# A Personal Privacy Risk Assessment Framework based on Disclosed PII

Ningning Wu
*Dept. of Information Science*
*University of Arkansas at*
*Little Rock*
Little Rock, AR, USA
nxwu@ualr.edu

Robinson Tamilselvan
*Dept. of Information Science*
*University of Arkansas at*
*Little Rock*
Little Rock, AR, USA
rtamilselvan@ualr.edu

***Abstract*—Protecting personal identifiable information (PII) is essential for personal privacy and data protection. The leakage of PII can lead to privacy and safety issues like personal embarrassment, workplace discrimination, and identity theft. Driven by privacy laws and regulations, business is becoming more diligent in privacy protection when handling PII. Individual users, on the other hand, are free to produce and share contents online that might contain sensitive information. This paper proposed a personal privacy risk assessment framework from user's perspective. The risk score would help PII owners assess their privacy risks so that they can be more actively control their information release and protect their privacy.**

***Keywords—PII, data privacy, personal privacy risk assessment***

## I. INTRODUCTION

PII or Personally Identifiable Information is any information that permits the identity of an individual to be directly or indirectly inferred, including any information that is linked or linkable to that individual. Examples of PII that directly identifies an individual include name, address, SSN, phone number, email address, biometrics, etc. Examples of PII that indirectly identifies an individual may include a combination of gender, race, birthdate, and other descriptors.

Guarding PII is very important to ensure the integrity of an individual's identity. With just a few bits of personal information, thieves can create false accounts in their name, start racking up debt, or even create a falsified passport and sell their identity to criminals. Indeed, with new advances in Web, wireless communication, IoT, and Cloud technologies, PII is much easier to access and share than before. The gathering of PII becomes a common practice by information technology companies, government, and organizations for all kinds of purposes, such as monitoring public health, analyzing consumer shopping behaviors and trends, predicting individual's political preferences, and so on. In addition, the wide use of smart devices and social media has contributed the broad scale PII gathering, some is with user's knowledge or permission and some not. As PII is collected anytime and anywhere, the unregulated dissemination and usage of PII can lead to serious privacy violations. When PII is getting into the wrong hands, it can result in devastating consequences such identity theft. Too many cases show that it is a huge hassle and extremely time-consuming for an identity theft victim to recover his true identity, not to mention the possible huge financial loss the victim may have suffered.

Protecting personal privacy is a collaborated effort that involves any entity that owns or handles PII. Data companies, organizations, and governments need abide by privacy laws and regulations when collecting, storing, and using PII. They should have appropriate security policies and mechanisms in place to control access to sensitive PII. PII owners should also be diligent in protecting their own information.

In the last two decades, social media becomes a popular platform for people to connect with friends and family members, make new friends, and to share their lives and experiences. When people share their thoughts and lives, their posts, pictures and videos might be minded and sold. In addition, users, when creating an account, might willingly relinquish some of their personal data such as name, birthdate, address, and so on. One privacy concern with the social media usage is oversharing, which happens when people share too much personal information[2,3]. Oversharing not only hurts a person's reputation if he says something inappropriate, but also leads serious security problems like identity theft and fraud.

According to a recent study from Viasat Savings[1], nearly 50% of social media users surveyed chose to keep their accounts open and public. People generally focus on breaches that involve passwords, but the social media data breach would compromise a significant amount of PII and thus have a more severe consequence. A study by Atlas VPN shows that 41% of all compromised records in 2021 originated from social media data leaks, which is a significant upsurge compared to 25% in 2020. One of the key factors that make social media becoming a primary target for data breach is users, whose oversharing make it simple for thieves to locate personal information. So PII owners should be vigilant in protecting their information.

It is estimated that in 2021 traditional identity fraud losses—those involving any use of a consumer's personal information to achieve illicit financial gain—amounted to $24 billion (USD) and ensnared 15 million U.S. consumers, and that losses involving identity fraud scams—involving direct contact with victims by criminals—totaled $28 billion and affected 27 million consumers in the United States[4]. Identity thieves typically use a combination of data from different sources to get the information needed to open credit cards, take out loans, and make erroneous purchases in the victim's name. Unfortunately, Cyberspace users often post and share information (texts, images, vlogs) that may contain private information. Moreover, there are more than abundant mobile apps and web applications that collect customer information (disclosed or undisclosed) through different channels. Some of that information is publicly accessible and searchable. On one hand, there is a need for effective regulation of those applications to collect and disseminate personal information. At the same time, there is a need for helping users: 1) to monitor what personal information has been collected, and 2) be provided with decision-making tools to help them identify information to be shared publicly; both in an effort to safeguard their privacy.

In the past decades, data breach incidents and illegal PII gathering practices by data companies have caused increasing concern over privacy violation, and thus had led to the enactment of several privacy laws in US, including the Health Insurance Portability and Accountability Act (HIPPA), the Fair Credit Report Act (FCRA), the Family Education Rights and Privacy Act (FERPA), the Gramm-Leach-Bliley Act (GLBA), the Electronic Communications Privacy Act (ECPA), the Children's Online Privacy Protection Act (COPPA), and the Video Privacy Protection Act (VPPA). Driven by the laws, companies are becoming more careful in protecting their data and restrictive in releasing personal information. Still personal information is spilled everywhere and can be found in public documents, government records, news outlets, individual's websites and social media accounts, etc.

In addition, there are people search companies that collect and sell personal information. These companies provide online people search services for free or a fee. When searching a person by name, the free service would show very limited information of a searched person, such as name, age, and relatives. The fee-based services provide more comprehensive information including credit report, property records, criminal and traffic records, and so on.

Our preliminary study on eight online people search companies showed that data companies are becoming more restrictive in releasing personal information. For instance, all companies in our study have a privacy policy that clearly states their practices of using and sharing PII and the privacy laws they abide by. They also have a user rights policy that allows user to opt out of the sale of their information to third parties. When searching a person's information by name, 7 companies just provide a profile preview that only discloses name and age information. Any users who want more information have to provide their identity information along with credit card in order to purchase the information.

While data companies are pushed to abide by federal laws and regulations in their practices, individual users, especially PII owners, on the other hand are free to produce contents that might contain sensitive information. The wide use of social media and online forums not only promotes data sharing and dissemination, but also increases the risks of privacy breach. If users don't have a clear understanding of privacy protection, they may disclose information that should not be revealed. Unfortunately once data is on the Internet, it is always on the Internet. The debate about what should happen to personal information published on the Internet is not a new one. The notion of "posting remorse" refers to the situation that things posted by a user can seem funny and harmless one day would come back and haunt him the next. Thus individual users become a weak link in personal privacy protection.

To protect individual user's privacy, we need to raise user's awareness about privacy protection so that they would be more cautious when sharing PII. In addition, users should actively monitor the dissemination of their information in the public domain, especially on the Internet. This research aims to design a privacy risk assessment framework from a person's perspective. The framework assesses an individual's privacy risk based on the amount of his/her PII information is disseminated publicly on Internet. The framework is part of an ongoing research project that aims to provide users with a tool that not only monitors their information on the Internet and assesses their privacy risks, but also allows them to proactively protect their private information.

## II. OUR RESEARCH

Internet becomes a major source for people to seek information. When googling a person's name, the amount of PII returned is alarming. This private information can be potentially used against the individual for multiple purposes. This research is motivated by the questions "how much a person's PII information can be found on Internet and how this information would impact on a person's privacy?"

The research considers personal privacy from owner's perspective. Personal privacy is different from data privacy in the business settings in that a person can only control what information they would like to release or share, but they have little control over how their information is collected, stored, released, and used by a third party. For the privacy protection purpose, this study only uses publicly accessible data.

### *A. PII information on Internet*

It would be interesting to see how much a person's information can be found on Internet. We developed an information retrieval framework that employs natural language processing and entity identification and resolution techniques to retrieve PII attributes in web documents. Given a name, the framework searches for the web documents containing the name, scrapes those documents, and then extracts PIIs.

We compiled a sample group of 50 persons that include authors, their friends and relatives, and random people. We searched these persons' information by name on Internet, and then applied our information retrieval framework to extract their information from the top 20 webpages. Figure 1 shows the PII attributes that are found from public documents along with the percentage of people have values on those attributes.

Identify applicable funding agency here. If none, delete this text box.

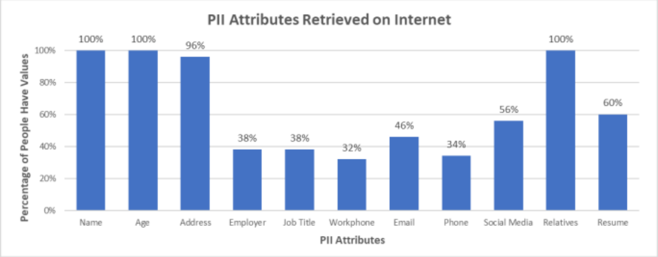

Figure 1. PII attributes found on Internet and the estimated percentage of people having values

### B. *Personal Privacy Risk Assessment Framework*

The personal privacy risk assessment framework aims to analyze the privacy risk of a person when some of his PII attributes are disclosed. PII attributes have different privacy levels. Some PII attributes are very sensitive such as SSN and biometric, while some are not such as a person's name and business email. We categorize PII into three groups: public PII, protected PII, and private PII.

**Public PII:** Public PII attributes are those that are intended for public use or can be obtained from public records such as phonebook, news outlets, public websites, social media, real estate transaction, etc. Most public PII attributes are used in everyday life and hard to keep private. They include person's name, address, personal and business phone numbers and emails, social media accounts, etc.

**Protected PII**: Protected PII attributes should be protected and not disclosed arbitrarily. These attributes are often required to verify a person's identity such as driver's license number and passport number. In addition, tax ID, employee ID, credit card numbers, bank account numbers, past addresses, date of birth, and birth place should be protected due to their sensitivity and common use for identity proof.

**Confidential PII:** These are the attributes that should be kept confidential. Data collection of confidential PII should be strictly regulated and controlled. The use of confidential PII in general applications should be prohibited. When storing or transmitting data that contains the information of confidential PII, data encryption should be used to protect their secrecy. Access to confidential PII should be strictly controlled. Ideally confidential PII should be stored separately from other PII so that disclosure of one would not lead to the total privacy loss.

PII attributes have different properties and their assumed usages. Some are an integral part of a person's identity such as person's names, birth place, birth date, and biometrics; others are related to person's life and activities in a society such as address, phone numbers, emails, social media accounts, bank accounts, etc. Some PII attributes never change, while others might change overtime and have a limited life time. Access to public PII attributes will not pose a direct threat to a person's privacy unless they are connected to non-public PII attributes. Similarly, when a confidential PII is revealed alone, it may not pose a severe privacy threat; however, if a confidential PII is disclosed along with other PII, then it could incur a big privacy risk.

The proposed framework considers two factors when assessing the privacy level of a PII attribute: the willingness of people giving out the PII's information, and the PII's power for resolving the owner's identity. The willingness of people giving out an PII information reflects their perception about the degree of the PII's privacy. When an PII is deemed of having high privacy, the owner would be reluctant to share it. In general people are assumed not willing to disclose the protected and confidential PII, so the willing measure for these PII are close to zero. The willingness measure of a PII could be estimated by the percentage of people disclosing it. Table 1 shows the willingness measures of some PIIs that are obtained from a sample of accounts on a popular social media platform.

TABLE 1: THE WILLINGNESS MEASURES OF PII ATTRIBUTES

| Attribute Name | Percentage |
|---|---|
| Name | 100% |
| Address | 10% |
| DoB | 83% |
| Personal Cell | 16% |
| Gender | 98% |
| employer | 50% |
| education | 80% |

A PII's power of resolving a person's identity is defined as its ability to uniquely identify the person. It is measured as the probability to uniquely identify a person based on the PII value, and it takes values between 0 and 1. If a PII has a unique value for each person, then its resolution power is 1; otherwise, its resolution power is less than 1. For example, the resolution power for SSN, biometrics, passport number, and driver's license number is 1, and the resolution power for birthdate and birth place could be much lower. The resolution power for a person's name depends on the uniqueness of the name. Indeed, human names carry deep personal, cultural, familial, and historical connection. Some names may also carry the information about the time and origin they are from. For example, in an English-speaking country, a typical popular English name may have low resolution power; however, a relatively rare and non-English name would have a much higher resolution power. So the resolution power of a name is determined by its uniqueness. The resolution power of a person's first name, middle name, and last name is assessed separately, and the highest probability will be a person's name resolution power.

Given a PII attribute $i$, its privacy score is defined as

$$s_i = \frac{1}{e^{-\beta_i(1-w_i)r_i}} \qquad (1)$$

Where $w_i$ and $r_i$ are the willingness measure and the resolution power of *i* respectively, and $\beta_i$ is a coefficient.

Figure 2 shows the privacy scores in terms of the willingness and resolution measures with $\beta_i = 1$. It shows that the more the willingness of an owner disclosing a PII, the less privacy score it is. Given the same willingness measure, a PII with high resolution power will have a high privacy score.

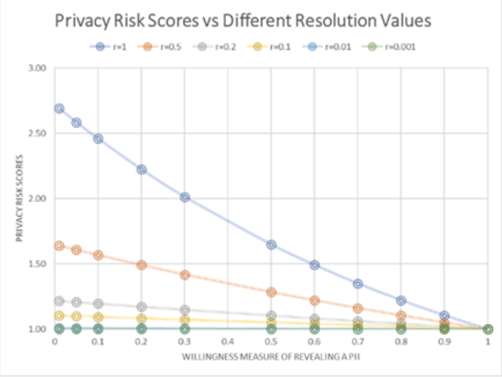

Figure 2: Privacy risk scores vs different resolution values

Personal information contains many attributes. When evaluating a person's privacy risk, the amount of personal information that is publicly accessible or searchable is used to calculate the risk score as follows:

$$R = \sum_i \alpha_i s_i \qquad (2)$$

where $s_i$ is the privacy score of attribute $i$; $\alpha_i$ is the weight of $i$, and it indicates the privacy degree of the PII. The weights for the confidential PIIs are the highest, and the weights for the public PIIs are the lowest.

To examine the risk scores of different scenarios, we created a synthetic dataset. The dataset is not intended to cover all PII attributes, rather it is used to study the effectiveness of privacy risk scores. The dataset contains eight public PIIs, five protected PIIs, and one confidential PIIs, as shown in the table below:

TABLE 2: THE PII ATTRIBUTES IN THE SYNTHETIC DATASET

| | |
|---|---|
| **Public PII** | name, address, personal phone number, email, business phone, social media accounts including facebook, twitter, and Instagram |
| **Protected PII** | DDL, passport number, credit card number, DoB, birth place |
| **Confidential PII** | SSN |

The attribute weights are assigned as follows:

TABLE 3: THE WEIGHT OF PII ATTRIBUTES

| | |
|---|---|
| Name | 1 |
| Address | 1 |
| DoB | 2 |
| Birth Place | 2 |
| Personal Cell | 0.5 |
| Email | 0.1 |
| business phone | 0.1 |
| Facebook account | 1 |
| Twitter account | 0.1 |
| Instagram account | 0.1 |
| DDL | 2 |
| Passport # | 2 |
| Credit card | 2 |
| SSN | 10 |

An attribute's weight is mainly determined by its privacy levels. The weights for confidential PII attributes and protected PII attributes are set as 10 and 2 respectively. The weights for public PII attributes vary depending on their abilities to identify a person.

The synthetic dataset enumerates all possible scenarios of the PII attributes being revealed, with 11 name resolution values. It has 180,224 scenarios in total. The purpose of the synthetic dataset is to simulate various scenarios that a person's PII attributes being revealed and to evaluate the corresponding privacy risk scores. Ideally the privacy risk scores should be able to distinguish the high-risk situations from low-risk ones. The following results are based on the scenarios with name resolution as 1. These scenarios are considered as the extreme cases when a person's name is unique, and their corresponding risk score values define the upper bound of the risk scores.

Table 4 shows the average, maximum, and minimum of risk scores by the number of disclosed public PII attributes, when none of the protected and confidential PII attributes is disclosed. It shows the minimum risk score is 0.1 when only one public PII attribute is revealed, and the maximum risk score is 6.87 when all eight public attributes are disclosed. Figure 3 shows the graph representation of table 4.

TABLE 4: THE AVERAGE, MAXIMUM, AND MINIMUM RISK SCORES BY THE NUMBER OF PUBLIC PII DISCLOSED

| # of public PII | Avg. Scores | Max. Scores | Min. Scores |
|---|---|---|---|
| 1 | 0.75 | 2.63 | 0.10 |
| 2 | 1.37 | 3.83 | 0.20 |
| 3 | 2.03 | 4.84 | 0.30 |
| 4 | 2.74 | 5.87 | 0.55 |
| 5 | 3.66 | 6.12 | 1.55 |
| 6 | 4.72 | 6.19 | 2.75 |
| 7 | 5.48 | 6.29 | 4.66 |
| 8 | 6.87 | 6.87 | 6.87 |

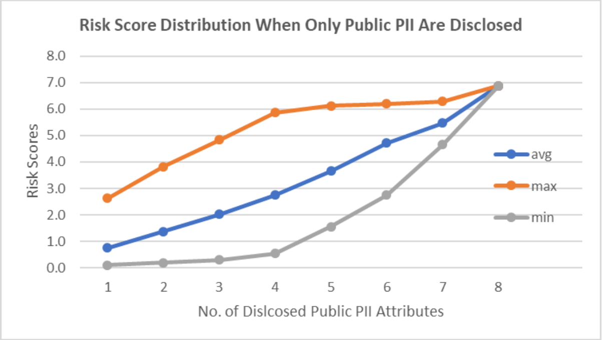

Figure 3: The risk scores by the number of public PII disclosed

Table 5 and Figure 4 show the average, maximum, and minimum of risk scores of scenarios by the number of revealed protected/confidential attributes. Generally, when a person's protected/confidential PII attribute is disclosed with other PII attributes, the person's privacy risk is considered high. If a confidential PII attribute is disclosed alone without any other information, the privacy risk is low. This would explain the minimum risk score is only 1.04 when 1 protected/confidential attribute is disclosed alone. The more protected/confidential PII attributes are disclosed, the higher the privacy risk is.

TABLE 5: THE RISK SCORES BY NUMBER OF PROTECTED/CONFIDENTIAL PII ATTRIBUTES

| # of protected/confidential PII | Avg. Scores | Max. Scores | Min. Scores |
|---|---|---|---|
| 1 | 9.82 | 31.19 | 1.04 |
| 2 | 17.65 | 38.93 | 2.26 |
| 3 | 25.08 | 44.31 | 7.65 |
| 4 | 33.81 | 49.62 | 13.03 |
| 5 | 40.60 | 50.81 | 18.51 |
| 6 | 48.56 | 51.79 | 45.57 |

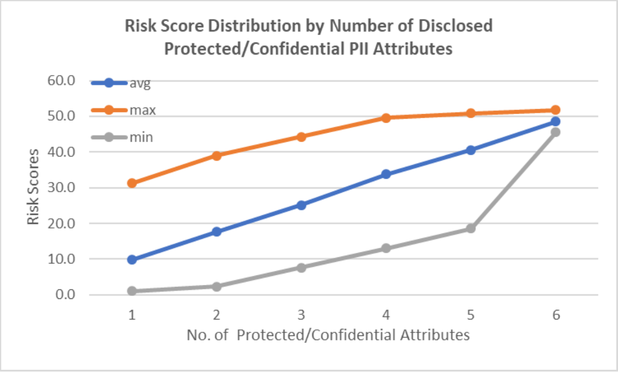

Figure 4: The risk scores by the number of protected/confidential PII disclosed

We used a heuristic approach, knee of a curve, to identify boundaries for partitioning the risk levels. In mathematics, a knee of a curve is a point where the curve visibly bends, and it is often used in optimization to find the optimum point for some decision. Figure 3 shows the knee of the minimum score curve is at 4, and the knee of the maximum score curve are 4 and 7; Figure 4 shows the knee of the minimum score curves are 1 and 5 respectively, the knee of the maximum scores curve is at 4. The average scores of the knee points in Figure 3 are used to categorize the risk scores shown as Table 6:

TABLE 6: CATEGORIZATION OF PRIVACY RISK LEVELS

| **Risk Levels** | **Risk Score Range** |
|---|---|
| Very low | *r* ≤2.74 |
| low | 2.74 <*r*≤5.48 |
| Medium | 5.48 <*r*≤ 6.87 |
| high | 6.87<*r***≤12.25** |
| Very high | *r*>12.25 |

The knees in Figure 4 are not used as the privacy risk is considered high when more than one protected/confidential PII attributes are disclosed. The score range of each risk category is mainly determined by the knees in Figure 3, except the categories of *high* and *very high*, where 12.25 represents the maximum score when one protected PII attribute is disclosed. When a risk score is higher than 12.25, it means more than one protected/confidential PII attributes are revealed, and thus the privacy risk would be very high. Table 7 shows some scenarios and their corresponding risk scores. It shows that the risk score is effective in indicating whether a scenario has a high privacy risk or not.

The proposed personal privacy risk assessment framework provides a strategy for evaluating privacy risk based on the revealed PII. Although the synthetic dataset contains only 14 attributes, the framework can be extended to include more PII attributes. Our future research will explore deep learning methods to leverage the correlations among PII attributes for better risk analysis.

## III. RELATED WORK

Risk assessment has been studied in the security domain where risk is often measured as a function of likelihood and impact of a security threat[13,14]. An important difference between security and privacy risk is that harm to individuals is a primary consideration for privacy risk, whereas it is of secondary importance for security risk[7]. French data protection regulator (CNIL) published a guidance on privacy risk management that assesses risk severity based on the possible prejudice effects and the level of identifiability of data[11]. Wagner suggests a method of quantifying and representing privacy risk that considers a collection of factors as well as a variety of contexts and attacker models. The method decomposes the impact and likelihood of privacy risk into fine-grained components that are measured individually and then combined visually to assess the overall privacy risk of a system[7].

Privacy risk assessment has also been studied from different aspects. Liu and Terzi proposed a method that computes the privacy score in social networks as the product of the sensitivity of profile items and their visibility[12]. Srivastava and Geethakumari used naïve approach to describe and calculate the privacy quotient to measure the privacy of the user's profile on social networks[17].

TABLE 7: RISKS OF SOME SCENARIOS

| name | address | DoB | birth place | cell | email | work phone | fb | tw | ins | DL | passport | credit | SSN | Risk Scores | Risk Level |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.11 | Very low |
| 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.21 | Very low |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2.73 | Low |
| 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 5.43 | Low |
| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 5.88 | Medium |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 6.87 | Medium |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 6.87 | Medium |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 12.05 | High |
| 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 12.02 | High |
| 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 13.67 | Very High |
| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 44.7 | Very High |

The OWASP top-10 list of privacy risks in web applications ranks privacy risks by their ratings for impact and likelihood[1]. Ahamdian *et al.* proposed a model-based privacy analysis to support privacy impact assessment in the early phases of information system development[6]. Bhattacharjee *et al.* proposed a technique for identifying the components of a business process and quantitatively assessing their security risks[8]. Zaeem *et al*. present the mathematical representation and implementation of a model of Personally Identifiable Information attributes for people, so called named Identity Ecosystem. The Identity Ecosystem is utilized to predict and to explain the risk of losing PII and the liability associated with fraudulent uses of these PII attributes[9,18].

Most existing research studied privacy risk from the business perspectives. This research tries to study the privacy risk from a private person's point of view.

## IV. Conclusion and Future Work

Guarding PII is important to ensure the integrity of an individual's identity and to protect people from falling into a victim of identity theft. Strict privacy laws and regulations push data providers to be more diligent in avoiding data breaches and privacy violations in their practices. Individual users become to a weak link in privacy protection. The paper proposed a personal privacy risk assessment framework based on the disclosed PII. The framework allows individuals to assess their privacy risks so that they can be more active in protecting their privacy. It's hoped that the risk score would help raise the user's awareness about privacy protection, and allow them to proactively control the future release/sharing of information to keep the risk low.

Our future research will work on developing a user-oriented privacy guard tool that allows users to monitor the dissemination of their information on Internet and compute their privacy risk scores. Extracting PII from Internet is a challenge task which involves natural language processing, entity resolution and identification, and entity disambiguation. We will investigate advanced natural language processing and deep learning techniques for PII retrieval and entity disambiguation. In addition, deep learning techniques will be studied to improve the risk calculation and categorization.

## Acknowledgment

Part of this research work is sponsored by NSF EPSCoR grant No. OIa-1946391

## References

[1] Lily Wachtor, "Are More People Public or Private on Social Media?" Viasat blog, May 27, 2020.

[2] L. E. Heffernan, "Oversharing: Why Do We Do It and How Do We Stop?". https://www.huffpost.com/entry/oversharing-why-do-we-do-it-and-how-do-we-stop_b_4378997

[3] "Oversharing and social Media." https://apps.illinoisworknet.com/ArticleViewer/Article/Index/257/%7Blink%7D

[4] J. Buzzard, "2022 Identity Fraud Study: The Virtual Battleground", Javelin Stategy.

[5] Dave Moore, "Once on the Internet, always on the Internet", The Norman Transcript.

[6] A. S. Ahmadian, D. Strueber, V. Riediger, and J. Juerjens, "Supporting Privacy Impact Assessment by Model-Based Privacy Analysis," Proceedings of the 33rd Annual ACM Symposium on Applied Computing, April 2018, pages 1467–1474.

[7] I.Wagner, E. Boiten, "Privacy Risk Assessment: From Art to Science, by Metrics." In: Garcia-Alfaro, J., Herrera-Joancomartí, J., Livraga, G., Rios, R. (eds) Data Privacy Management, Cryptocurrencies and Blockchain Technology.

[8] J. Bhattacharjee , A. Sengupta, and C. Mazumdar, " A Quantitative Methodology for Security Risk Assessment of Enterprise Business Processes," the 2nd International Conference on Information Systems Security and Privacy (ICISSP 2016), pages 388-399.

[9] R. N. Zaeem, S. Budalakoti, K. S. Barber, M. Rasheed and C. Bajaj, "Predicting and explaining identity risk, exposure and cost using the ecosystem of identity attributes," *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, 2016, pp. 1-8.

[10] F. Stahl and S. Burgmair, "OWASP Top 10 Privacy Risks Project (2017)," https://www.owasp.org/index.php/OWASP_Top_10_Privacy_Risks_Project

[11] Commission Nationale de l'Informatique et des Libert_es: Privacy impact assessment, (PIA) 1: Methodology (2018)

[12] K. Liu and E. Terzi, "A Framework for Computing the Privacy Scores of Users in Online Social Networks," ACM Trans. Knowl. Discov. Data 5(1), 6:1{6:30 (Dec 2010)

[13] National Institute of Standards and Technology (NIST): Guide for Conducting Risk Assessments. NIST special publication 800-30 r1 (Sep 2012)

[14] Open Web Application Security Project: OWASP Risk Rating Methodology (2018), https://www.owasp.org/index.php/OWASP_Risk_Rating_Methodology

[15] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen. "A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements," *Requirements Engineering* 16 (2011).

[16] M. C. Oetzel and S. Spiekermann, "A systematic methodology for privacy impact assessments: a design science approach," *European Journal of Information Systems* (2014), 126--150.

[17] A. Srivastava and G. Geethakumari, "Measuring privacy leaks in Online Social Networks," *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2013, pp. 2095-2100.

[18] Kai Chih Chang, Razieh Nokhbeh Zaeem, K. Suzanne Barber, "A Framework for Estimating Privacy Risk Scores of Mobile Apps", *Information Security*, vol.12472, pp.217, 2020