Blog : TheKeepersOfData PrivacyReview

Introduction

Online Social Media networks are expanding exponentially and have become an essential part of our lives. Although the services provided by social media platforms add great convenience in our life this comes with a few inherent costs and problems. Social media allows users to share a part of their life with the world, but it also raises the security threats to their personal information. This, in turn, affects the personal privacy of an individual as a whole.

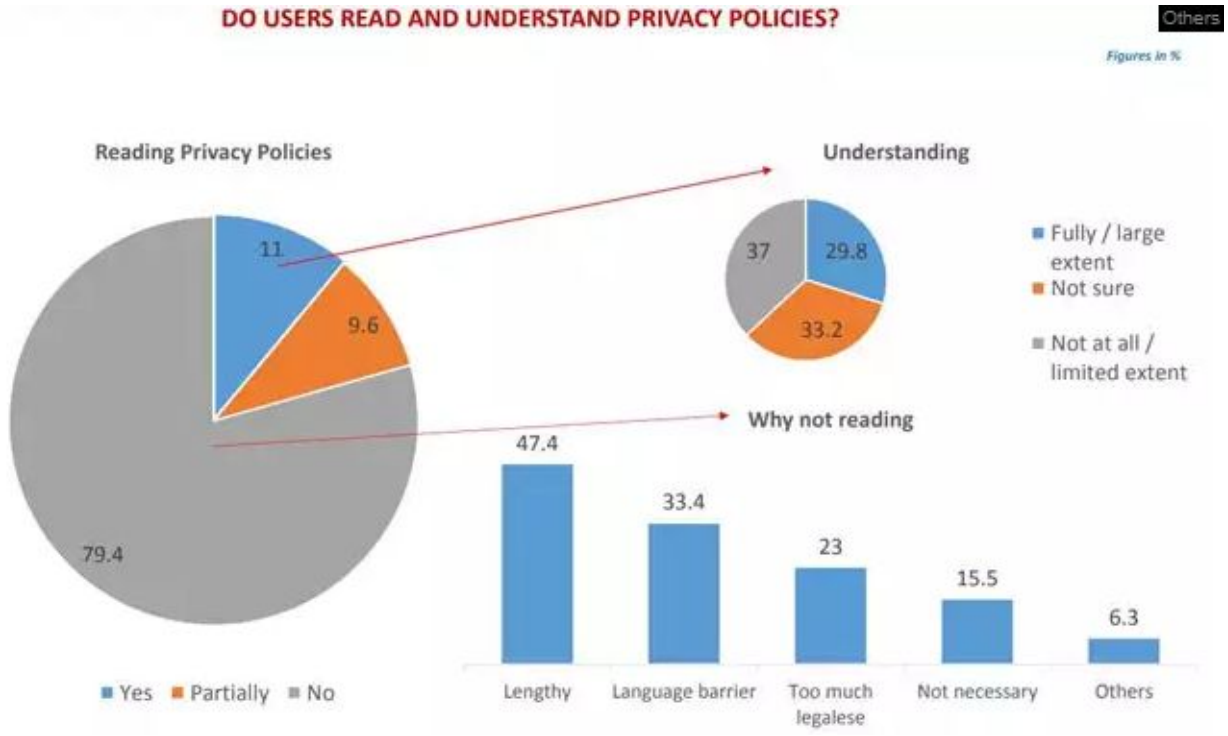
A significant portion of the world's population uses social media platforms. One of the most popular social media networks in the world, "Facebook", claims to have about 2.5 billion monthly active users. The user base consists of a wide proportion of population ranging from children and teenagers to the elderly. Social media has turned into an extension of their offline interactions

and is used not just as a way to stay connected but as a creative outlet for those brimming with talent.

Data has become a new currency in today's world, and the insights that the data can provide the new gold. Most of us won't like to have our data collected, but Social media makes it easier. Smart businesses usually take advantage of this data to better understand the behavior of individuals, groups, and society, terming it an analysis of consumer attitude: '*market segmentation and identification of market trends'*.

Motivation

Despite the ever-increasing usage of digital technology, it was astonishing to know that only 11 percent of the users read the privacy policies, revealed in a survey by CUTS International. The major findings of the survey are depicted in the below chart -



It precisely describes the major factors that prevent the user to read the privacy policies, viz. -

- Length of privacy policies, that make it cumbersome for users to read.
- The language barrier.
- Too many legal terms which the normal users skip or pay less heed to, because it becomes difficult for them to comprehend the legalities of the privacy policies.

These points motivate us to phrase the problem and find a solution, that takes into account all three reasons, forbidding a user to look at the privacy policies of an app before opting for it.

Problem

The extent to which users can access user profiles has become a recent topic of ethical consideration as anything we publicly share will be available in the public domain. Tools such as Social media crawlers/scrapers can be used to automatically collect data from social media and other websites in the public domain. There are various examples that show how the data of social media users are being breached. In 2019, the personal data of 267 million Facebook users were exposed online.

But it is not an all “Blame the website game”. We opt to give out our contact information and personal details to sign up, Similarly, our activity and preferences are tracked and our purchase patterns are stored. The location is logged continuously and can be used to find patterns in our routines. We knowingly or unknowingly give out information to these social media websites and their 3rd party connections and end up being kept in the dark about the information that is being shared.

Solution

Now we might have all skipped the Terms and Conditions and Privacy Policy pages for the better part of our lives and not realize what a grave mistake that is. Privacy Review is a service that aims at providing the end-user with complete details of what he signs up for in a meaningful manner. Our tool provides users with a **Credibility Score** (CS) (Range: 0 - 5) of top social media services. This score is a testimony to inform the users of how credible the social media app is in terms of its privacy policy. The higher the Credibility score, the more reliable the social media app is. As simple as that!

We took the painstaking process of reading the multi-page legal jargon called a privacy policy and summed it down to a score that everyone can relate to. For all the people who need additional info, a dedicated page with a detailed review is also present for scrutiny. Now with just a single click, users can see what kind of personal information the social media platform collects and how they plan on using it. Our application will rate social media platforms on a scale of 1 to 5. One representing that the platform collects the users' personal information and using it for various other activities, while a score of 5 representing what we call a Great privacy standard maintenance.

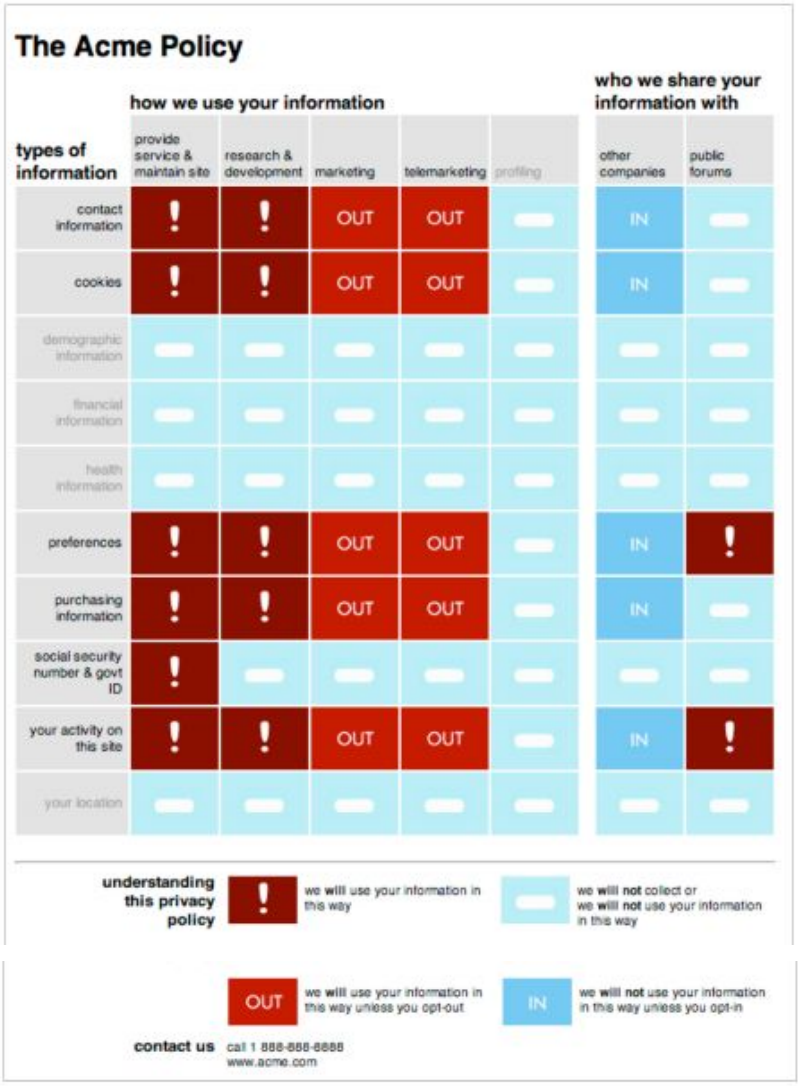
Initial Research

We analyzed the iterative design process mentioned in the paper **A "Nutrition Label" for Privacy** that developed a privacy label that presents to consumers the ways organizations collect, use, and share our personal information. Various surveys have shown that consumers are very concerned about online privacy. But there haven't been any substantial or viable solutions to get rid of the unnecessary data collection by OSM platforms.

There exists a major gap in the communication and understanding of privacy policies for the users. We are devising a bridge to this gap, by creating an information design that improves the visual presentation and comprehensibility of privacy policies. When compared to existing natural language privacy policies, the proposed privacy label allows participants to find relevant information more quickly and accurately and provides a more enjoyable information-seeking experience.

We collected all the information about the privacy policies of the top 50 social media platforms, according to the "proposed Privacy Nutrition Label". We manually read the 'privacy policies' and 'terms of service' of these apps and divided all the data that is being collected in the format given below The link of the google drive is here. The format also integrates the various uses an OSM platform collects our data for, with a four scaled option on our choice which is as follows:-

- ! – We will use your information in this way
- OUT – we will use your information in this way "unless you opt-out"
- -- – we will not collect or use your information in this way
- IN – we will not use your information in this way "unless you opt-in"



A bold title is used to set the context for the information.

Short labels are used for column and row headers, with longer definitions on our Useful Terms page.

Information that is not collected has a saturated label and a row full of the lightest symbol.

Four symbols on a scale from light to dark are used to indicate the severity of certain privacy practices.

Row and column locations are consistent so that two policies side-by-side can be easily visually compared.

A legend provides information about what each symbol means.

Figure 5. Our *proposed Privacy Nutrition Label.* This label is the one we tested in the second focus group and the laboratory study.

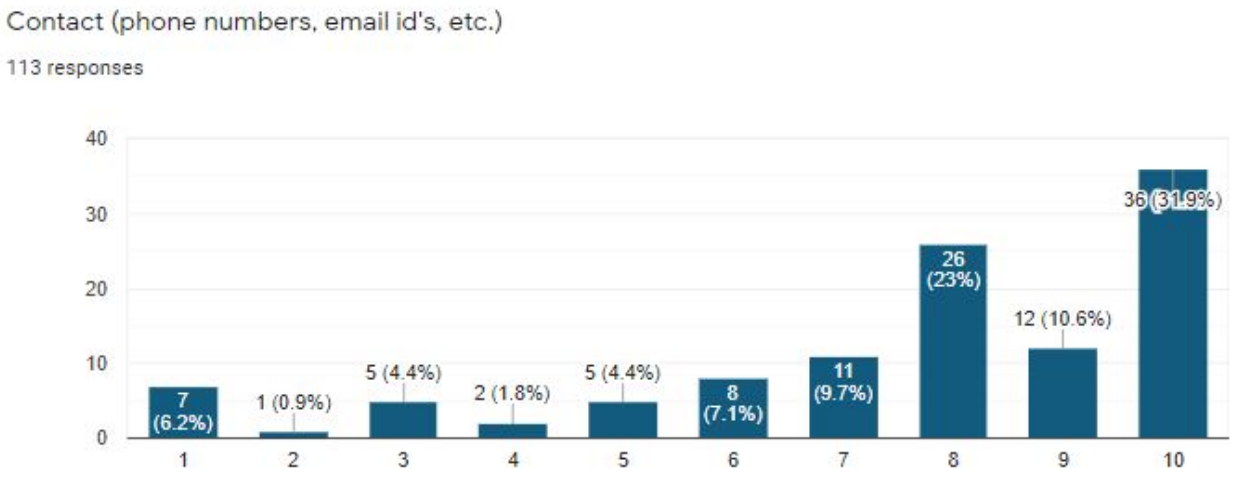
In bringing the privacy policies in this format we overcome the “Too many legal terms which the normal users skip or pay less heed to because it becomes difficult for them to comprehend the legalities of the privacy policies“ problem which users face.

Now our focus was to target the length of the privacy policies for each social media app. So we tried to convert the ‘terms of service’ and ‘privacy policies’ into a 1d format the link is given here-

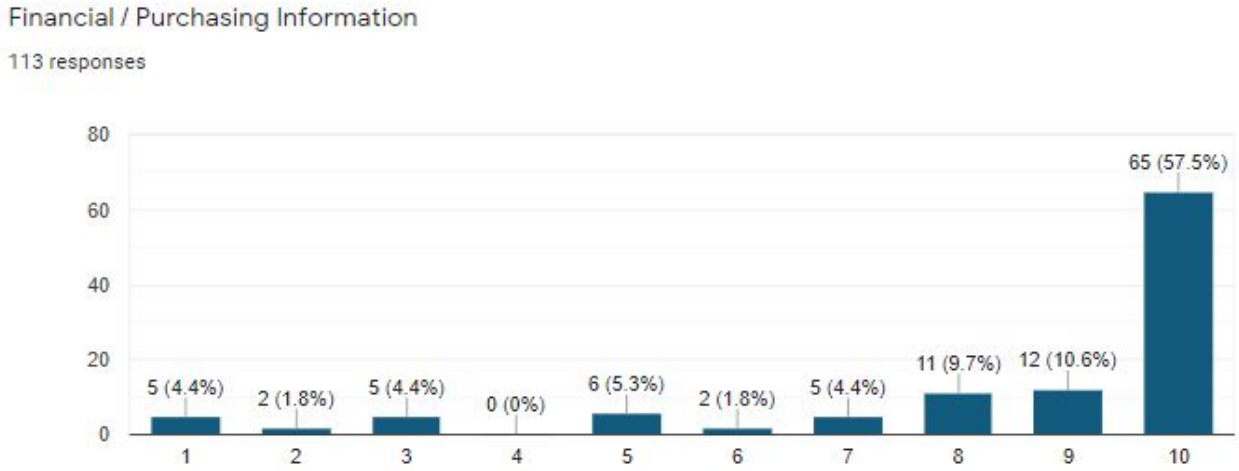
To remove the language barrier we thought of bringing the long texts in a numerical format. To accomplish this, we sought after the main highlights and commonalities among all the apps. Following this, we put out a survey, questioning about those specific highlights. This helped us consolidate the users’ needs and preferences in a graphical format, which became a major catalyst in coming up with a viable solution for this grave data security predicament.

Survey

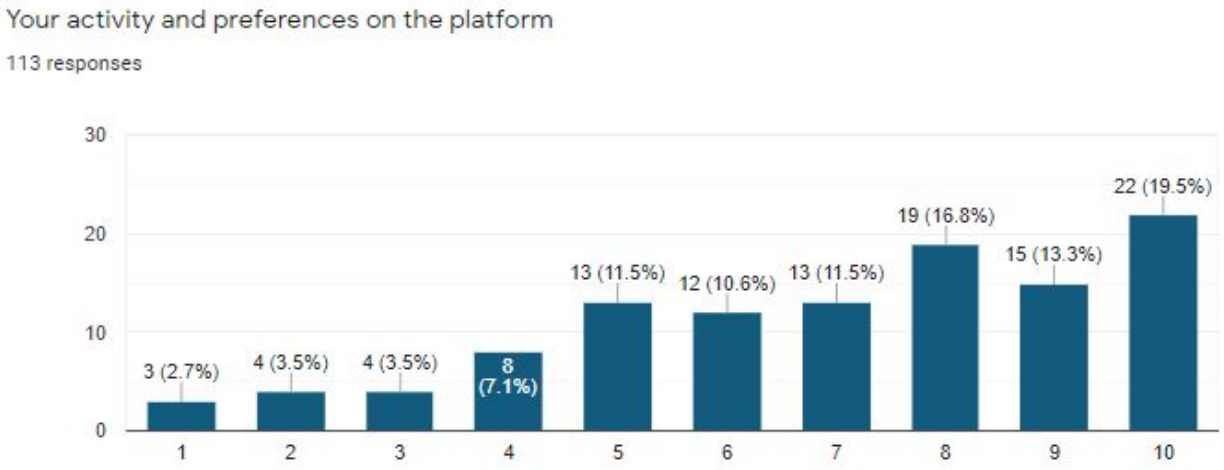
To further improve the visual presentation and comprehensibility of privacy policies, we surveyed to get some insights into what people think about their privacy preferences on social media. The results helped us to come up with metrics that show us the credibility of the social media application.



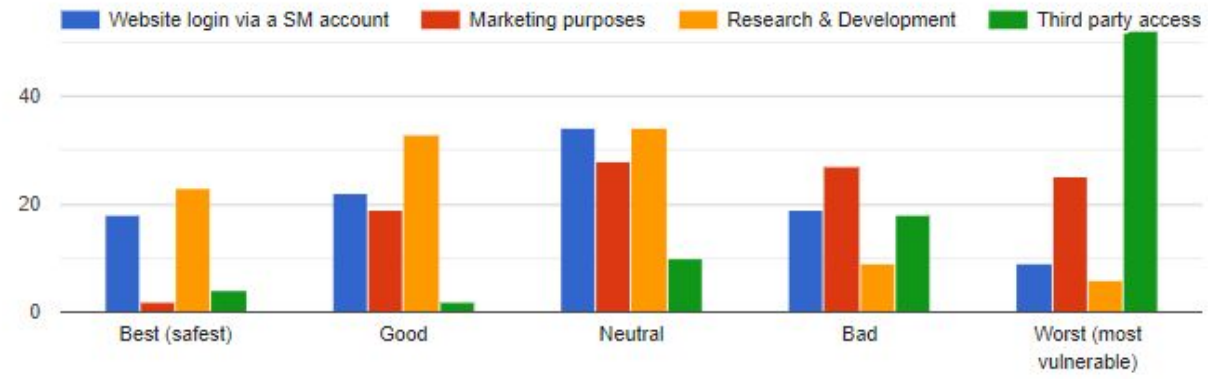
This graph depicts that the majority of the people who participated in the survey believe that breaching of contact information is highly intolerable.



This graph depicts that the breaching of financial information is highly intolerable for 65% of the people that participated in the survey.



The breaching of users' activity and preference information is intolerable.



This graph shows that people consider the sharing of their data with third-party organizations is very dangerous, whereas sharing their data for research and development is regarded as the safest in terms of the possibility of data breaching.

Formulation of Credibility Score

We gained insights from the data visualization and used it to form a new formula that will generate the score. The higher the Credibility Score, the more the social media app is trustable for not performing any kind of data breaching. On a scale of 1 to 5, our platform gives the Credibility Score to the social media applications where 0 is the worst and 5 is the best.

We assigned absolute integer values to the variables that project how the social media authorities use our personal information.

The Methodology of Calculating The Credibility Score:

1. From the Privacy Nutrition label model one could easily guess the areas where social media authorities can use our data(columns of the table). Now to find the amount of influence that each of these has, we looked at our survey results and figured out the appropriate weights as follows:

Areas where social media authorities can use our data	**Weightage**
Research and Development	10%
Other companies / third party app login	40%
Marketing	30%
Provide service and maintain the site	20%

The above weighted average values were crucial in developing the Uncertainty Score which is the next step.

The next step was to calculate the weighted average of all the rows in the Privacy Nutrition label model. For this, as above, the survey results were analyzed to get the following weights:

Types of information	**Weightage**
Purchasing information	40%
Location information	25%
Contact information	20%

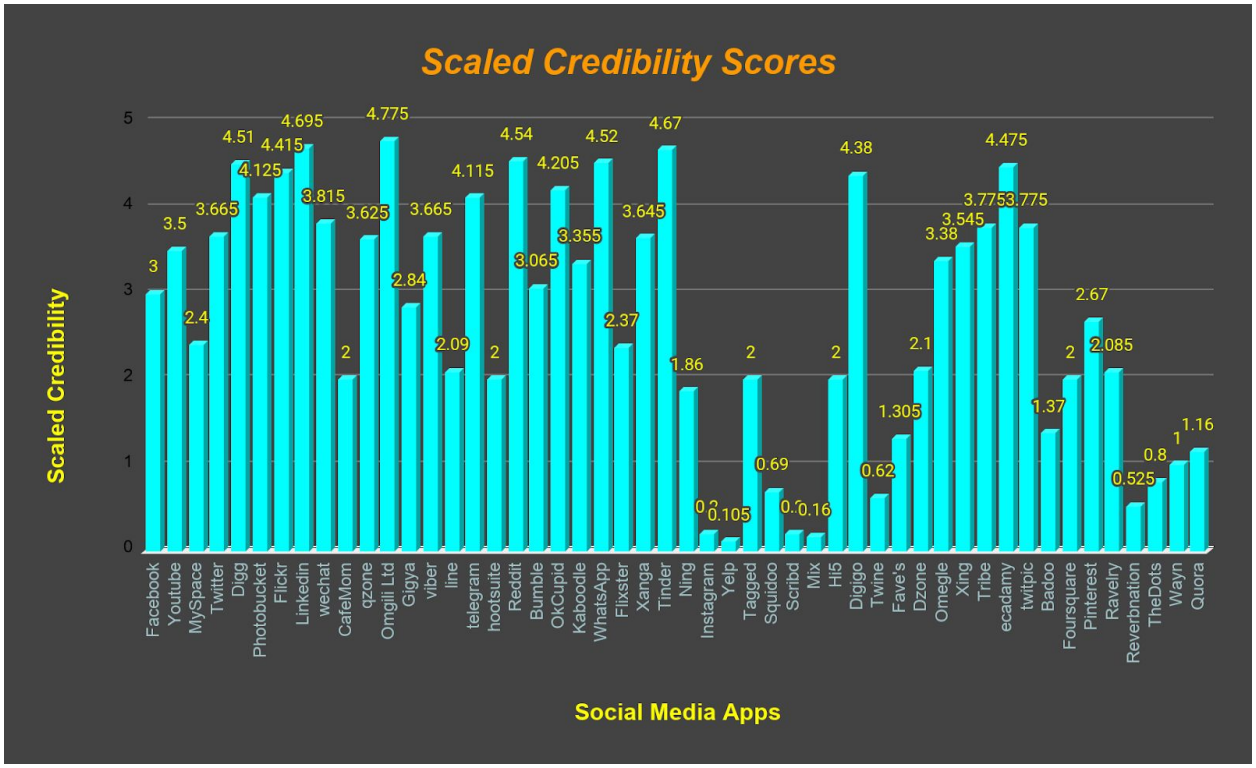
Users' activity	10%
Demographic information	3%
Cookies	2%

2. Now we have all prerequisites needed for calculating our uncertainty score, which is given by the formula:

FalseHood Score: (0.4*(purchasing information)+ 0.25*(location information) + 0.20*(contact information)+ 0.10*(users' activity) + 0.03*(demographic information) + 0.02*(cookies))/50

3. Now our final Credibility Score can be calculated by the simple formula:

Credibility Score: 5 * (1 - Falsehood Score)



In the above graph we can clearly notice how the apps that we use on a daily basis don't have a score greater than 4. This is because all of those apps are very data-hungry and our privacy is at stake while we use these apps. Moreover, the apps we lay our whole trust in and are our *'bread n butter'* when we come to online content consumption, are not all that safe.

Conclusion

The project was intriguing and fun at the same time. We got a better understanding of the apps that we use in our day to day lives, and how unconscious we are on how that app is using our data. We got a more holistic view of how companies tweak their 'privacy policies' and 'terms of service' to make them look secure to the general public. Consequently, this project helped increase our awareness, in addition to helping us build a new path to view the data privacy predicaments encountered in the usage of various applications, through a completely new lens.

Team Members

- Akansha Gautam, 2016221
- Akshat Joshi, MT18064
- Aditya Gupta, 2017325
- Vyshakh, 2016120
- Anushika Verma, 2016015
- Ankur Rangi, 2017331