy risk score using visual features and computer vision is developed to investigate and discover visual privacy leakage. However, using computer vision methods for visual privacy mitigation introduces additional privacy and fairness risks while developing and deploying visual privacy systems and machine learning algorithms. This necessitates the creation of interactive audit strategies to consider the broader impacts of research on the community. Overall, this dissertation contributes to the advancement of visual privacy solutions in social media networks and smart environments by investigating

and quantifying the visual privacy concerns and perspectives of individuals and stakeholders, advocating for the need for responsible visual privacy mitigation methods in these environments. It also strengthens the ability of researchers, stakeholders, and companies to protect individuals from visual privacy risks throughout the machine learning pipeline.

Chapter 1

Introduction

Visual privacy has risen to the forefront of individual and stakeholder concerns with the growth of technology in Social Media Networks (SMN) and smart environments. The concept of privacy can be user-subjective and task-oriented, but in this dissertation, privacy is defined as the ability of an individual to withhold information that is considered private, personal, and includes any content that an individual does not want to be shared. This definition allows for flexibility across environments and technologies and is inclusive of possible individual subjectivity. Personally identifiable information for individuals can include images, documents, and geographic location, among others. The visual content, image and video data, captured in these environments can contain privacy leaks regarding the individual or someone else. Privacy leaks include any instance in which a transfer of personal identifying visual content is captured in SMNs and smart environments. Private visual content exposes intimate information that can be detrimental to finances, personal life, and reputation. Private visual content includes baby faces, credit cards, phone numbers, social security cards, house keys, etc. The consequences of privacy leaks can include identity theft, burglary, and kidnapping.

With the use of SMNs and smart environments, there is an increased risk of an individual's private information being leaked, an infrastructure or system being breached by an

attacker, and an increased risk of physical threats because of these platforms. The technology developed to serve others may pose a grave risk to individuals, stakeholders, and researchers engaging with those platforms. The ongoing collection, connection, and storage in these environments serve as a catalyst for the rise in asset, location, and personal attacks. With these vulnerabilities continually being exploited, there is a need for visual mitigation strategies in these platforms to secure the privacy of individuals.

1.1 Research Questions and Objectives

In this dissertation, I address visual privacy leakage within the realm of SMNs and smart city environments. This research aims to assess the effects of visual privacy leakage in SMNs and smart cities and its' potential impact on individuals and stakeholders. The following research objectives facilitate the achievement of this aim:

1. Understanding and identifying individual and stakeholder perspectives and concerns of visual privacy in SMNs and smart cities (Chapter 3, Chapter 4)

2. Exploring and developing visual privacy risk scoring algorithms (Chapter 5)

3. Proposing interactive auditing strategies for Visual Privacy research (Chapter 6)

These objectives are addressed with these methodologies: machine learning, statistics, and user studies. This dissertation shows that there is a need for visual privacy mitigation and improvement in individual, stakeholder, and researcher understanding of visual privacy in smart cities and SMNs. The development of visual privacy research has brought rise to proposing an analysis of the visual privacy machine learning pipeline, which allows further mitigation of potential biases and privacy concerns from the model and modeler. Table 1.1 outlines the research questions that this dissertation seeks to answer.

Table 1.1: Research Questions and Summary of Contributions.

Research Question	**Contributions**	**Chapter**
What are the privacy-related experiences and concerns of social media users regarding visual content and threats on these platforms?	• Subjectivity of users' privacy attitude and perspective investigated with age and gender demographics. • Explores potential visual content privacy leaks on Twitter via keyword search. • Users' privacy attitude and perspective provide insight into the dangers concerning users and influenced a hierarchy of dangers in correlation to visual privacy on SMNs.	3

* **Table 1.1 – Continued on the next page** *

Research Question	Contributions	Chapter
What considerations do smart city stakeholders give for privacy and cost in technology and infrastructure?	• Survey smart city stakeholders on technology, cost, and privacy • Explores Smart City Challenge Finalist to understand common themes, technology requests, and privacy considerations in environment development • Proposes the use of visual privacy mitigation and Delay Tolerant Networks in smart city environments • Provides a case study of a low-cost and privacy-enabled smart technology deployed	4
How can object importance, prominence, and identifiability contribute to visual privacy risk scoring methodologies?	• Adaption of existing privacy scoring methodology to utilize computer vision • Proposes the *Vango* privacy risk scoring framework that includes features of importance, prominence, and identifiability	5

* **Table 1.1 – Continued on the next page** *

Research Question	Contributions	Chapter
How can privacy and fairness risks created by the development and deployment of visual privacy systems be mitigated to protect individuals and stakeholders?	• Proposes interactive audit strategies for privacy and fairness in visual privacy research	6

1.2 Contributions

To provide a comprehensive analysis of the research questions outlined in Table 1.1, this dissertation begins with participant surveys and quantitative analysis, which aims to identify the visual privacy risks associated with social media networks and smart environments. Additionally, this research endeavors to examine the unforeseen implications of the technology developed, including the potential risks posed by data collection, connection, and storage within these environments. By utilizing a comprehensive approach, this dissertation aims to provide a thorough understanding of visual privacy risks and offer viable mitigation strategies. The findings of this research have the potential to inform researchers, stakeholders, and individuals about methods for safeguarding visual privacy in social media networks and smart city environments.

1.2.1 User and Stakeholder Perspectives of Visual Privacy

Chapters 3 and 4 of this dissertation comprise qualitative and quantitative studies that investigate the privacy-related experiences and concerns of social media users regarding visual content, as well as the privacy considerations of smart city stakeholders for technology and

infrastructure. These chapters form the foundation of the dissertation by shedding light on issues faced by individuals in these environments. Through a combination of surveys and analysis, the studies demonstrate that the lack of visual privacy mitigation and the rapid advancement of technology can impact an individual's behavior, risk perception, and privacy concerns.

Chapter 3 focuses on exploring the visual privacy concerns of social media network users with regard to engagement, ownership, and the risks associated with sharing visual data. Based on the surveys conducted, several common themes emerged. Firstly, users' privacy attitudes are subjective and can supersede potential privacy concerns of others or the platform itself. Secondly, users' opinions about visual private data are contingent upon whether they anticipate any harm to themselves or their families. Finally, many users express high levels of concern about the physical dangers of kidnapping, burglary, and stalking due to the private data exposed on social media networks.

These findings have implications for social media network creators and personnel, as well as for future research directions. The subjectivity of users' privacy concerns means that privacy mitigation strategies should be customizable, which poses a challenge for many platforms and interfaces. Furthermore, continued research in visual privacy mitigation systems that reduce visual privacy risks and its' associated dangers are important to support visual privacy management for individuals.

Chapter 3 underscores the significance of prioritizing users in social media networks. This emphasis on users serves as a catalyst for Chapter 4, which delves into the influence of stakeholders' decisions regarding technology and infrastructure on privacy and cost for citizens in smart city environments. Departing from an individual-focused approach, this chapter examines the perspectives of stakeholders and policymakers on privacy and its implications, which may be beyond the control of citizens. Smart city technologies can potentially compromise citizens' visual privacy and rights. As such, stakeholders' decisions on infrastructure, technology, and data collection practices must factor in visual privacy risks and cost concerns

that arise from these developments, alongside considerations regarding how to incorporate visual privacy mitigation.

The findings in Chapter 4 reveal that privacy, simplicity, and convenience are frequently described as top priorities by stakeholders. Stakeholders have expressed concerns about privacy in smart cities but often neglect the integration of privacy protocols in technologies. Additionally, beyond the survey, this chapter explores the Smart City Challenge Finalist (DOT 2015) to identify shared themes, technology requests, and privacy considerations in developing smart city environments. These results corroborate the necessity of addressing visual privacy risks to guarantee that citizens are protected while benefiting from these technologies.

Chapters 3 and 4 establish the groundwork of this dissertation by means of qualitative and quantitative studies that highlight visual privacy challenges for social media networks and smart city environments. This research informs recommendations for researchers and infrastructure to develop visual privacy mitigation that accounts for the diversity of expectations, preferences, and circumstances.

1.2.2 Exploring Visual Features and Measuring Visual Privacy Risk

Chapter 5 addresses the research question, "*How can importance, prominence, and identifiability contribute to visual privacy risk scoring methodologies?*" This chapter introduces a visual privacy scoring method that is designed and applied, followed by a comparison with an adapted privacy scoring method. The proposed method is grounded in computer vision, with applications to the theories of the golden spiral and term frequency-inverse document frequency (TF-IDF). The severity of privacy leakage is derived from the hierarchy of visual privacy dangers outlined in Chapter 3. The hierarchical danger framework provides a theoretical foundation for understanding the intensity of visual privacy risk in terms of detecting private items and the potential user subjectivity toward visual content.

Additionally, this chapter explores several visual datasets to examine the efficacy of Object Importance Weights (OIW), Golden Spiral Distance (GSD), and Object Area Ratio (OAR) approaches in measuring visual privacy risk. The findings in Chapter 5 reveal that importance, prominence, and identifiability show promise as components of a visual privacy risk score. The results show an opening in necessity to exploiting visual features to understand and score visual content.

1.2.3 Issues with Visual Privacy Mitigation: Vulnerabilities and Regulation

This dissertation asserts the significance of exploring an interactive auditing pipeline for visual privacy research, as proposed in Chapter 6. The development and deployment of visual privacy systems necessitate a fundamental shift in research procedures, developer practices, and policies to prioritize strong visual privacy mitigation that account for fairness issues throughout the machine learning pipeline. The surveys conducted in Chapters 3 and 4 demonstrate that individuals are apprehensive about visual privacy leakage and data collection procedures. These chapters also reveal that many individuals say that they are unable or unwilling to engage with social media networks and smart city environments if privacy measures are not established. Furthermore, the protection methods employed in these environments intensify visual privacy and fairness implications during collection, processing, and storage. The development and deployment of visual privacy systems can potentially give rise to technologies that violate privacy and fairness. Therefore, researchers must consciously strive to implement best practices to safeguard all users from additional fairness and privacy risks by reconsidering data collection processes, machine learning algorithms, and the supervision of the machine learning pipeline.

1.3 Dissertation Outline

The chapters of this dissertation can be read independently. Table 1.1 provides the order of the chapters organized by the research questions and provides the progression of qualitative user studies to visual privacy risk scoring strategies to auditing visual privacy pipelines. In Chapter 2, I present the related work for this dissertation by relevant themes and concepts. Chapter 3 and Chapter 4 explore individual and stakeholder perspectives and concerns of visual privacy in SMNs and smart city environments and investigates the relationship between danger, cost, and leakage. Building on the exploration of visual privacy in various domains, Chapter 5 explores visual datasets to demonstrate a visual privacy risk scoring methodology, *Vango*, which implements computer vision techniques to detect objects in images and calculate the visual privacy score of visual data by exploiting visual features. This dissertation also presents a visual privacy auditing pipeline to mitigate unexpected risks due to visual privacy algorithms and systems in Chapter 6. Lastly in Chapter 7, this dissertation concludes by reflecting on the need and risks of visual privacy research applied in SMNs and smart city environments in order to provide inclusive individual and stakeholder protection.

Chapter 2

Related Work

Visual content (images and videos) that contain privacy leaks may expose intimate information harmful to your finances, personal life, and reputation (Gross and Acquisti 2005). Visual privacy leaks include any instance in which a transfer of personal identifying content is shared via visual content. Anything posted to Social Media Networks (SMNs) or captured in smart city environments can be exposed even after the removal of the content. From visual content, attackers can also extract textual information, including credit card numbers, social security numbers, place of residence, phone numbers, and other information (Gross and Acquisti 2005; Li et al. 2017c). This chapter lays the foundation across several areas of computer science and provides the framework for the studies and discussions in the following chapters.

2.1 Social Media's role in exposing Visual Privacy

According to Pew Research Center, 81 percent of online Americans use YouTube, 69 percent of online Americans use Facebook, and 40 percent of online Americans use Instagram (Auxier and Anderson 2021). Among adults under 30, the most commonly used platforms are Instagram, Snapchat, and Tiktok (Auxier and Anderson 2021). The users share images and videos daily. Understanding and protecting visual privacy is important in the growth of technology

and online social engagement. As SMNs continue to grow in popularity, they become a powerhouse for privacy leakage whether intentional or unintentional. To many, privacy on social media networks is user-dependent. People tend to share different content, have different privacy settings, and have subjective perspectives on privacy. Several papers have examined the privacy settings of users' accounts in correlation to their privacy leakage (Madejski et al. 2011; Gross and Acquisti 2005; Krishnamurthy and Wills 2008); however, there are privacy concerns that go beyond privacy settings (Gross and Acquisti 2005). By exploring users' attitudes and intentions on social media platforms, upcoming developments consider that the privacy settings of social networks are failing the users (Madejski et al. 2011). With efforts to understand users' attitudes and behaviors, authors (Knijnenburg 2017) suggest six privacy profiles that categorized social media users by their privacy settings and attitudes. Investigating users' privacy settings on social media is important, but it is also important to explore the disclosed information from or about users (Gross and Acquisti 2005; Veiga and Eickhoff 2016).

Ninety percent of Facebook profiles contain at least one image, 87.8% of users share their birth date, 39.9% of users list phone numbers (including 28.8% that contain cell phone numbers), and 50.8% of users share their current residency (Gross and Acquisti 2005). Additionally, revealing information such as birthdate, hometown, current residence, and phone number can be used to estimate the user's social security number and exposes them to potential financial and identity threats (Gross and Acquisti 2005). The textual content can also be extracted from posts that contain visual content. On SMNs, a user's profile information and visual content can intentionally or unintentionally be shared even though a privacy risk may arise (Gross and Acquisti 2005).

Previous work studying user privacy on SMNs has focused on multiparty privacy conflicts (Such et al. 2017b; Zhong et al. 2018), images or text content from users (Abdulhamid et al. 2014; Gross and Acquisti 2005; Squicciarini et al. 2014; Srivastava and Geethakumari 2013; Zerr et al. 2012c; Buschek et al. 2015; Tierney et al. 2013; Gurari et al. 2019; Kuang

et al. 2017), third-party applications that supervise privacy (Buschek et al. 2015; Zerr et al. 2012b; Gurari et al. 2019; Kuang et al. 2017), the influence of culture in settings and communities (Gross and Acquisti 2005), a user's privacy setting on SMNs (Krishnamurthy and Wills 2008; Gross and Acquisti 2005; Rosenblum 2007; Madejski et al. 2011), users' attitudes, intention, or behaviors on these platforms (Abdulhamid et al. 2014; Madejski et al. 2011), and understanding children and teenagers' interactions with SMNs (Boyd 2014; Boyd and Marwick 2011). Studies exploring teenager attitudes towards privacy note that teenagers tend to be more open about their lives on social media when compared to older users (Boyd 2014; Boyd and Marwick 2011) and emphasize the importance of stranger danger and insider threat for minors on SMNs (Johnson et al. 2012).

These studies focus on social media attitudes and privacy settings, but unlike Chapter 3, they do not evaluate users' attitudes about visual privacy in relation to user behavior and keywords that correlate to visual privacy leakage on social media. Chapter 3 builds on this research by utilizing users' visual privacy attitude and perspective to create a hierarchy of dangers from SMNs users. Furthermore, Chapter 3 re-emphasizes privacy concerns demonstrated by recent research across a broader range of social media platforms.

2.2 Addressing Visual Privacy Concerns in Smart City Environments

The concept of a smart city has recently led people, cities, and governments to pursue automated improvements to municipal infrastructure. Stakeholders may have different expectations for how their city should invest in improvements. Currently, no standard definition for a smart city exists, causing variable expectations of residents, city governments, and other community stakeholders. Smart city environments can heavily depend on wireless connections, servers, and storage which can be an attack vector for threats. Researchers have suggested using ontological security frameworks (Malkawi et al. 2022), smart contracts (Siddiqui

et al. 2023; Uchani Gutierrez and Xu 2023), block-chain (Uchani Gutierrez and Xu 2023), heterogeneous networks (Al-Turjman et al. 2019), and machine-learning architectures (Kunchala et al. 2023; Malkawi et al. 2022) to protect smart city technology, infrastructure, and services.

Studies have focused on security and privacy concerns of organizations (Aslam et al. 2022), and in the deployed technologies (Azhar 2020), however, the citizens' concerns are equally important to keep the smart city environment thriving (Martinez-Balleste et al. 2013). Citizens expect privacy, affordability (DeHart et al. 2020b), and timely and interactive information from a smart city (Cortez and Larios 2015). Recent studies have shown that popular smart technologies may not provide the stated benefits that consumers expect (Brandon et al. 2021). While technological innovations continue, citizens are critical about how unvetted smart cities can violate intrinsic rights (Smith 2019). However, laws are being proposed and passed to ensure the responsibility of the city or company to protect the privacy of the citizens (Doctorow 2020; Devlin 2020). Researchers have conducted surveys to understand citizens' feelings and attitudes about privacy toward the technology deployed in smart city environments (Wahyudi et al. 2022). To actively protect privacy in smart environments, citizens can depend on other products to curtain themselves from smart devices (Morse; Chen et al. 2020).

In an effort to protect visual privacy from cameras, researchers have invented methods to disguise themselves from surveillance systems using fashionable masks (Harvey 2012). The concept of visual privacy protection can further be extended to citizens' privacy protection in smart city environments. Recent works have shown an interest in developing visual privacy-preserving methods with the use of encryption schemes for surveillance cameras (Al-Husainy and Al-Shargabi 2020; Li et al. 2021a), and privacy-preserving video mitigation for visual pedestrian data (Kunchala et al. 2023). High costs can be incurred when equipping and maintaining privacy-preserving mechanisms in smart cities due to the cost of deploying technology, employing experts, continuous personnel security training, and frequent environment

assessments (Aslam et al. 2022). The transformation into a smart city is expensive (e.g., between $30 Million and $40 Billion), and only a few cities can obtain the resources required for upgrades (DeHart et al. 2020b).

Chapter 4 seeks to contribute to this field by understanding stakeholder considerations and perspectives for technology, privacy, and cost for smart city development. Additionally, Chapter 4 defines some common themes of smart cities, describes examples of technology deployed in smart cities, and investigates privacy considerations of smart cities by analyzing the 2015 Smart City Challenge. Related works focus on protecting organizations and infrastructure but lack the emphasis on citizens' protection beyond textual personally identifying information collected in smart city environments. Chapter 4 emphasizes the importance of visual privacy and cost in smart city environments with proposed solutions and showcases an example of low-cost and privacy-enabled smart technology.

2.3 Methods of Detecting, Protecting, and Scoring Visual Privacy Leakage

The perception of privacy is highly subjective and user-dependent (Zerr et al. 2012c; Martinez-Balleste et al. 2013), which is shown by literature focusing on users' attitudes towards privacy (Madejski et al. 2011; Wahyudi et al. 2022). As people engage in SMNs and smart cities, the visual content shared and collected in these environments can contain potential privacy leaks (Zerr et al. 2012c; Hoyle et al. 2015). SMNs and smart city personnel need to take meaningful action to decrease the exposure of personal information via visual content. The privacy risks of engaging on SMNs and in smart cities could outweigh the benefits. Several studies have found that visual content posted on SMNs (Squicciarini et al. 2014; Such et al. 2017b; Gross and Acquisti 2005; Rosenblum 2007) and captured in smart city environments (Kunchala et al. 2023; Martinez-Balleste et al. 2013) can pose a danger to the consumers. With the visual content exposed in these environments, attackers can uncover

personal identifying information that can be collected (Squicciarini et al. 2014; Al-Turjman et al. 2019). Visual content can also become a gateway for multiparty conflicts among consumers (Such et al. 2017b; Martinez-Balleste et al. 2013; Thomas et al. 2010). These conflicts can arise due to feelings of ownership, privacy boundaries, and privacy perspectives of the individuals in the visual content. Looking further into self-censoring and reduction of multiparty conflict, users can implement privacy-preserving procedures to reduce identity, association, and content disclosure (Loukides and Gkoulalas-Divanis 2009).

The identification of leaked content requires researchers to understand private items or objects that can be found in visual content (Zheng et al. 2022; Krishnamurthy and Wills 2008; Vishwamitra et al. 2022). Once potential visual private categories are identified, researchers can implement techniques to identify sensitive objects in the visual content. To detect visual privacy leakage in visual content, researchers have suggested the use of several machine learning algorithms (De Luca 2019): Recurrent Neural Networks (Neerbek 2020), Deep Neural Networks (Tonge and Caragea 2020, 2016), and Convolution Neural Networks (Vishwamitra et al. 2022; Gurari et al. 2019; Orekondy et al. 2018).

To protect the visual privacy of individuals, researchers have suggested the use of concepts like obfuscation (Li et al. 2017b; Padilla-López et al. 2015) for mitigating objects and faces. When a visual content privacy leak is detected, researchers have emphasized the use of mitigating visual privacy leaks in the visual content with blocking (DeHart and Grant 2018), blurring (Li et al. 2017b,c), censoring (DeHart and Grant 2018), wireframing (Kunchala et al. 2023), pixelate (von Zezschwitz et al. 2016; Orekondy et al. 2018), crystallize (von Zezschwitz et al. 2016), oil paint (von Zezschwitz et al. 2016), and adversarial noise (DeHart and Grant 2018; Goodfellow et al. 2014). Researchers have developed mitigation techniques for visual privacy that range between intervention methods and data hiding (Padilla-López et al. 2015). These methods can be implemented to protect visual privacy before posting the content or after identifying private objects. The visual content mitigation techniques can be incorporated into deployable systems for use in SMNs and smart environments. Studies

have provided privacy-aware systems that seek to reduce the number of visual privacy leaks in infrastructure and technology (Buschek et al. 2015; von Zezschwitz et al. 2016; Mazzia et al. 2012; Li et al. 2017c; Tierney et al. 2013; Zerr et al. 2012b; Tierney et al. 2013; Li et al. 2017a; Orekondy et al. 2018; Li et al. 2017b; Zhao and Stasko 1998; Boult 2005). Visual privacy mitigation should be centered around adaptability and cater to the individualistic concepts of privacy (DeHart and Grant 2018; DeHart et al. 2020c).

The assessment of privacy leakage for visual content allows the quantification of risk with a privacy score. A privacy score computes probable exposure (Becker 2009; Grandison et al. 2017) of data leaks (Liu and Terzi 2010) in that environment. Privacy scoring methods can include item-based (Wang et al. 2019; Chen et al. 2021; Aghasian et al. 2020), individual-based (Aghasian et al. 2017; Srivastava and Geethakumari 2013; Liu and Terzi 2010; Halimi and Ayday 2022; Pensa and Blasi 2016), and network-based (Pensa et al. 2019; Liu and Terzi 2010) techniques. Each privacy scoring method incorporates unique metrics and features to quantify the risk. Item-based privacy scoring metrics focus on an object or textual leakage from the content. Individual privacy scoring methods can include behavior (Ali et al. 2013; Caliskan Islam et al. 2014; Li et al. 2020), privacy settings (Liu and Terzi 2010; Coban et al. 2022), network structure (Pensa and Di Blasi 2016; Alemany et al. 2018; Li et al. 2021b; Kilic and Inan 2019), friends (Coban et al. 2022; Akcora et al. 2012), and accounts that can exist in several platforms (Aghasian et al. 2017; Li et al. 2018; Aghasian 2019). The assessment of network-based privacy scoring can be assessed with centrality measurements (Pensa et al. 2019) or by aggregation of user risk in the network (Liu and Terzi 2010). Privacy scores can be used as a method to reduce visual privacy risk.

In Chapter 5, I return to the visual privacy mitigation and protection concepts; proposing visual privacy risk scoring methodologies in SMNs and smart city infrastructure. The research presented in Chapter 5 extends privacy scoring methodologies, adapting existing privacy scoring methodologies to include visual privacy features and creating visual privacy risk scores to include visual features, computer vision, and privacy severity weighing. Fur-

thermore, Chapter 5 explores the efficacy of the visual features and privacy risk scoring across three visual datasets.

2.4 Ethics, Fairness, and Privacy

Visual privacy research creates and develops technologies and algorithms that can be integrated into various environments. Machine learning (ML) algorithms used for visual privacy and end to end privacy-aware systems are developed and deployed with the hopes of mitigating individual, stakeholder, and platform risks. Researchers have explored visual content to understand how attackers can extract textual information, including credit card numbers, social security numbers, residence, phone numbers, and other information (Li et al. 2017c). Visual privacy algorithms and systems have been implemented for individuals in their daily lives (von Zezschwitz et al. 2016; Dimiccoli et al. 2017; Gurari et al. 2019; Korayem et al. 2016), on social media networks (Zerr et al. 2012a; Tierney et al. 2013; Kuang et al. 2017), and in smart cities (Kunchala et al. 2023; Li et al. 2021a). Studies have also built visual privacy-aware systems to aid blind people who use social media networks (Gurari et al. 2019). In this study, researchers have collected a dataset of visual data with the use of a mobile application that allows the participants to consent to their photos being used in this study. Before making the visual dataset public, the authors removed private objects to protect each participant. However, it is not always the case that researchers obtain consent or clean raw data of private and sensitive information.

From these works, there is a range of applications and the broad impact that they can have on society. When building these algorithms and systems, issues with fairness and privacy can seep into the pipeline. One of the most widely used models for computer vision is the Convolutional Neural Network (CNN) (Hendricks et al. 2018; Simonyan and Zisserman 2014; Ranjan et al. 2017). To understand bias in visual recognition tasks, CNN models have been explored and strategies have been development to mitigate bias (Wang et al. 2020).

Measuring social biases in vision and language models leads to a direction of studying biases from a mixture of language and vision (Ross et al. 2020). A comparison study using multiple visual datasets was performed to help to understand how biases could be in datasets and affect object recognition task (Torralba and Efros 2011). Ethical concerns arise in facial processing technology. are described in (Raji et al. 2020). Specifically, auditing products need to be cautious about the ethical tension between privacy and representation. A framework for protecting users' privacy and fairness has been proposed by Soklic et al. (2017). The framework blocks harmful tasks, such as gender classification, which can generate sensitive information to certify privacy and fairness in the face verification tasks.

The accuracy and precision of these systems can depend on (1) the data collection process, (2) fairness forensics performed on the data, (3) human-*over*-the-loop techniques during model training, and (4) post-training evaluation. Chapter 6 approaches this problem by addressing these existing limitations on traditional visual privacy systems and suggests auditing strategies for the ML pipeline. The proposed solution considers privacy and fairness issues at each phase of the pipeline. In Chapter 6, I explore the application of human-*over*-the-loop and extend that technique to incorporate fairness and visual privacy auditing systems to allow researchers to create safe and fair systems to mitigate further risks for individuals, stakeholders, and developers.

2.5 Summary

The scope of visual privacy research seeks to understand, investigate, and explain visual privacy leakage in several environments and technologies. Advancements in this field happen in several areas of research including qualitative research studies, the development of privacy mitigation techniques, increasing privacy awareness through privacy risk quantification, and system evaluation using interactive auditing strategies. With the advancement of technology and vast quantities of visual data being uploaded or captured in technology and environ-

ments, this field has continued to grow. These factors also contribute to the importance of visual privacy research for increased consumer and stakeholder privacy awareness, scalable privacy-preserving techniques, and technological adaption as the world advances.

Typically, privacy on social media allows users to control privacy settings and permissions. Previous research focusing on SMNs explores user attitudes, user-centric privacy settings, and interpersonal conflicts, but does not consider cultural trends and keywords that correlate to visual privacy leakage. Unfortunately, there is little to no protection for the privacy of the bystanders in visual content for smart city environments. Additionally, there lacks a concise and universal definition of a smart city; which results in varied expectations, concerns about individual visual privacy, and a lack of legal and technical regulations for visual privacy. The development of detection, mitigation, and scoring methodologies will not be enough to adequately reduce privacy leakage without the consideration of the subjectivity of privacy, increased privacy awareness, and direct applications of auditing evaluations in each phase of the research process.

Chapter 3

Discovering User Attitudes and Beliefs about Visual Privacy on Social Media

Online privacy has become immensely important with the growth of technology and the expansion of communication. Social media networks have risen to the forefront of current communication trends. Users of social media networks share billions of images and videos daily. Users share private information within visual content, intentionally or unintentionally. This chapter explores (1) the users' perspective of privacy, (2) the pervasiveness of privacy leaks on Twitter, and (3) the threats and dangers of these platforms.

Through this investigation, the state of privacy on social media networks is explored, focusing on the occurrence of visual privacy leaks and the future of privacy for its users. This investigation creates a foundation to understand the users' concerns about the possibility of not having secrets in the future and the threats that emerging technologies will expose them to. In summary, the purpose of this chapter is to understand:

- The privacy perspective of a user can be subjective (Section 3.1). With this investigation, I uncover users' privacy subjectivity within age and sex demographics. This

section also demonstrates the differences in user perspectives between privacy and visual privacy.

- This study shows visual content privacy leaks that are common among users on Twitter (Section 3.2). With this investigation, I explore *severe* and *moderate* visual privacy leaks on Twitter.

- Several threats and dangers are heightened due to the visibility, accessibility, and sensitivity of visual content posted on social media (Section 3.1.4). This work provides an understanding of the most threatening dangers to SMN users and a hierarchy of dangers in correlation to the survey participants' rankings.

In this investigation, I uncover the importance of privacy and the growing need for evolving technologies to combat online threats. Previous works have discussed concerns with visual content, focusing on multiparty conflicts, third-party applications, privacy settings, and the danger of this content. The current state of this field shows the importance of continuing investigation and development of visual privacy and mitigation techniques to protect SMN users. The future of this field is in developing visual privacy mitigation techniques and helping users understand the correlation of privacy to threats and dangers on these networks. With this foundation, I explore visual privacy on SMNs through participant surveys, data collection, and analysis from Twitter. This work details the attitudes and perspectives toward visual privacy and the data collection results from Twitter.

3.1 Attitudes and Perspectives towards Visual Private Information

The user's perspective of freedoms, beliefs, ownership, and vulnerabilities aids in guiding their decision-making process when engaging on social media. On these platforms, users determine what information to share based on their perceived freedoms and feelings of security.

Each user's perspective of privacy will vary based on their subjectivity.

3.1.1 Survey Overview

The online surveys were conducted among users of social media. Respondent recruitment and data collection were conducted between January 2019 to November 2019. The surveys asked participants about their knowledge, experiences, and perspectives on social media networks. It further investigated participants' social engagement behaviors and visual privacy leaks. Questions in these surveys were influenced by related works (Abdulhamid et al. 2014; Madejski et al. 2011; Srivastava and Geethakumari 2013).

Participants were recruited through an invitation email with the survey link to various groups and organizations requesting their voluntary participation in an online survey. A total of N = 268 respondents took the survey and those responses were used for data analysis purposes. The participants were not required to answer every question. The order of the survey questions were randomized and the survey concluded with demographic questions (Schaeffer and Presser 2003; Kalton and Kasprzyk 1986). The first survey (IRB #10299) is used to gauge participants engagement across various platforms. The second survey (IRB #11349) focused on participants use of Twitter and their engagement with visual privacy. No financial incentive was provided to participants for these surveys. With exploring engagement on Twitter I can further understand what type of information disseminates, a user's network or community privacy leakage, and explore trends.

Participants and procedure

Of the $N = 268$ respondents, $n = 199$ (74%) respondents identified their sex. From the $N = 199$ respondents, $n = 96$ (48%) identified as male, $n = 102$ (51%) identified as female, and $n = 1$ (0.5%) identified as other. The average age of respondents was between 18–25. In terms of race and ethnicity, N = 199 (74%) identified their race. From the $N = 199$ (74%) responses, $n = 98$ (49%) were White, $n = 80$ (40%) were Black or African American, $n = 6$

(3%) were American Indian or Alaskan Native, $n = 12$ (6%) were Asian, $n = 2$ (1%) were Indian, $n = 15$ (7.5%) identified as other, and $n = 4$ (2%) preferred not to answer.

Social Media Engagement Measurements

From the surveys, I look at the participant's use of social media networks, definitions of privacy, and observations of privacy leaks on social media networks. Of the $N = 268$ respondents, $n = 111$ responded to how many hours a week spent on social media. In terms of the frequency of hours per week, $n = 39.6\%$ (42/111) participants use social media for 10–20 hours per week, $n = 44\%$ (49/111) participants use social media for 11–20 hours per week, and $n = 18\%$ (20/111) participants use social media for more than 21 hours per week. On average, participants engage in SMNs for 11–20 hours per week (Table 3.1B).

Most participants have multiple SMN profiles from different platforms (e.g., Facebook, Twitter, Reddit, Snapcat, Instagram, Pinterest, Tumblr, Flickr, LinkedIn, Twitch, and YouTube). When exploring which social media platforms participants use, N = 120 responded. Of the $N = 120$, $n = 96\%$ (116/120) identified that they have multiple social media accounts, and $n = 4\%$ (4/120) identified that they have one social media account. In this survey, the leading platforms are YouTube ($n = 98$ participants), Instagram ($n = 80$ participants), and Snapchat ($n = 65$ participants), which are all image and video-based platforms (Table 3.1A). From this survey, respondents ($N = 161$) seem to post images and videos the most across SMNs. Forty-seven percent of participants post images ($n = 75$), and 6% of participants post videos ($n = 11$); only 42% of participants post textual content (Table 3.1C).

Table 3.1 recaps the questions asked of the participants in the first survey. Each response that contains a text entry from participants was analyzed using text analysis methods, which are discussed in later sections of this chapter. The responses were segmented into categories: age and sex classification; then analyzed using Elbow–Knee plots, clustering, and feature weights.

Table 3.1: This table displays the outline of the IRB #10299 survey that was completed by participants. The survey was compromised of multiple-choice questions and short answers.

Item	**Question**
A	Of what Social Media Networks (SMNs) do you consider yourself a frequent user? (Multiple Choice)
B	How many hours per week do you spend on social media networks? (Multiple Choice)
C	What type of content do you usually post on social media? (Multiple Choice)
D	Do you post any of these types of images or videos on your SMNs? (Multiple Choice)
E	How would you define privacy? (Text Entry)
F	Would you define privacy the same for social media networks? (Multiple Choice)
G	Personally identifying information is information that can be used to uniquely identify, contact, or locate a person. Agree or Disagree? (Multiple Choice)
H	Privacy leaks include any instance in which a transfer of personal identifying visual content is shared on Social Media Networks. Private visual content exposes intimate information that can be detrimental to your finances, personal life, and reputation. Agree or Disagree? (Multiple Choice)
I	Would you consider any of these images to have identifying information? (Multiple Choice)
J	As a typical user of Social Media Networks (SMNs), if you were to post these items would you consider these items to be private? (Multiple Choice)
K	Drag and drop the following dangers in order of most threatening. (Multiple Choice)
L	Do you believe there are other dangers on Social Media Networks? (Text Entry)
M	What type of threat would these items fall under? (Multiple Choice)
N	Do you believe that conflict (e.g., bullying, domestic disputes) can increase the occurrence of privacy leaks? (Multiple Choice)

3.1.2 Pre-Processing Raw Survey Responses

Once the data collection process was complete, the raw data was preprocessed using text analysis (Pedregosa et al. 2011b), natural language processing (Loper and Bird 2002), and regular expressions (Sarkar 2019). The preprocessing phase of the data was conducted using *word tokenization*, *lemmatization*, *stopword removal*, and by combining words/acronyms with the same meaning. These methods are further discussed and defined in Appendix B. During preprocessing, each text entry was split into sentences and tokenized to return several word tokens. The tokens were split on the whitespace between words. After getting tokens, there were words with the same meaning but spelled differently. In these cases, I manually create a bag of similar words to make explicit substitutions (i.e., birthdate, birthday, bday). Finally, lemmatization was performed using the WordNet dictionary (Miller 1998).

The application of these techniques serves to decrease the margin of error and enhance the outcomes of the ML pipeline in the context of training and analysis. Preprocessing the raw data converts the data into a digestible form due to the machine learning algorithms' inability to grasp unstructured text completely. These preprocessing steps aid the algorithm in better understanding the underlying concepts. As text data may contain noise, such as emojis and punctuation, appropriate cleaning measures become imperative.

3.1.3 Surveyed Definitions of Privacy

To begin the analysis of these survey results, I analyzed the common themes from their definitions of privacy. Among the 250 participants that completed the survey, 154 participants responded to Question E in Table 3.1. To obtain the most meaningful terms from the participants' answers, I compute the Term Frequency-Inverse Document Frequency (TF-IDF) scores for each term and then take the average score across all documents where the terms appear. This process is further described in Appendix B.1.

In participants' responses, the words `information`, `personal`, `private`, and `share` are

the most relevant words used to detail how the participants' envision privacy on social media and in the real world (Table 3.2). *Privacy* can be defined as the ability to preserve or withhold information that is considered private, personal, and includes any content that an individual does not want to be shared. Of the $N = 154$, $n = 144$ participants were adamant that their definition of privacy would not change regarding their physical life or a digital one.

Table 3.2: The weight for each term is computed by averaging the Term Frequency and the Inverse Document Frequency (TF-IDF) weights over all the responses.

Term	**Avg. (TF-IDF)**
information	0.1418
personal	0.1254
private	0.0785
share	0.0655

Is Visual Privacy Defined Differently?

From the $N = 154$, $n = 10$ (6.5%) participants state that their definitions of privacy and visual privacy would not be the same. This definition change could be attributed to the participant's feelings about the levels of privacy, the unknown factors that exist on a digital network, or fears of exploitation by companies or scammers on these platforms. The keywords listed in Table 3.3 are similar to the words referenced by other groups however, these terms were combined to form a different definition that represents their concerns. They focus on the *information gain* of companies, *lack of control* over your privacy, and the *risks* on those platforms. This group emphasized how social media and visual privacy creates more risks for users. From this insight, *visual privacy* can be defined as a breach of sensitive information being accessed through visual data that leads to heightened privacy risk for individuals due to revealed information and the inability to exert control over data. The prevalence of visual content and the growth of social media are perceived to open more doors for attacks and

dangers.

Table 3.3: Top words used to define *visual privacy* using TF-IDF weights for each document.

Term	**Avg. (TF-IDF)**
private	0.1513
information	0.1280
media	0.1083
social	0.1083
share	0.0952

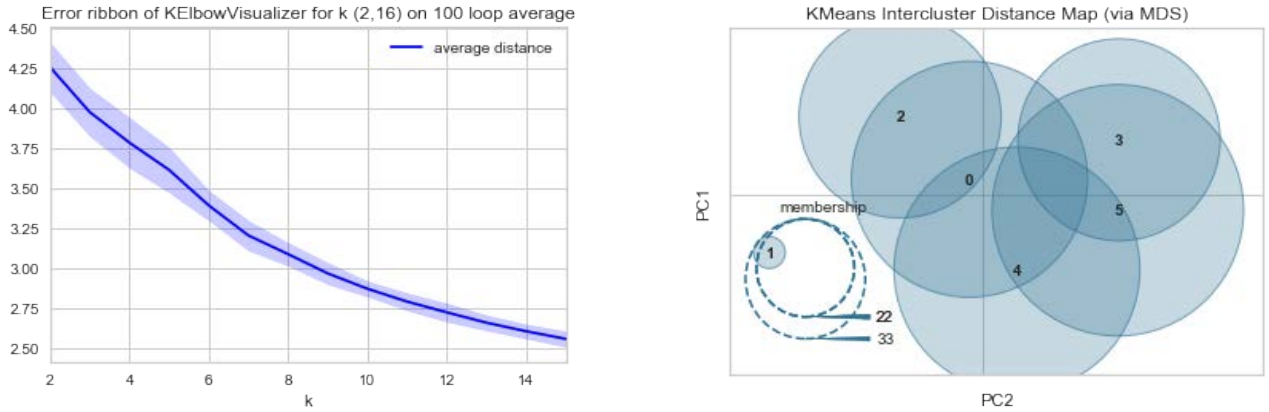
Privacy Definitions by Cluster

The definitions were further clustered into two groups using the K-means clustering algorithm and *yellowbrick* clustering via KElbowVisualizer (Bengfort et al. 2018). I deployed a K-means clustering model and found the elbow of the data using the *yellowbrick* clustering package. For evaluation, I used the *Calinski Harabasz* method to find the optimal cluster size (Caliński and Harabasz 1974). This method computes the ratio of distribution between clusters and the distribution of points within the clusters. The other scoring metrics provided by *yellowbrick* include distortion and silhouette. I chose the *Calinski Harabasz* method because it gave separation focusing on intracluster similarity and intercluster differences — rewarding the best clustering based on the total size and number of clusters. This method uses Equation 3.1.

$$\frac{SS_B}{SS_W} * \frac{N-k}{k-1} \tag{3.1}$$

In this equation, SS_B is the overall intercluster distance, SS_W is the overall intracluster distances, N is the total number of data points, and k is the number of clusters. Using this scoring method, I ran the KElbowVisualizer to find cluster values ranging from 2 to 16. In this process, I noticed that the elbow occurred at different k values across the runs.

To alleviate this issue, I averaged 100 runs from the model to find the optimal elbow point. Based on the lowest error and consistency of performance, the average elbow was at k= 6 (Figure 3.1a). The intercluster distance shows how strong the correlation is between the clusters and keywords. The embeddings of the cluster centers in 2D are shown in the intercluster distance maps, where the proximity between centers is maintained, representing their original feature space closeness. The cluster sizes are determined based on a scoring metric, *membership*, which reflects the number of instances assigned to each cluster's center. From Figure 3.1b, Cluster 1 completely differs from the remaining 5 clusters. However, cluster 0 and clusters 2–5 have a strong overlap of feature space.



(a) (b)

Figure 3.1: Diagrams of the Elbow–Knee scores and errors using Calinski Harabasz method. (**a**) Diagram of the Error Ribbon for the Elbow–Knee Plots using Calinski Harabasz method with cluster size (k) ranging from 2 to 16. (**b**) Diagram of Intercluster Distance Map using YellowBricks Calinski Harabasz method with cluster size (k) = 6.

In those clusters, the zeroth cluster included 36 definitions and the words: `information`, `share`, `private`, and `want` (Table 3.4). In Cluster 0, privacy revolves around the users' authorization, freedoms, and rights on these social media platforms. Cluster 1 included 33 definitions and included the words: `social`, `address`, `security`, and `information`. In this cluster, I derived the theme of protecting personal information regarding physical boundaries, digital security, and personal identifying information. Cluster 2 included 27 definitions and included the words: `know`, `want`, `people`, and `information`. In Cluster 2, I hypothe-

size the definition of privacy emphasized security in the connections/relationships on social media and their ability to share information at the owner's discretion. Cluster 3 included 30 definitions and included the words: `information`, `right`, `ability`, and `control`. The definition for Cluster 3 focuses on accessibility and knowledge of others. The participants in this cluster want to protect themselves from their information being shared in the public domain and keep the information disseminated in a controlled environment. Cluster 4 included 16 definitions and included the words: `personal`, `information`, `passwords`, and `control`. The participants in group four defined privacy as the access and use of information. It is the user's right to control access to data to keep their information safe. Cluster 5 included 13 definitions and included the words: `personal`, `private`, `identify`, and `share`. The last group of participants focus on individual subjectivity about privacy with a focus on personally identifiable information (e.g., social security number, address). Table 3.4 provides a synopsis of each of the cluster keywords and scores to show their level of importance to each respective cluster.

Privacy Definitions by Sex

These definitions were further broken down into sex classifications. Of the $N = 154$ responses, $n = 82$ identified as male, and $n = 71$ identified as female. Of the female participants ($n = 71$), the most important words were `information`, `personal`, `private`, `share`, and `social`. Of the male participants ($n = 82$), the most important words were `information`, `personal`, `want`, `control`, and `private`. It is further noted that the female participants are more concerned with personally identifying information regarding harm and hacking, while the male participants are concerned with financial attacks and exposed information.

In Table 3.5, the top five keywords are identified for each sex classification. For further investigation, I began to look at the statistical analysis to find the significance of the words for each sex classification subgroup. In this process, I used the analysis of variance (ANOVA) method to analyze the differences between the groups of my participants. In this

Table 3.4: Cluster breakdown of the top terms and definitions of privacy using the average TF-IDF weights.

Cluster (Total Size)	**Keyword**	**Weight**	**Interpretation**
0 (36)	information	0.1651	User authorization, freedoms, and rights on SMNs
	share	0.1086	
	private	0.0870	
	want	0.0770	
1 (33)	social	0.0984	Protecting personal information regarding physical boundaries, digital security, and personal identifying information
	address	0.0938	
	security	0.0883	
	information	0.8171	
2 (27)	know	0.1420	Security in the connections or relationships on social media and their ability to share information at the owner's discretion
	want	0.1315	
	people	0.1201	
	information	0.1075	
3 (30)	information	0.1813	Protection of self from information being shared in the public domain and keep the information disseminated in a controlled environment
	right	0.1442	
	ability	0.0898	
	control	0.0874	
4 (16)	personal	0.3792	User's right to control access to data to keep their information safe
	information	0.3785	
	passwords	0.1119	
	control	0.09364	
5 (13)	personal	0.2380	Individual subjectivity about privacy with a focus on personally identifiable information
	private	0.1271	
	identify	0.0769	
	share	0.0769	

work, ANOVA is used to observe the statistical variance for several data groups by different components. This methodology helps uncover information about the relationships between the dependent and independent variables. Table 3.6 explores the statistical values produced from this analysis.

To understand the significance of each category, I look at the null hypotheses and p-values associated with the independent and dependent variables. The null hypothesis states that there are no significant differences in sex demographics and the associated keywords. The p-value threshold is set at 0.05. In the sex comparisons of Female vs. Male; I reject the null hypothesis because differences exist in the keywords for sex demographics.

Table 3.5: Top five terms used to define privacy using average TF-IDF weights for each by sex demographic.

Sex Classification (Total Size)		**Keyword**	**Score**
Female	(71)	information	0.1361
		personal	0.0985
		private	0.0751
		share	0.0705
		social	0.0400
Male	(82)	information	0.1346
		personal	0.1038
		want	0.0631
		control	0.0545
		private	0.0542

Table 3.6: Calculated ANOVA score for top words used among sex classification to define visual privacy.

Sex Classification Comparison	***f*-Value**	***p*-Value**
Female vs. Male	5.9749	**0.0061**

Privacy Definitions by Age Group

The clusters in this section were determined by how many individuals were in each group. As the age increased, the population of the participants decreased significantly. Because of

this, the definitions of privacy were clustered into two age groups: 18–25 and 26 & up. Of the N = 154 responses, n = 111 were between the age of 18–25 and n = 43 were at least 26 years of age and up. In participants whose ages ranged between 18–25, the most important words were `information`, `personal`, `right`, `control`, and `private`. The definitions of this group revolve around the preservation of information from hackers and government surveillance. Of the participants that are age 26 and over, the most important words were `information`, `personal`, `anything`, `private`, and `share`. The second group's definitions seem to center around a common theme of personal information in relation to external sources while considering alternative factors that could play a role in the dissemination of information. The word `anything` asserted that the control or dissemination of any content is defined by the owner.

In Table 3.7, the top five keywords are identified for each age group. For further investigation, I began to look at the statistical analysis to find the significance of the words for each age subgroup. In this process, I used the ANOVA method to analyze the differences between the groups of my participants. Table 3.8 explores the statistical values produced from this analysis. The null hypothesis states that there are no significant differences in age groups with respect to the keywords. In the age comparison of 18–25 v. 26 & over; I retain the null hypothesis because differences do not exist in the keywords for age.

3.1.4 Attack Vectors and Existing Dangers

From this survey, I investigated what users perceived to be privacy leaks and the dangers of exposed leaks on social media networks. I asked participants if they would consider certain items to be privacy leaks. From this question (Table 3.1J), I see that 97% of participants identify credit or debit cards, driver's licenses, social security numbers, and passports as the highest-ranked privacy leaks. Following close behind are birth certificates (96%), phone numbers (90%), personal letters (85%), and keys (83%). Participants did not consider images of babies and children to be a privacy leak if posted on social media by their guardians.

Table 3.7: Top words used to define privacy using TF-IDF weights for each age demographic. This table includes the cluster label and the top five words by the highest TF-IDF weight.

Age (Total Size)		**Keyword**	**Score**
18–25	(111)	information	0.1167
		personal	0.0777
		right	0.0523
		control	0.0512
		private	0.0492
26 & over	(43)	information	0.1395
		personal	0.1212
		anything	0.0869
		private	0.0805
		share	0.0585

Table 3.8: Calculated ANOVA score for top words used among age groups to define visual privacy.

Age Group Comparison	***f*-Value**	***p*-Value**
18–25 v. 26+	0.3275	0.5776

Sixty-five percent of participants state that they have seen these types of privacy leaks on social media networks. From this, participants identify keywords or phrases that correlate to those privacy leaks. With this investigation, I uncovered hashtags and words such as *#stayoffthesidewalk, #licensedtodrive*, and *#racisttwitter*. The majority of participants stated that they do not recall the phrase that was used in correlation to the images but did notice privacy leaks on their news feeds. The words collected from this survey were used for the data collection process in Section 3.2.

Next, I ask participants to rank dangers (e.g., burglary, kidnapping, stalking) in reference to what seems to be most threatening (Table 3.1K). The dangers listed in the survey were defined by the author. The top dangers are kidnapping, burglary, and stalking. Table 3.9 displays the percentage of votes for the threat in each position. Along with these dangers, participants also mentioned cyberbullying, echo chambers, and social isolation (Table 3.1L).

For further investigation, I began to look at the statistical analysis to find the significance

Table 3.9: Dangers are listed in their respective order based on survey results. The column shows a rank between 1 – 5, and the rows indicate the dangers. The associated vote percentage for each threat is shown; the highest vote for each rank is highlighted. The underscored values denote the highest vote value for each threat.

Threat	**Rank 1**	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**	**Rank 6**
Kidnapping	**52.38**%	15.48%	10.71%	3.57%	9.52%	8.33%
Burglary	20.24%	**35.71**%	17.86%	10.71%	7.14%	8.33%
Stalking	5.95%	14.29%	**25.00**%	16.67%	23.81%	14.29%
Financial	4.76%	14.29%	23.81%	**30.95**%	17.86%	8.33%
Identity	14.29%	17.86%	14.29%	25.00%	**23.81**%	4.76%
Explicit Sites	2.38%	2.38%	8.33%	13.10%	17.86%	**55.95**%

of the dangers for each subgroup. In this process, I used the ANOVA to gain insight into the perspective between the sex classification and dangers from the participants. Tables 3.10 and 3.11 explore the statistical values produced from this analysis.

To understand the significance of each category, I look at the null hypotheses and p-values associated with the independent and dependent variables. For each category, my null hypothesis states that there are no significant differences by sex in the respective category. The p-value threshold is set at 0.05. In the categories of burglary, explicit websites, and identity theft; I reject the null hypothesis because differences exist by sex. For the categories of kidnapping, financial theft, and stalking; I retain the null hypothesis because no significant differences exist for sex identities. Within the male and female clusters, the female group displayed a higher concern for the threat of being posted on an explicit website unlike their male counterparts.

Table 3.10: Statistical analysis of sex-related differences of danger assessment results using the ANOVA method. The table shows the f-value and p-value for the Female and Male. Asterisks(**) denote the p-values that are significant (¡ 0.05).

ANOVA analysis of danger distribution among sex classification						
Statistic Value	**Burglary**	**Kidnapping**	**Explicit websites**	**Financial Theft**	**Identity Theft**	**Stalking**
f-Value	5.2662	2.8248	6.0343	1.8150	4.8928	2.9822
p-Value	****0.0063****	0.0629	****0.0031****	0.1668	****0.0089****	0.0541

Table 3.11: Statistical analysis of age-related differences of danger assessment results using the ANOVA method. The table shows the f-value and p-value for the age groups: 18–25 & 26 and over. Asterisks(**) denote the p-values that are significant (¡ 0.05).

ANOVA analysis of danger distribution among age groups						
Statistic Value	**Burglary**	**Kidnapping**	**Explicit websites**	**Financial Theft**	**Identity Theft**	**Stalking**
f-Value	0.3491	0.4125	4.1532	3.7922	3.5000	5.2348
p-Value	0.7059	0.6628	****0.0178****	****0.0250****	****0.0330****	****0.0064****

For each category, my null hypothesis states that there are no significant differences in age for the respective category. In the categories of explicit websites, financial theft, identity theft, and stalking; I reject the null hypothesis because differences exist in age. For the categories of burglary and kidnapping; I retain the null hypothesis because no significant differences exist for the age demographic. With this investigation, I found that the age group 26 & over has a higher concern for identity theft. While their younger counterparts tend to have a higher concern for financial theft, explicit websites, and stalking.

The participants allocated privacy leaks into three possible attack vectors (Table 3.1, M). The location attack vector is used to find out where an individual lives and/or current location. The participants classified keys, passports, driver's licenses, social security cards, and personal letters as an item in location threat. The identity attack vector is used to exploit an individual's identity, even to the intimate details. The participants classified credit/debit cards, children's images, driver's licenses, social security cards, passwords, and personal letters as an item in identity threat. The asset attack vector is used to gain access to an individual's possessions and valuables. The participants classified credit/debit cards, keys, passports, driver's licenses, social security cards, passwords, and personal letters as an item in asset threat.

3.2 Data Collection via web Crawling

To understand the pervasiveness of privacy leaks on SMNs, I ingested tweets from the Twitter API using the participants' described keywords. Each key term was given by a survey participant. In the initial survey, I asked users to define categories of privacy leaks based on keywords. Next, I examined the keywords that are related to those categories. I collected tweets and images from Twitter, resulting in approximately 1.4 million tweets collected and 18,751 images. I collected data using notable keywords derived from the survey participants' responses. This data was collected over a two-month time period.

3.2.1 Tweet Collection

The initial dataset was a collection of 1.4 million tweets that were analyzed by the associated hashtags (Table 3.12). Twitter was searched with keywords derived from the privacy leak categories and the words given by participants. From the survey, the participants gave keywords, hashtags, or phrases that they have seen used on Twitter that were related to a perceived visual privacy leakage (e.g., *#stayoffthesidewalk* includes images of license plates). To find a correlation between related images and hashtags, I searched these phrases using the Twitter API. The top hashtags from this search were *#racisttwitter* and *#wikileaks*. Of the tweets collected (N = 1,465,091), n = 18,751 contained images.

From the tweets collected, the most relevant results are from the college search, which includes the keywords college acceptance, college bound, and college letter. In this search, I collected trending hashtags in reference to college searches: `#neumannscholarship`, `#nmsubound`, and `#hu24`. These hashtags are associated with college acceptances, scholarship acceptances, and college letters.

3.2.2 Image Collection

Beyond collecting basic tweets, I searched for images associated with keywords and hashtags in the collected tweets. With this search, I collected 18,751 images. The images collected were classified into three categories based on risk: *severe*, *moderate*, and *no risk*. (1) *Severe* risk content contains images that have more than one attack vector (Section 3.1.4). These images include items that show government-issued identification (i.e., social security numbers, driver's license, etc.), items that can be used to identify a person and/or used for facial recognition (driver's license, identification cards), or items that contain information aobut a person's location and/or place of residence. (2) *Moderate* risk content refers to images from the asset or identity attack vectors. This content includes images that feature items that can be used to identify a person and/or can be used for facial recognition. However,

Table 3.12: Results of keyword crawling on Twitter. The keywords searched are under Keywords/Phrases, and the total amount of collected images from all the searched keywords is on the right side of the table.

Keywords/Phrases	**# of Tweets Collected**
credit card debit card	364,825
job offer job acceptance job letter	107,470
key house key car key	174,348
license licensed to drive driver's licenses	109,520
passport	183,048
password passwords	166,835
#racisttwitter	121,638
college acceptance college bound college letter	100,199
#wikileaks	137,208
Total	**1,465,091**

this content will not provide the user's location, or place of residence, nor feature any of their government-issued identification. (3) *No risk* content encompasses images that do not include any of the above items. The images were classified by three individuals and placed into categories based on the average agreement. Table 3.13 shows the total number of images in each category and its respective category after agreement and assignment.

Table 3.13: Risk Classification from keyword search with Twitter.

Category	**# of Images Collected**
Severe	160
Moderate	327
No risk	18,264

In each category, I examine the privacy risk for each image by keyword. In *Severe*, car keys, license plates, and job offers are the most prevalent images. In *Moderate*, the most prevalent images are work identification, school information, and job promotion letter images. In the *No risk* category, I observed that the search contained advertisements and spam content. Table 3.14 shows the keyword distribution of privacy leakage among the *Severe* and *Moderate* privacy risk categories.

The prevalence of images has a higher frequency for the terms `baby`, `hospital`, `medication`, and `medical records`. *Severe risk* includes images containing finances and keys, unlike *Moderate risk*, which contained more college and work-related images. When asking users to define privacy and identify threats, the participants did not identify hospitals, medical records, or medications as significant concerns. I find that these images trending on Twitter about medical information and hospitals have a higher chance of occurring than the other keywords.

3.3 Conclusion

This chapter addresses the first research question of this dissertation, "*What are the privacy-related experiences and concerns of Social Media users regarding visual content and threats*

Table 3.14: Distribution of content for privacy risk categories. This table includes the keyword and the content frequency.

Category		**Keyword (Count)**	
Severe	(160)	Baby	71
		Driver's License	12
		Financial Document	2
		Hospital	54
		Job	4
		Keys	1
		License Plate	4
		Medication	10
		Medical Records	6
Moderate (327)		Baby	45
		College Letter	6
		Driver's License	24
		Hospital	123
		Job Promotion	7
		Medical Information	52
		Medication	43
		Work Identification	12
		Workplace	15

on these platforms?" (Table 1.1). The analysis confirms that age groups have different levels of concern regarding explicit websites, financial theft, identity theft, and stalking. It also confirms that female and male participants have differences in the level of concern regarding burglary, explicit websites, and identity theft. The threats on these platforms are heightened because of the accessibility of social media. From this analysis, I find that cyberbullying and explicit content can be seen as low threats to participants. The results do not fit the hypothesis that visual privacy leaks are common on Twitter; however, rare breaches in privacy may still be devastating. The reliability of the data from Twitter is limited by the keyword search terms used.

As new trends arise and challenges appear, the keyword associations for the appropriate images change. From the survey, 65% of participants stated that they had seen visual privacy leaks on social media networks; however, I could not collect a corresponding amount of visual privacy leaks. In this study, I collected words regarding trends and challenges associated with visual content. From this data, I see that the most accurate keyword search was regarding `college bound` and `college acceptance`. I also find that images trending on Twitter about medical information and hospitals have a higher chance of occurring than the other keywords.

These results build on existing evidence of previous work regarding the dangers of social media (Rosenblum 2007; Gross and Acquisti 2005; Veiga and Eickhoff 2016) and the subjectivity of privacy (Rosenblum 2007; Such et al. 2017b). In this chapter, I explored the user's thoughts on social media privacy, particularly visual privacy. As new technologies arise, developers must implement mitigation techniques that allow users to explore the trade-offs between privacy and freedom. This will be increasingly important for non-text and visual sharing methods across SMNs.

Chapter 4

Understanding Stakeholder Perspectives of Smart City Environments

Citizens hope that a smart city can improve their quality of life, provide transparency about the city's data, consider the cost to citizens, and implement strategies that protect their privacy. The city governments emphasize concerns of the expenses to implement a smart city, the engagement of citizens in the city, the amount of data collected in the city, and the physical safety of citizens in a smart city. The term "smart city" is widely used, but no comprehensive definition exists. This can leave citizens and stakeholders are unsure what a smart city means for their community and how it affects cost and privacy.

In this chapter, I conduct a study to understand the stakeholder perspectives on privacy and cost in deploying smart cities. Additionally, I investigate the finalists' applications from the 2015 Smart City Challenge (DOT 2015) to understand the similarity about the concepts of a smart city, the common technologies that were requested, and the privacy considerations of each of the finalists' proposes. This analysis emphasizes understanding potential visual privacy leakage and cost considerations that can arise in smart cities through their technology

and infrastructure. I use text analysis techniques to investigate themes using document similarity, and topic modeling. Furthermore, I discuss proposed solutions to cost and visual privacy for smart city environments.

In Section 4.1, I conduct a survey to understand the perspectives of smart city stakeholders in relation to privacy and cost. In Section 4.2, I perform a detailed textual analysis of the finalists' smart city applications. Proposed solutions to the cost and visual privacy issues can be found in Section 4.3.1 and Section 4.3.2, respectively. Furthermore, I describe a case study of a privacy-enabled low-cost smart city technology implemented in a U.S. city in Section 4.3.3. Finally, this chapter is summarized in Section 4.4.

4.1 Surveying Stakeholder perspectives of Smart Cities

Smart city officials and stakeholders were surveyed through an online survey (IRB #13565) to gain insights into their perceptions of smart city technologies and infrastructure. Respondent recruitment and data collection were conducted in July 2021 and August 2021. Despite the global development of smart cities, the technological advancements associated with them can create difficulties for citizens, stakeholders, and governments. The stakeholders in this study refer to individuals who are tangentially involved with smart city implementation or governance in developing smart cities across the country. The survey focused on stakeholders' knowledge of privacy, their respective government or company involvement with developing smart cities, and the current cost of data collection with pedestrian counting devices. Participants were asked about projected costs spent on pedestrian counting technologies, their knowledge of smart city efforts, and their understanding of privacy expectations (see Table 4.1).

The survey was comprised of 26 questions and the average time to complete the survey was 27 minutes. Stakeholders were recruited through email invites. An invitation email with the survey link was sent to the stakeholders in local government, utility companies,

and industry companies of developing smart cities requesting their voluntary participation in the online survey. A total of $N = 100$ email invites were sent out via email and a total of 9 responded. Of the $N = 9$ respondents, $N = 5$ completed the survey and were therefore used for data analysis purposes.

4.1.1 Participants and procedure

Of the $N = 5$ completed surveys, $n = 2$ identified as female and $n = 3$ identified as male. The average age range of 45–54. In terms of race, $n = 5$ where White and n =1 identify Hispanic ethnicity. Of the respondents, $n = 4$ work for local government, $n = 1$ work for a utility company, and $n = 1$ works for a non-profit. The positions of the respondents in their organization were Transportation Engineers ($n = 1$), Information Technology staff ($n = 3$), and Operations ($n = 1$). Looking at the geographic information provided by the respondents, n = 2 work in Louisville, KY, $n = 1$ work in Greensboro, NC, $n = 1$ work in Aurora, IL, and $n = 1$ work in Cornelius, NC. Of the $N = 5$ respondents, $n = 3$ participants work in a city (Louisville, KY and Greensboro, NC) that applied to the Smart City Challenge (DOT 2015).

4.1.2 Analysis

Respondents had insightful answers when asked to define smart cities (Table 4.1, Q1). One participant highlighted "pressing issues for its residents and businesses" in their response and defined a smart city as:

> "One that employs technologies to improve services to the community and/or make government operations more efficient and effective. A truly smart city/community should also be targeting the most important and *pressing issues for its residents and businesses* not just applying technology for technology's sake." – Survey Respondent A

Another respondent stated that a smart city is:

> "A city whose residents are connected by technology, high-speed broadband, providing services online and interactively, telehealth services, using IoT and AI in traffic management, air quality management, parking, waste management, public safety, utilities, autonomous vehicles, etc." – Survey Respondent B

Beyond understanding what people define as a smart city, I wanted to gain insight into the privacy concerns in potential and deployed smart cities (Table 4.1, Q2). When asked to define privacy, participants highlighted the need to be able to revoke access to their data.

> "...I would include the ability to control or at least delete personal data as well that has been collected especially if the data has become obsolete or inaccurate." — Survey Respondent C

When asked about what data privacy protection methods would help improve their willingness to participate in city data sharing (Table 4.1, Q3), several participants stated they would like the "ability to review" any data collected by the city concerning them. If data is collected anonymously, there is an inherent difficulty when designing systems to review personalized data requests. To solve this, respondents suggest using blockchain or smart contract techniques to provide anonymous keys that support audit requests.

Analyzing the current results, I found common concerns around privacy. The words *personal, private, uninvited surveillance and protect* are frequently used to describe and articulate how privacy is visualized for pedestrians and companies in smart cities. The survey further asks about data sharing (Table 4.1, Q4). Participants were asked if they were comfortable sharing their data for developing and enhancing smart cities. However, the results show participants are skeptical about sharing their data with smart cities. The reasons provided by the participants included possible increased policing in under-served communities, vulnerability to data leakage, and not being aware of the purpose of data collection (Table 4.1, Q5).

Table 4.1: The survey was compromised of multiple-choice questions and short answers. This table shows selected questions from the survey related to defining a smart city, perspectives of privacy, technology, and spending.

Number	**Question**
1	How would you define a smart city/community?
2	How would you define privacy?
3	What data privacy protection methods would increase your willingness to share data with the city?
4	Would you be comfortable sharing personal data within these smart communities?
5	What makes you feel uncomfortable with sharing your personal data within smart communities?
6	How do you use the pedestrian counting data – for what purpose(s)?
7	How much do you spend annually on pedestrian counting data?
8	Where are the locations you need to have pedestrian counting data?

Although companies use pedestrian counting for marketing, economic development, safety, and infrastructure development (Table 4.1, Q6), I also found that some of the challenges concerning pedestrian counting are the cost and frequency of pedestrian counting. The most common places for pedestrian counting include intersections, downtown, or shopping areas (Table 4.1, Q8). Participants also believe that pedestrian counting devices will be useful at streetlights. Within the scope of smart city infrastructure, the survey respondents stated that privacy is the most valued feature, with simplicity as the second most valued. Privacy focuses on protecting data and individuals, while simplicity focuses on the technology's ease of use in the smart city. According to the survey responses, companies spend between $11,000 - $20,000 annually on counting pedestrians (Table 4.1, Q7).

4.2 Analyzing Smart City Finalist Applications

In 2015, the United States Department of Transportation announced the Smart City Challenge, which asked cities in the U.S. to create an integrated, smart, and efficient transportation system built on data, applications, and technology in an effort to improve the lives of their citizens (DOT 2015). The Smart City Challenge received 78 applicants describing what a smart city looked like for their community. From this challenge, the seven cities chosen as finalists include Columbus (Ohio), Austin (Texas), Denver (Colorado), Kansas City (Missouri), Pittsburgh (Pennsylvania), Portland (Oregon), and San Francisco (California). Figure 4.1 displays U.S. cities that were applicants for the 2015 Smart City Challenge, of these, the red circles denote the seven finalists (the circle area denotes the population size).

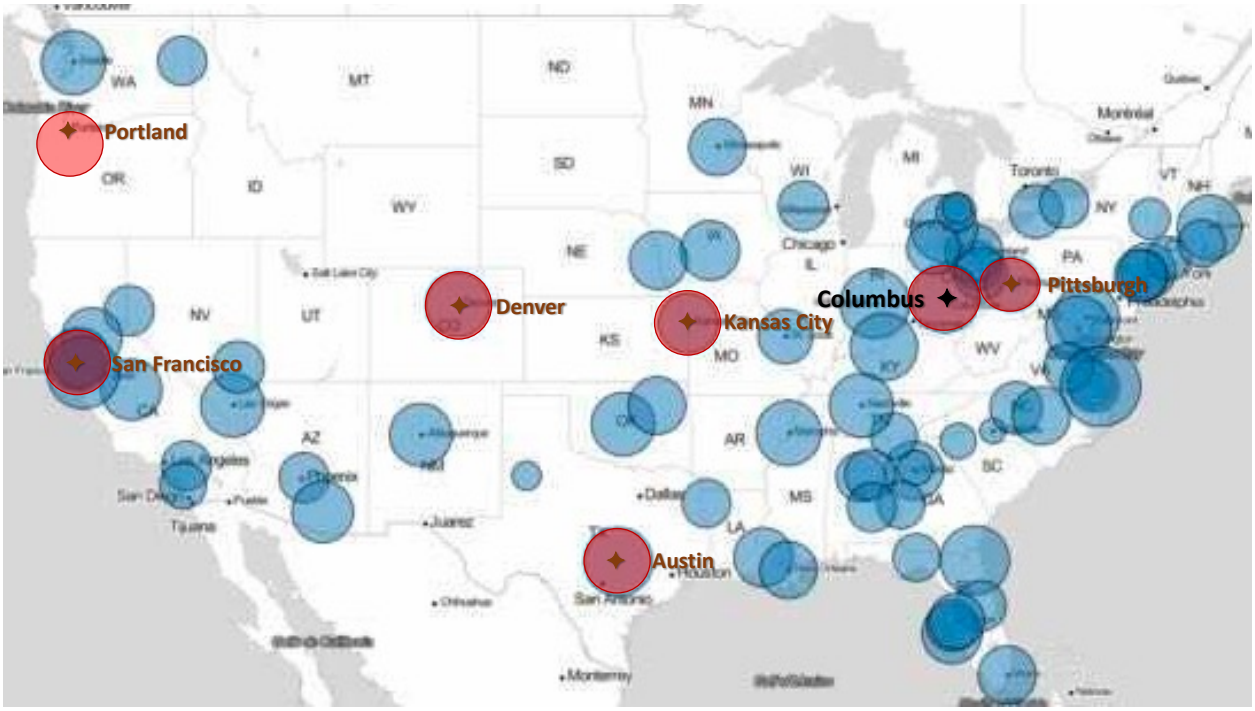


Figure 4.1: Smart City Challenge Applicant locations in the United States; red circles denoted the seven finalists.

To understand the potential technology, infrastructure, and policy designs for rising smart cities, I evaluate the proposals of the finalists from the Smart City Challenge. I perform text analysis for each of the finalist applications. I first describe the document preprocessing

methods used to transform the PDFs into a usable format (Section 4.2.1). I then perform cluster analysis to group the finalist applications, and I analyze overlap in the application requests (Section 4.2.2). Additionally, I performed topic modeling to derive the dominant themes present across the documents and provide insights on what a "smart" city is compromised of (Section 4.2.3). Furthermore, I provide details on the requested technology (Section 4.2.4) and a discussion of privacy mechanisms (Section 4.2.5) that the Smart City Challenge finalists considered for implementation.

4.2.1 Preprocessing Smart City Finalist Applications

Each of the finalists' applications was downloaded from the Smart City Challenge website, where their vision statements were made publicly accessible as a PDF file (of Transportation 2016). The textual content was extracted from the files with Python code using the PyPDF2 PDF manipulation library (Fenniak 2013). Figure 4.2 shows the distribution of

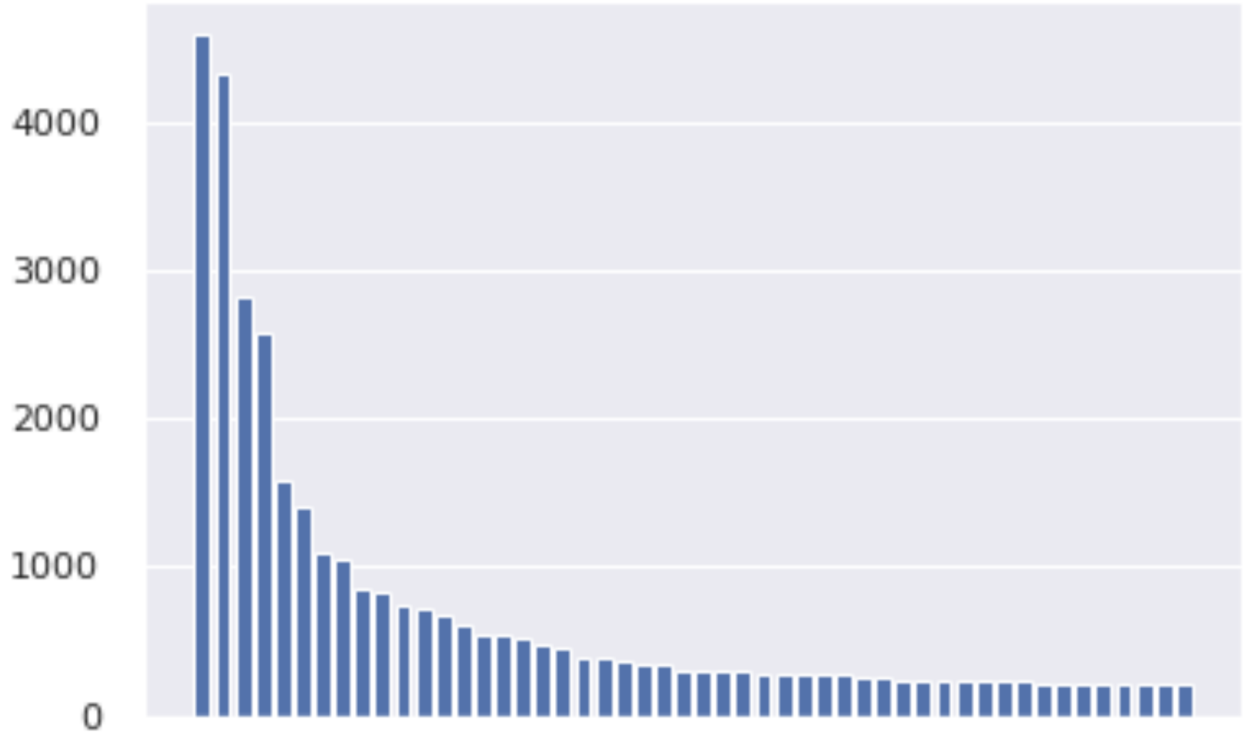


Figure 4.2: Token Frequency Distribution across the Smart City Finalist Corpus. The higher frequency tokens are conjunctions and common words.

word tokens across all documents with a truncated tail. The documents were cleaned by removing stopwords, alphanumeric text, unstemmed words, and words with high TF-IDF weights. Stopwords are removed using a list of typically infrequent words, misspellings (e.g. "asd", "buisness"), common words (e.g., "the", "a", "is"), and specific city names. The process of removing alphanumeric terms can alleviate typos as well as unsupportive words. Another pre-processing method I used was stemming. I used the Porter Stemmer to remove the endings of words to set them to the root (Porter 1980). When using this Stemmer, you will notice endings such as "ing", "ed", and "es" being removed. TF-IDF weights are calculated (Jones 1972) using the equation in Appendix B.1. The methods mentioned in this section are further described in Appendix B.1. With these preprocessing techniques, I remove terms that add little to no meaning to the content of the topics and themes. The repetition and frequency of irrelevant text can influence the text analysis results if not handled. With this collection of cleaned documents, I created a corpus used in the analysis steps. In this study, I will continually refer to a *corpus* as the collection of the preprocessed Smart City finalists' applications. Using text analysis, I can extract the text from the applications to create machine-readable information to perform machine learning.

4.2.2 Analyzing the Clusters of the Smart City Finalists

Cluster analysis was performed to group similar documents together. The documents found in the same cluster are more similar than those in other clusters. The cluster analysis was completed using K-Means clustering (MacQueen 1967). K-Means is an iterative centroid-based clustering method that creates groups based on closeness or similarity. It uses expectation maximization to place the centroids at an optimal location in the data space such that similar documents are in a cluster and dissimilar documents are not clustered. For the K-Means algorithm, I must define a k value, which is the number of clusters the K-Means model should produce. To obtain the k value, I evaluated the elbow of the corpus by fitting the model to various values of k between two and six. This elbow analysis of a corpus helped

to determine the optimal number of clusters for the respective corpus (Satopaa et al. 2011). The optimal k value was found when the cluster size is set at 4. The documents were then passed into the K-Means model to cluster the documents.

To visualize the clusters, I used Principal Component Analysis (PCA). PCA is traditionally used as a dimension reductionality method. I employ PCA to create a visualization that helps understand the clusters – I choose the first two principal components as the axes of a two-dimensional plane. The axes show how far the intercluster and intracluster distances for each cluster. This cluster visualization is shown in Figure 4.3.

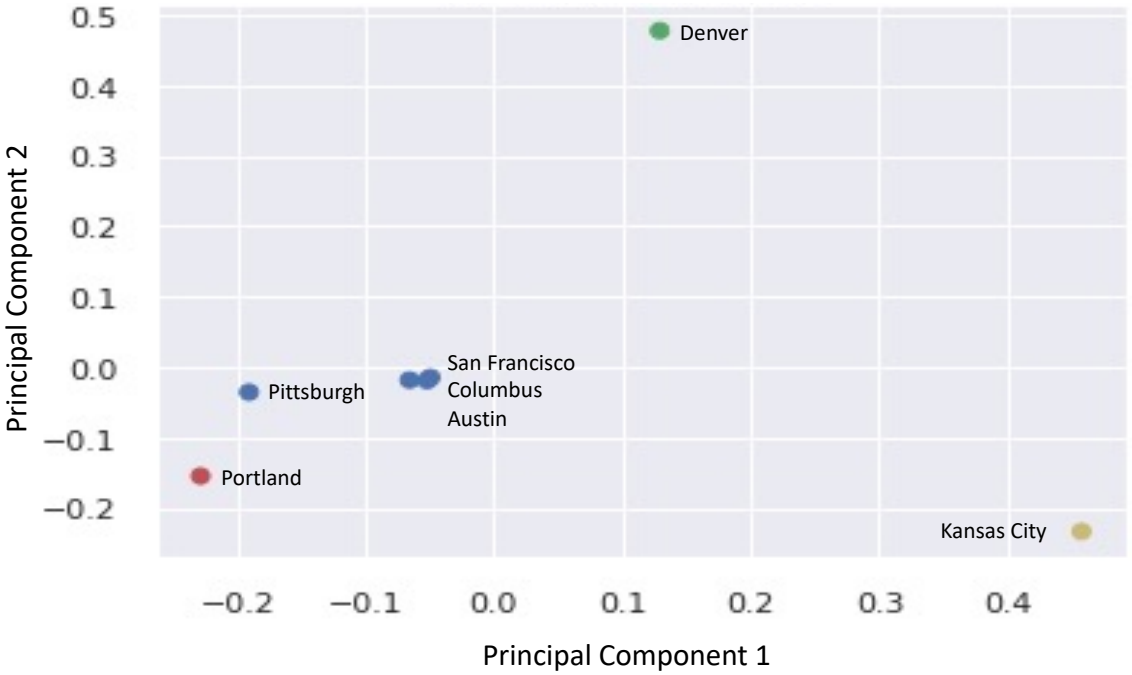


Figure 4.3: Two Component PCA for visualizing K-Means Clustering for Smart City Challenge Finalists.

The cities of `Denver`, `Portland`, and `Kansas City` are individual clusters which imply that they differ significantly from one another as well as from the large cluster. The larger cluster is comprised of the cities: `Pittsburgh`, `San Francisco`, `Columbus`, and `Austin`. The content in these documents is closer in similarity to each other and further from the other clusters. The centroid of this cluster is `Columbus`, with that applicant having the average document similarity in the cluster. It is also noted that there is heavy overlap in

`San Francisco`'s and `Austin`'s applications.

4.2.3 Deriving Common Themes with Topic Modeling

To start the topic modeling process, I define phrases and vocabulary from the corpus. When building the word dictionary for this model, I choose the words that appear in more than two documents but less than 90% of all documents. After this process is complete, I create a Latent Dirichlet Allocation (LDA) Topic Model (Blei et al. 2003). LDA can produce weighted topics based on the analysis of the corpus. The corpus consists of seven documents and has a vocabulary size of 2,282 words. With this model, I can derive themes and topics representative of the corpus. Topics are a list of weighted terms; I utilize the top-k. The LDA model creates three topics that are used to discover themes for the corpus. When the number of topics was larger than three, there were several duplicates which yielded less unique or distinct themes and meanings. In Table 4.2, the three topics are displayed with their respective words and themes.

Table 4.2: Topics and themes of the smart city finalist derived from the LDA model. The groups are listed with associated cities, topics, and themes. The topics listed contain the top ten words.

Group	Cities	Topics	Theme
1	Columbus, Kansas City, San Francisco, Austin	grant, proposal, event, digital, automated, university, demonstration, automated vehicle, deploy, tool	Autonomous Technology and Tools
2	Denver, Pittsburgh	component, grant, department transportation, university, benefit, consortium, efficiency, foundation, percent, avenue	Building Partnerships and Infrastructure
3	Portland	device, efficiency, equity, percent, market place, university, cloud, engineering, payment, benefit	Connecting the Collegiate Experience to the City

The terms denoted in gray have little to no contribution to the theme of the cluster. These

terms are used based on their assigned weight from the output values of the LDA model and the assigned topics. From Table 4.2, Topic #1 includes four cities (`Columbus, Kansas City, San Francisco, Austin`), Topic #2 includes two cities (`Denver and Pittsburgh`), and Topic #3 includes one city (`Portland`). These applications focus on several topics, but the overall similarity of the document content allowed the model to group the documents in the corpus and create themes to represent the groups. The documents were assigned to these groups by their dominant topic. The themes derived from these groups encompass the goals that these smart cities have. From these themes, it is implied that cities can become "smarter" with the use of autonomous technology (Topic 1), building partnerships and infrastructure (Topic 2), and connecting to the local universities in the city (Topic 3).

4.2.4 Understanding a Smart City through Technology

To define the essence of a smart city, I investigate the universal technologies requested by smart cities. I introduce definitions needed to build a basis for understanding the foundation of the technologies requested by these Smart City finalists. These definitions provide a foundation to understand the type of connectivity and technology smart cities require to be operational. There are additional technologies, networks, and sensor infrastructures that are not mentioned in these findings that cities can implement in their community.

Many cities are interested in *Dedicated Short-Range Communications* (DSRC), which allows vehicles to communicate with each other and other road users directly (Wu et al. 2013; Tokuda 2004). It is a wireless communication technology that can function properly without involving cellular or other infrastructures. It can save lives by cautioning drivers of a looming, threatening situation or occurrence in time to take necessary actions to help evade the situation.

Cities are also interested in technologies that improve efficiency for travelers. *Traffic Signal Priority* (TSP) can be defined as a technological set of operational improvements to shorten the wait time at traffic signals for vehicles and prolong the time for green light

signals (Smith et al. 2005; Hounsell and Wu 1995; Garrow and Machemehl 1999). This can be done by using the existence of vehicular locations and wireless communication to extend the time of the green light at a traffic signal. TSP can be implemented at street intersections. Additionally, pedestrian counters can be implemented in these intersections as well. *Pedestrian counters* can be defined as an electronic device used to classify, count, and measure pedestrian traffic along roads (Yuan et al. 1993; Alahi et al. 2022). These counters can also be used to measure the direction of the traffic by time and location. With this technology, corporations can find peak traffic times, identify entry and exit points of travelers, and set travel management protocols. Smart kiosks can serve as a gateway for pedestrian counting as well. A *smart kiosk* is an information kiosk that can detect and track pedestrians; it can also send and store information about pedestrians and engagements as data for usage (Sánchez-Corcuera et al. 2019). These kiosks can serve as a connectivity vector between the citizens, the city, and surrounding technologies like looking for available parking. *Smart parking* technologies can be defined as a strategy that infuses technology to inform citizens about free and occupied parking spaces over the web or applications (Fahim et al. 2021). These technologies can be a quick resource for travelers and reduce the time and consumption of fuel.

Cities are also considering transportation methods to reduce vehicle emissions and air pollution in the community. *Electric transportation* is any vehicle whose propulsion and accessory systems are powered exclusively by a zero-emissions electricity source. Electric transportation vehicles have rechargeable batteries. One electronic transportation method is E-bikes, which use rechargeable batteries battery mounted on the bike frame for power. and the Another transportation method is electric buses that have a battery located under the hood or in a protective barrier. Cities are interested in planning personal and public charging stations to support electric vehicles. Similarly, cities are interested in promoting *autonomous transportation*, or vehicles that drive with minimal human intervention. Also called driverless or self-driving vehicles, autonomous transportation requires detailed real-

time environmental sensing to detect surrounding objects along navigation pathways. Cities should also understand the evolving transportation regulations around the public deployment of automated vehicles. These electric and autonomous vehicles can include cars, scooters, bikes, and buses (Azad et al. 2019; Campbell et al. 2010).

Table 4.3: Requested Technologies from Smart City Challenge Finalist. The technologies are listed in descending order. Technologies can be requested by all cities.

Technology request	**Number of Cities**
Smart Traffic Signals	7
Web Applications	7
Electric Vehicle Charging Station	7
Use of Sensors	7
Use of WiFi/Communications	7
Use of Cameras	7
Autonomous Vehicles	6
Connected Vehicles: DSRC technology	5
Smart grid	3
Use of GPS	3
Kiosks	3
Use of Cellphone signals	3
Autonomous home delivery	3
Smart Parking	3
Bike and/or pedestrian Counters	2
Electric Bus	2
Information screens for bus stops	2
Road condition monitors	2
SMART roadside lights	2
Traffic Management Centers	1
Universal smart access card	1
Bike sharing	1
Transportation Hubs	1
Interactive Voice Response	1
Smart Pedestrian Guides	1

In Table 4.3, I display the requested technology from the Smart City Challenge finalist applications. Among the seven finalists, 25 unique technologies were requested in their proposals. The average city requested 12 technologies to be used in their smart city. The amount of technology requested by a city could depend on the population, as seen in Fig-

ure 4.4. The city requesting the least amount of technology is `San Francisco, CA`, with

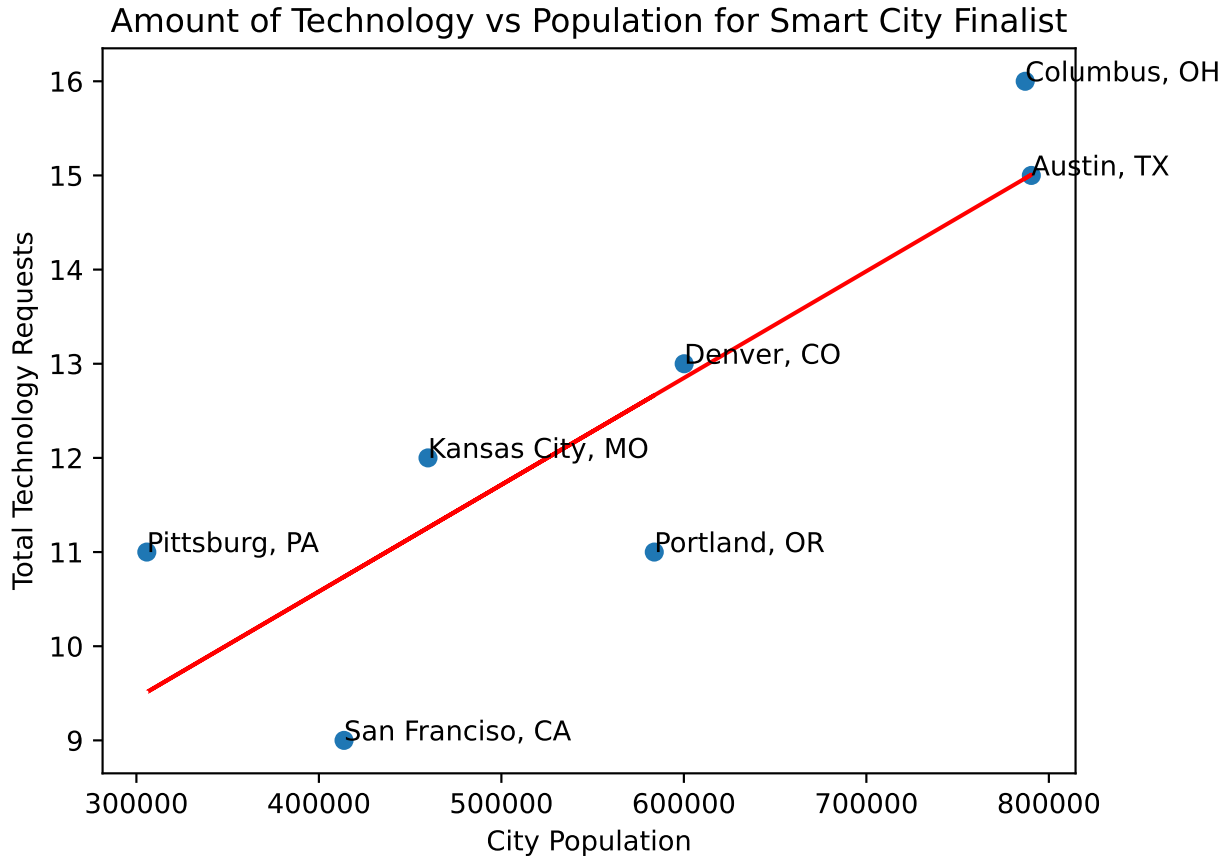


Figure 4.4: Comparing the city's population size with the amount of technology requested. A linear regression line shows the projected fit for the cities.

nine potential technology integrations into their smart city. Following close behind are the cities of `Portland, OR` and `Pittsburgh, PA` with 11 technology requests. The remaining cities had 12 technology requests (`Kansas City, MO`), 13 technology requests (`Denver, CO`), 15 technology requests (`Austin, TX`), and 16 technology requests (`Columbus, OH`). To integrate these technologies, the cities use sensors, video, Global Positioning Systems (GPS), and radio signals from pedestrians, vehicles, and equipment. These cities also use these video and GPS feeds for license plate recognition and to track crime-related incidents. The technology requested by the cities supports the goal of becoming a smarter city revolves around connecting communities to opportunities, decreasing health disparities, reducing air pollution, and increasing the mobility of citizens by relieving congestion of roadways. Assisting

low socioeconomic and disabled citizens have risen to the forefront of smart city development strategies, as mentioned in the finalists' applications To make these advancements more inclusive of those communities, the finalists have proposed the use of the following technology:

- *Smart kiosks* enable advanced payment options by incorporating additional features, such as braille and voice feedback
- *Electronic signs* can provide visual and audio cues to pedestrians crossing intersections
- *Autonomous car sharing* allows commuters first and last-mile transportation with a reduction in costs
- *Information screens* provide real-time transportation updates through audio and video

With the incorporation of these additional technologies, these cities hope to become more inclusive and smarter for all. On top of an already costly smart city, these specialized technologies introduce additional expenses tied to continuous maintenance and privacy integrations for supporting the aforementioned technologies.

4.2.5 Privacy Considerations in Smart Cities

A major concern for citizens in literature is understanding how increased city technologies will affect their privacy (Smith 2019; Harvey 2012; Doctorow 2020; Devlin 2020). Furthermore, cities will become a 24-hour hub for collecting information about the mobility and efficiency of transportation, but also personally identifying information of its' travelers (Sánchez-Corcuera et al. 2019). In the Smart City Challenge (DOT 2015), I examine how the finalists describe privacy risk and mitigation strategies for deploying technologies to their cities. The main privacy concerns listed by the cities include data sharing, individual privacy, system security, data privacy, and data management.

In Table 4.4, the Smart City Finalist proposals were reviewed and assessed a score based on a Likert Scale (Excellent, Average, Poor) from the five central privacy concerns found in the documents: Data Sharing, Individual Privacy, System Security, Data Privacy, & Data Management. From the proposals, I rated each city's proposal for the five privacy categories:

- Excellent: The proposal thoroughly discusses the privacy risks, mitigation strategies related to the topic, and a thorough plan of action for incidents that can arise.
- Average: The proposal has moderate to little discussion about the privacy risks, mitigation strategies related to the topic, and a general plan of action for incidents that can arise.
- Poor: The proposal has little to no discussion about the privacy risks, mitigation strategies related to the topic, and no plan of action for incidents that can arise.

I provide a definition for each of the five categories to describe the clarity of the topic in the documents. **Data management** outlines access control procedures, storage schema, and storage policies for smart city data and databases. **Data privacy** entails the encryption of items in the data and what information is stored from the citizens and anonymization schemes. **Data sharing** includes the procedures and policies by which the smart city data will be shared with organizations, entities, or the public. **Individual privacy** focuses on protecting citizens in the city. This protection could include but is not limited to, encryption schemes, consent documents, and privacy mitigation techniques. **System security** details the overall protection mechanisms for the smart city infrastructure.

Data sharing and data privacy concerns are addressed by the majority (4 of 7) of the cities. Strategies for addressing data sharing included access management, encryption, and anonymization. Individual privacy, system security, and data management categories are each addressed by three of the cities. The winning city, Columbus, is the only city with no discussion about these privacy concerns in its proposal. Of the finalists, none of these cities provide a detailed discussion of the privacy protection they will provide their citizens

Table 4.4: Rating of Privacy discussion by City. Each city receives a rating (poor, average, or excellent) based on five categories.

City	Data Sharing	Individual Privacy	System Security	Data Privacy	Data Management
Columbus, OH	Poor	Poor	Poor	Poor	Poor
Austin, TX	Poor	Poor	Excellent	Excellent	Excellent
Denver, CO	Poor	Poor	Poor	Average	Poor
Kansas City, MO	Poor	Average	Excellent	Poor	Poor
Pittsburgh, PA	Poor	Poor	Poor	Average	Poor
Portland, OR	Average	Poor	Poor	Average	Average
San Francisco, CA	Poor	Poor	Poor	Poor	Poor

in their proposals. These proposals focused on infrastructure protection and security. The discussion of citizens' privacy protections focused on (1) implementing standards from the government and industry, (2) anonymizing or masking sensitive textual personal data, and (3) partnering with cyber-security experts and the government to support protection efforts.

4.3 Discussion: Proposed Solutions and Privacy-Enabled Technology Case Study

In this section, I provide additional interpretations and considerations for smart city infrastructure. Based on the insights from the analysis in Sections 4.1 and 4.2, I propose methods to create a low-cost and privacy-enabled smart city. Sections 4.3.1 and 4.3.2 describes potential solutions to accomplish these features. Furthermore, in Section 4.3.3, I describe the possibility of incorporating these features with an existing smart city technology and discuss its' privacy-enabled and low-cost features.

4.3.1 Proposed Solution: Low-Cost Smart Cities

The Smart City Challenge finalists' provided no insight or discussion on the projected cost of development or maintenance of the environments; however, smart city projects can be expensive to deploy and manage. The technology in the survey (Section 4.1) focused on the cost of pedestrian counting technology, but that is not the only technology that can be implemented in a smart city. Cities around the world, such as San Diego, New Orleans, London, and Songdo, have either proposed or invested in smart city projects that cost between $30 Million and $40 Billion (DeHart et al. 2020a). With the smart city requesting, on average, 12 technologies to be implemented, stakeholders and citizens could expect the cost of smart city implementation and maintenance to soar. In addition to the cost of deploying and maintaining the IoT devices, a significant portion of the expense is due to providing Internet connectivity via 5G or WiFi to those devices. These costs are a major barrier to the widespread deployment of smart city technology and the social benefits that may ensue from that technology (Madamori et al. 2019).

To alleviate the costs, opportunistic communication, such as Delay Tolerant Networks (DTNs), can be used as a backbone for smart city communication to facilitate data that does not have real-time Quality of Service constraints. DTNs traditionally provide opportunistic networking connections in areas with little to no infrastructure. Messages are delivered with some delay directly correlated with the layout, density, and mobility of nodes in the network (Keränen et al. 2009; Hui et al. 2011). Recognizing that some data are needed in real-time, edge computing can be utilized as long as the placement of internet-connected nodes is optimized in the network. For data that can tolerate delays, the natural movement of people and vehicles through a city transfers data between nodes. In this way, the citizens become an integral part of the smart city network itself.

For low-cost smart cities to flourish with the use of DTNs as the backbone to be practical, the technical questions regarding the devices and network, as well as the social aspects of how

people and vehicles move through a city must be addressed. For almost 20 years, there has been a substantial amount of research in opportunistic communications and delay-tolerant networks; unfortunately, real-world deployments traditionally fall short of their simulated counterparts (Baker et al. 2017). Related efforts (Cabaniss et al. 2013; Costa et al. 2008; Daly and Haahr 2009; Gang et al. 2012; Hsu et al. 2012; Hui et al. 2011; Lindgren et al. 2004; Musolesi and Mascolo 2009; Gupta et al. 2019) have proven the ability to deliver messages when connections are intermittent but generally are limited to performing within simulation environments (Picu and Spyropoulos 2014).

4.3.2 Proposed Solution: Visual Privacy Enabled Smart Cities

From Section 4.1, survey respondents highlighted concerns for data privacy, data sharing, and surveillance, emphasizing a need for visual privacy considerations. Privacy concerns of stakeholders have been further explored by analyzing the finalist from the 2015 Smart City Challenge. The results from Section 4.2.5 show that stakeholders are not carefully considering privacy in their proposals and show no concern for visual privacy mitigation in their infrastructure and their citizens. From this investigation, smart cities have requested approximately 12 technologies and emphasized the integration of cameras throughout the infrastructure (see table 4.3). Cameras can be integrated into several technologies throughout a smart city, which can make them widely used in that environment. A city where facial recognition systems are used can lead to visual privacy leaks due to consent and individual privacy rights. While existing in a smart city environment, pedestrians carry identification, purchase items with credit or debit cards, use physical keys to enter restricted areas, and use virtual passcodes to access sensitive information. These types of private content will be captured in videos and image feeds (Hoyle et al. 2015; Korayem et al. 2016) in the environment. I have investigated the concerns of privacy leaks and the types of privacy leaks on social media (DeHart et al. 2020c). These privacy leak concerns can be expected in a smart city where citizens are continually being monitored. Previous works have provided a

foundation for visual privacy mitigation techniques used for social media networks; however, these same technologies can be implemented to protect citizens from surveillance concerns and privacy issues in smart cities.

Beyond the citizen's concern for anonymity or protection of minors, there is a concern for the type of information that is revealed in a public setting. I propose using visual privacy mitigation strategies for videos and images in smart cities based on existing literature (DeHart and Grant 2018). With the use of visual privacy mitigation techniques, there can be additional measures to ensure privacy and security for data sharing, individual privacy, system security, and data privacy. Studies have shown that obfuscation methods (Orekondy et al. 2018; Li et al. 2017b; Boult 2005), such as blocking and blurring objects, can protect individual privacy. These obfuscation methods can include blurring, blocking, adversarial noise, or replacing items in visual content. Methods such as blurring and blocking alter the pixelation of the visual content to provide distortion to the human eye. These methods can be added to objects, faces, and text in visual content. The technique of adversarial noise (Kurakin et al. 2016) adds a few pixels that can (1) impede a computer's ability to learn anything from the visual content even if it is in their possession, and (2) still allow the images to be visible to humans. To protect individuals' identities, studies have suggested face swapping (Korshunova et al. 2017; Zhu et al. 2020; Mahajan et al. 2017), which can switch detected faces of citizens with a collection of replacement faces.

To implement this solution, I suggest the deployment of visual privacy mitigation strategies to allow smart cities to implement mitigation techniques that are integrated into their technologies and systems. I propose mitigation techniques can be integrated into mobile applications, servers, IOT devices, and comprehensive systems (DeHart and Grant 2018). Visual privacy mitigation techniques can facilitate active privacy risk strategies for authorized personnel for analyzing pertinent privacy occurrences. This can provide additional security and privacy to the data and storage methods in smart city environments. Visual privacy mitigation can provide safety, security, and peace of mind to the citizens that reside

in those areas.

4.3.3 Case Study: Deployed Low-cost and Privacy-enabled Technology in a U.S. City

Smart city technology must be reliable, low-cost, and consider privacy to attract citizens to engage with those platforms. The Smart City Applications Platform (SCAP) is an example of a privacy-aware system coupled with reliable and effective management (DeHart et al. 2021a). It serves as a strong example for organizations to model pedestrian counting and computer vision technologies in smart cities. In this case study, SCAP is deployed in city *C*. SCAP was created by a major utility company. This platform consists of a complete hardware and software solution that identifies various moving objects common to an outdoor urban environment, such as bicycles, pedestrians, and scooters. At its core, SCAP is a Field Node with computer vision software that analyzes data from a high-definition camera on an edge compute device and transforms it into object count data (Figure 4.5A). The Field Node is available in a stand-alone enclosure or as an integrated subsystem of a digital information kiosk, as seen in Figure 4.6. The Field Node kiosk is integrated into city *C*'s downtown infrastructure. This data can be uploaded to the Cloud (Figure 4.5B) as anonymized statistics after data analysis is complete (Figure 4.5C). The data can then be viewed in a portal or accessed via an Application Programming Interface (Figure 4.5D).

To provide real-time data, the video analytics data is sent from the local device to the Cloud. Should the network connection be lost, data is queued in the Field Node compute device and transmitted once the network returns. This connection uses Message Queuing Telemetry Transport (MQTT) between the edge and cloud for communication. MQTT is a standard publish and subscribe technology that uses machine-to-machine communication with low bandwidth requirements. The cloud database is set up in a cluster for backup and redundancy purposes. SCAP utilizes a cloud-based user management system to control Portal and the Cloud API access. A username and password must be created to access

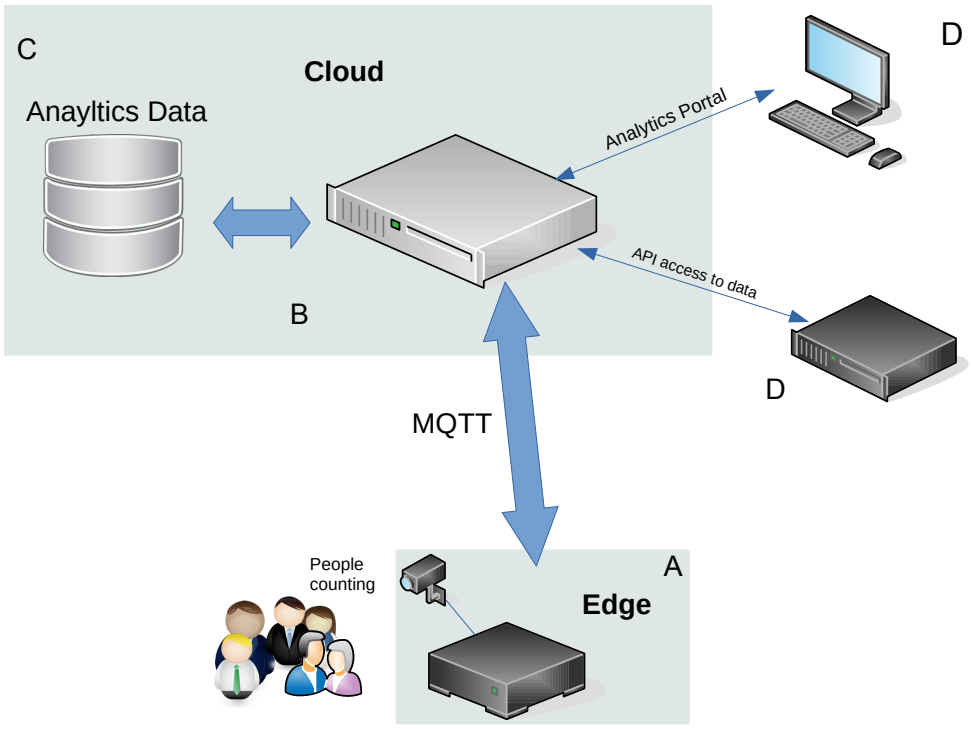


Figure 4.5: High-level overview of the deployed Smart City Applications Platform (SCAP).

any system information or data. The Platform is designed for utility-grade cybersecurity and network security standards. It is important to note that the SCAP software does not collect or record personally identifiable information, such as facial images, phone numbers, or mobile phone MAC addresses. Rather anonymized target object count data is collected and provided to the user. Furthermore, all video is processed on a local computer, and no images are recorded or stored, ensuring peace of mind for citizens and visitors.

Considering robust physical security, the SCAP Field Node or digital kiosk features an enclosure with a specially keyed locking system. Both the incoming and outgoing data to the Field Node are encrypted. Through the monitoring and control software, licenses for the Field Nodes can be remotely enabled or disabled. Each Field Node utilizes a computing

(a) Sidewalk view = (b) Opened view (c) Field Node

Figure 4.6: Field Node Designs provided by Smart City Applications Platform. (a) Field Node Integrated in a Kiosk, (b) Opened Field Node Kiosk deployed in the city, (c) Rendering of Field Node Integrated with a Light Pole (refer to Pole-Mountable Camera Support Structure, US Design Patent D902,985 S) (Cleveland et al. 2020)

device with storage capability. As a result, the larger system is unaffected if the Field Node becomes compromised.

While the SCAP Field Nodes can work with a variety of wired and wireless data back-haul networks, the most common type is anticipated to be cellular. One of the major advantages of the SCAP is that it has low bandwidth requirements. This allows the use of the lower bandwidth CAT-M1 network when cellular communications are required. As the Smart City Applications Platform is still in its infancy and undergoing field trials, there will be ample opportunity to reduce the cost of both system deployment and operations. For example, the complexity of mounting the Field Node equipment to appropriate street furniture or buildings will be simplified. As system requirements are better understood, optimization of the Field Node components will allow for a reduction in the Bill of Material costs as well as annual operating costs.

4.4 Summary

This chapter addresses the second research question of this dissertation, *"What considerations do smart city stakeholders give for visual privacy protection and cost in technology and infrastructure?"*. The insights from this chapter were drawn from deploying surveys and analyzing the Smart City Challenge applicants using text analysis methods. From the survey and Smart City Challenge analysis, I find that privacy and cost can continue to concern citizens and stakeholders in these environments. The Smart City Challenge analysis also alluded that the typical smart city will require 12 new technologies on average to become a smart city, which is more than a city with smart technology. With the creativity and development of smart city infrastructure, it is increasingly important to consider cost and privacy. In the Smart City Challenge finalist's proposals analysis, I found that their discussion of privacy and cost is not at the forefront of developer concerns; but rather technological innovation. The analysis and evaluation of smart cities using the 2015 Smart City Challenge and surveys are important to understand smart cities' infrastructure and the perspectives of individuals and stakeholders in those cities. It further demonstrates the disconnect between citizens and organizations who develop these smart cities in regard to privacy and cost. With citizens' input for smart cities, the organizations will be able to create inclusive, adaptable, and trusted relationships to aid in the acceptance and assimilation of the futuristic growth of the city.

In summary, this chapter argued that smart cities could be private and inexpensive in deployment and long-term sustainability. During the planning and implementation of these cities, officials and citizens should further consider the cost and privacy concerns associated with their development choices. The need for visual privacy mitigation in smart cities extends from the protection of personally identifying information to the choice of anonymity and protection of minors. Beyond the integration of visual privacy mitigation into infrastructure, I propose using DTNs to lower the cost of smart cities and allow citizens to assist in the

transmission of data across the city. Deploying traditional IOT infrastructure is prohibitively expensive for most cities, and expanded development introduces more privacy risks. However, low-cost smart cities and privacy-enabled technologies can achieve the goals of smart cities while allowing citizens to feel secure and protected.

Chapter 5

Proposing a Visual Privacy Risk Scoring Framework with Visual Feature Measurements

The growth and development of mobile devices, networks, and connected environments, exponentially increase the ease of capturing and sharing private visual content. Understanding privacy risk in visual content can be difficult and requires methods to describe and evaluate domain-specific intricacies. The definition of privacy risk can apply differently to people, to stakeholders, and within environments. The reduction of privacy risks calls for a need for mitigation strategies to support leakage across several domains. Researchers have begun quantifying individual, content, and network privacy risk scores using models and measurements. A privacy risk score is a common measurement used by researchers (Section 2.3). This score is a quantitative estimate of the privacy risk associated with the given information for content, a user, or a network. By evaluating the visual privacy risk, quantitative and qualitative techniques can derive meanings and show trends about content, individuals, and the network.

There have been efforts to gauge privacy risk in unstructured data, such as textual posts in

social media networks, user profiles, and networks (reviewed in Section 2.3). However, there exists a gap in quantifying the privacy risks of visual content. I explore three features of visual content in an effort to understand the correlation between visibility, appeal, and sensitivity in privacy risk scoring for images. These features were explored across various datasets to understand how the combination of these features can be applied to the quantification of visual privacy risk scoring methodologies.

This chapter seeks to improve the understanding of privacy risk in visual content, and explore the impact of quantifying visual features in privacy risk scoring. For the scope of this chapter, I narrow the discussion of privacy to existing dataset labels. The proposed concepts and algorithms in this chapter have the potential to benefit individuals by providing an idea of their current visual privacy risk score and what features contribute to the impact of their visual privacy risk score. It can also benefit stakeholders by providing privacy risk scores of visual content allowing time for adaptation and network management consolidation of privacy risk scores across the visual content. The visual privacy risk score proposed may be incorporated by social media networks and smart city environments infrastructure and can also be manually calculated by stakeholders engaging in these environments. Visual privacy risk scores with explainable feature components will enable individuals and stakeholders to regain control of their privacy leakage quicker and to mitigate risks sooner.

First, this chapter adapts an existing privacy risk scoring methodology to score visual content and incorporates object detection techniques to create a dichotomous application for scoring, *VPScorer*. This chapter considers visual content features as an important component of individual and stakeholder visual privacy risk. Second, I propose a visual privacy risk scoring framework, **V**isual **A**rea, e**N**coding, and **G**olden spiral **O**bject distance (Vango), that focuses on the privacy risk scoring of visual content using its features. Thus, providing a privacy risk scoring algorithm that uses a pre-trained object detection model and potential mitigation strategies. The chapter is organized as follows: Section 5.1 describes the visual content datasets, object detection model, and object labels. The privacy scoring algorithms

and visual feature metrics are discussed in Sections 5.2 and 5.3. Figure 5.1 depicts how each of the components interacts. Experiments and results are described and presented in Section 5.4. Furthermore, I discuss the main takeaways in Section 5.5. Throughout this chapter, I use the words privacy risk score and privacy score interchangeably.

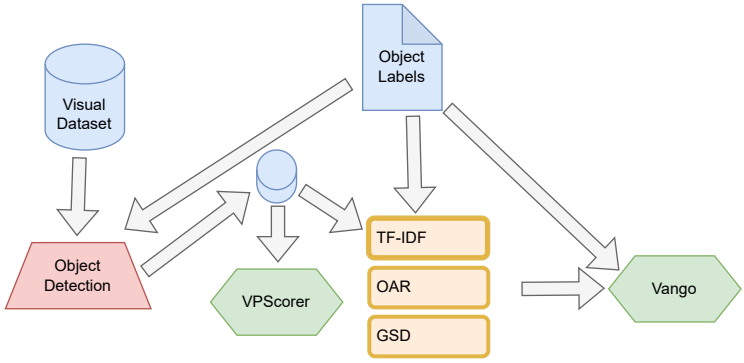


Figure 5.1: High-level overview of the Privacy Risk Scoring Pipeline.

5.1 Methodology

In this section, I describe the visual datasets used for the experiments, define private object labels using pre-existing datasets, outline the background and use of the object detection model, and expound on the types and applications of visual features for visual privacy.

5.1.1 Visual Content Datasets

I define a visual content dataset as a dataset that contains images and/or videos. For the scope of this chapter, I will focus on visual datasets that only contain images. The selected visual datasets are shown in Table 5.1. These datasets are used later to extract visual features and used in experiments for the visual privacy risk scoring algorithms. The datasets can be divided into two types: private image datasets and common image datasets. Private

Dayquan is a frequent user of a social media network called Snapcat. Dayquan has been on the social media site for years and has many family and friends on the site. Due to the economy, Dayquan lost his job and has been hunting for another gig. On Snapcat, Dayquan shares his daily job search attempts and outcomes to garner support and encouragement from his closest connections. One bright sunny day, Dayquan goes out to his mailbox to check the mail. To his surprise, he had received a job acceptance letter! With eagerness and pride, Dayquan quickly takes a picture of this letter and posts it to his social media. The caption reads, "*Never Quit! All praises be to God. #workingman #manofprayer #newmoney #igotajob*". Inadvertently, Dayquan has leaked his privacy. This job acceptance letter included his full name, home address, work address, start date, and employer. Once this letter was shared on Snapcat, his connections began to show their support for Dayquan by liking, commenting, and reposting. From these interactions, a friend of a friend sees Dayquan's post and decides to collect Dayquan's personal information. In this scenario, I shall call this friend of a friend, OPP. The privacy leak in the image allows OPP to choose a location-based attack vector, posing a threat to Dayquan's physical safety (burglary, kidnap, and stalking, among others). In retrospect, Dayquan did not recall posting an image with a privacy leak, so he was unaware of the risk and dangers that could potentially arise. Using visual privacy risk scoring methods can allow for quantifying privacy risk in images for a user and allow for more extensive privacy protection methods using visual privacy mitigation such as blurring, blocking, and censoring.

Figure 5.2: Scenario 1 – Social Media Privacy Risk Scoring

image datasets are collected focusing on images containing privacy leakage intended for visual privacy research. Computer vision image datasets are collected to be practical and contain common images with labels that focus on real-life entities intended for the machine learning community. The selected image datasets are used for object detection tasks to extract visual features from each object in the image. Across these datasets, the exploration of visual features can quantify potential privacy risks and leakage that exist within them.

A lawsuit is filed in a city, *C*. This city recently transitioned into a smart city and uses technology as a part of its community improvement initiatives. This year, a citizen was misidentified and wrongfully imprisoned as a criminal due to the video footage from the community improvement initiative, *Crime Stopper*. The lawsuit litigation process was started by filing a complaint about using smart city technology and data to accuse citizens of crimes in city C. The plaintiff believed that visual data about them was private information and had not agreed to disseminate the data. The settlement required city *C* to hire a Chief Privacy Investigator, create a system that included considerations of citizens' privacy, and make settlement payments to the plaintiff. City *C* brings in Dayquan as the Chief Privacy Investigator to create a privacy-centric system that can be integrated into the current smart city infrastructure. Dayquan implements a visual privacy risk scoring methodology that helped analyze the visual feeds from all over the city. By understanding the intensity and prevalence of leaked private content across the city, Dayquan implemented secure data management and storage mechanisms to keep the citizens' data safe.

Figure 5.3: Scenario 2 – Smart City Privacy Risk Scoring

Table 5.1: This table provides an overview of the visual content datasets and contains the year, number of images used from the dataset, the initial research domain of the dataset, and the machine learning task designed for the dataset.

Dataset	Year	Used Images	Domain	Tasks
Open Images v7 (Ferrari et al. 2022)	2022	77,891	Vision	Object detection
PrivacyAlert (Zhao et al. 2022)	2022	5,017	Privacy	Prediction
VISPR (Orekondy et al. 2017)	2017	10,221	Privacy	Prediction

Open Images v7 Dataset

The Open Images v7 dataset is the most recent update to the Open Images datasets (Kuznetsova et al. 2020; Ferrari et al. 2022). The Open Images datasets are used for computer vision tasks and contain over 9 million images. Open Images v7 dataset contains bounding box annotations for all images. The images were labeled and annotated using crowdsourcing. The images in this dataset contain (on average) 8.3 objects in each image and have diverse image scenes across the dataset. Some of the images include class labels like footwear, person, clothing, hair, and flower. The images contained in Open Images are collected from

Flickr and also overlap with other existing image datasets such as Flickr30k (Young et al. 2014) and Microsoft Common Objects in Context (MS COCO) (Lin et al. 2014).

PrivacyAlert Image Dataset

PrivacyAlert is one of the most recent image datasets for visual privacy research (Zhao et al. 2022). This dataset contains 6,400 images. The data is annotated into four categories: clearly public, public, private, and clearly private. In this study, private images should be kept confidential to the owner only and/or for selected people (e.g., nudity/sexual, other people, medical, drinking/partying). Public images were defined as any image that their entire social network could see (e.g., food, kitchen table, eat, dishes). Each image's privacy category was annotated using crowd-sourcing and inter-annotator agreement to select the final category. This dataset was originally collected from Flickr and used for image privacy prediction tasks.

Visual Privacy (VISPR) Image Dataset

The VISPR Image Dataset was created in 2017 (Orekondy et al. 2017). The dataset has 22k images with annotations. In this study, the researchers gathered a sample of images from Open Images v4 (Young et al. 2014) and manually gathered images from Flickr and Twitter. Privacy prediction is based on the list of privacy attributes that an image can disclose. The private attributes were identified into 68 categories which include physical disability, receipts, sports, date of birth, license plate, occupation, and religion. The multi-label task used crowd-sourcing with a small group of annotators that each annotated unique sets of images.

5.1.2 Defining and Identifying Private Objects in Visual Content

In this section, I discuss the YOLOv5 object detection model used to identify potential private objects in visual content. This object detection model has been pre-trained with

the MS COCO dataset (Lin et al. 2014). Since the model has been pre-trained with the MS COCO dataset, I define and discuss private labels of objects in visual content using labels from the MS COCO classes. The private class labels are chosen from MS COCO since currently there is no privacy or social media dataset that includes known private images and classes with bounding box annotations for training. Some objects in the images might not have a label detected since YOLOv5 was trained on 80 of the MS COCO labels. This framework is used to show the application and scope of using object detection models and bounding box annotation for creating visual features and incorporation into visual privacy risk scoring algorithms.

Identifying Objects in Visual Content

Each object in an image can influence the overall visual privacy risk score. The privacy image datasets above focus on a single classification for an image and do not include object detection models. This approach has drawbacks when trying to create an explainable and intuitive visual privacy risk scoring metric since quantitative visual features from the images may be ignored. To achieve this, I propose the use of the YOLOv5 object detection model (Redmon et al. 2016; Jocher et al. 2022) to identify privacy risks and enable the discovery of quantitative visual feature measures in images.

YOLO is a state-of-the-art object detection algorithm developed in 2015. The YOLO algorithm makes one pass across an image to predict object labels and to bound the objects in an image. With this process, objects in an image are separated into bounding boxes; each object that is bounded is associated with probabilities of its respective classifications. Since the initial release of YOLO, there have been several updates and improvements such as YOLOv5. The YOLOv5 model was evaluated and trained with the MS COCO dataset (Jocher et al. 2022). With the YOLOv5 model, I can detect objects in images across a wide range of objects, including labels such as people, traffic signs, and cars. In the experiments, I use the YOLOv5s pre-trained model (Jocher et al. 2022).

Labeling Private Object in Visual Datasets

The labeling of private classes in images can be difficult to do based on user subjectivity and relevance to the data collected. Due to limited visual datasets with private labels and bounding boxes, I generalize private content from the MS COCO labels. Since the YOLOv5 object detection algorithm is trained with MS COCO, I can directly apply the conceptual framework of extracting visual features and calculating visual privacy risk scoring across several datasets.

MS COCO focuses on object recognition tasks based on the context of an image's scene. These object recognition tasks can include object detection, segmentation, and captioning. For the scope of the chapter, I focus on bounding box applications to show the application of visual features. The YOLOv5 algorithm is pre-trained with photos of 80 class types from MS COCO. The MS COCO data contains a considerable amount of object instances per image, approximately 7.7 objects per image. From the object labels used to train YOLOv5, I create a simple hierarchy of privacy risk labels. For the scope of this chapter, the privacy risk classification can be sectioned into three levels: *No risk* (public), *moderate* privacy risk, and *severe* privacy risk. The schema of this hierarchy is influenced by the taxonomy in Section 3.2.2. The private label hierarchy created from the MS COCO dataset is outlined in Table 5.2.

The privacy risk classification scope is re-defined in this chapter as: (1) *Severe* privacy risk contains items that can contain or carry personally identifying information (backpack, handbag), personal devices or vehicles, or items that contain insight into a person's location and place of residence. (2) *Moderate* risk objects include public transportation, household items, or item. This content might not provide an individual's exact location, or place of residence, or contain any of their government-issued identification. (3) *No risk* content encompasses objects that do not include any of the above items. The labels were classified and placed into categories based on these definitions. Appendix B contains the list of private classes and a short description of privacy risks.

Table 5.2: Outline of private object class labels used for the visual features and visual privacy risk scoring methods. In this table, the object labels are separated based on severity categories. Only the moderate and severe privacy risk labels are shown.

Severe Privacy Risk Label	**Moderate Privacy Risk Label**
car	airplane
motorcycle	bus
traffic light	train
stop sign	truck
parking meter	boat
handbag	backpack
suitcase	wine glass
laptop	toilet
cell phone	tv
	mouse
	remote
	keyboard
	clock

The MS COCO labels are used to create a privacy classification to categorize objects in visual content. The object detection model is used to identify objects in the images and the output creates a one-hot encoding or data feature file. The object detection results for each visual content dataset and the defined privacy risk classification and labels are used in both of the privacy risk scoring methodologies.

5.2 Dichotomous Privacy Risk Score

The Dichotomous Privacy Risk Score (*DPScorer*) algorithm is a fundamental algorithm in the field of privacy scoring methodologies (Liu and Terzi 2010). The algorithm is used to compute the privacy score of a user, which indicates the potential risk caused by their shared information in the network. The algorithm focuses on the sensitivity and visibility of the information shared by a user, where sensitivity refers to the degree of sensitivity of the information and visibility refers to the extent to which the information spreads. This privacy risk scoring methodology converts unstructured data (textual information) in social

media profiles to binary values (shared, not shared).

In the context of online social networks, each user is associated with a profile consisting of n profile items. For each item, the user sets a privacy level that reflects their willingness to disclose the associated information. The privacy levels of all N users for all n profile items are stored in a response matrix R, which has dimensions $n \times N$. Specifically, the element $R(i, j)$ represents the privacy setting of user j for profile item i. If the entries in R take values in the set $0, 1$, then R is said to be dichotomous. In such a matrix, a value of 0 in $R(i, j)$ indicates that the user has chosen to keep the information associated with profile item i private, while a value of 1 indicates that the user has made the item publicly available. The calculation for the privacy score of User j due to Profile Item i is as follows:

$$\mathrm{PR}(i, j) = \beta_i \cdot V(i, j) \tag{5.1}$$

In equation 5.1, $\mathrm{PR}(i, j)$ represents the privacy score of a user j due to profile item i, while β_i represents the sensitivity of profile item i and $V(i, j)$ represents the visibility of profile item i. The dot symbol $(\cdot)$ denotes multiplication. The overall privacy score of User j can be calculated with a summation:

$$\mathrm{PR}(j) = \sum_{i=1}^{n} \mathrm{PR}(i, j) = \sum_{i=1}^{n} \beta_i \cdot V(i, j) \tag{5.2}$$

In equation 5.2, $\sum$ represents the summation symbol, n is the total number of items being considered, $\mathrm{PR}(i, j)$ represents the predicted privacy score of an item i for a user j, while β_i and $V(i, j)$ denote the sensitivity of item i and the visibility of item i that belongs to a user j respectively. The dot symbol $(\cdot)$ denotes multiplication. The equation computes the sum of privacy scores for all items i to User j. A disadvantage of this approach is that the sensitivity values obtained are significantly biased by the user population contained in the response matrix, R.

5.2.1 Adaptation of Dichotomous Privacy Scoring for Visual Privacy Risk Scoring

The *DPScore* algorithm has been modified to calculate the privacy score for Image I based on Object O. This adaptation utilizes the YOLOv5 object detection algorithm to identify object o in an image i. The privacy hierarchy is used to define public and private objects in images across the datasets and set the sensitivity values for each object independently. By defining the sensitivity based on privacy hierarchy, the values will not have a bias based on the response matrix, R.

In the context of the visual content datasets, each image is associated with objects consisting of n object labels. For each object label, an image sets a binary object feature flag that reflects if that object exists within the image. The object feature flag of all N images for all n object labels are stored in a response matrix R, which has dimensions $n \times N$ Specifically, the element $R(o, i)$ represents the feature flag of image i for object label o. In such a matrix, a value of 0 in $R(o, i)$ indicates that the image does not contain object o, while a value of 1 indicates that the image i does contains object o. Similar to the Equation (5.1), the calculation for the privacy score of image i due to object label o is as follows:

$$\mathrm{PR}(o, i) = \beta_o \cdot V(o, i) \tag{5.3}$$

In equation 5.3, $\mathrm{PR}(o, i)$ represents the privacy score of an image i due to an object label o, while β_o represents the sensitivity of object label o from the privacy hierarchy and $V(o, i)$ represents the visibility of object label o. The dot symbol $(\cdot)$ denotes multiplication. Following a similar framework as Equation (5.4), the privacy score of Image i can be calculated with this summation:

$$\mathrm{PR}(i) = \sum_{i=1}^{n} \mathrm{PR}(o, i) = \sum_{i=1}^{n} \beta_o \cdot V(o, i) \tag{5.4}$$

In equation 5.4, $\sum$ represents the summation symbol, n is the total number of object labels being considered, $\text{PR}(o, i)$ represents the predicted privacy score of an object o for an image i, while β_o and $V(o, i)$ denote the sensitivity of object label o and the visibility of a object label o that belongs to image i respectively. The dot symbol $(\cdot)$ denotes multiplication. The equation computes the sum of privacy scores for all Object O to Image I. For the remainder of this chapter, I will refer to this scoring method as the *VPScorer*.

5.3 Visual Privacy Risk Scoring Methodology using Visual Features

The analysis of visual data has become important due to its diverse range of applications (Prevedello et al. 2019; Taverner et al. 2020; Zatelli et al. 2019). Visual privacy can be interpreted from visual data analysis focusing on quantifying visual feature attributes that are present in an image. The proposed visual privacy risk scoring methodology in this chapter focuses on the visual features using a combination of techniques, including Object Importance Weight (OIW), Object Area Ratio (OAR), and Golden Spiral Distance (GSD). These techniques are used to introduce the *Vango* framework, which calculates an image's visual privacy risk score based on three components. These visual feature attributes are further described below.

5.3.1 Object Importance Weight

The ability to quantify the frequency of objects that occur in images from a dataset can be an important visual feature to analyze. A method commonly used for unstructured data (e.g., text) is feature vectorization, which involves the extraction of features from text and representing them as a vector. However, a common issue with feature vectorization is that frequently occurring objects tend to dominate these vectors, resulting in a biased data representation. TF-IDF vectorization has been proposed to address this issue as a

solution (Banweer et al. 2018; Pedregosa et al. 2011a).

The inverse document frequency (IDF) weighting reduces the weight of commonly occurring objects by assigning a higher weight to rare terms. The TF-IDF methodology is further described in Appendix B. The concept of OIW can be applied to weighting objects in images across a dataset. I assume private objects are less common in images and the dataset in this approach. Applying inverse document frequency will reduce the weight of public objects and increase the weight of rare objects (i.e., private objects).

In order to process the objects obtained from images, I convert the objects from that image into a sentence string. Once each image in the dataset has a sentence string, these sentences are passed into a TF-IDF pipeline to calculate the OIW across the entire dataset. In this algorithm, the object's frequency (i.e., term frequency) is quantified by the sensitivity of the object label from the privacy risk classification. Further, the objects are analyzed as unique sets with an *n-grams* approach. An *n-gram* is a sequence of words that occur in a given window. The *n-gram* object frequency is quantified by the sum of the sensitivity of each object label from the privacy risk classification. It is assumed that *n-grams* can be used to quantify the privacy risk of an image due to multiple private objects appearing. The study of *ngrams* can show trends, frequency, and weighting across images and datasets.

The formula for calculating OIW is given by Equation 5.5, where each object o is extracted from image i in a visual dataset I. Specifically, the element $\mathrm{OIW}(o, I)$ is the product of the privacy risk of the object ρ_o, the frequency of the object appearing in the image $tf(o, i)$, and the inverse of the frequency that the object appears across the image dataset $idf(o, I)$.

$$\mathrm{OIW}(o, i) = \rho_o \cdot \mathrm{tf}(o, i) \cdot \mathrm{idf}(o, i). \tag{5.5}$$

This approach can potentially improve the accuracy of object quantification in image datasets. By finding the importance of the objects in visual data, this visual feature can achieve an accurate and comprehensive weighing of objects to gauge privacy risk. I discuss the results of OIW as a component of the *Vango* privacy risk score in Experiment 5.4.1.

5.3.2 Object Area Ratio

Object Area Ratio (OAR) calculates the size of objects in relation to the entire image. Calculating the size of objects can provide valuable information about the overall composition of the image and help to identify key features and attributes that may be important for visibility. Assuming that the size of private objects influences visibility, it is important to consider the relative size of objects in the image when calculating visual privacy risk scores. For example, smaller objects may be less visible and have a lower impact on the score than larger objects due to their size relative to the overall image. Similarly, larger objects may be more visible and have a high impact on the score due to their dominant size and visual prominence.

To calculate the size of objects in relation to the entire image, various techniques can be used, such as object recognition and segmentation. Once the objects have been identified, their size can be calculated using an object's area or perimeter. The formula for calculating OAR is given by Equation 5.6, where each object o is extracted from an image i. Specifically, the element $OAR(o, i)$ represents the ratio between the computed area of an object, $A(o)$, and is divided by the total image size, $Area(i)$.

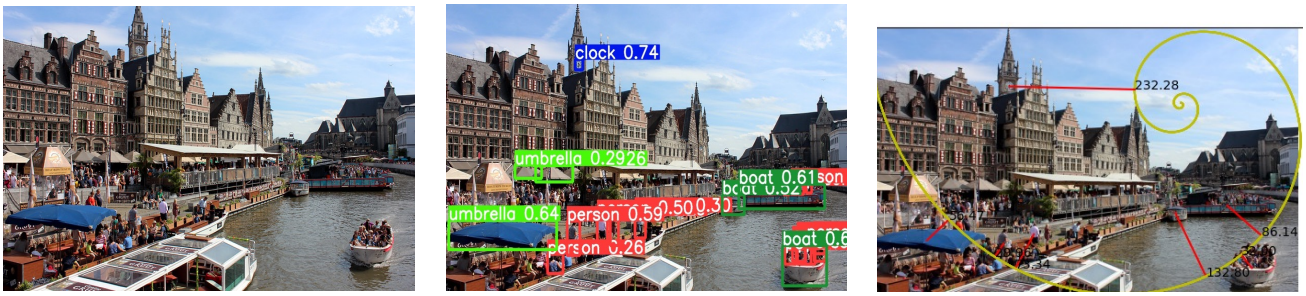
$$\text{OAR}(o, i) = \frac{\text{Area}(o)}{\text{Area}(i)} \tag{5.6}$$

This technique is particularly useful when dealing with images that contain complex objects or scenes with multiple objects. By using techniques such as object recognition, this method can calculate the size of objects and use this information to gain a deeper understanding of object visibility in visual privacy scoring. I discuss the results of OAR as a component of the *Vango* privacy risk score in Experiment 5.4.1.

5.3.3 Golden Spiral Distance

The golden spiral, a logarithmic spiral derived from the Fibonacci sequence, is a visual feature that can be applied to visual privacy risk. The golden spiral is commonly found in nature and art, and it has been studied for its visual appeal and mathematical applications (Neperud and Freedman 1988; Green 1995; Saraswathi 2007). During visual data analysis, the golden spiral can be used to highlight important objects within an image (Katukuri 2019). Objects towards the center of the spiral can be considered more visually appealing (Neperud and Freedman 1988); in this work, I extend this notion to include all objects along the curve of the golden spiral.

When analyzing images for potential privacy risks, the golden spiral can be a valuable tool for identifying and mitigating private objects close to the spiral. Figure 5.4 shows a photo of a lively and vibrant day in a town. In this example, the people and boats are immediately visible due to their proximity to the curve. In Figure 5.4c, the clock is the



(a) Original Image (b) YOLOv5 Detection (c) GSD Annotation

Figure 5.4: This figure shows an image from the PrivacyAlert Dataset and how the image's visual features for GSD are transformed through the pipeline: (a) The original image from the PrivacyAlert dataset, (b) The image is annotated with bounding boxes for the objects detected by YOLOv5, (c) The image is annotated, showing the Golden Spiral and the calculated distance (in pixels) for each object from the curve.

farthest distance away from the curve. Since the GSD of the clock is high, the object's impact on the *Vango* privacy score will be low. The algorithm for calculating the GSD is shown in Listing 1. The function `golden_spiral` takes an image dataset object. It iterates over each image and computes the golden spiral using the function `compute_spiral`. After

each spiral is computed, the distance between each object in the image and the golden spiral is calculated. This function returns the image, the closest coordinate pair in the golden spiral, and the GSD from the object's center to the closest coordinate.

The formula for calculating GSD is given by Equation 5.7, where each image i is associated with an object o. Specifically, the element $GSD(o, i)$ represents the distance between the computed golden spiral of an image, $G(i)$, to the center coordinate of an object, $center(o)$.

$$\text{GSD}(o, i) = ||\text{G}(i) - \text{center}(o)|| \tag{5.7}$$

The GSD could be an important visual feature when analyzing images for privacy risk. By understanding and utilizing visual features such as the golden spiral, researchers can better quantify privacy and understand the visual features of images. I discuss the results of GSD as a component of the *Vango* privacy risk score in Experiment 5.4.1.

5.3.4 Combining Visual Features into the *Vango* privacy risk score

The visual privacy risk scoring framework does a weighted combination of the OIW (Equation 5.5), OAR (Equation 5.6), and GSD (Equation 5.7) visual feature measurements. The formula for calculating the *Vango* privacy risk score is given by Equation 5.8, where each object o is extracted from image i in a visual dataset I. Specifically, the element Vango(I) is the weighted product of the object importance in the image dataset I, OIW(I), the average visibility of the object i in an image i, GSD(o, i), and the average prominence an object i in an image i, OAR(o, i). Inside of the weights (γ_{1-3}), there is a term to average the component across all the objects. Each score component is also normalized using max scaling to ensure the range is between 0 and 1. The total *Vango* score of images i in a visual dataset I can be calculated as:

$$\text{Vango}(I) = \gamma_1 \cdot \sum_{i \in I, o \in \text{Yolo}(i)} \text{OIW}(i) + \gamma_2 \cdot \sum_{i \in I, o \in \text{Yolo}(i)} \text{OAR}(o, i) + \gamma_3 \cdot \sum_{i \in I, o \in \text{Yolo}(i)} \text{GSD}(o, i). \tag{5.8}$$

Listing 1 Calculating Golden Spiral distance for objects in an image

```
import numpy as np

def compute_spiral(topleft, topright, bottomright, pic, resolution=1000):

    line1 = find_slope_intercept(topleft, bottomright)
    line2 = find_slope_intercept((bottomright[0] / 1.6, bottomright[1]), topright)
    x0, y0 = find_intersection(line1, line2)
    phi = (1 + 5**0.5) / 2
    theta0 = np.arctan2(-y0, -x0)
    k = 2 * np.log(phi) / np.pi
    a = -x0 / (np.exp(k * theta0) * np.cos(theta0))
    t = np.linspace(-20, theta0, resolution)

    def x(t):
        return x0 + a * np.exp(k * t) * np.cos(t)
    def y(t):
        return y0 + a * np.exp(k * t) * np.sin(t)

    return list(zip(x(t), y(t)))

def closest_point_and_distance(point_list, target_point, box_width, box_height):

    center_x,  center_y  = box_width / 2, box_height / 2
    box_center_x, box_center_y  = target_point[0] + center_x, target_point[1] + center_y
    min_distance = math.inf
    closest_point, closestp = None, None

    for p, point in enumerate(point_list):
        distance = math.sqrt(
            (point[0] - box_center_x) ** 2 + (point[1] - box_center_y) ** 2
        )
        if distance < min_distance:
            min_distance = distance, closest_point = point, closestp = p

    return closest_point, min_distance

def golden_spiral(imagedataset):

    for m, img in imagedataset:
        topleft = (0, img["width"])
        topright = (img["height"], img["width"])
        bottomright = (img["height"], 0)
        point_list = compute_spiral(topleft, topright, bottomright)

        for i, item in enumerate(img["object_data"]):
            start_coordinates = (item["x"], item["y"])
            closest, dist, point = closest_point_and_distance(
                point_list, start_coordinates, item["width"], item["length"])

            yield (img, closest, dist)
```

5.4 Experiments and Results

This section explores the visual features, visual dataset features, and visual privacy risk scoring algorithms. The experiments employ the object detection algorithm, YOLOv5, as the backbone and are coupled with pre-trained MS COCO Labels and privacy risk labels for objects to investigate the efficacy of OIW, GSD, and OAR calculations. The study examines the object frequency and visual feature calculations of objects across the datasets. Furthermore, the section compares two visual privacy risk scoring methods by running them across three datasets. Experiment 2 offers insight into the effectiveness of visual privacy risk methodologies for visual datasets.

5.4.1 Experiment 1: Exploring of the Efficacy of OIW, GSD, and OAR as Visual Features for Visual Privacy Risk Scoring

This study investigates the effectiveness of the OIW, GSD, and OAR visual features for analyzing objects within images at scale. To this end, the state-of-the-art object detection algorithm, YOLOv5, is leveraged, along with a privacy risk classification for objects in MS COCO. The evaluation of the proposed methods is conducted across three distinct visual datasets: PrivacyAlert, VISPR, and Open Images v7.

The central task of this experiment is to identify trends in the frequency of objects and trends regarding the visual features across the visual datasets. This task involves object detection for visual data, as well as the calculation of an image's visual features and their respective measurements across the datasets. By systematically analyzing the performance of these visual features across the datasets, this study aims to provide a comprehensive understanding of their efficacy for visual privacy risk.

Visual Feature: The efficacy of Object Importance Weights

With Object Importance Weight (OIW), I analyze visual data by quantifying the object's weight that occurs in an image for a given dataset. The results below show the frequency occurrence of *n-gram* objects in the datasets and the object weights for each dataset using the OIW methodology. In this experiment, I consider *n-grams* to understand how the occurrence of objects in images correlates with trends, frequency, and object importance weighting. The *n-gram* objects contain one public or one private; *n-gram* objects containing two objects can contain public-private, private-public, or private-private bigram objects. The order of this varies based on the sequence of detection. Instances of the `person` class have been excluded due to the exponentially high frequency across the datasets.

Open Images v7 Dataset: Frequency occurrence of *n-gram* objects. This analysis examines the Open Images v7 dataset and presents the results of the object frequency distribution. The findings are presented in Figure 5.5, which reports the top 30 unigram and bigram objects in the dataset, irrespective of their privacy risk level. The unigram graph illustrates that the objects with the highest frequency counts are `chair` (34,210), `car` (27,856), and `tie` (19,778), while `backpack` (3,349), `traffic` (2,904), `light` (2,904), and `motorcycle` (2,748) are least frequent. Notably, the remaining unigram objects occur between 105 and 2,025 times across the dataset, with the `snowboard` object having the lowest frequency count.

The bigram graph shows that the most frequently occurring bigram objects are `chair person` (13,090), `car car` (12,734), and `person chair` (10.971), while `umbrella person` (1,929), `person laptop` (1,889), and `book book` (1,731) have a lower frequency. The bigram objects after these have between 1 to 1,722 occurrences across the dataset, with `airplane bird` having the lowest frequency count. Based on the results in Figure 5.5, it is evident that certain objects are more frequently depicted in images than others. The occurrence of `person`, `chair`, and `car` are the highest among all of the objects in the dataset, including

the unigram and bigram objects. However, it is important to note that this analysis only considered object frequency and did not consider the privacy implications of these objects.

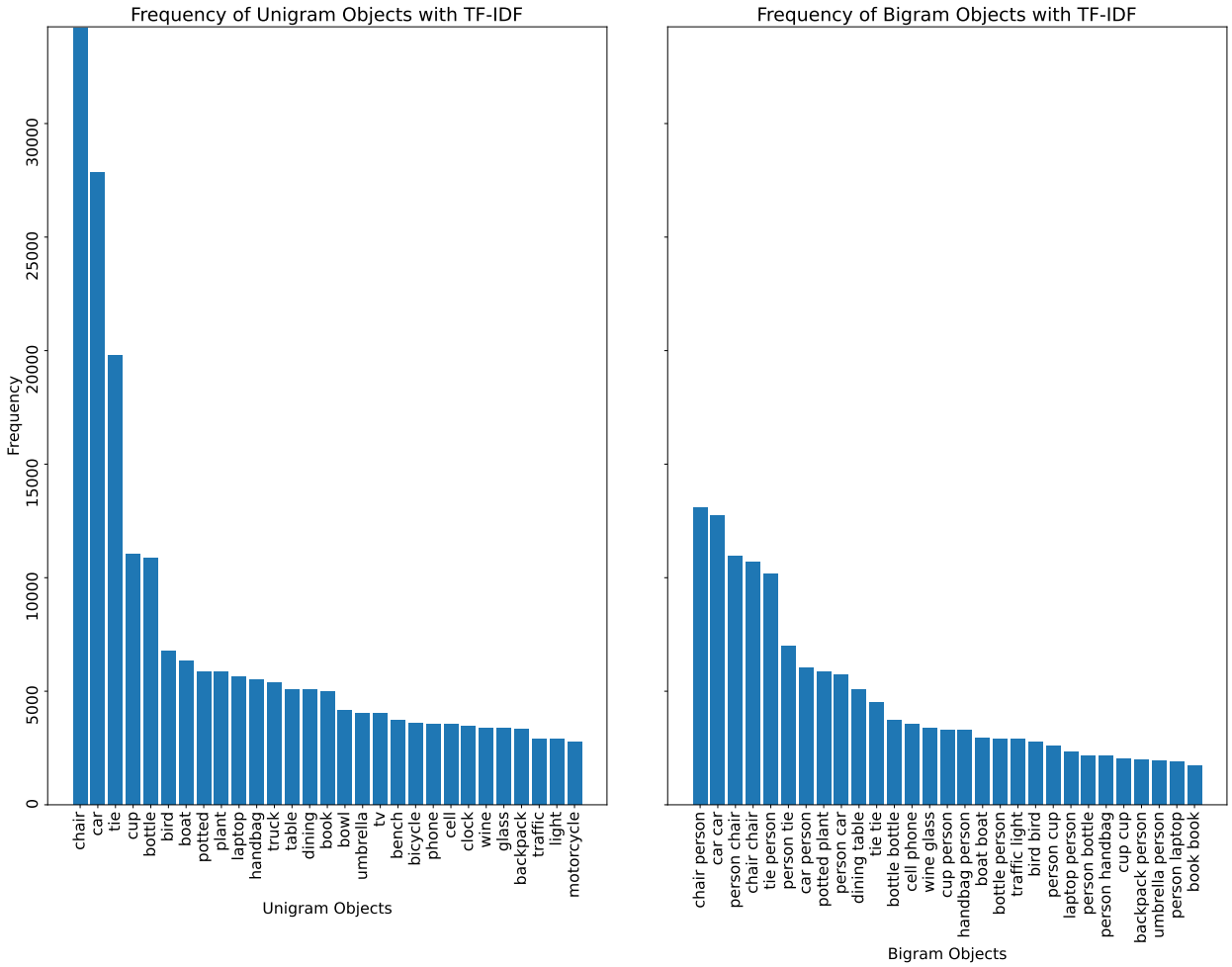


Figure 5.5: This figure displays the frequency (y-axis) of *n-gram* objects (x-axis) across the Open Images v7 dataset. The figure excludes the `person` label due to its extremely high occurrence across the dataset.

Open Images v7 Dataset: Object importance weights for *n-grams*. This analysis examines the Open Images v7 dataset and presents the results of the object weights. The findings are presented in Figure 5.6, which reports the 27 unigram and the top 30 bigram objects in the dataset. The unigram graph illustrates that the private objects with the highest importance weights counts are `train`, `airplane`, and `toilet`, while `traffic and light`, `wine`, and `glass` have a lower OIW. Notably, of the unigram private objects, the

lowest weight is `glass` with a score of 0.234.

The bigram objects `laptop truck`, `car oven`, and `oven car` have the heaviest weights for public-private and private-private bigram objects. Of the 30 private object bigrams, 6 of those contain two private object classes. Those bigrams are `laptop truck`, `bus laptop`, `truck toilet`, `suitcase remote`, `tv boat`, and `airplane toilet`. These object bigrams indicate that these objects are more important than others across the entire dataset. The private object bigram objects show uniquely weighted occurrences compared to unigram objects.

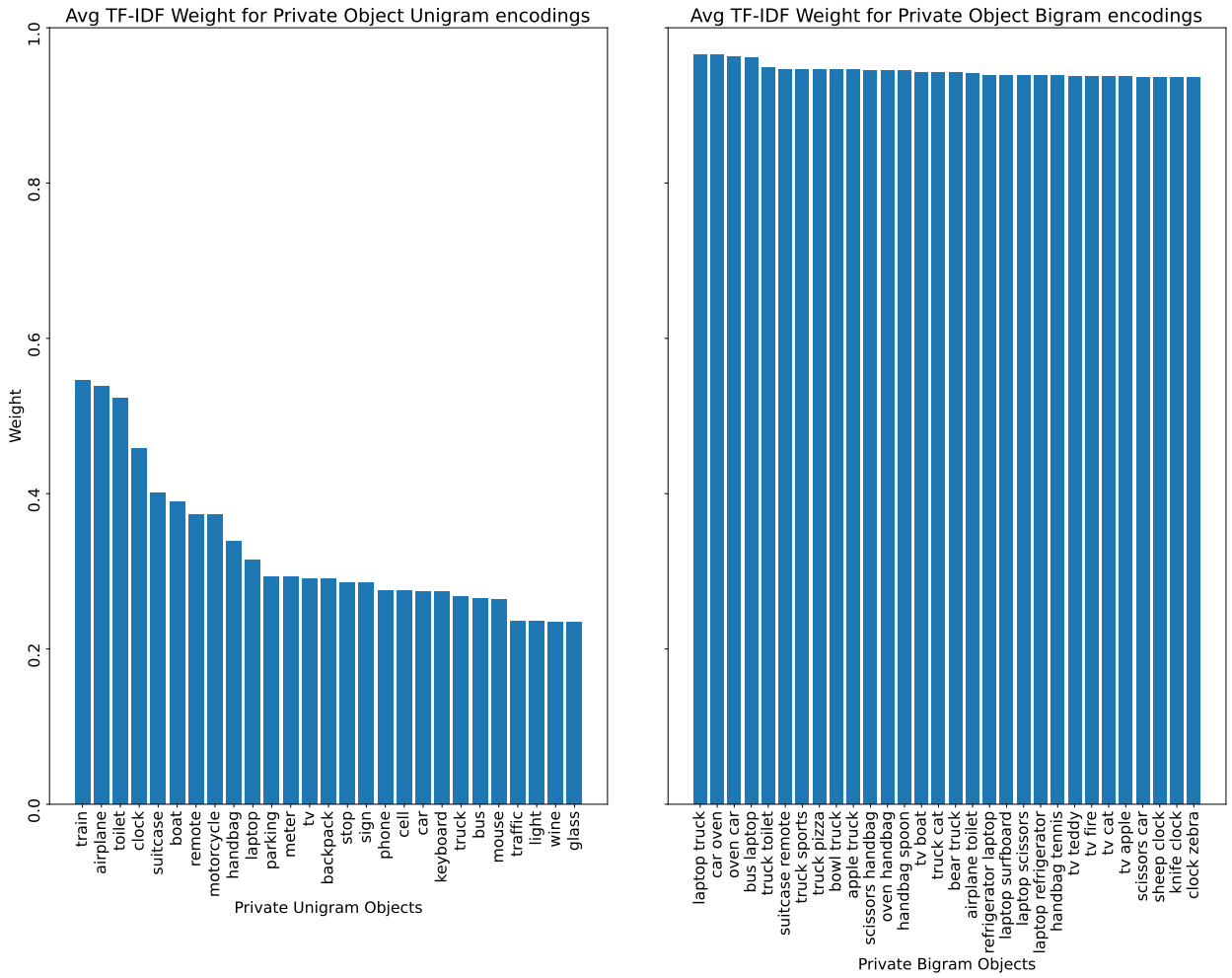


Figure 5.6: This figure displays the average object importance weight (y-axis) of *n-gram* object (x-axis) across the Open Images v7 dataset. The first graph displays all of the private object unigrams and their OIW. The second graph displays the bigrams that contain at least one private object label.

PrivacyAlert Dataset: Frequency occurrence of *n-gram* objects. This analysis examines the PrivacyAlert dataset and presents the results of the object frequency distribution. The findings are presented in Figure 5.7, which reports the top 30 unigram and bigram objects in the dataset, irrespective of their privacy risk. The unigram graph illustrates that the objects with the highest frequency counts are `car` (1389), `chair` (994), and `bottle` (483), while `bicycle` (136), `backpack` (134), and `tv` (115) are the least frequent. Notably, the remaining unigram objects occur between 5 and 114 times across the dataset, with the `snowboard` object having the lowest frequency count.

The bigram graph shows that the most frequently occurring bigram objects are `car car` (588), `chair person` (345), and `car person` (298), while `backpack person` (82), `person bottle` (77), and `bicycle person` (75) are the least frequent. The bigram objects after these have between 1 to 73 occurrences across the dataset, with `airplane bird` having the lowest frequency count. Similar to Open Images v7 frequency distribution Figure 5.5, the occurrence of `person`, `chair`, and `car` are the highest among all of the objects in the dataset, including unigram and bigram objects.

PrivacyAlert Dataset: Object importance weights for *n-grams*. This analysis examines the PrivacyAlert dataset and presents the results of an analysis of its object weights. The findings are presented in Figure 5.8, which reports the 27 unigram and the top 30 bigram objects in the dataset. The unigram graph illustrates that the private objects with the highest weights counts are `train`, `airplane`, and `toilet`, while `traffic`, `light`, and `bus` have a lower OIW weight. Notably, the lowest weight of the unigram private objects is `bus` with a score of 0.236.

The bigram objects `car vase`, `truck airplane`, and `toilet toilet` have the heaviest weights for public-private and private-private bigrams. Of the 30 private object bigrams, 10 of those contain two private object classes. Those bigrams are `truck airplane`, `toilet toilet`, `airplane boat`, `airplane airplane`, `suitcase suitcase`, `suitcase keyboard`,

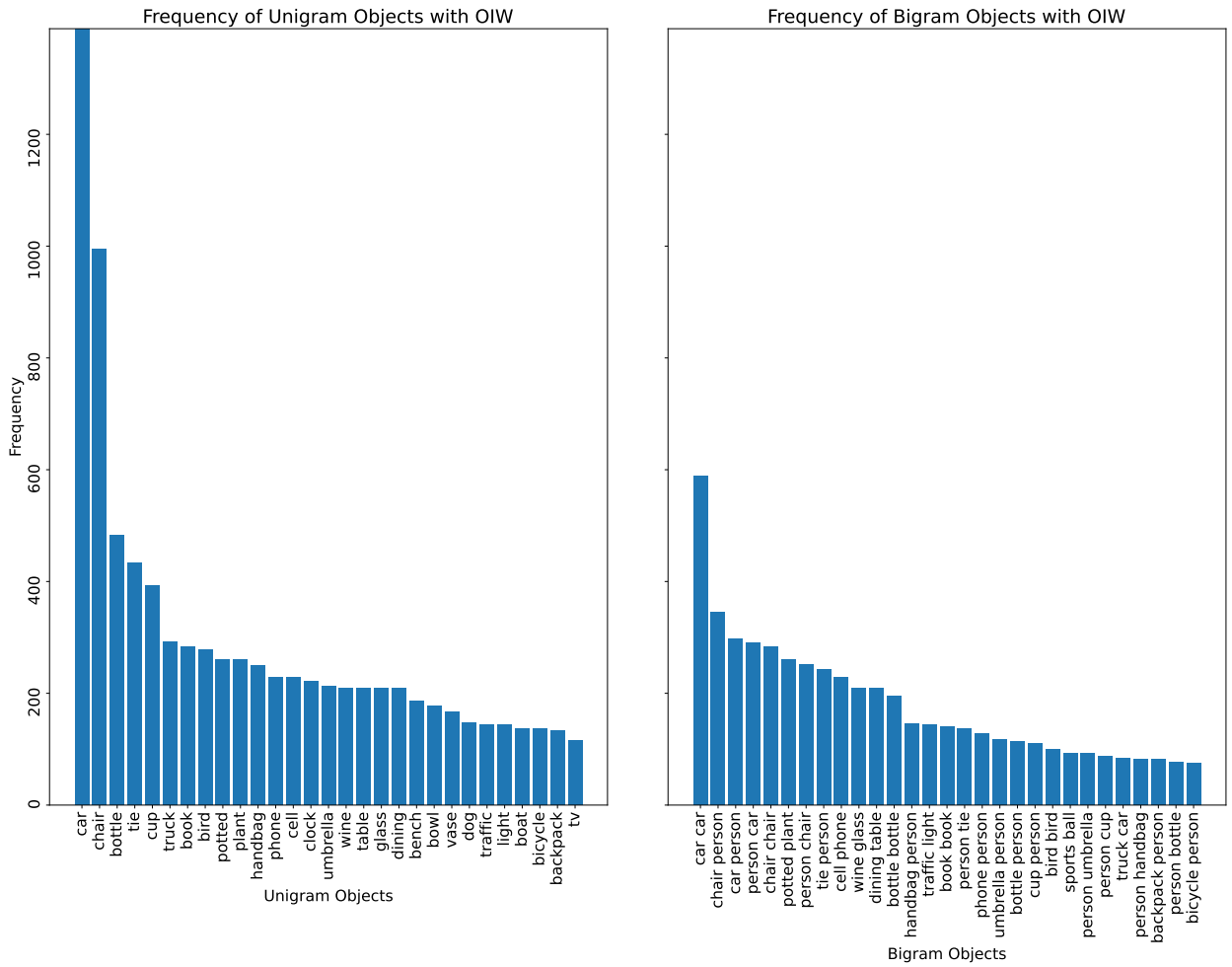


Figure 5.7: This figure displays the frequency (y-axis) of *n-gram* objects (x-axis) across the PrivacyAlert dataset. The figure excludes the `person` label due to its extremely high occurrence across the dataset.

`toilet mouse`, `boat motorcycle`, `clock book`, and `train handbag`. These bigram objects indicate that these *n-grams* objects are more important than others across the entire dataset. The bigrams show that the combination of private objects can lead to uniquely weighted occurrences in comparison to their unigram counterparts.

VISPR Dataset: Frequency occurrence of *n-gram* objects. This analysis examines the VISPR Dataset and presents the results of an analysis of its object frequency distribution. The findings are presented in Figure 5.9, which reports the top 30 unigram and bigram object possibilities in the dataset, irrespective of their privacy risk. The unigram graph illustrates

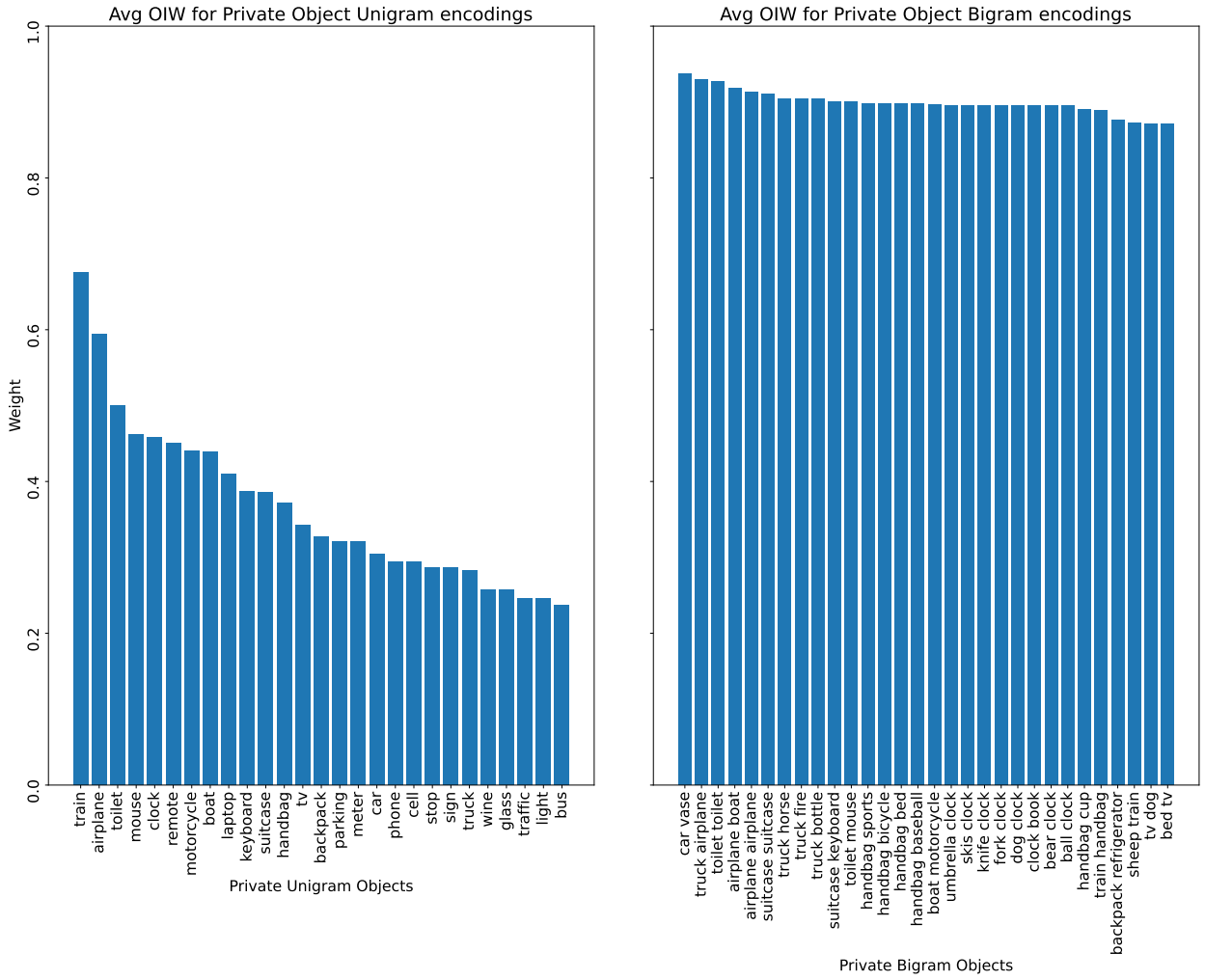


Figure 5.8: This figure displays the object importance weights (y-axis) of *n-gram* objects (x-axis) across the PrivacyAlert dataset. The first graph displays the private object unigrams and their OIW weights. The second graph displays bigrams that contain at least one private object label.

that the objects with the highest frequency counts are `car` (2,730), `chair` (2,471), and `handbag` (1,116), while `wine` (302), `glass` (302), and `sports` (275) are the least frequent. Notably, the remaining unigram objects occur between 1 and 275 times across the dataset, with `toaster` having the lowest frequency count.

The bigram graph shows that the most frequently occurring bigram objects are `car car` (1,190), `chair person` (997), and `person chair` (816), while `phone person` (235), `person cup` (231), and `bottle bottle` (227) are the least frequent. The bigram objects after these

have between 1 to 201 occurrences across the dataset, with `airplane dog` having the lowest frequency count. Similar to Open Images v7 (Figure 5.5) and PrivacyAlert (Figure 5.7) frequency distributions, the occurrence of `person`, `chair`, and `car` are the highest among all of the objects in the dataset including unigram and bigram objects.

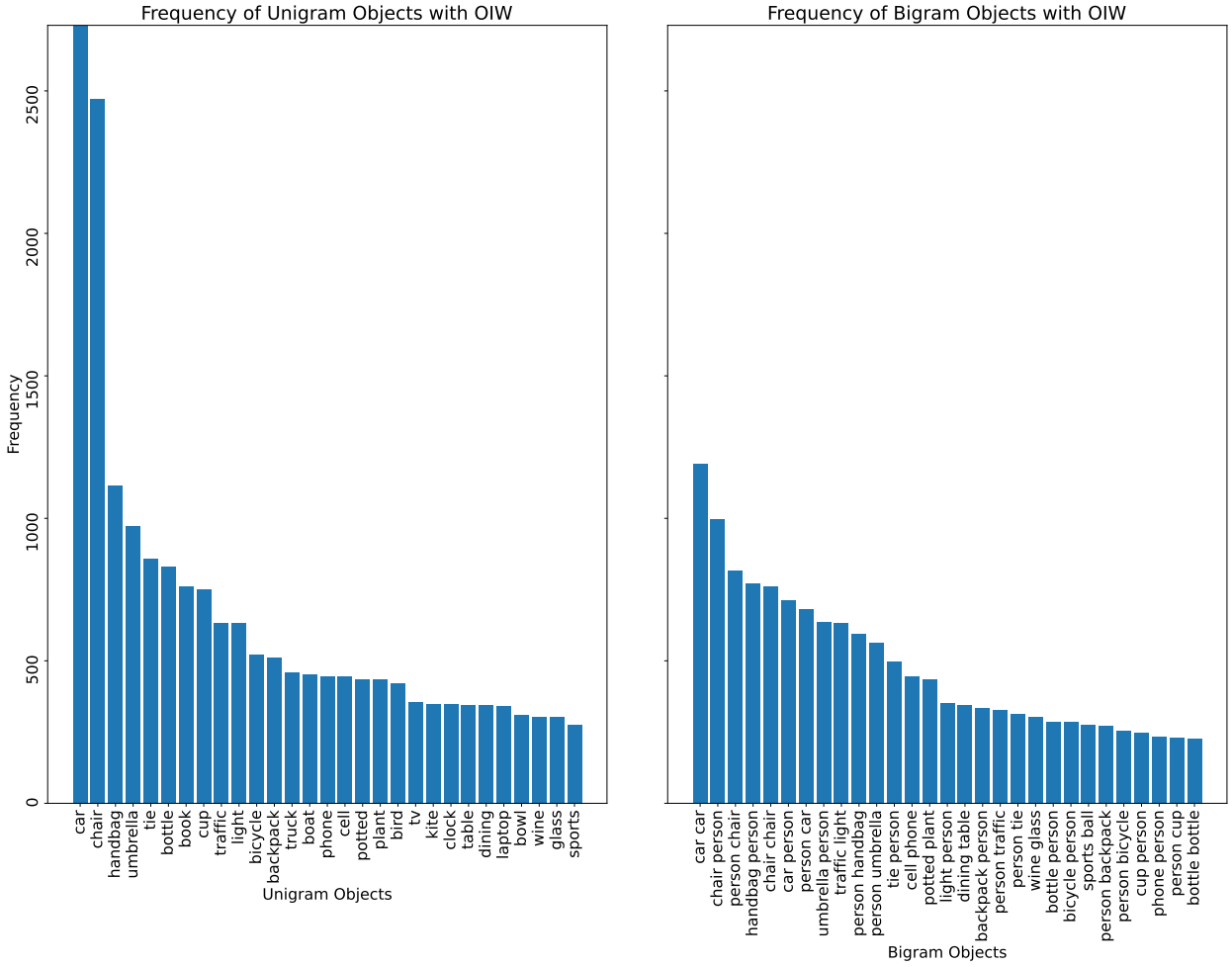


Figure 5.9: This figure displays the frequency (y-axis) of *n-gram* objects (x-axis) across the VISPR Dataset. The figure excludes the `person` label due to its extremely high occurrence across the dataset.

VISPR Dataset: Object importance weights for *n-grams*. This analysis examines the VISPR Dataset and presents the results of an analysis of its object weights. The findings are presented in Figure 5.10, which reports the 27 unigram and the top 30 bigram object possibilities in the dataset. The unigram graph illustrates that the private objects with the

highest weights counts are `toilet`, `airplane`, and `train`, while `light`, `wine`, and `glass` have a lower OIW. Notably, of the unigram private objects, the lowest weight is `glass` with a score of 0.231.

The bigram objects `horse car`, `couch car`, and `cat car` have the heaviest weights for public-private and private-private bigram combinations. Of the 30 private object bigrams, 9 of those contain two private object classes. Those bigrams are `clock boat`, `tv backpack`, `clock motorcycle`, `clock handbag`, `suitcase tv`, `laptop remote`, `backpack tv`, `handbag boat` and `toilet toilet`. These object weights indicate that these bigrams are more important than others across the entire dataset. The bigrams show that private objects can lead to uniquely weighted occurrences in comparison to unigrams.

Discussion. The analyses of the Open Images v7, PrivacyAlert, and VISPR datasets have provided valuable insights into the frequency and weight distribution of n-gram objects in these datasets. The results reveal that certain objects, such as `chair`, `car`, and `person`, are more frequently depicted in images than others, while private objects such as *traffic*, *light*, and *wine* occur less frequently in the datasets and do not have a significant weight. The occurrence of these objects and their combinations have implications for privacy risk and preservation. Moreover, the use of term frequency and inverse document frequency (OIW) can help to identify the importance of objects in these datasets. In the context of image analysis, OIW can be used to identify the most important objects and their combinations based on their frequency and weight. By using OIW, it is possible to extract significant information from visual content datasets and help identify and prioritize important objects for various image-related tasks. The findings suggest that visual content datasets can use OIW vectorization to understand object importance in images, while also highlighting the importance of visual privacy when analyzing the frequency and unique objects among the datasets.

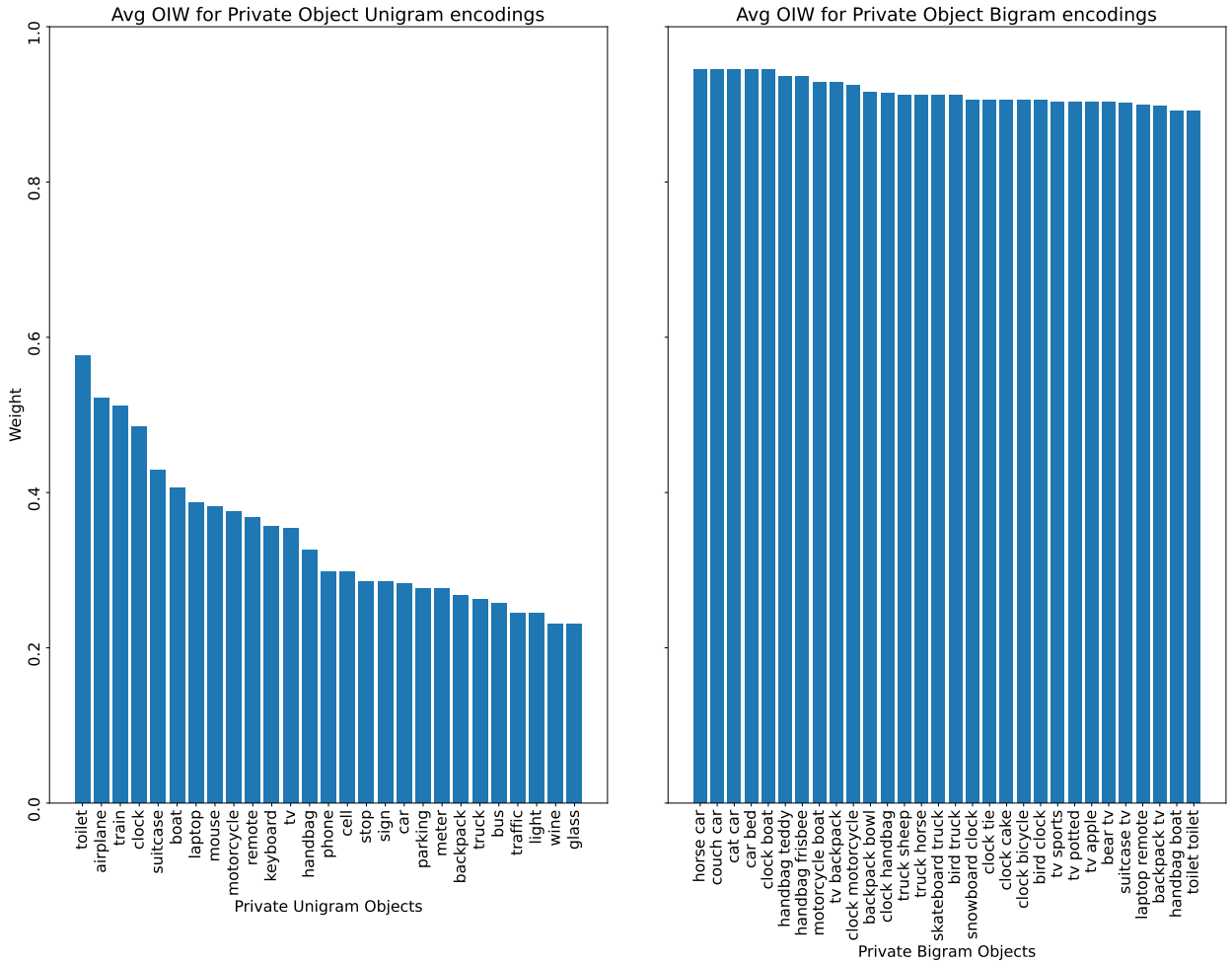
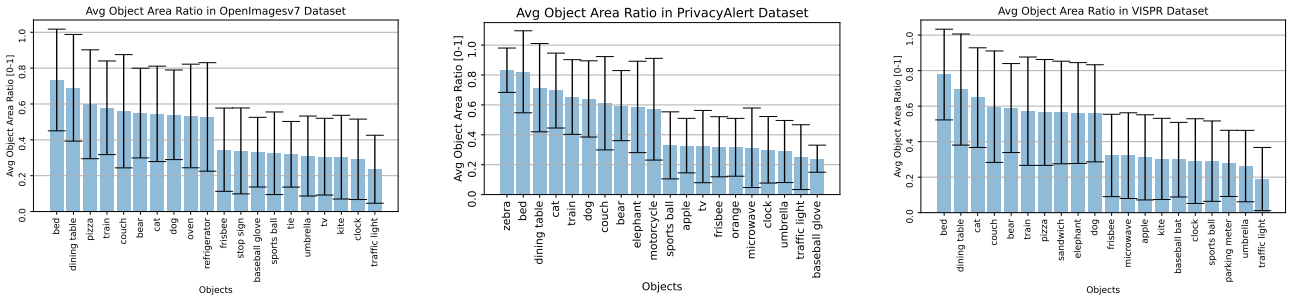


Figure 5.10: This figure displays the object importance weights (y-axis) of *n-gram* objects (x-axis) across the VISPR Dataset. The first graph displays the private object unigrams and their OIW. The second graph displays bigrams that contain at least one private object label.

Visual Feature: Object Area Ratio

The Object Area Ratio (OAR) measurement is a valuable technique for analyzing images that contain complex objects or scenes with multiple objects. It quantifies the visibility of objects across the image by calculating the size of the objects detected, which can provide crucial information about the image's overall composition and identify key features and attributes. When analyzing the visibility of objects, it is important to consider the relative size of private objects, assuming that their size influences privacy risk. The results of the OAR methodology show the object ratios for both public and private categories, providing



(a) Open Images v7 (b) PrivacyAlert (c) VISPR

Figure 5.11: This figure displays the average OAR measurement of the top 10 and bottom 10 objects across the Open Images v7 Dataset. The average standard deviation of the object's ratio is shown in the error line. (a) Open Images v7, (b) PrivacyAlert, (c) VISPR

insights into the ratio of objects in the visual content datasets.

Open Images v7 Dataset: Object Area Ratio of objects. The analysis of the Open Images v7 dataset yielded interesting findings regarding the relative size of objects in images. As depicted in Figure 5.11a, certain objects tend to occupy a larger area in images compared to others. For instance, the ratio of `bed` (0.73), `dining table`(0.68), and `pizza` (0.59) objects take up a larger portion of the entire image among all of the objects in the dataset. Interestingly, only one private object, `train`, appears in the top 10 OAR measurements. Furthermore, the OAR analysis highlights that private objects such as `stop sign`(0.33), `tv` (0.30), `clock` (0.29), and `traffic light` (0.23) have smaller average OARs compared to other objects in the dataset.

Open Images v7 Dataset: Object Area Ratio of the Private Objects. Looking at the OAR of the private object classes from the Open Images v7 dataset (shown in Figure 5.12), the results show that vehicle objects are frequently larger in images than others. For instance, the occurrence of `train` (0.57), `bus` (0.50), `airplane` (0.49) and *motorcycle* (0.49) are the highest among all of the objects in the dataset. All of the vehicle objects are considered private for the scope of this chapter. The *Moderate* privacy risk graph further

shows that images containing `toilet` are taken in close proximity to the object leading to a larger OAR measurement of 0.52 on average. On the other hand, the *Severe* privacy risk graph reveals that traffic signals such as `stop sign` (0.33) and `traffic light` (0.23) have smaller OAR measurements.

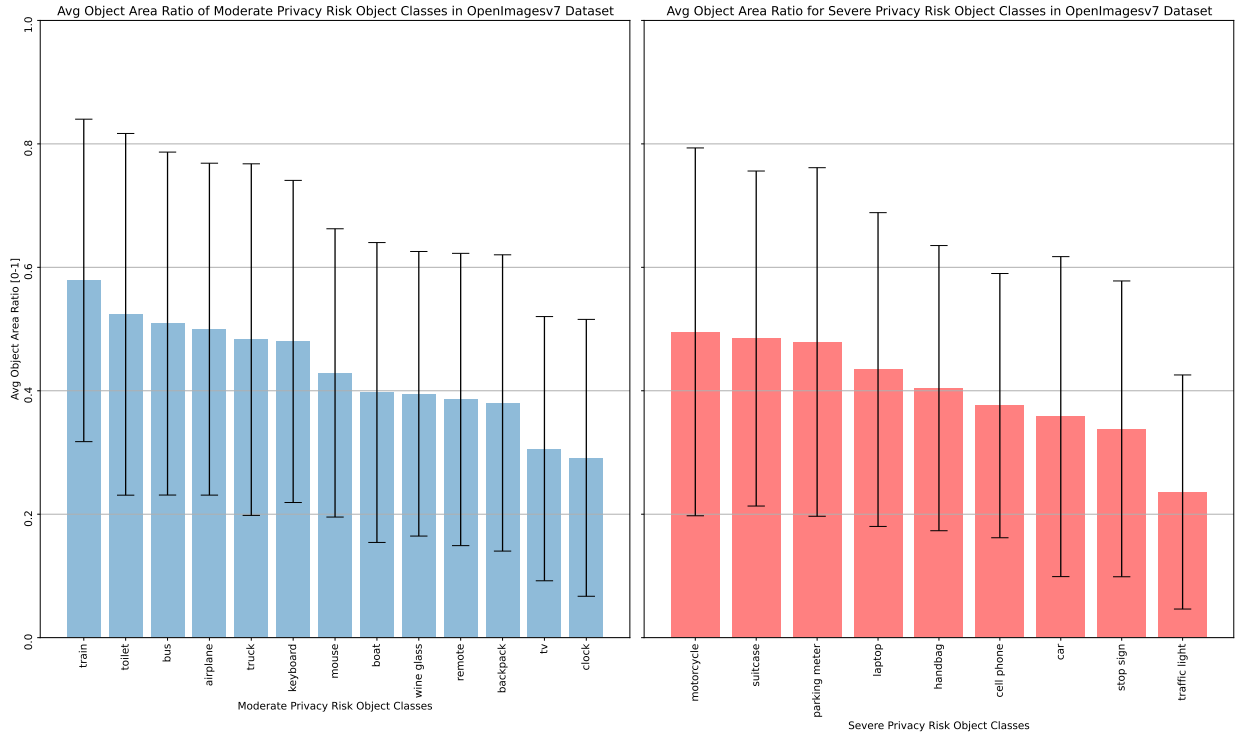


Figure 5.12: This figure displays the average OAR measurement of the moderate and severe privacy risk object labels across the Open Images v7 Dataset. The standard deviation of the object's measurement is shown in the error line.

PrivacyAlert Dataset: Object Area Ratio of objects. The results of the PrivacyAlert dataset analysis show that animals, on average, are larger than other objects, shown in Figure 5.11b. For instance, the average ratio of `zebra` (0.83), `cat`(0.69), `bear` (0.59), and `elephant` (0.58) cover more than 50% of the image. Of the top 10 OAR measurements, two private objects, `train` and `motorcycle`, appears to have a large OAR on average. This object has a consistently larger OAR measurement as seen in the Open Images v7 analysis. Moreover, the OAR analysis of the PrivacyAlert dataset highlights that private objects, such as `tv` (0.32), `clock` (0.29), and `traffic light` (0.24) have small averages across the objects,

which is consistent with the Open Images v7 OAR scores.

PrivacyAlert Dataset: Object Area Ratio of the Private Objects. Through an analysis of the PrivacyAlert dataset, the private object classes were examined for their Object Area Ratio (OAR) in Figure 5.13. The results indicate that vehicle objects, such as `train` (0.65), `motorcycle` (0.57), `airplane` (0.50), and `truck` (0.50) are often larger in images than other objects in the dataset. Additionally, the study observed that images containing computer accessories (`mouse`, `laptop`, `keyboard`) are photographed in closer proximity to the object, leading to a higher OAR measurement of approximately 0.45 to 0.51 on average. The *Severe* privacy risk graph reveals that objects, such as `cellphones` (0.33) and `traffic light` (0.23), have smaller OAR measurements, indicating that they are less frequently photographed in close proximity to the object.

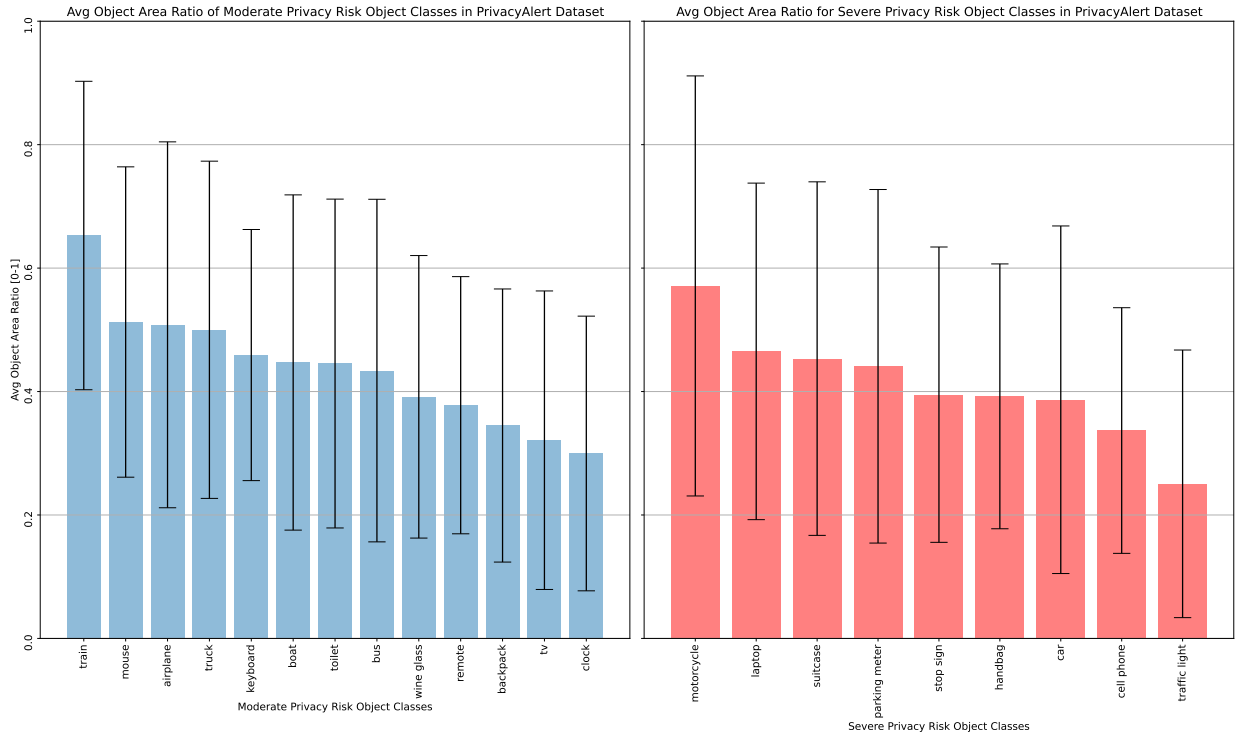


Figure 5.13: This figure displays the average OAR measurement of the moderate and severe privacy risk object labels across the PrivacyAlert Dataset. The standard deviation of the object's measurement is shown in the error line.

VISPR Dataset: Object Area Ratio of objects. The analysis of the Object Area Ratio (OAR) of private object classes within the VISPR dataset is displayed in Figure 5.11c. The results indicate that certain non-private objects, namely `bed`, `dining table`, `cat`, and `couch`, occupy a larger portion of the image, with an average OAR greater than 0.59, accounting for more than 55% of the images. Interestingly, only one private object, namely `train`, displayed a large OAR measurement on average, consistent with previous analyses of Open Images v7 and PrivacyAlert datasets. Furthermore, the OAR analysis of the VISPR dataset revealed that private objects, such as `clock`, `parking meter`, and `traffic light`, have small average OARs across the object classes, which is in line with previous findings from Open Images v7 and PrivacyAlert OAR scores.

VISPR Dataset: Object Area Ratio of the Private Objects. From the analysis of the VISPR dataset, the private object classes were examined for their Object Area Ratio (OAR) in Figure 5.14. The results follow similar trends to Open Images and PrivacyAlert dates. Vehicle objects are often larger in images than other objects in the dataset: `train` (0.57), `motorcycle` (0.51), `airplane` (0.49), and `truck` (0.47). Private objects related to computer accessories (`mouse`, `laptop`, `keyboard`) had a high average OAR measurement as well. There is also a trend of the objects having lower OAR scores across the datasets which includes such as `tv`, `clock`, `cellphones`, and `traffic light` (0.23).

Discussion. The Object Area Ratio (OAR) measurement is a technique used to analyze images with complex objects or multiple scenes. It quantifies the visibility of objects by calculating the size of objects detected, providing vital information about the overall composition of the image, and identifying key features and attributes for analysis. The OAR methodology results show the ratio of objects in the visual content datasets, with the analysis revealing that private objects tend to occupy a larger area in images. These results highlight the importance of considering the OAR of private object classes in image privacy analyses and suggest that these objects may pose a higher privacy risk due to their larger

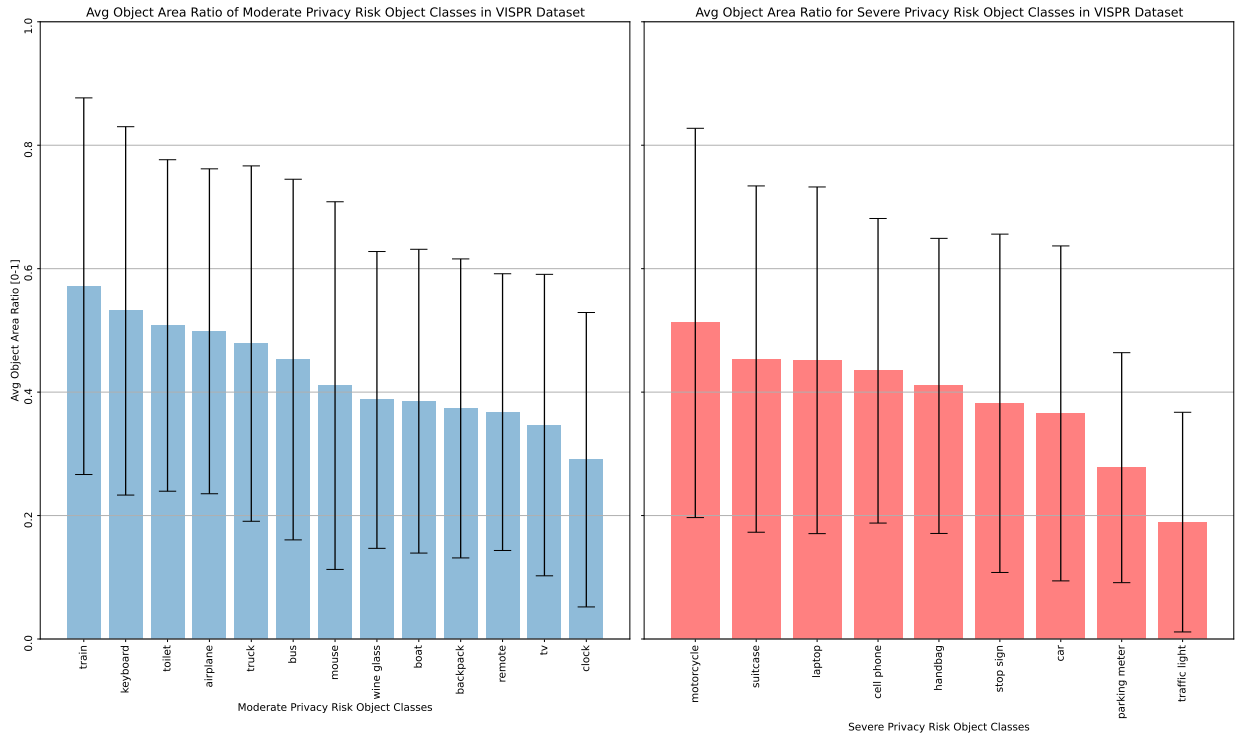
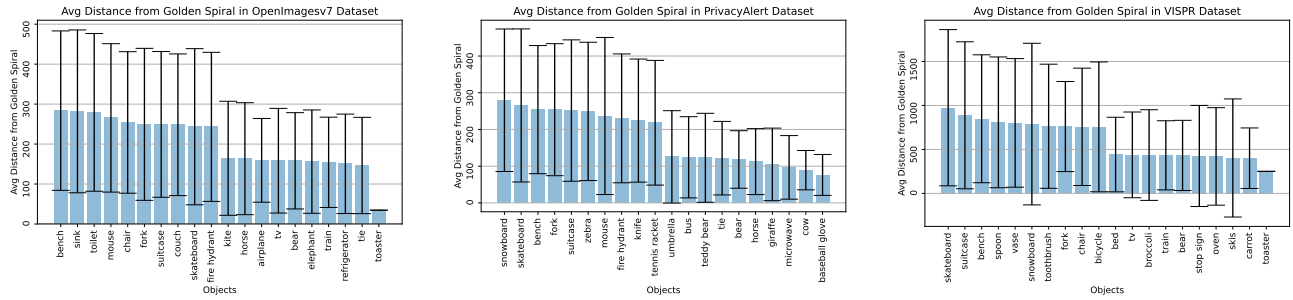


Figure 5.14: This figure displays the average OAR measurement of the moderate and severe privacy risk object labels across the VISPR Dataset. The standard deviation of the object's measurement is shown in the error line.

spatial coverage in images. Automotive objects have large OAR measurements which could make items like license plates more visible and even increase the privacy risk due to noticing landmarks, signs, or objects to reveal their location.

Visual Feature: Golden Spiral Distance

The Golden Spiral Distance (GSD) measurement analyzes visual data by quantifying the visibility of the object based on how close an object is to the most appealing portions of an image. The distance is measured from the center of the objects' bounding box to the closest point on the golden spiral. For these experiments, the golden spiral has a fixed starting location starting from the lower left corner of the image. The results below show the average distances for the objects that are considered both public and private in the datasets using the GSD methodology.



(a) Open Images v7 (b) PrivacyAlert (c) VISPR

Figure 5.15: This figure displays the average GSD measurement of the top 10 and bottom 10 objects across the visual content datasets. The standard deviation of the object's measurement is shown in the error line. (a) Open Images v7, (b) PrivacyAlert, (c) VISPR

Open Images v7 Dataset: Golden Spiral Distance for Objects. The average distance of an object from the golden spiral for the Open Images v7 dataset is shown in Figure 5.15a. This figure shows the objects that are the farthest and closest to the golden spiral. The objects that are farthest away are `bench`, `sink`, and `toilet`. When analyzing the OAR measurements, the private object *toilet* had a larger GSD score; however, it can be noted that the object on average, is farther away from the spiral. The objects that are the closest to the spiral are `refrigerator`, `tie`, and `toaster`. Three of the private classes are typically farther away from the golden spiral; there are also three private objects that are close to the spiral.

Open Images v7 Dataset: Golden Spiral Distance of the Private Objects. Through an analysis of the Open Images v7 dataset, the private object classes were examined with respect to GSD in Figure 5.16. The results indicate that automotive objects, such as `train`, `airplane`, and `truck`, are often closer to the golden spiral in images than most private objects in the dataset. The *Severe* privacy risk graph reveals that objects, such as `stop sign`, `parking meter`, and `traffic light`, have smaller distance from the golden spiral, indicating that they are more likely to identified.

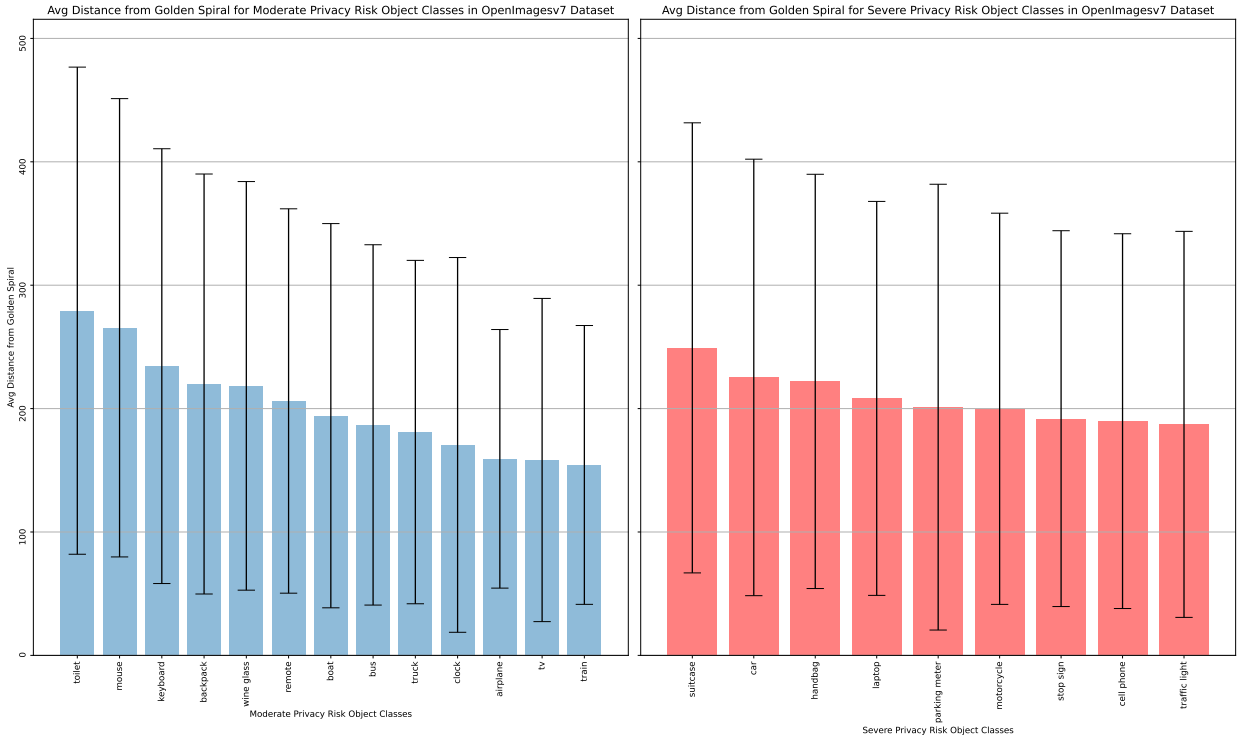


Figure 5.16: This figure displays the average GSD measurement of the moderate and severe privacy risk object labels across the Open Images v7 Dataset. The standard deviation of the object's measurement is shown in the error line.

PrivacyAlert Dataset: Golden Spiral Distance for Objects. The analysis of the PrivacyAlert dataset yielded interesting findings regarding the GSD of objects in images. As depicted in Figure 5.15b, `snowboard`, `skateboard`, and `bench` objects tend to be farther away from the golden spiral in images compared to objects. Similar to the Open Images v7 dataset, the `mouse` object also has a larger GSD. Interestingly, of the 10 closest objects only one private object, `bus`, has closest the average GSD seen in Figure 5.15b.

PrivacyAlert Dataset: Golden Spiral Distance of the Private Objects. Through an analysis of the PrivacyAlert dataset, the private object classes were examined with respect to GSD in Figure 5.17. The results show a similar trend to Open Images v7 dataset where the vehicle objects are closer to the golden spiral in images. The figure also shows `toilet`, `mouse`, and `wine glass` as having the highest distance in the *Moderate* privacy risk classes.

`suitcase`, `parking meter`, and `motorcycle` objects were the furthest from the golden spiral in the *Severe* privacy risk classes.

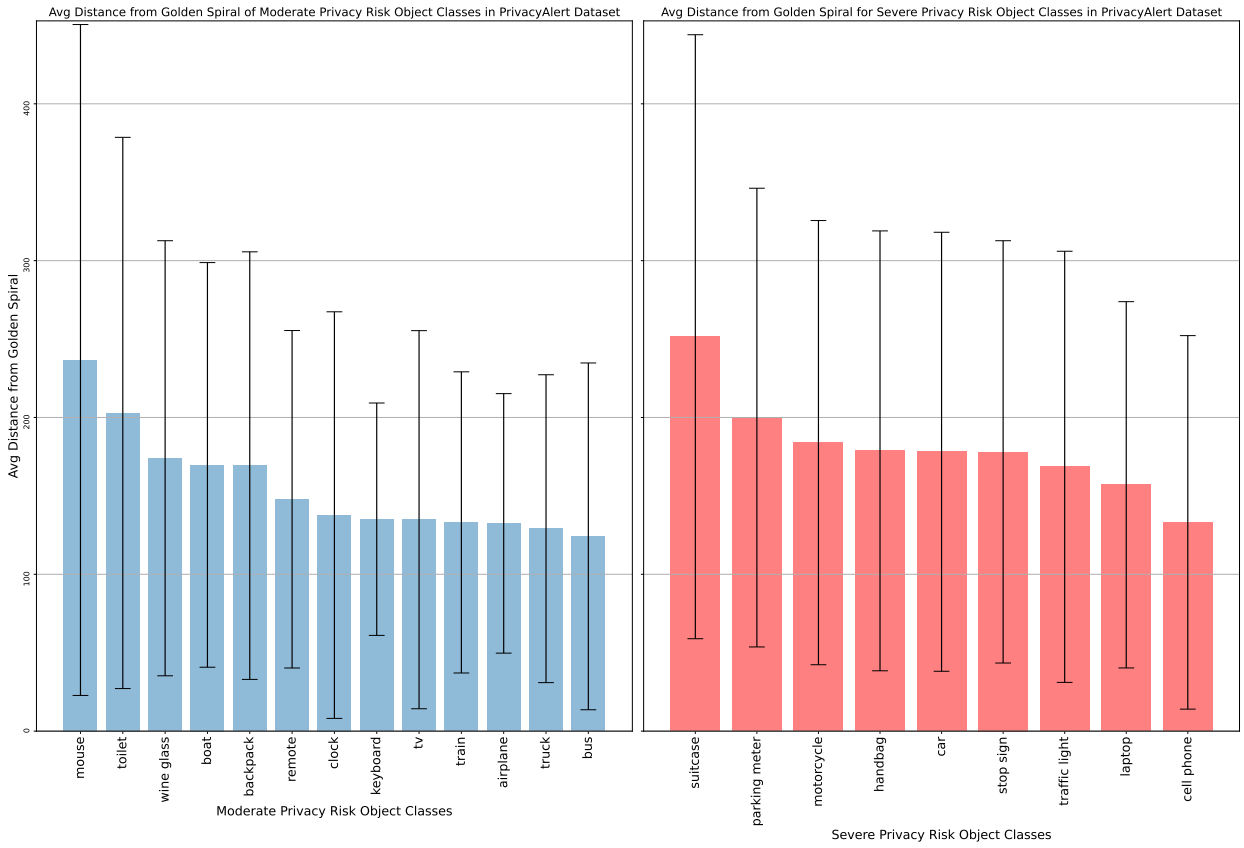


Figure 5.17: This figure displays the average GSD measurement of the moderate and severe privacy risk object labels across the PrivacyAlert Dataset. The standard deviation of the object's measurement is shown in the error line.

VISPR Dataset: Golden Spiral Distance for Objects. This analysis investigates the distribution of objects in the VISPR dataset with respect to the golden spiral. The analysis reveals the average distance of objects from the golden spiral, as depicted in Figure 5.15c, and identifies the objects that are farthest and closest to it. In Figure 5.15c, `skateboard`, `suitcase`, and `bench` objects tend to be farther away from the golden spiral in images compared to objects. The graph follows similar trends to the Open Images v7 (Figure 5.16) and

PrivacyAlert (Figure 5.17) graphs with these objects being in the top 10 distances that are the farthest away.

VISPR Dataset: Golden Spiral Distance of the Private Objects. Through an analysis of the VISPR dataset, the private object classes were examined with respect to GSD in Figure 5.18. The results show a similar trend to Open Images v7 and PrivacyAlert datasets where the vehicle objects are closer to the golden spiral in images. The figure also shows `toilet`, `mouse`, and `keyboard` as having the highest distance in the *Moderate* privacy risk classes. The *Severe* privacy risk graph further corroborates that objects dealing with traffic are closer to the golden spiral.

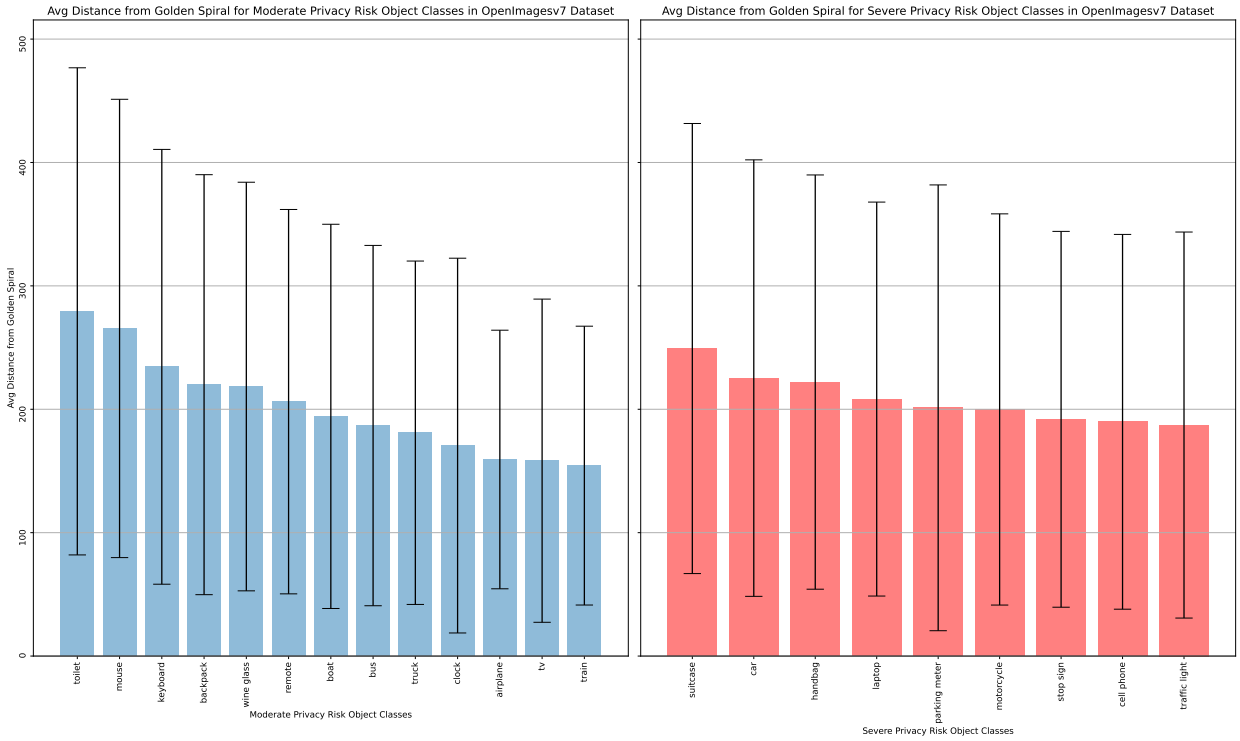


Figure 5.18: This figure displays the average GSD measurement of the *moderate* and *severe* privacy risk object labels across the VISPR Dataset. The standard deviation of the object's measurement is shown in the error line.

Discussion. The analysis of the three datasets (Open Images v7, PrivacyAlert, VISPR) yielded interesting findings regarding the distance of objects from the golden spiral in images. The Open Images v7 dataset showed that vehicle objects tended to be closer to the golden spiral, while objects such as the `toilet` and `sink` were the farthest away. In the PrivacyAlert dataset, `snowboard`, `skateboard`, and `bench` objects were farthest from the golden spiral, while vehicle objects were again closer to it. The visual datasets showed that the average distance of objects from the golden spiral varied significantly across the object classes. Overall, these findings suggest that the location of an object in an image with respect to the golden spiral could provide insights into the visual privacy risks in the image, as well as the object's identifiability.

5.4.2 Experiment 2: An Empirical Comparison of *VPScore* and *Vango* Privacy Risk Scoring Algorithms for Visual Dataset Analysis

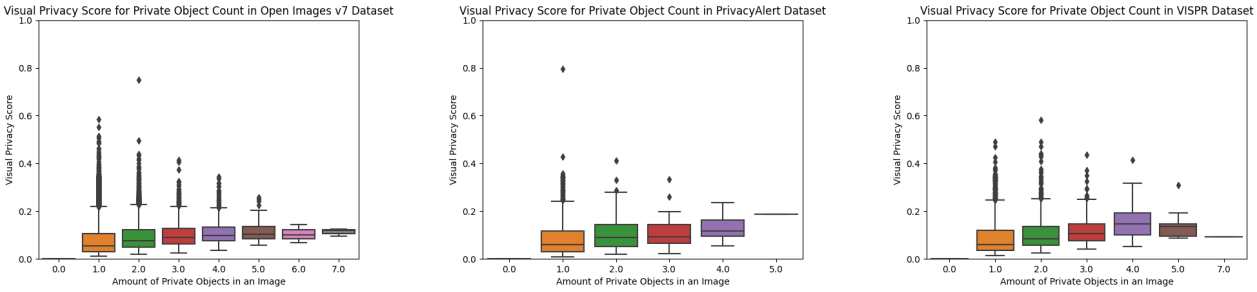
This study aims to assess the efficacy of two privacy risk scoring algorithms, *VPScore* and *Vango*, for analyzing images at scale. To obtain the *Vango* privacy risk score, I combine visual feature attribute scores with weights. The evaluation of these methods is conducted across three diverse datasets: PrivacyAlert, VISPR, and Open Images v7. The primary objective of this research is to compare the *VPScore* and *Vango* privacy risk scoring algorithms in identifying potentially sensitive content in the datasets.

Visual Privacy Risk Scores for Private Objects

In this study, I evaluate the visual privacy score for n private objects across the datasets. The x-axis is the n amount of private objects in an image from the dataset. The range of n is between 0 to 7 private objects. The y-axis shows the visual privacy risk score for images in the dataset containing n private objects. The range of the visual privacy risk score is

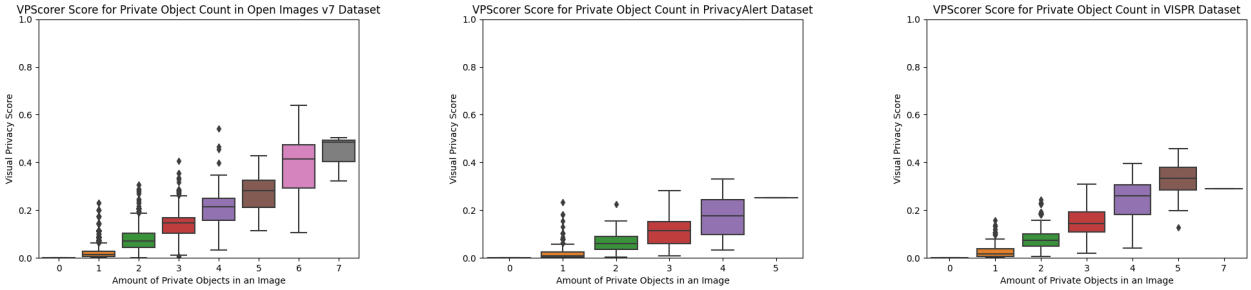
between 0 to 1. The lower the privacy score is the less privacy implications are; the higher the privacy score the more risk an image is in the dataset. In Figures 5.19 and 5.20, we look at the visual privacy risk scores for the *VPScore* and *Vango* algorithms.

The visual privacy risk scoring algorithm proposed in this chapter, *Vango*, has a similar score range across the datasets with respect to the n private objects. As the number of private objects increases, the visual privacy risk scores shown have a consistent score as it approaches the largest n private objects in an image. The score range from *Vango* is between 0 to 0.9. The *VPScore* privacy risk scorer also shows consistent score ranges across all three datasets. The high increase in privacy risk scores is in correlation to the number of objects in the images across the dataset. From the three datasets, the score ranges between *VPScore* is between 0 to 0.7. With both algorithms, it can be noted that the privacy scores increase as more objects are present; however, in the VISPR dataset when the number of objects per image is 7, both scores show a drop.



(a) Open Images v7 (b) PrivacyAlert (c) VISPR

Figure 5.19: This figure displays the *Vango* visual privacy score using objects across the visual content datasets. The standard deviation of the object's measurement is shown in the error line. (a) Open Images v7, (b) PrivacyAlert, (c) VISPR



(a) Open Images v7 (b) PrivacyAlert (c) VISPR

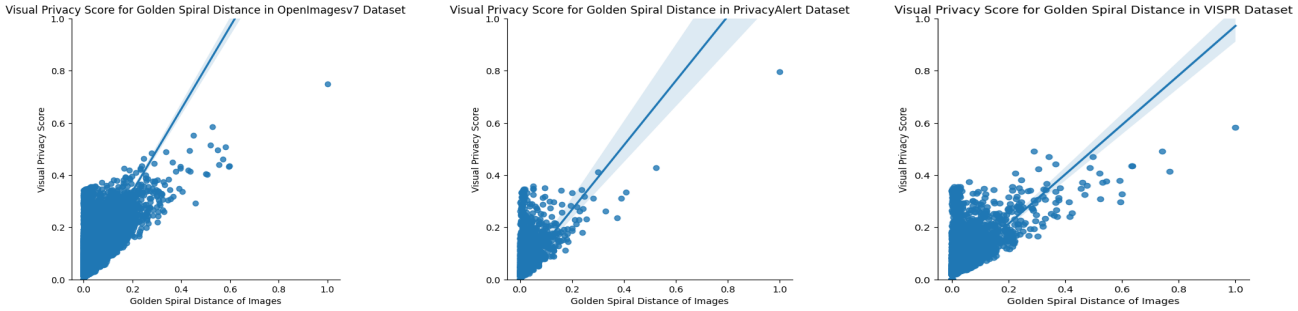
Figure 5.20: This figure displays the *VPScorer* visual privacy score using objects across the visual content datasets. The standard deviation of the object's measurement is shown in the error line. (a) Open Images v7, (b) PrivacyAlert, (c) VISPR

Understanding Visual Feature Influence in the *Vango* Algorithm

For this analysis, I look at the influence of visual features in the visual privacy score for the *Vango* algorithm is shown across three datasets. In Figures 5.21 to 5.23, the x-axis ranges are over the OIW weights, golden spiral distance, and object area ratios respectively. The y-axis shows the typical visual privacy risk score for images in the dataset in a specific range of the respective feature measurement. The range of the visual privacy risk score is between 0 to 1. Looking at the trends in Figure 5.21, it can be predicted that the lower the OIW weight the larger the image privacy score is. This visual feature also correlates with the assumption that the private objects across a dataset occur less, thus giving the object more importance. The object area ratio measurements show a closer trend in the graph, indicating that the OAR impact on the visual privacy score directly increases (shown in Figure 5.23). The graphs in Figure 5.22 show a different correlation. The closer an object is to the golden spiral will decrease the visual privacy score. This insight implies that private objects will not be close to the spiral.

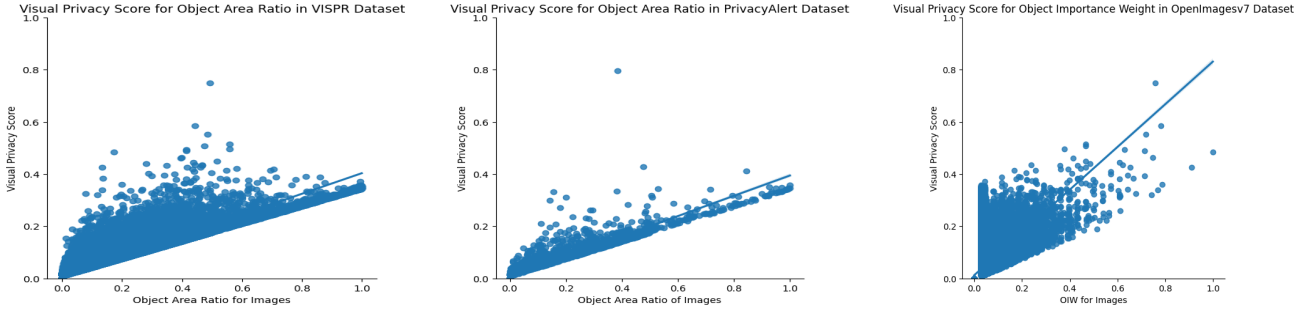
(a) Open Images v7 (b) PrivacyAlert (c) VISPR

Figure 5.21: This figure displays the *Vango* visual privacy score in respect to the Object Importance Weights across the images in the visual content datasets: (a) Open Images v7, (b) PrivacyAlert, (c) VISPR



(a) Open Images v7 (b) PrivacyAlert (c) VISPR

Figure 5.22: This figure displays the *Vango* visual privacy score in respect to the Golden Spiral Distance across the images in the visual content datasets: (a) Open Images v7, (b) PrivacyAlert, (c) VISPR



(a) Open Images v7 (b) PrivacyAlert (c) VISPR

Figure 5.23: This figure displays the *Vango* visual privacy score in respect to the Object Area Ratio across the images in the visual content datasets. (a) Open Images v7, (b) PrivacyAlert, (c) VISPR

5.5 Summary

This chapter addresses the third question of this dissertation, "*How can importance, prominence, and identifiability contribute to visual privacy risk scoring methodologies?*". The insights from this chapter were drawn with the use of object detection using YOLOv5, by designing a privacy hierarchy using MS COCO labels and adopting the privacy taxonomy defined in Section 3.2.2, and by incorporating existing visual content datasets to investigate the efficacy of visual features and applications two visual privacy risk scoring algorithms: *VPScorer* and *Vango*. It further demonstrates the application of visual features (i.e., Object Importance Weight, Golden Spiral Distance, Object Area Ratio) in visual privacy risk scoring, showing a trend between the use of OIW and OAR visual features and *Vango* privacy scoring methodology.

This analysis affirms that OIW identifies sensitive objects across a visual dataset by increasing the weight of private objects due to them being less frequent. OIW uses labels generated from the object detection algorithm to extract significant objects from within the visual data set. The experiments show that using OIW Equation (5.5) contributes to object importance and subsequently the sensitivity values of the privacy score. The OAR measurement (Equation 5.6) computes the size of the bounding box of each object in each image. The results show that automotive objects have large OAR measurements which makes license plates, landmarks, and signs increased privacy risks. With further exploration, the OAR measurement could show promise to enhance privacy scoring methodologies by considering object spatial coverage in images. Additionally, the analysis confirms that OAR can be applied to privacy risk scores, showing the visibility of the object correlates to its' privacy score. The findings in this study did not show a correlation between GSD and the visual privacy risk score. GSD computes a score representing the likelihood that the object would be viewed (shown in Equation 5.7 and Listing 1). This visual feature could provide insight into the privacy risks associated with the object's identifiablility.

In Section 5.2.1, an existing privacy score was adapted to be applied to visual content datasets. In addition to adapting an existing privacy risk scoring method, the *Vango* privacy risk scoring method was developed using the visual features as a backbone. The visual privacy algorithms were analyzed over n private objects across three datasets. The results show that across the visual datasets, the visual privacy risk scores have a consistent score increase as it approaches the largest n private objects in an image. In summary, this chapter argued that privacy risk in visual content can be accomplished by quantifying visual features. I apply a number of theories and concepts to establish visual feature measurements that can be incorporated into a visual privacy risk score. The need for visual privacy risk scoring is not limited to binary or dichotomous approaches but can leverage computer vision and visual analysis to improve understanding of visual content.

Chapter 6

An Interactive Audit Pipeline for Investigating Privacy and Fairness in Visual Privacy Research

As society progresses, people can become more dependent on the accessibility and convenience that technology offers. Every day a large amount of visual content is uploaded to SMNs and collected by smart city servers across the globe, which can explain the large amounts of sensitive data that is available online. While these ecosystems have goals that revolve around helping people build connections with others; there are gaps in the methods used to protect the information of individuals and corporations who share or collect content (Krishnamurthy and Wills 2008; Gross and Acquisti 2005; Rosenblum 2007; Madejski et al. 2011; Van Zoonen 2016; Elmaghraby and Losavio 2014; DeHart et al. 2020c). A need for visual privacy has emerged from SMNs and the integration of technology in smart cities that can expose sensitive information through visual content (Korayem et al. 2016; Hoyle et al. 2015; Sánchez-Corcuera et al. 2019). The constant sharing and storing of videos and images bring skepticism about individual privacy and rights (Such et al. 2017b; Zhong et al. 2018). Visual privacy techniques extend from SMNs, smart cities, lifelogging, and much

more (DeHart et al. 2020a).Various harms can occur as a result of sensitive information being displayed, which makes visual privacy a growing area of concern (Gross and Acquisti 2005; Li et al. 2017c; Rosenblum 2007). Existing technologies in the industry can show a disregard for protecting the information of individuals who share visual content or for individuals who are captured in the content (Madejski et al. 2011; Elmaghraby and Losavio 2014).

Researchers have created datasets, models, and deployed applications that they believe will provide privacy to its' users (Arlazarov et al. 2019; Tonge and Caragea 2016, 2020; Li et al. 2017b; Zhong et al. 2018; Zerr et al. 2012c; Tierney et al. 2013). Within these algorithms and systems, researchers should continually make decisions to assess the fairness, privacy, and accessibility of the data and model in regard to the communities they serve. Bias can be curated from the data collection process, reinforced in the model's training, and systematically imposed in the deployment phase (Suresh and Guttag 2019). While research is being done to address these concerns, a gap exists in understanding the overlap between fairness, privacy, and human feedback for visual privacy issues in the machine learning (ML) pipeline. With privacy and bias issues arising throughout the ML pipeline, it provokes the question: *can visual privacy systems bring rise to additional privacy and fairness risks for individuals and stakeholders?* In an ideal world, deployed ML models will enhance our society. Researchers hope that those models will provide unbiased and ethical decisions that will benefit everyone. However, this is not always the case; issues arise during the data curation process and throughout the steps leading to the models' deployment. The continued use of biased datasets and biased processes will adversely damage communities and increase the cost of fixing the problem later.

The goal of this chapter is two-fold. First, it aims to understand visual privacy and fairness as their issues intrude into the ML pipeline and potentially impact the stakeholders and community where the ML pipeline is deployed. Secondly, I provide a comprehensive pipeline indicating fairness and privacy issues and propose auditing strategies to reduce these

effects in visual privacy research. I examine visual privacy research and draw lessons that can apply broadly to artificial intelligence (AI) and I observe the critical decisions that are often overlooked when deploying AI. In this chapter, I walk through the decision-making process that a researcher must make before, during, and after their project to consider the broader impacts of their research on the community. Throughout this chapter, I discuss several privacy, fairness, and ownership issues that can arise in the ML pipeline (Sections 6.1, 6.2). I argue for the use of human-*over*-the-loop strategies to discover privacy and fairness issues in the ML pipeline. I extend this technique to suggest two auditing processes: Fairness Forensics Auditing System (FASt) and Visual Privacy (ViP) Auditor (Section 6.3). Finally, reflect on the need to review research agendas focusing on harmful societal impacts (Section 6.4).

6.1 Defining the Machine Learning Pipeline

I describe the ML pipeline as having three phases (Figure 6.1). Phase 1 is the **Data Preparation** process. This phase includes considerations of (1) raw data sources, (2) data collection processes, (3) data storage, and (4) data cleaning processes that a researcher should explore before entering into the next phase. Data can come from anywhere and everywhere. With so many data source possibilities available, the researcher should consider which sources are relevant to them. The data collection process for researchers can include using existing image datasets, social media datasets, or web scraping methods with respect to the visual privacy research task. Once a dataset is collected, a researcher could employ data cleaning tasks (e.g., crowd-sourced labeling) to derive an optimal dataset and labels.

In Phase 2, shown in Figure 6.1, I begin the **Modeling** process. The cleaned data from Phase 1 can be divided into three datasets: training, testing, and validation. Training data is used as input for the ML algorithm. After training with the researcher's desired ML algorithm, the researcher receives a model to run testing and validation datasets on. The model provides the output of the performance with several metrics. This new information

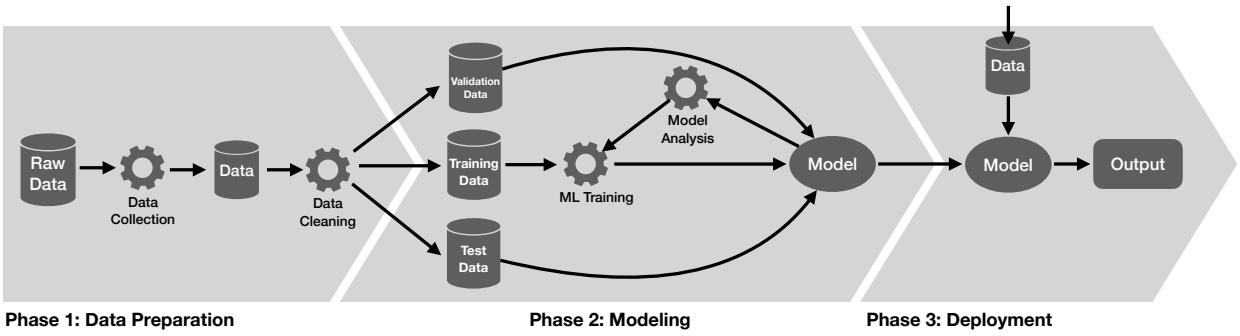


Figure 6.1: This figure illustrates the traditional ML pipeline. The pipeline has three phases: data preparation, modeling, and deployment.

can be used for refining the model before entering the final phase of the pipeline.

The last phase of the proposed machine learning pipeline is **Deployment**. The **Deployment** phase uses real-world data as input for the selected model. The researcher or end-user will see the real-world results and impact of their selected model from Phase 2. This phase allows the researcher to evaluate their model's performance and impact on the communities they serve.

6.1.1 The Guise of Pipeline Ownership

Researchers must consider who has ownership of the data and model at each phase before beginning these processes. These considerations are essential when protecting the privacy of individuals and biases that someone could impose.

At the **Data Preparation** and **Deployment** phases, the researcher should consider who are the owners of the data and how they are receiving the content. This can explore if online visual content belongs to the users or a corporation, if existing datasets belong to the proprietary researchers, or if agreements on volunteered data belong to companies or individuals. Furthermore, if the researcher is using online resources for data processing, the researcher should consider: how the data is stored, does it still belong to the researcher, and what information is being stored on these platforms.

In the **Modeling** and **Deployment** phases, the researchers should consider who holds

ownership of the model. Considerations should be given to understand if the rights and ownership of the model are owned by the researcher, company, or third party (Neyaz et al. 2020).The data uploaded to the model could lead to an individual perceiving that their privacy has been breached due to the authorization and ownership of the model. If a third party owns the model, it is important to consider what information they are collecting from the use of it and who they share this information with. Throughout each phase of the ML pipeline, the stakeholders should continue to ask tough questions and make critical decisions that are ethical, fair, and in the best interest of those the technology is meant to serve.

6.2 Exploring Privacy and Fairness Concerns in the Visual Privacy ML Pipeline

Efforts to implement technology that serves to mitigate harm is the motivation behind visual privacy research. Positive outcomes are desired from systems that are created to help solve society's most pressing issues, such as visual privacy leakage on social media and in smart cities. In this section, I will discuss privacy and fairness issues frequently occurring with developing and deploying visual privacy systems. Examples of these issues are shown in Figure 6.2 and Table 6.1. I suggest that when evaluating visual privacy systems, researchers should consider bias issues as they arise in the ML pipeline. As apparent from Figure 6.2, fairness issues are involved with all stages of the ML pipeline. This investigation will comprise three over-arching visual privacy issues and describe how they could affect the ML pipeline.

6.2.1 Privacy

Visual privacy issues can arise at any point in the ML pipeline. The stakeholders and researchers must be aware of these issues and develop ways to solve them proactively as they arise. This section discusses three visual privacy issues that can arise in the ML pipeline:

Table 6.1: This table displays the privacy and fairness issues in various phases of the machine learning pipeline. The description provides a high-level overview of what those issues are. The checkmark (✓) indicates that those issues could arise in that part of the pipeline.

			Phase 1				Phase 2				Phase 3		
		Description	Raw Data	Data Collection	Data	Data Cleaning	Training Data	ML Training	Model	Model Analysis	Data	Model	Output
Privacy issues	Obtaining Content Consent	Exploring the ethics and privacy methodology of researchers to obtain consent for collected visual content on a public domain.		✓		✓					✓		✓
	Multiparty Conflict	Understanding privacy concerns for images and videos which are owned by multiple persons.	✓	✓	✓	✓		✓		✓	✓		✓
	Image Removal Request	Determining when and how visual consent should be removed from the pipeline via requests.			✓	✓		✓	✓	✓	✓		✓
Fairness issues	Historical bias	The inherent bias from a biased world is absorbed by the source data. Suresh and Guttag 2019	✓								✓		
	Algorithmic bias	The bias relates to the algorithm in the ML pipeline, and it could have different bias sources and types. Bantilan 2018 Calmon et al. 2017 Danks and London 2017					✓	✓	✓	✓		✓	✓
	Software Discrimination	The output from a predictive software used to aid in decision making may lead to unfair consequences. Galhotra et al. 2017										✓	✓
	Individual fairness	Similar individuals should be treated as similarly as possible. Dwork et al. 2012					✓	✓	✓	✓		✓	✓
	Group fairness	The groups defined by protected attributes should obtain similar treatments or equal opportunity as the privileged group. Hardt et al. 2016					✓	✓	✓	✓		✓	✓
	Disparate treatment	Protected attributes are directly applied in modeling where unfairness occurs. Zafar et al. 2017					✓	✓	✓	✓		✓	✓
	Disparate impact	Even though the protected feature is not directly used, its relevant features still could lead a selection process to make unfair outputs. Feldman et al. 2015					✓	✓	✓	✓		✓	✓

visual content consent, multiparty conflict, and image removal requests.

Obtaining Visual Content Consent

Researchers use large public image datasets (Zerr et al. 2012c; Lin et al. 2014; Deng et al. 2009) to train ML algorithms to perform various visual privacy research tasks (Tonge and Caragea 2016, 2020; Zerr et al. 2012a). Additionally, when collecting a large amount of data, many researchers question the use of *web scraping* methods to obtain this data (Zimmer 2010; Zimmer and Kinder-Kurlanda 2017; Mancosu and Vegetti 2020; Krotov and Silva 2018) and the use of crowd-sourcing methods to label data (Lin et al. 2014; Deng et al. 2009; Xiao et al. 2010; Torralba et al. 2008). While researchers' efforts can focus on creating systems to help with visual privacy, their approach in collecting data can bring rise to privacy and ethical concerns in Phase 1 of the machine learning pipeline. The methods that researchers use to collect this data can overlook individuals' privacy, consent, and protection. When collecting visual content or using existing datasets, researchers can un-intentionally collect private content containing minors or bystanders (Perez et al. 2017; Dimiccoli et al. 2017; Hasan et al. 2020; Birhane et al. 2021).

The topic of consent is essential to gauge participants' willingness to participate in the study or research. For traditional studies that include people or living subjects, specific procedures and policies need to be followed according to a governing entity (i.e., an institutional review board). The visual data collection processes do not abide by any standard practice policies or procedures when using personal data. Visual content consent issues begin to arise in the **Data Preparation** phase and can continue to be a pressing issue during the **Deployment**phase. There is no quick or easy way to handle this issue if consent is collected too late in the ML pipeline. If this visual privacy issue is resolved early, researchers can tangentially reduce issues with multiparty conflicts and image removal requests.

Multiparty Conflict (MPC)

Visual content can seem to be owned by multiple people or entities (Such et al. 2017a). Co-ownership issues can arise in various situations, a few of these scenarios can include: (1) group photos, (2) reposting images or videos of others (e.g., children, pets), or (3) a person having physical possession of images of other people on posting on social media (Zemmels and Khey 2015). Multiparty conflicts can affect the privacy of minors (Lwin et al. 2008; Batool 2020) and bystanders (Li et al. 2019; Perez et al. 2017) when discussing ownership and consent. Co-owned visual content can cause visual privacy leakage for others without it being the individual's intent (DeHart and Grant 2018). Ownership can also extend to the stakeholders, organizations, and companies who collect, store, and host this content in their environments. In the ML pipeline, the researcher should consider possible issues for MPCs in all phases.

Considerations for content ownership and individual rights should be made early in the ML pipeline. When working with visual content, it can be necessary to seek permission from all parties involved. Multiparty conflicts can enter the ML pipeline as early as the **Data Preparation** phase. In the **Deployment** phase, the real-world data used for the ML task can bring additional concerns for this issue.

Image Removal Requests

When collecting data or using existing datasets, ownership issues will arise and should be addressed early and appropriately. Instead of using public resources, researchers should seek participation consent from individuals. This becomes important when using data for research and in deployed systems. This raises the issue of what to do if an individual's visual content is requested or petitioned to be removed from the dataset and the model's training phase. In July 2020, MIT decided to remove the 80 Million Tiny images dataset because of the bias and offensive labels that occurred in the dataset (Torralba et al. 2020). If researchers have used this dataset, these issues can affect the credibility of their work and

the deployed system. Image removal requests can affect all phases of the ML pipeline and should be handled accordingly.

6.2.2 Fairness

In this section, I discuss three typical fairness issues. These fairness issues sneak into most phases of the ML pipeline. These issues can lead researchers to consider where or when bias can occur. Later, in the algorithmic bias section, I will discuss additional biases (i.e., *individual fairness* versus *group fairness*, *disparate treatment* versus *disparate impact*) that explore who is affected and how those issues arise in the pipeline.

Historical Bias

When data is generated, the inherent bias from the world could stealthily engrave into data. Historical bias can enter the ML pipeline at the start of the **Data Preparation** phase through the **Deployment** phase. Even under ideal sampling and feature selection, historical bias could still exist and cause concern. When the historical bias proliferates through the ML pipeline, it can impact modeling and decision-making in the deployment stage (Hellström et al. 2020; Suresh and Guttag 2019).

Algorithmic Bias

Algorithmic biases are bound together with each process in the ML pipeline. Roughly, algorithmic bias is focused in the **Modeling** phase. Because algorithms are connected with every part of ML systems, there are different bias sources and types from different components of the ML pipeline. The algorithm's bias could be sourced from biased training data, a biased algorithm, or misinterpreting the algorithm's output (Danks and London 2017). Identifying the source of algorithmic bias contributes significantly to dissolving fairness issues. In addition, researchers must also consider the types of algorithmic bias. It is typical to think about who is the victim impacted by algorithmic bias. For example, similar individuals are treated

inconsistently based on the predictions of the model, while *individual fairness* requires that each similar individual should be treated as similarly as possible (Dwork et al. 2012). As a more general example, *group fairness* considers groups defined by protected attributes (e.g., gender, race), and it requires that the protected groups should obtain similar treatment as the privileged group (Hardt et al. 2016). *Group fairness* is also referred to as statistical parity or demographic parity.

After identifying who suffers from the algorithmic bias, it becomes increasingly important to understand how fairness issues arise in the ML pipeline. *Disparate treatment*, also known as direct discrimination or intentional discrimination, occurs when protected attributes are used explicitly in ML systems. Consequently, disadvantaged groups identified by the protected attributes are deliberately treated differently. *Disparate impact* is pervasive and entrenched in our society (Feldman et al. 2015). Regarding *disparate impact* in the ML pipeline, it exists under the guise of correlated variables that implicitly correspond to protected attributes.

Software Discrimination

Last but not least, software discrimination appears at the end of the entire ML pipeline, which is the **Deployment** phase. Bias could exist due to a problematic model. After an ML model is passed to its end-users, the interpretability and transparency of the model can benefit from identifying and mitigating potential bias generated by the software. Researchers have developed many tools that audit fairness for deployed ML models. Tools like IBM's AI Fairness 360 toolkit (Bellamy et al. 2019) implement fairness metrics and bias mitigation algorithms. Other works have generated test suites to measure software fairness from a causality-based perspective (Galhotra et al. 2017).

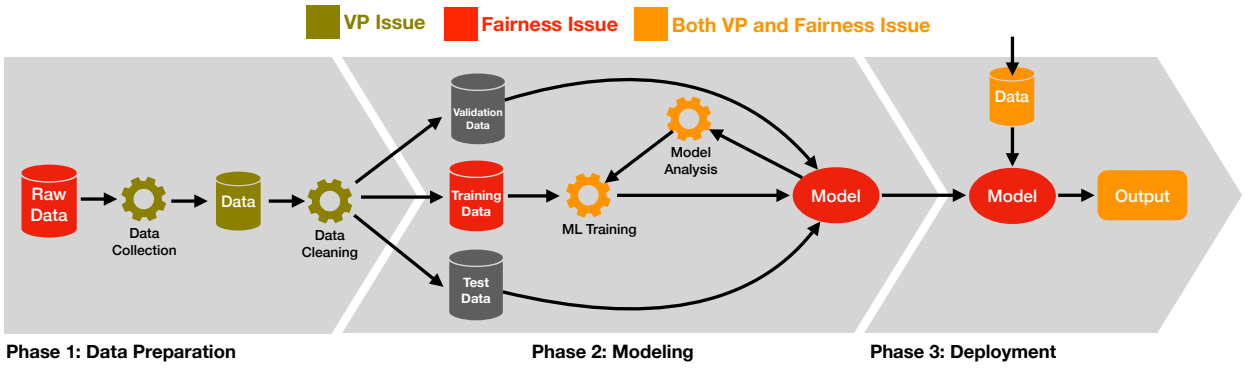


Figure 6.2: In the ML pipeline, I indicate where privacy (green) and fairness (red) issues could arise. Possible overlaps in the system are defined in orange.

6.2.3 Overlaps in Privacy and Fairness Issues

Figure 6.2 shows overlapping phases that contain visual privacy and fairness issues. When both issues arise, researchers should be ready to deal with them; otherwise, they will affect the system's outcome. For instance, a model builder perceives that the protected groups could be affected by the fairness issues in a facial recognition system. Consequently, the modeler strives to collect more data to make up for the disproportion. However, the increased visual privacy risk for individuals during the data collection process could be an unexpected problem and is increased for the underrepresented group (Raji et al. 2020).

It is essential to understand the relationship between visual privacy issues and fairness issues, since solving one issue could have a negative impact on the other. For instance, a user uploaded a picture to a biased ML model in the cloud. The user could experience unfair decisions from the biased model with simultaneous loss of privacy to the service provider. A goal of this chapter to raise awareness of such worse cases. The trade-off analysis between privacy and fairness will develop an in-depth understanding of building a process for visual privacy systems.

6.3 Integration of Interactive Audit Strategies for the Machine Learning Pipeline

ML models are constantly being updated once deployed to the real world; regular updates help avoid and minimize costly errors. Differences in the time between error discovery and model correction for the deployed model is crucial. Systems should be able to respond to unexpected bias before, during, and after deployment. It could be impossible to erase the damage caused by the aftermath of a system; however, stakeholders could start making a change now. One way to do this would be using an interactive ML approach, human-*in*-the-loop (Fails and Olsen Jr 2003; Amershi et al. 2014, 2015; Lee et al. 2019). Training in the human-*in*-the-loop framework requires humans to make incremental updates to anticipate issues (Bond et al. 2016). Traditional ML pipelines conduct training on their own without interference from humans. To debug these models, the researcher must begin a thorough investigation of the model's predictions, parameters, and data after the learning phase has been completed. An interactive approach would allow a human in the **Modeling** phase, which will reduce debugging and runtime. The human is able to check the learning for the model and coach the model to meet the desired results in a feedback cycle. Feedback cycles allow the researcher to provide positive feedback iteratively to the model after viewing the processes. This can allow the researcher to understand the possible bias and privacy issues in the model and mitigate it immediately. This approach can be extended to various ML research areas in fairness, computer vision, and privacy.

In traditional human-*in*-the-loop approaches, the human becomes a bottleneck for the feedback process. In light of this, I suggest using a **human-*over*-the-loop** approach (Graham et al. 2017). Human-*over*-the-loop allows researchers to step into the pipeline as needed to perform corrections. This removes the necessity of a human approving each iteration of the model. With this feature integrated in the ML pipeline, the researchers should consider having multiple humans to monitor the training. This, in turn, can lower response times

to resolve biases that may be imposed from humans during learning. Based on the human-*over*-the-loop technique, I propose the use of two interactive auditing strategies that can reduce fairness and privacy issues to allow researchers to conceptualize, develop, and deploy safer visual privacy systems.

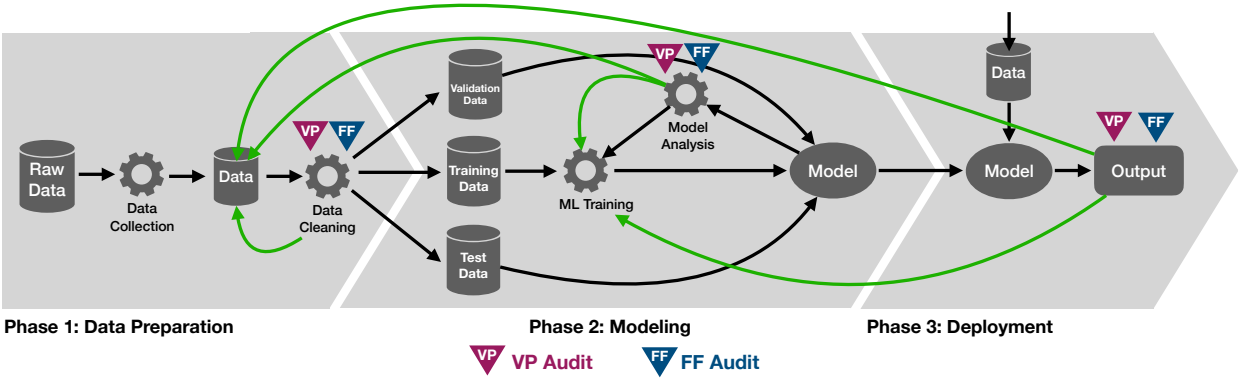


Figure 6.3: This figure illustrates the feedback loops when human-*over*-the-loop techniques are implemented. The green lines denote audit traces for feedback loops. The loops that are suggested to have a ViP Audit are denoted with a VP marker. The loops that are suggested to have FASt are denoted with a FF marker. Data Preparation has one feedback loop from Data to Data Cleaning. The Modeling phase has two feedback loops: (1) from model analysis to ML training, and (2) from Model Analysis to Data (in the Data Preparation phase). In the final phase of the pipeline, the deployment loops are from (1) Output to Data (in the Data Preparation phase) and (2) Output to ML training in the Modeling phase.

6.3.1 Incorporating a Fairness Forensics Auditing System (FASt)

ML bias is a rising threat to justice, and it has been investigated in broad areas, including employee recruitment, criminal justice, and facial detection. ML research can cause unanticipated and harmful consequences on daily life. Decision-makers begin to utilize the result of the output from ML algorithms without considering fairness. Fairness forensics focuses on supporting researchers and modelers to inspect a dataset or a new model by techniques and tools evaluating for bias.

Fairness forensics requires an overarching understanding of the types of bias, the ML pipeline, and analysis processes for bias at different stages of the ML pipeline. It is vital to understand how biases have harmful impacts on different communities of people when

deploying ML systems related to visual privacy. Fairness forensics has three major tasks: bias detection, bias interpretation, and bias mitigation. Researchers and domain experts can use fairness metrics to evaluate the input or output of ML models for bias detection. Bias report generator tools and bias visualization tools can facilitate analysis and interpretation of bias to understand the meaning and impact of bias detection results. Once the bias is discovered, bias mitigation strategies can be applied by the interventions to the input data, the algorithm, or the decision-making. Bias mitigation algorithms can be categorized into three types: pre-processing, in-processing, and post-processing algorithms (Bellamy et al. 2019).

6.3.2 Proposing a Visual Privacy (ViP) Auditor

Visual privacy content consent and leakage are important for individuals, stakeholders, and researchers. Mitigating visual privacy issues proactively can improve the ML systems' impact on the community. The auditor in this section focuses on supporting researchers and modelers to inspect the learning process for evaluating privacy.

For actively investigating visual privacy research, I propose the use of a human-*over*-the-loop technique specifically designed to handle privacy and computer vision issues. Most visual privacy systems are comprised of visual data and mitigation techniques that employ ML techniques. I envision the ViP Auditor as a comprehensive auditing tool that will enable researchers to use visual analytics (Liu et al. 2017) to understand the models' learning process. During learning, the modeler will be able to enhance the feedback process by using similar schemes as ModelTracker (Amershi et al. 2015) or Crayons Classifier (Fails and Olsen 2003). With a visual privacy auditor, the modeler will protect an individual's privacy in the ML process by incorporating visual privacy mitigation strategies built into the auditor. For model analysis, the researcher can obfuscate objects in visual content, understand the dataset attributes (e.g., number of faces, number of privacy leaks for each category), the models' classification performance, and the perceived privacy risk score of the model.

6.3.3 Integrating FASt and ViP Auditors in the ML Pipeline

From the **Data Preparation** phase in Figure 6.3, researchers examine the dataset, data labels, and ownership for the content regarding privacy concerns. This loop allows a researcher to consider the initial privacy concerns (in Section 6.2) and develop other strategies to mitigate them. In the **Modeling** phase, the researcher should employ auditors at both feedback loops (see Figure 6.3). The first feedback loop allows the researcher to conduct a privacy evaluation from the model's output. Evaluating the results from this feedback loop enables the human-*over*-the-loop to step in and make changes to achieve the desired level of privacy in the model. Auditing at this phase of the pipeline allows researchers to accurately correct recognition errors (bounding boxes, instance segmentation) from the models' learning. The second feedback loop conducts a privacy evaluation that allows the researcher to identify issues within the dataset from the Model Analysis. When the dataset issues are identified, the researcher can collect more data, remove the data from the pipeline, or add more tags/labels to mitigate privacy concerns that arise. The **Deployment** phase feedback loops consider the real-world output from the model. With auditors in place at this phase, the stakeholders can understand privacy issues as they arise. The stakeholders can fix issues in deployment as they arise by sanitizing the data and re-training the model. The ViP Auditor will produce a privacy risk score based on the models' performance and flag potential privacy issues.

The feedback loop from FASt is similar to the loop from the ViP Auditor. Fairness forensics system feedback occurs at different steps in all three phases of the ML pipeline (see Figure 6.3), and it can encourage researchers to sanitize their data or adjust the model. The process of fairness forensics allows the human-*over*-the-loop to determine the need for human intervention and assess for fairness in order to achieve social justice. Imperfect fairness metrics or conflicting fairness objectives (Friedler et al. 2021) means humans will need to intervene to maintain performance guarantees.

6.4 Summary

This chapter addresses the fourth research question of this dissertation, "*How can privacy and fairness risks created by the development and deployment of visual privacy systems be mitigated to protect individuals and stakeholders?*". Researchers should closely monitor data preparation, modeling, and deployment processes to avoid harming communities and stakeholders. The decision-making process for researchers can be challenging, but it is imperative to continually evaluate to improve the model's learning process and the deployment outcomes for the communities they serve. When building a visual privacy model needing large amounts of data, it can be easy to obtain datasets that are already widely distributed but may not have been examined for discriminatory, private, or fairness issues. This work discusses privacy and fairness issues that frequently occur in the ML pipeline that could emerge at various phases. I also assert the need for a responsible auditing system to bring accountability into model training and the deployed system. To do this, I propose using human-*over*-the-loop strategies to introduce interactive auditing for fairness and privacy. With ML pipeline audits and engaged researchers, the evaluation and consideration given to project development and deployed systems can become a standard procedure. These proposed mitigation strategies are the first steps of a much-needed effort to address privacy and fairness issues in the ML pipeline.

Being mindful of the societal impacts, evaluation methods (i.e., FASt and ViP) and monitoring strategies (i.e., human-*over*-the-loop) have been presented as mitigation techniques to reduce errors in the ML pipeline and in the deployed system's life cycle. However, there are no full-proof techniques for ensuring that the software is exempt from producing harm. For a stakeholder to know when to halt deployment implies that they have developed a plan for the system and require human intervention throughout the ML pipeline for proactive decision-making. Monitoring for privacy and fairness issues and their potential to harm the community throughout the software's life is an essential part of this. When evaluating the

fairness of a model, a researcher can explore the model's training data and performance metrics to decipher sub-trends and anomalies. From this evaluation, the researcher can generate an idea of what success can look like from their model.

It might also be helpful to pivot directions for the machine learning model to avoid going too far down a path that could prove disastrous for marginalized communities. There may be a point at which the model is beyond recognition. It may be worth completely re-imagining the ML pipeline or abandoning the effort altogether when it has strayed far from its intended goal. Before completely re-imaging or abandoning the model, the researcher could integrate human-*over*-the-loop techniques to improve the ML pipeline's consideration for privacy and fairness. The decision of which route to go ultimately involves the researcher evaluating the trade-off between the safety of the impacted communities or the potential accomplishments of producing innovative software. Success should be inspired by the ability to impact society positively, not by a system's ability to quickly solve an idea. Halting deployment on a project that has gone awry should be seen as a successful learning result, not as a failed project. If permissible, the stakeholder should consider opening up the research project or system for external review to cultivate a meaningful conversation around learning from the harm that development and deployment could have caused.

Chapter 7

Conclusion

With the increasing popularity and advancements in SMNs and smart environments, there has been a significant challenge in protecting visual privacy, thereby necessitating a better understanding of the visual privacy implications in these environments. These concerns can arise intentionally or unintentionally from the individual, other entities in the environment, or a company. To address these challenges, it is necessary to understand visual privacy leakage in various domains, create viable strategies to aid in quantifying visual privacy risk, and design fair and privacy data collection processes and ML pipelines. In this dissertation, myrgued that visual privacy is a point of critical concern for SMNs and smart cities because of the growth, advancement, and visual data shared in these domains. In particular, I demonstrated that there is a need to visual privacy leakage and risk to provide mitigation strategies that consider subjectivity, methods to quantify privacy leakage, and the creation of privacy and fair visual privacy systems. The consideration of these needs will improve the individual, stakeholder, and researcher's understanding of visual privacy in SMNs and smart environments. The studies, methods, and algorithms provided by this dissertation explore, investigate, and propose visual privacy mitigation and interactive auditing strategies for SMNs and smart environments.

Chapters 3 and 4 of this dissertation aimed to identify visual privacy challenges for social

media networks and smart city environments. Based on qualitative and quantitative analysis of privacy-related experiences, concerns of social media users regarding visual content, and infrastructure privacy considerations of smart city stakeholders, it can be concluded that the development of visual privacy mitigation strategies should be introduced to reduce privacy leakage and dangers in these domains. The risks and dangers that could arise in these domains require individuals and stakeholders to understand how visual privacy leaks can affect them and those around them. The implementation of these visual privacy mitigation systems can be incorporated into the infrastructure of SMNs and smart environments. While these are useful findings, they are limited by the sampling size and collection method. Broadly, issues with research samples and selection can lead to skewed results and bias. These issues do not capture the spectrum of visual privacy perspectives and experiences.

Chapter 5 endeavors to explore visual privacy risk scoring methodologies as an initial step of visual privacy mitigation. With computer vision as the backbone of these scoring methods, I investigate the necessity to exploit visual features to understand visual content and a case for using visual features in privacy risk scoring. I explored visual privacy risk scoring methods by adapting an existing privacy scoring methodology, *VPScore*, and developing a visual privacy scoring method using visual features as components, *Vango*. The visual privacy risk scores used a quantitative privacy severity weight inspired by Section 3.2.2. The results show that TF-IDF, Golden Spiral, and Object Area Ratio approaches can be used as visual privacy risk score components. This chapter adds to the methods privacy risk scoring is applied across domains, focusing on visual privacy scoring methods with computer vision support. The findings of this research are limited to using visual datasets that do not contain annotations for object detection or that are not focused on privacy research, utilizing pre-trained object detection models, and applying existing class labels for the generalizability of privacy risk scoring methodologies.

Lastly, this dissertation asserts the need for responsible auditing systems in the ML pipeline to reduce privacy and fairness issues that can occur, as proposed in Chapter 6. This

chapter walks through privacy and fairness issues that can arise in visual privacy research and broadly extends the findings to other fields of ML research. Due to technologies and practices that can violate the privacy and fairness of individuals, this chapter proposed interactive auditing strategies in an effort to reduce privacy leakage and harm to individuals. The interactive auditing strategies are introduced with a *human-over-the-loop* framework to allow an engaged approach to bring accountability to the ML training and deployment phase. This chapter provides a high-level overview of considerations and the impact of visual privacy and fairness issues. It provides a comprehensive auditing pipeline indicating fairness and privacy issues to reduce these effects in visual privacy research. There may be possible limitations in this study since there are no full-proof techniques for ensuring that the software is exempt from producing harm. This chapter suggests an interactively auditable ML pipeline to help reduce the risks and implications of the ML pipeline, but these suggestions are only those of the author.

Appendix A

Publications & Bibliographic Notes

The majority of the content included in **Chapter 2** was originally published in the related work sections of the following publications: DeHart et al. 2020c, DeHart et al. 2021a, DeHart et al. 2021b.

Chapter 3 was originally published with the following citation: Jasmine DeHart, Makya Stell, and Christan Grant. Social media and the scourge of visual privacy. *Information*, 11 (2):57, 2020c

Chapter 4 was originally published with the following citation: Jasmine DeHart, Oluwasijibomi Ajisegiri, Greg Erhardt, Jamie Cleveland, Corey E. Baker, and Christan Grant. Becoming a smart city: A textual analysis of the us smart city finalists. *International Journal on Advances in Intelligent Systems*, 14(3 and 4):94–103, 2021a

Chapter 6 was originally published with the following citation: Jasmine DeHart, Chenguang Xu, Lisa Egede, and Christan Grant. Proposing an interactive audit pipeline for visual privacy research. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 1249–1255, 2021b. doi: 10.1109/BigData52589.2021.9671478

Appendix B

Terms and Definitions

B.1 Relevant Definitions

This section introduces preliminary concepts needed to understand this dissertation. This section will include definitions and examples of the terms included in this dissertation. Throughout this dissertation, I refer to these terms frequently.

Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF is a measure of how relevant a word is to a given document from a collection of documents. That is, for a term t that appears in an answer d among the set of all answers D. The term frequency (TF) is the frequency at a term t that appears among any term t in the answer (Equation B.1).

$$\mathrm{tf}(t, d) = \frac{\sum_{|t \in d|} 1}{\sum_{|t' \in d|} 1} \tag{B.1}$$

The document frequency is the number of times a term t appears across all answers ($d \in D$). The document frequency is given by Equation B.2.

$$\mathrm{df}(t, D) = \sum_{d \in D} \mathbb{1}(t \in D) \tag{B.2}$$

The inverse document frequency (IDF) is a factor that down weights terms that appear too often across all documents. These words are viewed as less important. The idf formula is given by Equation B.3.

$$\text{idf}(t, d) = 1 + \log \frac{1 + n}{\text{df}(t, D)} \tag{B.3}$$

The average importance of each term in the data set is calculated by the product of the inverse document frequency and average term frequency for each document the term appears as shown in Equation B.4.

$$\text{avg(TF-IDF)}(t, D) = \text{idf}(t, D) * \text{avg}_{t \in d}(\text{tf}(t, d)). \tag{B.4}$$

Word Tokenization.

Word tokenization is the process of splitting each sentence into smaller units. It takes a raw data string and converts it into useful data. These smaller units are referred to as *tokens*. Here is an example string of data that can be tokenized: "`What is visual privacy?`". Word tokenization is performed in order for the machine learning model to understand the data. The string is broken down into several parts by tokenizing into words: `what`, `is`, `visual`, `privacy`, `?`. Word tokenization helps the machine learning model understand each word individually and the word's functionality in a larger text. This also allows the machine learning model to count the frequencies of words as they appear in the document.

Lemmatization.

Lemmatization is the process of replacing words that contain prefixes and suffixes with their root word with the use of dictionaries. Lemmatization allows treating the list of words with different inflections or derivatives of meaning as the same word. For example, to lemmatize the words `images`, `image's`, and `images'` means to remove the suffixes `s`, `'s`, and `s'` to bring out the root word `image`. Lemmatization aids the machine learning model in understanding the context around the words from the document and helps the model approximate the

meaning of the sentences in the document.

Stemming.

Stemming is a rule-based process that reduces words to their root form. A word can be mapped to a stem by removing prefixes and suffixes even if the word is not a real word. A limitation of using stemming is shown with the word `troubled`, which is stemmed into the word `troubl`. With this method, we try to avoid over-stemming to preserve the meaning of the word. For example, to stem the words `likes`, `likely`, and `liking` means to remove the suffixes `s`, `ly`, and `ing` to bring out the root word `like`.

Alphanumeric characters and Stopword removal.

Stopwords are commonly used words like `the`, `is`, `a`, and `are`. These stopwords include prepositions, pronouns, and conjunctions and are removed based on words from the standard English language. *Alphanumeric characters* can consist of letters (from A to Z), symbols (e.g., +, -, ¡), and numbers (0 to 9). In this process, we remove the low-level information from the data to focus on information that is more insightful for themes and keywords. The removal of these words did not show a negative impact on the algorithms used.

B.2 Selected Private Items from COCO Labels

In Chapter 5, the private classes were selected from an existing dataset: MS COCO. From MS COCO, the classes used were the object categories. This section discusses the choices of private items for the scope of this dissertation. The private classes were divided into three categories: *public*, *moderate*, and *severe*. The lists below only cover *moderate* and *severe* privacy labels.

Moderate risk objects include public transportation, household items, or time-based items. This content might not provide an individual's exact location, place of residence, or contain any of their government-issued identification.

- airplane - public transportation vehicle that can give location information
- bus - public transportation vehicle that can give location information
- train - public transportation vehicle that can give location information
- truck - public transportation vehicle that can give location information
- boat - public transportation vehicle that can give location information
- backpack - a personal item that holds or contains documentation
- wine glass - a personal activity
- toilet - an item that is used privately
- tv - a household item that could let someone know your location or interests
- mouse - an item that is an accessory to a computer/laptop
- remote - an item that is connected to personal electronic devices
- keyboard - an item that is an accessory to a computer/laptop

- clock - an item that can let others know the time of an event or location

Severe privacy risk contains items that can contain or carry personally identifying information, personal devices or vehicles, or items that contain insight into a person's location and place of residence

- car - a personal vehicle that can contain license plates
- motorcycle - a personal vehicle that can contain license plates
- traffic light - an item that can contain street names and/or reveal a location
- stop sign - an item that can contain street names and/or reveal a location
- parking meter - an item that can contain street names and/or reveal a location
- handbag - a personal item that holds or contains private items
- suitcase - a personal item that can indicate traveling plans or location
- laptop - a personal device that can show revealing information
- cell phone - a personal device that can show revealing information

B.3 Acronyms

AI Artificial Intelligence

ANOVA Analysis of Variance

CNN Convolutional Neural Network

DPScorer Dichotomous Privacy Score

DSRC Dedicated Short Range Communication

DTN Delay Tolerant Network

FASt Fairness Forensics Auditing System

GSD Golden Spiral Distance

IDF Inverse Document Frequency

IOT Internet of Things

LDA Latent Dirichlet Allocation

ML Machine Learning

MS COCO Microsoft Common Objects in COntext

OAR Object Area Ratio

OIW Object Importance Weight

PCA Principal Component Analysis

SCAP Smart City Applications Platform

SMN Social Media Network

TF Term Frequency

TF-IDF Term Frequency-Inverse Document Frequency

TSP Traffic Signal Priority

Vango Visual Area, eNcoding, and Golden spiral Object distance

ViP Visual Privacy Auditor

VISPR VISual PRivacy Dataset

VPScorer Visual Privacy Score

YOLO You Only Look Once

Appendix C

Study Instruments

C.1 Social Media & Privacy Survey Interview Protocol

C.1.1 Survey 1: Privacy Attitudes and Perspectives

- Are you over 18 years old?
  - Yes
  - No
- Have you used social media in the past two months?
  - Yes
  - No
- Of what Social Media Networks (SMNs) do you consider yourself a frequent user?
  - Facebook
  - Snapchat
  - Instagram
  - Pinterest

  - Tumblr
  - Flickr
  - LinkedIn
  - Twitter
  - Reddit
  - Twitch
  - YouTube
  - No Social Media Accounts

- How many hours per week do you spend on social media networks?
  - 0 -10
  - 11- 20
  - 21 +

- What type of content do you usually post on social media?
  - Images
  - Videos
  - Text

- Do you post any of these types of images or videos on your SMNs? (Check all that apply)
  - Selfies
  - Scenery
  - Food
  - Animals

  - Art

- How would you define privacy?
- Would you define privacy the same for social media networks?
- If not, why?
- Personally identifying information is information that can be used to uniquely identify, contact, or locate a person.
  - Agree
  - Disagree
- Privacy leaks include any instance in which a transfer of personal identifying visual content is shared on Social Media Networks. Private visual content exposes intimate information that can be detrimental to your finances, personal life, and reputation.
  - Agree
  - Disagree
- Would you consider any of these images to have identifying information?
- As a typical user of Social Media Networks (SMNs), if you were to post these items (images and/or videos) on your social media page would you consider them a privacy leak? (Answer Yes or No for each option)
  - Baby Face
  - Credit Card
  - Driver's License
  - House Keys
  - Phone Number

  - Social Security Card
  - Passport
  - Birth Certificate
  - Personal Letters

- Drag and drop the following dangers in order of most threatening (most threatening ranked 1 and so on).
  - Burglary (Home invasions)
  - Kidnapping (Physically or digitally kidnapping by means of taking visual content and posting as their own)
  - Explicit Websites (Images or Videos exported to explicit sites)
  - Financial Threat (Imposes on credit, bank account, loans and etc)
  - Identity Theft (Impersonation, Fraudulent accounts)
  - Stalking (Closely follows everything you do/post on SMNs and/or in real life)
- Do you believe there are other dangers on Social Media Networks? If so, list them.
- What type of threat would these items fall under? - Location
  - Credit Cards
  - House Keys
  - Baby faces
  - Passport
  - Driver's License
  - Social Security Card
  - Passwords

  - Personal Letters

- What type of threat would these items fall under? - Identity
  - Credit Cards
  - House Keys
  - Baby faces
  - Passport
  - Driver's License
  - Social Security Card
  - Passwords
  - Personal Letters
- What type of threat would these items fall under? - Asset
  - Credit Cards
  - House Keys
  - Baby faces
  - Passport
  - Driver's License
  - Social Security Card
  - Passwords
  - Personal Letters
- Do you believe that conflict (e.g. bullying, domestic disputes) can increase the occurrence of privacy leaks?
  - Strongly agree

  - Agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Disagree
  - Strongly disagree

C.1.2 Survey 2: Twitter Users and Privacy

- Are you over 18 years old?
  - Yes
  - No
- Do you have a Twitter Account?
  - Yes
  - No
- Why haven't you created a Twitter Account?
- What made you create a Twitter Account?
- How many hours per week do you use Twitter?
  - 0 - 5 hours
  - 5 - 10 hours
  - 10 - 20 hours
  - Over 20 hours
- What is your definition of sensitive information?

- Which of the following images do you think contains sensitive information?
- Sensitive information is defined as data that should be guarded from unauthorized access and unwarranted disclosure in order to maintain the information security of an individual or organization.
  - I agree
  - I do not agree
- While using Twitter have you ever seen sensitive information that was publicly posted online?
  - Yes
  - No
- What are the hashtag(s), keyword(s), and/or username(s) where the sensitive information was publicly posted on Twitter?

C.2 Smart City Survey Interview Protocol

- What type of organization do you work for?
  - Local Government
  - State Government
  - Federal Government
  - Utility
  - Transportation Agency
  - Non-Profit
  - Live, Work, Play Community

  - Other Private Sector Business
  - Not Employed
  - Comments
- What best describes your role in that organization?
  - Business Owner
  - Elected Official
  - Transportation Engineer/Planner
  - Information Technology Staff
  - Senior/Policy Advisor
  - Public Works Staff
  - Other
  - Not Applicable
  - Operations or Property Manager
- About how many employees work for your organization?
- Is your organization involved in any Smart City/Community work?
  - Yes
  - No
- Are you familiar with the term "smart city" or "smart community"?
  - Yes
  - No
- How would you define a smart city/community?

- How would you define privacy?
- How worried are you about Smart City/Community applications impacting your privacy?
  - Not close to worrying
  - Slightly worried
  - Moderately worried
  - Very worried
  - Extremely worried
- Describe any additional applications that would be valuable for your organization.
- Do you have other comments regarding privacy concerns in a Smart City?
- Where are the locations you need to have pedestrian, bicycle, and e-scooter counting data? (Check all that apply)
  - Downtown/main shopping street
  - Busy intersection
  - Parks
  - Greenwalks
  - Bus/Train stations
  - Stadiums
  - Government buildings (courthouse, fire & rescue, town hall, etc.)
  - Hospitals and healthcare facilities
  - Other
  - Not Sure

- What are the biggest challenges or pain points associated with pedestrian, bicycle, and e-scooter counting data? (Check all that apply)
  - Cost
  - Accuracy
  - Frequency
  - Data Analysis
  - Other, please list:
  - Not Sure
- If using today, what do you value most about your current pedestrian, bicycle, e-scooter counting data? On a scale of 1-5, please rate these features.
  - Accuracy
  - Simplicity
  - Frequency
  - Privacy
  - Distinguishing transportation types
- How important is pedestrian, bicycle, and e-scooter privacy?
  - Not at all important
  - Could be important
  - Moderately important
  - Very important
  - Extremely important
- How important is providing electric vehicle charging stations in your community?

- Consider each of the potential Smart City/Community applications below. For each, please rank how valuable these applications could be to your organization.
  - Discovering trends related to how pedestrians move throughout a city or community
  - Getting vehicular traffic data in real time to redirect traffic if there is an accident or lane closure
  - Solutions that help direct pedestrian and vehicle traffic
  - Adjusting traffic signal timing in real-time
  - Detecting potholes, burnt out streetlights, clogged drainage inlets, and other maintenance issues
  - Directing people to local businesses and events
  - Providing severe air quality and severe weather alerts
  - Directing drivers to open parking spots
  - Managing e-scooter and bicycle traffic and parking
  - Making retail location decisions
  - Making residential or commercial development decisions
  - Sizing and designing infrastructure
- For each of the following situations, please describe the level of importance of local data collection to your community or organization.
  - Counting the number of people passing a public location using a technology that could not identify individuals (such as radar, a pressure pad, etc.).
  - Counting the number of people passing a public location using video image recognition technology, as long as only counts are stored. Images would not be stored, and would not be available to view.

  - Storing video images from a public location with faces and license plates obscured.
  - Live video of a public location that is not stored.
  - Live video of a public location that is stored for future use.
  - Video or data is shared with a police department.
  - Video or data is shared with a police department only with a court order.
  - Video or data is shared with the public under a Freedom of Information Act request.
  - Recording mobile phone information without the ability to identify individuals.
  - Recording mobile phone information with the ability to identify individuals.
  - Recording credit card transactions without the ability to identify individuals.
  - Recording mobile phone screen interactions without the ability to identify individuals.
  - Automated parking enforcement (violators are sent a ticket in the mail).

- Consider each of the following types of data that could be collected on every street corner in a Smart City/Community. For each, please indicate how valuable the appropriate storage is for such data.
  - Traffic Counts
  - Pedestrians counts on streets or passing retail business
  - Air Quality Monitoring
  - Noise
  - Maintenance issues (e.g., potholes, burnt out street lights, clogged storm sewer inlets)
  - Public Transit Arrivals
  - Public Transit Ridership

  - Parking availability and cost
  - E-scooter and bicycle traffic

- Consider each of the following types of data collection. For each, indicate the processes your organization currently uses. Check all that apply.
  - Traffic Counts
  - Pedestrians counts on streets or passing retail business
  - Air Quality Monitoring
  - Noise

- Describe how difficult it is to mount a camera enclosure on your community streetlights?
  - Extremely difficult
  - Somewhat difficult
  - Neither easy nor difficult
  - Somewhat easy
  - Extremely easy

Appendix D

Additional Data

D.1 Smart City Challenge Finalist Application Data

Table D.1: This table contains the city name, population, and requested technologies for each Smart City Challenge finalist. The list is in ascending order based on population.

City	**Population Size**	**Technology Requested**
Pittsburgh, PA	305,704	1. Smart Grid 2. Electric Vehicle Charging Station 3. Autonomous Vehicles 4. Connected Vehicles: DSRC 5. Autonomous home delivery 6. Use of Cellphone signals 7. Use of Cameras 8. Use of WiFi/Communications 9. Use of Sensors 10. Web Applications 11. Smart Traffic Signals

* **Table D.1 – Continued on the next page** *

City	Population Size	Technology Requested
San Francisco, CA	413,775	1. Bike/Pedestrian Counters 2. Electric Vehicle Charging Station 3. Electric bus 4. Autonomous Vehicles 5. Smart Parking 6. Autonomous home delivery 7. Use of Cameras 8. Use of WiFi/Communications 9. Use of Sensors 10. Web Applications 11. Smart Traffic Signals 12. Smart roadside lights

* **Table D.1 – Continued on the next page** *

City	Population Size	Technology Requested
Kansas City, MO	459,787	1. Smart Grid 2. Electric Vehicle Charging Station 3. Autonomous Vehicles 4. Smart Parking 5. Use of Cameras 6. Use of WiFi/Communications 7. Use of Sensors 8. Web Applications 9. Kiosks 10. Smart Traffic Signals 11. Smart roadside lights

* **Table D.1 – Continued on the next page** *

City	Population Size	Technology Requested
Portland, OR	583,776	1. Electric Vehicle Charging Station 2. Autonomous Vehicles 3. Connected Vehicles: DSRC 4. Use of GPS 5. Use of Cameras 6. Use of WiFi/Communications 7. Use of Sensors 8. Web Applications 9. Smart Traffic Signals

* **Table D.1 – Continued on the next page** *

City	Population Size	Technology Requested
Denver, CO	600,158	1. Electric Vehicle Charging Station 2. Electric bus 3. Connected Vehicles: DSRC 4. Traffic Management Centers 5. Use of Cellphone signals 6. Use of Cameras 7. Use of WiFi/Communications 8. Use of Sensors 9. Web Applications 10. Kiosks 11. Information screens for bus stops 12. Smart Traffic Signals 13. Road condition monitors

* **Table D.1 – Continued on the next page** *

City	Population Size	Technology Requested
Columbus, OH	787,033	1. Smart Grid 2. Electronic Signs 3. Bike Sharing 4. Electric Vehicle Charging Station 5. Autonomous Vehicles 6. Connected Vehicles: DSRC 7. Smart Parking 8. Universal smart access card 9. Use of GPS 10. Use of Cameras 11. Use of WiFi/Communications 12. Use of Sensors 13. Web Applications 14. Kiosks 15. Information screens for bus stops 16. Smart Traffic Signals

* **Table D.1 – Continued on the next page** *

City	**Population Size**	**Technology Requested**
Austin, TX	790,390	1. Bike/Pedestrian Counters 2. Travel Hub 3. Electric Vehicle Charging Station 4. Autonomous Vehicles 5. Connected Vehicles: DSRC 6. Autonomous home delivery 7. Use of Cellphone signals 8. Use of GPS 9. Use of Cameras 10. Use of WiFi/Communications 11. Use of Sensors 12. Web Applications 13. Interactive Voice Response 14. Smart Traffic Signals 15. Road condition monitors

D.2 Mean (std) of Group differences for Danger Assessments

Table D.2: Mean and Standard deviations of sex-related danger assessment results. The table shows the mean (standard deviations) for Female and Male

Gender	Burglary	Kidnapping	Explicit websites	Financial Theft	Identity Theft	Stalking
Female	3.125 (1.699)	2.625 (1.809)	4.593 (1.623)	3.812 (1.281)	3.500 (1.606)	3.343 (1.658)
Male	2.52 (1.358)	2.00 (1.551)	5.40 (0.903)	3.58 (1.310)	3.40 (1.428)	4.10 (1.297)

Table D.3: Mean and Standard deviations of age-related differences of danger assessment results. The table shows the mean (standard deviations) for the age groups: 18–25 & 26 and over.

Age	Burglary	Kidnapping	Explicit websites	Financial Theft	Identity Theft	Stalking
18–25	2.708 (1.505)	2.236 (1.682)	5.069 (1.292)	3.694 (1.296)	3.513 (1.482)	3.777 (1.493)
26 and over	3.00 (1.699)	2.30 (1.702)	5.20 (1.316)	3.60 (1.429)	2.70 (1.337)	4.20 (1.316)

Bibliography

Shafii M Abdulhamid, Sulaiman Ahmad, Victor O Waziri, and Fatima N Jibril. Privacy and national security issues in social networks: the challenges. *arXiv preprint arXiv:1402.3301*, 2014.

Erfan Aghasian. *A privacy-based mechanism for users' information scoring and anonymisation across multiple online social networks.* PhD thesis, University of Tasmania, 2019.

Erfan Aghasian, Saurabh Garg, Longxiang Gao, Shui Yu, and James Montgomery. Scoring users' privacy disclosure across multiple online social networks. *IEEE access*, 5:13118–13130, 2017.

Erfan Aghasian, Saurabh Garg, and James Montgomery. An automated model to score the privacy of unstructured information—social media case. *Computers & Security*, 92:101778, 2020.

Cuneyt Akcora, Barbara Carminati, and Elena Ferrari. Privacy in social networks: How risky is your social graph? In *2012 IEEE 28th International Conference on Data Engineering*, pages 9–19. IEEE, 2012.

Mohammed Al-Husainy and Bassam Al-Shargabi. Secure and lightweight encryption model for iot surveillance camera. *International Journal of Advanced Trends in Computer Science and Engineering*, 9:1840–1847, 04 2020. doi: 10.30534/ijatcse/2020/143922020.

Fadi Al-Turjman, Hadi Zahmatkesh, and Ramiz Shahroze. An overview of security and privacy in smart cities' iot communications. *Transactions on Emerging Telecommunications Technologies*, 33(3):e3677, 2019.

Md Eshrat E Alahi, Fowzia Akhter, Anindya Nag, Nasrin Afsarimanesh, and Subhas Mukhopadhyay. Internet of things (iot)-enabled pedestrian counting in a smart city. In *Proceedings of International Conference on Computational Intelligence and Computing: ICCIC 2020*, pages 89–104. Springer, 2022.

José Alemany, E Del Val, J Alberola, and Ana García-Fornes. Estimation of privacy risk through centrality metrics. *Future Generation Computer Systems*, 82:63–76, 2018.

Shaukat Ali, S Anwar, and S Solehria. User interaction based framework for protecting user privacy in online social networks. *Proceedings of the ICISO*, 2013.

Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. Power to the people: The role of humans in interactive machine learning. *Ai Magazine*, 35(4):105–120, 2014.

Saleema Amershi, Max Chickering, Steven M Drucker, Bongshin Lee, Patrice Simard, and Jina Suh. Modeltracker: Redesigning performance analysis tools for machine learning. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 337–346, 2015.

V.V. Arlazarov, K. Bulatov, T. Chernov, and V.L. Arlazarov. Midv-500: a dataset for identity document analysis and recognition on mobile devices in video stream. *Computer Optics*, 43(5):818–824, Oct 2019. ISSN 0134-2452. doi: 10.18287/2412-6179-2019-43-5-818-824. URL `http://dx.doi.org/10.18287/2412-6179-2019-43-5-818-824`.

Mudassar Aslam, Muhammad Abbas Khan Abbasi, Tauqeer Khalid, Rafi us Shan, Subhan Ullah, Tahir Ahmad, Saqib Saeed, Dina A Alabbad, and Rizwan Ahmad. Getting smarter about smart cities: Improving data security and privacy through compliance. *Sensors*, 22 (23):9338, 2022.

Brooke Auxier and Monica Anderson. Social media use in 2021. *Pew Research Center*, 1: 1–4, 2021.

Mojdeh Azad, Nima Hoseinzadeh, Candace Brakewood, Christopher R Cherry, and Lee D Han. Fully autonomous buses: A literature review and future research directions. *Journal of Advanced Transportation*, 2019, 2019.

Ishaq Azhar. Security, privacy and risks within smart cities: Literature review and development of a smart city interaction framework. *International Journal of Creative Research Thoughts (IJCRT)*, pages 2320–2882, 2020.

Corey E Baker, Allen Starke, Tanisha G Hill-Jarrett, and Janise McNair. In vivo evaluation of the secure opportunistic schemes middleware using a delay tolerant social network. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pages 2537–2542. IEEE, 2017.

Niels Bantilan. Themis-ml: A fairness-aware machine learning interface for end-to-end discrimination discovery and mitigation. *Journal of Technology in Human Services*, 36(1): 15–30, 2018.

Keerti Banweer, Austin Graham, Joe Ripberger, Nina Cesare, Elaine Nsoesie, and Christan Grant. Multi-stage collaborative filtering for tweet geolocation. In *Proceedings of the 2Nd ACM SIGSPATIAL Workshop on Recommendations for Location-based Services and Social Networks*, LocalRec'18, pages 4:1–4:4, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-6040-1. doi: 10.1145/3282825.3282831. URL `http://doi.acm.org/10.1145/3282825.3282831`.

Saqba Batool. *Exploring vulnerability among children and young people who experience online sexual victimisation*. PhD thesis, University of Central Lancashire, 2020.

Justin Lee Becker. *Measuring privacy risk in online social networks*. University of California, Davis, 2009.

Rachel KE Bellamy, Kuntal Dey, Michael Hind, Samuel C Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, Aleksandra Mojsilović, et al. Ai fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, 63(4/5):4–1, 2019.

Benjamin Bengfort, Rebecca Bilbro, Nathan Danielsen, Larry Gray, Kristen McIntyre, Prema Roman, Zijie Poh, et al. Yellowbrick, 2018. URL `http://www.scikit-yb.org/en/latest/`.

Abeba Birhane, Vinay Uday Prabhu, and Emmanuel Kahembwe. Multimodal datasets: misogyny, pornography, and malignant stereotypes. *arXiv preprint arXiv:2110.01963*, 2021.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

Jared Bond, Christan Grant, Joshua Imbriani, and Erik Holbrook. A framework for interactive t-sne clustering. *International Journal of Software & Informatics*, 10(3), 2016.

Terrance Edward Boult. Pico: Privacy through invertible cryptographic obscuration. In *Computer Vision for Interactive and Intelligent Environment (CVIIE'05)*, pages 27–38. IEEE, 2005.

Danah Boyd. *It's complicated: The social lives of networked teens*. Yale University Press, 2014.

Danah Boyd and Alice E Marwick. Social privacy in networked publics: Teens' attitudes, practices, and strategies. 2011.

Alec Brandon, Christopher M Clapp, John A List, Robert Metcalfe, and Michael Price. Smart tech, dumb humans: The perils of scaling household technologies. *Work*, 2021.

Daniel Buschek, Moritz Bader, Emanuel von Zezschwitz, and Alexander De Luca. Automatic privacy classification of personal photos. In *IFIP Conference on Human-Computer Interaction*, pages 428–435. Springer, 2015.

Roy Cabaniss, Srinivasa S Vulli, and Sanjay Madria. Social group detection based routing in delay tolerant networks. *Wireless networks*, 19(8):1979–1993, 2013.

Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

Aylin Caliskan Islam, Jonathan Walsh, and Rachel Greenstadt. Privacy detective: Detecting private information and collective privacy behavior in a large social network. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, WPES '14, page 35–46, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450331487. doi: 10.1145/2665943.2665958. URL https://doi.org/10.1145/2665943.2665958.

Flavio P Calmon, Dennis Wei, Bhanukiran Vinzamuri, Karthikeyan Natesan Ramamurthy, and Kush R Varshney. Optimized pre-processing for discrimination prevention. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 3995–4004, 2017.

Mark Campbell, Magnus Egerstedt, Jonathan P How, and Richard M Murray. Autonomous driving in urban environments: approaches, lessons and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928): 4649–4672, 2010.

Yuxin Chen, Huiying Li, Shan-Yuan Teng, and Steven Nagels Zhijing Li. Wearable microphone jamming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.

Zhang Chen, Thivya Kandappu, and Vigneshwaran Subbaraju. Privattnet: Predicting privacy risks in images using visual attention. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 10327–10334, 2021. doi: 10.1109/ICPR48806.2021.9412925.

James Cleveland, Gregory S Tribbe, Louis Lombardi, Gilbert DeFreitas, and Peter Henderson. Pole-mountable camera support structure, November 24 2020. US Patent App. 29/713,374.

Onder Coban, Ali Inan, and Selma Ayse Ozel. Inverse document frequency-based sensitivity scoring for privacy analysis. *Signal, Image and Video Processing*, 16(3):735–743, 2022.

César R Cortez and Victor M Larios. Digital interactive kiosks interfaces for the gdl smart city pilot project. 2015.

Paolo Costa, Celicia Mascolo, Mirco Musolesi, and Gian Pietro Picco. Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 26(5):748–760, 2008. doi: 10.1109/JSAC.2008.080602.

Elizabeth M Daly and Mads Haahr. Social network analysis for information flow in disconnected delay-tolerant manets. *Mobile Computing, IEEE Transactions on*, 8(5):606–621, 2009.

David Danks and Alex John London. Algorithmic bias in autonomous systems. In *IJCAI*, volume 17, pages 4691–4697, 2017.

Pasquale De Luca. Detecting private information in large social network using mixed machine learning techniques. 2019.

Jasmine DeHart and Christan Grant. Visual content privacy leaks on social media networks. *arXiv preprint arXiv:1806.08471*, 2018.

Jasmine DeHart, Corey E Baker, and C Grant. Considerations for designing private and inexpensive smart cities. In *ICWMC 2020-2020 IARIA The Sixteenth International Conference on Wireless and Mobile Communications (ICWMC)*, pages 30–33. IARIA, 2020a.

Jasmine DeHart, Corey E Baker, and Christan Grant. Considerations for designing private and inexpensive smart cities. *The Sixteenth International Conference on Wireless and Mobile Communications*, 2020b.

Jasmine DeHart, Makya Stell, and Christan Grant. Social media and the scourge of visual privacy. *Information*, 11(2):57, 2020c.

Jasmine DeHart, Oluwasijibomi Ajisegiri, Greg Erhardt, Jamie Cleveland, Corey E. Baker, and Christan Grant. Becoming a smart city: A textual analysis of the us smart city finalists. *International Journal on Advances in Intelligent Systems*, 14(3 and 4):94–103, 2021a.

Jasmine DeHart, Chenguang Xu, Lisa Egede, and Christan Grant. Proposing an interactive audit pipeline for visual privacy research. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 1249–1255, 2021b. doi: 10.1109/BigData52589.2021.9671478.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009.

Hannah Devlin. AI systems claiming to 'read' emotions pose discrimination risks. *The Guardian*, February 2020. ISSN 0261-3077. URL `https://www.theguardian.com/technology/2020/feb/16/ai-systems-claiming-to-read-emotions-pose-discrimination-risks`. last accessed on 09/01/2020.

Mariella Dimiccoli, Juan Marín, and Edison Thomaz. Mitigating bystander privacy concerns in egocentric activity recognition with deep learning and intentional image degradation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1: 1 – 18, 2017.

Cory Doctorow. The case for ... cities that aren't dystopian surveillance states. *The Guardian*, January 2020. ISSN 0261-3077. URL `https://www.theguardian.com/cities/2020/jan/17/the-case-for-cities-where-youre-the-sensor-not-the-thing-being-sensed`. last accessed on 09/01/2020.

U.S. DOT. Smart city challenge, Jan 2015. URL `https://www.transportation.gov/smartcity`. last accessed on 09/01/2020.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 214–226, 2012.

Adel S Elmaghraby and Michael M Losavio. Cyber security challenges in smart cities: Safety, security and privacy. *Journal of advanced research*, 5(4):491–497, 2014.

Abrar Fahim, Mehedi Hasan, and Muhtasim Alam Chowdhury. Smart parking systems: Comprehensive review based on various aspects. *Heliyon*, 7(5):e07050, 2021.

Jerry Fails and Dan Olsen. A design tool for camera-based interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 449–456, 2003.

Jerry Alan Fails and Dan R Olsen Jr. Interactive machine learning. In *Proceedings of the 8th International Conference on Intelligent User interfaces*, pages 39–45, 2003.

Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 259–268, 2015.

Mathieu Fenniak. Home page for the pypdf2 project, 12 2013. URL `https://github.com/mstamy2/PyPDF2`.

Vittorio Ferrari, Jordi Pont-Tuset, Alina Kuznetsova, and Ashlesha Sadras. Open images v7 dataset, October 10 2022.

Sorelle A Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. The (im) possibility of fairness: Different value systems require different mechanisms for fair decision making. *Communications of the ACM*, 64(4):136–143, 2021.

Sainyam Galhotra, Yuriy Brun, and Alexandra Meliou. Fairness testing: testing software for discrimination. In *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering*, pages 498–510, 2017.

Wang Gang, Wang Shigang, Liu Cai, and Zhang Xiaorong. Research and realization on improved manet distance broadcast algorithm based on percolation theory. In *2012 International Conference on Industrial Control and Electronics Engineering (ICICEE)*, pages 96–99. IEEE, 2012.

Michael Garrow and Randy Machemehl. Development and evaluation of transit signal priority strategies. *Journal of Public Transportation*, 2(2):65–90, 1999.

Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

Austin Graham, Yan Liang, Le Gruenwald, and Christan Grant. Formalizing interruptible algorithms for human over-the-loop analytics. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 4378–4383. IEEE, 2017.

Tyrone WA Grandison, Sherry Guo, Kun Liu, Michael Maxmilien, Dwayne L Richardson, and Tony Sun. Providing and managing privacy scores, July 11 2017. US Patent 9,704,203.

Christopher D Green. All that glitters: A review of psychological research on the aesthetics of the golden section. *Perception*, 24(8):937–968, 1995.

Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, pages 71–80. ACM, 2005.

Amit Kr Gupta, Jyotsna Kumar Mandal, and Indrajit Bhattacharya. Comparative performance analysis of dtn routing protocols in multiple post-disaster situations. In *Contemporary Advances in Innovative and Applicable Information Technology*, pages 199–209. Springer, 2019.

Danna Gurari, Qing Li, Chi Lin, Yinan Zhao, Anhong Guo, Abigale Stangl, and Jeffrey P. Bigham. Vizwiz-priv: A dataset for recognizing the presence and purpose of private visual information in images taken by blind people. In *CVPR*, 2019.

Anisa Halimi and Erman Ayday. Real-time privacy risk quantification in online social networks. ASONAM '21, page 74–81, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391283. doi: 10.1145/3487351.3488272. URL `https://doi.org/10.1145/3487351.3488272`.

Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 3323–3331, 2016.

Adam Harvey. Cv dazzle: Camouflage from computer vision. *Technical report*, 2012.

Rakibul Hasan, David J. Crandall, Mario Fritz, and Apu Kapadia. Automatically detecting bystanders in photos to reduce privacy risks. *2020 IEEE Symposium on Security and Privacy (SP)*, pages 318–335, 2020.

Thomas Hellström, Virginia Dignum, and Suna Bensch. Bias in machine learning-what is it good for? In *International Workshop on New Foundations for Human-Centered AI (NeHuAI) co-located with 24th European Conference on Artificial Intelligence (ECAI 2020), Virtual (Santiago de Compostela, Spain), September 4, 2020*, pages 3–10. RWTH Aachen University, 2020.

Lisa Anne Hendricks, Kaylee Burns, Kate Saenko, Trevor Darrell, and Anna Rohrbach. Women also snowboard: Overcoming bias in captioning models. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 771–787, 2018.

NB Hounsell and J Wu. Public transport priority in real time traffic control systems. 1995.

Roberto Hoyle, Robert Templeman, Denise Anthony, David Crandall, and Apu Kapadia. Sensitive lifelogs: A privacy analysis of photos from wearable cameras. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1645–1648. ACM, 2015.

Wei-jen Hsu, Debojyoti Dutta, and Ahmed Helmy. Csi: A paradigm for behavior-oriented profile-cast services in mobile networks. *Ad Hoc Networks*, 10(8):1586–1602, 2012.

Pan Hui, Jon Crowcroft, and Eiko Yoneki. Bubble rap: Social-based forwarding in delay-tolerant networks. *IEEE Transactions on Mobile Computing*, 10(11):1576–1589, 2011.

Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, NanoCode012, Yonghye Kwon, Kalen Michael, TaoXie, Jiacong Fang, imyhxy, Lorna, Zeng Yifu, Colin Wong, Abhiram V, Diego Montes, Zhiqiang Wang, Cristi Fati, Jebastin Nadar, Laughing, UnglvKitDe, Victor Sonck, tkianai, yxNONG, Piotr Skalski, Adam Hogan, Dhruv Nair, Max Strobel, and Mrinal Jain. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation, November 2022. URL `https://doi.org/10.5281/zenodo.7347926`.

Maritza Johnson, Serge Egelman, and Steven M Bellovin. Facebook and privacy: it's complicated. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 9. ACM, 2012.

Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 1972.

Graham Kalton and Daniel Kasprzyk. The treatment of missing survey data. *Survey methodology*, 12(1):1–16, 1986.

Sathish K. Katukuri. *Viewpoint Recommendation for Aesthetic Photography*. PhD thesis, 2019. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2023-03-06.

Ari Keränen, Jörg Ott, and Teemu Kärkkäinen. The one simulator for dtn protocol evaluation. In *Proceedings of the 2nd international conference on simulation tools and techniques*, page 55. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009.

Y Kilic and A Inan. Privacy scoring over professional osns: More central users are under higher risk. pages 157–161, 2019.

Bart P Knijnenburg. Privacy? i can't even! making a case for user-tailored privacy. *IEEE Security & Privacy*, 15(4):62–67, 2017.

Mohammed Korayem, Robert Templeman, Dennis Chen, David Crandall, and Apu Kapadia. Enhancing lifelogging privacy by detecting screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4309–4314. ACM, 2016.

Iryna Korshunova, Wenzhe Shi, Joni Dambre, and Lucas Theis. Fast face-swap using convolutional neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3677–3685, 2017.

Balachander Krishnamurthy and Craig E Wills. Characterizing privacy in online social networks. In *Proceedings of the First Workshop on Online Social Networks*, pages 37–42. ACM, 2008.

Vlad Krotov and Leiser Silva. Legality and ethics of web scraping. *Twenty-fourth Americas Conference on Information Systems*, 2018.

Zhenzhong Kuang, Zongmin Li, Dan Lin, and Jianping Fan. Automatic privacy prediction to accelerate social image sharing. *2017 IEEE Third International Conference on Multimedia Big Data (BigMM)*, pages 197–200, 2017.

Anil Kunchala, Mélanie Bouroche, and Bianca Schoen-Phelan. Towards a framework for privacy-preserving pedestrian analysis. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4370–4380, January 2023.

Alexey Kurakin, Ian Goodfellow, and Samy Bengio. Adversarial machine learning at scale. *arXiv preprint arXiv:1611.01236*, 2016.

Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, et al. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *International Journal of Computer Vision*, 128(7):1956–1981, 2020.

Doris Jung Lin Lee, Stephen Macke, Doris Xin, Angela Lee, Silu Huang, and Aditya G. Parameswaran. A human-in-the-loop perspective on automl: Milestones and the road ahead. *IEEE Data Eng. Bull.*, 42(2):59–70, 2019.

F. Li, Zhe Sun, Ang Li, Ben Niu, H. Li, and G. Cao. Hideme: Privacy-preserving photo sharing on social networks. *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pages 154–162, 2019.

Hao Li, Tianhao Xiezhang, Cheng Yang, Lianbing Deng, and Peng Yi. Secure video surveillance framework in smart city. *Sensors*, 21(13):4419, 2021a.

Xuan Li, Dehua Li, Zhi Yang, and Weiwei Chen. A patch-based saliency detection method for assessing the visual privacy levels of objects in photos. *IEEE Access*, 5:24332–24343, 2017a.

Xuefeng Li, Yixian Yang, Yuling Chen, and Xinxin Niu. A privacy measurement framework for multiple online social networks against social identity linkage. *Applied Sciences*, 8(10): 1790, 2018.

Xuefeng Li, Yang Xin, Chensu Zhao, Yixian Yang, Shoushan Luo, and Yuling Chen. Using user behavior to measure privacy on online social networks. *IEEE Access*, 8:108387–108401, 2020.

XueFeng Li, Chensu Zhao, and Keke Tian. Privacy measurement method using a graph structure on online social networks. *Etri Journal*, 43(5):812–824, 2021b.

Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. Blur vs. block: Investigating the effectiveness of privacy-enhancing obfuscation for images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1343–1351. IEEE, 2017b.

Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):67, 2017c.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer, 2014.

Anders Lindgren, Avri Doria, and Olov Schelen. Probabilistic routing in intermittently connected networks. In *Service Assurance with Partial and Intermittent Resources*, pages 239–254. Springer, 2004.

Kun Liu and Evimaria Terzi. A framework for computing the privacy scores of users in online social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(1):6, 2010.

Shixia Liu, Xiting Wang, Mengchen Liu, and Jun Zhu. Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics*, 1(1):48–56, 2017.

Edward Loper and Steven Bird. Nltk: the natural language toolkit. *arXiv preprint cs/0205028*, 2002.

Grigorios Loukides and Aris Gkoulalas-Divanis. Privacy challenges and solutions in the social web. *XRDS: Crossroads, The ACM Magazine for Students*, 16(2):14–18, 2009.

May O Lwin, Andrea JS Stanaland, and Anthony D Miyazaki. Protecting children's privacy online: How parental mediation strategies affect website safeguard effectiveness. *Journal of Retailing*, 84(2):205–217, 2008.

James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

Oluwashina Madamori, Esther Max-Onakpoya, Christan Grant, and Corey Baker. Using delay tolerant networks as a backbone for low-cost smart cities. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 468–471. IEEE, 2019.

Michelle Madejski, Maritza Lupe Johnson, and Steven Michael Bellovin. The failure of online social network privacy settings. 2011.

Sachit Mahajan, Ling-Jyh Chen, and Tzu-Chieh Tsai. Swapitup: A face swap application for privacy protection. In *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, pages 46–50. IEEE, 2017.

Ola Malkawi, Nadim Obaid, and Wesam Almobaideen. Toward an ontological cyberattack framework to secure smart cities with machine learning support. *International Journal of Advanced Computer Science and Applications*, 13(11), 2022.

Moreno Mancosu and Federico Vegetti. What you can scrape and what is right to scrape: A proposal for a tool to collect public facebook data. *Social Media+ Society*, 6(3): 2056305120940703, 2020.

Antoni Martinez-Balleste, Pablo A. Perez-martinez, and Agusti Solanas. The pursuit of citizens' privacy: a privacy-aware smart city is possible. *IEEE Communications Magazine*, 51(6):136–141, 2013. doi: 10.1109/MCOM.2013.6525606.

Alessandra Mazzia, Kristen LeFevre, and Eytan Adar. The pviz comprehension tool for social network privacy settings. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 13. ACM, 2012.

George A Miller. *WordNet: An electronic lexical database*. MIT press, 1998.

Jack Morse. There's a privacy bracelet that jams smart speakers and, hell yeah, bring it. URL `https://mashable.com/article/bracelet-jams-alexa-smart-speakers/`. last accessed on 09/01/2020.

Mirco Musolesi and Cecilia Mascolo. Car: context-aware adaptive routing for delay-tolerant mobile networks. *IEEE Transactions on Mobile Computing*, 8(2):246–260, 2009.

Jan Neerbek. Sensitive information detection: Recursive neural networks for encoding context. *arXiv preprint arXiv:2008.10863*, 2020.

Ronald W Neperud and Kerry Freedman. Bases of children's visual preferences and discriminations. *Visual Arts Research*, pages 83–88, 1988.

Ashar Neyaz, Avinash Kumar, Sundar Krishnan, Jessica Placker, and Qingzhong Liu. Security, privacy and steganographic analysis of faceapp and tiktok. *International Journal of Computer Science and Security (IJCSS)*, 14(2):38, 2020.

U.S. Department of Transportation. Smart city challenge vision statements, 04 2016. URL `https://www.transportation.gov/smartcity/visionstatements/index`.

Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

Tribhuvanesh Orekondy, Mario Fritz, and Bernt Schiele. Connecting pixels to privacy and utility: Automatic redaction of private information in images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8466–8475, 2018.

José Ramón Padilla-López, Alexandros Andre Chaaraoui, and Francisco Flórez-Revuelta. Visual privacy protection methods: A survey. *Expert Systems with Applications*, 42(9): 4177–4195, 2015.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011a.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct): 2825–2830, 2011b.

Ruggero G Pensa and Gianpiero Di Blasi. A semi-supervised approach to measuring user privacy in online social networks. In *International Conference on Discovery Science*, pages 392–407. Springer, 2016.

Ruggero G. Pensa and Gianpiero Di Blasi. A centrality-based measure of user privacy in online social networks. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1438–1439, 2016. doi: 10.1109/ASONAM.2016.7752439.

Ruggero G Pensa, Gianpiero Di Blasi, and Livio Bioglio. Network-aware privacy risk estimation in online social networks. *Social Network Analysis and Mining*, 9(1):1–15, 2019.

A. J. Perez, S. Zeadally, and S. Griffith. Bystanders' privacy. *IT Professional*, 19(03):61–65, May 2017. ISSN 1941-045X. doi: 10.1109/MITP.2017.42.

Andreea Picu and Thrasyvoulos Spyropoulos. Dtn-meteo: Forecasting the performance of dtn protocols under heterogeneous mobility. *IEEE/ACM Transactions on Networking*, 23 (2):587–602, 2014.

Martin F Porter. An algorithm for suffix stripping. *Program*, 1980.

Luciano M Prevedello, Safwan S Halabi, George Shih, Carol C Wu, Marc D Kohli, Falgun H Chokshi, Bradley J Erickson, Jayashree Kalpathy-Cramer, Katherine P Andriole, and Adam E Flanders. Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions. *Radiology: Artificial Intelligence*, 1 (1):e180031, 2019.

Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. Saving face: Investigating the ethical concerns of facial recognition auditing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 145–151, 2020.

Rajeev Ranjan, Swami Sankaranarayanan, Carlos D Castillo, and Rama Chellappa. An all-in-one convolutional neural network for face analysis. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 17–24. IEEE, 2017.

Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

David Rosenblum. What anyone can know: The privacy risks of social networking sites. *IEEE Security & Privacy*, 5(3):40–49, 2007.

Candace Ross, Boris Katz, and Andrei Barbu. Measuring social biases in grounded vision and language embeddings. *arXiv preprint arXiv:2002.08911*, 2020.

Ruben Sánchez-Corcuera, Adrián Nuñez-Marcos, Jesus Sesma-Solance, Aritz Bilbao-Jayo, Rubén Mulero, Unai Zulaika, Gorka Azkune, and Aitor Almeida. Smart cities survey: Technologies, application domains and challenges for the cities of the future. *International Journal of Distributed Sensor Networks*, 15(6):1550147719853984, 2019.

P Saraswathi. The golden proportion and its application to the human face. *European Journal of Anatomy*, 11(3):177, 2007.

Dipanjan Sarkar. *Text analytics with Python: a practitioner's guide to natural language processing*. APress, 2019.

Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st International Conference on Distributed Computing Systems Workshop*, pages 166–171. IEEE, 2011.

Nora Cate Schaeffer and Stanley Presser. The science of asking questions. *Annual Review of Sociology*, 29(1):65–88, 2003.

Shahbaz Siddiqui, Sufian Hameed, Syed Attique Shah, Abdul Kareem Khan, and Adel Aneiba. Smart contract-based security architecture for collaborative services in municipal smart cities. *Journal of Systems Architecture*, 135:102802, 2023.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Harriet R Smith, Brendon Hemily, and Miomir Ivanovic. Transit signal priority (tsp): A planning and implementation handbook. 2005.

Joshuas Emerson Smith. As San Diego increases use of streetlamp cameras, ACLU raises surveillance concerns, August 2019. URL `https://lat.ms/33AzG7I`. last accessed on 09/01/2020.

Anna C Squicciarini, Cornelia Caragea, and Rahul Balakavi. Analyzing images' privacy for the modern web. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 136–147. ACM, 2014.

Agrima Srivastava and G Geethakumari. Measuring privacy leaks in online social networks. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2095–2100. IEEE, 2013.

Jose M. Such, Joel Porter, Sören Preibusch, and Adam Joinson. Photo privacy conflicts in social media: A large-scale empirical study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 3821–3832, New York, NY, USA, 2017a. Association for Computing Machinery. ISBN 9781450346559. doi: 10.1145/3025453.3025668. URL `https://doi.org/10.1145/3025453.3025668`.

Jose M Such, Joel Porter, Sören Preibusch, and Adam Joinson. Photo privacy conflicts in social media: A large-scale empirical study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3821–3832. ACM, 2017b.

Harini Suresh and John V Guttag. A framework for understanding unintended consequences of machine learning. *preprint arXiv:1901.10002*, 2019.

Joaquin Taverner, Ramon Ruiz, Elena del Val, Carlos Diez, and Jose Alemany. Image analysis for privacy assessment in social networks. In *Distributed Computing and Artificial Intelligence, Special Sessions II, 15th International Conference 15*, pages 1–4. Springer, 2020.

Kurt Thomas, Chris Grier, and David M Nicol. unfriendly: Multi-party privacy risks in social networks. In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 236–252. Springer, 2010.

Matt Tierney, Ian Spiro, Christoph Bregler, and Lakshminarayanan Subramanian. Cryptagram: Photo privacy for online social media. In *Proceedings of the First ACM Conference on Online Social Networks*, pages 75–88. ACM, 2013.

Kiyohito Tokuda. Dsrc-type communication system for realizing telematics services. *Oki Technical Review*, 71(2):64–67, 2004.

Ashwini Tonge and Cornelia Caragea. Image privacy prediction using deep neural networks. *ACM Transactions on the Web (TWEB)*, 14(2):1–32, 2020.

Ashwini Kishore Tonge and Cornelia Caragea. Image privacy prediction using deep features. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

A. Torralba, R. Fergus, and W. T. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1958–1970, 2008.

Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *CVPR 2011*, pages 1521–1528. IEEE, 2011.

Antonio Torralba, Rob Fergus, and Bill Freeman. 80 million tiny image dataset, Jun 2020. URL `https://groups.csail.mit.edu/vision/TinyImages/`.

Omar Cliff Uchani Gutierrez and Guangxia Xu. Blockchain and smart contracts to secure property transactions in smart cities. *Applied Sciences*, 13(1):66, 2023.

Liesbet Van Zoonen. Privacy concerns in smart cities. *Government Information Quarterly*, 33(3):472–480, 2016.

Maria Han Veiga and Carsten Eickhoff. Privacy leakage through innocent content sharing in online social networks. *arXiv preprint arXiv:1607.02714*, 2016.

Nishant Vishwamitra, Yifang Li, Hongxin Hu, Kelly Caine, Long Cheng, Ziming Zhao, and Gail-Joon Ahn. Towards automated content-based photo privacy control in user-centered social networks. In *Proceedings of the Twelveth ACM Conference on Data and Application Security and Privacy*, pages 65–76, 2022.

Emanuel von Zezschwitz, Sigrid Ebbinghaus, Heinrich Hussmann, and Alexander De Luca. You can't watch this!: Privacy-respectful photo browsing on smartphones. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4320–4324, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858120. URL `http://doi.acm.org/10.1145/2858036.2858120`.

Arif Wahyudi, Mirza Triyuna Putra, Dana Indra Sensuse, Sofian Lusa, Prasetyo Adi, Assaf Arief, et al. Measuring the effect of users' privacy concerns on the use of jakarta smart city mobile application (jaki). *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(6):1014–1020, 2022.

Qiaozhi Wang, Hao Xue, Fengjun Li, Dongwon Lee, and Bo Luo. # donttweetthis: Scoring private information in social networks. *Proceedings on Privacy Enhancing Technologies*, (4):72–92, 2019.

Zeyu Wang, Klint Qinami, Ioannis Christos Karakozis, Kyle Genova, Prem Nair, Kenji Hata, and Olga Russakovsky. Towards fairness in visual recognition: Effective strategies for bias mitigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8919–8928, 2020.

Xinzhou Wu, Sundar Subramanian, Ratul Guha, Robert G. White, Junyi Li, Kevin W. Lu, Anthony Bucceri, and Tao Zhang. Vehicular communications using dsrc: Challenges, enhancements, and evolution. *IEEE Journal on Selected Areas in Communications*, 31(9): 399–408, 2013. doi: 10.1109/JSAC.2013.SUP.0513036.

J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3485–3492, 2010.

Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics*, 2:67–78, 2014.

Xidong Yuan, Yean-Jye Lu, and Semaan Sarraf. A computer vision system for measurement of pedestrian volume. In *Proceedings of TENCON'93. IEEE Region 10 International Conference on Computers, Communications and Automation*, volume 2, pages 1046–1049. IEEE, 1993.

Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1171–1180, 2017.

Pe Zatelli, S Gobbi, C Tattoni, N La Porta, M Ciolli, et al. Object-based image analysis for historic maps classification. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:247–254, 2019.

David R Zemmels and David N Khey. Sharing of digital visual media: privacy concerns and trust among young people. *American Journal of Criminal Justice*, 40(2):285–302, 2015.

Sergej Zerr, Stefan Siersdorfer, and Jonathon Hare. Picalert!: A system for privacy-aware image classification and retrieval. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pages 2710–2712, New York, NY, USA, 2012a. ACM. ISBN 978-1-4503-1156-4. doi: 10.1145/2396761.2398735.

Sergej Zerr, Stefan Siersdorfer, and Jonathon Hare. Picalert!: a system for privacy-aware image classification and retrieval. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 2710–2712. ACM, 2012b.

Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, and Elena Demidova. Privacy-aware image classification and search. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 35–44. ACM, 2012c.

Chenye Zhao, Jasmine Mangat, Sujay Koujalgi, Anna Squicciarini, and Cornelia Caragea. Privacyalert: A dataset for image privacy prediction. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 1352–1361, 2022.

Qiang Alex Zhao and John T Stasko. The awareness-privacy trade-off in video supported informal awareness: A study of image-filtering based techniques. Technical report, Georgia Institute of Technology, 1998.

Tengfei Zheng, Tongqing Zhou, Qiang Liu, Kui Wu, and Zhiping Cai. Characterizing and detecting non-consensual photo sharing on social networks. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, CCS '22, page 3209–3222, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450394505. doi: 10.1145/3548606.3560571. URL `https://doi.org/10.1145/3548606.3560571`.

Haoti Zhong, Anna Squicciarini, and David Miller. Toward automated multiparty privacy conflict detection. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1811–1814. ACM, 2018.

Bingquan Zhu, Hao Fang, Yanan Sui, and Luming Li. Deepfakes for medical video de-identification: Privacy protection and diagnostic information preservation. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 414–420, 2020.

Michael Zimmer. "but the data is already public": on the ethics of research in facebook. *Ethics and information technology*, 12(4):313–325, 2010.

Michael Zimmer and Katharina Kinder-Kurlanda. *Internet research ethics for the social age: New challenges, cases, and contexts*. Peter Lang International Academic Publishers, 2017.

DEDICATION

to

I dedicate this dissertation to my family, my late grandmothers, Gucci Ramone, and, lastly, to *myself* – with pride, relief, and love. This is dedicated to my hometown, Saint Louis, Missouri, my high school, Jennings Senior High, and my Alma Mater, Philander Smith College. With a future dedication to all the little *Jasmine*'s that find themselves in the realm of Computer Science.

Hard work + Dedication = Success